Fathers´ smoking in different time windows as related to offspring's epigenetic and phenotypic outcomes

Gerd Toril Mørkve Knudsen

Thesis for the degree of Philosophiae Doctor (PhD)
University of Bergen, Norway
2023

UNIVERSITY OF BERGEN

Fathers´ smoking in different time windows as related to offspring's epigenetic and phenotypic outcomes

Gerd Toril Mørkve Knudsen

Thesis for the degree of Philosophiae Doctor (PhD)
at the University of Bergen

Date of defense: 17.02.2023

Year: 2023

Title: Fathers´ smoking in different time windows as related to offspring's epigenetic and phenotypic outcomes

Name: Gerd Toril Mørkve Knudsen

Print: Skipnes Kommunikasjon / University of Bergen

Scientific environment

The work presented in this thesis was performed at the Department of Global Public Health and Primary Care at the University of Bergen (UIB) and at the Department of Occupational Medicine at Haukeland University Hospital, Bergen. The PhD grant was provided by the University of Bergen. The analysis of biological material was conducted at the University of Southampton and the project has involved scientific collaboration with colleagues in Europe and Australia who are part of the researcher networks of the population-based surveys: European Community Respiratory Health Survey (ECRHS), Respiratory Health In Northern Europe study (RHINE) as well as the Respiratory Health In Northern Europe, Spain and Australia study (RHINESSA).

Supervisor

Professor Cecilie Svanes, Centre for International Health, Department of Global Public Health and Primary Care, University of Bergen, Norway;
Department of occupational Medicine, Haukeland University Hospital, Bergen, Norway.

Co-Supervisors

Professor Ane Johannessen, Centre for International Health, Department of Global Public Health and Primary Care, University of Bergen, Norway.

Svein-Magne Skulstad, MD PhD, Department of Occupational Medicine, Haukeland University Hospital, Bergen, Norway

Professor John W Holloway, Human Development and Health, Faculty of Medicine, University of Southampton, UK, NIHR Southampton Biomedical Research Center, University Hospitals Southampton, UK.

Acknowledgements

Throughout the course of this PhD period, I have been faced with life events that have disrupted the expected progress of my work. I am deeply indebted to the University of Bergen for assisting me above and beyond expectations during these years, and to the department of Occupational Medicine for providing me with a comfortable workplace in a congenial environment. The suggested workplan has enabled me to continue with my PhD project.

The same holds true for my supervisor Cecilie Svanes. I am extremely grateful for your unwavering and continuous support over the years. Your ceaseless belief in my abilities and patient guidance have been indispensable for finalising this project. I will always be thankful for how you have nurtured my scientific curiosity and given me the freedom to pursue my own ideas yet providing me with a sense of direction and overview that has been important in seeing this project through. Thank you for all the enthusiastic discussions, and that you so generously have shared your knowledge and experience. Surely, it has been priceless for my understanding of epidemiological and cross-generational research. Your scientific capacity and aspirations are truly inspiring.

My sincerest thanks must also be given to my co-supervisors Ane Johannessen and Svein Magne Skulstad. Ane, your skills in statistical and epidemiological methodologies and data management have been crucial throughout my PhD training. Thank you for providing me with solid data sets, and for helping with all kinds of statistical bottle necks, always with a smile on your face and with a comforting word. Your talent for scientific writing and constructive advice has truly improved my scientific skills. Moreover, your emotional support, both at work and in life itself, has helped me at times where I have faltered. Svein Magne, your interest in the biological roles of epigenetic systems has truly been catching, and of great help to my work. Thank you for all the conversations and your valuable feedback and comments throughout the years.

Furthermore, I am heavily indebted to my co-supervisor John Holloway, who with his extensive background and knowledge within the field of epigenetics and bioinformatics methodologies has been indispensable for this PhD work. I am grateful for all the help you have provided, and for your suggested directions and focus which has aided and for sure improved my understanding of epigenetic epidemiological research. I am also grateful for the research stay in your lab in Southampton, where I got the opportunity to sit down with and learn from my co-author fellow and bioinformatician Faisal Ibne Rezwan. He deserves a warm thanks for patiently introducing me to the analytical concepts of epigenome-wide association studies, and for assisting in interpretation of results. I am also grateful that I got the opportunity to collaborate and be co-authoring together with Negusse Kibata. I have truly been impressed by your great work capacity, and your bioinformatic knowledge. It would be a privilege if I ever got the chance to cooperate with you again.

I owe a special thanks to my former and initial supervisors Roy Miodini Nilsen and Shyamali Dharmage for investing time and effort in my research. Roy, your statistical guidance was for sure an eye-opener and woke my interest for statistical computing. Shyamali, your commended and constructive responses and scientific thoughts have been most helpful. I am also grateful for the inspiring and fruitful scientific inputs I have been given by Francisco Gomez Real and Randi Bertelsen throughout these years. The same should be said to all co-authors and international colleagues in the research consortium of ECRHS, RHINE and RHINESSA. I have sincerely appreciated your contributions to the work presented in this thesis.

Moreover, there are many former PhD-colleagues I want to thank, and that I have enjoyed the company of throughout this period. Marianne Lønnebotn, Ingrid Nordeide Kuiper, Hilde Kristin Vindenes, Marie Johanne Wåtevik, Kai Triebner, Oskar Jogi, Eirunn Waatevik Saure and Trude Duelien Skorge. These years have truly been more fun with you there. Thanks for all the good memories and laughs. Marianne and Ingrid- I will always cherish your friendship- and I hope we will be able to meet up soon. Also thank you Vilde Svanes Sørbye and Benedikte Svanes for your open

hearted natures and for always making people feel included. Furthermore, I am deeply grateful for all former and present colleagues and fieldworkers I have had the privilege to meet during my time in the RHINESSA research group. Thanks for all the inspiring meetings and enthusiastic discussions, not at least, for all the fun we have had on trips and social events.

I owe a heartfelt thanks to friends outside the scientific community as well. Nothing in life happens in isolation – neither does this thesis. Åse, Sidsel, Kristin, Tonje, Maggi and Anne Line. You have cheered me on all these years, and life is for sure better with you in it. Thanks to new friends and neighbours for social gatherings and truly positive distractions. Rita, "joggegruppe med gode intensjoner" has certainly been a soundness of mind and helped me to stay in good spirits- even at times with writer's block.

Finally, and most importantly, I am truly grateful for the support and love of my family. Liv Marit and Svein Harald, my akin, when we meet, you certainly make my day better. Mum and dad, you have always encouraged me to go my own ways and to pursue my own aspirations, yet your everlasting support has been invaluable for seeing this through. Dad, you have been my solid rock during the ups and downs throughout these years. Thank you for being there, for reading endless notes and manuscripts, patiently listen to all my concerns, and for backing me at times I have lacked the confidence to go on. I simply couldn't have done this without you. Petter, thank you for being my sanctuary. You have encouraged me to stay in this to the very end. At last, my beloved children, thank you for having filled my life with love and reason. In your own peculiar ways, you remind me to live and embrace all the opportunities and second chances that life offers.

Contents

Abbreviations

3'UTR	3 prime untranslated region
5'UTR	5 prime untranslated region
A	Adenine
ACSF3	Acyl-CoA Synthetase Family Member 3
AGO	Argonaut protein
AHRR	Aryl Hydrocarbon Receptor Repressor
ALSPAC	Avon Longitudinal Study of Parents and Children
ATG	Methionine
ATP6V1E1	A TPase H + Transporting V1 Subunit E1
BCAS1	Brain Enriched Myelin Associated Protein 1
BIA	Bioelectrical Impedance Analysis
BMI	Body mass index
BMIQ	Beta mixture quantile dilatation
bp	Base pair
BPA	Bisphenol A
BRINP3	BMP/ Retinoic Acid Inducible Neural Specific 3
C	Cytosine
C2	Complement 2
C2orf39	Chromorome 2 Open Reading Frame 39
CCNYL1	Cyclin Y Like 1
cd	Coding sequence
CDK3	Cyclin Dependent Kinase 3
CDO1	Cysteine Dioxygenase Type 1
CENPP	Centromere Protein P
CHARGE	Cohorts for Heart and Aging Research in Genomic Epidemiology
Chr	Chromosome
CI	Confidence interval
CNTNAP2	Contactin Associated Protein 2
CpG	Cytosine-phosphate Guanine di-nucleotide
CSF1R	Colony Stimulating Factor 1 Receptor
CTNNA2	Catenin Alpha 2
CYP1A1	Cytochrome P450 Family 1 Subfamily A member 1
DAG	Directed Acyclic Graph
DLGAP1	DLG Associated Protein 1
dmCpG	Differentially methylated CpG
DMP	Differentially methylated position
DMR	Differentially methylated region
DNA	Deoxyribonucleic acid
DNAJC14	DNAJ Heat Shock Protein Family (Hsp40) Member C14
DNMT	DNA methyltransferase
DNMT1	DNA methyltransferase 1

DNMT3A	DNA methyltransferase 3α
DNMT3B	DNA methyltransferase 3 β
DNMT3L	DNA methyltransferase 3-like
DNTT	DNA Nucleotidylexotransferase
DoHaD	Developmental Origins of Health and Disease
ECRHS	European Community Respiratory Health Survey
ENCODE	Encyclopedia of DNA Elements
EPIC	Infinium Methylation EPIC BeadChip array
EWAS	Epigenome-wide association study
F2RL3	F2R Like Thrombin Or Trypsin Receptor 3
FAM53B	Family With Sequence Similarity 53 Member B
FDR	False Discovery Rate
FFM	Fat free mass
FMI	Fat mass index
FREM2	FRAS1 Related Extracellular Matrix 2
FSH	Follicle-Stimulating Hormone
FWER	Family-wise-error rate
G	Guanine
GFI1	Growth Factor Independent Transcriptional Repressor
GNG12	G Protein Subunit Gamma 12
GnRH	Gonadotropin-releasing hormone
GO	Gene Ontology
GPR15	G Protein-Coupled Receptor 15
GRAMD4	GRAM Domain Containing 4
HAT	Histone acetyltransferase
HC1	Heteroskedasticity-consistent estimator 1
HDAC	Histone deacetylase
HMT	Histone methyltransferase
HPG axis	Hypothalamic-Pituitary-Gonadal axis
i.e.	*Id est*/ that is
IAP	Intracisternal A-particle
IgE	Immunoglobulin E
IOW	Isle of Wight
IOWBC	Isle of Wight birth cohort
IRS1	Insulin Receptor Substrate 1
KCNJ1	Potassium Inwardly Rectifying Channel Subfamily J Member 1
KEGG	Kyoto Encyclopedia of Genes and Genomes
KLHDC1	Kelch Domain Containing 1
LEPROTL1	Leptin Receptor Overlapping Transcript Like 1
lncRNA	Long non-coding RNA
mb	Megabase
MBIP	MAP3K12 Binding Inhibitory Protein 1
MFGE8	Milk Fat Globule EGF And Factor V/VIII Domain Containing

MGAT3	Beta-1,4-Mannosyl-Glycoprotein 4-Beta-N-Acetylglucosaminyltransferase
miRNA	MicroRNA
mRNA	Messenger RNA
MTSS1	MTSS I-Bar Domain Containing 1
MYADML2	Myeloid Associated Differentiation Marker Like 2
MYO1G	Myosin IG
nc	Non-coding
NCAPG2	Non-SMC Codesin II Complex Subunit G2
ncRNA	Non-coding RNA
NIH	National Institutes of Health
nt	Nucleotide
NTRK2	Neurotrophic Receptor Tyrosine Kinase 2
PACE	The Pregnancy and Childhood Epigenetic Consortium
PCR	Polymerase chain reaction
PEBP4	Phosphatidylethanolamine Binding Protein 4
PGC	Primordial germ cell
PHF12	PHD Finger Protein 12
piRNA	Piwi-interacting RNA
PRAP1	Proline Rich Acidic Protein 1
PRR5	Proline Rich 5
PSTPIP2	Proline-Serine Threonine Phosphatase Interacting Protein 2
RALB	Ras Like Proto-Oncogene B
RARA	Retinoic Acid Receptor Alpha
RGS9	Regulator Of G Protein Signaling 9
RHINE	Respiratory Health in Northern Europe
RHINESSA	Respiratory Health in Northern Europe, Spain and Australia
RISK	RNA-induced silencing complex
RNA	Ribonucleic acid
ROS	Reactive Oxygen Species
RPC	Robust partial correlation
rRNA	Ribosomal small RNA
SD	Standard deviation
SE	Standard error
siRNA	Short-interfering RNA
slk correction	Stouffer-Liptak-Kechris correction
SNED1	Sushi, Nidogen And EGF Like Domains 1
SNP	Single nucleotide polymorphism
sRNA	Small RNA
T	Thymine
TBW	Total body water
TET	Ten-eleven translocation enzymes
TF ChIP-seq	Transcription factor chromatin immunoprecipitation sequencing

TLR9	Toll Like Receptor 9
TMEM51	Transmembrane Protein 51
TPCN1	Two Pore segment Channel 1
TRAF3IP2-AS1	TRAF3IPS Antisense RNA 1
tRFs	tRNA fragments
TRIM2	Tripartite Motif Containing 2
tRNA	Transfer RNA
TSHR	Thyroid Stimulating Hormone Receptor
tsRNA	Transfer derived sRNA
TSS	Transcription start site
UCSC	The University of California Santa Cruz
UNC93B1	UNC-93 Homolog B1, TLR Signaling Regulator
VTI1A	Vesicle Transport Through Interaction With T-SNAREs 1A
WDR60	WD Repeat-Containing Protein 60
ZNF689	Zinc Finger Protein 689

Thesis at a glance

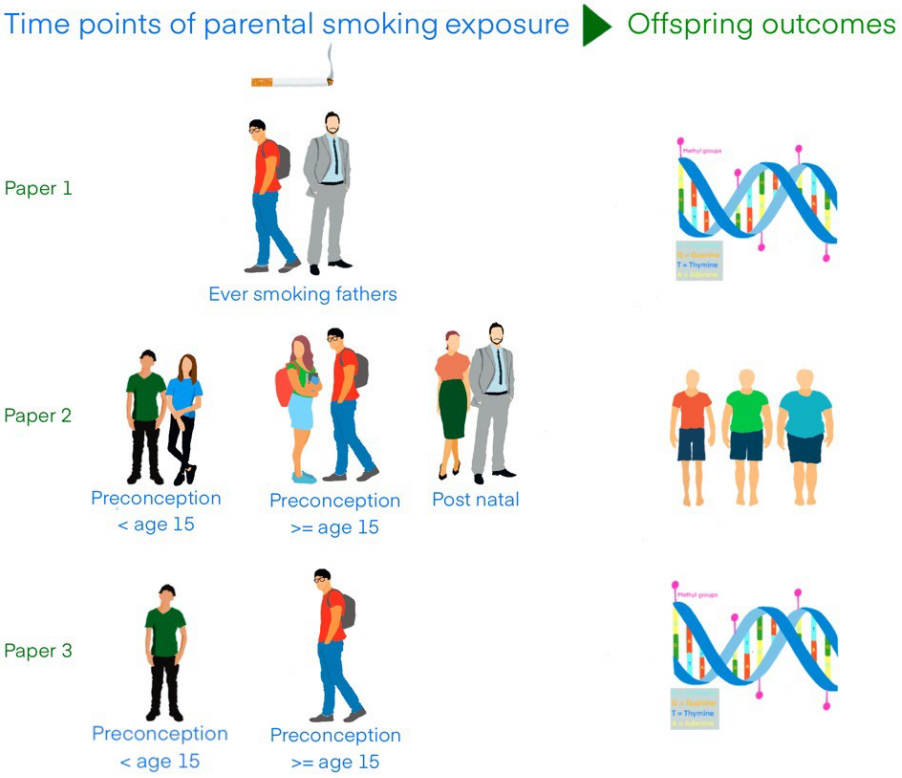


Figure 1: Overview of the three papers included in the thesis: Paper I and III: Epigenome-wide association studies investigating offspring's DNA methylation patterns in relation to fathers' ever smoking, and according to time points of smoking commencing in preconception and early adolescent years. Paper II: Analyses of sex-specific associations between parents' smoking onset at three preconception and postnatal time points and adult offspring's body mass index and fat mass; with additional mediation analyses exploring whether the observed associations are mediated by parental smoking intensity and adult BMI status, or the offspring's own smoking behaviour. Illustration by G. Toril Mørkve Knudsen.

Sammendrag

Bakgrunn: Epidemiologiske studier tyder på at fars røyking, særlig hvis han begynner å røyke når han er svært ung (<15 år), kan påvirke overvekt og lungehelse hos fremtidige barn. Stoffene i tobakksrøyk kan føre til endring i det epigenetiske apparatet i sædcellene, og dyrestudier viser mekanismer for hvordan endring i epigenetisk status kan overføres over generasjoner. Epigenetiske mekanismer kan tenkes å forklare hvordan fars røykevaner lenge før konsepsjon kan påvirke barns tidliglivsutvikling og videre helse. Likevel har studier av røykerelaterte DNA-metyleringsendringer fram til nå stort sett fokusert på egen røyking og mors røyking, ikke røyking hos far. Det er også svært få studier som har undersøkt fars røyking i spesifikke tidsperioder før konsepsjon og etter fødselen. Gitt sårbarhetsfaser i utvikling av sædcellene, er det biologisk grunnlag for å stille spørsmål om tenår og tidlig ungdomstid utgjør en kritisk eksponerings-sensitiv periode for at røyking kan påvirke neste generasjons helse.

Formål: I) Å identifisere potensielle DNA-metyleringssignaler hos (voksne) barn assosiert med fars røykevaner. II) Å undersøke tidspunkt for foreldres røykestart prekonsepsjon og postnatalt i relasjon til deres barns kroppsmasseindeks og fettmasse, og å undersøke om assosiasjoner er modifisert av barnas kjønn, eller mediert av ulike faktorer hos foreldre og barn (foreldres BMI og røyking pakkeår; barnas egen røyking og fødselsvekt). III) Å identifisere DNA-metyleringssignaler hos (voksne) sønner og døtre relatert til fars røykestart før konsepsjon og i tidlig ungdomsalder, å undersøke om de metyleringssignalene man finner er forskjellige fra signaler assosiert med personlig røyking og mødres røyking, og å studere om noen av de identifiserte metyleringssignalene også er relatert til kroppsmasseindeks og lungehelse hos barna.

Materiale og metoder: Vi utførte epigenomvide assosiasjonsstudier (EWAS) for å undersøke DNA-metyleringsmønstre hos (voksne) barn i forhold til fars eksponering for røyking. I artikkel I studerer vi 195 barn (11-54 år) som deltok i Bergen RHINESSA eller ECRHS. I artikkel III studerer vi barn (7-50 år) fra 6 RHINESSA studiesentre, med data for fars røyking fra ECRHS; vi analyserer spesifikt fars røyking som startet før konsepsjon (N=875; kjønnsspesifikke strata med 457 sønner og 418 døtre) eller før 15 års alder (N=304). I begge artikler ble barnas DNA-metyleringstatus kvantifisert i perifert blod og ved bruk av Illumina Infinium MethylationEPIC teknologimatriser. Vi brukte Comp-p (artikkel I) og dmrff or DMRcate (artikkel III) for å søke etter

differensielt metylerte regioner (DMR) (artikkel I), og vi anvendte robuste lineære regresjoner, justert for mors røykestatus, fars alder, barnas røyking/kjønn/alder/predikerte celletype proporsjoner (artikkel I) og mors røykestatus, studiesenter, barnas røyking/kjønn/alder/predikerte celletypeproporsjoner (artikkel III), for å påvise differensielt metylerte CpG-posisjoner (dmCpGs eller DMPs). I begge artikler ble det justert for inflasjon og metodefeil i teststatistikk, og i tilleggs-analyser studerte vi funksjonalitet av metyleringssignalene, samt molekylære og biologiske signalveier tilknyttet genene annotert til metyleringssignalene. I artikkel III ble det også utført EWAS i forhold til mors og barnas egen røykestatus for sammenligning med metyleringssignaler relatert til fars røykestatus. Signifikante metyleringssignaler fra fars prekonsepsjon og tidlig ungdomsrøyking ble til sist undersøkt i forhold til følgende helseutfall hos barna: noen gang hatt astma, noen gang hatt piping i brystet, vekt og BMI.

I artikkel II undersøkte vi ulike prekonsepsjon (<15 år, ≥15 år) og postnatale tidspunkter for røykestart hos mødre og fedre, i forhold til kroppsmasseindeks (BMI) og fettmasseindeks (FMI) hos deres voksne barn. Vi benyttet data for mødre (N=2569) og fedre (N=2111) som hadde deltatt i de befolkningsbaserte RHINE- og ECRHS-studiene i alderen 39-65 år, og data for deres voksne barn (18-49 år, N=6487) som hadde deltatt i RHINESSA studien. BMI ble beregnet fra selvrapportert høyde og vekt, og FMI var basert på bioelektriske impedansemål tilgjengelig for en undergruppe. Assosiasjoner ble analysert med generaliserte lineære regresjonsmodeller med hensyn til korrelasjon av observasjoner innenfor studiesenter og familier, justert for foreldres utdanning, og med barnas kjønn inkludert som interaksjonsterm. Medieringsanalyser ble brukt for å undersøke om observerte assosiasjoner ble mediert via foreldrenes røyking (i pakkeår), foreldrenes BMI, barnas egen røykestatus og barnas fødselsvekt.

Resultater

Artikkel I: Vi identifiserte seks DMRs (Sidak-korrigerte P-verdier: 0,0006-0,01739) assosiert med fars røykeeksponering, annotert til gener involvert i medfødt og adaptiv immunitet (ATP6V1E1, C2), fettsyresyntese (ACSF3), utvikling av nevrale system (CTNNA2) og cellulære prosesser (WDR60). Ingen DMPs oppfylte genomvidt signifikansnivå (FDR < 0,05) etter kontroll for genomisk inflasjon (λ=1,46).

Artikkel III: Vi identifiserte 2 dmCpGs (FDR<0,05 med λ = 1,29) assosiert med fars røykestart før konsepsjon, og 19 dmCpGs (FDR<0,05 med λ = 1,29) assosiert med fars røyking som startet før 15 års alder. I separate analyser av sønner og døtre, fant vi fire dmCpGs (KCNJ1, GRAMD4/DIP, TRIM2 og MYADML2) hos sønner og én dmCpG (LEPROT1) hos døtre relatert til fars røyking før konsepsjon (FDR <=0,05). Ingen av EWAS analysene avdekket signifikante DMR regioner. dmCpGs assosiert med fars tidlige ungdomsrøyking var beriket i promotorregioner, CpG-øyer og genkropper, og annotert til gener involvert i medfødt og adaptiv immunitet, inflammatoriske responser (TLR9, DNTT, PSTPIP2, CSF1R), og glukose- og fettmetabolisme (IRS1). I tillegg var noen av disse dmCPGs assosiert med vekt- og BMI-relaterte utfall hos barna (cg03380960 i FAM53B; cg12053348 (NA), og cg22402007 i NTRK2) og til astma (cg22402007 i NTRK2) og piping i brystet (cg11380624 (DNAJC14) og cg10981514 i TPCN1) hos barna. Metyleringssignalene knyttet til fars røyking var tydelig forskjellige fra signalene knyttet til mors og barnas egen røykeeksponering. Imidlertid passet metyleringssignalene fra våre EWAS for mors og egen røyking med det andre studier har vist. Dette styrker tilliten til våre funn vedrørende fars røyking, noe som er av særlig betydning fordi der ikke finnes kohorter med tilstrekkelige data for å gjøre en tilfredsstillende replikasjonsanalyse av fars prekonsepsjon røyking og DNA metylering hos hans (voksne) barn.

Artikkel II: begge foreldres prekonsepsjon røykestart var assosiert med økt BMI hos voksne barn (fars røykestart ≥15 år; β 0,551, 95 % KI: 0,174-0,929, p=0,004, n=2916; mors røykestart <15 år; β 1,161, 95 % KI: 0,378-1,944, p=0,004; debut ≥15 år; β 0,720, 95 % KI: 0,293-1,147, p=0,001, n=3531). I analysene av mors røykeeksponering ble det også observert en assosiasjon med økt BMI for røyking initiert postnatalt (β 2,257, 95 % KI: 1,220-3,294, p<0,001). Imidlertid var bare fars røykeeksponering også assosiert med økt fettmasse hos avkom, og viste en mer konsistent sammenheng med sønnenes fettmasse (fars røykestart <15 år; β 1,604, 95 % KI: 0,269-2,939, p=0,019; røykedebut ≥15 år; β 2,590, 95 % KI: 0,544-4,636, p=0,013, og debut etter fødsel, β 2,736, 95 % KI: 0,621-4,851, p= 0,001, n=129). Vi kunne ikke identifisere om fars alder ved røykestart hadde en betydning i dette begrensede datasettet med fettmassedata, og vi utførte ikke medieringsanalyser i denne undergruppen. Medieringsanalyser vedrørende BMI i hele datasettet indikerte at de observerte assosiasjonene mellom foreldres prekonsepsjon røykestart og barnas BMI var fullstendig mediert via foreldrenes røyking

i pakkeår i barnas oppvekst (fars røykestart ≥ 15 år; indirekte effekt: β 0,482, p=0,044, mors røykestart <15 år; indirekte effekt: β 1,059, p<0,001; mors røykestart ≥15 år; indirekte effekt: β 0,833, p<0,001), og delvis mediert via foreldrenes BMI samt barnas egen røykeeksponering.

Konklusjon: Våre EWAS-resultater viste at fars røyking, særlig fars røykestart før 15 års alder, var assosiert med spesifikke DNA metyleringssignaler (dmCpGs) hos hans (voksne) barn. Videre analyser gav holdepunkter for at de identifiserte signalene er av betydning for funksjonalitet, og noen av signalene var også knyttet til kroppsmasseindeks og lungehelse hos barna. Signalene var forskjellige fra signaler knyttet til mors røyking og personlig røyking. Funnene kan tyde på at fars røyking kan påvirke fenotype hos hans fremtidige barn via påvirkning på epigenetiske mekanismer. I vår epidemiologiske studie fant vi at fars røyking var assosiert med økt fettmasse hos hans sønner. Dette støtter hypotesen om overføring via farslinje, med betydning for metabolsk fenotype hos barna. Våre medieringsanalyser passet med at en rekke aspekter bidrar til overvekt, og at vedvarende og kumulativ eksponering for foreldres røyking - ikke bare foreldres røyking før konsepsjon, er av betydning for (voksne) barns risiko for overvekt. Avhandlingen indikerer altså at fars røyking kan påvirke både det epigenetiske mønster og fenotype hos hans fremtidige barn. Imidlertid bør de identifiserte metyleringssignalene om mulig replikeres i andre studier, og ytterligere studier er nødvendig bl.a. for å analysere om sammenhengen mellom fars røyking og fenotype til fremtidige barn faktisk er mediert via spesifikke epigenetiske signaler.

Abstract

Background: Epidemiological studies suggest that fathers' smoking, particularly smoking commencing in early adolescent years, can affect their offspring's metabolic and respiratory health. Tobacco smoke constituents have been demonstrated to induce alterations to the sperm epigenetic machinery and negatively affect the regulation of embryo development. It has also been suggested that the developmental stage of the sperm precursor cells may be important for their susceptibility to environmental agents. These observations provide plausible evidence for a cross-generational transmission of altered epigenetic states, and a potential epigenetic pathway by which the fathers' preconception and adolescent smoking exposures can affect the early life development and health trajectories in his offspring. Yet, to date, in humans, study of the association of smoking exposures on DNA methylation changes have largely focused on personal and maternal smoking exposures. There are also few epidemiological reports that have assessed the effects of parental smoking exposures in specific time windows, commencing in preconception and postnatal years, to establish whether early adolescence is a critical exposure-sensitive period for smoking exposure to potentiate cross-generational impacts on adult offspring's body composition and risk of obesity.

Objectives: I) To identify potential DNA methylation signals in offspring associated with fathers' ever smoking behaviours. II) To investigate time points of parents' preconception and postnatal smoking exposure onset in relation to phenotypic outcomes on offspring's body mass index and fat mass, and to investigate whether associated outcomes are modified by the sex of the offspring or mediated by parental and offspring factors (parental BMI and pack years of smoking, offspring's personal smoking and birthweight). III) To identify DNA methylation signals in male and female offspring related to fathers' preconception and early adolescent smoking onset, to investigate whether detected methylation sites are different from signals associated with personal and maternal smoking, and to further investigate if identified dmCpGs are associated with BMI and respiratory outcomes in offspring.

Material and methods: We conducted epigenome-wide association studies (EWAS) to investigate DNA methylation patterns in relation to fathers' ever smoking exposures (N=195) in offspring (11-54 years) participating in the RHINESSA and ECRHS studies

(paper I), and in relation to fathers' smoking commencing during preconception (N=875) and early adolescent (< age 15) years (N=304) in offspring (7-50 years) originating from 6 RHINESSA study centres (Paper III). In both papers offspring's DNA methylation was quantified in peripheral blood using Illumina Infinium MethylationEPIC Beadchip arrays. Differentially methylated regions were detected using Comp-p (paper I) and dmrff and DMRcate (Paper III), and robust linear regressions, adjusted for mothers smoking, fathers age, offspring smoking/sex/age/cell-type proportions (paper I) and mothers smoking, study centre, offspring smoking/sex/age/cell-type proportions (paper III), were used to detect differentially methylated CpG sites (dmCpGs). In additional analyses, associations between fathers' preconception smoking and offspring's DNA methylation were also investigated in strata of male (N=457) and female (N=418) offspring (paper III). Both papers adjusted for inflation and bias of test statistics, and searched for enrichment of regulatory regions, gene interactions and pathways to gain insight into the molecular and biological processes of the differentially methylated sites and their annotated genes. Replication of findings was pursued in the Isle of Wight (IoW) (paper I) and the Avon Longitudinal Study of Parents and Children (ALSPAC) cohorts (paper III). In paper III, EWAS of maternal and offspring's personal smoking were also performed for comparison with fathers' smoking related methylation signals. In sensitivity analyses, identified dmCpGs were regressed against the following offspring outcomes; ever-asthma, ever-wheezing, weight and BMI.

In paper II we investigated preconception and postnatal time points of smoking onset in mothers (N=2569) and fathers (N=2111) aged 39-65, of the population based RHINE and ECRHS studies, in relation to adult RHINESSA participating offspring's (18-49 years, N=6487) body mass index (BMI) and fat mass index (FMI). BMI was calculated from self-reported height and weight, and FMI was based on bioelectrical impedance measures in a subsample. Associations were analysed with generalized linear regression models, adjusted for parental education and clustered by study centre and family origin, and offspring sex was included as an interaction term. Mediation analyses were employed to investigate whether observed associations were mediated via parental pack years of smoking, parental BMI, offspring smoking and offspring birthweight.

Results:
Paper I: we identified six DMRs in offspring (Sidak corrected P-values: 0.0006-0.01739) associated with fathers' ever smoking exposures, annotated to genes involved in innate and adaptive immunity (ATP6V1E1, C2), fatty acid synthesis (ACSF3), as well as to neural system development (CTNNA2) and cellular processes (WDR60). No DMPs passed epigenome significance (FDR < 0.05) after controlling for genomic inflation (λ=1.46).

Paper III: we identified 2 dmCpGs in offspring (FDR<0.05 with λ=1.29) associated with fathers' preconception smoking onset, and 19 dmCpGs (FDR<0.05 with λ=1.29) associated with fathers' smoking commencing in early adolescent years. Sex-stratified analyses detected four dmCpGs (KCNJ1, GRAMD4/DIP, TRIM2 and MYADML2) in males and one dmCpG (LEPROT1) in females related to fathers' preconception smoking (FDR ≤0.05). Significant DMRs were not detected in either EWAS. Of note, differentially methylated sites related to fathers' early adolescent smoking, were enriched for promotor regions, CpG islands and gene bodies. They were distinctly different from methylation signals identified in the EWAS on maternal and personal smoking, and annotated to genes with roles in innate and adaptive immunity and inflammatory responses (TLR9, DNTT, PSTPIP2, CSF1R), as well as with glucose and fat metabolism (IRS1). Some of the identified dmCpGs were additionally associated with weight and BMI related outcomes in the offspring (cg03380960 in FAM53B; cg12053348 (NA), and cg22402007 in NTRK2) and to offspring's ever-asthma (cg22402007 in NTRK2) and ever-wheeze (cg11380624 (DNAJC14) and cg10981514 in TPCN1). Our EWAS results have not yet been successfully replicated in an independent cohort and warrant further conformation in order to be verified as true positive findings.

Paper II: both parents' preconception smoking onset was associated with increased BMI in adult offspring (Fathers' onset ≥15 years; β 0.551, 95% CI: 0.174-0.929, p=0.004, n=2916; Mothers' onset <15 years; β 1.161, 95% CI: 0.378-1.944, p=0.004; onset ≥15 years; β 0.720, 95% CI: 0.293-1.147, p=0.001, n=3531). In the maternal lineage an association was also observed when smoking was initiated in postnatal years (β 2.257, 95% CI: 1.220-3.294, p<0.001). However, only fathers' smoking exposures were also associated with increased fat mass, and demonstrated a more consistent

impact on the sons (onset <15 years; β 1.604, 95% CI: 0.269-2.939, p=0.019; onset ≥15 years; β 2.590, 95% CI: 0.544-4.636, p=0.013; and onset after birth; β 2.736, 95% CI: 0.621-4.851, p= 0.001, n=129). This relationship was not found to be more pronounced if the fathers started to smoke in early adolescent years. Although not high enough numbers to pursue in the subsample with fat mass data, independent mediation analysis indicated that the observed associations between parents' preconception smoking onset and adult offspring BMI were fully mediated via the parents' pack years smoked during childhood years (father onset ≥15 years; indirect effect: β 0.482, p=0.044, mother onset <15 years; indirect effect: β 1.059, p<0.001; mother onset ≥15 years; indirect effect: β 0.833, p<0.001), and partially mediated via parental BMI and offspring own smoking exposure.

Conclusion: Our novel EWAS results indicated that fathers' smoking, particularly smoking commencing during early adolescent years, was associated with differentially methylated CpG sites in offspring. Further analyses suggested that the identified signals are functionally important, and several of the identified dmCpGs were related to BMI, weight and respiratory outcomes in the offspring suggesting father's smoking might influence offspring phenotype through epigenetic mechanisms. The epigenetic signals related to father's smoking were distinct from those related to mother's or personal smoking, while our EWAS of personal and mother's smoking showed results comparable with previous studies. This lends support to the validity of the EWAS results of father's smoking, in the absence of available data for appropriate replication analyses. Our epidemiological study found that fathers' smoking was associated with increased fat mass in their sons, which lends support to a specific paternal lineage transmission of male-specific responses on offspring's body composition and obesity related phenotypes. Our mediation analyses support the multifactorial aspects contributing to obesity, and that the sustained and cumulative exposures of parental smoking trajectories, and not parental preconception smoking alone, are important for offspring risk of obesity. In conclusion, this thesis indicates that father's smoking, in particular early onset (adolescent) smoking, may influence both the epigenetic patterns and the phenotype of his future offspring. However, the identified novel methylation signals should be replicated, if possible, in other studies, and future studies are needed in order to explore whether the associations of father's smoking with offspring phenotype are mediated via epigenetic alterations.

List of papers

Paper I

Mørkve Knudsen G.T[*], Rezwan F.I[*], Johannessen A, Skulstad S.M, Bertelsen R.J, Real F.G, Krauss-Etschmann S, Patil V, Jarvis D, Arshad S.H, Holloway J.W, Svanes C. **Epigenome-wide association of father's smoking with offspring DNA methylation: a hypothesis-generating study**

Environmental Epigenetics, 2019, Dec 6;5(4): dvz023. doi: 10.1093/eep/dvz023. eCollection 2019

Paper II

Mørkve Knudsen G.T, Dharmage S, Janson C, Abramson M.J, Beneditsdóttir B, Malinovschi A, Skulstad S.M, Bertelsen R.J, Real F.G, Schlünssen V, Jõgi N.O, Sánches-Ramos J.L, Holm M, Garcia-Aymerich J, Forsberg B, Svanes C, Johannessen A. **Parents' smoking onset before conception as related to body mass index and fat mass in adult offspring: Findings from the RHINESSA generation study**

PLoS One. 2020 Jul &;15(7): e 0235632. doi: 10.1371/ journal.pone. 0235632. eCollection 2020

Paper III

Kitaba NT[*], Mørkve Knudsen G.T[*], Ane Johannessen A, Faisal I. Rezwan F.I, Malinovschi A, Oudin A, Benediktsdottir B, Martino D, Callejas González F.J, Gómez L.P, Holm M, Jõgi N.O, Dharmage S, Skulstad S.M, Watkins S.H, Suderman M, Ilen-Caven Y, Golding J, Real F.G, Schlünssen V, Svanes C[#], Holloway J.W[#]. **Fathers' preconception smoking and offspring DNA methylation: A two generation study**

Submitted October 2022. [*]Equal joint first authors, [#] Equal joint senior authors

1. Introduction

1.0 Researcher perspective

When I started this PhD project, I was very much aware that my background as a physiotherapist with a master degree in global health was a far cry from epigenetic epidemiology, and that I now entered uncharted waters to a field of scientific concepts that was completely new to me. It has been a complex but above all fascinating undertaking to get an understanding of the biological roles of epigenetic systems, how their propagation of regulatory states, albeit essential for normal cell tissue function and development, also are highly responsive to exogeneous influences and confer potential molecular mechanisms involved in the development of complex diseases.

With an overall focus on paternal exposures in relation to offspring outcomes, and a *priori* hypothesis of a potential epigenetic inheritance through the male germline, it has been instructive to gather knowledge of the sequence-specific periods of epigenetic reprogramming events throughout germ cell development, in which the sperm epigenome is increasingly dynamic and thus susceptible for environmental exposures to modify germline epigenetic processes that is crucial for normal sperm function and for maintaining embryonic development.

A valuable contribution to this learning process has been the opportunity to participate as coauthor on a rostrum paper on transgenerational and intergenerational epigenetic inheritance in allergic diseases (Journal of Allergy and Clinical Immunology. 2018; 142: 765-72). Although, the paper is not part of my PhD thesis, it increased my understanding of the regulatory role of epigenetic mechanisms, their involvement in the development of sperm capable of fertilization and for modulating gene expression patterns during embryonic development, and not at least, how these epigenetic systems may serve as potential candidates for a cross-generational inheritance of altered epigenetic states. As the rationale for the thesis proceeds from this molecular context, the next paragraphs aim to describe the theoretical framework, and as such try to set the stage for the objectives of the PhD work.

1.1 Concepts of epigenetics

The recognition that there is more to heredity than genes is not a new concept, neither is the discipline of epigenetics. It dates to the early 1940s when Conrad Waddington introduced the term and conceptualized an epigenetic landscape (fig 2) to illustrate how the genotype and developmental environment interacted in directing cellular trajectories during embryonic development to bring the phenotype into being [1, 2].

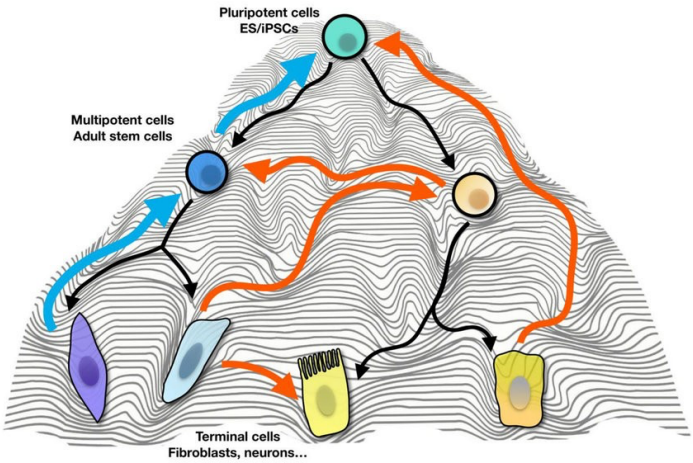


Figure 2: Waddington´s visual metaphor of the various developmental pathways a cell might take toward differentiation. On its way to a final tissue type, cells, depicted as balls, roll through an epigenetic landscape of bifurcating valleys and ridges. Each valley corresponded to a possible cell fate and the ridges separating the valleys represented barriers to maintain the committed cell fate once chosen. Extracted and reused from Gam R, Sung M and Pandurangan AP. Cells, 2019; 8(10), 1189 [3]. Illustration adapted from Waddington C.H. 1957. The strategy of the Genes (London: Geo Allen & Unwin).

Although Waddington laid the groundwork for our present understanding of the relationship between genes and development, epigenetics has over the last decades

evolved and developed into a discipline that today has taken centre stage within biochemical research [4]. The enhanced interest in epigenetics accommodates the technological breakthroughs and accelerated discoveries of the molecular mechanisms that control gene activity and the inheritance of cell phenotypes. From a contemporary and operational view, epigenetics can be defined as the study of dynamic and chemical modifications that occur to our DNA, through alterations in the chromosome rather in the DNA sequence, and how these modifications affect an individual cell or organism's expression of genetic information in a potentially heritable way, both through cell division (mitosis) and between generations (meiosis) [5].

1.2 Epigenetic mechanisms

The role and function of many epigenetic mechanisms have been identified, and additional players involved in the intricacy of gene expression regulation are likely to surface as research proceeds. Although this PhD works pertains to methylomic alterations, the following section will describe three main epigenetic phenomena; DNA methylation, post-translational histone modifications, and non-coding RNAs (figure 3). Notwithstanding these modifications act through distinct mechanisms, their control on gene expression is closely interconnected and are mechanistically dependent on each other.

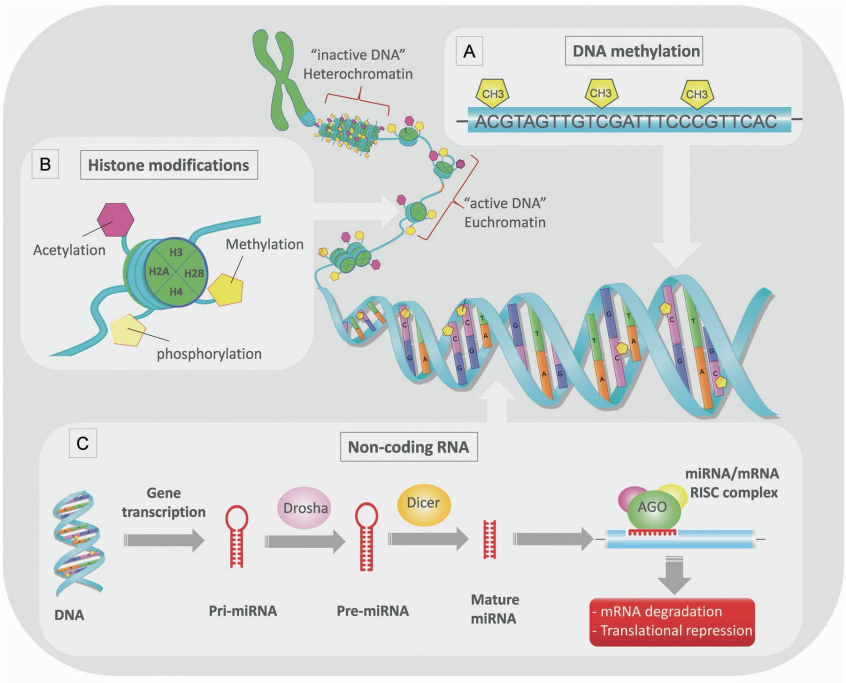


Figure 3: Three fundamental epigenetic modifications. **A** Catalysed by DNA methyltransferases, DNA methylation is the covalent addition of a methyl group to the 5-position of cytosine in the context of cytosine guanine dinucleotides to form 5-methylcytosine. In a genomic context-dependent manner cytosine methylation influence on gene expression regulation through inhibition or activation of transcription, either by directly interfering with binding of DNA and transcription factors, or through recruitment of proteins that bind to modify DNA and thus blocking other transcription factors from binding to the site. **B** Post-translational modifications such as acetylation, methylation and phosphorylation of histone proteins control gene expression through different mechanisms, such as changing the electrostatic charges of histones which affect the binding capacity of DNA-histone interactions and influence on DNA accessibility and transcription activity. Modified histones can also interact and bind to various proteins which further modify the histones and dynamically regulate chromatin structure and hence gene expression. **C** Non-coding RNAs can affect gene expression trough transcriptional or post-transcriptional mechanisms. Here exemplified by the biogenesis of micro-RNAs, a process in which miRNA processing enzymes (Drosha and Dicer) executes the cleavage of precursor primary miRNAs (pri-miRNAs) and pre-miRNAs to form mature miRNAs. miRNAs can mediate gene silencing by their loading into the Argonaut (AGO) protein within the RNA-induced silencing complex (RISC), which promote translational inhibition or degradation of mRNA transcripts. Reused with permission from the Japanese Society of Hypertension Research. Originally published by Arif M, Sadayappan S, Becker RC, Martin LJ, Urbina EM, in Hypertension research 2019; 42: 1099-1113 [6].

1.2.1 DNA methylation

Catalysed by DNA methyltransferase enzymes (DNMTs), cytosine methylation (5-methylcytosine, 5mC) is the most widely studied DNA modification, and involves the transfer of a methyl group directly to a cytosine nucleotide within a cytosine phosphate guanine (CpG) dinucleotide [7]. Although DNA methylation is highly flexible during development and cell differentiation [7-11], methylation patterns can be stably inherited across multiple cell generations and play a vital role for establishing and maintaining cellular identity through the control of chromatin structure and gene expression [12, 13]. This dynamic process is mediated by enzymes that add ("write"), recognize ("read") or remove ("erase") methyl groups onto cytosine residues.

Three DNMTs, including DNMT1, DNMT3A and DNMT3B, catalyse the addition of methyl groups to genomic DNA and act as writers of DNA methylation [14]. Based on a classical DNA methylation model, DNMT1 functions as a maintenance methyltransferase that repairs DNA methylation [15] and faithfully copies and maintains methylation marks from the parental to the synthesized daughter strand during DNA replication [15, 16] . DNMT3A and DNMT3B, in conjunction with the regulatory role of the catalytic inactive DNMT3L, are considered as *de novo* methyltransferases [17, 18]. However, it has become increasingly apparent that maintenance and de novo methyltransferase enzymes have overlapping roles, and that they are all necessary for proper DNA methylation initiation and maintenance [19].

DNA methylation influences gene expression by either directly interfering with transcription factor binding [20], or indirectly, by recruiting "reader" proteins that recognize and bind to methyl groups and regulate DNA transcription through chromatin remodelling and DNA accessibility [21, 22]. Removal, and erasure, of methyl groups from cytosines eventuate through either active or passive demethylation processes. Passive DNA demethylation occurs when the maintenance

methyltransferase DNMT1 does not copy methylation patterns during successive rounds of DNA replication in mitosis. By contrast, active demethylation requires enzymatic reactions, mediated by enzymes, such as the ten-eleven translocation (TET) family, that convert 5mC back to cytosine [23].

The majority of CpG residues across the eukaryotic genome are methylated (70-80%) [7]. Heavily methylated genomic regions are generally associated with repressed chromatin states, such as intergenic regions, where DNA methylation is mainly involved in silencing retroviral elements and to maintain genomic stability [24]. CpG islands are clusters of CpG dinucleotides, often found at the promotor regions of genes. Cytosine methylation within a promoter region is believed to prevent transcriptional initiation by recruiting gene suppressor proteins and inhibiting transcription factor binding to DNA [21]. However, CpG islands, and notably, those associated with gene promotors, are normally free from DNA methylation [25]. Genes with promoters containing CpG islands are found to have a distinct chromatin organization associated with active transcription [26]. Methylated cytosines in regions downstream of a transcription start site are also shown to be informative of gene expression, such as gene body methylation, which is associated with increased gene expression in dividing cells [27-29]. This demonstrates the dual roles DNA methylation exerts on both transcriptional activation and repression of genes, depending on the genomic location and context.

Cytosine methylation is a stable epigenetic component, and due to its ability to remain intact during DNA extraction, processing and long term storage, DNA methylation is widely used as an epigenetic marker in epidemiological studies [30]. As whole blood is easily accessible, it has become a commonly assayed tissue for studying DNA methylation signatures in humans. Although bisulphite-DNA sequencing of human chromosomes has indicated that DNA methylation levels between sexes and across different tissues exert considerable similarities, especially in developmentally close tissues, such as different types of lymphocytes which are all derived from mesoderm

[31], different cell types and tissues nonetheless have their own distinct DNA methylation profiles [32]. As interindividual cellular proportions can introduce systematic differences in the methylation profiling between cases and controls, the diverse cellular composition in whole blood therefore represents a major caveat which needs careful methodological consideration when it comes to interpreting DNA methylation differences in relation to extrinsic environmental influences or disease states [32]. In addition to being influenced by cellular identity, DNA methylation variability also arises in response to aging and intrinsic genomic characteristics such as DNA sequence patterns and single nucleotide polymorphisms [33, 34]. Consequently, DNA methylation is influenced by several factors that need consideration when characterising and interpreting methylomic variability in humans.

1.2.2 Histone modification and chromatin structure

The most well-understood post-translational modifications to the N-terminal tails of histone proteins include enzyme-catalysed acetylation, methylation, and phosphorylation. In combination these chemical modifications have distinct effects on chromatin regulation. By altering the electrostatic charge of histones and thereby influencing the binding capacity between DNA-histone interactions in nucleosomes, they can affect the recruitment and binding of regulatory proteins [35, 36]. Nucleosomes, which are the building blocks of chromatin, are composed of an octamer of histone proteins in which a section of DNA is wrapped around [37]. By recruiting chromatin remodelling complexes, histone modifications can regulate DNA accessibility by engaging in either opening the chromatin (euchromatin), which facilitates transcription, or closing the chromatin (heterochromatin), generally associated with transcriptional repression [38]. Whereas histone acetylation, catalysed by histone acetyltransferases (HATs), promotes structural relaxation of chromatin by weakening the electrostatic interactions with DNA and is usually associated with increased transcription factor binding and gene expression, removal of acetyl groups by histone deacetylases (HDACs) opposingly condense the nucleosomes and function as a repressor of gene expression [35, 39]. Histone methylation, mediated by histone

methyltransferases (HMTs), and removed by demethylases, can either activate or repress transcription, depending on where the methylation site is located and/or the number of methyl groups that are added [35, 36]. Contrary to histone acetylation, histone methylation does not alter the histone charge or directly affect DNA-histone interactions, but rather impacts on the recruitment of regulatory binding proteins to chromatin [40, 41]. Histone phosphorylation, catalysed by kinases, has besides regulating transcriptional activity, pivotal roles in DNA damage repair and chromatin condensation during cell divisions [36, 42]. Histone phosphorylation also interacts with other histone modifications and generates a complex regulatory network that interfere with chromatin function and gene expression [43].

1.2.3 Non-coding RNA

Non-coding RNAs (ncRNAs), are molecules that are transcribed from DNA but not translated into proteins. Based on their size, they can be categorised as short non-coding RNAs, and long non-coding RNAs. Short ncRNAs less than 200 nucleotides (nt) include microRNAs (miRNAs), short interfering RNAs (siRNAs) and piwi-interacting RNAs (piRNAs), whereas ncRNAs longer than 200 nt are known as long non-coding RNAs (lncRNAs) [44]. Although their role in the crosstalk of transcriptional regulation continues to emerge, it is increasingly recognised that non-coding RNAs, coordinate with DNA methylation and histone proteins to govern gene regulation by modulating chromatin structure and affecting transcription (reviewed in [45-48]). The mechanisms by which the different classes of non-coding RNA molecules guide crucial biological processes during development and cellular differentiation are distinct, yet they appear to function as a regulatory network and interact with each other to influence on chromatin enzyme activity, DNA accessibility, and transcriptional inhibition or activation through transcriptional and post-translational regulatory mechanism in cells (reviewed in [45-48]). siRNAs have important roles in maintaining genomic stability and can induce transcriptional gene silence by influencing on DNA methylation and histone modifications in cells [49-51]. Similarly, piRNAs are involved in protecting germline integrity by reducing

transposon methylation, and thus suppressing retroviral transposon elements in the testes, preventing germ line mutations [52]. Mature miRNAs can mediate gene silencing through their incorporation into the RNA-induced silencing complex (RISC), which promotes translational inhibition or degradation of mRNA transcripts [53]. miRNAs have been shown to alter the DNA or chromatin state through regulation of histone modifying enzymes and chromatin remodelling enzyme activity, and by impacting on DNA methylation in cells via regulation of DNA methylases [54-59]. The highly heterogeneous class of long non-coding RNAs (lncRNAs) regulate gene expression of nearby and distant genes at multiple levels and exert their functions by interacting with DNA, RNA and proteins (Reviewed in [47]). They can affect transcription by regulating chromatin structure [60, 61], or by interfering with histone-DNA interactions [62]. lncRNAs also recruit chromatin modifying factors and histone modifications enzymes [63, 64] which in turn can inhibit [65-67] or promote [68, 69] activation of genes. Some lncRNAs directly interact with DNA to affect chromatin accessibility in order to activate [70-72], or repress [71, 73, 74] gene expression. lncRNAs also engage in post-transcriptional regulation of gene expression, such as binding to RNA sequences and influence on mRNA splicing processes and gene modulation (highlighted in Figure 5A in [47], or through regulation of other ncRNAs, such as miRNAs, by binding to their sites, and as such reduce their regulatory effect on their target mRNAs (highlighted in figure 5C in [47].

1.3 The role of epigenetics in epidemiological research on health and disease

The epigenome, which collectively comprises all the epigenetic modifications in a single cell, is undoubtedly essential for stabilising gene expression patterns required for preservation of cellular identity through differentiation and for normal cell function. However, epigenetic mechanisms also display a high degree of structural adaption and can be modified by exogeneous influences through molecular events which may affect gene expression profiles and phenotypic outcomes [75, 76]. Although such flexible and dynamic responses to intra- and extracellular stimuli allow

an organism to adapt in response to a changing environment which are crucial for normal development and health maintenance, epigenetic regulators have also been suggested as a possible mechanistic link between the environment, our genomic function and susceptibility to disease [77]. This has led to the development of a new science – epigenetic epidemiology- which can be defined as the study of the associations between epigenetic variation and human disease etiology [78]. It aims to identify epigenetic biomarkers related to various environmental exposures or disease states, and seeks to explore underlying epigenetic mechanisms associated with risk factors and health outcomes [30]. The role of epigenetic modifications in risk of disease has particularly been embraced by epidemiological research within the framework of "Developmental Origins of Health and Disease" (DOHaD) [79], which in numerous studies have reported relationships between environmental conditions during early life and diseases manifesting in adulthood (reviewed in [80]). In the last decade, a growing body of epidemiological literature has also suggested that the environmental exposures and lifestyle habits of one generation can modify the risk of disease initiation and progression in subsequent generations (reviewed in [81]). Moreover, it has become evident that epigenetic mechanisms involved in conveying environmental induced plastic responses can profoundly influence gamete formation, thus providing a route through which a father's environment can affect the development and phenotypic variation in his offspring [82]. This has generated a substantial interest in investigating the cross-generational effects that fathers' environmental exposures may produce on epigenetic states and developmental plasticity.

1.4 Epigenetic inheritance

Cross-generational inheritance of altered epigenetic states is theoretically inferred as being either intergenerational or transgenerational, depending on whether the transmitted epigenetic marks, which may be referred to as epimutations [83], are directly exposed to an environmental stressor, or transferred through the germline and persist to generations transcending the initial exposure [84]. In case of an exposed

pregnant female (figure 4A), intergenerational epigenetic inheritance involves epigenomic changes that affect the pregnant (F0), the fetus (F1) as well as the fetal germline cells (F2) as these can all be accounted for by direct effects of exposure on the DNA that will form that generation (either the unborn child or the oocytes that will contribute to the grandchild). A transgenerational transmission can only be proven to occur if the epigenetic changes persist to the third generation (F3), which would be the first generation not directly affected by the initial environmental stressor. Intergenerational inheritance in an exposed male and non-pregnant female (figure 4B) encompasses epigenetic changes in the individual (F0), and their germ cells that eventually will produce their progeny (F1). In this scenario, epigenetic modifications persisting to the second generation (F2), and transcending the direct environmental exposure, would be considered a true transgenerational effect. At present, the molecular basis for transgenerational inheritance is still controversial, although being increasingly recognised, as technological advances in profiling epigenetic events of germ cells and embryos at multiple stages are improving our understanding of the cellular mechanisms that may perpetuate a germline transmission of epigenetic signatures.

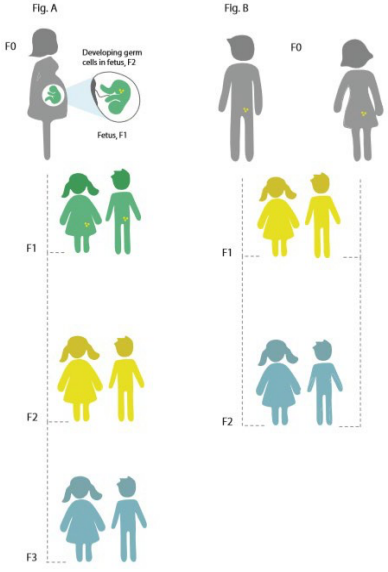


Figure 4: Illustration of intergenerational versus transgenerational inheritance: **A** If a pregnant female is exposed to an environmental stressor, epigenetic perturbations in the mother (F0), the fetus (F1, green) and the fetal germ cells (F2, yellow) would all be directly exposed from the external stimulus. Thus, transmission from F0-F2 encompass intergenerational inheritance. Epigenetic effects persistent to the F3 generation (blue) would represent true transgenerational inheritance. **B** If a male or non-pregnant female is exposed, epigenetic perturbations in the individuals (F0) and their germ cells that eventually will shape their offspring (F1, yellow) would represent intergenerational inheritance. Epigenetic modifications observed in the F2 generations (blue) would be considered a transgenerational effect. Reused from Mørkve Knudsen T, Rezwan FI, Jiang Y, Karmaus W, Svanes C, Holloway JW, et al. J Allergy Clin Immunol 2018; 142:765-72

1.5 Epigenetic reprogramming during the male germline and embryonic development

In the life cycle of mammals there are at least two developmental periods in which genome-wide reprogramming occurs and epigenetic stability is thoroughly perturbed and thus increasingly responsive to environmental experiences: during gametogenesis, and shortly after fertilization when the parental gametes fuse to form the zygote [85, 86]. During these developmental stages, a concerted action of epigenetic mechanisms participates in a highly ordered transcriptional and posttranscriptional regulation of cell proliferation, differentiation and testis-specific gene expression that is essential for the formation of functional gametes that possess the capacity of fertilization and generating totipotency in the zygote [87-91]. Conversely, if this coordinated and sequence-specific epigenetic regulation is disrupted, germline reprogramming may fail and foster an aberrant embryonic development. Thus, epigenetic reprogramming events across the various stages of male germ cell development are thought to represent periods of increased susceptibility for environmental exposures to influence the epigenome (figure 5) [92]. Moreover, if perturbed epigenetic marks (epimutations) are retained throughout gametogenesis and in the preimplantation embryo, they provide a mechanism for a germline transmission of epigenetic states, with the potential to elicit cross-generational effects and phenotypic alterations in subsequent generations [93]. In the following section these germline reprogramming events, and how they may facilitate epigenetic inheritance across generations, will be further elaborated.

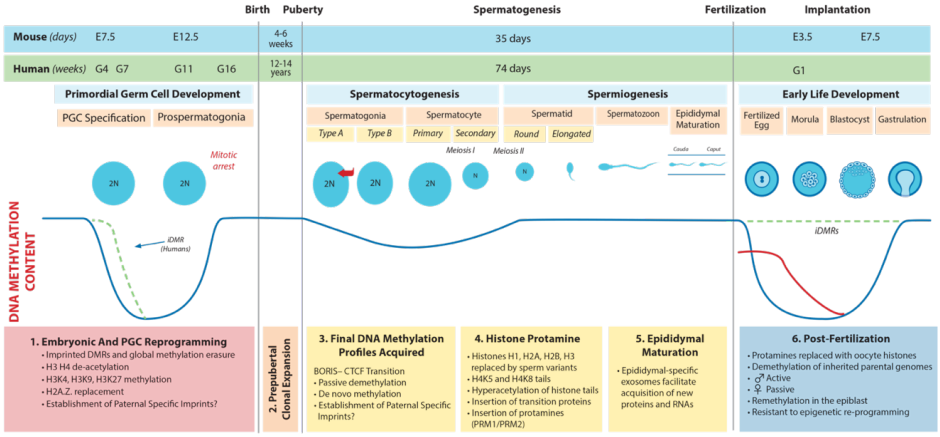


Figure 5: Multistage-specific epigenetic remodelling events occur throughout gametogenesis and in the pre-implantation embryo in order to re-establish pluripotency between each generation. **1** Primordial germ cells (PGCs), specified from the epiblast, undergo global DNA demethylation, dynamic histone modifications with transient loss of heterochromatin histone marks, and erasure of parental imprints to ensure genetic totipotency. Several ncRNAs, specifically miRNAs, and presumably piRNAs and lncRNA contribute to control DNA and chromatin remodelling during PGC specification through interacting with histone modifying enzymes and transcription factors. piRNAs components are specifically engaged to protect genomic stability from early phases of germ cell specification and throughout germline development. In mitotically arrested prospermatogonia, de novo methylation occurs. **2** Approximately 2 years before puberty onsets, expansion of prospermatogonia occurs in a gonadotropin independent manner as the expression of puberty activating genes in the hypothalamus is repressed by a complex network of DNA methylation, repressive histone marks and histone demethylase enzymes. **3** At puberty, the HPG axis is fully activated and initiates spermacytogenesis, in which mitotic proliferation of immature spermatogonia produces spermatocytes and take on meiosis to proliferate and differentiate into spermatids. The process is tightly regulated by DNA modifications and chromatin remodelling events in which final acquisition and erasure of methylation marks is achieved, including de novo methylation of paternal imprints. Histone variants are incorporated. miRNAs are also contributing to regulate differentiation and induce meiosis. **4** During spermiogenesis spermatids differentiate and mature to spermatozoa. This process is characterized by global reorganization of chromatin, and extensive nuclear shaping and condensation through testis-specific histone replacement and histone-protamine exchange. Through their engagement in chromatin remodelling complexes and by post-transcriptional regulating of gene expression, miRNAs are also involved in chromatin condensation and protamine targeting. **5** To gain motility and fertilization potential spermatozoa mature and modify during transit through the epididymis. The epididymis harbour unique small ncRNAs which are transported to the mature sperm by exosomes. **6** After fertilization, parental specific epigenetic marks of gametes undergo reprogramming through active and passive demethylation and protamine-histone (sperm protamines are replaced with oocyte histones) exchange to establish totipotency in the developing embryo. Reused from Wu H, Hauser R, Krawetz S.A, Pilsner J.R. Current Environmental Health Reports, 2015. 2(4): p.356-366.

As embryonic primordial germ cells (PGCs) are derived from epiblast cells, which have begun a on a course of somatic fate, a complex regulatory network of DNA remodelling and chromatin reorganization participates in repressing or activating germline genes that drives the erasure of parental imprints and re-establishment of totipotency for sex-specific epigenetic programming (reviewed in ([87, 92, 94, 95]). Concurrent with global and imprint specific demethylation, a transient loss of heterochromatin histone marks occurs [96, 97]. Non-coding RNA components also engage in this coordinated process, where specifically miRNAs are important for triggering PGS specification [98], and for precursor germ cells to properly proceed mitotic proliferation and subsequent formation of prospermatogonia [99, 100] (Figure 5,1). Studies have also indicated a potential regulatory role for piRNAs [101], and possibly lncRNAs [102, 103] in controlling DNA remodelling and transcription factors during PGC specification.

Despite the comprehensive methylation loss in primordial germ cells, erasure is not complete, predominantly due to the resistance of repeat sequences such as intracisternal A-particle (IAP) retrotransposons and their proximal genes to undergo demethylation [104] . Although this is presumably an important mechanism for maintaining genomic stability during exhaustive erasure [105], methylation marks in repeat sequences confer a potential mechanistic candidate for transgenerational inheritance. In prospermatogonia, methylation patterns are gradually re-established, although final acquisition takes place during spermatogenesis [106]. Beyond the methylation status of repetitive elements, experimental models have also demonstrated that methyltransferases involved in the establishment and maintenance of DNA methylation (Dnmt1 and Dnmt3l) throughout gametogenesis can induce aberrant methylation patterns in gametes and alter gene expression in subsequent generations [107, 108], and as such also have a role in initiating cross-generational inheritance.

Several rodent studies have evidenced the exposure sensitive nature of this developmental stage, in which various environmental factors have been found to alter the epigenetic components of primordial germ cells and negatively affect normal

sperm development. Whereas the endocrine disruptor bisphenol A (BPA) has been associated with changes to histones and DNA methylation patterns in precursor germ cells and an impaired PGC migration [109, 110], the agricultural fungicide vinclozolin has been demonstrated to elicit altered DNA methylation patterns [111] and miRNA profiles in primordial germ cells [112]. Similarly, have PGCs shown to exhibit changes in DNA methylation in response to hypoglycemic conditions in the uterus [113]. In utero caloric restriction during the time of re-establishment of prospermatogonia methylation patterns has also been found not only to cause differentially methylated regions in the sperm of the F1 generation, but also to induce metabolic-related disorders in the subsequent F2 and F3 generations [114], which strongly support that environmentally-induced perturbations to the sperm epigenome, indeed can have severe impacts on future offspring's development and health.

During the slow growth period, a prepubertal expansion of undifferentiated prospermatogonia proliferation occurs in a relatively gonadotropin independent manner [115], as expression of puberty activating genes in the hypothalamus and the secretory activity of gonadotropin-releasing hormone (GnRH) neurons is predominantly under an inhibitory control at this developmental stage [116]. A complex network of DNA methylation, repressive histone marks and histone demethylase enzymes act as regulatory mechanisms that control this repressive state [116-119] (fig 5, 2). Although as yet not investigated in sperm epigenetic studies, there are epidemiological data that associate exposures such as excess food supply [120-122] and smoking [123] during the slow growth period and prepubertal years with metabolic and cardiovascular health and obesity risk in subsequent generation(s). Moreover, dioxin exposure during childhood years (up to the age of 9) has been found to reduce sperm concentration and motility, and permanently alter levels of estradiol and follicle-stimulating hormones (FSH) [124].

In the pubertal transitional phase, there is a regulatory switch of puberty activating gene expression from an inhibitory to an excitatory state, which activates the

hypothalamic-pituitary-gonadal (HPG) axis to release GnRH and initiate puberty [116]. These processes are epigenetically controlled through various mechanisms, including loss of repressor enzymes and insertion of DNA modifications and activating histone marks that facilitate expression [116, 119]. Concurrently, posttranscriptional regulation of miRNAs contributes to aid normal pubertal transition by controlling pituitary development and proper HPG axis function [116, 125]. With the onset of puberty, immature spermatogonia enter spermacytogenesis, and differentiate through mitosis and two rounds of meiosis to produce spermatocytes and round spermatids, respectively [92]. These multistep developmental events are regulated by the coordinated actions of non-coding RNAs, chromatin remodelling and DNA modifications. miRNAs have pivotal roles in germ cell differentiation and meiosis initiation [99, 126], and throughout this stage histone variants are incorporated [127], accompanied by final erasure and re-establishment of methylation patterns, including de novo methylation of paternal imprinted regions [128] (fig 5, 3).

Although few epigenetic studies have time targeted this phase of testicular development, pubertal exposure to fungicides have been found to alter histone methylation levels in mice sperm [129, 130]. There are also epidemiological findings that clearly suggest that early adolescence (below the age of 15) represents an important period of susceptibility for environmental exposures to negatively impact on the next generation's lung health [131-133]. Moreover, dioxin exposure during pre/pubertal years (age of 10-17) has been associated with alterations in sperm parameters as well as to permanent changes to estradiol and FSH levels [124], which can affect the initiation of spermatogenesis, and the ability to maintain normal sperm production in adulthood [134, 135]. Interestingly, whereas dioxin exposure in childhood years (1-9) resulted in decreased sperm motility and concentrations, dioxin exposure in pre/pubertal years elicited an opposite effect, and neither sperm parameters nor endocrine levels were affected if the dioxin exposure occurred in adulthood [124].

During spermiogenesis, spermatids differentiate and mature to spermatozoa [92]. This process is characterized by global reorganization of chromatin, and extensive nuclear shaping and condensation through testis-specific histone replacement and histone-protamine exchange [136-138]. Through their engagement in chromatin remodelling complexes and by post-transcriptional regulating of gene expression, miRNAs are also involved in chromatin condensation and protamine targeting [139-141]. The process through which histones are replaced with protamines enables compact packaging of DNA and heterochromatinization that restricts transcriptional activity. This is critical for the maturation and motility of sperm as well as to provide a safe environment for the paternal genome before encountering the epididymis and female reproductive tract [142] (fig 5, 4). However, histone-protamine exchange is not complete, 5-15 % of histones are retained in mature sperm [143, 144]. As histone retention is not randomly distributed in the genome and also have been found to be specifically enriched in regulatory regions of developmental and imprinted genes [143], it has been suggested to provide a structural framework to govern reprogramming events within the paternal genome.

Animal studies have demonstrated that the sperm retention process can be altered in response to a variety of environmental exposures, such as toxicants [145], cigarette smoke [146], heavy pollution [147] as well as in utero caloric restriction [114]. Experimental models have also demonstrated that manipulation of histone retention can affect genomic function and development in the zygote and subsequent generations, which raises the possibility for a heritable gametic chromatin state [148-150]. Skinner and colleagues have even demonstrated a transgenerational effect of toxicants on histone retention in a rodent F3 generation [145, 151, 152]. Intriguingly, although the F1, F2 and F3 generations all displayed altered DNA methylation patterns, the set of DNA methylation changes varied between the directly exposed (F1, F2) and the non-exposed (F3) generations. Similarly, as histone retention changes were only seen in the F3 generation, this may indicate that the initial exposure-induced sperm epimutations can promote additional, and possibly stochastic, epigenetic changes to embryonic and germ cell development in subsequent generations [153].

To gain motility and fertilization potential, spermatozoa mature and modify during transit through the epididymis [154]. The epididymis harbour unique profiles of non-coding small RNAs (sRNAs), including miRNAs, piRNAs, ribosomal small RNAs (rRNAs) and transfer derived sRNAs (tsRNAs) also known as tRNAs fragments (tRFs) [155], which are transported to the mature sperm by epididymosomes. It has been increasingly acknowledged that this epididymosomal sRNA transfer is responsible for a significant remodelling of sperm sRNA load during the post-testicular maturation process in the epididymis [156], and may have critical roles for the achievement of competent sperm required for normal embryogenesis [157-159] (fig 5, 5). Epididymosomal sRNAs, specifically tsRNAs are also believed to regulate the transitions in gene expression patterns in the preimplantation embryo [156], one possible mechanism through their influences on histone mRNA processing and chromatin structure [160]. Epididymosomal and sperm sRNAs levels are influenced by lifestyle related factors such as paternal diet, stress [156, 161-169] (reviewed in [170]), as well as smoking [171], and have also been shown to affect phenotypic outcomes in offspring, such as early life weight [171], susceptibility to metabolic disorders [172-174] and neuro-behavioural development [168]. Rodent studies have also demonstrated that in utero exposure to vinclozolin induces changes of sperm piRNAs and tRNA upon reaching adulthood [151, 152]. This has increasingly acknowledged epididymosomal sRNAs as a potential mechanism of epigenetic inheritance from fathers to offspring.

Several human studies have also demonstrated that a broad range of environmental factors can influence on sperm DNA methylation patterns in spermatozoa and mature sperm (reviewed in [95]), which may potentiate a cross-generational inheritance of epigenetic abnormalities and influence on early life development. This includes exposure to phthalates [175], alcohol consumption [176], flame retardants [177, 178], chemotherapy treatment [179], obesity [180], and exercise [181].

In the early zygote, following fertilization, a second wave of genome-wide reprogramming occurs in which the paternal genome undergoes extensive chromatin remodelling and decondensation through protamine removal and deposition of the

histone variant H3.3 [182]. This is followed by active and passive demethylation of parental methylation patterns, which are essential to ensure totipotency in the developing embryo [104]. However, erasure of epigenetic marks is not complete, as populations of sperm RNAs are present and are found to be functional important in the post-fertilization zygote [156, 158, 183]. Moreover, a subset of parental imprints and methylated loci near transposon elements acquired during gametogenesis are also faithfully retained in the zygote and are proposed to have pivotal roles in regulating and support preimplantation embryo development [184, 185]. On the other hand, if aberrant epigenetic marks are retained in the developing embryo, it can affect the trajectory of offspring development and health and contribute to the inheritance of altered epigenetic states (fig 5, 6).

Indeed, several experimental studies have reported transgenerational effects of epimutations in response to a wide spectre of environmental toxicants (reviewed in [95, 153], and psychological stress [186]. However, as there is a substantial difference in the clonal expansion of rodent and human male germ cells, cross-generational effects of environmentally induced sperm epimutations in mice may not necessarily translate to humans. Whereas one rodent spermatogonium undergoes intense cell divisions, and may produce over 4000 spermatids, one human spermatogonium produces 32 spermatozoa [187]. Consequently, if an epigenetic error occurs during early germ cell divisions, and is not corrected during the fate of gametogenesis, this impact would be expected to be much more pronounced and affect many more sperm cells in mice than in humans [92].

1.6 Cigarette smoke exposure and the sperm epigenome

Since the early planning and development of this PhD project in 2011 there has been an accumulating body of evidence on the mechanistic pathways by which cigarette smoking metabolites influence the epigenetic components of sperm [188]. It is well established that constituents of tobacco smoke can penetrate the blood-testis barrier

[189-191], and can cause reduced sperm quality, altered sperm function and impaired reproductive health through increased levels of seminal reactive oxygen species (ROS) and oxidative stress [192-194]. However, with the recent advances in omics technologies and the profiling of molecular factors and regulatory mechanisms underlying spermatogenesis, it has become increasingly apparent that the adverse effects of cigarette smoke also have severe implications for the epigenetic reprogramming machinery during spermatogenesis, and may negatively affect the paternal contribution to embryo development and subsequent postnatal health [192-196].

Several human methylome analyses have identified genome-wide alterations in sperm methylation profiles of smokers [197-201], and that DNA methylation changes in spermatozoa of smoking men also affect the transcriptional level of sperm genes [202]. Smoking-induced DNA methylation changes in mice testes have been shown to alter the expression of proteins involved in spermatogenesis [203, 204].

Moreover, cigarette smoking is correlated with aberrant histone-protamine transition and transcription of protamine genes, which may lead to a defective chromatin condensation and spermatogenesis [146, 205]. Smoking-induced alterations in the expression and function of noncoding RNAs may also negatively affect spermatogenesis through alterations in the expression and function of noncoding RNAs [206], and smokers have been found to exhibit altered miRNA and mRNA expression patterns of miRNAs involved in regulating signalling pathways that have pivotal roles for preserving sperm integrity and normal embryonic development [207, 208]. These findings clearly demonstrate that the toxicogenic effects of cigarette smoke metabolites can modulate the sperm epigenetic system and induce aberrant epimutations, which in turn may adversely impact on embryogenesis and early life development in the next generation.

2. Rationale for thesis and study aims

The mechanistic rationale for this thesis arises from the supposition that the male sperm epigenome, although being essential for regulation of transcriptional activity during gametogenesis and in the early zygote, is yet highly plastic and responsive to tobacco smoke constituents throughout reprogramming stages during germ cell development. Consequently, there may be several exposure sensitive periods throughout a man's life where the chemical components in tobacco possibly can affect epigenetic states and give rise to pleiotropic effects, not only in himself, but also in his future offspring if transmitted to the next generation at fertilization.
Although Epigenome-Wide Association Studies (EWAS) have identified a substantial number of methylation biomarkers associated with personal and mothers' (in utero) smoking exposure, potential methylation changes in response to fathers' tobacco use is yet little investigated. Moreover, to our knowledge, no EWAS studies have so far time targeted exposures that are thought to concur with epigenetic reprogramming events when sperm maturate and enter spermatogenesis in adolescent and preconceptual years.

The present PhD work has aimed to combine epigenetic and epidemiological data and analyses to investigate potential mechanisms on how paternal smoking trajectories can affect the development and phenotypic variation in his offspring.

2.1 Main and specific objectives

The overall objective of the thesis has been to identify and characterize associations between paternal smoking at different preconception and postnatal time points in relation to genomic methylation patterns and phenotypic outcomes in his offspring.

We have sought to address this by the following specific objectives:

1. To investigate a potential association between fathers' smoking exposures and offspring DNA methylation, and to explore the biological impact of methylated loci and annotated genes. (Paper I)

2. To investigate parental smoking onset in specific preconception and postnatal time windows in relation to offspring body mass index (BMI) and in a subsample, fat mass index (FMI), and to explore whether associated outcomes are modified by offspring sex or mediated by the following factors: parental pack years of smoking, parental BMI, offspring smoking, and offspring birthweight. (Paper II)
3. To identify DNA methylation changes in offspring associated with fathers' smoking commencing at any time during preconceptual years, and in a subsample, fathers' smoking in early adolescence (before age 15), and to further investigate whether the identified differentially methylated sites (dmCpGs) are different from methylation signals related to personal and maternal smoking and additionally, whether they are associated with BMI and respiratory outcomes in the offspring (Paper III).

3. Method summary

3.1 Cohort descriptions

This thesis used data from three international population-based studies- the European Community Respiratory Health Survey (ECRHS), the Respiratory Health In Northern Europe study (RHINE), and the RHINESSA generation study. An overview of centres for the three cohort studies is given in figure 6. Inclusion criteria for the individual papers are given in figures 7-9.

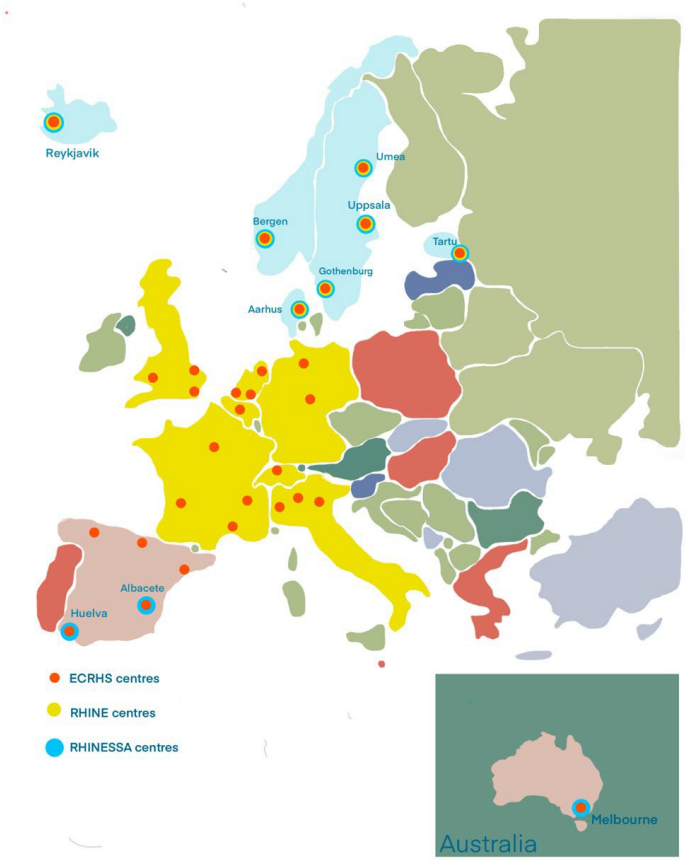


Figure 6: The red dots represent ECRHS centres, yellow dots represent RHINE centres, and blue dots represent centres involved in the RHINESSA study. The light blue areas represent the Northern European countries with study centres involved in all three cohort studies, whereas the pink areas represent countries and study centres engaged in ECRHS and RHINESSA. Last, the yellow areas illustrate countries that solely have ECRHS study centres.

3.1.2 European Community Respiratory Health Survey (ECRHS)

In 1990-1994 the ECRHS was conducted; a multi-centred study in 16 European countries and beyond, with the overall aim to describe variations in childhood or current exposure to risk factors proposedly or known to be important for the development of asthma and allergy. Adult men and women, 20-44 years of age, from general populations were randomly selected to summon an initial questionnaire stage. A subsample was subsequently invited for clinical examination with, amongst other

things, collection of blood samples, anthropometric measurements and lung function testing (www.erhs.org). Responders from 29 centres were thereafter followed-up with questionnaires and clinical investigations in ECRHS II (1998-2004, mean follow-up time 9 years) and ECRHS III (2011-2014, mean follow-up time 11 years). Response rate for follow-up after 20 years was 49 % [209].

3.1.3 Respiratory Health in Northern Europe study (RHINE)

RHINE is a questionnaire-based follow-up survey of Northern European responders from seven study centres (Bergen in Norway, Umea, Uppsala and Gothenburg in Sweden; Aarhus in Denmark; Reykjavik in Iceland; and Tartu in Estonia) who initially participated in the questionnaire stage of ECRHS I. The study population was followed-up with questionnaries in two subsequent waves, RHINE II (1999-2001, response rate 75 %), and RHINE III, 2010-2012, response rate 61 %) [210] (www.rhine.nu). At each study wave, information on lifestyle habits, body composition, sociocultural factors and childhood and adult environmental exposures were collected.

3.1.4 The RHINESSA study

The RHINESSA study is an international research project aiming to investigate the offspring of participants from the seven RHINE study centres and three additional ECRHS centres (Huelva and Albacete in Spain, and Melbourne in Australia) (www.rhinessa.net). Offspring were sent web-based and/or postal questionnaires that had been harmonised with RHINE protocols in the years 2013-2015 (response rate 34.7 %) [211]. A subsequent subsample of offspring with parents who had provided clinical information in ECRHS, were also invited for clinical investigation and interview according to standardised protocols that had been harmonised with the ECRHS protocols (response rate 34.9 %) [211].

3.2 Selection of study subjects

In paper I, we used DNA data from 100 subjects who had participated in the clinical phase of ECRHS II, and 95 participants enrolled in the RHINESSA study. For all 195 subjects, blood samples had been collected at the Bergen study centre. In the present analysis, information on offspring from the two population cohorts were merged and analysed together (figure 7). Personal smoking information was obtained from interview performed in the ECRHS (appendix 1) or RHINESSA study, respectively (appendix 2). For RHINESSA offspring, parental smoking information was retrieved directly from their parents who participated in ECRHS. For the offspring who were ECRHS participants, on the other hand, parental smoking information was provided by themselves.

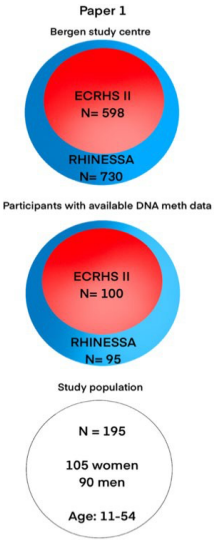


Figure 7: Schematic illustration of eligible and included study subjects in paper 1, and the population cohorts and study centre they originated from.

In paper II, we used parental data from 2111 fathers and 2569 mothers who had participated in RHINE II or ECRHS II, and who had provided extensive information on smoking habits (appendix 3 and 1 for RHINE and ECRHS questionnaires/interview, respectively). Information on their 6487 offspring, 2777 sons and 3710 daughters, was obtained from the RHINESSA study (figure 8).

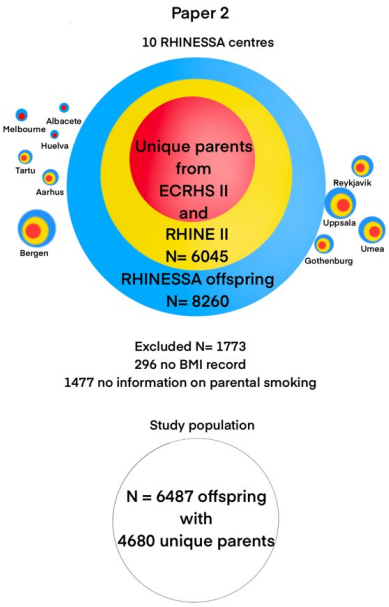


Figure 8: Schematic illustration of eligible and included study subjects in paper 2, and the population cohorts and study centres they originated from. The size of the study centre dots reflects the number of participants from that study centre.

In paper III, the study population for EWAS 1 comprised 875 offspring from six RHINESSA study centres with available peripheral blood data for DNA methylation measurement and who had provided complete information on parental and personal smoking. The subpopulation in EWAS 2 comprised 304 offspring. Information on their parents was collected from questionnaire data from RHINE II (appendix 3) or standardised interviews (appendix 1) in the ECRHS II studies (figure 9).

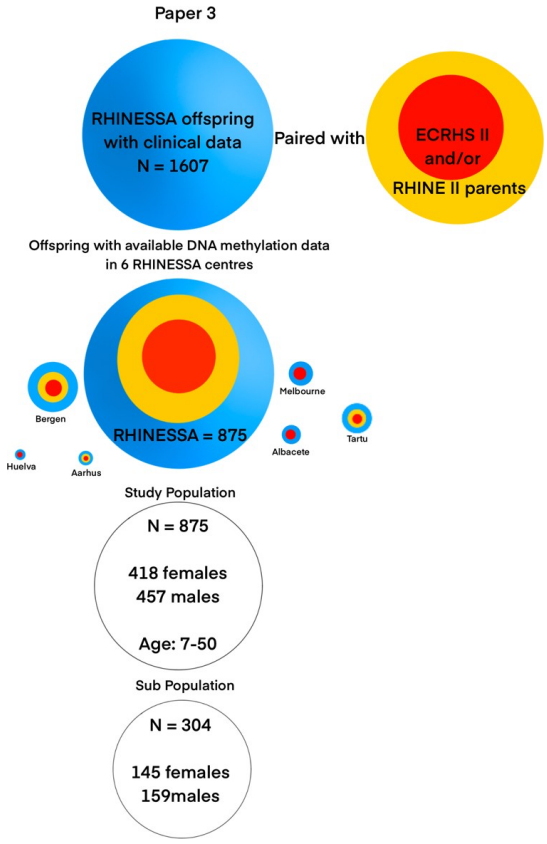


Figure 9: Schematic illustration of eligible and included study subjects in paper 3, and the population cohorts and study centres they originated from. The size of the study centre dots reflects the number of participants from that study centre.

3.3 Exposure variables

The main exposure of interest throughout this thesis has been parental smoking trajectories – and most specifically fathers' smoking. In **paper 1**, we defined fathers' smoking as a binary exposure variable, and classified it as either ***(i) ever having smoked or (ii) never having smoked during offspring's childhood***. Information on fathers' smoking habits was in the ECRHS cohort based on offspring's responses to to the question *"did your father ever smoke regularly during your childhood?"*. In the RHINESSA cohort, information on fathers' smoking exposure was obtained from longitudinal data given by their fathers themselves when participating in the ECRHS, and responding to the question "*have you ever smoked for as long as a year?"*.

In **paper 2**, we extended the smoking exposure to additionally include mothers' smoking trajectories, and we obtained detailed information on parental smoking habits based on their responses to the following questions in ECRHS: *i. "Are you a smoker?" ii. "Are you an ex-smoker?" iii. "if yes, how old were you when you started smoking?"* (given in whole years) *iiii. "Smoked for …years" iv. "Stopped smoking in (year)"*. Ever-smokers were categorised according to whether smoking commenced in early (<15) or later adolescence (≥15), and whether they started to smoke in preconception years (defined as smoking initiation at least 2 years before the offspring's birthyear), or after the offspring was born (defined as at least 1 year after the offspring's birthyear). This yielded a four-level exposure variable with the following categories: ***(i) never smoked; (ii) started smoking before age 15 years, (iii) started smoking between age 15 years and prior to conception of offspring, and (iv) started smoking after birth of offspring***. Maternal and paternal lines were investigated separately.

In **paper 3**, fathers' smoking exposure was determined by their responses to the following questions in ECRHS: *i."Have you ever smoked for as long as a year?" ii."If yes, how old were you when you started smoking?"* (given in whole years). *iii. "Have you stopped or cut down smoking?" iv. "How old were you when you stopped or cut down smoking?"*. By relating this information to offspring's birthyear, and whether smoking started before conception (≥2 years before offspring birthyear), or

after the offspring was born (≥1 year after the offspring birthyear), we constructed the following two sets of fathers' smoking exposure groups: In **EWAS 1** we classified fathers' smoking exposure as ***(i) any preconception smoking onset versus (ii) fathers' postnatal smoking onset or never smoking***. In **EWAS 2** fathers' smoking exposure were categorised as ***(i) preconception smoking before age 15 or (ii) never smoking***.

3.4 Outcome variables

3.4.1 DNA methylation

In paper 1 and 3 the primary outcome was offspring DNA methylation.

3.4.1.2 Laboratory processing

Briefly, DNA was extracted from peripheral blood, by use of a standard salting out procedure [211, 212]. DNA was bisulfite-converted[1] at the Oxford Genomics Centre (Oxford, UK) using the EZ 96-DNA methylation kit. DNA was hybridized to Infinium MethylationEPIC Beadchips, which capture more than 850.000 methylation sites per sample, and scanned according to the manufacturer's protocol (Illumina, Inc. CA, USA). To control for batch effects[2], all samples were randomly distributed onto the BeadChips, which accomodated eight samples per chip and DNA methylation was measured using the Infinium protocol.

3.4.1.3 Methylation quality control and normalisation

Methylation data was evaluated, quality controlled and filtered based on standard approaches. CpG probes were eliminated if (1) their intensity levels were at or near

[1] Bisulfite ions deaminates unmethylated cytosines into uracils while methylated cytosines remain unchanged. Uracils converts to thymines during subsequent polymerase chain reaction (PCR) amplification. This provides a basis for how methylation arrays can distinguish between methylated (C) and unmethylated (T) CpG sites.
[2] Batch effects are systematic, non-biological differences between groups of samples that are related to experimental factors and sampling handling, such as laboratory conditions, experiment time and chip position.

the background intensity levels and therefore had non-significant detection p-values[3] (2) if they failed to measure DNA methylation in less than 2% of the samples, and thus had a call rate <98% or (3) if three or more beads affixed to the probe in more than 5% of the samples failed detection on the array (4) they were located on X and Y chromosomes and therefore had a different methylation value distribution to that of autosomal chromosomes and (5) they were located to DNA sequences with known single nucleotide polymorphisms (SNPs), which may disrupt probe binding at the site and lower the intensity signals, or (6) were known to hybridise at multiple genomic locations.

In paper 1, probes that did not reach a detection p value of 10E-16 were set to missing, and samples with call rates less than 98% were excluded. After probes located on sex chromosomes, and at 0 distance to known SNPs were removed, 765,082 CpG sites remained for subsequent analysis.

In paper 3, probes were excluded based on the following criteria: detection p-value above 0.01, probes associated with SNPs, probes with a beadcount below three in at least 5% of samples, probes at multiple locations, non-cg probes, probes on X or Y chromosomes, and cross-reactive probes[4] on the microarray [213]. Following processing steps, 726,661CpGs were retained for analysis.

3.4.1.4 Data normalisation

Data normalisation was applied to minimise unwanted variation within and between samples, and to reduce signal bias between the two probe design types.[5] The

[3] Various processing pipelines apply different significance thresholds for detection, i.e., from p<10E-16 to p<0.05. Very small detection p-values are indicative of a reliable signal, whilst large p-values generally indicate a poor signal quality.

[4] Cross-reactive probes are probes that map to multiple places in the genome.

[5] The Methylation EPIC microarray uses two different probe types (Type I and Type II) with distinct experimental characteristics. Type I probes have two beads on the same color channel (one for each of the methylated and unmethylated signal). Type II probes have one bead and two color channels (one color for each of the methylated and unmethylated signal). The two probe types have different dynamic range and therefore display different distributions for DNA methylation.

Methylation EPIC microarray uses oilgonucleotides coupled to beads in a single nucleotide extension assay to detect methylation at a single locus. As the magnitude of the measured methylation depends on the underlying bead assay chemistry, normalisation steps are crucial to avoid any enrichment driven by probe type in the differential methylation analysis. Both quantile normalisation and beta mixture quantile dilatation (BMIQ) are commonly used to reduce between sample variation and to reduce the bias between probe types. To further reduce technical variation and minimize confounding between potential sources of batch efffects, batch-adjustment tools, such as ComBat were also applied to the data.

Following data normalisation, the proportion of DNA methylation at a particular CpG site (beta value) was calculated from normalised intensity values, by taking the ratio of the methylated (C) to the unmethylated (T) signal. Thus the methylation level estimates represented the percentage of methylation for each CpG site, and followed a beta distribution ranging from 0 to 1, with 0 inferring an unmethylated signal, and 1 representing a fully methylated CpG site.

In paper 1, the CPACOR pipeline [214] was used to pre-process and normalise the methylation data. Illumina background correction[6] was applied to all intensity values, and Limma was applied for quantile normalisation to minimize unwanted variation within and between samples, and to account for different probe types. ComBat was used to correct for batch effects [215].

In paper 3, Minfi [216] and Meffil [217] packages were used to process and quality assess the methylation intensity data, and BMIQ was applied for normalisation [218]. ComBat was used to correct for batch effects [219].

[6] Background correction can be performed to correct for non-specific/random contributions to the overall signal.

3.4.1.5 Cell-type adjustment methods

Molecular profiles derived from peripheral blood samples represent a mixture of many different cell types. Given that methylation signatures are cell-type specific, cellular heterogeneity between study participants can confound methylation results in EWAS studies, if not accounted for. Commonly applied correction methods are reference-based deconvolution algorithms, which use a *priori* database of cell-type specific methylation signatures to infer cell-type composition. An example of such an approach is the algorithm by Houseman [220], in which cell proportions of white blood cells populations (CD8T cells, CD4T cells, Natural Killer cells, B cells, monocytes, and granolucytes) are predicted and estimated. The cell composition coefficients that are derived from the algorithm can subsequently be accounted for in the methylation analysis. This cell-type adjustment method was applied in paper 1. In paper 3, the more recent reference-based approach in EpiDISH [221] was used to obtain estimates of cellular proportions (CD8T cells, CD4T cells, Natural Killer cells, B cells, Monocytes, Neutrophils). This method uses a reference database with cell-type specific methylation data from the NIH Roadmap and ENCODE and applies a robust partial correlation (RPC) technique to estimate cell-type proportions.

3.4.2 Body mass index and fat mass index

In paper 2 our main outcomes were body mass index (BMI) and fat mass index (FMI). BMI was calculated from self-reported height and weight using the formula [weight(kg)/height(m)^2]. Fat mass was estimated from bioelectrical impedance data, measured using BodyStat 1500 MDD (https://www.bodystat.com/medical/). FMI was calculated by the formula [fat mass(kg)/height(m)^2].

3.5 Covariates

3.5.1 Paper 1

Information on mothers' smoking was collected from the participants' responses to the questions *"Did your mother ever smoke regularly during your childhood"* (Appendix 2, RHINESSA questionnaire) or *"Did your mother ever smoke regularly during your childhood, or before you were born?"* (Appendix 1, ECRHS questionnaire). Mothers' smoking was classified as either having smoked or never having smoking during offspring's childhood.

Personal smoking was categorised as current, ex or never smoking, and was based on the questions *i."Have you ever smoked for as long as a year?" ii."If yes, how old were you when you started smoking?" iii."Have you stopped or cut down smoking?" iv."How old were you when you stopped or cut down smoking?"* (same phrasing for both study cohorts, RHINESSA and ECRHS main questionnaires, appendix 2 and 1, respectively).

Parental education was classified as lower (primary school), intermediate (secondary school) and higher education (college or unversity), and was based on the questions *i."What was the highest level of education your mother had?" ii."What was the highest level of education your father had?"* (Same phrasing in both study cohorts, appendix 3 and 1 for RHINESSA and ECRHS questionnaires respectively). Fathers' educational level was used as a surrogate measure of socioeconomic status.

Personal education was classified in concordant categories according to the participants' responses to the questions *"At what age did you complete full time education?"* (ECRHS questinnaire appendix 1) and *"Please mark the educational level which best describes your level"* (RHINESSA questionnaire, appendix 2).

The participants' age as well as their fathers' age, were estimated from their date of birth. In the present study we included the offspring's age at the time they undertook the clinical assessment, and their fathers' age the year the offspring were born.

3.5.2 Paper 2

Parental and offspring education was categorised as lower (primary school), intermediate (secondary) or higher (college or university) based on their responses to the questions *"At what age did you complete full time education?* (ECHRS questionnaire appendix 1) and "*Please mark the eduactional level which best describes your level"* (same phrasing in RHINE and RHINESSA questionnaires, appendix 3 and 2 respectively). Educational attainment was further used as a proxy for socioeconomic status and included in the analysis as a potential confounder.

The following mediators were included in the paper:

Parents' pack years 1) preconception and 2) up to the offspring's age 18 were constructed by multiplying the number of 20-packs of cigarettes smoked daily 1) by the number of years they had smoked up to ≥2 years before the offspring's birth year and 2) from the offpsring's birth year up to their eighteenth birth year.

Parental BMI was calculated from self-reported height and weight when participating in RHINE III or ECRHS III using the formula [weight(kg)/height $(m)^2$].

Offspring's own smoking was categorised as either ever having smoked (current or ex-smoker), or never having smoked, and was based on the questions *i."Do you smoke?" ii."Did you smoke previously?"* (RHINESSA questionnaire, appendix 2). (Same phrasing in both study cohorts, appendix 2 and 1 for RHINESSA and ECRHS questionnaires, respectively).

Offspring birthweight were collected from national registry data for a subsample of 813 mother-offspring pairs.

3.5.3 Paper 3

Personal smoking was categorised as ever having smoked (current or ex smoker) or never having smoked, and was based on the questions *i."Have you ever smoked for as long as a year?" ii."If yes, how old were you when you started smoking?" iii."Have you stopped or cut down smoking?" iv."How old were you when you stopped or cut down smoking?"* (RHINESSA questinnaire, appendix 2). Personal smoking was both included as a potential confounder in the EWAS analyses on fathers' smoking, as well as an exposure variable in the EWAS on offspring's own smoking in relation to their methylation patterns.

Information on mothers' smoking was collected from the offspring's responses to the question *"Did your mother ever smoke regularly during your childhood"* (Appendix 2, RHINESSA questionnaire), and was classified as ever having smoked or never having smoked during the offspring's childhood. Maternal smoking was also included as both a potential confounder in the fathers' smoking EWAS analyses, and an exposure variable in the EWAS on mothers' smoking exposure in relation to offspring's DNA methylation.

Offspring's age was estimated by date of birth, and it was included in the analysis as their age when undertaking the clinical examination. Offspring age was both included as a potential confounding variable in the EWAS analyses, as well as in sensitivity analyses investigating correlations with fathers' smoking associated dmCpGs as well as age-related CpG markers (cg1686765, cg24724428 (ELOVL2); cg22454769 (FHL2) and cg131083 (DNAH9)) identified from an EWAS in the RHINESSA cohort for age .

Fathers' educational level was categorised as lower (primary school), intermediate (secondary) or higher (college or university) based on their responses to the questions *"At what age did you complete full time education?* (ECHRS questionnaire appendix 1) and "*Please mark the eduactional level which best describes your level"* (RHINE questionnaire appendix 3). Educational attainment was further used as a proxy for socioeconomic status and included as an exposure variable in a sensitivity regression analysis with fathers' smoking associated dmCpGs as outcome measures.

Similarly, we conducted sensitivity regression analyses to investigate whether fathers' smoking associated dmCpGs were related to the following offspring variables: weight, BMI, wheeze and asthma.

Clinical data on offpsring's weight and height was based on antropometric measures collected at the time of clinical examination. BMI was calculated by the formula [weight(kg)/height(m)^2]

Information on offspring's wheeze and asthma symptoms was based on their own responses to the questions *"Have you ever had wheezing or whistling in you chest?"* and *"Do you have or have you ever had asthma?"* (RHINESSA questionnaire, appendix 2), and was further classified as ever having had wheeze or asthma or never having had wheeze or asthma.

3.6 Replication cohorts

3.6.1 The Isle of Wight (IoW) Cohort

In paper 1, we pursued replication of our findings in an independent study population based on the IoW third-generation study. This prospective multicohort study was established to investigate the natural history and risk factors for the development of asthma, allergic diseases and lung function over three generations. The first generation (F0) was enrolled between the years of 1989 and 1990 and at the time of childbirth of the second generation (F1), which constituted the original Isle of Wight birth cohort (n=1536) (IOWBC). Since 2010 and onwards, the children (F2) born to the second-generation parents have been enrolled [222]. Participants from the F1 and F2 generation have been extensively assessed at multiple timepoints through clinical examinations and interview/questionnaires. The present replication cohort comprised 159 study subjects with available DNA methylation measurements, obtained from cord blood DNA, and assessed using the Illumina Infinium HumanMethylation450 Beadchip array. Information on their fathers' smoking exposure were collected from responses given by the fathers themselves in the IOWBC.

3.6.2 The Avon Longitudinal Study of Children and Parents (ALSPAC) cohort

In paper 3, we pursued replication in the ALSPAC longitudinal birth cohort. This multigenerational prospective observational study was established to study health and developmental outcomes throughout the lifecourse of parents and children [223, 224]. The study recruited pregnant women who attended health districts in the Bristol area of the UK between the years of 1991 and 1992, and in a subsequent enrollment period (n= 15,247). Participants have been extensively investigated through clinical examinations and self- and parent-reported questionnaires. The present replication cohort comprised 542 participants with DNA methylation measurements obtained from peripheral blood, and assessed with the Infinium MethylationEPIC beachip array at age 15-17. Information on their fathers' smoking exposures and age of smoking initiation was based upon their own responses in the study.

3.7 Statistical analyses

3.7.1 Paper 1

For identification of Differentially methylated regions (DMRs), which comprise of multiple successive differentially methylated positions (DMPs or CpG sites) we used the Python based software tool, Comb-p [225] (Python version 2,7). This approach accounts for non-evenly spaced spatial correlation between CpG sites by first combining the separate P values derived from the site-specific methylation analysis with an autocorrelation adjusted test called the Stouffer-Liptak-Kechris (slk) correction. It then employs a peak detection method to identify potential DMRs, and autocorrelates adjacent P values within each region by applying the Stouffer-Liptak-Kechris (slk) correction for each region. After an adjusted P value is assigned to each

dynamically sized region, a one-step Sidak correction[7] is performed to adjust for multiple testing. Regions that contained at least two probes and had a Sidak-corrected P value<0.05 were considered statistically significant.

To identify differentially methylated probes (DMPs), we applied robust multivariate linear regression models[8], which were adjusted for fathers' and offspring age, offspring sex, offspring and mothers' smoke exposure, and cell type proportions.

Multiple hypothesis testing was accounted for by controlling the false discovery rate (FDR)[9] using Benjamini and Hochberg's algorithm [226]. CpG sites with FDR-corrected P-value<0.05[10] were considered to be statistically significant.

To adjust for inflation[11] and bias[12] of test statistics, which are presumably thought to arise from unmeasured technical and biological confounding, we applied the correcting procedure implemented in the R/bioconductor package BACON [227]. Based on the observed test statistics, BACON estimates the empirical null distribution[13] and obtains estimates to correct for the amount of inflation and bias, represented by the distribution's mean and standard deviation, respectively.

In sensitivity analyses on top CpGs, we accounted for study participants who originated from the same family, and performed linear mixed model analysis with

[7] A method to control the family-wise error rate (FWER). i.e., the probability of making false discoveries/type I errors and mistakenly reject an actually true null-hypothesis when performing multiple hypotheses test.

[8] Compared to standard linear regression, robust linear regression is less sensitive to outliers and data points that significantly differs from other observations.

[9] FDR is the expected proportion of false positive findings among all the rejected hypotheses (false positives and true positives). Thus, FDR controls for a low proportion of false positives, and provides a less stringent method for controlling type I errors compared to other FWER controlling procedures which guard against making any false positive conclusions at all. FDR correction methods therefore have greater power.

[10] A FDR corrected p-value of 0.05 implies that 5% of significant tests will result in false positives.

[11] Inflated test statistics overestimates the level of statistical significance and therefore increase the number of false positive findings.

[12] Bias of test statistics affects the distribution of effect sizes

[13] Multiple and large scale testing situations permit an empirical estimation of the null distribution which can be considerably different and more dispersed than the usual theoretical null distribution (the probability distribution of test statistics when the null hypothesis is true).

family id as random effect[14]. We additionally tested whether top CpGs were potentially confounded by social class, by adding fathers' educational level as a proxy and covariate for socioeconomic status.

For CpG annotation, we used the UCSC Genome Browser provided in the Infinium MethylationEPIC manifest, and SNIPPER (version 1.2, http://csg.sph.umich.edu/boehnke/snipper/) to annotate the nearest gene within 10 megabases (Mb)[15] of each CpG.

To gain insight into the regulatory context of the top differentially methylated probes, we applied Enrich [228] to investigate potential enrichment in annotated regulatory targets identified from TF Chip-seq and histone modification data from the Encyclopedia of DNA elements [229] (ENCODE) and the NIH Roadmap Epigenomics Mapping Consortium [230].

To investigate whether annotated genes were overrepresented in any biochemical or cellular pathways, we employed KEGG (Kyoto Encyclopedia of Genes and Genomes) [231] and GO (Gene Ontology) [232] databases implemented in the gometh function in the R package missMethyl [233].

Replication of top CpGs was pursued in a comparable subsample from the IoW cohort (n=159).

3.7.2 Paper 2

We analysed maternal and paternal lines separately. We used generalized linear regressions[16] to investigate the associations between parental smoking at different timepoints and offspring BMI, and in a subsample (n=240), FMI.

[14] In addition to fixed effect (variation) predictors, linear mixed models additionally account for random variation of the independent variables where it is expected that observations within groups, such as family origin, may be correlated.
[15] A megabase has a unit of length equal to one million base pairs
[16] Generalized linear models have less stringent assumptions and allow for data to be nonlinear and heteroscedastic, i.e., with non-constant variance.

We applied a Directed Acyclic Graph (DAG) (figure 10) to identify and explore potential variables that could confound the relationship between parental smoking onset and offspring BMI/ FMI, i.e., being a common cause of both the exposure and the outcome, and therefore important to control for in the regression analyses (e.g., parental education, parental age, offspring education, the other parent's[17] smoking habits and BMI).

We adjusted for parents educational level, and applied a two-dimensional clustering by study centre and family. We estimated clustered robust standard errors using the R package jtool [234] and employed the HC1[18] variant to account for heteroskedasticity across clusters of observations. We included offspring sex as an interaction term, with significance level for interaction effect set at 0.05, to investigate whether the associations with fathers' or mothers' smoking on offspring BMI/ FMI varied between daughters and sons.

We also applied the DAG to identify potential intermediate variables that could affect the association between parental smoking and offspring BMI, i.e., being mediators that might be caused by the exposure while also actively affecting the the outcome. We constructed mediation models with the following potential mediators: i. parental packyears, ii. parental BMI, iii. offspring smoking (dichotomized as never- and ever smoked), and iv. offspring birthweight (available for a subsample).

[17] The other parent refers to information on parents who did not participate in RHINE/ECRHS themselves, but instead were based on responses given by the offspring.
[18] There are many heteroskedasticity consistent covariance matrix estimators, ranging from HC0 to HC5 which all aim to correct for underestimation of the true variances. HC1 adjust for degrees of freedom and is a commonly used robust standard error estimator, as well as being Stata's default robust option.

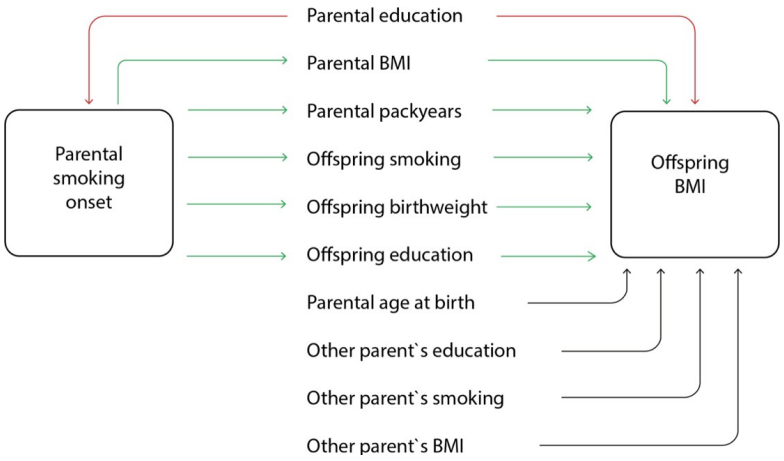


Figure 10: The figure presents potential covariates (confounders (marked with red arrows), mediators (marked with green arrows, and other baseline covariates (marked with black arrow) that were considered for inclusion in the statistical regression models. Here exemplified with the outcome measure BMI.

As a first step in the mediation analyses, we conducted regression analyses to investigate whether there was mediator-outcome or exposure-mediator confounding that would violate the assumptions required for estimating mediation effects. We applied the R package Medflex [235], anchored within the counterfactual framework[19] which provide functions to fit natural effect models[20] regardless of the data distribution. This approach facilitates for inference and parameterization of the pathways by which an exposure affects an outcome in non linear settings by decomposing total effects into natural direct, i.e., the exposure effect that does not go through a given set of potential mediators, and indirect components, i.e., the exposure effect that goes through a given set of potential mediators.

[19] Within the context of the counterfactual framework, the causal effect of the exposure on the outcome is conceptualized as the hypothetical contrast between two or more counterfactuals (potential outcomes) under alternative exposure levels.
[20] Opposed to controlled effects where the value of the mediator is fixed and assumed to be the same for all subjects, natural effect models provide a mediation tool that allows the value of the mediator to vary as a function of the exposure.

We chose an imputation-based[21] model for expanding[22] and imputing[23] the dataset, and fitted working models for the outcome mean for each of the potential mediators conditional on parental education and offspring sex. Separate natural effect models were then specified to enable estimation of natural direct and indirect effects[24]. To account for correlation due to duplicated observations in the extended dataset, we specified robust variance estimates, based on the sandvich estimator, and generated confidence interval plots, to visualise effect estimates and their uncertainty.

In all mediation models we additionally wanted to explore whether direct or indirect effects differed between male and female offspring, and fitted a new set of mediation models with offspring sex included as an interaction term (significance level set at 0.05). All analyses were performed using R, version 3.5.2, downloaded at the Comprehensive R Archive Network (Cran) at http:/www.R-project.org/.

3.7.3 Paper 3

We ran two separate EWAS to identify differentially methylated probes (DMPs) related to fathers' any preconception smoking, and fathers' adolescent smoking commencing before age 15, respectively. We constructed robust multiple linear regression models on methylation beta values, using the limma package [236], and fitted models adjusted for the following offspring covariates: sex, age, smoking habits and cell-type proportions (B-cells, Natural killer cells, CD4 T-cells, CD8 T-cells, Monocyte, Neutrophils), mother's smoking and study centre. In additional analyses, associations between fathers' preconception smoking and offspring's DNA

[21] An imputation-based approach requires fitting a working model for the outcome mean and does not require specification of a mediator model. It can therefore deal with both categorical and continuous mediation variables.
[22] The dataset was expanded by constructing replicates for each subject, corresponding to the number of unique levels of the categorical exposure variable
[23] Counterfactual outcomes for each subject were imputed, based on the observed (X) and all other potential exposure combinations, enumerated in (X*).
[24] Direct and indirect effect estimates were given by the coefficients of the observed (X) and potential (X*) exposure, respectively.

methylation were also stratified by offspring sex. CpG sites with FDR-corrected P-value<0.05 were considered to be statistically significant.

Manhattan plots displaying the significance level of each CpG site associated with fathers' smoking and its chromosome location throughout the genome, were generated using the R package CMplot [237].

We applied BACON [227] to adjust for inflation from systematic bias.

We applied dmrff [238] and DMRcate [239] to identify differentially methylated regions (DMRs). Dmrff employs an inverse-variance weighted[25] meta-analysis of EWAS effects sizes to account for correlation between CpG sites. In a two-step process, it first identifies candidate regions by grouping nominally significant (P<0.05) CpG sites that have methylation changes in a consistent direction and are within a 500 bp distance of another. Meta-analysis test statistics are then calculated for each region, and sub- regions displaying the strongest test statistics are identified [238].

DMRcate identifies DMRs by combining EWAS summary t-statistics from nearby CpG sites, by using a Gaussian kernel smoother[26] of a specified width. Based on this, the p-values for for each CpG site is recomputed and a new genome-wide significance threshold is selected, corresponding to the number of CpGs that survided multiple testing correction in the original EWAS. Consecutive CpGs with recomputed p-values below this treshold are then identified as DMRs [239].

We applied eFORGE TF [240] to investigate whether differentially methylated sites overlapped with cell-type specific regulatory elements such as transcription factor binding sites, open chromatin states and histone marks.

We compared our identified differentially methylated CpG sites with known biological traits of previously published epigenome wide association studies in the

[25] The most common approach to calculate average effects in meta-analyses. The standard error is squared to obtain the variance of each effect size. Since a lower variance indicates higher precision, the inverse of the variance is used to determine the weight of each study.
[26] Kernels define the shape of the function that is applied when averaging over neighbouring points. The Gaussian kernel has the shape of a normally distributed curve, where points around 0 will be weighted higher than points further away.

EWAS Atlas knowledge base [241], and with gene-disease associations in The Open Targets Platform [242]. We applied STRING [243] to identify dmCpG annotated genes that overlapped or interacted in specific biological functions. To gain further insight into the molecular and biological processes of annotated genes, we performed functional enrichment analyses to detect enriched pathways using UniprotR [244] and gometh [245].

To explore whether DNA methylation marks identified in relation to fathers' preconception and adolescent smoking exposure differed from methylation signals associated with personal and maternal smoking, we contructed two additional EWAS studies to identify differentially methylated sites related to 1) offspring's own smoking exposure and 2) their mothers' smoking exposure. Top hits for the smoking-associated methylation signatures identified in our four epigenome-wide association studies (i.e., fathers' any preconception smoking, fathers' adolescent smoking commencing before age 15, personal smoking and mother's smoking) were compared to previously published meta analysed association results on offspring DNA methylation with personal [246, 247][27], and maternal cigarette smoking [248][28].

Replication was carried out in a comparable subsample from the ALSPAC cohort (N=542), with similar statistical modeling and adjustments including the following covariates: offspring's sex, age, smoking status, predicted cell count proportions, maternal smoking, and batch effects. T-tests were applied to investigate whether the beta coefficients of differentially methylated sites in the RHINESSA study (both at a FDR rate of <0.05 as well as the top 100 dmCpGs) were significantly associated with the dmCpGs obtained in the ALSPAC replication cohort. Signed binomial tests were used to test the strength and direction of the association.

In a sensitivity analysis, we adressed whether the fathers' smoking related CpGs were potentially confounded by the effect of social class, by adding fathers' educational

[27] The meta-analysis by Joehanes et al. 2016, used DNA profiles from 16 cohorts, assessed with the Illumina Methylation bead Chip 450 k array, whereas the meta-analysis by Christiansen et al. 2021, used EPIC DNA profiles from four population-based cohorts.
[28] The meta-analysed association results by Joubert et al. 2016, used DNA profiles measured with the 450 k microarray from 13 cohorts recruited into the Pregnancy and Childhood Epigenetics Consortium (Pace).

level, in being a surrogate measure of socioeconomic background, as an independent variable and regressed with the identified top dmCpGs. The potential impact of offspring's age was also more extensively investigated in subsequent analyses, by correlating known age-related CpG markers from the RHINESSA EWAS study, with both the top CpGs identified as related to fathers' smoking, as well as to the age of the offspring.

We also conducted sensitivity regression analyses to investigate whether the fathers' smoking associated dmCpGs were related to the following offspring variables: weight, BMI, ever-asthma and ever-wheeze. In these analyses, offspring sex was included as a covariate.

3.8 Ethical considerations

The ECRHS, RHINE and RHINESSA studies were approved by regional committees of medical research ethics and adhered to national legislations in each study centre [249]. Data collection complied with the ethical principles for medical reasearch asserted by the World Medical Association Declaration of Hesinki [250]. All participants gave their written informed consent prior to participation, which covered consent to collect questionnaire data, to retrieve data from national registries and to collect clinical data (applicable for a sub-sample). Participants were informed that they could withdraw from the study at any time. The risks and inconveniences for participants were regarded to be minimal as data collection primarily comprised questionnaire and registry data, and low-risk clinical examinations, such as lung function testing.

Appropriate Data Protection measures were highly prioritized to ensure safe storage of information, and to avoid non-authorized access and misuse. The study database was stored on a designated research server at the Haukeland University Hospital, in complience with the hospital's research regulations [251]. The research server was developed by the IT department at the Haukeland University hospital to ensure a secure processing of sensitive personal data for research purposes. The storage system

complied to the "Norwegian Code of conduct for information security in the healthcare and care services sector" [252] and ensured that sensitive personal data preserved its confidentiality, integrity, and availability when being processed.

4. Paper Summaries

4.1 Paper I: Epigenome-wide association of father's smoking with offspring DNA methylation: a hypothesis-generating study

This hypothesis-generating study arises from excisting literature suggesting that paternal line exposures may affect offspring health, and the increasingly genome-wide evidence of altered DNA methylation patterns in response to personal and maternal tobacco smoke exposure. Based on the molecular properties of DNA methylation to be stably propagated during successive cell divisions, we hypothesized it as biological plausible that a father's smoking exposure, beyond affecting his own epigenome, also might transmit to the subsequent generation and affect the methylome in his offspring. We set out to test this hypothesis, by conducting an epigenome-wide association study investigating a potential association between fathers' ever smoking exposure and offspring EPIC DNA methylation profiles[29], using data and peripheral blood samples of 195 male and female adolescent and adult offspring participating in the population-based cohorts ECRHS and RHINESSA.

We used Comp-p to detect differentially methylated regions (DMRs), and robust multivariate linear regressions, adjusted for maternal smoking, paternal age and offspring's sex/age/smoking status and cell-type proportions, to detect differentially methylated probes (DMPs). We adjusted for inflation and bias of test statistics, and

[29] DNA methylation was quantified using the Illumina Infinium MethylationEPIC Beadchip array.

performed enrichment and pathway analyses to explore the biological processes and functions of annotated genes.

We identified six differentially methylated regions (DMRs) associated with fathers' smoking. Five DMRs, which span between 3 and 5 DNA methylation sites, co-localized with genomic regions indicative of a potentially regulatory function, and consisted of consequtive CpGs not previously identified in epigenome-wide studies on maternal and personal smoking. This made us suggest that the smoking exposures from paternal and maternal smoking might influence their offspring's methylome through different biological mechanisms. The novel smoking methylation signatures were annotated to genes involved in innate and adaptive immunity (*ATP6V1E1*, *C2*), lipid metabolism and fatty acid biosynthesis (*ACSF3*), as well as to cellular processes such as cell cycle progression, signal transduction and gene regulation (*WDR60*). We also identified one differentially methylated region that overlapped with a gene previously associated with smoking (*CTNNA2*), whose expression is crucial for neural system development in the brain and also implicated in a spectre of behavioral disorders and addiction.

Although, at a single probe level, none of the differentially methylated probes (DMPs) passed epigenome significance (FDR<0.05) after controlling for genomic inflation (λ=1.46), we observed that 5 of the top DMP annotated genes had similar putative functions as seen in the DMR analysis, specifically in relation to innate and adaptive immunity (*BCAS1*, *MFGE8*, *UNC93B1*, *RALB*), as well as to neural systems and behavioral dysfunction (*DLGAP1*).

Due to differences in the methylation array platforms used in the present study and the Isle of Wight replication cohort, the amount of missing CpG sites in addition to the low number of exposed individuals, made it unfeasable to attain replication of our results.

In conclusion, this epigenome-wide association study is the first to report novel father smoking signatures in offspring of adolescent and adult age. Although this may indicate a potential persistent effect of fathers' smoking exposure on their offspring's

methylome, subsequent studies are needed in order to verify our findings, as are further investigations on whether the methylated loci associated with fathers' smoking are biologically relevant and in fact able to impact on the offspring's phenotypic diversity.

4.2 Paper II: Parents' smoking onset before conception as related to body mass index and fat mass in adult offspring: Findings from the RHINESSA generation study

In part this epidemiological study evolved from findings and biological plausibilities proceeding from our epigenome-wide analysis, and *a priori* hypothesis that putative gene functions related to our novel father smoking DMRs might correlate to certain phenotypic outcomes in the offspring. In part, the rationale for the study also built on prior observations from our research group, where particularly exposures occurring during fathers' prepubertal and pubertal years were related to adverse outcomes in the offspring, thus potentially conferring early adolescence as an exposure sensitive time period of critical importance. This was also accentuated by Northstone and colleagues from the Avon Longitudinal Study of Parents and Children (ALSPAC), who observed striking sex-specific associations, in which sons of early smoking fathers had higher BMI and fat mass in their teens. Given that we in our previous EWAS study had identified fathers' smoking methylation marks related to immunity and metabolic regulating genes, whose putative functions also have been implicated in obesity, we were intrigued to pursue whether we could identify similar sexual dimorphistic patterns of parental smoking exposure as related to adult offspring's owerweight in the population cohorts ECRHS, RHINE and RHINESSA. The present study therefore set out to investigate whether fathers' and mothers' smoking were associated with sex-specific outcomes in their adult offspring's BMI and, in a subsample, fat mass, and whether these associations were more pronounced if smoking commenced in early

adolescence as opposed to initiated at later preconceptional or postnatal time points. Secondly, due to the social patterning and inequalities related to smoking behavior, as well as to the multifactorial aspects contributing to obesity, we also aimed to investigate whether factors such as parental pack years, parental BMI, offspring smoking, and, in a subsample, offspring birthweight, might mediate the potential associations between parental smoking onset at different time points and offspring's body composition. Our study used data from 10 study centres and comprised 4680 unique parents (n= 2111 fathers and n=2569 mothers) enrolled in the RHINE/ ECRHS study and 6487 offspring from the RHINESSA study (n=2777 sons and n=3710 daughters).

BMI was calculated from self-reported height and weight, and FMI was estimated in a sub-sample with available bioelectrical impedance measures. Associations with parental smoking were analysed with generalized linear regressions, adjusted for parental education and clustered by study centre and family. We checked for interaction by offspring sex, and whether the observed associations on offspring's BMI with parental smoking were mediated by parental pack years, parental BMI, offspring smoking and offspring birthweight.

We observed that fathers' preconception smoking onset was associated with increased BMI in adult offspring (onset ≥15 years; β 0.551, 95% CI: 0.174-0.929, p=0.004, n=2916), as well as increased fat mass in his sons (onset <15 years; β 1.604, 95% CI: 0.269-2.939, p=0.019; onset ≥15 years; β 2.590, 95% CI: 0.544-4.636, p=0.013; and onset after birth; β 2.736, 95% CI: 0.621-4.851, p= 0.001, n=129). However, discordantly to Northstone et al. and previous observations in our research group, we did not find this relationship to be more pronounced if the fathers started to smoke in early prepubertal years.

Also mothers' smoking commencing at either preconceptional or postnatal time points was associated with higher BMI in her offspring (onset <15 years; β 1.161, 95% CI: 0.378-1.944, p=0.004; onset ≥15 years; β 0.720, 95% CI: 0.293-1.147, p=0.001; and

onset after birth; β 2.257, 95% CI: 1.220-3.294, p<0.001, n=3531). There was no association between mother's smoking exposure and offspring's fat mass.

Although not possible to pursue in the subsample with fat mass data, independent mediation analysis indicated that the observed associations between parents' preconception smoking onset and adult offspring BMI were fully mediated via the parents' pack years smoked during chilhood years (father onset ≥15 years; indirect effect: β 0.482, p=0.044, mother onset <15 years; indirect effect: β 1.059, p<0.001; mother onset ≥15 years; indirect effect: β 0.833, p<0.001). Moreover our mediation analyses suggested that the association on offspring BMI with parental preconception and/or postconception smoking onset were partially mediated via parental BMI and offspring's personal smoking. There was no effect modification by offspring sex.

In conclusion, we found that both fathers' and mothers' smoking were associated with increased BMI in their adult offspring, yet indicating that the exposure from fathers' smoking may have a particularly profound impact on their sons' fat mass and body composition. However, in contrast to previous reports, our results suggested that the associations between parental preconception and/or postnatal smoking onset and offspring's BMI were mediated via a cumulative smoking exposure during the offspring's childhood. As such, our findings may indicate that the potential long lasting influence on offspring's BMI and risk of obesity in response to parents' smoking exposure could reflect shared familial environments and lifestyle-related factors, or a higher susceptibility to such. Other investigations should explore further the observed association between fathers' smoking exposure and sons' fat mass, particularly given that this is a more specific outcome parameter for assessment of both obesity as well as metabolic health.

4.3 Paper III: Fathers' preconception smoking and offspring DNA methylation: A two generation study

Resting on the recent advances in omics technology and experimental findings suggesting that the male sperm epigenome may be particular exposure sensitive during sperm differentiation and maturation, the rationale for paper III proceeded from the surmise that there may be several periods throughout a man's life where the chemical components in tobacco possibly can affect his sperm epigenetic states and elicit pleiotropic effects, not only in himself, but also in his future offspring if transmitted to the next generation at fertilization. Although we in paper I were able to identify novel differentially methylated regions in offspring with ever smoking fathers, to our knowledge, no EWAS studies have so far time targeted exposures that are thought to concur with epigenetic reprogramming events when sperm maturate and enter spermatogenesis in adolescent and preconceptual years. The present paper therefore set out to examine whether periods of fathers' preconception smoking exposures could be associated with differential DNA methylation in his male and female offspring, whether the identified sites would differ from methylation signals related to personal and maternal smoking, and whether identified signals could be related to respiratory and BMI outcomes in the offspring. We used data and peripheral blood samples from 875 offspring, who originated from six RHINESSA study centres, and who had detailed parental data which had been collected in the ECRHS/RHINE studies.

We ran two epigenome-wide association studies investigating epigenetic signatures in offspring related to 1) fathers' smoking commencing at any time during preconceptual years (n=875), and in a subsample 2) fathers' adolescent smoking onset before age 15 (n=304), using Illumina Infinium MethylationEPIC Beadchip arrays. We constructed robust multivariate linear regressions, adjusted for mothers' smoking, study centre, offspring smoking/sex/age and cell-type proportions, to detect differentially

methylated CpG sites (dmCpGs), and employed dmrff and DMRcate to detect differentially methylated regions (DMRs). In sex-stratified analyses of males (N=457) and females (N=418) we additionally investigated whether the patterns of associations between fathers' preconception smoking and offspring's DNA methyation were different for sons and daughters. We adjusted for inflation, and searched for enrichment of regulatory regions, gene interactions and pathways to gain insight into the molecular and biological processes of the differentially methylated sites and annotated genes.

We investigated whether epigenetic signals in offspring associated with fathers' preconception smoking exposures differed to methylation marks related to personal and maternal smoking, by constructing two additional epigenome-wide association studies investigating 1) offspring DNA methylation in relation to their mothers' smoking exposure and 2) offspring DNA methylation associated with personal smoking.

In sensitivity analyses, we investigated whether the identified fathers' preconception smoking associated dmCpGs were related to the following offspring phenotypic outcomes; weight, BMI, ever-asthma and ever-wheeze.

After adjusting for inflation, we identified 2 differentially methylated CpG sites (dmCpGs) (FDR <0.05 with λ=1.19) associated with fathers' any preconception smoking, and 19 dmCpGs (FDR <0.05 with λ=1.29) associated with fathers' smoking before age 15. Sex-stratified association analyses on fathers' preconception smoking onset identified four male-specific dmCpGs mapped to *KCNJ1*, *GRAMD4*/*DIP*, *TRIM2* and *MYADML2*, and one dmCpG in females located to *LEPROT1* at a FDR level of $\leq$ 0.05. Significant DMRs were not detected in either EWAS. Particularly for fathers' adolescent smoking, several of the differentially methylated sites were enriched for promoter regions, CpG islands and gene bodies, which may add plausability for the CpG sites to have a regulatory role and be functionally important. The novel smoking methylation signatures were distinctly different from the methylation signals we identified in the mother smoking EWAS (14 dmCpGs,

FDR<0.05) and the personal smoking EWAS (33 dmCpGs, FDR<0.05), and annotated to genes with roles in innate and adaptive immunity and inflammatory responses (*TRL9*, *DNTT*, *PSTPIP2*, *CSF1R*), and with glucose and fat metabolic function (*IRS1*). Some of the identified dmCpGs were additionally associated with weight and BMI related outcomes in the offspring (cg03380960: FAM53B, cg12053348 (NA) and cg22402007: NTRK2) and to offspring's ever-asthma (cg22402007: NTRK2) and ever-wheeze (cg11380624: DNAJC14 and cg10981514: TPCN1). When we compared our EWAS results with previously published meta-analyses results on maternal and personal smoking, 16 of our 19 dmCpGs associated with fathers' adolescent smoking had not previously been associated with maternal or personal smoking exposures. In contrast, 10 of our mother smoking associated dmCpGs and 25 of our personal smoking related dmCpGs had also been reported in the meta-analyses results. Subsequent sensitivity analyses of the fathers' smoking related CpGs revealed no confounding by social class, when measured at the level of the fathers' educational attainment, or correlations with the offspring's age (maximum correlation r= 0.2 with 9 CpGs showing correlation at r=0). When pursuing replication of the true positive status of the dmCpGs sites associated with fathers' smoking before age 15 (FDR<0.05) in the ALSPAC cohort, the identified dmCpGs did not overlap, however revealed nominal replication (at a threshold of p<0.05) with similar direction of effects (correlation r=0.49, p=0.12).

In conclusion, our EWAS results showed that fathers' smoking, particularly smoking exposures commencing during early adolescent years, were associated with differentially methylated signatures in the offspring. The identified dmCpGs were distinct to those we identified in relation to mothers' and personal smoking, and as previously reported in meta-analyses on maternal and personal smoking. Moreover, several of the dmCpGs were associated with respiratory and BMI related outcomes in the offspring. We suggest these novel smoking-associated methylation biomarkers may be specific for fathers' preconception smoking exposures occurring in early adolescence. However, as the identified dmCpGs could not be appropriately investigated in an independent replication cohort, further reasearch is needed in order to verify our preliminary findings.

5. Discussion

5.1 Methodological considerations

The fundamental premise of all epidemiological investigations is to produce valid[30] and reliable[31] measurements that are applicable to a more general population than the specific population under investigation. However, all measurements are prone to some degree of error, which can introduce bias and affect the results of an epidemiological study. Thus, awareness of, and means to reduce measurement errors throughout all stages of a study - from the selection of study design and participants, through procedures of data collection and handling, and to the analyses of exposure, outcome and other covariates - is paramount to minimise systematic errors and to assure precise and correct estimates.

5.1.1 Random error (chance)

When drawing an inference of an entire population based on the evaluation of a sample population, the effects of random variation from sample to sample- referred to as the sampling error and measured by the standard error [253], may affect the results and precision of a given exposure -outcome association, and produce an estimate that is different from the true underlying value [254]. Random sampling errors have no preferred direction and may result in both type I[32] and type II[33] errors, and consequently, an over- or underestimation of the true value, respectively [255]. Although the estimate may be imprecise, it is not expected to be inaccurate, and the effect is presumed to negate towards zero, when averaging over a large number of observations [256]. Thus random effects are not generally considered a threat to the

[30] Validity corresponds to a study's ability to accurately measure what it purports to measure, and the degree to which the results are accurate for the study population (internal validity), and generalisable and thus representative of a wider population (external validity).
[31] Reliability refers to the precision and consistency of a measurement, and to which extent the result will be replicated under repeated measurements and among different observers.
[32] Type I errors occur when an actually true null hypothesis is mistakenly rejected and are equivalent to false positive findings.
[33] Type II errors are caused by accepting a null hypothesis when it is not true and are equivalent to false negative findings.

validity of a study. The impact of random error can be decreased with an increased sample size [253].

5.1.2 Systematic error (bias)

Systematic error or bias, on the other hand, is not due to chance alone, but arises from any errors in the collection, analysis, and interpretation of data that systematically distorts the true relationship between a given exposure and outcome [257]. Bias is an inevitable issue in epidemiological research and its net direction and magnitude will not be eliminated when averaging over a large number of observations [256]. Thus bias can lead to inaccurate estimates and poses a threat to the internal validity of a study. Most systematic errors can be attributed to selection bias, information bias and confounding [258].

5.1.2.1 Selection bias

This bias eventuates from errors in procedures used to recruit and select individuals for the study, and from factors affecting the study participation [258]. As a consequence, the participants included in the study will be systematically different from the target population, including those unwilling or failing to respond, which potentially can bias the estimates and distort the exposure-outcome association [259]. For example, more women than men, and individuals with higher educational levels are more likely to participate in survey studies, frequently leading to this kind of selection bias [260, 261].

5.1.2.2 Information bias (misclassification)

Misclassification is a systematic error that occurs during data collection or from inaccurate exposure-outcome assessment, in which an individual, a value or an attribute is classified into a category other than that to which it should be classified in [254]. Misclassification is considered to be random or non-differential if exposure-outcome misclassifications are equal for all the study groups being compared, such as cases and controls, i.e., the probability of an exposure status being misclassified would be independent of the outcome, and vice versa [255]. Conversely, in differential misclassification, the proportion of subjects being misclassified is systematically different between the study groups, because the misclassification is related to the subjects' exposure status, or wether they have or may not have a specific health outcome [255]. This could lead to a biased estimate in the direction of either an over- or underestimation of the exposure-outcome association [253]. A particular source of misclassification in survey studies emanates from recall bias and the ability or willingness of study participants to accurately or completely report past exposures and events [258]. As a notable example, due to increased stigma associated with smoking, are survey based mesures of actual cigarette consumption commonly thought to be underreported and thus prone to misclassification bias [262].

5.1.2.3 Confounding

Bias by counfounding may preclude an actual exposure-outcome association, or more commonly, falsely indicate that such an association exists, because the exposure effect on an outcome is mixed with the effect of an additional factor, or set of factors [263]. Confounding variables are characterised by being associated with both the exposure and outcome, and not being on the causal pathway between the exposure and the outcome [264]. To avoid erroneus conclusions, measurement and quantification of potential confounding should be considered in the implementation and design of a study, and be controlled for during analyses, through statistical adjustment methods and mathematical modelling, such as stratification or multivariate analyses [263].

The following paragraphs will assess the potential impact and sources of random and systematic errors, at first instance within the context of the overall design, recruitment, and data collection of the study populations emplyed in this thesis, and second, according to the study samples of each individual paper.

5.1.3 Cohort designs

This thesis used data from the ECRHS, RHINE and RHINESSA studies, which are multi-centred prospective cohort studies established to study respiratory health and developmental outcomes over time and across generations. Prospective study designs provide a better quality on the temporal sequence between an exposure, a given set of confounding variables and an outcome, as data points and specific expoures in the study population can be gathered prior to the collection of potential outcome information [265]. However, although exposures precede outcomes in time in the three papers included in this thesis, we have not employed statistical approaches to infer causality. Consequently the relationship between the exposures and outcomes of interest is restricted to ascertainment of associations rather then cause or effects.

Multi-centre study designs are faced with challenges related to potential biological and methodological differences between centres, in our case both over time and across generations. Although the heterogeneity in human populations is presumed to increase the random variation and error in the data, which could attenuate an exposure-outcome association, it is not expected to introduce bias and pose a threat to the validity of the results. However, if a study is subject to systematic methodological differences between centres, over time, and across cohorts included in the study, bias can occur and lead to incorrect and invalid estimates. As a mean to reduce bias arising from such systematic discrepancies, standardised and harmonised protocols and questionnaires were developed and applied in all the study centres and across the three population cohorts.

5.1.4 Recruitment of study cohort participants

Participants in the ECRHS and RHINE were initially recruited from random selections of available population registries, with an overall response rate of 78% [266] and 86% [210], respectively. As such, we would anticipate that the study samples to a large extent were representative and generalisable of the populations from which they were drawn, given they were within the same age range (20-64) and demographic areas (mainly European and Nordic countries) as those eligible for, and included in the study. However, the studies do face some degree of selection bias from factors affecting the study participation, in particular those lost to follow-up after the initial study waves. In the ECRHS and RHINE, more men than women were lost to follow-up, along with those who were youngest at baseline [210]. On condition that these loss-to- follow ups were selective and not random with respect to both the exposure and outcome, this selection bias could potentially affect the internal validity of a study [258]. However, a previous cohort profile publication of the RHINE study did show that, alhough prevalence estimates were somewhat affected by selection bias in follow-up stages, exposure-outcomes and risk-associations were mainly unaffected [210]. As the overall aim in this thesis has been to elucidate exposure-outcome associations, we therefore do not consider this selection bias to pose a large threat to our results.

In contrast to the sampling method employed in the ECRHS and RHINE studies, participants in the RHINESSA study were recruited based on being the offspring of the ECRHS and RHINE participants. Consequently, the RHINESSA study population constituted a purposive and non-random sample, potentially not adequately representative and generalisable of the population it was intended to study, or to the wider target population. The survey response rate of 35% [211], was also considerably lower than in the foregoing population studies, which potentially could have introduced errors and non-response bias, if those who were unwilling or failed to respond were selective with regard to both the exposure and outcome of a study [258]. However, responders and non-responders in the RHINESSA study have beeen demonstrated to be similar in respect of parental characteristics and exposure

information, such as smoking [211]. As follows, we would not expect that a potential selection bias would affect the exposure distribution in the present thesis. Although we cannot rule out that non-responders were potentially, and independently of the exposure, related to the outcomes of interest, we would not consider this to pose a major risk of distorting the effect estimates or affect the internal validity of the study.

5.1.5 Sample sizes of the included papers

The study described in paper 1 comprised 195 male and female offspring, aged 11-54 years, who originated from either the RHINESSA or ECRHS study in the Bergen study centre. Information on their fathers was collected in the ECRHS study, and the study sample was enriched by offspring with smoking fathers for DNA methylation measurements. Consequently, the sample was made with some degree of selection bias, and may not have been representative of the target population, or generalisable to a wider European/ Australian population. However, when the study sample was compared to demographic characteristics in the noticeably larger sample in paper 2, as well as to data of the wider Norwegian population obtained from Statistics Norway and reported in a previous thesis [267], they were all similar with respect to sex, age, educational level as well as smoking habits. Furthermore, as previously investigated in the RHINE [210] as well as in the Norwegian Mother and Child Cohort study [268], a strictly representative sample may not be essential when the aim is to investigate risk association.We therefore do not consider this to major concern with respect to the internal validity of our resuts.

The study inevitably faced sample size constraints, which may have increased the random variability in the data. However, the study sample was pragmatic, based on blood DNA methylation data that was available and extracted in the population studies at that time, and as such, reflects a common imbalance between the availability of epigenetic versus phenotypic information in adequately large samples from epidemiological surveys, not originally designed to conduct epigenetic analyses [269]. Studies constrained with low sample sizes are encumbered by an increased risk of

producing imprecise estimates and committing an either type I or type II error. The study may therefore not be sufficiently statistically powered to provide precise effect estimates, which possibly can have lead to an either over – or under-estimation of the true value. On the other hand, sample size limitations are not considered to result in inaccurate effect sizes even if they affect variability, and we therefore do not regard this to have affected the validity of our results.

In paper 3, the sample size populations of the two EWAS studies were considerably larger, and included 875 and 304 male and female offspring (age 7-50 years), respectively. The subjects originated from six RHINESSA study centres, and were linked with parental data obtained in the ECRHS and RHINE studies. Thus, the study sample was more likely to be more representative of the source population, and also potentially generalisable to a wider European and Australian population. With an increased sample size, we would furthermore expect that the statistical power and precision of the estimates increased, and were less affected by the potential of producing either false positive or false negative results. That said, the study may still not have been adequately powered to detect small differences at CpG methylation sites.

Moreover, type 1 error rates and significance thresholds are also impacted by the vast numbers of CpG sites that are tested simultaneously in EWAS studies. In paper 1 and 3, we applied a FDR threshold of $P<0.05$, to allow for a balanced compromise between type I and type II error rates. We may therefore not have been adequately stringent to control the false positive rate for EPIC arrays. The FDR rate further assumes that p-values across measured CpGs are uniformly distributed [270], which might have been violated due to the heteroskedasticity and non-constant variance of methylation levels between groups, indicated by the skewed p-value distribution and inflated test statistics in the two papers. Consequently, our studies could be constrained with false positive associations and biased results, which could threaten the validity of the studies. However, a study recently demonstrated that although DNA methylation data may be variable, and do not satisfy the assumptions of equal variances, or normally distributed error terms, neither heteroskedasticity nor

data distributions prevailing excess skewness or kurtosis, generated false positive or false negative associations, and did not seem to produce biased results [271].

Although the study in paper 3 accounted for the nested structure of study participants within study centres, it did not account for participants nested within families. As such, the regression coefficients were estimated without correctly accounting for the potential correlation between siblings, due to both shared familial environment and genetics, which was the case for 73 of the study subjects. This may have lead to underestimated standard errors, and imprecise effect estimates, which potentially could have mislead the statistical inference of the paper [272]. However, when family origin was added as a random effect in a sensitivity regression analysis in paper 1, the potential effect of the within similarities in this cluster, did not change the original results.

In paper 2, the sample size included 6487 adult male and female offspring (mean age 30 years) from 10 RHINESSA centres, and 2111 and 2569 unique fathers and mothers, respectively, who were participating in the RHINE and ECRHS studies. The study sample was large and regarded as representative of the source population, as well as generalisable to the wider European and Australian population. We also considered the sample size to be sufficiently powered to detect precise effect estimates and correctly identify true positive results. In addition we accounted for clusters within study centres and families, and thus accounted for bias of standard errors in the analyses.

That said, the subsample with available electrical bioimpedance data (n=240) was of limited size, and may not have been adequately powered to avoid producing either false positive or false negative results. This may have affected the precision and reliability of our effect estimates in these particular sub-analyses.

5.1.6 Data collection

The ECRHS, RHINE and RHINESSA studies, have all measured a wide range of characteristics, which have enabled the assessment of a variety of exposures, outcomes, and potential mediating and confounding factors. However, measurement instruments, such as self-administered questionnaires, interviews, labaratory tests or physical measurements, may all be subject to some degree of error and bias if they are not able to correctly and reliably assess and classify what they purports to measure [255], which consequently can pose a threat to the validity and repeatibility of our results. A pertinent step towards minimizing systematic distortion and measurement errors during data collection, is therefore to apply research instruments and procedures that are accurate and consistent. To surpass and reduce potential misclassification, particularly those who arise with the use of self-reported data, the ECRHS, RHINE and RHINESSA have all employed standard and harmonized operating procedures, coordinated field-work training, as well as extensive interview guides and systematic procedures for translating and backtranslating questionnaires and interviews [210, 211].

5.1.7 Exposure assessment of fathers' and parents' smoking trajectories

The exposure assignments for the papers in this thesis, were for the most part obtained from the parents' own responses to interview or questionnaires in the ECRHS and RHINE studies, or when not feasable, based on the offspring's report on childhood smoking exposures, which was the case for the offspring originating from the ECRHS cohort in paper 1. A previous study from the RHINESSA has found good agreement between offspring's and parents' report of smoking exposure (κ=0.79 (0.78-0.80)), thus assessing this as a valuable measure in the absence of parents' direct reports [273].

Nonetheless, although commonly applied in survey studies, the use of self-reported data and measures is hampered with limitations and potential bias, such as those

related to the responders' tendency to report experiences they consider to be more socially desirable [274]. Such a social desirability bias may be particularly present when querying about participants' smoking habits, and may lead to underreported measures of actual cigarette consumption [262]. Another source of misclassification could derive from recall bias and imprecise memories of past exposures.

Consequently, the smoking measures could be inaccurate, and may fail to correctly classify offspring as truly exposed or unexposed to parental smoking. However, we would not expect such a misclassification to be differential and dependent on the offspring's DNA methylation signatures, nor their BMI and FMI levels. Although the effect of non-differential and random misclassification could have increased the similarities between exposed and nonexposed offspring in the three papers, and thus result in an underestimation and attenuation of the true exposure outcome association, we would not consider this to have biased our estimates and distorted the validity in the studies.

That said, the three papers may not have been able to correctly define and classifiy the exposures of fathers' and parental smoking as purported by the studies' aims, which potentially also can have introduced errors and lead to erroneous conlusions of our results. This may be of particular relevance for paper 2 and 3, as the smoking exposures were not only crudely classified by the parents' ever and never smoking status (as were the case in paper 1), but were further refined and categorised according to various pre-and postconceptual time points of smoking onset. However, these exposure definitions have not been able to mutually exclude, and address, other potential aspects of smoking exposures, and may therefore be biased by the presence of subsequent accumulating second hand smoke exposure. Moreover, in paper 2, the exposure classifications of parental smoking, particularly with regard to that of postnatal smoking onset, yielded groups of few observations, which potentially could have caused misleading inference and affected the reliability of our study results.

Although we ran independent mediation analyses to additionally investigate and account for a potential mediation by the amount of packyears the parents had smoked in paper 2, the fathers' smoking exposure variables employed in paper 3 cannot truly

distentangle whether the offspring's methylation patterns associated with fathers' smoking are reflective to the time of smoking initiation, or rather attribute to other intermediating factors, such as smoking intensity, and passive smoking exposures during the offspring's childhood. However, in a previous epidemiological report, assessing the phenotypic impact of various aspects of smoking in more than 20,000 father-offspring pairs, the age of the fathers' onset was found to be of considerably greater importance than any other measures of fathers' smoking [133]. Even though this may add support to the validity of our study results, the constraints and potential bias inherent in the fathers' smoking definitions still merit caution when interpreting and drawing conclusions of the observed exposure-outcome associations.

5.1.8 Outcome assessment

5.1.8.1 DNA methylation

Genome-wide investigations of offspring DNA methylation patterns were carried out using the Infinium Methylation EPIC (EPIC) BeadChip array, which is the most current array-based detection method for assessing DNA methylation. It quantifies methylation at 853,307 CpG sites, particularly enriched in regulatory regions such as enhancers[34], transcription factor binding sites[35], open chromatin regions[36] and DNase1 hypersensitive sites[37] [275].

Compared to the former Illumina Infinium HumanMethylation450K BeadChip assay, the EPIC array is augmented with additionally 413, 745 CpG sites [276], and is regarded a valid and reliable tool for methylation measurements [213, 275]. However, BeadChip arrays such as the EPIC assay, still quantifie a small fraction of the total

[34] Short (50-1500 bp) regions of DNA that function to enhance – and increase- likelihood of transcription to occur
[35] Binding sites for proteins with DNA binding activity (transcription factors) that are involved in the regulation of transcription. Transcription factor binding sites are often located to a gene's promoter or to enhancers, and sites involved in the regulation in transcription
[36] Regions that can be accessed by regulatory elements.
[37] Special regions that easier allow for DNase I cleavage and chromatin breakdown, which makes the chromatin less condensed and thus makes the DNA accessible.

number of CpG sites in the genome (~28 million) [277], and are further limited with a proportion of less reliable probes that may fail to produce consistent and replicable signals during repeated measurements, and consequently can give rise to bias, spurious associations and false negative results [278]. Although this potentially could pose a threat to the validity and reliability of the EWAS results in paper 1 and 3, the overall data from the EPIC array at single loci have been validated to be highly reproducable across technical and biological replicates [213].

Furthermore, to ensure robust and valid results and avoid bias attributed to batch effects and other experimental steps of the Infinium assay, the DNA methylation data in both the studies have been thoroughly preprocessed, and analysed according to well established methodological pipelines. Moreover, we would not expect that any potential measurement errors arising from technological artefacts and variation would be systematically differential between offspring exposed and non-exposed to paternal smoking. We therefore do not expect that any imprecision and errors in the outcome assessment would result in distorted methylation patterns between cases and controls. None-the less future studies should explore the use of alternative methodologies e.g nanopore sequencing[38] to confirm differential methylation at CpG sites identified in our studies and its relationship to methylation at adjacent CpG sites adjacent to the loci identified utilising the Illumina EPIC array.

5.1.8.2 Body Mass Index

Assessment of offspring's body mass index (BMI) was based on self-reported measures of height and weight in the RHINESSA study, and was calculated by the formula: (weight [kg]/ ((height [m])^2). Although correlation coefficients for self-reported and technician-measured weight and height have been shown to be high (0.89 and 0.94, respectively) [279], the data are nonetheless prone to the same

[38] Nanopore sequencing is a third generation sequencing technique that uses electrical signal profiles to detect the sequence of a DNA molecule. Compared to microarrays, it provides a much more comprehensive genome-wide coverage and can detect differential methylation directly from sequence data that has not been bisulphite-converted.

misclassification bias and errors as those previously discussed when measurements are based on the participants' own responses. Although participants not accurately reporting their weight can have lead to a tendency of both overweight and underweight [280, 281], we would not expect that these inaccurate responses were differential, or dependent on whether their parents smoked or not. We therefore do not consider this potential misclassification bias to pose a threat to the internal validity of the study

Another challenge with BMI is that it does not distinguish between lean and fat mass, and has been found to have a limited accuracy for diagnosing obesity and correctly identify individuals with excess fat mass [282]. BMI may not necessarily reflect the differences in percentage of body fat between men and women, or changes in body fat and muscle mass that occur with age [283]. Consequently, as our study has utilized BMI as a measure of obesity, this may have introduced misclassification errors which potentially can have biased our effect estimates. This would be of a particular concern if the proportions of subjects potentially misclassified as being overweight, would be differential, and systematically different across the parental smoking exposure categories. Given the sex and age dependent variations in BMI measures, we tested whether the associations between parents´ time points of smoking onset and offspring BMI were modified according to the sex of the offspring in all the regression analyses, which they were not. Although we did not include the offspring's age as an interaction term in the original study analyses, when running subsequent analyses on the study data, we found no evidence of interaction by the offspring's age, neither in the father-offspring (interaction p= 0.6) nor the mother-offspring (interaction p= 1.0) analyses. We therfore find it likely that any measurements errors related to the use of BMI as outcome would be random, and non-differential according to the various parental smoking exposure groups.

5.1.8.3 Fat Mass Index

Measures of body composition and fat mass were assessed by Bioelectrical Impedance Analysis (BIA), using the Bodystat 1500 MDD (https://www.bodystat.com/medical/). Due to its noninvasive, quick, and fairly inexpensive technique, BIA is particularly suitable for assessing body composition in large population cohorts, and the instrument provide precise and reliable estimates of fat-free mass (FFM) and total body water (TBW) [284-286], which are used to calculate absolute and relative body fat amounts [284]. BIA derived prediction equations have been validated in both children and adults and are considered to be reliable estimates of adiposity in human populations [284]. Based on these estimates, we calculated fat mass index (FMI) by the formula: ((fat mass [kg]/ (height [m])^2).

To ensure accurate and consistent BIA measurements among different technicians and between participants, the BIA analysis was performed by trained personnel following detailed instruction protocols in adherence to the manufacturer's guidelines. Calibration checks on the instrument were performed daily according to the manufaturer's recommended standards. These aspects assure us that the data on fat mass in the paper 2 are both valid and reliable estimates. If in any case the mesaurement would be subject to some degree of error, we would expect such a misclassification to be random and just as likely to occur in the offspring, regardless of whether they had smoking parents or not, and as such potentially dilute, but not distort, the true strength of the association between parental smoking and offspring's body composition.

5.1.9 Covariates and potential confounders included in the papers

5.1.9.1 The EWAS studies in paper 1 and 3

In paper 1 and 3, the covariates selected for the regression models, were foremost guided by convential preceps of variables known to affect the methylome, commonly ascertained in epigenetic studies in order to avoid spurious associations [269]. The

papers included the following offspring covariates: sex, age, offspring's own smoking and cell-type proportions. In paper 3, the potential impact of DNA methylation variability by offspring's sex was also more thoroughly investigated using sex-stratified EWAS analyses.

In addition, both the papers included maternal smoking in the regression analyses, as we regarded this a potential confounding factor that could be predictive of both the offspring's DNA methylation patterns, as well as likely be associated with the fathers' smoking exposures.

Furthermore, besides adjusting for maternal and offspring's own smoking status in the EWAS analysis on fathers' smoking exposures, in paper 3 we also conducted EWAS analyses on mothers' and the offspring' own smoking behaviours to allow for comparison of the smoking associated dmCpGs identified for each exposure (personal, maternal and paternal).

As paper 1 explored associations between fathers' smoking exposure and DNA methylation levels in offspring originating from two population cohorts investigated at different decades and therefore with different mean ages of 26 and 44 years, respectively, we additionally included fathers' age at the offspring's birthyear, to account for a potential cohort effect modification on offspring's methylation levels, by age and time dependent variations in smoking rates [287]. In paper 3, the offspring span from 7 to 50 years of age, thus a potential age dependent variation in methylation patterns was also more extensively assessed in a sensitivity analysis, by investigating correlations between offspring's age according to aging-related CpG markers identified from the RHINESSA EWAS, as well as to dmCpGs identified as related to fathers' preconception smoking onset before the age of 15.

Given that paper 3 had an extended study population of offspring-parent pairs originating from different European/Australian countries, we also adjusted for study centre origin in the regression model, to account for potential similarities of observations within each study centre cluster.

Both papers did adress the potential confounding impact of social class in sensitivity analyses, in which the fathers' education level was used as a proxy for socioeconomic status, and added either as a coviarate (paper 1), or used as a independent variable (paper 3) in regression analysis of top father smoking associated dmCpGs.

Although we have aimed to account for and diminish the impact of these potential biases in the EWAS studies, epigenetic markers are dynamic factors that vary in response to a wide range of environmental influences. Thus, we are aware that even after applying statistical correction methods for common covariates, there will likely be additional confounding by factors other than the measures reported in the current papers, which potentially can have biased our study results and lead to erroneous conclusions [152]. However, in a recent study, utilizing highly advanced probabilistic simulation techniques and based on data from the ECRHS population cohort, unmeasured confounding was demonstrated to exert only minor impact on the associations between fathers' preconception smoking and offspring's phenotypic outcomes [132]. This implies that the study results may still be valid, despite the presence of unknown and unmeasured confounding factors, which is likely to be the case in most epidemiological studies.

Yet, in EWAS studies, the complexity of residual confounding is further augmented by the fact that DNA methylation variation between individuals may reflect that of cell subtype effects, not adequately accounted for in the analyses, or of even greater concern, be attributable to unmeasured influences of transcriptional or DNA sequence effects on DNA methylation [288]. Consequently, these factors may have lead to inaccurate estimates and can have precluded the true associations between fathers' smoking exposure and offspring DNA methylation [263].

Similar to that of many epigenetic epidemiological studies, our data are faced with the tissue-constraints of banked peripheral blood material, and are thus not capable of adressing potentially subtle cell type composition effects, beyond what is estimated and controlled for by the use of statistical deconvolution algorithms [289]. Moreover, DNA methylation studies, based on large-scale population study cohorts, such as ours, are rarely perfomed with concurrent genotyping of the same study subjects, or

accompanied by transcriptional studies of the same cells. This make us unable to truly distinguish intra-individual epigenetic variation from that of DNA methylation influences arising from genetic variants or transcriptional changes [269]. However, as the present papers have seeked to detect epigenetic profiles in offspring as biomarkers of fathers' smoking exposures, and have not aimed to understand causal mechanisms, it may not be crucial whether these biomarkers are due to epigenetic changes in the cells tested, or influenced by distinct effects of subtle cell sub-types, or by genomic or transcriptomic variablility [290].

Nonetheless, given the range of potential confounding sources in EWAS studies – known or unknown – it is particularly pertinent to ascertain validation and corroboration of study results in independent, but comparable study samples, with similar measures and statistical modelling of the exposure and outcome [269, 291]. In both papers, we pursued to replicate and confirm our association results. However, obtaining suitable replication cohorts for EWAS studies often merit difficulties, particularly when the exposure of interest is rarely quantified or reported- such as time points or fathers' preconception and/or postnatal smoking exposure, or when different assay platforms are used and therefore exhibit differences in genomic coverage and probe sites. For this reason, too few fathers' smoking dmCpGs were present for replication in the Isle of Wight Cohort in paper 1. In paper 3, 11 of the 19 significant dmCpGs associated with fathers' preconception smoking before age 15, showed nominal replication (correlation =r 0,49, p-value 0.12) and concordant direction of effects in the ALSPAC replication cohort. Still, we did not identify overlapping significant dmCpGs associated with fathers' preconception smoking, and acknowledge that the EWAS results in this thesis are yet to be confirmed and validated, and as such, should be considered as preliminary findings that warrant further investigation in order to be considered as reliable and generalisable to a wider population.

5.1.9.2 Paper 2

In paper 2, we selected potential mediating and confounding factors for the regression analysis based on both reviewing relevant literature, and by inspecting the dependent structure and direction of effects of multiple variables using a Directed Acyclic Graph (DAG). Although we considered a broad range of covariates in the DAG, including parental age, offspring's education, as well as the smoking status and BMI of the other parent who did not participate in the RHINE/ECRHS, we only regarded the parents' educational level, in being a proxy for socioeconomic status, as a factor that potentially could confound the relationship between pre- and postconceptional time periods of parental smoking onset and offspring's BMI and FMI levels.

As the study comprised study participants originating from the same family and from ten various study centres, we implemented study centre and family origin as cluster variables, to account for the heteroskedasticity across clustered observaions.

We further added offspring sex as an interaction term to investigate whether any of our observed offspring's BMI and FMI outcomes related to time points of parental smoking exposures, varied in strength or direction between male and female offspring.

Based on the identified paths between variables incorporated in the DAG, we considered the following variables to be potentially intermediate variables on the causal pathway between parental smoking onset and offspring BMI outcomes: i.parental pack years of smoking, ii. parental BMI, iii. offspring's smoking habits, and iv. offspring's birthweight, which was available for a subsample of offspring. To further identify whether the associations differed between male and female offspring, we included offspring sex as an interaction term in the independent mediation analyses.

Although these methodological attentions have aimed to reduce the presence and impact of confounding bias, there are likely to be other factors than those considered in the paper, that may have influenced the study results. Given that the offspring population had already reached adulthood, a broad range of lifestyle related factors, such as dietary habits and physical activity would be expected to impact on their BMI

levels and risk of becoming obese. However, for these factors to potentially have confounded the observed exposure-outcome associations, they should per defention not only have preceeded the outcome in time, but also the exposure, i.e., parental smoking onset in preconception and postnatal years, which therefore rules out many of the environmental and behavioral factors that would have occurred during the time the offspring grew up and became adults. On the other hand, these factors could be potential intermediating factors of the relationship between parental smoking onset and offspring BMI outcomes.

In addition, there may be genetic and biological components in both the parents and offspring that are involved in the aetiology of obesity [292, 293], which could have constituted potentially unmeasured confounders. Although we in mediation analyses investigated the impact of the parents' adult BMI on the observed exposure-outcome associations, we did not have information on the parents' BMI in childhood and early adolescent years, which we consider to be a potentially important confounding factor in these analyses. Although the statistical probabilistic simulation techniques employed in a previously mentioned study found that unmeasured confounding excerted a limited impact on the study's findings [132], we can not rule out the possibility that these factors may have impacted our results, or lead to erroneous conclusions.

Within this context, the constraints of the applied Medflex mediation package, may compose a particular methodological concern. Although, we chose this tool due to its flexibility in handling non-linear parametric models, it does not offer commands to assess the sensitivity of the mediation results to possible violations by the existence of potentially unobserved confounding covariates [294]. This is a pertinent step when conducting mediation analyses embedded within the context of causal inference, as they are based upon strong assumptions that are not always possibly to verify from the observed data, particularly with regard to that of sequential ignorability, i.e. that all potential variables are independent, and no unmeasured confounding is present in neither the exposure-mediator, exposure-outcome, nor the mediator-outcome relationships [295]. However, as we rather presume it most likely that our data are

constrained with some degree of unmeasured confounding, the lack of sensitivity analyses beget caution when interpreting the mediation results, and they should be regarded as preliminary findings, that warrant further investigation in order to be confirmed.

In the present paper we assessed the mediation effect of each of the hypothesized mediators in separate models. In initial analyses, we additionally estimated a joint mediation effect by assessing all the mediators within one single effect model. However, this analysis was complicated due to the amount of mediators, and did not allow to make inference on each of the hypothesized mediators' contributions and effects. For this reason, we chose to investigate each mediator in independent analysis. That said, yet, these mediators may not have been conditionally independent, but rather be linked through a sequential causal chain [235].

Lastly, the subpopulation with FMI data was not suffciently large for conducting mediation analyses. We have therefore not been able to investigate whether the potential intermediating factors excerted a similar influence on the association between parental smoking onset and offspring's body composition, when the outcome parametre was a considerably more accurate measure of fat mass and obesity. As such, our results may not be adequately valid and applicaple to a wider population.

5.1.10 Missing data bias

The three papers included in the thesis, were confined to complete case analyses, where participants with missing data in the exposure and outcome variables were excluded from the analyses. This has decreased the sample sizes of the studies. We considered the study population with BMI measures in paper 2 to be adequately large for running a complete case analysis without being negatively affected by a loss of statistical power. However, this assumption would not hold for the small subsample with FMI measures, or for the limited EWAS population in paper 1. The missing data in the EWAS studies in paper 3 might also have caused loss of precision and power,

particularly for the smaller subsample investigating DNA methylation levels related to fathers' smoking onset before age 15.

Although lack of information and amount of missing in the data is an inevitable problem in epidemiological reasearch, unless they occur completely at random- and are missing purely by chance- they may introduce bias and weaken the validity of our reasearch results [296]. Although the clinical outcome measurements in the papers, such as DNA methylation levels and FMI values, were foremost restricted due to the limited number of subjects attending the clincial studies, we can not rule out the possibility that the missing data, particularly those related to loss of information on parental smoking exposures, might have occured from some other observed or unobserved sources not verifiable in the observed data [296]. We are aware that multiple imputation methods may have been the preferred approach to avoid unnecessary deletion of observations due to missing values in both our epidemiological as well as high-dimensional DNA methylation analyses, which also would have increased the statistical power and precision of the estimated effects [296]. These statistical techniques use the distribution of the observed data to create multiple data sets that are imputed and combined to obtain a set of plausible values for the missing data [297]. However, in neither of the studies, this approach was considered when planning or conducting the analyses. We acknowledge that the missing observations in the studies may constitue potential bias, that can have influenced on the accuracy and the precision of our results.

5.2 Discussion of main findings

The overall focus of this thesis was to explore potential intergenerational outcomes related to parents', and most specifically fathers', previous smoking exposures. Firstly, by adressing whether fathers' smoking exposures may be related to altered methylation patterns in offspring, which to a large extent have only been investigated for personal and maternal smoking exposures. In paper 1, our intial effort aimed to

explore this hypothesis, by investigating whether offspring with fathers who had ever been smoking, exhibited altered DNA methylation patterns compared to offspring with never smoking fathers. Secondly, given that germ cells may be particularly plastic and sensitive to tobacco smoke constituents during germ cell development and maturation, paper 2 and 3 further aimed to specifically target parental smoking exposures occuring at preconception and adolescent time points, and to explore whether these potential vulnerability periods were related to phenotypic and methylomic and variations in the offspring.

5.2.1 Summary of top DNA methylation signals related to fathers' ever smoking

The 33 differentially methylated positions (DMPs) (FDR<0.05) in our initial hypothesis generating EWAS study, did not remain epigenome-wide significant after correcting for genomic inflation (λ=1.462). However, when accounting for spatial correlation of the single site EWAS P-values in the region based analysis, we were able to identify six genomic regions (DMRs) that were significantly associated with fathers' ever smoking exposure (Sidak-corrected P value <0.05). Among these, five were mapped to known genes, and are listed in table 1.

To our knowledge, none of the the consequtive methylation sites in these DMRs have so far been associated with any smoking exposure. However, when undertaking an updated review of the current literature from large epigenome-wide meta-analyses on personal and sustained maternal smoking during pregnancy, we note that most of the annotated genes have also been reported by these studies. We suggest this further supports that the methylation signals detected in our study indeed are related to smoking, and may indicate that certain common loci are particularly susceptible for tobacco smoke induced variation in DNA methyation, and thus consequently implicated in various sources of smoking exposures. Yet, we find that the signals associated with fathers' ever smoking seem to be related to methylation changes at

distinct, and to our knowledge, novel CpG sites, not previously identified in EWAS studies on personal and maternal smoking (table 1).

Table 1: DMRs (Sidak<0.05) associated with fathers' ever smoking exposure

Chr location and no. probes	Pregion*	Sidak**	Gene annotation[a]	Context[b]	Association	Previous evidence linking CpG site/ gene to smoking exposure
Chr22:18111277-18111521 4 probes	6.01E-07	0.0019	ATP6V1E1 Intron, 5´UTR, cds	Island	**Fathers' ever smoking**	Novel sites. Gene previously associated with personal smoking by Qiu et al[298]
Chr6:31865522-31865578 5 probes	2.49E-06	0.0055	C2 TSS, intron, exon, 5ÚTR	Shore		Novel sites and novel gene not previously associated with smoking
Chr2:80752765-80752967 4 probes	1.69E-06	0.0006	CTNNA2 Intron	NA		Novel sites. Gene has been associated with personal smoking by Joehanes et al[246]; CHARGE[299]; Christiansen et al[247]; Hannon et al[300], and to current and former smoking by Dugué et al[301]
Chr16:89180587-89180443 3 probes	5.83E-06	0.0173	ACSF3 Intron, cds, nc_intron, nc_exon, nc_intron	NA		Novel sites. Gene has previously been linked to personal smoking by Joehanes et al[246]; CHARGE[299]; Hannon et al [300], and with sustained maternal smoking during pregnancy: Joubert et al[248]
Chr7:158766826-158767135 3 probes	5.24E-06	0.0129	WDR60 Intergenic	Island		Novel sites. Gene has been reported as related to personal smoking: Joehanes et al[246]; CHARGE[299]; Christiansen et al[247], and to current and former smoking by Dugué et al[301]

*Pregion: P-value of DMR calculated by the Stouffer-Liptak-Kechris (slk) correction **Sidak: P-value of DMR corrected for multiple testing a:5´UTR: 5 prime untranslated region; cds: coding sequences; TSS: transcription start site, nc_intron: non-coding intron; nc_exon: non-coding exon b: Genomic context relative to CpG islands: Shore: up to 2 kb from flanking CpG islands

The annotated genes and their related pathways are involved in innate immunity (*ATP6V1E1*, *C2*) [302, 303], neural system development (*CTNNA2*) [304], lipid metabolism and fatty acid biosynthesis (*ACSF3*) [305], as well as to cell cycle progression, signal transduction and gene regulation (*WDR60*) [306]

As methylated cytosines across a genomic region tend to exhibit similar levels of methylation, differentially methylated regions are more conceivably implicated in chromatin remodelling and transcriptional regulation compared to isolated CpG sites [269]. When inspecting the genomic context for potential functional implications, two of the identified regions (*ATP6V1E1*, *WDR60*) harboured CpG islands, and the CpG sites within *ATP6V1E1* covered parts of the 5 prime untranslated region (5´UTR) as well as the coding sequence (cds) of the gene. The DMR on chromosome 6 (*C2*) co-localized with a CpG island shore, and overlapped with the transcription start site (TSS), the 5´UTR, and exon 1 of the gene, whereas the methylated sites within the *CTNNA2* region overlapped with intron 11 of the gene (table 1). Although this indicates that the methylation variations within the identified regions are located at functionally relevant genetic elements, the DMR analysis is based on the single site EWAS analysis with P-values of only nominal significance. Thus, our results may have underestimated dependencies between CpG sites within the genomic regions, and potentially may have failed to avoid false positive findings.

Given that the inflated test statistics in the single site level analysis did not allow us to properly account for the multiple testing burden and control for type I errors, the potentially correlated structure and co-methylation of proximate CpG sites inevitably makes our DMP analysis prone to false discoveries. We are well aware that this merit caution when interpreting the differentially methylated positions. That said, when reviewing the top ten DMPs (table 2) according to published meta analyses, we do observe a similar overall trend as that of the region based analysis: the methylated positions are novel, but a majority of the annotated genes are previously related to personal smoking (*BCAS1*, *MFGE8*, *ZNF689*, *PEBP4*, *UNC93B1*, *PHF12*, *RALB*,

FREM2, *DLGAP1*) [246, 299-301] as well as with sustained maternal smoking during pregnancy (*MFGE8*, *PEBP4*, *UNC93B1*, *DLGAP1*) [248, 299].

Table 2: Top 10 DMPs

CpG	Chr	Gene	P-value	Adj P*	Beta	SE
cg05019203	20	BCAS1	2.83E-08	4.40E-06	-0.018	0.003
cg25727029	15	MFGE8	3.56E-08	5.15E-06	0.013	0.002
cg00626693	16	ZNF689	6.27E-08	7.64E-06	-0.014	0.003
cg19754387	2	CCNYL1	1.33E-07	1.29E-05	0.006	0.001
cg24534854	8	PEBP4	2.09E-07	1.76E-05	-0.013	0.003
cg20272935	11	UNC93B1	3.02E-07	2.27E-05	0.024	0.005
cg04164584	17	PHF12	3.44E-07	2.49E-05	-0.010	0.002
cg06876354	2	RALB	4.65E-07	3.07E-05	0.017	0.003
cg25012097	13	FREM2	4.74E-07	3.11E-05	-0.012	0.002
cg07217718	18	DLGAP1	6.17E-07	3.73E-05	0.025	0.005

*Inflation-adjusted P-value <0.0001

Furthermore, we do observe that several of the top DMP annotated genes have similar putative functions and pathways as seen in the region based analysis, specifically in relation to innate and adaptive immunity (*BCAS1*, *MFGE8*, *UNC93B1*, *RALB*) [307-311], metabolism (*MFGE8*) [312], as well as with neural systems and behavioral dysfuntion (*DLGAP1*) [313].

That said, the effect estimates of the differentially methylated probes were small, with beta values ranging from -0.02 to 0.03, which of necessity, make reason to question their potential biological importance. Although previous studies have shown that even small variations in DNA methylation can influence on transcriptional activity [291, 314], neither of the differentially methylated positions were significantly enriched for regulatory targets, such as histone modification signatures and transcription factor sites, identified from the ENCODE and Epigenomic roadmap. Thus, the functional relevance of the methylated sites is yet to be determined, and needs further investigation. However it also needs to be considered that methylation levels were assessed in mixed blood cell populations, and small differences in average methylation may result from larger methylation differences in a cell type that is present at a small proportion in blood.

Taken together, these hypothesis-generating results do not necessarily conflict with what was reported by Joubert el al, who found no evidence of fathers smoking exposure being related to CpG sites previously associated with maternal smoking during pregnancy [315]. However, we argue this lack of overlapping sites does not necessarily imply that the exposure effects from paternal smoking do not affect his offspring's methylome, rather contradictorily, our findings suggest that fathers' smoking is associated with novel differentially methylated sites in the offspring, even detectable at adolescent and adult age (age of offspring ranging from 11-54 years). Given that DNA methylation can be stably propagated through mitotic, and possibly meiotic cell divisions [7, 76], and the extensive experimental evidence demonstrating father to offspring transmission of gametic methylation changes [316], it may be theoretically plausible that fathers' smoking exposures can persistently alter the offspring's methylation patterns.

However, our study only investigated variations in offspring's DNA methylation in relation to fathers who had ever smoked, and consequently cannot disentangle whether the observed methylomic differences potentially have been transmitted through gametic epigenetic alterations, or rather reflect smoking exposures ensued from fathers' secondhand smoking during the offspring's gestational period or in post-natal years. Although we have controlled for the potential impact of mothers' and the offspring's own smoking exposure, as well as for socioeconomic influences and the potential effects of shared familial environment and genetics, nonetheless the skewed P-value distribution implicated from the inflated test statistics of the single site EWAS analysis, might reflect the presence of unaccounted confounders, which can have influenced on the observed exposure associations. Given the small sample size, we are also aware that the study is constrained by low statistical power and increased error rates. The study results are therefore yet to be confirmed in an independent cohort, as we acknowledge they are preliminary findings that need further validation.

In the subsequent EWAS studies in paper 3, we aimed to accommodate these limitations, by specifically investigating time points of fathers' preconception smoking exposures in a substantially larger study population of offspring. In addition, by running separate EWAS analysis on personal as well as maternal smoking, we further aimed to elucidate whether the methylation signals associated with fathers' smoking exposures truly were distincly different to those identified in relation to the offspring's own and their mothers smoking behaviours.

5.2.2 Summary of top DNA methylation signals related to preconception time points of fathers' smoking onset

The top differentially methylated positions, detected at epigenome-wide significance in the second EWAS paper are summarized in table 3. Sex-stratified DMPs (FDR ≤ 0.05) are presented in table 4. The reported associations are further marked in blue and green, according to whether the signals are related to fathers' any preconception smoking onset and fathers' early adolescent (<15 years of age) smoking onset, respectively. Several more methylation sites surpassed significance in the EWAS on fathers' adolescent smoking onset, as opposed to the analysis investigating differential methylation related to fathers' smoking commencing at any time during preconceptional years.

To our knowledge, this study is the first to investigate potential epigenetic alterations in offspring in relation to fathers' adolescent smoking exposures. Except for cg20728490 in *DNTT*, cg12053348 (missing gene annotation), and cg11380624 in *DNAJC14* (marked in bold in table 3), the detected methylation sites are novel and not previously associated with smoking exposures. However, similar to our preliminary findings in paper 1, remarkably many of the annotated genes are previously reported in large meta-analyses and consortia based EWAS studies on personal or sustained maternal smoking during pregnancy, which adds to the credibility of the differentially

methylated sites truly to be related to smoking exposures, and thus representing potential candidates for validation in other studies.

Table 3: DMPs (FDR<0.05) associated with fathers' preconception and early adolescent smoking onset

CpG	Gene annotation[a]	Context*	Beta	SD	Adj.P**	Association	Previous evidence linking CpG sites and annotated genes to smoking exposures
cg00870527	PRR5, 5´UTR	shelf	-0.024	0.070	0.028	**Any preconception smoking onset**	Novel site. Gene associated with personal smoking by Joehanes et al[246]; CHARGE[299] and sustained maternal smoking during pregnancy by Joubert et al[248]; PACE[299]
cg08541349	CENPP, Body	opensea	-0.012	0.023	0.028		Novel site. Gene associated with personal smoking by Joehanes et al[246]; CHARGE[299]
cg23021329	TLR9, Body	shore	0.015	0.021	0.026	**Smoking onset before age 15**	Novel site. Loci linked to personal smoking: Joehanes et al[246]; CHARGE[299]
cg12053348	**NA**	**opensea**	**0.036**	**0.056**	**0.026**		**CpG site previously related to personal smoking: Christiansen et al.[247]**
cg03380960	FAM53B, Body	opensea	0.019	0.045	0.034		Novel site. Loci related to personal smoking: Joehanes et al[246]; CHARGE[299], current and former smoking: Dugué et al[301] and sustained maternal smoking during pregnancy: Joubert et al[248]
cg26274304	NCAPG2, 5´UTR	shore	0.018	0.027	0.037		Novel site. Gene linked to personal smoking: Joehanes et al[246]; CHARGE[299]; Hannon et al[300]
cg16730908	PSTPIP2, TSS1500	shore	0.021	0.032	0.037		Novel site. Gene associated with personal smoking: Joehanes et al[246]; CHARGE[299]. Current and former smoking: Dugué et al[301]
cg13904562	NA	opensea	0.041	0.056	0.037		Novel site
cg07508217	NA	opensea	0.026	0.042	0.037		Novel site
cg03516318	MBIP, TSS1500	shore	0.028	0.039	0.037		Novel site. Loci linked to sustained maternal smoking during pregnancy: PACE[299]
cg10883621	C2orf39, TSS200	shore	0.020	0.032	0.037		Novel site. Gene associated with personal smoking: Joehanes et al[246]; CHARGE[299]. Current and former smoking: Dugué et al[301]
cg22402007	NTRK2, TSS1500	opensea	0.022	0.031	0.041		Novel site and novel gene not previously reported related to smoking exposure
cg15882605	NA	opensea	0.025	0.051	0.041		Novel site
cg03818156	NA	opensea	0.017	0.028	0.041		Novel site
cg13288863	CDO1, Body	opensea	0.020	0.049	0.048		Novel site. Gene related to sustained maternal smoking during pregnancy: Joubert et al[248]; PACE[299]

cg03743584	PRAP1, 1st Exon	opensea	0.018	0.025	0.048	Novel site and novel gene not previously reported related to smoking exposure
cg10981514	TPCN1, Body	opensea	0.023	0.042	0.048	Novel site. Gene associated with personal smoking: Zeilinger et al[317]; Joehanes et al[246]; CHARGE[299]; Christiansen et al[247], current and former smoking: Dugué et al[301], and sustained maternal smoking during pregnancy: Joubert et al[248]
cg06600694	IRS1, TSS200	island	0.005	0.008	0.048	Novel site. Gene linked to personal smoking: Joehanes et al[246]; CHARGE[299] and sustained maternal smoking during pregnancy: Joubert et al[248]
cg14700085	CSF1R, Body	opensea	0.016	0.024	0.050	Novel site. Gene linked to personal smoking: Joehanes et al[246]; CHARGE[299]; Hannon et al[300]
cg20728490	**DNTT, 5´UTR**	**opensea**	**0.032**	**0.049**	**0.026**	**CpG and loci reported related to personal smoking by Christiansen et al[247]. Gene reported by Joehanes et al[246]; CHARGE[299]; Dugué et al[301]**
cg11380624	**DNAJC14, 5´UTR**	**shore**	**0.024**	**0.036**	**0.041**	**CpG and loci reported related to personal smoking by Joehanes et al[246]; CHARGE[299]**

a: 5´UTR: 5´untranslated region located between Transcription Start Site (TSS) and the start codon (ATG); TSS1500: 200-1500 bases upstream of TSS; TSS200: 0-200 bases upstream of TSS, *Genomic context relative to CpG islands: Shelf: up to 2 kb from flanking shores; Open sea: >4 kb from CpG islands; Shore: up to 2 kb from flanking CpG islands **DMPs corrected for multiple testing at a FDR treshold of <0.05

Table 4: Sex-stratified DMPs (FDR<=0.05) associated with fathers' preconception smoking onset

Offspring	CpG	Gene annotation[a]	Context[*]	Beta	Adj.P[**]	Association	Previous evidence linking CpG sites and annotated genes to smoking exposure
Males (N=457)	cg05193832	KCNJ1, 5´UTR	opensea	-0.010	0.038	**Any preconception smoking onset**	Novel site. Gene associated with personal smoking by Joehanes el al [246]
	cg25390635	GRAMD4, TSS200	island	-0.009	0.039		Novel site. Gene associated with personal smoking by Joehanes el al [246] and sustained maternal smoking during pregnancy by Joubert et al [248] and PACE [299]
	cg22905274	TRIM2, Body	island	-0.005	0.052		Novel site. Gene reported related to personal smoking: Joehanes et al [246]; CHARGE [299]; Christiansen et al [247] and with sustained maternal smoking during pregnancy: Joubert et al [248]; PACE [299]
	cg02518394	MYADML2, TSS1500	opensea	-0.010	0.052		Novel site and novel gene not previously reported related to smoking exposure
Females (N=418)	cg09801901	LEPROTL1, Body	opensea	-0.022	0.023		Novel site. Gene linked to personal smoking by Joehanes et al [246]

a: 5´UTR: 5´untranslated region located between Transcription Start Site (TSS) and the start codon (ATG); TSS200: 0-200 bases upstream of TSS; TSS1500: 200-1500 bases upstream of TSS, *Genomic context relative to CpG islands: open sea: >4 kb from CpG islands, **DMPs corrected for multiple testing at a FDR treshold of <=0.05

The detected methylated signals are mainly located in open sea genomic regions, but also reside at CpG islands (*IRS1*, *C2orf39, GRAMD4, TRIM2*), CpG island shores (*TLR9*, *NCAPG2*, *PSTPIP2*, *NTRK2*, *DNAJC14*, *CDO1*), and CpG island shelf regions (*CENPP*), and are enriched for functionally relevant gene structures such as promoter regions and gene bodies (table 3 and 4), which add support for the methylation variations associated with fathers preconception and early adolescent smoking onset to have potential functional implications.

Our single site EWAS findings were not reinforced in the subsequent DMR analyses, which only showed a suggestive hit for a region in DNTT (FDR adjusted p-value of 0.08) (table E8 in supplementary material, paper III). However, in additional *in silico* analyses exploring the biological context of methylation variations associated with fathers' adolescent smoking onset, 7 dmCpGs were significantly enriched for transcription factor binding sites (q-value<0.05) (table E6 in supplementary material, paper III), and several dmCpGs overlapped with promotor regions and correlated with gene expression in a variety of tissues in the EWAS atlas, as shown in figure 11A and B, respectively. Altogether, this provide additional support that our findings, even with relatively small magnitude, may be of functional and regulatory relevance.

Figure 11A: Correlation between dmCpGs and gene expression regulation in relation to genomic context

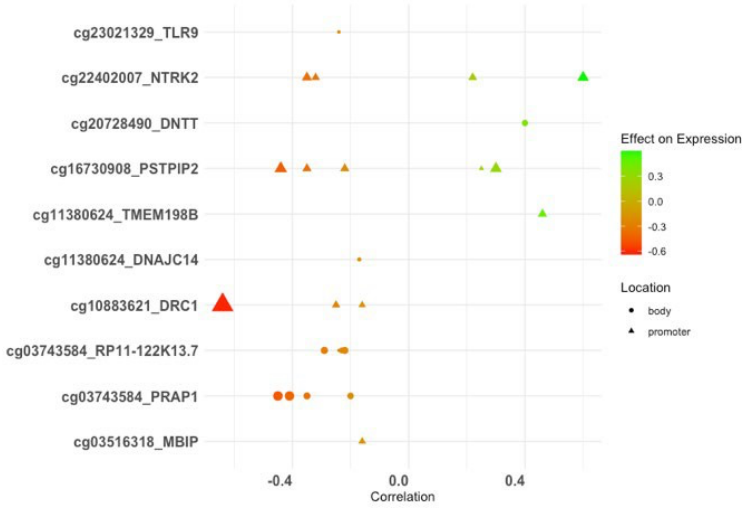


Figure 11B: Correlation between dmCpGs and gene expression regulation across different tissue types

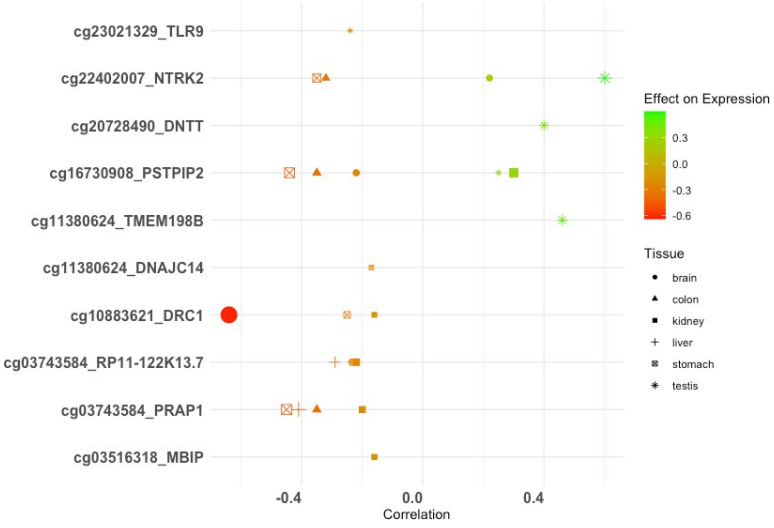


Figure 11A and B are reused with permission from the authors of submitted paper by Kibata N.T. et al., 2022

The top differentially methylated sites furthermore map to genes with putative roles in innate and apaptive immunity and inflammatory responses (*TLR9*, *DNTT*, *PSTPIP2*, *CSF1R*, *GRAMD4* , *TRIM2*) [318-323], as well as with glucose and fat metabolism (*IRS1*) [324, 325]. Of note, some of the annotated genes have also previously been reported as related to asthma development (*TLR9*, *CSF1R*, *NTRK2*) [326-328], lung function measures (KCNJ1, LEPROTL1) [329], as well as with body mass- and fat mass index (*GRAMD4*, *MYADML2*) [330], and other BMI related phenotypes in GWAS studies (*MBIP*, *NTRK2*, *TPCN1, KCNJ1, MYADML2*) [331-335].

The functional exploration of the top differentially methylated sites further support that the annotated genes are enriched in biological processes and molecular functions related to inflammatory responses, and innate immunity (figure 12), as well as with gene ontology pathways linked to immune regulation and insulin signaling (table E9 in supplementary material paper III).

According to previously published EWAS studies available from the EWAS atlas knowledgebase, the differentially methylated sites related to fathers' early adolescent smoking have also been associated with other traits such as autoimmune diseases, atopy, smoking and puberty, whereas the top 23dmCpGs, prevailing the lowest FDR corrected p-values in the EWAS study on fathers' any preconception smoking onset, have been found enriched for traits including immunoglobulin E (IgE) level, muscle hypertrophy, maternal smoking, and birthweight (figure 3B and 3A in paper III, repsectively). Although, this entails conjectures that our findings may provide additional epigenetic evidence for the previous epidemiological observations that link fathers' early adolescent smoking exposures to both increased asthma risk and lower lung function in the offspring [131, 133], we are aware that differentially methylated sites related to traits identified in previous association studies do not necessarily apply to our data, or our cohort. Moreover, the functional enrichment analyses are based on a list of genes annotated by the Genomic Browser in presence of being in closest

proximity to the methylated CpG sites. Even though this is commonly used for genomic annotation in EWAS studies, we are aware that CpG methylation and gene expression are not as straightforwardly correlated, and that metylated cytosines may exert their regulatory functions and impact the expression of both adjacent as well as distal genes.

Figure 12: Functional enrichment of methylation signals associated with fathers' adolescent smoking onset

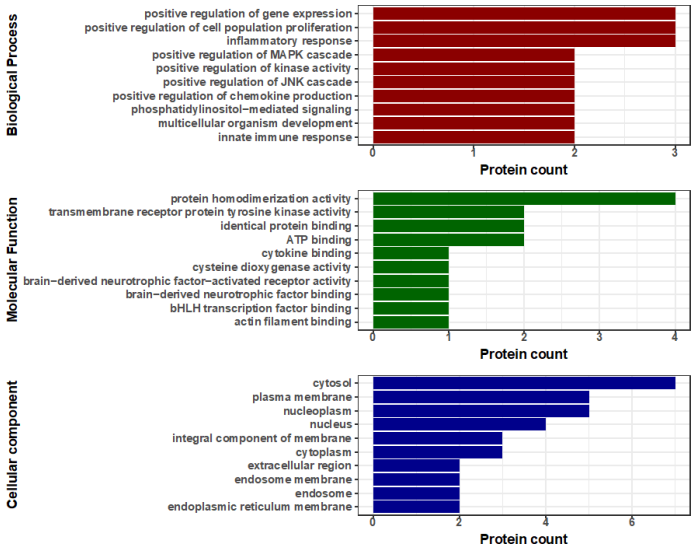


Figure 12 is reused with permission from the authors of submitted paper by Kibata N.T. et al., 2022

For this reason, we additionally investigated whether the identified dmCpGs associated with fathers' early adolescent smoking could also be related to phenotypic outcomes in the offspring. As shown in table 5, cg03380960 in FAM53B, cg12053348 (missing annotation), and cg22402007 in NTRK2 were associated with offspring's weight and BMI status, whereas cg22402007 in NTRK2 was associated with ever-asthma, and cg11380624 in DNAJC14 and cg10981514 in TPCN1 were related to ever-wheeze in the offspring.

Table 5: Associations between fathers' adolescent smoking related dmCpGs and phenotypic outcomes in the offspring (p-value <0.05)

Offspring outcomes	CpG	Beta	P-value	Gene
Asthma	cg22402007	10.790	0.014	NTRK2
Ever-wheezing	cg11380624	14.401	<0.001	DNAJC14
Ever-wheezing	cg10981514	-8.335	0.012	TPCN1
Weight	cg12053348	-51.081	0.001	Intergenic
Weight	cg03380960	42.575	0.023	FAM53B
Weight	cg22402007	-67.167	0.013	NTRK2
BMI	cg12053348	-0.001	0.004	Intergenic
BMI	cg03380960	0.002	0.008	FAM53B
BMI	cg22402007	-0.002	0.025	NTRK2
BMI	cg23021329	0.002	0.067	TLR9
BMI	cg11380624	-0.001	0.074	DNAJC14

Analyses are adjusted for offspring sex

Although the effect estimates associated with offspring's BMI are of small magnitude, the findings add support for the detected methylation sites to potentially be implicated in the development of metabolic and respiratory related phenotypes. Yet, future studies are needed in order to investigate whether associations between fathers' smoking exposures and offspring's risk of respiratory diseases and obesity truly are mediated via epigenetic alterations.

5.2.3 Fathers' smoking methylation markers in comparison to personal and maternal smoking signals

To further assess whether our identified methylated signals related to fathers' any preconception and early adolescent smoking onset were distinctly different from other smoking exposures, we ran two additional EWAS analyses to investigate the epigenome-wide associations on the offspring's own smoking, as well as their mothers' smoking during pregnacy and childhood. The top differentially signals detected in

relation to personal and maternal smoking are summarized in table 6 and 7, respectively.

Table 6: DMPs (FDR<0.05) associated with personal smoking

CpG	Chr.	Gene annotation*	Context**	Beta	P value	Adj.P***	Previous evidence of CpG site
cg05575921	5	AHRR	Body	Shore	-0.04	2.68E-54	1.94E-48	Joehanes et al.[246]; CHARGE[299]; Christiansen et al.[247]
cg21566642	2			Island	-0.04	8.21E-29	2.97E-23	Joehanes et al.[246]; CHARGE[299]; Christiansen et al.[247]
cg06644428	2			Island	-0.03	5.35E-20	1.29E-14	Joehanes et al.[246]; CHARGE[299]
cg01940273	2			Island	-0.03	1.50E-16	2.72E-11	CHARGE[299]; Christiansen et al.[247]
cg17739917	17	RARA	5'UTR	Shelf	-0.03	4.85E-16	6.77E-11	Christiansen et al.[247]
cg03636183	19	F2RL3	Body	Shore	-0.03	5.61E-16	6.77E-11	Joehanes et al.[246]; CHARGE[299]
cg21911711	19	F2RL3	TSS1500	Shore	-0.02	7.79E-12	8.05E-07	
cg25189904	1	GNG12	TSS1500	Shore	-0.03	6.96E-11	6.30E-06	Joehanes et al.[246]; CHARGE[299]; Christiansen et al.[247]
cg12806681	5	AHRR	Body	Shore	-0.01	1.31E-10	1.06E-05	Joehanes et al.[246]; CHARGE[299]; Christiansen et al.[247]
cg21161138	5	AHRR	Body	Open sea	-0.02	1.60E-10	1.16E-05	Joehanes et al.[246]; CHARGE[299]
cg09338374	22			Shelf	0.02	9.29E-10	6.11E-05	Christiansen et al.[247]
cg24838345	8	MTSS1	Body	Shelf	-0.02	1.44E-09	8.13E-05	Joehanes et al.[246]; CHARGE[299]; Christiansen et al.[247]
cg26703534	5	AHRR	Body	Shelf	-0.02	1.56E-09	8.13E-05	Joehanes et al.[246]; CHARGE[299]; Christiansen et al.[247]
cg14753356	6			Open sea	-0.022	1.57E-09	8.13E-05	Joehanes et al.[246]; CHARGE[299]; Christiansen et al.[247]
cg18110140	15			Open sea	-0.024	1.95E-08	0.00094207	Christiansen et al.[247]
cg26718213	2	SNED1	Body	Island	0.035	3.50E-08	0.00158405	Joehanes et al.[246]; CHARGE[299]
cg05086879	22	MGAT3	5'UTR	Open sea	-0.016	6.03E-08	0.00256768	Christiansen et al.[247]
cg21322436	7	CNTNAP2	TSS1500	Shore	-0.01	6.71E-08	0.00269941	Joehanes et al.[246]; CHARGE[299]; Christiansen et al.[247]
cg25648203	5	AHRR	Body	Open sea	-0.015	9.27E-08	0.00352967	Joehanes et al.[246]; CHARGE[299]; Christiansen et al.[247]
cg20832643	6	TRAF3IP2-AS1	Body	Open sea	-0.014	1.23E-07	0.00428032	Christiansen et al.[247]
cg09935388	1	GFI1	Body	Island	-0.026	1.24E-07	0.00428032	Joehanes et al.[246]; CHARGE[299]; Christiansen et al.[247]
cg18096787	21			Open sea	0.016	2.44E-07	0.00803438	
cg24090911	5	AHRR	Body	Open sea	-0.018	3.31E-07	0.01041909	Joehanes et al.[246]; CHARGE[299]; Christiansen et al.[247]
cg19859270	3	GPR15	1stExon	Open sea	-0.006	4.17E-07	0.01258376	Joehanes et al.[246]; CHARGE[299]
cg25159376	14	KLHDC1	TSS200	Island	-0.002	5.53E-07	0.01600192	
cg17025708	10	VTI1A	Body	Open sea	-0.006	6.49E-07	0.01783236	
cg02978227	3			Open sea	-0.006	6.65E-07	0.01783236	Christiansen et al.[247]
cg26707709	2	SNED1	Body	Island	0.017	1.28E-06	0.03257775	Joehanes et al.[246]; CHARGE[299]; Christiansen et al.[247]
cg23577033	10			Shelf	-0.005	1.31E-06	0.03257775	
cg25401612	12			Open sea	-0.03	2.02E-06	0.04794569	
cg13525276	14	TSHR	Body	Open sea	0.02	2.05E-06	0.04794569	Joehanes et al.[246]; CHARGE[299]; Christiansen et al.[247]

cg17260354	17	CDK3	3'UTR	Shore	0.014	2.13E-06	0.04819969	
cg06117824	1	TMEM51	Body	Shelf	-0.004	2.28E-06	0.04990539	
cg25949550	7	CNTNAP2	Body	Shore	-0.006	2.43E-06	0.05178064	Joehanes et al.[246]; CHARGE[299]
cg15342087	6			Open sea	-0.006	2.74E-06	0.05448434	Joehanes et al.[246]; CHARGE[299]
cg22635676	2	SNED1	Body	Island	0.046	2.79E-06	0.05448434	Joehanes et al.[246]; CHARGE[299]; Christiansen et al.[247]
cg21747070	5			Shore	-0.01	2.79E-06	0.05448434	Joehanes et al.[246]; CHARGE[299]

* TSS1500: 200-1500 bases upstream of TSS; 5´UTR: 5´untranslated region located between TSS and the start codon (ATG); TSS200: 0-200 bases upstream of TSS; 3´UTR: 3´untranslated region found immediately after the translation stop codon, **Genomic context relative to CpG island: Shore: up to 2 kb from flanking CpG islands; Shelf: up to 2 kb from flanking shores; Open sea: >4 kb from CpG islands; ***DMPs corrected for multiple testing at a FDR treshold of <0.05

Table 7: DMPs (FDR<0.05) associated with mothers' smoking

CpG	Chr.	Gene annotation*	Context**	Beta	P value	Adj.P***	Previous evidence of CpG site
cg19089201	7	MYO1G	3'UTR	Island	0.03	1.47E-10	0.0001	Joubert et al. [248]; PACE [299]
cg12803068	7	MYO1G	Body	Shore	0.06	3.31E-10	0.0001	Joubert et al. [248]; PACE [299]
cg05549655	15	CYP1A1	TSS1500	Island	0.02	6.91E-09	0.0015	Joubert et al. [248]; PACE [299]
cg14179389	1	GFI1	Body	Island	-0.03	8.48E-09	0.0015	Joubert et al. [248]; PACE [299]
cg04180046	7	MYO1G	Body	Island	0.03	1.73E-08	0.0025	Joubert et al. [248]; PACE [299]
cg25949550	7	CNTNAP2	Body	Shore	-0.01	2.10E-08	0.0025	Joubert et al. [248]; PACE [299]
cg13305373	17	RGS9	Body	Open sea	0.01	5.71E-08	0.0059	
cg22549041	15	CYP1A1	TSS1500	Island	0.03	1.63E-07	0.0147	Joubert et al. [248]; PACE [299]
cg05009104	7	MYO1G	Body	Shore	0.03	2.53E-07	0.0180	
cg11924019	15	CYP1A1	TSS1500	Island	0.02	2.63E-07	0.0180	Joubert et al. [248]; PACE [299]
cg12101586	15	CYP1A1	TSS1500	Island	0.03	2.74E-07	0.0180	Joubert et al. [248]; PACE [299]
cg05777089	1	BRINP3	5'UTR	Open sea	0.01	6.30E-07	0.0380	
cg13570656	15	CYP1A1	TSS1500	Island	0.03	6.84E-07	0.0381	Joubert et al. [248]; PACE [299]
cg04785284	16			Open sea	0.01	8.18E-07	0.0423	

*3´UTR: 3´untranslated region found immediately after the translation stop codon; TSS1500: 200-1500 bases upstream of TSS; 5´UTR: 5´untranslated region located between TSS and the start codon (ATG), **Genomic context relative to CpG island: Shore: up to 2 kb from flanking CpG islands; Open sea: >4 kb from CpG islands, ***DMPs corrected for multiple testing at a FDR treshold of <0.05

These results clearly emphasise what we have previously noted when comparing the methylation variants of fathers' smoking exposures to that of previously published signatures related to personal and sustained maternal during pregnancy; the 33 and 14 differentially methylated sites surpassing significance (FDR<0.05) in the EWAS analyses on offspring's own and their mothers' smoking exposure, respectively, are different to those associated with fathers' smoking. Moreover, we replicate common methylation sites associated with personal and maternal smoking previously detected in large meta-analyses and consortia based EWAS studies. Other than adding validity to our analytical methods, our findings further strengthen the credibility that the methylation signals identified in relation to fathers' smoking exposures are truly unique and different from the smoking related sites that are so far discovered.

Mounting evidence has demonstrated that exposure to tobacco smoke constituents can significantly alter epigenetic regulatory marks in gametes and be transmitted to the embryo [192-196]. Cigarette smoke has been shown both to significantly alter DNA methylation profiles in sperm [197], to persist during epigenetic reprogramming in embryonic development, and affect the methylome, expression levels, and metabolic function in F1 progeny [336]. Given that we particularly find fathers' early adolescent smoking onset to be associated with altered methylation levels in the offspring, and that these dmCpGs additionally are related to offspring's weight, BMI, ever-asthma and ever-wheeze, our results support observations from previous epidemiological studies, where exposures occurring specifically during pre- and pubertal years are associated with adverse health outcomes in the offspring [123, 131-133, 337]. Yet, whether the potential underlying mechanism involves modulation of sperm epigenomic reprogramming events during a pre- and pubertal transition phase, and whether this time period in fact confers a critically plastic and exposure sensitive age for smoking exposures to induce aberrant sperm DNA methylation signals that escape genome-wide erasure and are re-established in the subsequent generation, is left to conjecture, and needs to be addressed in future studies.

When inspecting the smoking induced sperm DNA methylation signals reported by Jenkins et al. [197], we found no evidence that these overlapped with the peripheral

blood based methylation sites we identified in relation to fathers' preconception and early smoking onset. However, experimental studies have observed that even though sperm DNA methylation variations in the parental F0 generation are found to exert a similar overall phenotype as the epigenetic and transcriptomic effects in somatic cells of the F1 offspring, they are not necessarily identified at the same genes and locations [338-340]. This concept is further supported by the observed variable DNA methylation states between individual sperm samples and the plausibility that these can give rise to an epigenetic inherited effect that is not constricted to exert a single gene specific effect on embryonic development [338].

Although we are left to speculate whether the exposure effects from fathers' smoking in fact may be propagated to the offspring, we are aware that the time points investigated of preconception and adolescent smoking are based on when the fathers' started to smoke, and consequently cannot disentangle whether the observed methylation changes truly are related to smoking exposures occurring in early adolescence and prior to conception, or rather reflects other smoking related parameters, such as an accumulating smoking exposure and number of years smoked. However, as commonly seen in population cohorts investigating smoking trajectories over time, smoking rarely occurs at only one specific time point, and most human studies are therefore underpowered to study smoking exposures solely transpiring in preconception years. Furthermore, as previously discussed in the methodological discussion, our EWAS studies have not been able to fully control for the potential confounding effects of fathers' accumulating second hand smoke exposure during the offspring's gestational period or in their postnatal years. However, a recent murine study found that exposure to cigarette smoke from the onset of puberty until 2 days prior to mating elicited altered miRNA expression levels in the spermatozoa and affected the body weights of the F1 progeny [171].

As recently highlighted by Breton and colleagues, evidence of a cross-generational inheritance of epigenetic marks, does not necessarily imply that these are exclusively environmentally induced or epigenetically transmitted [81]. As we have no concurrent genomic or transcriptomic information of the study subjects, we are aware that the

identified differentially methylated signals may be genetically driven and influenced by directly single nucleotide polymorphisms also associated with propensity to smoke, or indirectly by genetically determined transcriptional variation [290]. Beyond shared genetics, our study may be additionally constrained by factors attributable to that of shared familial environments. Although we found no evidence that our top differentially methylated signals were related to fathers' educational level in a sensitivity analysis, there may be other unmeasured aspects related to social class, such as diet, lifestyle habits and housing conditions, which all could potentially have influenced on our findings.

Finally, although 11 of the 19 CpG sites identified in relation to fathers early adolescent smoking showed nominal replication and concordant directional effects (correlation r=0.49, p-value=0.12) in the ALSPAC replication cohort, there was no overlap of differentially methylated sites (FDR<0.05). Given that relatively few offspring had fathers who started to smoke before age of 15 in both cohorts (n=64 RHINESSA, n=86 ALSPAC), we were most likely underpowered to confirm validation of our results and acknowledge that further investigations are required to demonstrate whether our preliminary and novel findings are replicable in other datasets.

5.2.4 Intergenerational impacts on BMI and fatmass in relation to preconceptional and postnatal time points of parental smoking onset

In this epidemiological paper we first and foremost set out to pursue whether parental smoking trajectories occuring in early adolescence (< age 15), as opposed to commencing at later preconception (≥ age 15) and postnatal time points, conferred potential critical periods for smoking exposures to potentiate cross-generational impacts on their adult offspring's body composition. To assess whether the phenotypic outcomes prevailed a sexual dimorphistic pattern, as attentuated by

findings from the ALSPAC cohort [120, 123], we additionally tested for potential interaction effects by the sex of the offspring.

We found that fathers' preconception smoking onset was associated with increased BMI as well as with increased fat mass (FMI) in adult offspring (age span 18-49 years of age), as shown in table 8 and 9, respectively. There was no sex-specific effect on offspring's BMI (p=0.395), but in the analysis on offspring's fat mass, we observed a tendency of effect modification by the sex of the offspring (p=0.014), indicative of a more consistent impact on the sons' body composition. However, discordantly to the previous findings in ALSPAC [120, 123, 337, 341], we did not find this male-specific association to be more pronounced if fathers' smoking commenced in early adolescence, but rather with smoking initiation occuring in later preconception years, from the age 15 and up to 2 years prior to birth of the offspring. Similarly, we also found fathers' postnatal smoking onset to be associated with increased fat mass in the sons (interaction p=0.020). Yet, the subsample with available bioimpedance measurements was inevitably constrained by too few observations within the various exposure groups, specifically for fathers' smoking onset occuring before age 15 as well as in postnatal years, which can have compromised our results, and thus leaves this analysis inconclusive and in need of further confirmation.

Table 8: Associations between fathers' smoking onset and offspring BMI (n= 2916)

	Sons' and daughters' BMI		
Fathers' smoking onset*	**Adj. BMI diff.**	**95% CI**	**P-value**
Preconception smoking onset < age 15	0.486	-0.196 – 1.169	0.162
Preconception smoking onset ≥ age 15	0.551	0.174 – 0.929	0.004**
Postnatal smoking onset	0.763	-0.692 – 2.217	0.304

observations within exposure groups: <15: n=303; ≥15: n=1162; postnatal: n=57, Estimates from generalized linear regression models with adjustment for offspring sex and fathers' education. Clustered by family id and study centre. P-value sign. level: .05, **.01, ***.001.

Table 9: Associations between fathers' smoking onset and offspring FMI (n= 129)

	Sons' and daughters' FMI			
Fathers' smoking onset*	**Adj. FMI diff.**	**95% CI**	**P-value**	**Int. sex P**
Preconception smoking onset < age 15	1.604	0.269 – 2.939	0.019**	0.982
Preconception smoking onset ≥ age 15	2.590	0.544 – 4.636	0.013**	0.014**[a]
Postnatal smoking onset	2.736	0.621 – 4.851	0.011**	0.020**[b]

observations within exposure groups: <15: n=12; ≥15: n=68; postnatal: n=8, a: smoking onset ≥15 female offspring: β: -2.797, CI: -5.023, -0.571, b: postnatal smoking onset female offspring: β: -3.041, CI: -5.599, -0.483 Estimates from generalized linear regression models with adjustment for offspring sex and fathers' education. Clustered by family id and study centre. P-value sign. level: .05, **.01, ***.001

We also found that mothers' preconceptional and postnatal time points of smoking onset were significantly related to increased BMI in the offspring. The adjusted regression coefficients are presented in table 10, and show that mothers' preconception smoking onset, both occuring in early adolescence as well as from age 15 and up to 2 years prior to the offspring birth, were associated with increased BMI in the offspring. We further found that the observed increase in BMI related to mothers preconception smoking commencing from age 15 were modified by the sex of the offspring, and were specific to males only (interaction p=0.010). There were no observed associations with mothers' preconception or postnatal smoking onset on offspring's fat mass. However, the subsample comprised only 111 subjects, and was as notably constrained as the fat mass analysis in the paternal lineage, and with comparably few observations of smoking commencing before the age of 15 as well as in postnatal years, and thus probably underpowered to detect true differences between the exposure groups.

Table 10: Associations between mothers' smoking onset and offspring BMI (n=3531)

	Sons' and daughters' BMI			
Mothers' smoking onset*	**Adj. BMI diff.**	**95% CI**	**P-value**	**Int. sex P**
Preconception smoking onset < age 15	1.161	0.378 – 1.944	0.004**	0.338
Preconception smoking onset ≥ age 15	0.720	0.293 – 1.147	0.001**	0.010**[a]
Postnatal smoking onset	2.257	1.220 – 3.294	< 0.011***	0.952

observations within exposure groups: <15: n=383; ≥15: n=1368; postnatal: n=91, a: smoking onset ≥15 female offspring: β: -0.720, CI: -1.264, -0.170, Estimates from generalized linear regression models with adjustment for offspring sex and fathers' education. Clustered by family id and study centre. P-value sign. level: .05, **.01, ***.001

Altogether, our study suggests that parental smoking, occuring at both preconception as well as postnatal time periods, is associated with increased BMI in the adult offspring. Furthermore, associations between smoking exposures and increased fat mass were only observed in the paternal lineage, and demonstrated as more consistent in sons than daughters. Even though the subsample with available fat mass data merit caution in interpretation, it lends support to previous reports of a potential paternal lineage tranmission of male-specific responses in offspring's obesity and metabolic related phenotypes [120, 123].

Indeed, this has also been substantiated by recent studies, which consistently have reported paternal smoking exposures occurring during preconceptional years to be largely associated with male offspring's obesity [337, 342, 343]. In a subsequent paper from the ALSPAC cohort, sons' of prepubertal smoking fathers were additionally found to have excess fat mass even detectable in their early adult years [337]. Accordingly, fathers' preconception smoking exposures have also been related to an earlier onset of metabolic syndrome in adult male offspring [344]. As our study also found increased BMI and FMI levels in offspring's who had reached adult years, taken together these findings suggest that fathers' smoking exposures may impose long-term impacts on their offspring's body composition and metabolic health.

Moreover, the preconception time period has been particularly underpinned in a recent study from the Children Lifeway Cohort in Shenzen, Southern China, which found that male children of fathers whose smoking exposures had solely occurred during preconceptional years, had an increased risk of childhood obesity [343]. The authors also found that preconception smoking, in conjuction with sustained smoking in postnatal years, were associated with increased overweight in the male offspring, but not with smoking exposures occurring in postnatal years alone [343]. Similar to what we find, this study also indicates that besides the prepubertal age, there might be other exposure sensitive periods in preconception years where fathers' smoking exposures can convey developmental and potentially sex specific modifications in the offspring's body composition. Experimental and human epigenetic studies have undeniably demonstrated that cigarette smoke metabolites can induce alterations and modulate the sperm epigenetic machinery [192-196]. Given that it has become increasingly apparent that the sperm epigenome, other than allowing sperm cells to develop and mature, also has pivotal roles in regulating and supporting embryo development [184, 185], this adds mechanistic and biological plausibility of a potential epigenetic pathway by which the fathers' preconception smoking exposures can affect the early life development and health trajectories in his offspring.

Although we find no evidence of increased fat mass in the maternal lineage, mothers' preconception and postnatal smoking exposures are associated with increased BMI in her adult offpsring. This is not surprising, given the substantial epidemiological evidence linking maternal smoking during pregnancy with adverse offspring outcomes and increased risk of obesity [345-348]. Nicotine and other tobacco constituents can cross the placenta, and affect fetal growth and development [349, 350], which besides factors like low birth weight and subsequent rapid weight gain in postnatal years, also have been associated with long-term risks of adiposity and metabolic changes later in life [351, 352]. This could be underpinning the increased BMI outcomes observed in our data, as well as to those reported from longitudinal studies, where the risk of offspring's obesity has been shown to increase by age [346, 353, 354], and even to persist into adulthood [355-357]. Moreover, sustained maternal smoking during

pregnancy has been associated with strong and extensive epigentic changes in the offspring [248, 299, 315], and with consistent effects across epigenetic marks and tissues [358], which also add plausibility for an epigenetic pathway by which mothers' smoking exposures can affect the future health of her children.

Although we find that both male and female offspring of mothers who started to smoke at either preconception or postnatal time points had higher BMI in adulthood, also in the maternal lineage we observed an indication of a male-specific BMI increase related to mothers' preconception smoking onset from age 15 and up to 2 years prior to the offspring's birth. Although there is limited evidence of sex-specific effects in relation to maternal smoking, some epigenome-wide association studies have suggested that male offspring may be more susceptible to DNA methylation changes at sites thought to be implicated in offspring's birtweights outcomes [314, 359]. However, to our knowledge, there are no comparaple findings on cross-generational body composition impacts from maternal smoking exposures occurring during early adolescent and preconception years.

Even so, obesity is a multifactorial condition, reflective of a complex combination of both shared genetic as well as lifestyle related factors [360-362]. This was also indicated in our independent mediation analyses, where the observed associations with time points of parental smoking onset on offspring's BMI outcomes were partially mediated via the parents' BMI status (table 5 and supplementary table S4 in paper II), as well as to the offspring's own smoking behaviours (table 5 supplementary table S5 in paper II). Similar to what we find, other studies have also emphasized the significance of genetic predisposition and shared environmental influences in the development of obesity, in that the joint effects of maternal smoking and obesity have been found to increase adult offspring's risk of obesity two-fold [363]. Due to the apparent effects maternal smoking has on fetal and infant growth, low birthweights have been suggested to be on the intermediate pathway between mothers' smoking exposures and subsequent offspring risk of obesity [351]. However, we found no mediation effects either via the sons' or the daughters' birthweights.

When we further investigated whether particularly parents' smoking initiation during preconceptional time points were related to other intermediating aspects of smoking exposures, such as to smoking intensity, we found that parents' packyears of smoking during the offspring's childhood fully mediated the associations between parental preconception smoking onset and the offspring's BMI (table 5 and supplementary table S3 in paper II). There was no mediation effect when packyears of smoking were restricted to occur in preconception years alone. In contrast to these results, others have found that the associations with fathers' preconception smoking on offspring's obesity and metabolic related outcomes showed a dose-dependent realtionship with other aspects of smoking exposures, and that the risk of obesity and an early onset of metabolic syndrom in adulthood were additionally increased by the great number of both years and cigarettes the fathers had smoked, as well as if smoking was initiated at an earlier age [343, 344]. This may be reflective of the direct toxicogenic effects of cigarette smoke, and to its impact on reactive oxygen species (ROS) production in germ cells [208, 364], which in experminental studies have been demonstrated to induce epigenetic changes and mediating metabolic phenotypes to the progeny generation [364-367]. As such, this may also be why our results indicate that rather the sustained and cumulative exposures associated with parental smoking, and not parental preconception smoking onset alone, are necessary in order to elicit long lasting influences on their adult offspring's BMI and risk of obesity.

With that being said, our mediation analyses were limited to offspring's BMI outcomes, which are based on the height and the weight of the study subjects, and as such is a rather inaccurate measure of body fat and obesity [282, 283]. Although underpowered to do so, we acknowledge it would be of greater pertinence to investigate potential intermediating factors in relation to offspring's fat mass, and with an outcome measure more specifically related to obesity and metabolic phenotypes [284, 368], particularly since the findings in the ALSPAC cohort largely and consistently have reported associations between fathers' prepubertal smoking onset

with excess fat mass in the subsequent offspring, and not BMI [120, 123, 337, 341, 369].

Furthermore, smoking behaviours and early smoking initiation are unavoidably related to patterns of socioeconomic inequalities [370]. Although we adjusted for parental education, and additionally clustered by family origin to account for the potential impact of shared familiar environments, particularly given the multifactorial aspects contributing to obesity [292, 293], we are aware that far more factors, attributable to social class or genetic background, not measured or investigated in this study, may underlie our associations. Moreover, early adolescent smoking of the grandpaternal mother has been associated with excess fat mass in the granddaughters [342, 369], which may very well have confounded our findings. Although probabilistic simulation techniques have previously found that unaccounted factors exerted limited impact on the association between parental preconception smoking exposures and offspring's phenotypic outcomes [132], this surmise may not hold for our findings, and as previously addressed in the methodological discussion, the presence of unmeasured confounding merit pertinent caution when interpreting the mediation results, as the sequential ignorability assumption may has been violated. As we have not been able to address the robustness of the mediation results in subsequent sensitivity analyses, future studies are needed in order to confirm whether the investigated intermediating factors are truly implicated in the observed associations between time points of parental smoking onset and offspring's BMI outcomes and potential risk of obesity.

6. Conclusions

Collectively, the three papers included in this thesis underpin the importance of accounting for paternal lineage smoking trajectories when investigating cross-generational impacts on epigenetic and phenotypic outcomes.

Our novel EWAS results indicate that fathers' smoking exposures are associated with differentially methylated signals in the offspring; this has to a large extent only been investigated in relation to personal and maternal smoking exposures.

Specifically, we find more methylation sites surpassing epigenome significance when investigating fathers' smoking exposures commencing in early adolescent years, before the age of 15. To our knowledge, no EWAS studies have previously explored associations between adolescent onset smoking and DNA methylation characteristics in offspring. The differentially methylated sites identified in relation to fathers' preconception and adolescent smoking onset were different to those we found in our EWAS analyses on personal and maternal smoking. Yet, although different CpG sites, they mapped to similar genes as those previously reported in large consortia-based and meta-analyses on personal and sustained maternal smoking during pregnancy. This adds to the credibility that the methylation signals truly reflect an effect of smoking. We suggest these signals present as canditates for validation in future studies.

Parents' smoking exposures were associated with increased BMI outcomes in adult offspring. However, only fathers' smoking was also related to increased fat mass, with more consistent associations observed in their sons. This lends support to previous studies reporting a specific paternal lineage transmission of male-specific responses on offspring's body composition and obesity-related phenotypes. However, future studies are needed in order to pursue whether the cross-generational impacts from fathers' smoking are mediated via epigenetic alterations.

7. Future implications

The possibility that paternal exposures during early adolescent years might impact future offspring health, has wide public health implications – prioritising resources to improve the environment and health of older children may possibly improve the health of both those targeted by the intervention and their future offspring. This research thus opens for particularly efficient novel strategies to improve public health.

Although it has been increasingly regognized that the sperm epigenome is higly plastic and responsive to tobacco smoke constituents during germ cell development and maturation, and thus confers a potential molecular mechanism for a cross-generational transmission of altered epigenetic states, to date intergenerational cohort studies have been limited by a scarcity of detailed data on future fathers' cigarette smoke consumption, particularly with regard to early adolescent years, in cohorts with DNA methylation data. This thesis supports the urgency for intergenerational population studies to obtain exposure data from fathers and for this distinct time period. Moreover, obtaining exposure data on early adolescent smoking initiation in ongoing two- generation cohorts will be critically important for conducting paternal lineage studies with samples that are sufficiently large to detect small to moderate associations. Besides improving generalisability of study results, this will also facilitate the use of auxiliary approaches such as Mendelian Ransomization to allow for statistical inference of the causal effects of CpG sites associated with fathers' smoking. The collection of adequately powered studies will also be of particular importance for pursuing replication of our preliminary findings. To date, to our knowledge, there are no suitable replication cohorts that allow us to test the replicability of methylation sites identified in relation to fathers' early adolescent smoking.

In order to further unravel mechanisms for potential father to offspring transmission of gametic methylation changes, it will be of great importance to investigate sperm epigenetic signals in samples of preconception smoking fathers' versus non-smoking controls. By concurrent investigating epigenetic DNA methylation somatic changes in

the offspring, this will advance our knowledge on whether aberrant sperm epigenetic marks associated with fathers' smoking, in fact escape genome-wide erasure and are re-established in the subsequent generation, and affect their epigenomic and phenotypic outcomes. However due to clearly apparent ethical reasons, human studies will not be able to address smoking specific sperm epigenetic alterations occurring during early adolescent years. Neither do human studies allow for exposures to be investigated in isolation, and there will always be the presence of a range of potential confounding factors. For this reason, collaborative experimental resarch in experimental model systems such as mice will be important and provide mechanistic insights, as there are non-invasive approaches, such as the balano-preputial separation test that can be used to assess pubertal onset. This will be important to advance our knowledge on the sperm epigenomic impacts of smoking exposures solely occurring during early adolescent years.

Finally, by integrating multiple omics analyses in EWAS studies, and simultaneosuly extract genotype information from the individuals as well as transcriptional RNA from the same cell populations used to assess DNA methylation patterns, will allow to assement of DNA methylation variability due to DNA sequence differences and transcriptional changes, and thus dimish confounding by such biological influences. Taken together this will generate a comprehensive mechanistic understanding of the cross-generational epigenomic and phenotypic consequences of fathers' smoking exposures.

8. References

1. Jablonka, E. and M.J. Lamb, *The changing concept of epigenetics.* Ann N Y Acad Sci, 2002. **981**: p. 82-96.
2. Esteller, M., *Epigenetics in evolution and disease.* The Lancet, 2008. **372**: p. S90-S96.
3. Gam, R., M. Sung, and A. Prasad Pandurangan, *Experimental and Computational Approaches to Direct Cell Reprogramming: Recent Advancement and Future Challenges.* Cells, 2019. **8**(10): p. 1189.
4. Tost, J., *10 years of Epigenomics: a journey with the epigenetic community through exciting times.* Epigenomics, 2020. **12**(2): p. 81-85.
5. Berger, S.L., et al., *An operational definition of epigenetics.* Genes Dev, 2009. **23**(7): p. 781-3.
6. Arif, M., et al., *Epigenetic modification: a regulatory mechanism in essential hypertension.* Hypertension Research, 2019. **42**(8): p. 1099-1113.
7. Bird, A., *DNA methylation patterns and epigenetic memory.* Genes Dev, 2002. **16**(1): p. 6-21.
8. Bernstein, B.E., A. Meissner, and E.S. Lander, *The mammalian epigenome.* Cell, 2007. **128**(4): p. 669-81.
9. Hemberger, M., W. Dean, and W. Reik, *Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal.* Nat Rev Mol Cell Biol, 2009. **10**(8): p. 526-37.
10. Shipony, Z., et al., *Dynamic and static maintenance of epigenetic memory in pluripotent and somatic cells.* Nature, 2014. **513**(7516): p. 115-9.
11. Reik, W., *Stability and flexibility of epigenetic gene regulation in mammalian development.* Nature, 2007. **447**(7143): p. 425-32.
12. Probst, A.V., E. Dunleavy, and G. Almouzni, *Epigenetic inheritance during the cell cycle.* Nat Rev Mol Cell Biol, 2009. **10**(3): p. 192-206.
13. Smith, Z.D. and A. Meissner, *DNA methylation: roles in mammalian development.* Nat Rev Genet, 2013. **14**(3): p. 204-20.
14. Lyko, F., *The DNA methyltransferase family: a versatile toolkit for epigenetic regulation.* Nat Rev Genet, 2018. **19**(2): p. 81-92.
15. Mortusewicz, O., et al., *Recruitment of DNA methyltransferase I to DNA repair sites.* Proc Natl Acad Sci U S A, 2005. **102**(25): p. 8905-9.
16. Hermann, A., R. Goyal, and A. Jeltsch, *The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites.* J Biol Chem, 2004. **279**(46): p. 48350-9.
17. Ooi, S.K., et al., *DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA.* Nature, 2007. **448**(7154): p. 714-7.
18. Jia, D., et al., *Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation.* Nature, 2007. **449**(7159): p. 248-51.
19. Jeltsch, A. and R.Z. Jurkowska, *New concepts in DNA methylation.* Trends Biochem Sci, 2014. **39**(7): p. 310-8.
20. Tate, P.H. and A.P. Bird, *Effects of DNA methylation on DNA-binding proteins and gene expression.* Curr Opin Genet Dev, 1993. **3**(2): p. 226-31.
21. Schübeler, D., *Function and information content of DNA methylation.* Nature, 2015. **517**(7534): p. 321-6.

22. Nan, X., et al., *Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex.* Nature, 1998. **393**(6683): p. 386-9.
23. Kohli, R.M. and Y. Zhang, *TET enzymes, TDG and the dynamics of DNA demethylation.* Nature, 2013. **502**(7472): p. 472-9.
24. Walsh, C.P., J.R. Chaillet, and T.H. Bestor, *Transcription of IAP endogenous retroviruses is constrained by cytosine methylation.* Nat Genet, 1998. **20**(2): p. 116-7.
25. Bird, A., et al., *A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA.* Cell, 1985. **40**(1): p. 91-9.
26. Vavouri, T. and B. Lehner, *Human genes with CpG island promoters have a distinct transcription-associated chromatin organization.* Genome Biol, 2012. **13**(11): p. R110.
27. Hellman, A. and A. Chess, *Gene body-specific methylation on the active X chromosome.* Science, 2007. **315**(5815): p. 1141-3.
28. Ball, M.P., et al., *Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells.* Nat Biotechnol, 2009. **27**(4): p. 361-8.
29. Aran, D., et al., *Replication timing-related and gene body-specific methylation of active human genes.* Hum Mol Genet, 2011. **20**(4): p. 670-80.
30. Michels, K.B., *The promises and challenges of epigenetic epidemiology.* Exp Gerontol, 2010. **45**(4): p. 297-301.
31. Eckhardt, F., et al., *DNA methylation profiling of human chromosomes 6, 20 and 22.* Nat Genet, 2006. **38**(12): p. 1378-85.
32. Jaffe, A.E. and R.A. Irizarry, *Accounting for cellular heterogeneity is critical in epigenome-wide association studies.* Genome Biol, 2014. **15**(2): p. R31.
33. Langevin, S.M., et al., *The influence of aging, environmental exposures and local sequence features on the variation of DNA methylation in blood.* Epigenetics, 2011. **6**(7): p. 908-19.
34. Bock, C., et al., *CpG island methylation in human lymphocytes is highly correlated with DNA sequence, repeats, and predicted DNA structure.* PLoS Genet, 2006. **2**(3): p. e26.
35. Swygert, S.G. and C.L. Peterson, *Chromatin dynamics: interplay between remodeling enzymes and histone modifications.* Biochim Biophys Acta, 2014. **1839**(8): p. 728-36.
36. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications.* Cell Res, 2011. **21**(3): p. 381-95.
37. Finch, J.T., et al., *Structure of nucleosome core particles of chromatin.* Nature, 1977. **269**(5623): p. 29-36.
38. Allis, C.D. and T. Jenuwein, *The molecular hallmarks of epigenetic control.* Nat Rev Genet, 2016. **17**(8): p. 487-500.
39. Roth, S.Y., J.M. Denu, and C.D. Allis, *Histone acetyltransferases.* Annu Rev Biochem, 2001. **70**: p. 81-120.
40. Kaniskan, H., M.L. Martini, and J. Jin, *Inhibitors of Protein Methyltransferases and Demethylases.* Chem Rev, 2018. **118**(3): p. 989-1068.
41. Hyun, K., et al., *Writing, erasing and reading histone lysine methylations.* Exp Mol Med, 2017. **49**(4): p. e324.
42. Rossetto, D., N. Avvakumov, and J. Côté, *Histone phosphorylation: a chromatin modification involved in diverse nuclear events.* Epigenetics, 2012. **7**(10): p. 1098-108.

43. Zippo, A., et al., *Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation.* Cell, 2009. **138**(6): p. 1122-36.
44. Dinger, M.E., et al., *Differentiating protein-coding and noncoding RNA: challenges and ambiguities.* PLoS Comput Biol, 2008. **4**(11): p. e1000176.
45. Wei, J.W., et al., *Non-coding RNAs as regulators in epigenetics (Review).* Oncol Rep, 2017. **37**(1): p. 3-9.
46. Panni, S., et al., *Non-coding RNA regulatory networks.* Biochim Biophys Acta Gene Regul Mech, 2020. **1863**(6): p. 194417.
47. Statello, L., et al., *Gene regulation by long non-coding RNAs and its biological functions.* Nat Rev Mol Cell Biol, 2021. **22**(2): p. 96-118.
48. Chen, J. and Y. Xue, *Emerging roles of non-coding RNAs in epigenetic regulation.* Sci China Life Sci, 2016. **59**(3): p. 227-35.
49. Suzuki, K., et al., *Prolonged transcriptional silencing and CpG methylation induced by siRNAs targeted to the HIV-1 promoter region.* J RNAi Gene Silencing, 2005. **1**(2): p. 66-78.
50. Morris, K.V., et al., *Small interfering RNA-induced transcriptional gene silencing in human cells.* Science, 2004. **305**(5688): p. 1289-92.
51. Bayne, E.H. and R.C. Allshire, *RNA-directed transcriptional gene silencing in mammals.* Trends Genet, 2005. **21**(7): p. 370-3.
52. Kuramochi-Miyagawa, S., et al., *DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes.* Genes Dev, 2008. **22**(7): p. 908-17.
53. Gregory, R.I., et al., *Human RISC couples microRNA biogenesis and posttranscriptional gene silencing.* Cell, 2005. **123**(4): p. 631-40.
54. Denis, H., M.N. Ndlovu, and F. Fuks, *Regulation of mammalian DNA methyltransferases: a route to new mechanisms.* EMBO Rep, 2011. **12**(7): p. 647-56.
55. Yuan, J.H., et al., *The histone deacetylase 4/SP1/microrna-200a regulatory network contributes to aberrant histone acetylation in hepatocellular carcinoma.* Hepatology, 2011. **54**(6): p. 2025-35.
56. Tuddenham, L., et al., *The cartilage specific microRNA-140 targets histone deacetylase 4 in mouse cells.* FEBS Lett, 2006. **580**(17): p. 4214-7.
57. Fabbri, M., et al., *MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B.* Proc Natl Acad Sci U S A, 2007. **104**(40): p. 15805-10.
58. Benetti, R., et al., *A mammalian microRNA cluster controls DNA methylation and telomere recombination via Rbl2-dependent regulation of DNA methyltransferases.* Nat Struct Mol Biol, 2008. **15**(9): p. 998.
59. Sinkkonen, L., et al., *MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells.* Nat Struct Mol Biol, 2008. **15**(3): p. 259-67.
60. Isoda, T., et al., *Non-coding Transcription Instructs Chromatin Folding and Compartmentalization to Dictate Enhancer-Promoter Communication and T Cell Fate.* Cell, 2017. **171**(1): p. 103-119.e18.
61. Mumbach, M.R., et al., *HiChIRP reveals RNA-associated chromosome conformation.* Nat Methods, 2019. **16**(6): p. 489-492.

62. Dueva, R., et al., *Neutralization of the Positive Charges on Histone Tails by RNA Promotes an Open Chromatin Structure.* Cell Chem Biol, 2019. **26**(10): p. 1436-1449.e5.
63. Tsai, M.C., et al., *Long noncoding RNA as modular scaffold of histone modification complexes.* Science, 2010. **329**(5992): p. 689-93.
64. Rinn, J.L. and H.Y. Chang, *Genome regulation by long noncoding RNAs.* Annu Rev Biochem, 2012. **81**: p. 145-66.
65. Kotzin, J.J., et al., *The long non-coding RNA Morrbid regulates Bim and short-lived myeloid cell lifespan.* Nature, 2016. **537**(7619): p. 239-243.
66. Beckedorff, F.C., et al., *The intronic long noncoding RNA ANRASSF1 recruits PRC2 to the RASSF1A promoter, reducing the expression of RASSF1A and increasing cell proliferation.* PLoS Genet, 2013. **9**(8): p. e1003705.
67. Marín-Béjar, O., et al., *The human lncRNA LINC-PINT inhibits tumor cell invasion through a highly conserved sequence element.* Genome Biol, 2017. **18**(1): p. 202.
68. Fanucchi, S., et al., *Immune genes are primed for robust transcription by proximal long noncoding RNAs located in nuclear compartments.* Nat Genet, 2019. **51**(1): p. 138-150.
69. Wang, K.C., et al., *A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression.* Nature, 2011. **472**(7341): p. 120-4.
70. Mondal, T., et al., *MEG3 long noncoding RNA regulates the TGF-β pathway genes through formation of RNA-DNA triplex structures.* Nat Commun, 2015. **6**: p. 7743.
71. O'Leary, V.B., et al., *PARTICLE, a Triplex-Forming Long ncRNA, Regulates Locus-Specific Methylation in Response to Low-Dose Irradiation.* Cell Rep, 2015. **11**(3): p. 474-85.
72. Grote, P., et al., *The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse.* Dev Cell, 2013. **24**(2): p. 206-14.
73. Schmitz, K.M., et al., *Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes.* Genes Dev, 2010. **24**(20): p. 2264-9.
74. Martianov, I., et al., *Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript.* Nature, 2007. **445**(7128): p. 666-70.
75. Jaenisch, R. and A. Bird, *Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals.* Nat Genet, 2003. **33 Suppl**: p. 245-54.
76. Feinberg, A.P., *Phenotypic plasticity and the epigenetics of human disease.* Nature, 2007. **447**(7143): p. 433-40.
77. Kanherkar, R.R., N. Bhatia-Dey, and A.B. Csoka, *Epigenetics across the human lifespan.* Front Cell Dev Biol, 2014. **2**: p. 49.
78. Waterland, R.A. and K.B. Michels, *Epigenetic Epidemiology of the Developmental Origins Hypothesis.* Annual Review of Nutrition, 2007. **27**(1): p. 363-388.
79. Gluckman, P.D., M.A. Hanson, and T. Buklijas, *A conceptual framework for the developmental origins of health and disease.* J Dev Orig Health Dis, 2010. **1**(1): p. 6-18.
80. Heindel, J.J., et al., *Review of developmental origins of health and disease publications in environmental epidemiology.* Reproductive Toxicology, 2017. **68**: p. 34-48.

81. Breton, C.V., et al., *Exploring the evidence for epigenetic regulation of environmental influences on child health across generations.* Communications Biology, 2021. **4**(1): p. 769.
82. Soubry, A., *Epigenetics as a Driver of Developmental Origins of Health and Disease: Did We Forget the Fathers?* Bioessays, 2018. **40**(1).
83. Holliday, R., *The inheritance of epigenetic defects.* Science, 1987. **238**(4824): p. 163-170.
84. Mørkve Knudsen, T., et al., *Transgenerational and intergenerational epigenetic inheritance in allergic diseases.* Journal of Allergy and Clinical Immunology, 2018. **142**(3): p. 765-772.
85. Hales, B.F., et al., *Epigenetic programming: from gametes to blastocyst.* Birth Defects Res A Clin Mol Teratol, 2011. **91**(8): p. 652-65.
86. Ly, L., D. Chan, and J.M. Trasler, *Developmental windows of susceptibility for epigenetic inheritance through the male germline.* Seminars in Cell & Developmental Biology, 2015. **43**: p. 96-105.
87. Messerschmidt, D.M., B.B. Knowles, and D. Solter, *DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos.* Genes Dev, 2014. **28**(8): p. 812-28.
88. Li, E., *Chromatin modification and epigenetic reprogramming in mammalian development.* Nat Rev Genet, 2002. **3**(9): p. 662-73.
89. Santos, F. and W. Dean, *Epigenetic reprogramming during early development in mammals.* Reproduction, 2004. **127**(6): p. 643-51.
90. Morgan, H.D., et al., *Epigenetic reprogramming in mammals.* Hum Mol Genet, 2005. **14 Spec No 1**: p. R47-58.
91. Reik, W. and M.A. Surani, *Germline and Pluripotent Stem Cells.* Cold Spring Harb Perspect Biol, 2015. **7**(11).
92. Wu, H., et al., *Environmental Susceptibility of the Sperm Epigenome During Windows of Male Germ Cell Development.* Current Environmental Health Reports, 2015. **2**(4): p. 356-366.
93. Heard, E. and R.A. Martienssen, *Transgenerational epigenetic inheritance: myths and mechanisms.* Cell, 2014. **157**(1): p. 95-109.
94. Hill, P.W., R. Amouroux, and P. Hajkova, *DNA demethylation, Tet proteins and 5-hydroxymethylcytosine in epigenetic reprogramming: an emerging complex story.* Genomics, 2014. **104**(5): p. 324-33.
95. Ben Maamar, M., E.E. Nilsson, and M.K. Skinner, *Epigenetic transgenerational inheritance, gametogenesis and germline development†.* Biol Reprod, 2021. **105**(3): p. 570-592.
96. Kagiwada, S., et al., *Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice.* Embo j, 2013. **32**(3): p. 340-53.
97. Gkountela, S., et al., *The ontogeny of cKIT+ human primordial germ cells proves to be a resource for human germ line reprogramming, imprint erasure and in vitro differentiation.* Nat Cell Biol, 2013. **15**(1): p. 113-22.
98. Jinhyuk, B., et al., *PGC-Enriched miRNAs Control Germ Cell Development.* Molecules and Cells, 2015. **38**(10): p. 895-903.
99. He, C., et al., *Roles of Noncoding RNA in Reproduction.* Front Genet, 2021. **12**: p. 777510.

100. Zhu, Q., et al., *Small Noncoding RNAs in Reproduction and Infertility.* Biomedicines, 2021. **9**(12).
101. von Meyenn, F., et al., *Comparative Principles of DNA Methylation Reprogramming during Human and Mouse In Vitro Primordial Germ Cell Specification.* Dev Cell, 2016. **39**(1): p. 104-115.
102. Zhang, C., L. Gao, and E.Y. Xu, *LncRNA, a new component of expanding RNA-protein regulatory network important for animal sperm development.* Semin Cell Dev Biol, 2016. **59**: p. 110-117.
103. Taylor, D.H., et al., *Long non-coding RNA regulation of reproduction and development.* Mol Reprod Dev, 2015. **82**(12): p. 932-56.
104. Hackett, J.A., et al., *Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine.* Science, 2013. **339**(6118): p. 448-52.
105. Lane, N., et al., *Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse.* Genesis, 2003. **35**(2): p. 88-93.
106. Oakes, C.C., et al., *Developmental acquisition of genome-wide DNA methylation occurs prior to meiosis in male germ cells.* Dev Biol, 2007. **307**(2): p. 368-79.
107. Chong, S., et al., *Modifiers of epigenetic reprogramming show paternal effects in the mouse.* Nat Genet, 2007. **39**(5): p. 614-22.
108. Basu, A., et al., *DNMT3L enables accumulation and inheritance of epimutations in transgenic Drosophila.* Sci Rep, 2016. **6**: p. 19572.
109. Lombó, M., et al., *Embryonic Exposure to Bisphenol A Impairs Primordial Germ Cell Migration without Jeopardizing Male Breeding Capacity.* Biomolecules, 2019. **9**(8).
110. Zhang, X.F., et al., *Bisphenol A exposure modifies DNA methylation of imprint genes in mouse fetal germ cells.* Mol Biol Rep, 2012. **39**(9): p. 8621-8.
111. Skinner, M.K., et al., *Environmentally induced transgenerational epigenetic reprogramming of primordial germ cells and the subsequent germ line.* PLoS One, 2013. **8**(7): p. e66318.
112. Brieño-Enríquez, M.A., et al., *Exposure to endocrine disruptor induces transgenerational epigenetic deregulation of microRNAs in primordial germ cells.* PLoS One, 2015. **10**(4): p. e0124296.
113. Ren, J., et al., *Intrauterine hyperglycemia exposure results in intergenerational inheritance via DNA methylation reprogramming on F1 PGCs.* Epigenetics Chromatin, 2018. **11**(1): p. 20.
114. Radford, E.J., et al., *In utero effects. In utero undernourishment perturbs the adult sperm methylome and intergenerational metabolism.* Science, 2014. **345**(6198): p. 1255903.
115. Simorangkir, D.R., et al., *Prepubertal expansion of dark and pale type A spermatogonia in the rhesus monkey (Macaca mulatta) results from proliferation during infantile and juvenile development in a relatively gonadotropin independent manner.* Biol Reprod, 2005. **73**(6): p. 1109-15.
116. Toro, C.A., C.F. Aylwin, and A. Lomniczi, *Hypothalamic epigenetics driving female puberty.* J Neuroendocrinol, 2018. **30**(7): p. e12589.
117. Vazquez, M.J., et al., *SIRT1 mediates obesity- and nutrient-dependent perturbation of pubertal timing by epigenetically controlling Kiss1 expression.* Nat Commun, 2018. **9**(1): p. 4194.
118. Motti, M.L. and R. Meccariello, *Minireview: The Epigenetic Modulation of KISS1 in Reproduction and Cancer.* Int J Environ Res Public Health, 2019. **16**(14).

119. Lomniczi, A., et al., *Epigenetic control of female puberty.* Nat Neurosci, 2013. **16**(3): p. 281-9.
120. Pembrey, M.E., et al., *Sex-specific, male-line transgenerational responses in humans.* Eur J Hum Genet, 2006. **14**(2): p. 159-66.
121. Bygren, L.O., G. Kaati, and S. Edvinsson, *Longevity determined by paternal ancestors' nutrition during their slow growth period.* Acta Biotheor, 2001. **49**(1): p. 53-9.
122. Kaati, G., L.O. Bygren, and S. Edvinsson, *Cardiovascular and diabetes mortality determined by nutrition during parents' and grandparents' slow growth period.* Eur J Hum Genet, 2002. **10**(11): p. 682-8.
123. Northstone, K., et al., *Prepubertal start of father's smoking and increased body fat in his sons: further characterisation of paternal transgenerational responses.* Eur J Hum Genet, 2014. **22**(12): p. 1382-6.
124. Mocarelli, P., et al., *Dioxin exposure, from infancy through puberty, produces endocrine disruption and affects human semen quality.* Environ Health Perspect, 2008. **116**(1): p. 70-7.
125. Messina, A., et al., *A microRNA switch regulates the rise in hypothalamic GnRH production before puberty.* Nat Neurosci, 2016. **19**(6): p. 835-44.
126. Zhu, Q., et al., *Small Noncoding RNAs in Reproduction and Infertility.* Biomedicines, 2021. **9**(12): p. 1884.
127. Montellier, E., et al., *Chromatin-to-nucleoprotamine transition is controlled by the histone H2B variant TH2B.* Genes & development, 2013. **27**(15): p. 1680-1692.
128. Klenova, E.M., et al., *The novel BORIS + CTCF gene family is uniquely involved in the epigenetics of normal biology and cancer.* Semin Cancer Biol, 2002. **12**(5): p. 399-414.
129. Li, H., et al., *Low doses of carbendazim and chlorothalonil synergized to impair mouse spermatogenesis through epigenetic pathways.* Ecotoxicol Environ Saf, 2020. **188**: p. 109908.
130. Gao, Y., et al., *Pubertal exposure to low doses of zearalenone disrupting spermatogenesis through ERα related genetic and epigenetic pathways.* Toxicol Lett, 2019. **315**: p. 31-38.
131. Accordini, S., et al., *Prenatal and prepubertal exposures to tobacco smoke in men may cause lower lung function in future offspring: a three-generation study using a causal modelling approach.* Eur Respir J, 2021. **58**(4).
132. Accordini, S., et al., *A three-generation study on the association of tobacco smoking with asthma.* Int J Epidemiol, 2018. **47**(4): p. 1106-1117.
133. Svanes, C., et al., *Father's environment before conception and asthma risk in his children: a multi-generation analysis of the Respiratory Health In Northern Europe study.* Int J Epidemiol, 2017. **46**(1): p. 235-245.
134. Zirkin, B.R., et al., *Is FSH required for adult spermatogenesis?* J Androl, 1994. **15**(4): p. 273-6.
135. Sinkevicius, K.W., et al., *Estrogen-dependent and -independent estrogen receptor-alpha signaling separately regulate male fertility.* Endocrinology, 2009. **150**(6): p. 2898-905.
136. Goudarzi, A., et al., *Genome-scale acetylation-dependent histone eviction during spermatogenesis.* J Mol Biol, 2014. **426**(20): p. 3342-9.

137. Goudarzi, A., et al., *Dynamic Competing Histone H4 K5K8 Acetylation and Butyrylation Are Hallmarks of Highly Active Gene Promoters.* Mol Cell, 2016. **62**(2): p. 169-180.
138. Bao, J. and M.T. Bedford, *Epigenetic regulation of the histone-to-protamine transition during spermiogenesis.* Reproduction, 2016. **151**(5): p. R55-70.
139. Chen, X., et al., *The roles of microRNAs in regulation of mammalian spermatogenesis.* Journal of Animal Science and Biotechnology, 2017. **8**(1): p. 35.
140. Dai, L., et al., *Testis-specific miRNA-469 up-regulated in gonadotropin-regulated testicular RNA helicase (GRTH/DDX25)-null mice silences transition protein 2 and protamine 2 messages at sites within coding region: implications of its role in germ cell development.* J Biol Chem, 2011. **286**(52): p. 44306-18.
141. Yu, Z., T. Raabe, and N.B. Hecht, *MicroRNA Mirn122a reduces expression of the posttranscriptionally regulated germ cell transition protein 2 (Tnp2) messenger RNA (mRNA) by mRNA cleavage.* Biol Reprod, 2005. **73**(3): p. 427-33.
142. Güneş, S. and T. Kulaç, *The role of epigenetics in spermatogenesis.* Turk J Urol, 2013. **39**(3): p. 181-7.
143. Hammoud, S.S., et al., *Distinctive chromatin in human sperm packages genes for embryo development.* Nature, 2009. **460**(7254): p. 473-8.
144. Gatewood, J.M., et al., *Sequence-specific packaging of DNA in human sperm chromatin.* Science, 1987. **236**(4804): p. 962-4.
145. Ben Maamar, M., et al., *Epigenetic Transgenerational Inheritance of Altered Sperm Histone Retention Sites.* Sci Rep, 2018. **8**(1): p. 5308.
146. Hamad, M.F., et al., *Impact of cigarette smoking on histone (H2B) to protamine ratio in human spermatozoa and its relation to sperm parameters.* Andrology, 2014. **2**(5): p. 666-77.
147. Lettieri, G., et al., *Discovery of the Involvement in DNA Oxidative Damage of Human Sperm Nuclear Basic Proteins of Healthy Young Men Living in Polluted Areas.* Int J Mol Sci, 2020. **21**(12).
148. Siklenka, K., et al., *Disruption of histone methylation in developing sperm impairs offspring health transgenerationally.* Science, 2015. **350**(6261): p. aab2006.
149. Lesch, B.J., et al., *Intergenerational epigenetic inheritance of cancer susceptibility in mammals.* Elife, 2019. **8**.
150. Ihara, M., et al., *Paternal poly (ADP-ribose) metabolism modulates retention of inheritable sperm histones and early embryonic gene expression.* PLoS Genet, 2014. **10**(5): p. e1004317.
151. Skinner, M.K., et al., *Alterations in sperm DNA methylation, non-coding RNA and histone retention associate with DDT-induced epigenetic transgenerational inheritance of disease.* Epigenetics Chromatin, 2018. **11**(1): p. 8.
152. Ben Maamar, M., et al., *Alterations in sperm DNA methylation, non-coding RNA expression, and histone retention mediate vinclozolin-induced epigenetic transgenerational inheritance of disease.* Environ Epigenet, 2018. **4**(2): p. dvy010.
153. Nilsson, E.E., I. Sadler-Riggleman, and M.K. Skinner, *Environmentally induced epigenetic transgenerational inheritance of disease.* Environ Epigenet, 2018. **4**(2): p. dvy016.
154. Jones, R.E. and K.H. Lopez, *Human reproductive biology*. 2013: Academic Press.
155. Donkin, I. and R. Barrès, *Sperm epigenetics and influence of environmental factors.* Mol Metab, 2018. **14**: p. 1-11.

156. Sharma, U., et al., *Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals.* Science, 2016. **351**(6271): p. 391-396.
157. Sharma, U., et al., *Small RNAs Are Trafficked from the Epididymis to Developing Mammalian Sperm.* Dev Cell, 2018. **46**(4): p. 481-494.e6.
158. Conine, C.C., et al., *Small RNAs Gained during Epididymal Transit of Sperm Are Essential for Embryonic Development in Mice.* Dev Cell, 2018. **46**(4): p. 470-480.e3.
159. Nixon, B., et al., *Profiling of epididymal small non-protein-coding RNAs.* Andrology, 2019. **7**(5): p. 669-680.
160. Boskovic, A., et al., *Control of noncoding RNA production and histone levels by a 5' tRNA fragment.* Genes Dev, 2020. **34**(1-2): p. 118-131.
161. Chen, Q., et al., *Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder.* Science, 2016. **351**(6271): p. 397-400.
162. Chan, J.C., et al., *Reproductive tract extracellular vesicles are sufficient to transmit intergenerational stress and program neurodevelopment.* Nat Commun, 2020. **11**(1): p. 1499.
163. de Castro Barbosa, T., et al., *High-fat diet reprograms the epigenome of rat spermatozoa and transgenerationally affects metabolism of the offspring.* Mol Metab, 2016. **5**(3): p. 184-197.
164. Dickson, D.A., et al., *Reduced levels of miRNAs 449 and 34 in sperm of mice and men exposed to early life stress.* Transl Psychiatry, 2018. **8**(1): p. 101.
165. Grandjean, V., et al., *RNA-mediated paternal heredity of diet-induced obesity and metabolic disorders.* Sci Rep, 2015. **5**: p. 18193.
166. Short, A.K., et al., *Exercise alters mouse sperm small noncoding RNAs and induces a transgenerational modification of male offspring conditioned fear and anxiety.* Transl Psychiatry, 2017. **7**(5): p. e1114.
167. Zhang, Y., et al., *Dnmt2 mediates intergenerational transmission of paternally acquired metabolic disorders through sperm small non-coding RNAs.* Nat Cell Biol, 2018. **20**(5): p. 535-540.
168. Short, A.K., et al., *Elevated paternal glucocorticoid exposure alters the small noncoding RNA profile in sperm and modifies anxiety and depressive phenotypes in the offspring.* Transl Psychiatry, 2016. **6**(6): p. e837.
169. McPherson, N.O., et al., *Preconception diet or exercise intervention in obese fathers normalizes sperm microRNA profile and metabolic syndrome in female offspring.* Am J Physiol Endocrinol Metab, 2015. **308**(9): p. E805-21.
170. James, E.R., et al., *The Role of the Epididymis and the Contribution of Epididymosomes to Mammalian Reproduction.* Int J Mol Sci, 2020. **21**(15).
171. Hammer, B., et al., *Preconceptional smoking alters spermatozoal miRNAs of murine fathers and affects offspring's body weight.* Int J Obes (Lond), 2021. **45**(7): p. 1623-1627.
172. Carone, B.R., et al., *Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals.* Cell, 2010. **143**(7): p. 1084-96.
173. Fullston, T., et al., *Paternal obesity initiates metabolic disturbances in two generations of mice with incomplete penetrance to the F2 generation and alters the transcriptional profile of testis and sperm microRNA content.* Faseb j, 2013. **27**(10): p. 4226-43.

174. Fullston, T., et al., *Paternal obesity induces metabolic and sperm disturbances in male offspring that are exacerbated by their exposure to an "obesogenic" diet.* Physiol Rep, 2015. **3**(3).
175. Wu, H., et al., *Preconception urinary phthalate concentrations and sperm DNA methylation profiles among men undergoing IVF treatment: a cross-sectional study.* Hum Reprod, 2017. **32**(11): p. 2159-2169.
176. Ouko, L.A., et al., *Effect of alcohol consumption on CpG methylation in the differentially methylated regions of H19 and IG-DMR in male gametes: implications for fetal alcohol spectrum disorders.* Alcohol Clin Exp Res, 2009. **33**(9): p. 1615-27.
177. Soubry, A., et al., *Human exposure to flame-retardants is associated with aberrant DNA methylation at imprinted genes in sperm.* Environ Epigenet, 2017. **3**(1): p. dvx003.
178. Greeson, K.W., et al., *Detrimental effects of flame retardant, PBB153, exposure on sperm and future generations.* Sci Rep, 2020. **10**(1): p. 8567.
179. Shnorhavorian, M., et al., *Differential DNA Methylation Regions in Adult Human Sperm following Adolescent Chemotherapy: Potential for Epigenetic Inheritance.* PLoS One, 2017. **12**(2): p. e0170085.
180. Soubry, A., et al., *Obesity-related DNA methylation at imprinted genes in human sperm: Results from the TIEGER study.* Clin Epigenetics, 2016. **8**: p. 51.
181. Denham, J., et al., *Genome-wide sperm DNA methylation changes after 3 months of exercise training in humans.* Epigenomics, 2015. **7**(5): p. 717-31.
182. Torres-Padilla, M.E., et al., *Dynamic distribution of the replacement histone variant H3.3 in the mouse oocyte and preimplantation embryos.* Int J Dev Biol, 2006. **50**(5): p. 455-61.
183. Yuan, S., et al., *Sperm-borne miRNAs and endo-siRNAs are important for fertilization and preimplantation embryonic development.* Development, 2016. **143**(4): p. 635-47.
184. Seisenberger, S., J.R. Peat, and W. Reik, *Conceptual links between DNA methylation reprogramming in the early embryo and primordial germ cells.* Current Opinion in Cell Biology, 2013. **25**(3): p. 281-288.
185. Guo, H., et al., *The DNA methylation landscape of human early embryos.* Nature, 2014. **511**(7511): p. 606-10.
186. Rodgers, A.B., et al., *Transgenerational epigenetic programming via sperm microRNA recapitulates effects of paternal stress.* Proc Natl Acad Sci U S A, 2015. **112**(44): p. 13699-704.
187. Schlatt, S. and J. Ehmcke, *Regulation of spermatogenesis: an evolutionary biologist's perspective.* Semin Cell Dev Biol, 2014. **29**: p. 2-16.
188. Gunes, S., et al., *Smoking-induced genetic and epigenetic alterations in infertile men.* Andrologia, 2018. **50**(9): p. e13124.
189. Pacifici, R., et al., *Environmental Tobacco-Smoke - Nicotine and Cotinine Concentration in Semen.* Environmental Research, 1995. **68**(1): p. 69-72.
190. Pacifici, R., et al., *Nicotine, Cotinine, and trans-3-Hydroxycotine Levels in Seminal Plasma of Smokers: Effects on Sperm Parameters.* Therapeutic Drug Monitoring, 1993. **15**(5).
191. Vine, M.F., et al., *Cotinine concentrations in semen, urine, and blood of smokers and nonsmokers.* American Journal of Public Health, 1993. **83**(9): p. 1335-1338.
192. Hammadeh, M., et al., *Protamine contents and P1/P2 ratio in human spermatozoa from smokers and non-smokers.* Human Reproduction, 2010. **25**(11): p. 2708-2720.

193. Gualtieri, R., et al., *Sperm Oxidative Stress during In Vitro Manipulation and Its Effects on Sperm Function and Embryo Development.* Antioxidants (Basel), 2021. **10**(7).
194. Wyck, S., et al., *Oxidative stress in sperm affects the epigenetic reprogramming in early embryonic development.* Epigenetics Chromatin, 2018. **11**(1): p. 60.
195. Beal, M.A., C.L. Yauk, and F. Marchetti, *From sperm to offspring: Assessing the heritable genetic consequences of paternal smoking and potential public health impacts.* Mutation Research/Reviews in Mutation Research, 2017. **773**: p. 26-50.
196. Esakky, P. and K.H. Moley, *Paternal smoking and germ cell death: A mechanistic link to the effects of cigarette smoke on spermatogenesis and possible long-term sequelae in offspring.* Molecular and cellular endocrinology, 2016. **435**: p. 85-93.
197. Jenkins, T.G., et al., *Cigarette smoking significantly alters sperm DNA methylation patterns.* Andrology, 2017. **5**(6): p. 1089-1099.
198. Hamad, M.F., et al., *The status of global DNA methylation in the spermatozoa of smokers and non-smokers.* Reprod Biomed Online, 2018. **37**(5): p. 581-589.
199. Laqqan, M., et al., *Aberrant DNA methylation patterns of human spermatozoa in current smoker males.* Reprod Toxicol, 2017. **71**: p. 126-133.
200. Fragou, D., et al., *Smoking and DNA methylation: Correlation of methylation with smoking behavior and association with diseases and fetus development following prenatal exposure.* Food Chem Toxicol, 2019. **129**: p. 312-327.
201. Song, B., et al., *Sperm DNA integrity status is associated with DNA methylation signatures of imprinted genes and non-imprinted genes.* J Assist Reprod Genet, 2021. **38**(8): p. 2041-2048.
202. Laqqan, M.M. and M.M. Yassin, *Influence of tobacco cigarette heavy smoking on DNA methylation patterns and transcription levels of MAPK8IP3, GAA, ANXA2, PRRC2A, and PDE11A genes in human spermatozoa.* Middle East Fertility Society Journal, 2021. **26**(1): p. 41.
203. Xu, W., et al., *Cigarette smoking exposure alters pebp1 DNA methylation and protein profile involved in MAPK signaling pathway in mice testis.* Biol Reprod, 2013. **89**(6): p. 142.
204. Dai, J., et al., *Nicotine elevates sperm motility and induces Pfn1 promoter hypomethylation in mouse testis.* Andrology, 2015. **3**(5): p. 967-78.
205. Yu, B., et al., *Cigarette smoking is associated with abnormal histone-to-protamine transition in human sperm.* Fertil Steril, 2014. **101**(1): p. 51-57.e1.
206. Maccani, M.A. and V.S. Knopik, *Cigarette smoke exposure-associated alterations to non-coding RNA.* Front Genet, 2012. **3**: p. 53.
207. Metzler-Guillemain, C., et al., *Sperm mRNAs and microRNAs as candidate markers for the impact of toxicants on human spermatogenesis: an application to tobacco smoking.* Syst Biol Reprod Med, 2015. **61**(3): p. 139-49.
208. Marczylo, E.L., et al., *Smoking induces differential miRNA expression in human spermatozoa: a potential transgenerational epigenetic concern?* Epigenetics, 2012. **7**(5): p. 432-9.
209. Triebener, K.P., *Respiratory health in aging women*, in *Department of Clinical Science*. 2017, University of Bergen: AiT Bjerch AS/ University of Bergen.
210. Johannessen, A., et al., *Longterm follow-up in European respiratory health studies - patterns and implications.* BMC Pulm Med, 2014. **14**: p. 63.

211. Svanes, C., et al., *Cohort profile: the multigeneration Respiratory Health in Northern Europe, Spain and Australia (RHINESSA) cohort.* BMJ Open, 2022. **12**(6): p. e059434.
212. Miller, S.A., D.D. Dykes, and H.F. Polesky, *A simple salting out procedure for extracting DNA from human nucleated cells.* Nucleic Acids Res, 1988. **16**(3): p. 1215.
213. Pidsley, R., et al., *Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling.* Genome Biol, 2016. **17**(1): p. 208.
214. Lehne, B., et al., *A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies.* Genome Biol, 2015. **16**(1): p. 37.
215. Johnson, W.E., C. Li, and A. Rabinovic, *Adjusting batch effects in microarray expression data using empirical Bayes methods.* Biostatistics, 2007. **8**(1): p. 118-27.
216. Aryee, M.J., et al., *Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays.* Bioinformatics, 2014. **30**(10): p. 1363-9.
217. Min, J.L., et al., *Meffil: efficient normalization and analysis of very large DNA methylation datasets.* Bioinformatics, 2018. **34**(23): p. 3983-3989.
218. Teschendorff, A.E., et al., *A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data.* Bioinformatics, 2013. **29**(2): p. 189-96.
219. Leek, J.T., et al., *The sva package for removing batch effects and other unwanted variation in high-throughput experiments.* Bioinformatics, 2012. **28**(6): p. 882-3.
220. Houseman, E.A., et al., *DNA methylation arrays as surrogate measures of cell mixture distribution.* BMC Bioinformatics, 2012. **13**: p. 86.
221. Zheng, S.C., et al., *EpiDISH web server: Epigenetic Dissection of Intra-Sample-Heterogeneity with online GUI.* Bioinformatics, 2019. **36**(6): p. 1950-1.
222. Arshad, S.H., et al., *Cohort Profile Update: The Isle of Wight Whole Population Birth Cohort (IOWBC).* Int J Epidemiol, 2020. **49**(4): p. 1083-1084.
223. Fraser, A., et al., *Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort.* Int J Epidemiol, 2013. **42**(1): p. 97-110.
224. Boyd, A., et al., *Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children.* Int J Epidemiol, 2013. **42**(1): p. 111-27.
225. Pedersen, B.S., et al., *Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values.* Bioinformatics, 2012. **28**(22): p. 2986-8.
226. Benjamini, Y. and Y. Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiple testing.* Journal of the Royal statistical society: series B (Methodological), 1995. **57**(1): p. 289-300.
227. van Iterson, M., E.W. van Zwet, and B.T. Heijmans, *Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution.* Genome Biol, 2017. **18**(1): p. 19.
228. Kuleshov, M.V., et al., *Enrichr: a comprehensive gene set enrichment analysis web server 2016 update.* Nucleic Acids Res, 2016. **44**(W1): p. W90-7.
229. *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.
230. Bernstein, B.E., et al., *The NIH Roadmap Epigenomics Mapping Consortium.* Nat Biotechnol, 2010. **28**(10): p. 1045-8.

231. Kanehisa, M., et al., *KEGG: new perspectives on genomes, pathways, diseases and drugs.* Nucleic Acids Res, 2017. **45**(D1): p. D353-d361.
232. *Gene Ontology Consortium: going forward.* Nucleic Acids Res, 2015. **43**(Database issue): p. D1049-56.
233. Phipson, B., J. Maksimovic, and A. Oshlack, *missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform.* Bioinformatics, 2016. **32**(2): p. 286-8.
234. Long, J.A., *Analysis and Presentation of Social Scientific Data.* 2019.
235. Steen, J., et al., *medflex: An R Package for Flexible Mediation Analysis using Natural Effect Models.* Journal of Statistical Software, 2017. **76**(11): p. 1 - 46.
236. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Res, 2015. **43**(7): p. e47.
237. Yin, L., et al., *rMVP: A Memory-efficient, Visualization-enhanced, and Parallel-accelerated Tool for Genome-wide Association Study.* Genomics, Proteomics & Bioinformatics, 2021. **19**(4): p. 619-628.
238. Suderman, M., et al., *dmrff: identifying differentially methylated regions efficiently with power and control.* bioRxiv, 2018: p. 508556.
239. Peters, T.J., et al., *De novo identification of differentially methylated regions in the human genome.* Epigenetics Chromatin, 2015. **8**: p. 6.
240. Breeze, C.E., et al., *eFORGE v2.0: updated analysis of cell type-specific signal in epigenomic data.* Bioinformatics, 2019. **35**(22): p. 4767-4769.
241. Li, M., et al., *EWAS Atlas: a curated knowledgebase of epigenome-wide association studies.* Nucleic Acids Res, 2019. **47**(D1): p. D983-d988.
242. Carvalho-Silva, D., et al., *Open Targets Platform: new developments and updates two years on.* Nucleic acids research, 2019. **47**(D1): p. D1056-D1065.
243. Szklarczyk, D., et al., *STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets.* Nucleic Acids Res, 2019. **47**(D1): p. D607-d613.
244. Soudy, M., et al., *UniprotR: Retrieving and visualizing protein sequence and functional information from Universal Protein Resource (UniProt knowledgebase).* J Proteomics, 2020. **213**: p. 103613.
245. Maksimovic, J., A. Oshlack, and B. Phipson, *Gene set enrichment analysis for genome-wide DNA methylation data.* Genome Biol, 2021. **22**(1): p. 173.
246. Joehanes, R., et al., *Epigenetic Signatures of Cigarette Smoking.* Circ Cardiovasc Genet, 2016. **9**(5): p. 436-447.
247. Christiansen, C., et al., *Novel DNA methylation signatures of tobacco smoking with trans-ethnic effects.* Clin Epigenetics, 2021. **13**(1): p. 36.
248. Joubert, B.R., et al., *DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis.* Am J Hum Genet, 2016. **98**(4): p. 680-96.
249. *Overview of ethics committees and approval numbers of RHINESSA centers. Available online: https://helse-bergen.no/seksjon/RHINESSA/Documents/Ethic%20Committees%20list.pdf. (Accessed on 25th of May 2022).*
250. *World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects.* Jama, 2013. **310**(20): p. 2191-4.

251. *Haukeland University Hospital regulations for research. Available online: https://helse-bergen.no/fag-og-forsking/forsking/forskingsrutinar#malar (Accessed on 25th of May 2022).*
252. *Norwegian Code of conduct for information security in the healthcare and care servises sector. Available online: https://www.ehelse.no/normen/documents-in-english (Accessed on 25th of May 2022).*
253. Kirkwood B, *Essentials of Medical Statistics. Blackwell Science, 2003.*
254. Hennekens CH, B.J., *Epidemiology in Medicine, Lippincott Williams & Wilkins, 1987.*
255. *Errors in epidemiological measurements. Available online: https://healthknowledge.org.uk (Accessed 9th June, 2022).*
256. *STAT 509: Design and Analysis of Clinical Trials. Penn State's Department of Statistics. Available online: https://online.stat.psu.edu (Accessed on 9th of June 2022).*
257. Last J, *(ed): A Dictionary of Epidemiology, ed 3. Oxford, Oxford University Press, 1988.*
258. Tripepi, G., et al., *Selection Bias and Information Bias in Clinical Research.* Nephron Clinical Practice, 2010. **115**(2): p. c94-c99.
259. Fowler, F., *Survey Research Methods (4th ed.)*. 2009, SAGE Publications, Inc.: Thousand Oaks, California.
260. Spitzer, S., *Biases in health expectancies due to educational differences in survey participation of older Europeans: It's worth weighting for.* The European Journal of Health Economics, 2020. **21**(4): p. 573-605.
261. Smith, W., *Does Gender Influence Online Survey Participation? A Record-Linkage Analysis of University Faculty Online Survey Response Behavior.* Online Submission, 2008.
262. Liber, A.C. and K.E. Warner, *Has Underreporting of Cigarette Consumption Changed Over Time? Estimates Derived From US National Health Surveillance Systems Between 1965 and 2015.* Am J Epidemiol, 2018. **187**(1): p. 113-119.
263. Skelly, A.C., J.R. Dettori, and E.D. Brodt, *Assessing bias: the importance of considering confounding.* Evidence-based spine-care journal, 2012. **3**(1): p. 9-12.
264. Rothman KJ, *Epidemiology-An Introduction. New York: Oxford University Press; 2002.*
265. *Advantages & Disadvantages of Cohort Studies. Available online: https://sphweb.bumc.bu.edu (Accessed on 13th of June, 2022).*
266. Janson, C., et al., *The European Community Respiratory Health Survey: what are the main results so far? European Community Respiratory Health Survey II.* Eur Respir J, 2001. **18**(3): p. 598-611.
267. Kuiper, I.N., *Effects of air pollution and greenness on asthma and allergy-over time and across generations*, in *Department of Occupational Medicine at Haukeland University Hospital; Department of Global Public Health and Primary Care, University of Bergen* 2020, University of Bergen: Skipnes Kommunikasjon/University of Bergen. p. 1-78.
268. Nilsen, R.M., et al., *Self-selection and bias in a large prospective pregnancy cohort in Norway.* Paediatric and perinatal epidemiology, 2009. **23**(6): p. 597-608.
269. Mill, J. and B.T. Heijmans, *From promises to practical strategies in epigenetic epidemiology.* Nat Rev Genet, 2013. **14**(8): p. 585-94.
270. Moskvina, V. and K.M. Schmidt, *On multiple-testing correction in genome-wide association studies.* Genet Epidemiol, 2008. **32**(6): p. 567-73.

271. Mansell, G., et al., *Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array.* BMC Genomics, 2019. **20**(1): p. 366.
272. Ntani, G., et al., *Consequences of ignoring clustering in linear regression.* BMC Med Res Methodol, 2021. **21**(1): p. 139.
273. Pape, K., et al., *Agreement of offspring-reported parental smoking status: the RHINESSA generation study.* BMC Public Health, 2019. **19**(1): p. 94.
274. Devaux, M. and F. Sassi, *Social disparities in hazardous alcohol use: self-report bias may lead to incorrect estimates.* Eur J Public Health, 2016. **26**(1): p. 129-34.
275. Moran, S., C. Arribas, and M. Esteller, *Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences.* Epigenomics, 2016. **8**(3): p. 389-99.
276. Carvan, M.J., 3rd, et al., *Mercury-induced epigenetic transgenerational inheritance of abnormal neurobehavior is correlated with sperm epimutations in zebrafish.* PLoS One, 2017. **12**(5): p. e0176155.
277. Bibikova, M., *Chapter 2 - DNA Methylation Microarrays*, in *Epigenomics in Health and Disease*, M.F. Fraga and A.F. Fernández, Editors. 2016, Academic Press: Boston. p. 19-46.
278. Sugden, K., et al., *Patterns of Reliability: Assessing the Reproducibility and Integrity of DNA Methylation Measurement.* Patterns, 2020. **1**(2): p. 100014.
279. Tehard, B., et al., *Anthropometric measurements and body silhouette of women: validity and perception.* Journal of the American Dietetic Association, 2002. **102**(12): p. 1779-1784.
280. Connor Gorber, S., et al., *A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review.* Obes Rev, 2007. **8**(4): p. 307-26.
281. Roberts, R.J., *Can self-reported data accurately describe the prevalence of overweight?* Public Health, 1995. **109**(4): p. 275-84.
282. Romero-Corral, A., et al., *Accuracy of body mass index in diagnosing obesity in the adult general population.* International journal of obesity (2005), 2008. **32**(6): p. 959-966.
283. Rothman, K.J., *BMI-related errors in the measurement of obesity.* Int J Obes (Lond), 2008. **32 Suppl 3**: p. S56-9.
284. Houtkooper, L.B., et al., *Why bioelectrical impedance analysis should be used for estimating adiposity.* The American Journal of Clinical Nutrition, 1996. **64**(3): p. 436S-448S.
285. Böhm, A. and B.L. Heitmann, *The use of bioelectrical impedance analysis for body composition in epidemiological studies.* European Journal of Clinical Nutrition, 2013. **67**(1): p. S79-S85.
286. Roubenoff, R., *Applications of bioelectrical impedance analysis for body composition to epidemiologic studies.* The American Journal of Clinical Nutrition, 1996. **64**(3): p. 459S-462S.
287. Keyes, K.M., et al., *What is a cohort effect? Comparison of three statistical methods for modeling cohort effects in obesity prevalence in the United States, 1971-2006.* Soc Sci Med, 2010. **70**(7): p. 1100-8.
288. Birney, E., G.D. Smith, and J.M. Greally, *Epigenome-wide Association Studies and the Interpretation of Disease -Omics.* PLOS Genetics, 2016. **12**(6): p. e1006105.

289. Houseman, E.A., et al., *DNA Methylation in Whole Blood: Uses and Challenges.* Current Environmental Health Reports, 2015. **2**(2): p. 145-154.
290. Birney, E., G.D. Smith, and J.M. Greally, *Epigenome-wide Association Studies and the Interpretation of Disease -Omics.* PLoS Genet, 2016. **12**(6): p. e1006105.
291. Breton, C.V., et al., *Small-Magnitude Effect Sizes in Epigenetic End Points are Important in Children's Environmental Health Studies: The Children's Environmental Health and Disease Prevention Research Center's Epigenetics Working Group.* Environ Health Perspect, 2017. **125**(4): p. 511-526.
292. Chooi, Y.C., C. Ding, and F. Magkos, *The epidemiology of obesity.* Metabolism, 2019. **92**: p. 6-10.
293. Han, T.S., et al., *Contributions of maternal and paternal adiposity and smoking to adult offspring adiposity and cardiovascular risk: the Midspan Family Study.* BMJ Open, 2015. **5**(11): p. e007682.
294. Gapp, K., et al., *Alterations in sperm long RNA contribute to the epigenetic inheritance of the effects of postnatal trauma.* Mol Psychiatry, 2020. **25**(9): p. 2162-2174.
295. Forastiere, L., A. Mattei, and P. Ding, *Principal ignorability in mediation analysis: through and beyond sequential ignorability.* Biometrika, 2018. **105**(4): p. 979-986.
296. Sterne, J.A., et al., *Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls.* Bmj, 2009. **338**: p. b2393.
297. White, I.R., P. Royston, and A.M. Wood, *Multiple imputation using chained equations: Issues and guidance for practice.* Stat Med, 2011. **30**(4): p. 377-99.
298. Qiu, W., et al., *The impact of genetic variation and cigarette smoke on DNA methylation in current and former smokers from the COPDGene study.* Epigenetics, 2015. **10**(11): p. 1064-1073.
299. Sikdar, S., et al., *Comparison of smoking-related DNA methylation between newborns from prenatal exposure and adults from personal smoking.* Epigenomics, 2019. **11**(13): p. 1487-1500.
300. Hannon, E., et al., *Assessing the co-variability of DNA methylation across peripheral cells and tissues: Implications for the interpretation of findings in epigenetic epidemiology.* PLoS genetics, 2021. **17**(3): p. e1009443-e1009443.
301. Dugué, P.A., et al., *Smoking and blood DNA methylation: an epigenome-wide association study and assessment of reversibility.* Epigenetics, 2020. **15**(4): p. 358-368.
302. SCR_002773), G.R., *Summary statistics on entrez gene: ATPase H+Transporting V1 Subunit E1 were downloaded from GeneCards (RRID: SCR_002773). Accessed on 23/08/2022*.
303. SCR_002773), G.R., *Summary statistics on entrez gene: Complement 2 were downloaded from GeneCards (RRID: SCR_002773). Accessed on 23/08/2022*.
304. SCR_002773), G.R., *Summary statistics on entrez gene: Catenin Alpha 2 were downloaded from GeneCards (RRID: SCR_002773). Accessed on 23/08/2022*.
305. SCR_002773), G.R., *Summary statistics on entrez gene: Acyl_CoA synthesase Family member 3 were downloaded from GeneCards(RRID:SCR_002773). Accessed on 23/08/2022*.
306. SCR_002773), G.R., *Summary statistics on entrez gene: WD Repeat_Containing Protein 60 were downloaded from GeneCards (RRID: SCR_002773). Accessed on 23/08/2022*.

307. Baghdadi, M., et al., *MFG-E8 regulates the immunogenic potential of dendritic cells primed with necrotic cell-mediated inflammatory signals.* PLoS One, 2012. **7**(6): p. e39607.
308. SCR_002773), G.R., *Summary statistics on entrez gene: Unc-93 Homolog B1, TLR Signaling Regulator were downloaded from GeneCards (RRID: SCR_002773). Accessed on 23/08/2022*.
309. SCR_002773), G.R., *Summary statistics on entrez gene: Brain Enriched Myelin Associated Protein 1 were downloaded from GeneCards (RRID: SCR_002773). Accessed on 23/08/2022*.
310. Pelka, K., et al., *The Chaperone UNC93B1 Regulates Toll-like Receptor Stability Independently of Endosomal TLR Transport.* Immunity, 2018. **48**(5): p. 911-922.e7.
311. Mantovani, A. and F. Balkwill, *RalB signaling: a bridge between inflammation and cancer.* Cell, 2006. **127**(1): p. 42-4.
312. Khalifeh-Soltani, A., et al., *Mfge8 regulates enterocyte lipid storage by promoting enterocyte triglyceride hydrolase activity.* JCI Insight, 2016. **1**(18): p. e87418.
313. SCR_002773), G.R., *Summary statistics on entrez gene: DLG Associated Protein 1 were downloaded from GeneCards (RRID: SCR_002773). Accessed on 23/08/2022*.
314. Murphy, S.K., et al., *Gender-specific methylation differences in relation to prenatal exposure to cigarette smoke.* Gene, 2012. **494**(1): p. 36-43.
315. Joubert, B.R., et al., *Maternal smoking and DNA methylation in newborns: in utero effect or epigenetic inheritance?* Cancer Epidemiol Biomarkers Prev, 2014. **23**(6): p. 1007-17.
316. Wei, Y., H. Schatten, and Q.Y. Sun, *Environmental epigenetic inheritance through gametes and implications for human reproduction.* Hum Reprod Update, 2015. **21**(2): p. 194-208.
317. Zeilinger, S., et al., *Tobacco smoking leads to extensive genome-wide changes in DNA methylation.* PLoS One, 2013. **8**(5): p. e63812.
318. SCR_002773), G.R., *Summary statistics on entrez gene: Toll Like Receptor 9 were downloaded from GeneCards (RRID: SCR_002773). Accessed on 30/08/2022*.
319. Haeryfar, S.M., et al., *Terminal deoxynucleotidyl transferase establishes and broadens antiviral CD8+ T cell immunodominance hierarchies.* J Immunol, 2008. **181**(1): p. 649-59.
320. Chitu, V., et al., *Primed innate immunity leads to autoinflammatory disease in PSTPIP2-deficient cmo mice.* Blood, 2009. **114**(12): p. 2497-2505.
321. SCR_002773), G.R., *Summary statistics on entrez gene: Colony Stimulating Factor 1 Receptor were downloaded from GeneCards (RRID: SCR_002773). Accessed on 29/08/2022*.
322. SCR_002773), G.R., *Summary statistics on entrez gene: GRAM Domain Containing 4 were downloaded from GeneCards (RRID: SCR_002773). Accessed on 07/10/2022*.
323. SCR_002773), G.R., *Summary statistics on entrez gene: Tripartite Motif Containing 2 were downloaded from GeneCards (RRID: SCR_002773). Accessed on 07/10/2022*.
324. Caruso, M., et al., *Increased interaction with insulin receptor substrate 1, a novel abnormality in insulin resistance and type 2 diabetes.* Diabetes, 2014. **63**(6): p. 1933-47.
325. Kubota, N., et al., *Differential hepatic distribution of insulin receptor substrates causes selective insulin resistance in diabetes and obesity.* Nat Commun, 2016. **7**: p. 12977.

326. Murakami, Y., et al., *TLR9-IL-2 axis exacerbates allergic asthma by preventing IL-17A hyperproduction.* Sci Rep, 2020. **10**(1): p. 18110.
327. Shin, E.K., et al., *Association between colony-stimulating factor 1 receptor gene polymorphisms and asthma risk.* Human genetics, 2010. **128**(3): p. 293-302.
328. Modena, B.D., et al., *Gene Expression Correlated with Severe Asthma Characteristics Reveals Heterogeneous Mechanisms of Severe Disease.* American journal of respiratory and critical care medicine, 2017. **195**(11): p. 1449-1463.
329. NHGRI_EBI GWAS Catalog (Buniello, M.e.a., 2019), *Summary statistics were downloaded from the NHGRI-EBI GWAS Catalog (Buniello, MacArthur et al., 2019) for study GCST007080 (Kichaev et al., 2018) accessed on 05/10/2022.*
330. Rzehak, P., et al., *DNA-Methylation and Body Composition in Preschool Children: Epigenome-Wide-Analysis in the European Childhood Obesity Project (CHOP)-Study.* Scientific reports, 2017. **7**(1): p. 14349-14349.
331. SCR_002773), G.R., *Summary statistics on entrez gene: MAP3K12 Binding Inhibitory Protein 1 were downloaded from GeneCards (RRID: SCR_002773). Accessed on 30/08/2022*.
332. SCR_002773), G.R., *Summary statistics on entrez gene: Neurotrophic Receptor Tyrosine Kinase 2 were downloaded from GeneCards (RRID: SCR_002773). Accessed on 30/08/2022*.
333. SCR_002773), G.R., *Summary statistics on entrez gene: Two Pore Segment Channel 1 were downloaded from GeneCards (RRID: SCR_002773). Accessed on 30/08/2022*.
334. NHGRI_EBI GWAS Catalog (Buniello, M.e.a., 2019), *Summary statistics were downloaded from the NHGRI-EBI GWAS Catalog (Buniello, MacArthur et al., 2019) for study GCST012227 (Christakoudi et al., 2021) accessed on 05/10/2022.*
335. NHGRI_EBI GWAS Catalog (Buniello, M.e.a., 2019), *Summary statistics were downloaded from the NHGRI-EBI GWAS Catalog (Buniello, MacArthur et al., 2019) for study GCST012230 (Christakoudi et al., 2021) accessed on 05/10/2022.*
336. Liu, Y., et al., *Effects of paternal exposure to cigarette smoke on sperm DNA methylation and long-term metabolic syndrome in offspring.* Epigenetics & Chromatin, 2022. **15**(1): p. 3.
337. Golding, J., et al., *Investigating Possible Trans/Intergenerational Associations With Obesity in Young Adults Using an Exposome Approach.* Frontiers in Genetics, 2019. **10**.
338. Katzmarski, N., et al., *Transmission of trained immunity and heterologous resistance to infections across generations.* Nat Immunol, 2021. **22**(11): p. 1382-1390.
339. Kazachenka, A., et al., *Identification, Characterization, and Heritability of Murine Metastable Epialleles: Implications for Non-genetic Inheritance.* Cell, 2018. **175**(5): p. 1259-1271.e13.
340. Oey, H., et al., *Genetic and epigenetic variation among inbred mouse littermates: identification of inter-individual differentially methylated regions.* Epigenetics Chromatin, 2015. **8**: p. 54.
341. Northstone, K., et al., *The Avon Longitudinal Study of Parents and Children (ALSPAC): an update on the enrolled sample of index children in 2019.* Wellcome Open Res, 2019. **4**: p. 51.
342. Mejia-Lancheros, C., et al., *Smoking habit from the paternal line and grand-child's overweight or obesity status in early childhood: prospective findings from the*

lifeways cross-generation cohort study. International Journal of Obesity, 2018. **42**(11): p. 1853-1870.
343. You, Y., et al., *Effect of Exposure to Paternal Smoking on Overweight and Obesity in Children: Findings from the Children Lifeway Cohort in Shenzhen, Southern China.* Obesity Facts, 2022. **15**(4): p. 609-620.
344. Yen, A.M.-F., et al., *Longer Duration and Earlier Age of Onset of Paternal Betel Chewing and Smoking Increase Metabolic Syndrome Risk in Human Offspring, Independently, in a Community-Based Screening Program in Taiwan.* Circulation, 2016. **134**(5): p. 392-404.
345. Wilson, R., et al., *The dynamics of smoking-related disturbed methylation: a two time-point study of methylation change in smokers, non-smokers and former smokers.* BMC genomics, 2017. **18**(1): p. 805-805.
346. Timmermans, S.H., et al., *Maternal smoking during pregnancy and childhood overweight and fat distribution: the KOALA Birth Cohort Study.* Pediatric Obesity, 2014. **9**(1): p. e14-e25.
347. Toschke, A., et al., *Childhood obesity is associated with maternal smoking in pregnancy.* European Journal of Pediatrics, 2002. **161**(8): p. 445-448.
348. Mizutani, T., et al., *Association of Maternal Lifestyles Including Smoking During Pregnancy with Childhood Obesity.* Obesity, 2007. **15**(12): p. 3133-3139.
349. Schneider, S., et al., *Smoking cessation during pregnancy: a systematic literature review.* Drug Alcohol Rev, 2010. **29**(1): p. 81-90.
350. Lisboa, P.C., E. de Oliveira, and E.G. de Moura, *Obesity and endocrine dysfunction programmed by maternal smoking in pregnancy and lactation.* Front Physiol, 2012. **3**: p. 437.
351. Oken, E., E.B. Levitan, and M.W. Gillman, *Maternal smoking during pregnancy and child overweight: systematic review and meta-analysis.* Int J Obes (Lond), 2008. **32**(2): p. 201-10.
352. Godfrey, K.M. and D.J. Barker, *Fetal programming and adult health.* Public Health Nutr, 2001. **4**(2b): p. 611-24.
353. Suzuki, K., et al., *Gender differences in the association between maternal smoking during pregnancy and childhood growth trajectories: multilevel analysis.* International Journal of Obesity, 2011. **35**(1): p. 53-59.
354. Riedel, C., et al., *Differences in BMI z-Scores between Offspring of Smoking and Nonsmoking Mothers: A Longitudinal Study of German Children from Birth through 14 Years of Age.* Environmental Health Perspectives, 2014. **122**(7): p. 761-767.
355. Power, C. and B.J.M.H. Jefferis, *Fetal environment and subsequent obesity: a study of maternal smoking.* International Journal of Epidemiology, 2002. **31**(2): p. 413-419.
356. Mattsson, K., et al., *Maternal smoking during pregnancy and daughters' risk of gestational diabetes and obesity.* Diabetologia, 2013. **56**(8): p. 1689-1695.
357. Maessen, S.E., et al., *Maternal smoking early in pregnancy is associated with increased risk of short stature and obesity in adult daughters.* Scientific reports, 2019. **9**(1): p. 4290-4290.
358. Nakamura, A., O. François, and J. Lepeule, *Epigenetic Alterations of Maternal Tobacco Smoking during Pregnancy: A Narrative Review.* International journal of environmental research and public health, 2021. **18**(10): p. 5083.
359. Witt, S.H., et al., *Impact on birth weight of maternal smoking throughout pregnancy mediated by DNA methylation.* BMC Genomics, 2018. **19**(1): p. 290.

360. Thaker, V.V., *GENETIC AND EPIGENETIC CAUSES OF OBESITY.* Adolescent medicine: state of the art reviews, 2017. **28**(2): p. 379-405.
361. Ambrosini, G.L., *Childhood dietary patterns and later obesity: a review of the evidence.* Proc Nutr Soc, 2014. **73**(1): p. 137-46.
362. Brug, J., et al., *Differences in weight status and energy-balance related behaviors among schoolchildren across Europe: the ENERGY-project.* PLoS One, 2012. **7**(4): p. e34742.
363. Zalbahar, N., et al., *Parental pre-pregnancy BMI influences on offspring BMI and waist circumference at 21 years.* Aust N Z J Public Health, 2016. **40**(6): p. 572-578.
364. Lane, M., et al., *Oxidative stress in mouse sperm impairs embryo development, fetal growth and alters adiposity and glucose regulation in female offspring.* PLoS One, 2014. **9**(7): p. e100832.
365. Ng, S.F., et al., *Chronic high-fat diet in fathers programs β-cell dysfunction in female rat offspring.* Nature, 2010. **467**(7318): p. 963-6.
366. Ng, S.F., et al., *Paternal high-fat diet consumption induces common changes in the transcriptomes of retroperitoneal adipose and pancreatic islet tissues in female rat offspring.* Faseb j, 2014. **28**(4): p. 1830-41.
367. Huypens, P., et al., *Epigenetic germline inheritance of diet-induced obesity and insulin resistance.* Nat Genet, 2016. **48**(5): p. 497-9.
368. Ozhan, H., et al., *Performance of bioelectrical impedance analysis in the diagnosis of metabolic syndrome.* J Investig Med, 2012. **60**(3): p. 587-91.
369. Golding, J., et al., *Human transgenerational observations of regular smoking before puberty on fat mass in grandchildren and great-grandchildren.* Scientific Reports, 2022. **12**(1): p. 1139.
370. Green, M.J., et al., *Socioeconomic position and early adolescent smoking development: evidence from the British Youth Panel Survey (1994-2008).* Tob Control, 2016. **25**(2): p. 203-10.

9. Appendices

Appendix 1. Questionnaire from the ECRHS study

Appendix 2. Questionnaire from the RHINESSA study

Appendix 3. Questionnaire from the RHINE study

Appendix 1. Questionnaire from the ECRHS study

THE EUROPEAN COMMUNITY RESPIRATORY HEALTH SURVEY II

★

ECRHS II

MAIN QUESTIONNAIRE

Project Leaders:
Prof Peter Burney
Dr Deborah Jarvis

For further information:

www.ecrhs.org

ECRHS II was funded by the European Commission as part of their Quality of Life Programme

Centre number
Personal number
Sample
Date
DAY MONTH YEAR

I AM GOING TO ASK YOU SOME QUESTIONS. AT FIRST THESE WILL BE MOSTLY ABOUT YOUR BREATHING. WHEREVER POSSIBLE, I WOULD LIKE YOU TO ANSWER 'YES' OR 'NO'.

1. Have you had wheezing or whistling in your chest at any time in the last ***12 months***? NO YES

IF 'NO' GO TO QUESTION 2, IF 'YES':

NO YES

1.1 Have you been at all breathless when the wheezing noise was present?

1.2. Have you had this wheezing or whistling when you did ***not*** have a cold? NO YES

2. Have you woken up with a feeling of tightness in your chest at any time in the last ***12 months***? NO YES

3. Have you had an attack of shortness of breath that came on during the day when you were at rest at any time in the last ***12 months***? NO YES

4. Have you had an attack of shortness of breath that came on ***following*** strenuous activity at any time in the last ***12 months***? NO YES

5. Have you been woken by an attack of shortness of breath at any time in the last ***12 months?*** NO YES

IF NO GO TO Q6, IF YES

5.1 Have you been woken by an attack of shortness of breath in the last ***3 months***? NO YES

IF NO GO TO Q6, IF YES

5.1.1 ***On average*** have you been woken by an attack of shortness of breath ***at least once a week*** *in the* ***last 3 months?*** NO YES

IF NO GO TO Q6, IF YES

5.1.1.1 How many times a week ***on average*** have you been woken by shortness of breath in the ***last 3 months***? TIMES

6. Have you been woken by an attack of coughing at any time *in the last* ***12 months***? NO YES

NO YES

7. Do you ***usually*** cough first thing in the morning in the winter?
[IF DOUBTFUL, USE QUESTION 8.1 TO CONFIRM]

NO YES

8. Do you ***usually*** cough during the day, or at night, in the winter?

IF 'NO' GO TO QUESTION 9, IF 'YES':

8.1 Do you cough like this on most days for as much as three months each year? NO YES

9. Do you ***usually*** bring up any phlegm from your chest first thing in the morning in the winter? NO YES

[IF DOUBTFUL, USE QUESTION 10.1 TO CONFIRM]

10. Do you ***usually*** bring up any phlegm from your chest during the day, or at night, in the winter? NO YES

IF 'NO' GO TO QUESTION 11, IF 'YES':

10.1 Do you bring up phlegm like this on most days for as much as three months each year? NO YES

NO YES

11. Do you ever have trouble with your breathing?

IF 'NO' GO TO QUESTION 12, IF 'YES':

11.1 Do you have this trouble — TICK ONE BOX ONLY

a) continuously so that your breathing is never quite right? 1 ☐

b) repeatedly, but it always gets completely better? 2 ☐

c) only rarely? 3 ☐

12. Are you disabled from walking by a condition ***other than*** heart or lung disease? NO YES

IF 'YES' STATE CONDITION _____ AND GO TO QUESTION 13, IF 'NO':

12.1 Are you troubled by shortness of breath when hurrying on level ground or walking up a slight hill? NO YES

IF 'NO' GO TO QUESTION 13, IF 'YES':

12.1.1 Do you get short of breath walking with other people of your own age on level ground? NO YES

IF 'NO' GO TO QUESTION 13, IF 'YES':

12.1.1.1 Do you have to stop for breath when walking at your own pace on level ground? NO YES

13. ***FOR WOMEN ONLY - MEN GO TO Q14***

Have you ever noticed that you had respiratory symptoms (such as wheeze, tightness in your chest or shortness of breath) at a particular time of your monthly cycle? — TICK ONE BOX ONLY

yes, in the week before my period	1 ☐
yes, during my period	2 ☐
yes, in the week after my period	3 ☐
yes, another time of the month	4 ☐
does not apply to me (i.e., amenorrhoeal)	5 ☐
No	6 ☐

NO YES

14. Have you ever had asthma?
IF 'NO' GO TO QUESTION 15, IF 'YES':

NO YES

14.1 Was this confirmed by a doctor?

YEARS

14.2 How old were you when you had your first attack of asthma?

YEARS

14.3 How old were you when you had your most recent attack of asthma?

14.4.1-6 Which months of the year do you usually have attacks of asthma?

	NO	YES
14.4.1 January / February	☐	☐
14.4.2 March / April	☐	☐
14.4.3 May / June	☐	☐
14.4.4 July / August	☐	☐
14.4.5 September / October	☐	☐
14.4.6 November / December	☐	☐

NO YES

14.5 Have you had an attack of asthma in the last ***12 months***?
IF NO GO TO 14.8, IF YES

ATTACKS

14.6 How many attacks of asthma have you had in the last ***12 months***?

ATTACKS

14.7 How many attacks of asthma have you had in the last ***3 months***?

14.8 How many times have you woken up because of your asthma in the last ***3 months?***

TICK ONE BOX ONLY

every night or almost every night	1	☐
more than once a week, but not most nights	2	☐
at least twice a month, but not more than once a week	3	☐
less than twice a month	4	☐
not at all	5	☐

14.9. How often have you had trouble with your breathing because of your asthma in the last ***3 months?***

TICK ONE BOX ONLY

continuously	1	☐
about once a day	2	☐
at least once a week, but less than once a day	3	☐
less than once a week	4	☐
not at all	6	☐

NO YES

14.10 Are you currently taking any medicines including inhalers, aerosols or tablets for asthma?

NO YES

14.11 Do you have a peak flow meter of your own?
IF 'NO' GO TO QUESTION 14.12 , IF 'YES':

14.11.1 How often have you used it over the last 3 months? TICK ONE BOX ONLY

never	1	☐
some of the days	2	☐
most of the days	3	☐

14.12 Do you have written instructions from your doctor on how to manage your asthma if it gets worse or if you have an attack? NO YES

14.13. ***FOR WOMEN ONLY - MEN GO TO Q15***

Have you ever noticed that your asthma got worse with your monthly cycle? TICK ONE BOX ONLY

Yes, in the week before my period	1	☐
Yes, during my period	2	☐
Yes, in the week after my period	3	☐
Yes, another time of the month	4	☐
Does not apply to me (i.e., amenorrhoeal)	5	☐
No	6	☐

NO YES

14.14 Have you been pregnant (at least 25 weeks) since your asthma started?

IF NO GO TO Q15, IF YES

14.14.1. What happened to your asthma during your pregnancies?

TICK ONE BOX ONLY

got better	1	☐
got worse	2	☐
stayed the same	3	☐
not the same for all pregnancies	4	☐
don't know	5	☐

NO YES

15. Do you have any nasal allergies, including hay fever?

IF NO GO TO Q16, IF YES

15.1 How old were you when you first had hay fever or nasal allergy? YEARS

16. Have you ever had a problem with sneezing, or a runny or a blocked nose when you did not have a cold or the flu? NO YES

IF NO GO TO Q17, IF YES

16.1. Have you had a problem with sneezing or a runny or blocked nose when you did not have a cold or the flu ***in the last 12 months***? NO YES

IF NO GO TO Q17, IF YES

16.1.1. Has this nose problem been accompanied by itchy or watery eyes? NO YES

16.1.2. In which months of the year did this nose problem occur? NO YES

January	☐	☐
February	☐	☐
March	☐	☐
April	☐	☐
May	☐	☐
June	☐	☐
July	☐	☐

August
September
October
November
December

17. ***Since the last survey*** have you used any medication to treat nasal disorders? NO YES

IF NO GO TO Q18, IF YES

17.1 Have you used any of the following nasal sprays for the treatment of your nasal disorder? NO YES

{SHOW LIST OF STEROID NASAL SPRAYS}

IF NO GO TO Q17.2, IF YES

17.1.1 How many years have you been taking this sort of nasal spray? YEARS

17.1.2 Have you used any of these nasal sprays ***in the last 12 months?*** NO YES

17.2 Have you used any of the following pills, capsules, or tablets for the treatment of your nasal disorder? NO YES

{SHOW LIST OF ANTIHISTAMINES}

IF NO GO TO Q18, IF YES

17.2.1 How many years have you been taking these sort of pills, capsules or tablets? YEARS

17.2.2 Have you used any of these pills, capsules or tablets ***in the last 12 months?*** NO YES

NO YES

18. Have you ***ever*** had eczema or any kind of skin allergy?

19. Have you ***ever*** had an itchy rash that was coming and going for at ***least 6 months***? NO YES

IF 'NO' GO TO QUESTION 20, IF 'YES': NO YES

19.1. . Have you had this itchy rash ***in the last 12 months?***

IF 'NO' GO TO QUESTION 20, IF 'YES':

19.1.1. Has this itchy rash ***at any time*** affected any of the following places: the folds of the elbows, behind the knees, in front of the ankles under the buttocks or around the neck, ears or eyes NO YES

NO YES

20. Have you ever had any difficulty with your breathing after taking medicines?

IF 'NO' GO TO QUESTION 21, IF 'YES':

20.1-2 Which medicines? _____ 20.1.

20.1.2

YEARS

21. How old was your mother when you were born?

22. How many times did you move house during the first five

years of your life? TICK ONE BOX ONLY

None	1 ☐
Once	2 ☐
more than once	3 ☐

NO YES

23. Were you hospitalised before the age of two years for lung disease?

YEARS

24. At what age did you first attend a school, play school, day care or nursery?

25. How many ***other*** children regularly slept in your bedroom before ***you were five years old?*** CHILDREN

I would now like to ask you some questions on the type of jobs that you have done.

I am interested in each one of the jobs that you have done for more than 3 consecutive months since the time we last contacted you (in 1991/2). These jobs may be outside the house or at home, full time or part time, paid or not paid, including self employment, for example in a family business. Please include part time jobs only if you had been doing them for more than 8 hours per week.

Q26. Are you currently

TICK ONE BOX ONLY

Employed (including military service)	1 ☐
Self employed	2 ☐
Unemployed, looking for work	3 ☐
Not working because of poor health	4 ☐
Full-time house-person	5 ☐
Full time student	6 ☐
Retired	7 ☐
Other	8 ☐

IF EMPLOYED OR SELF EMPLOYED OR A FULL TIME HOURSEPERSON GO TO Q28

27. Have you been employed in any job for three continuous months or longer since the last survey? NO ☐ YES ☐

IF YES NOW GO TO OCCUPATIONAL MATRIX

Centre number
Personal number
Sample
Date
DAY MONTH YEAR

Q 28. If you had more than one job in the same company, or if you were doing more than one job at the same time, we would like to talk about them separately. Please start with your current or last job.

JOB	Q28.1. What is (was) the title of your current (last) job? *OCCUPATION*	Q28.2. What did the firm, company or organisation do or what services did it provide? INDUSTRY	Q28.3. In what month and year did you start working in this job? MONTH YEAR	Q28.4. In what month and year did you stop working in this job? MONTH YEAR
JOB 1				
JOB 2				
JOB 3				
JOB 4				
JOB 5				
JOB 6				
JOB 7				
JOB 8				
JOB 9				
JOB 10				

NO YES

29. Have any of these jobs ever made your chest tight or wheezy?

IF YES, (tick no or yes for each job)

	NO		YES
Job 1?			
Job 2?			
Job 3?			
Job 4?			
Job 5?			
Job 6?			
Job 7?			
Job 8?			
Job 9?			
Job 10?			

30. Have you had to leave any of these jobs because they affected your breathing? NO YES

IF YES, (tick no or yes for each job)

	NO		YES
Job 1?			
Job 2?			
Job 3?			
Job 4?			
Job 5?			
Job 6?			
Job 7?			
Job 8?			
Job 9?			
Job 10?			

31.Since the last survey have you been involved in an accident at home, work or elsewhere that exposed you to high levels of vapours, gas, dust or fumes? NO YES

IF YES,

31.1 Did you experience respiratory symptoms immediately following this exposure? NO YES

IF YES

31.1.1 Could you describe to me what it was? _____

Appendix B 1 – ECRHS II Main Questionnaire

Centres performing the extra occupational modules should at this point introduce the modular introductory questionnaire and complete modules as appropriate.

YEARS

32. At what age did you complete full time education?

If full time student enter 88

33. How often do you usually exercise so much that you get out of breath or sweat?

TICK ONE BOX ONLY

every day	1 ☐
4-6 times a week	2 ☐
2-3 times a week	3 ☐
once a week	4 ☐
once a month	5 ☐
less than once a month	6 ☐
never	7 ☐

34. How many hours a week do you usually exercise so much that you get out of breath or sweat?

TICK ONE BOX ONLY

none	1 ☐
about ½ hr	2 ☐
about 1 hour	3 ☐
about 2-3 hours	4 ☐
about 4-6 hours	5 ☐
7 hours or more	6 ☐

NO YES

35. Do you avoid taking vigorous exercise because of wheezing or asthma?

YEAR

36. When was your present home built? ☐☐☐☐

NO YES

37. Do you live in the same home as when you were last surveyed?

IF YES GO TO QUESTION 38, IF NO

TIMES

37.1. How many times have you moved since you were last surveyed?

YEARS

37.2. How many years have you lived in your current home?

37.3 Where do you currently live?

TICK ONE BOX ONLY

a different home, but still in the study sampling area	1 ☐
outside the sampling area but still in the same country	2 ☐
a different country	3 ☐

37.3.1. **IF A DIFFERENT COUNTRY** Which country? ☐☐☐

37.4 Which best describes the building in which you live? TICK ONE BOX ONLY

a) a mobile home or trailer?	1 ☐
b) a one family house detached from any other house?	2 ☐
c) a one family house attached to one or more houses?	3 ☐
d) a building for two families?	4 ☐
e) a building for three or four families?	5 ☐
f) a building for five or more families?	6 ☐
g) a boat, tent or van	7 ☐
e) other: _____	8 ☐

38. Does your home have any of the following?

	NO	YES
38.1 central heating	☐	☐
38.2 ducted air heating (forced air heating)	☐	☐
38.3 air conditioning	☐	☐

39. Which of the following appliances do you use for heating or for hot water?

	NO	YES
39.1 open coal, coke or wood fire	☐	☐
39.2 open gas fire	☐	☐
39.3 electric heater	☐	☐
39.4 paraffin heater	☐	☐
39.5 gas-fired boiler	☐	☐
39.6 oil-fired boiler	☐	☐
39.7 portable gas heater	☐	☐
39.8 other: _____	☐	☐

40. What kind of stove do you ***mostly*** use for cooking? TICK ONE BOX ONLY

a) coal, coke or wood (solid fuel)?	1 ☐
b) gas (gas from the mains)?	2 ☐
c) electric?	3 ☐
d) paraffin (kerosene)?	4 ☐
e) microwave	5 ☐
f) gas (gas from bottles or other non-mains source)	6 ☐
g) other: _____	7 ☐

40.1 IF YOU USE GAS FOR COOKING Which of the following do you have?

	NO	YES
40.1.1 gas hob		
40.1 2.gas oven		

41. What kind of stove was mostly used for cooking in the home you lived in when you were five years old? TICK ONE BOX ONLY

a) coal, coke or wood (solid fuel)?	1
b) gas (gas from the mains)?	2
c) electric?	3
d) paraffin?	4
e) gas (gas from bottles or other non-mains source)	5
f) don't know	6
g) other: _____	7

MINUTES

42. ***On average*** how long have you spent cooking with your stove each day over the ***last four weeks***? ☐☐☐

43. ***Over the last four weeks*** when you were cooking did you have a door or window to the outside air open

TICK ONE BOX ONLY

a) most of the time 1 ☐
b) some of time 2 ☐
c) rarely (or only occasionally) 3 ☐
d) I do not have a door or window that opens to the outside in my kitchen 4 ☐

NO YES DK

44. Do you have an extractor fan over the cooker?

***IF 'NO' OR 'DON'T KNOW' GO TO QUESTION 45, IF 'YES'*:**

44.1 When cooking, do you use the fan

TICK ONE BOX ONLY

a) all of the time? 1 ☐
b) some of the time? 2 ☐
c) none of the time? 3 ☐

NO YES DK

44.2 Does the fan take the fumes outside the house?

45. Does the room which you use most at home during the day

NO YES

45.1 have fitted carpets covering the whole floor? ☐ ☐
45.2 contain rugs? ☐ ☐
45.3 have double glazing? ☐ ☐

46 How old is the oldest carpet or rug in the room which you use most at home during the day? ,

TICK ONE BOX ONLY

a) less than one year 1
b) 1-5 years old 2
c) more than 5 years old 3

47 On what floor is the room which you use most at home during the day?
(The lowest floor of a building is 00) _____

48. Does your bedroom

NO YES

48.1 have fitted carpets covering the whole floor?
48.2 contain rugs?
48.3 have double glazing?

49 How old is the oldest carpet or rug in your bedroom

TICK ONE BOX ONLY

a) less than one year
b) 1-5 years old
c) more than 5 years old

TICK ONE BOX ONLY

50 How old is your mattress
- a) less than one year
- b) 1-5 years old
- c) more than 5 years old

51 What floor of the building is your bedroom on? (lowest=00)

NO YES

52. Do you sleep with the windows open at night during winter?

IF 'NO' GO TO QUESTION 53, IF 'YES':

52.1 Do you sleep with the windows open — TICK ONE BOX ONLY
- a) all of the time? 1 ☐
- b) sometimes? 2 ☐
- c) only occasionally? 3 ☐

53. Has there been any water damage to the building or its contents, for example, from broken pipes, leaks or floods? NO ☐ YES ☐ DK ☐

IF YES NO YES DK

53.1 Has there been any water damage in the last 12 months ☐ ☐ ☐

54. ***Within the last 12 months*** have you had wet or damp spots on surfaces inside your home other than in the basement (for example on walls, wall paper, ceilings or carpets)? NO YES

55. Has there ever been any mould or mildew on any surface, other than food, inside the home? NO YES DK

IF 'NO' OR 'DON'T KNOW' GO TO QUESTION 56, IF 'YES':

55.1.1-6 Which rooms have been affected? NO YES
- 55.1.1 bathroom(s) ☐ ☐
- 55.1.2 bedroom(s) ☐ ☐
- 55.1.3 living area(s) ☐ ☐
- 55.1.4 kitchen ☐ ☐
- 55.1.5 basement or attic ☐ ☐
- 55.1.6 other: _____ ☐ ☐

55.2 Has there been mould or mildew on any surfaces inside the home in the last ***12 months***? NO YES

Appendix B 1 – ECRHS II Main Questionnaire

'This scale looks like a thermometer; it allows you to rate your personal opinion regarding the following question on annoyance from air pollution. You can indicate your level of annoyance on this scale between 0 and 10 where 0 mean does not annoy at all' and 10 means intolerable annoyance.'

56 . How much are you annoyed by outdoor air pollution (from traffic, industry, etc.) if you keep the windows open?

10	intolerable annoyance
9	
8	
7	
6	
5	
4	
3	
2	
1	
0	doesn't annoy at all

Appendix B 1 – ECRHS II Main Questionnaire

THOSE WHO HAVE NOT MOVED HOME SINCE LAST SURVEY (Check with response to question 37)

GO TO QUESTION 58

THOSE WHO HAVE MOVED SINCE LAST SURVEY – answer 57

57. How much were you annoyed by outdoor air pollution (from traffic, industry, etc.) in your previous home, if you kept the windows open?

.

10	intolerable annoyance
9	
8	
7	
6	
5	
4	
3	
2	
1	
0	doesn't annoy at all

58. How often do cars pass your house? TICK ONE BOX ONLY

a) constantly	1
b) frequently	2
c) seldom	3
d) never	4

59. How often do heavy vehicles (e.g. trucks/buses) pass your house? TICK ONE BOX ONLY

a) constantly	1
b) frequently	2
c) seldom	3
d) never	4

60. Have you taken any of the following measures to reduce allergen or exposure to allergen in your home since the last survey?

	NO	YES
60.1 changed from carpet to a wooden or other smooth surface on floor of the room you use most		
60.2 changed from carpet to a wooden or to a smooth surface on floor of your bedroom		
60.3 bought a new carpet for the room you use most		
60.4 bought a new carpet for your bedroom		
60.5 used antidust-mite sprays		
60.6 put an allergy-proof cover on your mattress		
60.7 sold, given away or destroyed a pet dog or cat		

	NO	YES
61. Do you keep a cat?		
IF 'NO' GO TO QUESTION 62, IF 'YES'	NO	YES
61.1 Is your cat (are your cats) allowed inside the house?		
61.2 Is your cat (are your cats) allowed in the bedroom?		

	NO	YES
62. Do you keep a dog?		
IF 'NO' GO TO QUESTION 63, IF 'YES':	NO	YES
62.1 Is your dog (are your dogs) allowed inside the house?		
62.2 Is your dog (are your dogs) allowed in your bedroom?		

	NO	YES
63. Do you keep any birds?		
IF 'NO' GO TO QUESTION 64, IF 'YES':	NO	YES
63.1 Are any of these birds kept inside the house?		

	NO	YES	DK
64. Was there a cat in your home?			
64.1 during your first year of life	☐	☐	☐
64.2 when you were aged 1 to 4 years	☐	☐	☐
64.3 when you were aged 5-15 years	☐	☐	☐

	NO	YES	DK
65. Was there a dog in your home?			
65.1 during your first year of life	☐	☐	☐
65.2 when you were aged 1 to 4 years	☐	☐	☐
65.3 when you were aged 5-15 years	☐	☐	☐

	NO	YES	DK
66. Was there a bird in your home?			
66.1. during your first year of life	☐	☐	☐
66.2 when you were aged 1 to 4 years	☐	☐	☐
66.3 when you were aged 5-15 years	☐	☐	☐

67. What term best describes the place you lived most of the time when you were under the age of five years? TICK ONE BOX ONLY

a) farm	1	☐
b) village in a rural area	2	☐
c) small town	3	☐
d) suburb of a city	4	☐
e) inner city	5	☐

68. When you are near animals, such as cats, dogs or horses, do you ***ever***

	NO	YES
68.1 start to cough?	☐	☐
68.2 start to wheeze?	☐	☐
68.3 get a feeling of tightness in your chest?	☐	☐
68.4 start to feel short of breath?	☐	☐
68.5 get a runny or stuffy nose or start to sneeze?	☐	☐
68.6 get itchy or watering eyes?	☐	☐

69. When you are in a dusty part of the house, or near pillows or duvets do you ***ever***

	NO	YES
69.1 start to cough?	☐	☐
69.2 start to wheeze?	☐	☐
69.3 get a feeling of tightness in your chest?	☐	☐
69.4 start to feel short of breath?	☐	☐
69.5 get a runny or stuffy nose or start to sneeze?	☐	☐
69.6 get itchy or watering eyes?	☐	☐

70. When you are near trees, grass or flowers, or when there is a lot of pollen about, do you ***ever***

	NO	YES
70.1 start to cough?	☐	☐
70.2 start to wheeze?	☐	☐
70.3 get a feeling of tightness in your chest?	☐	☐
70.4 start to feel short of breath?	☐	☐
70.5 get a runny or stuffy nose or start to sneeze?	☐	☐
70.6 get itchy or watering eyes?	☐	☐

IF 'YES' TO ANY OF THE ABOVE:

70.7.1-4 Which time of year does this happen?	NO	YES
70.7.1 winter	☐	☐
70.7.2 spring	☐	☐
70.7.3 summer	☐	☐
70.7.4 autumn	☐	☐

71. How often do you eat pre-packaged food, such as tinned food or pre-prepared frozen meals? TICK ONE BOX ONLY

a) every day or most days	1	☐
b) at least once a week	2	☐
c) less than once a week	3	☐

NO YES

72 Do you take snacks between meals?

IF 'NO' GO TO QUESTION 73, IF 'YES':

72.1.1-3 Which of the following would you have as a snack at least ***once a week***?

NO YES

72.1.1 savoury biscuits or crisps ☐ ☐

72.1.2 sweets, chocolates or sweet biscuits ☐ ☐

72.1.3 fruit or vegetables ☐ ☐

73. Have you ever had an illness or trouble caused by eating a ***particular*** food or foods? NO YES

IF 'NO' GO TO QUESTION 74, IF 'YES':

73.1 Have you nearly always had the same illness or trouble after eating this type of food? NO YES

***IF 'NO' GO TO QUESTION 74, IF 'YES'*:**

73.1.1 What type of food was this? [List up to 3]

_____ ☐ ☐

_____ ☐ ☐

_____ ☐ ☐

73.1.2.1-6 Did this illness or trouble include NO YES

73.1.2.1 a rash or itchy skin? ☐ ☐

73.1.2.2 diarrhoea or vomiting? ☐ ☐

73.1.2.3 runny or stuffy nose? ☐ ☐

73.1.2.4 severe headaches? ☐ ☐

73.1.2.5 breathlessness? ☐ ☐

73.1.2.6 other: _____ ☐ ☐

NO YES

74. Have you ever smoked for as long as a year?

['YES' means at least 20 packs of cigarettes or 12 oz (360 grams) of tobacco in a lifetime, or at least one cigarette per day or one cigar a week for one year]

IF 'NO' GO TO QUESTION 75, IF 'YES':

YEARS

74.1 How old were you when you started smoking?

NO YES

74.2 Do you **now** smoke, as of ***one month ago***?

***IF 'NO' GO TO QUESTION 74.3, IF 'YES'*:**

74.2.1-4 How much do you ***now*** smoke on average? NUMBER

74.2.1 number of cigarettes per day

74.2.2 number of cigarillos per day

74.2.3 number of cigars a week

74.2.4 pipe tobacco in a) ounces / week

b) grams / week

NO YES

74.3 Have you stopped or cut down smoking?

IF 'NO' GO TO QUESTION 74.4, IF 'YES': YEARS

74.3.1 how old were you when you stopped or cut down smoking?

74.3.2.1-4 ***on average*** of the entire time you smoked, before you stopped or cut down, how much did you smoke? NUMBER

74.3.2.1 number of cigarettes per day

74.3.2.2 number of cigarillos per day

74.3.2.3 number of cigars a week

74.3.2.4 pipe tobacco in a) ounces / week

b) grams / week

NO YES

74.4 Do you or did you inhale the smoke?

75. Have you been **regularly** exposed to tobacco smoke in the last ***12 months***? ['Regularly' means on most days or nights] NOYES

IF 'NO' GO TO QUESTION 76, IF 'YES':

75.1. Not counting yourself, how many people in your household smoke regularly? NUMBER

NO YES

75.2 Do people smoke regularly in the room where you work?

75.3 How many hours per day are you exposed to ***other people's*** tobacco smoke? HOURS

75.4 Please provide more information.

How many hours per day, are you exposed to other peoples tobacco smoke in the following locations? HOURS

at home

at workplace

in bars, restaurants, cinemas or similar social settings

elsewhere

76. Have you used any **inhaled** medicines to help your breathing at any time in the last ***12 months***? NO YES

IF NO' GO TO QUESTION 77, IF 'YES':

Which of the following have you used in the last ***12 months***? NO YES

76.1 short acting **beta-2-agonist inhalers**

(Please include combinations that include beta 2 and steroids in section 76.5)

76.1.1 If used, which one? _____

76.1.2 What type of inhaler do you use?

NUMBER

76.1.3. What is the dose per puff (in micrograms)?

76.1.4. In the last 3 months, how have you used them: TICK ONE BOX ONLY

a) when needed 1

b) in short courses 2

c) continuously 3

d) not at all 4

If answer to 76.1.4 is <u>when needed</u>: NUMBER

76.1.5 Number of puffs per month

If answer to 76.1.4 is <u>in short courses</u> NUMBER

76.1.6 number of courses
76.1.7 number of puffs per day
76.1.8 average number of days per month

If answer to 76.1.4 <u>is continuously</u> NUMBER

76.1.9 number of puffs per day

NO YES

76.2 **long acting beta-2-agonist inhalers**
(Please include combinations that include beta 2 and steroids in section 76.5)

76.2.1 If used, which one? _____
76.2.2 What type of inhaler do you use?

NUMBER

76.2.3. What is the dose per puff (in micrograms)?

76.2.4. In the last 3 months, how have you used them: TICK ONE BOX ONLY
a) when needed 1
b) in short courses 2
c) continuously 3
d) not at all 4

If answer to 76.2.4 is <u>when needed</u>: NUMBER
76.2.5 Number of puffs per month

If answer to 76.2.4 is <u>in short courses</u> NUMBER

76.2.6 number of courses
76.2.7 number of puffs per day
76.2.8 average number of days per month

If answer to 76.2.4 is <u>continuously</u> NUMBER

76.2.9 number of puffs per day

NO YES

76.3 **non-specific adrenoreceptor agonist inhalers**

76.3.1 If used, which one? _____

NO YES

76.4 **anti-muscarinic inhalers**

76.4.1 If used, which one? _____
76.4.2 What type of inhaler do you use?

NUMBER

76.4.3. What is the dose per puff (in micrograms)?

76.4.4. In the last 3 months, how have you used them: TICK ONE BOX ONLY

a) when needed	1
b) in short courses	2
c) continuously	3
d) not at all	4

If answer to 76.4 .4 is <u>when needed</u>: NUMBER

76.4.5 Number of puffs per month

If answer to 76.4.4 is in <u>short courses</u> NUMBER

76.4.6 number of courses	
76.4.7 number of puffs per day	
76.4.8 average number of days per month	

If answer to 76.4.4 is <u>continuously</u> NUMBER

76.4.9 number of puffs per day

NO YES

76.5 **inhaled steroids**

(***if combined B2 and steroid please insert inhaled steroid dose***)

76.5.1 If used, which one? _____

76.5.2 What type of inhaler do you use?

NUMBER

76.5.3. What is the dose per puff (in micrograms)?

76.5.4. In the last 3 months, how have you used them: TICK ONE BOX ONLY

a) when needed	1
b) in short courses	2
c) continuously	3
d) not at all	4

If answer to 76.5.4 is <u>when needed</u>: NUMBER

76.5.5 Number of puffs per month

If answer to 76.5.4 is in <u>short courses</u> NUMBER

76.5.6 number of courses	
76.5.7 number of puffs per day	
76.5.8 average number of days per month	

If answer to 76.5.4 is <u>continuously</u> NUMBER

76.5.9 number of puffs per day

NO YES

76.6 **inhaled cromoglycate/nedocromil**

76.6.1 If used, which one? _____

NUMBER

76.6.2. What is the dose per puff (in milligrams)?

76.6.3. In the last 3 months, how have you used them: TICK ONE BOX ONLY

a) when needed	1 ☐
b) in short courses	2 ☐
c) continuously	3 ☐
d) not at all	4 ☐

If answer to 76.6.3 is <u>when needed</u>: NUMBER

76.6.4 Number of puffs per month

If answer to 76.6.3 is in <u>short courses</u> NUMBER

76.6.5 number of courses	☐☐
76.6.6 number of puffs per day	☐☐
76.6.7 average number of days per month	☐☐

If answer to 76.6.3 is <u>continuously</u> NUMBER

76.6.8 number of puffs per day

NO YES

76.7 **inhaled compounds**

76.7.1 If used, which one? _____

76.7.2 What type of inhaler do you use?

NUMBER

76.7.3. What is the dose per puff (in micrograms)?

77. Have you used any **pills**, **capsules**, **tablets** or **medicines**, other than inhaled medicines, to help your breathing at any time in the last ***12 months***? NO YES

IF 'NO' GO TO QUESTION 78, IF 'YES':

Which of the following have you used in the last ***12 months***?

NO YES

77.1 **oral beta-2-agonists**

77.1.1 If used, which one? _____

77.1.2 what dose of tablet ☐☐

77.1.3. In the last 3 months, how have you used them: TICK ONE BOX ONLY

a) when needed	1 ☐
b) in short courses	2 ☐
c) continuously	3 ☐
d) not at all	4 ☐

If answer to 77.1.3 is <u>when needed</u>: NUMBER

77.1.4 number of tablets per month

If answer to 77.1.3 is in <u>short courses</u> NUMBER

77.1.5 number of courses	☐☐
77.1.6 tablets per day	☐☐
77.1.7 average number of days per month	☐☐

If answer to 77.1.3 is <u>continuously</u> NUMBER

77.1.8 tablets per day

NO YES

77.2 **oral methylxanthines**

77.2.1 if used, which one?
77.2.2 what dose of tablet

77.2.3. In the last 3 months, how have you used them: TICK ONE BOX ONLY

a) when needed	1 ☐
b) in short courses	2 ☐
c) continuously	3 ☐
d) not at all	4 ☐

If answer to 77.2.3 is <u>when needed</u>: NUMBER

77.2.4 number of tablets per month

If answer to 77.2.3 is in <u>short courses</u> NUMBER

77.2.5 number of courses ☐☐
77.2.6 tablets per day ☐☐
77.2.7 average number of days per month ☐☐

If answer to 77.2.3 is <u>continuously</u> NUMBER

77.2.8 tablets per day

NO YES

77.3 **oral steroids**

77.3.1 If used, which one? _____ ☐☐
77.3.2 what dose of tablet ☐☐.

77.3.3. In the last 12 months, how have you used them: TICK ONE BOX ONLY

a) when needed	1 ☐
b) in short courses	2 ☐
c) continuously	3 ☐

If answer to 77.3.3 is <u>when needed</u>: NUMBER

77.3.4 number of tablets per month

If answer to 77.3 3 is in <u>short courses</u> NUMBER

77.3.5 number of courses ☐☐
77.3.6 tablets per day ☐☐
77.3.7 average number of days per month ☐☐

If answer to 77.3.3 is <u>continuously</u> NUMBER

77.3.8 tablets per day

NO YES

77.3.9. Have you used them in the last ***3 months***?

NO YES

77.4 oral anti-leukotrienes

77.4.1 If used, which one? _____
77.4.2 what dose of tablet

77.4.3. In the last 3 months, how have you used them: TICK ONE BOX ONLY
a) when needed 1 ☐
b) in short courses 2 ☐
c) continuously 3 ☐
d) not at all 4 ☐

If answer to 77.4.3 is <u>when needed</u>: NUMBER
77.4.4 number of tablets per month

If answer to 77.4.3 is in <u>short courses</u> NUMBER
77.4.5 number of courses
77.4.6 tablets per day
77.4 .7 average number of days per month

If answer to 77.4.3 is <u>continuously</u> NUMBER
77.4.8 tablets per day

NO YES

77.5 **ketotifen**

77.5.1 If used, which one? _____
77.5.2 what dose of tablet

77.5.3. In the last 3 months, how have you used them: TICK ONE BOX ONLY
a) when needed 1 ☐
b) in short courses 2 ☐
c) continuously 3 ☐
d) not at all 4 ☐

If answer to 77.5.3 is <u>when needed</u>: NUMBER
77.5.4 number of tablets per month

If answer to 77.5.3 is in <u>short courses</u> NUMBER
77.5.5 number of courses
77.5.6 tablets per day
77.5.7 average number of days per month

If answer to 77.5.3 is <u>continuously</u> NUMBER
77.5.8 tablets per day

NO YES

78. Since the last survey have you ever used inhaled steroids (show list)?
IF NO GO TO QUESTION 79 YEARS
78.1. How old were you when you first started to use inhaled steroids?

NO YES

78.2. Have you used inhaled steroids ***every year*** since the last survey?

IF NO GO TO QUESTION 78.3, IF YES MONTHS

78.2.1. On average how many months each year have you taken them?

NOW GO TO Q79

78.3 How many of the years since the last survey have you taken inhaled steroids? YEARS

78.4. On average how many months of each of these years have you taken them? MONTHS

NO YES DK

79. Have you been vaccinated for allergy since the last survey?

NO YES

79.1 Have you been vaccinated for allergy in the last ***12 months***?

80. Have you had any other **injections** to help your breathing at any time in the last ***12 months***? NO YES

IF 'NO' GO TO QUESTION 81, IF 'YES':

80.1 What injections? _____

81. Have you had any suppositories to help your breathing at any time in the last ***12 months***? NO YES

IF 'NO' GO TO QUESTION 82, IF 'YES':

81.1 What suppositories? _____

82 . Have you used any other **remedies** to help your breathing at any time in the last ***12 months***? NO YES

***IF 'NO' GO TO QUESTION 83 IF 'YES'*:**

82.1. What remedies? _____

83. Has your doctor ever prescribed medicines, including inhalers, for your breathing? NO YES

IF 'NO' GO TO QUESTION 84, IF 'YES':

83.1 If you are prescribed medicines for your breathing, do you ***normally*** take TICK ONE BOX ONLY

a) all of the medicine? 1
b) most of the medicine? 2
c) some of the medicine? 3
d) none of the medicine? 4

83.2 ***When your breathing gets worse***, and you are prescribed medicines for your breathing, do you normally take — TICK ONE BOX ONLY

a) all of the medicine? 1
b) most of the medicine? 2
c) some of the medicine? 3
d) none of the medicine? 4

83.3 Do you think it is bad for you to take medicines all the time to help your breathing? NO YES

83.4 Do you think you should take as much medicine as you need to get rid of ***all*** your breathing problems? NO YES

84. Since the last survey have you visited a hospital casualty department or emergency room because of breathing problems? NO YES

IF NO GO TO Q85, IF YES

84.1 Have you visited a hospital casualty department or emergency room because of breathing problems in the ***last 12 months***? NO YES

IF NO GO TO 85, IF YES NO YES

84.1.1 Was this due to asthma, shortness of breath or wheezing?

84.1.2 How many times ***in the last 12 months*?** TIMES |

85. Since the last survey have you spent a night in hospital because of breathing problems? NO YES

IF NO GO TO Q86 IF YES

85.1 Have you spent a night in hospital because of breathing problems in the ***last 12 months***? NO YES

IF NO GO TO Q86, IF YES NO YES

85.1.1 Was this due to asthma, shortness of breath or wheezing?

85.1.2 How many nights have you spent on each of the following types of ward in ***the last 12 months?*** — NUMBER

General
Chest medicine
Rehabilitation
Intensive care unit
Other

86. Since the last survey have you been seen by a doctor because of breathing problems or because of shortness of breath? NO YES

IF NO GO TO Q87, IF YES

86.1 Have you been seen by a general practitioner because of breathing problems or shortness of breath in the ***last 12 months?*** NO YES

IF NO GO TO Q86.4, IF YES NO YES

86.2.Was this due to asthma, shortness of breath or wheezing?

86.3 How many times have you been seen by your general practitioner because of breathing problems or shortness of breath in each of these locations over the last 12 months?

	NUMBER
at home (excluding emergency visits)	
in his surgery	
at home in an emergency	
at another location	

86.4 Have you seen a specialist (chest physician, allergy specialist, internal medicine specialist, ENT doctor) because of your breathing problems or shortness of breath ***in the last 12 months***? NO YES

IF NO GO TO Q87 IF YES NUMBER

86.4.1 How many times?

87. Are you given regular appointments to be seen by a doctor (or nurse) for your asthma, wheezing or shortness of breath? NO YES

IF NO GO TO Q88 IF YES NO YES

87.1.Are you given regular appointments with a hospital doctor? NO YES

87.2 Are you given regular appointments with your general practitioner? NO YES

87.3. Are you given regular appointments with a nurse?

88. How many times have you visited the following because of breathing problems or shortness of breath ***in the last 12 months?***

	NUMBER
88.1 nurse	
88.2 physiotherapist	
88.3 practitioner of 'alternative' medicine	

89. Have you had any clinical or laboratory tests because of asthma wheezing or shortness of breath ***in the last 12 months***? NO YES

IF NO GOT Q90 ,IF YES

89.1. How many times have you had the following ***in the last 12 months?***

	NUMBER
Breathing test in a laboratory specially for lung function measures	
Skin test for allergy	
Blood test for allergy	
x-rays	

NO YES

90. Are you currently working?

IF NO GO TO Q90.2 IF YES

90.1. How many days of work have you lost because of asthma, shortness of breath or wheezing in the last **12 months?** NUMBER

90.2. Were you forced to **give up working** because of asthma, wheezing or shortness of breath in the last **12 months?** NO YES

IF NO GO TO 91. IF YES DAY MONTH YEAR

91.2.1. When?

91. Have there been any days when you have had to **give up activities other than work** (e.g. looking after children, the house, studying) because of your asthma, wheezing or shortness of breath in the ***last 12 months?*** NO YES

IF NO YOU HAVE FINISHED THE QUESTIONNAIRE IF YES

91.2. How many days on average each month?

M F

Subjects Gender

DAY MONTH YEAR

Subjects Date of Birth

INTERVIEW TYPE? TICK ONE BOX ONLY

a) At centre face to face 1
b) At home face to face 2
c) By telephone 3
d) Self completed at home 4

END FIELDWORKER NUMBER

Appendix B 2 – ECRHS II Main Questionnaire Instructions and Coding

Introduction

The use of a questionnaire to collect information makes it possible to obtain answers to important questions in a standardised way. The reliability of the questionnaire depends on the behaviour of the interviewer, and therefore it is important that the questions are read exactly as they are printed and that no non-verbal clues are given.

Basic rules

1. Interviews should take place where there is minimal disturbance, where both interviewer and subject can be comfortable, and where eye contact and hence the attention of the subject is maintained.
2. The interviewer is started when the interviewer has the subject's full attention, with the introductory sentence used in the questionnaire.
3. Occasionally, the interview may be complicated by one of the following difficulties:
   a) The subject will not understand the question.
   b) The subject or interviewer will find an ambiguity in the question.
   c) The subject's answer may be inappropriate to the question.
4. It is very important that all interviewers in all the centres follow the same procedure for solving problems, so that it is possible to compare the answers given in one centre with the answers given in another.
5. The following general rules should be obeyed when there is a problem:
   a) The question is repeated exactly as written, emphasising the wording where there is ambiguity,
   b) The subject is reminded that he/she should try to answer 'YES' or 'NO' to each of the questions.
   c) If an answer of 'YES' or 'NO' is required and the subject does not understand the question even when repeated, the answer is coded as 'NO', (unless a 'DON'T KNOW' option is specifically provided).
   d) Where an answer is required to a quantitative or semi-quantitative question, the subject's 'best guess' may be accepted.
   e) An explanation may be given to the subject, instructions for these are provided. Words in the question that should be stressed are underlined. Notes in square brackets are guidance and should not be read out.

6. Many questions ask 'since the last survey'. The interviewer should know the month and year of the last survey so that they can remind the subject when this was.

If, during the interview, a subject requests further information or clarification of a question that is not possible according to the questionnaire rules, the interviewer should explain to the subject that these points can be discussed at the end of the questionnaire.

Although this is essentially a study of asthma, the word 'asthma' is considered to be emotive and it is generally replaced by 'respiratory health' or 'breathing problems'. If the word 'asthma' does not appear in the question, it should not be used as any further clarification or discussion with the subject.

Training

Before starting the survey, the questionnaire and instructions should be studied and any difficulties discussed. Trainee interviewers must become familiar with the flow of questions. Interviewers should test the questionnaire on 10 or more subjects (such as hospital patients), who have at least some chest symptoms, as there is usually no difficulties with subjects who have no symptoms. These interviews should be witnessed by an experienced person who can identify mistakes or doubtful points that need clarification.

Recording the replies to the questions

Most of the questions are of the 'YES' or 'NO' type and where applicable 'DON'T KNOW'. If there is not provision for a 'DON'T KNOW' answer and the subject is uncertain of the answer it is recorded as 'NO'. If the answer to the question is a number, this should be recorded directly in the boxes provided. Where the answer is a date, this should be written out in full. The interviewer should follow instructions given in the questionnaire regarding which questions to ask according to the subject's response. In cases when further questions are irrelevant (and this can follow a 'YES' or a 'NO' answer) a 'skip' ('GO TO') will direct interviewers to the next question. Occasionally, there are 'skips' within sub-divisions of questions. For questions where

there is a choice of answers there are two formats. If there is only one possible or likely answer the format is 'TICK ONE BOX ONLY'. If the subject cannot decide between two options, then the choice which applies most of the time and most recently should be recorded. The second format is a 'YES' or 'NO' box to each of a number of possibilities or choices in cases where they could all apply. Some of these questions have as a final option 'OTHER'. If the subject chooses this option and, therefore, gives an unusual or unexpected answer, the box next to this option is ticked 'YES' and the answer written in freehand and left un-coded. The 'OTHER' option is also chosen if the subject is asked to list items and there is insufficient space, the most often used or the item the subject considered most important should be recorded.

Coding

Answers to questions are either chosen from a selection of options or written freehand. Sometimes not all the answers are coded, but the information is there for reference at a later date. All freehand answers are coded after the questionnaire has been administered.

Additional clarification of questions

QUESTION 1

These questions are intended to identify particpants who have occasionally and/or frequent wheezing. Subjects may confuse wheezing with snoring or bubbling sounds in the chest. '*Wheeze*' can be described as '*A whistling sound, whether high or low pitched and however faint*'. If the question is not understood, a vocal demonstration of wheezing by the interviewer can be helpful. No distinction is made between those who only wheeze during the day and those who only wheeze at night.

QUESTION 2

The question refers to waking with tightness in the chest at any time regardless of whether the subject has had a cold during that period.

QUESTION 3, 4, and 5

These questions distinguish between attacks of breathlessness during periods of inactivity, 'exercise-induced' breathlessness and night-time (or during 'sleep period')

breathlessness. In the question regarding breathlessness following activity, the word 'following' should be stressed. If the subject has not carried out any strenuous activity in the last 12 months for whatever reason, the answer is recorded as 'NO'. This includes those subjects who avoid strenuous activity because they would become breathless.

Supplementary questions have been added to question 5 to determine whether symptoms have been frequent in the last 3 months.

QUESTION 6, 7, and 8

In parts of the world where respiratory symptoms are most common at other times in the year, the appropriate word should be substituted for 'winter'. Where there is no seasonal variation in respiratory symptoms the word 'winter' should be omitted. When night shift workers are interviewed the words ' on getting up' should be used instead of 'first thing in the morning'. A cough with their first smoke or on going out of doors is included. Clearing the throat or a single cough is excluded. The word 'usually' should be emphasised. An occasional cough may be considered as normal and the answer should be recorded as 'NO'. As a rough guide single coughs at a frequency of less than six a day are 'occasional'. The words 'do you cough like this' refers to whatever kind of cough or frequency of cough the subject has already reported in the previous question and whenever it occurred. 'Three months' refers to three consecutive months, and 'each year' to the last two years. There are special rules for recording the answers to question 7. If the answer to question 7 is doubtful, the interviewer should then ask question 8.1. The answer to question 8.1 is recorded as the answer to question 7. The interviewer should then ask question 8, followed by 8.1 again and the answers recorded as they are given.

QUESTION 9 and 10

As with cough, phlegm with the first smoke or on going out of doors is included, but not mucoid discharge from the nose. Contrary to cough, however, 'occasional' phlegm production from the chest is considered abnormal if it occurs twice or more per day. The interviewer may use any suitable word that accords with local usage provided that it distinguishes phlegm from the chest or throat from pure nasal discharge. Some subjects admit to bringing up phlegm without admitting to coughing. This should be accepted without charging the replies to the questions about cough. A

claim that phlegm is coughed from the chest but swallowed counts as a positive reply. For question 9, question 10.1 is used to ascertain the answer to question 9, as described above.

QUESTION 11

The phrase 'trouble with your breathing' should not be elaborated upon. If the subject feels that there is something wrong with their breathing, whatever the reason, the answer is recorded as 'YES'.

QUESTION 12

This question refers to any physical disability other than chest or heart disease (for example, confined to a wheelchair) that prevents the subject from walking normally and that has been present for at least 12 months. This precise nature of disability should be recorded freehand but not coded. If the subject has a temporary physical disability that has not been present through the last 12 months, the questions are asked pertaining to the time when the subject was fit. In order to increase uniformity between surveys carried out a different breathlessness is at its worst. If the subject is disabled from walking (e.g. confined to a wheelchair or uses crutches continuously) these questions are omitted and the disabling condition is recorded freehand. 'Hurrying' implies walking quickly. These questions refer to the average condition during the previous two winters. If the subject avoids hurrying because they would become breathless and, therefore, the question is irrelevant, the answer is recorded as 'NO'.

QUESTION 13

This question assesses cyclical variation in breathing problems in women. Women should identify the most appropriate response for them.

QUESTION 14

14.1 Further explanation of the definition of 'asthma' should not be given. If the term is not understood, the answer should be recorded as 'NO'.

14.2 If the subject does not remember their age at time of their first of most recent attack of asthma, the interviewer should ask the subject to make a decision as to what age should be recorded. This is more likely with the first, rather than the most recent, but an estimate may also be given for most 'recent attack'.

14.4.1 All the relevant months when the subject commonly has asthma attacks should be recorded as 'YES'. If the subject replies 'all the time' or 'at any time' the 'YES' is recorded for all the months.

14.5-9 Subjects are asked how frequently they have symptoms and should choose the most appropriate response.

14.10 'Currently taking medication' is defined as 'having the medication available at home'. Alternative therapy is included if prescribed by a licensed practitioner.

QUESTION 15

The term nasal allergies includes all symptoms of rhinitis, whether seasonal or perennial, and whatever the allergens associated with symptoms. This question is the same question as in ECRHS I.

If the subject cannot remember how old they were when they first had hayfever or nasal allergy, then the interviewer should prompt the person to give an approximate answer.

QUESTION 16

These questions are similar to those adopted by ISAAC for the definition of hayfever in children. These questions are asked after question 15 in order to maintain similarity with ECRHS I. However, where someone has answered 'YES' to question 15 but 'NO' to question 16, the question should be repeated and the response recorded. However, the interviewer should not prompt the subject further, even if the subject again replies 'NO'.

QUESTION 17

17.1.1 For steroid nasal sprays, each country should make the lists of the drugs used in their country, and the interviewer should show these lists. The list <u>should not</u> include cromolyn and antihistamine sprays. If the participant reports having used any medication on the list, the answer is 'YES'.

17.1.2 Count the number of years since the first treatment even if the subject uses treatment only some months each year (e.g. seasonal rhinitis)

17.2 For antihistamines, each country should make a list of pills, capsules or tablets <u>used to treat nasal disorder</u> in their country, and the interviewer should show these lists. The list <u>should not</u> include compound syrups with antihistamines.

Subjects should only respond 'YES' if they have used these medications for the treatment of their nasal disorder.

17.2.1 Same as question 17.1.2

QUESTION 18

This question has been retained to allow comparison with ECRHS I. If the term eczema is not understood the answer should be recorded as 'NO'.

QUESTION 19

This question is designed following agreed working party definitions on eczema.

19.2 The answer should be recorded as 'YES' if any of the stated locations are affected.

QUESTION 20

If the response to breathing difficulties associated with the use of any medicine is 'YES', the appropriate group should not be recorded and the exact drug recorded freehand. Skin reactions to drugs are not included.

QUESTION 21

Subjects may need to use the 'best guess' to give their mothers age at the time they were born.

QUESTION 23

'Hospitalised' means spending a night as an inpatient in hospital. 'Lung disease' means any condition that was related to lower respiratory, chest or lung problems including chest infections, pneumonia and asthma.

QUESTION 24

Local terminology relevant to day care for children under five years can be used. If a child is looked after by a childminder or 'day-mother', together with children from other families this is considered to be 'day-care'. Interviewers should ask for age in years and if "x years y months" is written, only x years should be recorded. If interviewers have written for example "3-4 years", 3 years should be recorded (i.e., the lower figure).

QUESTION 25

'Regularly' sharing a bedroom means routinely at home for more than one year as opposed to when visiting relatives or for short holiday periods.

QUESTION 26 – QUESTION 31

A full-time student is defined as one currently attending an educational establishment and not having full-time employment. If the subject is a student, but works part-time this counts as full-time education.

QUESTION 28

This question is the occupational matrix and instructions on how to complete it are in Appendix C3.

QUESTION 32

Responses are recorded in years. When subjects give an answer in years and months, only the number of years should be recorded and should be rounded down. This question can be difficult if, for example, a subject has worked and then becomes a student. Should this occur, please contact the ECRHS II Co-ordinating Centre to advise on coding.

QUESTION 33 and 34

Some people may 'exercise' as part of their work. In this question 'exercise' at work is included, if it makes the subject 'get out of breath' or 'sweat'.

QUESTION 36

The age of the present home gives an indication of the amount of insulation and degree of air-tightness, but may not be known to individuals who have recently moved. If the subject is unsure of the year in which their house was built, the interviewer should record their 'best guess'.

QUESTION 37

This question is used to identify subjects who have moved house since the last survey.

37.2 The interviewer should request an answer in whole years and if "x years y months" is written, only x years should be recorded (i.e., rounded down).

37.3 The local questionnaire can be modified to identify the sampling area for ECRHS I or the interviewer may code directly from their knowledge of the sampling area of the ECRHS I.

37.4 If a subject owns more than one home or has a holiday caravan or boat, the question pertains to the dwelling in which the subject spends most time. Buildings that have been built or reconstructed behind old facade should count as new buildings (from the date of reconstruction).

QUESTION 38, 39, and 40

These questions refer to heating and cooking fuels and give some idea of indoor air pollution. Information on the type of heating will provide information on temperature differentials and humidity changes throughout the house, which can occur when there is no central heating. 'Central heating' is defined as a gas or oil fired boiler feeding radiators in every or nearly all the rooms in the house or electric storage heaters used throughout the house. Central heating includes radiators that are in most rooms and which maintain a regular temperature for most of the day. Heating of this type in part of the house, for example, in the living room only, also counts as 'YES'. Air conditioning is either 'central' air conditioning or 'individual units' in the windows of rooms. 'Open fires' as a form of heating refers to a 'fireplace' a 'stove' or a 'woodstove' used for heating or hot water, but not for cooking, in a room which is inhabited rather than in an unused basement, whether or not it is part of a ducted heating system. If the subject has additional forms of heating (for example, electric storage heaters) and they have been used at least once in the last 12 months, the answer is recorded as 'YES'. If other heaters are present but have never been used in the previous 12 months, the answer to the question is 'NO'. For countries where 'distance heaters' and 'electrical radiators' are commonly used, the answer should be recorded as 'YES' or 'OTHER' and the Fieldworkers should refer to the coding instructions.

QUESTION 42

This figure relates to the average time spent cooking with the main cooking appliance referred to in question 40. Subjects must think about the last four weeks and make an estimate of the time he/she prepares meals on their stove or spends cooking each day. Time when the oven is on should be included in this amount, but only if the subject is the one who is preparing the meal. The answer should be recorded in minutes.

QUESTION 43

If someone has responded 'never' – they should be coded as 3 – 'rarely or occasionally'

QUESTION 45 and 48

This question asks about the type of window insulation and furnishings that are present in the home. 'Double glazing' means double or triple windowpanes. If these are removable panes and are only used for part of the year and they have been used in the last 12 months, the answer is recorded as 'YES'.

QUESTION 46 and 49

If someone has no carpets or rugs code 4

QUESTION 47 and 51

The lowest floor of a building that is habitable is considered as 00, and all floors above this are numbered from there. Therefore, for some homes 00 will be equivalent to the ground floor and for others it will be equivalent to the first floor 01.

QUESTION 53, 54 and 55

These questions refer to the amount of damp or mould that is apparent in the subject's home. The interviewer should stress 'in the last 12 months'. Where appropriate 'basements' or 'cellars' are rooms that are below ground floor level that the subject has permanent access to and that are immediately below the subject's residence.

QUESTION 56 and 57

The interviewer should read out this paragraph as it is presented and the subject should provide a number that rates their response, while they look at the thermometer. This full number is entered.

QUESTION 60

This question records changes made to reduce allergen. The answer should only be coded as 'YES' if changes were specifically made to reduce allergen exposure.

QUESTION 61, 62 and 63

These questions are about pets currently owned and to establish the length of time spent indoors by the pet, which is thought to reflect the amount of animal-derived indoor allergen present.

QUESTION 68 and 69

These questions refer to symptoms related to exposure to aeroallergens, including animal dander and dust mite allergen. Each part of the question should be read out by the interviewer and a 'YES' or 'NO' answer recorded.

QUESTION 70

A question on seasonality of symptoms is included that requires a 'YES' or 'NO' answer to each season. If different symptoms occur at different seasons, the interviewer should record a 'YES' to all the relevant seasons. The seasons and months included may be adapted locally for different parts of the world.

QUESTION 71, 72 and 73

Questions on diet refer to food consumed at home and not in restaurants. These questions relate to the amount of convenience food and 'junk' food the subject is consuming, which will give an indication of sodium and food additive intake. The food 'categories' are:

1) savoury foods (salty/fatty)
2) sweet foods (may be fatty)
3) fruit and vegetables

Cheese as a snack is included as a 'savoury' food. 'Every day or most days' means four or more days a week. For the question on 'trouble after eating foods', the type or types of food are recorded freehand. If more than three foods are involved, three foods or types of food that cause the most severe problems should be recorded. In countries where food additives are not permitted in frozen foods, the words 'pre-prepared frozen meals' is omitted. Mineral water is not included as a 'fizzy drink'.

QUESTION 74.1

If the subject is in doubt about their smoking status the interviewer should read the definition of 'smoking'. If the subject answers 'YES' but does not remember when they started smoking, the interviewer should ask for an approximate age. Interviewers

should ask for age in years and if "x years y months" is written, only x years should be recorded. If interviewers have written for example "17-18 years", 17 years should be recorded (i.e., the lower figure).

QUESTION 74.2

The question on 'present' smoking status relates to the last month. For example, if the subject smoked their last cigarette two weeks ago the answer is 'YES'. The words 'as of one month ago' should be stressed. If the subject's smoking habits have changed, they will be asked how old they were when they cut down or stopped smoking. The tendency will be to remember 'how long ago' rather than 'at what age', so the interviewer will need to work out with the subject the age at cutting down. The subject should then be asked (QUESTION 74.3.2) how much he/she smoked on average the entire time that he/she smoked before cutting down. The questions are designed so that a consistent smoker answers only about what he/she smokes now and ex-smoker answers about what he/she now smokes and what he/she smoked before.

'Home' or 'self-rolled' cigarettes are included in 'number of cigarettes' smoked. The question on 'pipe tobacco' are to be answered in either ounces or grams, depending on which the subject is most familiar with. 2-3 cigars per month should be recorded as less than one per week and less than 7 cigarettes per week is less than one cigarette a day.

QUESTION 75

The question on inhalation of cigarette smoke refers to the way that the subject smoked for most of the time. The question on regular exposure to smoking is concerned with exposure to environmental tobacco smoke and related to the last 12 months only. The question may be irrelevant to a present smoker (where that answer is 'YES'), but should still be asked. People in the household (apart from the subject) who smoke regularly may include a babysitter/nanny or housekeeper/au pair, who are present most of the time or live in. It also includes regular visitors who smoke in the house at least five days a week. It does not include occasional visitors who smoke. If the subject works in a very large room (open planned office or factory) where people smoke some distance away, 10m (3ft) can be regarded as a cut-off. In order to obtain more information on the location, in which people are exposed to tobacco smoke, subjects are asked at what locations they have experienced their exposures. However,

in order to maintain complete comparability with ECRHS I question 75.3 remains unchanged. 'Elsewhere' may include the home of relatives or home of friends. If the interviewer has written "x hours and y minutes", x hours should be recorded (i.e., it should be rounded down).

QUESTION 76 and 77

The subject should be asked to bring along any medication that he/she is currently taking. The question refers to the last 12 months so it is possible that the subject no longer has the medicine or that it is not in its original container, so therefore, the interviewer can show the subject photographs of inhalers/medicines at the time of questioning. If two or more inhalers or medicines from the same group are simultaneously used, the one that is most often or most recently used should be recorded. Menthol rubs and similar 'inhaled' medicine are not counted as inhalers.

The general format of the question is to ask about use in the last 12 months, and then use in a shorter period of time. Subjects should identify where during recent usage these drugs are used when needed, in short courses or continuously. However some may not have used them at all in the recent period-this option is provided. Having done this, subjects are asked to describe their average use of these drugs over the specified time period.

QUESTION 78

Question 78 is designed to divide subjects into those who, since the last survey have

- never used inhaled steroids
- used inhaled steroids most months since the last survey
- used inhaled steroids every month every year since the last survey
- used inhaled steroids for only some months of some years since the last surveys

From the information provided the total months that people have taken steroids since the last survey can be determined.

QUESTION 79

These questions refer to desensitisation injections or immunotherapy. The subject may volunteer this information. If the question is not understood, the answer is recorded as 'NO'. Desensitisation injections should be distinguished from other injections to 'help

breathing', which can include penicillin shots in acute respiratory infection or depot steroids. It does not include antiviral vaccines and translations of the word 'immunotherapy' should ensure that there is no misunderstanding.

QUESTION 83

This question is about the subject's attitude to the use of medication for their breathing problems, and also distinguishes between subjects who have been prescribed medication and subjects who self-medicate or use 'over the counter' medication. The interviewer should try not to evoke any guilt in the subject if they are reluctant to take medication so that a false answer is not obtained.

QUESTION 84

The wording of the introductory statement is similar to the ECRHS I but in 84.2 subjects are asked whether attendance was due to asthma, shortness of breath or wheezing.

QUESTION 85

The wording of the introductory statement is similar to ECRHS I but in 85.2 subjects are asked whether attendance was due to asthma, shortness of breath or wheezing.

QUESTION 86

The wording of the introductory statement is similar to ECHRS I but in 86.2 subjects are asked whether attendance was due to asthma, shortness of breath or wheezing.

QUESTION 87

'Regular appointment' means that the subject is seen at specified periods by the health practitioner (i.e. every 3 months, or 4 months etc.) A 'regular appointment' is also one where at the end of a consultation a date is fixed for the next attendance.

Appendix B 2 – ECRHS II Main Questionnaire Instructions and Coding

Standard coding

Area number (as for ECRHS I)

Subject number (as for ECRHS I)

Sample (as for ECRHS I)

For all questions;

1 NO

2 YES

3 DON'T KNOW

Questions with 'TICK ONE BOX ONLY' instruction:
The number of the box ticked is the code for that answer.

General Instructions

8, 98 or 998	NOT CODED (details recorded on questionnaire)
9, 99 or 999	DON'T KNOW (or questions with an answer missing; 'DON'T KNOW' answers without a 'DON'T KNOW' option)

Unanswered boxes in questions 17.1, 57.2, 63.1 and 65.1 (or they may be left blank).

Questions other than the above

QUESTION 14.2 First attack of asthma

00 First attack of asthma as early as they can remember or less than one year old or as a baby

99 Don't know

If a fieldworker has not been able to obtain an accurate answer and recorded "less than 3 years", it should be coded as 2; if they have recorded "3-4 years" then it should be coded 3 (i.e., lower figure is used)

QUESTION 14.3 Most recent attack of asthma

99 Don't know

As in question 14.2, the lower figure should be used as the code.

QUESTION 14.6 and 14.7 Attacks of asthma in the last 12 months.

98 'Maximum' number

99 Don't know

QUESTION 20.1 Medicines

1 Aspirin

2 Beta-blockers

3 Non-steroidal anti-inflammatory agents

4 Mixture of the above

8 not coded (includes allergic reaction to penicillin involving breathing difficulties)

9 not known

QUESTION 21

99 Don't know

QUESTION 25

8 If 8 or more children in the room

QUESTION 32

88 Currently a full-time student

QUESTION 37.3.1 Countries and Territories

001	Afghanistan	063	Dominica
002	Albania	064	Dominican Republic
003	Algeria	065	Ecuador
004	American Samoa	066	Egypt
005	Andorra	067	El Salvador
006	Angola	068	Equatorial Guinea
007	Anguilla	069	Estonia

008	Antarctica (Australian Territory)	070	Ethiopia
009	Antigua & Barbuda	071	Falkland Islands
010	Antilles (Netherlands)	072	Faroe Islands
011	Argentina	073	Fiji
012	Armenia	074	Finland
013	Ascension Island	075	France
014	Australia	076	French Guinea
015	Austria	077	French Polynesia
016	Azerbaijan	078	Gabon
017	Azores	079	Gambia
018	Bahamas	080	Germany (former East)
019	Bahrain	081	Germany (former West)
020	Bangladesh	082	Georgia
021	Barbados	083	Ghana
022	Belgium	084	Gibraltar
023	Belize	085	Greece (Mainland)
024	Benin	086	Greek Islands
025	Bermuda	087	Greenland
026	Bhutan	088	Grenada
027	Bolivia	089	Guadeloupe
028	Botswana	090	Guam
029	Brazil	091	Guatemala
030	British Virgin Island	092	Guinea-Bissau
031	Brunei	093	Guinea
032	Bulgaria	094	Guyana
033	Burkina Faso	095	Haiti
034	Burma	096	Honduras
035	Burundi	097	Hong Kong
036	Byelorussia	098	Hungary
037	Cameroon	099	Iceland
038	Canada	100	India
039	Canary Islands	101	Indonesia
040	Cape Verde	102	Iran
041	Caroline Islands	103	Iraq
042	Cayman Islands	104	Irish Republic

043	Central African Republic	105	Israel and occupied territory
044	Chad	106	Italy (includes Vatican City)
045	Channel Islands	107	Jamaica
046	Chatham Islands	108	Japan
047	Chile	109	Johnston and Sand Island
048	China and Taiwan	110	Jordan
049	Christmas Island	111	Kampuchea (Cambodia)
050	Cocos (Keeling Island)	112	Kazakhstan
051	Colombia	113	Kenya
052	Comoros	114	Kirghizia
053	Congo	115	Kiribati
054	Cook Islands	116	Korea (North)
055	Corsica	117	Korea (South)
056	Costa Rica	118	Kuwait
057	Cote d'Ivoire (Ivory Coast)	119	Laos
058	Cuba	120	Latvia
059	Cyprus	121	Lebanon
060	Czechoslovakia	122	Lesotho
061	Denmark	123	Liberia
062	Dijbout	124	Libya
125	Liechtenstein	188	Saudi Arabia
126	Lithuania	189	Senegal
127	Luxembourg	190	Seychelles
128	Macao	191	Sierra Leone
129	Madagascar	192	Singapore
130	Madeira	193	Solomon Islands
131	Malawi	194	Somalia
132	Malaysia	195	South Africa
133	Maldives	196	Spain
134	Mali	197	Sri Lanka
135	Malta	198	Sudan
136	Marshall Island	199	Suriname
137	Martinique	200	Swaziland
138	Mauritiana	201	Sweden
139	Mauritius	202	Switzerland

140	Mexico	203	Syria
141	Micronesia (Federated States of)	204	Tadzhikistan
142	Midway Islands	205	Tanzania
143	Moldavia	206	Thailand
144	Monaco	207	Togo
145	Mongolia	208	Tonga
146	Monserrat	209	Trinidad and Tobago
147	Morocco	210	Tristan de Cunha
148	Mozambique	211	Tunisia
149	Namibia	212	Turkey
150	Nauru	213	Tukmenistan
151	Nepal	214	Turks and Caicos Island
152	Netherlands	215	Tuvalu
153	New Caledonia	216	Uganda
154	New Zealand	217	Ukraine
155	Nicaragua	218	United Arab Emirates
156	Niger	219	United Kingdom (England IOM)
157	Nigeria	220	United Kingdom (Scotland)
158	Niue Island	221	United Kingdom (Wales)
159	Norfolk Island	222	United Kingdom (N Ireland)
160	North Miriana Island	223	Uruguay
161	Norway	224	USA
162	Oman	225	Uzebikstan
163	Pakistan	226	Vanuatu
164	Palau	227	Venezuela
165	Panama	228	Vietnam
166	Papua New Guinea	229	Virgin Islands of the US
167	Paraguay	230	Wake Island
168	Peru	231	Wallis and Future Island
169	Philippines	232	Western Sahara
170	Pitcairn Islands	233	Western Somoa
171	Poland	234	Yemen Arab Republic
172	Portugal	235	Yemen (Peoples Democratic Republic)
173	Puerto Rico	236	Yugoslavia (Former)
174	Qatar	237	Zaire

175	Reunion	238	Zambia
176	Rodriguez Island	239	Zimbabwe
177	Romania		
178	Russia (see also other States)	998	Not coded
179	Rwanda		
180	St Christopher and Nevis		
181	St Helena and Dependencies		
182	St Lucia		
183	St Pierre and Miquelon		
184	St Vincent and the Grenadines		
185	San Marino		
186	Sao Tome Principe		
187	Sardinia		

QUESTION 39 Other fuels for heating

1	No
3	Distance heaters
4	Electrical radiators (containing heating coils)
5	Closed coal fire
8	Not coded
9	Not known

QUESTION 48 and 51

00	Lowest habitable floor (could be basement or ground floor)
01	Floor above lowest habitable floor
02	Floor, two floors above lowest habitable floor, etc.

03, 04, 05 – etc

Appendix B 2 – ECRHS II Main Questionnaire Instructions and Coding

QUESTION 73.1.1 Food

01	Fruits, fresh/frozen/canned	24	Alcohol (other than red or white wine - see below)
02	Fruits, juice	25	Seafood/shellfish/fish
03	Fruits, dried	26	Eggs
04	Vegetables, fresh/frozen/canned	27	Tea/coffee
05	Vegetables, dried	28	Red meat, fresh
06	Vegetable, pickled	29	Poultry
07	Dairy products (excluding cheese),but including milk/yoghurt/ice-cream	30	Herbs/spices/condiments, including garlic and chilli
08	Chocolate	31	seeds (e.g. sunflower, linseed)
09	Savoury snack foods (e.g. potato crisps, corn chips)	32	High fat foods
10	Confectionery, lollies, liquorice	33	High sugar foods
11	Biscuits/cake, sweet	34	Acidic foods
12	Biscuits/cake, savoury	35	Spicy foods
13	Biscuits/cake, unspecified	36	Artificial colours
14	fats/oils, butter/margarine/cream/salad dressing	37	Preservatives, incl. sulphites
15	Gluten	38	monosodium glutamate (MSG)
16	Wheat products, bread/plain cereal		
17	Mixed cereal products (e.g. muesli)	40	Miscellaneous mixed dishes
18	Soups	41	Soft drinks/cordial
19	Sauces, including tomato paste/seasoning	42	Processed meats, ham, bacon
20	Nuts, including peanut butter/coconut	43	Pastry/pastry dishes
21	Yeast and yeast extracts	50	Cheese

22	Sugar, including golden syrup/jam	60	Indian restaurant, takeaway meal
23	Honey	61	Chinese restaurant, takeaway meal
		62	Burger Meal, restaurant takeaway meal
		63	Other restaurant, take away meal **not elsewhere specified**
		70	White wine
		71	Red wine
		98	Not coded
		99	Not known

QUESTION 75.3

0 less than an hour

If more than 1 hour code as number of hours

QUESTION 76.1 Inhaled short acting beta-2-agonist inhalers

76.1.1 (Which one?)

01	Salbutamol		
02	Terbutaline		
03	Fenoterol		
04	Pirbuterol		
05	Reproterol		
06	Rimiterol		
07	Bitolterol		
08	Hexoprenaline		
09	Carbuterol		
98	Not coded		
99	Not known	-	If compound of B_2 and steroids please enter in question 76.5

QUESTION 76.1.2 (Type of inhaler?)

01 MDI
02 Dry powder
03 Nebuliser
08 Not coded
09 Not known

QUESTION 76.2 Inhaled long acting beta-2-agonist inhalers
76.1.1 (Which one?)

01 Salmeterol
02 Formoterol - If compound of B_2 and steroid please enter in
08 Not coded question 76.5
09 Not known

QUESTION 76.2.2 (Type of inhaler?)

01 MDI
02 Dry powder
03 Nebuliser
08 Not coded
09 Not known

QUESTION 76.3 Inhaled non-specific adrenoreceptor agonist
76.3.1 (Which one?)

01 Adrenaline
02 Isoprenaline
03 Orciprenaline
04 Isoetharine
08 Not coded
09 Not known

QUESTION 76.4 Antimuscarinic inhalers
76.4.1 (Which one?)

01 Ipratropium bromide
02 Atropine (any salt)

03 Oxytropium bromide

08 Not coded

09 Not known

QUESTION 76.4.2 (Type of inhaler?)

01 MDI

02 Dry powder inhaler

03 Nebuliser

08 Not coded

09 Not known

QUESTION 76.5 Inhaled steroids

76.5.1 (Which one?)

01 Beclomethasone diproprionate

02 Betamethasone valerate

03 Budesonide

04 Dexamethasone

05 Flunisolide

06 Triamcinolone

07 Fluticasone

08 Mometasone Furoate

09 Combination Salbutamol and beclamethasone

10 Combination of salmeterol and steroid

11 Symbicort

98 Not coded

99 Not known

QUESTION 76.5.2 (Type of inhaler?)

01 MDI

02 Dry powder inhaler

03 Nebuliser

08 Not coded

09 Not known

QUESTION 76.6 Inhaled cromoglycate/nedocromil

01 Sodium cromoglycate

02 Nedocromil sodium

03 Cromoglycate + beta-agonist

08 Not coded

09 Not known

QUESTION 76.7 Inhaled compounds

76.7.1 (Which one?)

01 Compounds of beta-2-agonists

02 Compounds of non-specific adrenoreceptor agonists (with/without local anaesthetic)

03 Beta-2-agonists with non-specific adrenoreceptor agonists

04 Beta-2-agonists with anti-muscarinics

05 Beta-agonists with steroids

06 Non-specific adrenoreceptor agonists with sodium cromoglycate

07 Beta-agonists with sodium cromoglycate

08 Not coded

09 Not known

QUESTION 76.7.2

01 MDI

02 Dry powder inhaler

03 Nebuliser

08 Not coded

09 Not known

QUESTION 77.1 Oral beta-2-agonists

77.1.1 (Which one?)

01 Salbutamol

02 Terbutaline

03 Fenoterol

04 Pirbuterol

05 Reproterol

06 Bambuterol
07 Tolbuterol
98 Not coded
99 Not known

QUESTION 77.2 Oral methylaxanthines
77.2.1 (Which one?)
01 Aminophylline
02 Choline theophyllinate
03 Theophylline
04 Etophylline
05 Bamifylline
06 Dyprophylline
98 Not coded
99 Not known

QUESTION 77.3 Oral steroids
77.3.1 (Which one?)
Betamethasone
Cortison acetate
Dexamethasone
Fludrocortisone
Hydrocortisone
Methylprednisolone
Prednisolone
Prednisone
Triamcinolone
Cortivazol
Celestamine
Deflazacort
98 Not coded
99 Not known

QUESTION 77.4 Oral anti-leukotrienes

77.4.1 (Which one?)

01 Montelukast

02 Zafirlukast

03 Pranlukast

04 Zileuton

08 Not coded

09 Not known

QUESTION 77.5 Ketotifen

77.5.1

01 Ketotifen

QUESTION 78.2.1

0 If less than a month

QUESTION 78.4

0 If less than a month

QUESTION 80 Injections

80.1 (What injections)

01 Subcutaneous adrenoreceptor agonist self administered

02 Long acting or depot steroid

03 Methylaxanthines

08 Not coded

09 Not known

QUESTION 81 Suppositories

81.1 (What suppositories?)

01 Aminophylline

02 Theophylline

08 Not coded

09 Not known

Appendix B 2 – ECRHS II Main Questionnaire Instructions and Coding

QUESTION 82 Remedies

82.1 (What remedies?)

01 Hypnotherapy

02 Acupuncture

03 Homeopathy (herbal remedies)

04 Diet control

05 Breathing exercises

06 Swimming or other exercises

07 Reflexology

08 Not coded

09 Not known

Appendix B 2 – ECRHS II Main Questionnaire Instructions and Coding

During this questionnaire you will be asked several times about what has happened to you since the last survey.

You took part in the last survey in _____ in _____

This questionnaire has been prepared so that short interviews can be conducted on the phone.

Subjects who are eligible for this questionnaire are subjects who

- *have completed the short stage 1 screening questionnaire*
- *have refused to come to the clinic for any testing*
- *have refused a home visit for the long questionnaire*
- *have refused to complete a longer more detailed questionnaire on the phone*

All subjects who complete this reduced questionnaire should be asked if they are prepared to complete the SF-36 (with the two 'chronic conditions' questions) if it is sent to them.

(Please note that for clarity numbers have been kept the same as in main questionnaire)

Centre number
Personal number
Sample
Date
DAY MONTH YEAR

I AM GOING TO ASK YOU SOME QUESTIONS. AT FIRST THESE WILL BE MOSTLY ABOUT YOUR BREATHING. WHEREVER POSSIBLE, I WOULD LIKE YOU TO ANSWER 'YES' OR 'NO'.

1. Have you had wheezing or whistling in your chest at any time in the last ***12 months***? NO YES

   IF 'NO' GO TO QUESTION 2, IF 'YES':

   1.1 Have you been at all breathless when the wheezing noise was present? NO YES

   1.2. Have you had this wheezing or whistling when you did ***not*** have a cold? NO YES

2. Have you woken up with a feeling of tightness in your chest at any time in the last ***12 months***? NO YES

3. Have you had an attack of shortness of breath that came on during the day when you were at rest at any time in the last ***12 months***? NO YES

4. Have you had an attack of shortness of breath that came on ***following*** strenuous activity at any time in the last ***12 months***? NO YES

5. Have you been woken by an attack of shortness of breath at any time in the last ***12 months?*** NO YES

   IF NO GO TO Q6, IF YES

   5.1 Have you been woken by an attack of shortness of breath in the last ***3 months***? NO YES

   IF NO GO TO Q6, IF YES

   5.1.1 ***On average*** have you been woken by an attack of shortness of breath ***at least once a week*** *in the* ***last 3 months?*** NO YES

   IF NO GO TO Q6, IF YES

   5.1.1.1 How many times a week ***on average*** have you been woken by shortness of breath in the ***last 3 months***? TIMES

6. Have you been woken by an attack of coughing at any time in the ***last 12 months***? NO YES

7. Do you ***usually*** cough first thing in the morning in the winter? NO YES
   [IF DOUBTFUL, USE QUESTION 8.1 TO CONFIRM]

8. Do you ***usually*** cough during the day, or at night, in the winter? NO YES

   IF 'NO' GO TO QUESTION 9, IF 'YES':

   8.1 Do you cough like this on most days for as much as three months each year? NO YES

9. Do you ***usually*** bring up any phlegm from your chest first thing in the morning in the winter? **[IF DOUBTFUL, USE QUESTION 10.1 TO CONFIRM]** NO YES

10. Do you ***usually*** bring up any phlegm from your chest during the day, or at night, in the winter? NO YES

IF 'NO' GO TO QUESTION 14, IF 'YES':

10.1 Do you bring up phlegm like this on most days for as much as three months each year? NO YES

NO YES

14. Have you ever had asthma?

IF 'NO' GO TO QUESTION 15, IF 'YES': NO YES

14.1 Was this confirmed by a doctor?

YEARS

14.2 How old were you when you had your first attack of asthma?

YEARS

14.3 How old were you when you had your most recent attack of asthma?

NO YES

14.5 Have you had an attack of asthma in the last ***12 months***?

IF NO GO TO 14.8, IF YES ATTACKS

14.6 How many attacks of asthma have you had in the last ***12 months***?

ATTACKS

14.7 How many attacks of asthma have you had in the last ***3 months***?

14.8 How many times have you woken up because of your asthma in the last ***3 months?*** TICK ONE BOX ONLY

every night or almost every night	1	
more than once a week, but not most nights	2	
at least twice a month, but not more than once a week	3	
less than twice a month	4	
not at all	5	

14.9. How often have you had trouble with your breathing because of your asthma in the last ***3 months?*** TICK ONE BOX ONLY

continuously	1	
about once a day	2	
at least once a week, but less than once a day	3	
less than once a week	4	
not at all	6	

NO YES

14.10 Are you currently taking any medicines including inhalers, aerosols or tablets for asthma?

NO YES

14.11 Do you have a peak flow meter of your own?

***IF 'NO' GO TO QUESTION 14.12 , IF 'YES'*:**

14.11.1 How often have you used it over the last 3 months? TICK ONE BOX ONLY

never	1	
some of the days	2	
most of the days	3	

14.12 Do you have written instructions from your doctor on how to manage your asthma if it gets worse or if you have an attack? NO YES

NO YES

15. Do you have any nasal allergies, including hay fever?

IF NO GO TO Q16, IF YES YEARS

15.1 How old were you when you first had hay fever or nasal allergy?

16. Have you ever had a problem with sneezing, or a runny or a blocked nose when you did not have a cold or the flu? NO YES

IF NO GO TO Q17, IF YES

16.1.Have you had a problem with sneezing or a runny or a blocked nose when you did not have a cold or the flu ***in the last 12 months***? NO YES

IF NO GO TO Q17, IF YES

16.1.1. Has this nose problem been accompanied by itchy or watery eyes? NO YES

16.1.2. In which months of the year did this nose problem occur? NO YES

January		
February		
March		
April		
May		
June		
July		
August		
September		
October		
November		
December		

17. ***Since the last survey*** have you used any medication to treat nasal disorders? NO YES

NO YES

18. Have you ***ever*** had eczema or any kind of skin allergy?

19. Have you ***ever*** had an itchy rash that was coming and going for at ***least 6 months***? NO YES

IF 'NO' GO TO QUESTION 20, IF 'YES': NO YES

19.1.. Have you had this itchy rash ***in the last 12 months?***

IF 'NO' GO TO QUESTION 20, IF 'YES':

19.1.1. Has this itchy rash ***at any time*** affected any of the following places: the folds of the elbows, behind the knees, in front of the ankles under the buttocks or around the neck, ears or eyes NO YES

NO YES

20. Have you ever had any difficulty with your breathing after taking medicines?

IF 'NO' GO TO QUESTION 21, IF 'YES':

20.1-2 Which medicines? _____ 20.1.1 20.1.2

YEARS

21. How old was your mother when you were born?

NO YES

23. Were you hospitalised before the age of two years for lung disease?

Appendix B 2 – ECRHS II Main Questionnaire Instructions and Coding

I would now like to ask you some questions on the type of jobs that you have done.
I am interested in each one of the jobs that you have done for more than 3 consecutive months since the time we last contacted you (in 1991/2). These jobs may be outside the house or at home, full time or part time, paid or not paid, including self-employment, for example in a family business. Please include part time jobs only if you had been doing them for more than 8 hours per week.

Q26. Are you currently

	TICK ONE BOX ONLY	
Employed (including military service)	1	☐
Self employed	2	☐
Unemployed, looking for work	3	☐
Not working because of poor health	4	☐
Full-time house-person	5	☐
Full time student	6	☐
Retired	7	☐
Other	8	☐

IF EMPLOYED OR SELF EMPLOYED OR A FULL TIME HOURSEPERSON GO TO Q28

27. Have you been employed in any job for three continuous months or longer since the last survey? NO YES

IF YES NOW GO TO QUESTION 28, OCCUPATIONAL MATRIX

YEARS

32. At what age did you complete full time education?

If full time student enter 88

NO YES

37. Do you live in the same home as when you were last surveyed?

58. How often do cars pass your house? TICK ONE BOX ONLY

a) constantly	1	☐
b) frequently	2	☐
c) seldom	3	☐
d) never	4	☐

59. How often do heavy vehicles (e.g. trucks/buses) pass your house? TICK ONE BOX ONLY

a) constantly	1	☐
b) frequently	2	☐
c) seldom	3	☐
d) never	4	☐

NO YES

74. Have you ever smoked for as long as a year?
['YES' means at least 20 packs of cigarettes or 12 oz (360 grams) of tobacco in a lifetime, or at least one cigarette per day or one cigar a week for one year]

IF 'NO' GO TO QUESTION 75, IF 'YES':

YEARS

74.1 How old were you when you started smoking?

NO YES

74.2 Do you **now** smoke, as of ***one month ago***?
***IF 'NO' GO TO QUESTION 74.3, IF 'YES'*:**

74.2.1-4 How much do you ***now*** smoke on average — NUMBER

74.2.1 number of cigarettes per day
74.2.2 number of cigarillos per day
74.2.3 number of cigars a week
74.2.4 pipe tobacco in a) ounces / week
b) grams / week

74.3 Have you stopped or cut down smoking? — NO YES

IF 'NO' GO TO QUESTION 74.4, IF 'YES':

74.3.1 how old were you when you stopped or cut down smoking? — YEARS

74.3.2.1-4 ***on average*** of the entire time you smoked, before you stopped or cut down, how much did you smoke? — NUMBER

74.3.2.1 number of cigarettes per day
74.3.2.2 number of cigarillos per day
74.3.2.3 number of cigars a week
74.3.2.4 pipe tobacco in a) ounces / week
b) grams / week

74.4 Do you or did you inhale the smoke? — NO YES

75. Have you been **regularly** exposed to tobacco smoke in the last ***12 s***? ['Regularly' means on most days or nights] — NO YES

IF 'NO' GO TO QUESTION 76, IF 'YES':

75.1 Not counting yourself, how many people in your household smoke regularly? — NUMBER

75.2 Do people smoke regularly in the room where you work? — NO YES

75.3 How many hours per day are you exposed to ***other people's*** tobacco smoke? — HOURS

75.4 Please provide more information.
How many hours per day, are you exposed to other peoples tobacco smoke in the following locations — NUMBER

at home
at workplace
in bars, restaurants, cinemas or similar social settings
elsewhere

76. Have you used any **inhaled** medicines to help your breathing at any time in the last ***12 months***? — NO YES

IF NO' GO TO QUESTION 77, IF 'YES':

What have you used in the last ***12 months***? _____

INTERVIEWER TO CODE UNDER THE FOLLOWING CATEGORIES

76.1 short acting **beta-2-agonist inhalers** — NO YES
(Please include combinations that include beta 2 and steroids in section 76.5)
76.1.1 If used, which one? _____

76.2 **long acting beta-2-agonist inhalers** — NO YES
(Please include combinations that include beta 2 and steroids in section 76.5)

76.2.1 If used, which one? _____

NO YES

76.3 **non-specific adrenoreceptor agonist inhalers**

76.3.1 If used, which one? _____

NO YES

76.4 **anti-muscarinic inhalers**

76.4.1 If used, which one? _____

NO YES

76.5 **inhaled steroids**
(***if combined B2 and steroid please insert inhaled steroid dose***)

76.5.1 If used, which one? _____

NO YES

76.6 **inhaled cromoglycate/nedocromil**

76.6.1 If used, which one? _____

NO YES

76.7 **inhaled compounds**

76.7.1 If used, which one? _____

77. Have you used any **pills**, **capsules**, **tablets** or **medicines**, other than inhaled medicines, to help your breathing at any time in the last ***12 months***? NO YES

IF 'NO' GO TO QUESTION 78, IF 'YES':
What have you used in the last ***12 months***? _____

INTERVIEWER TO CODE UNDER THE FOLLOWING CATEGORIES

NO YES

77.1 **oral beta-2-agonists**

77.1.1 If used, which one? _____

NO YES

77.2 **oral methylxanthines**

77.2.1 if used, which one? _____

NO YES

77.3 **oral steroids**

77.3.1 If used, which one? _____

NO YES

77.4 oral anti-leukotrienes

77.4.1 If used, which one? _____

NO YES

77.5 **ketotifen**

77.5.1 If used, which one? _____

NO YES

78. Since the last survey have you ever used inhaled steroids?
(GIVE NAMES ON LIST)
IF NO GO TO QUESTION 79 YEARS
78.1. How old were you when you first started to use inhaled steroids?

NO YES

78.2. Have you used inhaled steroids ***every year*** since the last survey?
IF NO GO TO QUESTION 78.3, IF YES MONTHS
78.2.1. On average how many months each year have you taken them?
NOW GO TO Q79 YEARS
78.3 How many of the years since the last survey have you taken inhaled steroids?

MONTHS

78.4. On average how many months of each of these years have you taken them?

NO YES DK

79. Have you been vaccinated for allergy since the last survey?
IF 'NO' OR 'DON'T KNOW' GO TO QUESTION 84, IF 'YES':

NO YES

79.1 Have you been vaccinated for allergy in the last ***12 months***?

84. Since the last survey have you visited a hospital casualty department or emergency room because of breathing problems? NO YES
IF NO GO TO Q85, IF YES
84.1 Have you visited a hospital casualty department or emergency room because of breathing problems in the ***last 12 months***? NO YES
IF NO GO TO 85, IF YES NO YES
84.2 Was this due to asthma, shortness of breath or wheezing?

TIMES

84.2.1 How many times ***in the last 12 months*****?**

85. Since the last survey have you spent a night in hospital because of breathing problems? NO YES
IF NO GO TO Q86 IF YES
85.1 Have you spent a night in hospital because of breathing problems in the ***last 12 months***? NO YES
IF NO GO TO Q86, IF YES NO YES
85.1.1 Was this due to asthma, shortness of breath or wheezing?

85.1.2 How many nights have you spent on each of the following types of ward in ***the last 12 months?*** NUMBER

General	
Chest medicine	
Rehabilitation	
Intensive care unit	
Other	

86. Since the last survey have you been seen by a doctor because of breathing problems or because of shortness of breath? NO YES
IF NO GO TO Q87, IF YES
86.1 Have you been seen by a general practitioner because of breathing problems or shortness of breath in the ***last 12 months?*** NO YES
IF NO GO TO Q86.4, IF YES
86.2.Was this due to asthma, shortness of breath or wheezing?

86.3 How many times have you been seen by your general practitioner because of breathing problems or shortness of breath in each of these locations over the last 12 months?

NUMBER

at home (excluding emergency visits)	
in his surgery	
at home in an emergency	
at another location	

86.4 Have you seen a specialist (chest physician, allergy specialist, internal medicine specialist, ENT doctor) because of your breathing problems or shortness of breath ***in the last 12 months***? NO YES

IF NO GO TO Q87 IF YES

NUMBER

86.4.1 How many times?

87. Are you given regular appointments to be seen by a doctor (or nurse) for your asthma, wheezing or shortness of breath? NO YES

IF NO GO TO Q88 IF YES NO YES

87.1.Are you given regular appointments with a hospital doctor? NO YES

87.2 Are you given regular appointments with your general practitioner? NO YES

87.3. Are you given regular appointments with a nurse? NO YES

90. Are you currently working?

IF NO GO TO Q90.2 IF YES

90.1. How many days of work have you lost because of asthma, shortness of breath or wheezing in the last **12 months?** NUMBER

90.2. Were you forced to **give up working** because of asthma, wheezing or shortness of breath in the last **12 months?** NO YES

IF NO GO TO 91. IF YES

91.2.1. When? DAY MONTH YEAR

91. Gender M F

92. Date of birth DAY MONTH YEAR

INTERVIEW TYPE? TICK ONE BOX ONLY

a) At centre face to face	1
b) At home face to face	2
c) By telephone	3
d) Self completed at home	4

END FIELDWORKER NUMBER

Appendix 2. Questionnaire from the RHINESSA study

Lungehelseundersøkelsens Generasjonsstudie

– translated «The lung health investigation's Generation Study"

Name chosen in order to be as similar as possible to

Airways symptoms and allergic symptoms

1. Have you had wheezing or whistling in your chest at any time **in the last 12 months?** ☐No ☐Yes

If NO go to question 2, if YES:

1.1. Have you been at all breathless when the wheezing noise was present? …… ☐No ☐Yes

1.2. Have you had this wheezing or whistling when you did not have a cold?…… ☐No ☐Yes

2. Have you woken up with a feeling of tightness in your chest at any time **in the last 12 months?** ..... ☐No ☐Yes

3. Have you been woken by an attack of shortness of breath at any time **in the last 12 months?** …….. ☐No ☐Yes

4. Have you been woken by an attack of coughing at any time **in the last 12 months?**… ☐No ☐Yes

5. Have you had an attack of asthma **in the last 12 months**? ..... ☐No ☐Yes

6. Are you currently taking any medicine (including inhalers, aerosols or tablets) for asthma?..... ☐No ☐Yes

7. Do you have any nasal allergies including hay fever?..... ☐No ☐Yes

8. What is your date of birth? (day/month/year) ……. ____dd ____mm_____yyyy

9. What is today's date? (day/month/year) …… ____dd ____mm _____yyyy

10. Gender ☐Man ☐Woman

11. How tall are you? ……. _____cm

12. How much do you weigh? ..... _____kg

13. In recent years, have you been troubled by a protracted cough?..... ☐No ☐Yes

14. Do you usually bring up phlegm or do you have phlegm in your lungs which you have difficulty bringing up? ☐No ☐Yes

If NO to question 13 and 14 go to question 15, if YES:

14.1. Do you cough or bring up phlegm in this way almost every day for at least three months every year?…….. ☐No ☐Yes

14.2. Have you had periods of this kind for at least two years in a row? ☐No ☐Yes

15. Do you have or have you ever had asthma?…….. ☐No ☐Yes

If NO go to question 16, if YES:

15.1. Have you ever had asthma diagnosed by a doctor?…… ☐No ☐Yes

15.2. How old were you when you first experienced asthma symptoms? _____years

15.3. How old were you when you last experienced asthma symptoms?.....……..… _____years

16. Has a doctor ever told you that you have chronic obstructive pulmonary disease (COPD)? ☐No ☐Yes

17. Have you been woken by an attack of shortness of breath at any time in **the last 3 days**? ☐No ☐Yes

18. Have you been woken by an attack of coughing at any time in **the last 3 days**? ☐No ☐Yes

19. Have you had wheezing or whistling in your chest in **the last 3 days**? ☐No ☐Yes

20. Have you **ever** had wheezing or whistling in your chest? ☐No ☐Yes

If NO go to question 21, if YES:

20.1 How old were you when you first noticed wheezing or whistling in your chest? ..... _____years

21. Have you ever experienced nasal symptoms such as nasal congestion, rhinorrhoea (runny nose) and/or sneezing attacks without having a cold? ☐No ☐Yes

If No go to question 22, if YES:

21.1. How old were you when you experienced such nasal symptoms for the first time? _____years

21.2. Have you had such nasal symptoms in **the last 12 months**?…… ☐No ☐Yes

21.3. Has this nose problem been accompanied by itchy or watery eyes? ☐No ☐Yes

21.4. In which months of the year did this nose problem occur?

January / February …… ☐

March / April …..... ☐

May / June…….. ☐

July / August …… ☐

September / October……. ☐

November / December…… ☐

22. Have you ever had eczema or any kind of skin allergy? ☐No ☐Yes

If NO go to question 23,if YES:

22.1. How old were you when you first had eczema or skin allergy? _____years

23. Have you ever had an itchy rash that was coming and going for at least 6 months? ☐No ☐Yes

If NO go to question 24, if YES:

23.1. Have you had this itchy rash in **the last 12 months**? ☐No ☐Yes

23.2. Has this itchy rash at any time affected any of the following places:
the folds of the elbows, behind the knees, in front of the ankles, under the buttocks or around the neck, ears or eyes? ☐No ☐Yes

23.3. Has this itchy rash affected your hands at any time in **the last 12 months?** ☐No ☐Yes

24. Have you ever had an illness or truoble caused by eating **a particular** food or foods? ☐No ☐Yes

If NO go to question 25, if YES:

24.1. Have you nearly always had the same illness or trouble after eating this type of food? ☐No ☐Yes

If NO go to question 25, if YES:

24.2. What type of food was this (*list up to three foods*)?

_____

24.3. Did this illness or truoble include:

24.3.1. a rash or itchy skin? ☐No ☐Yes
24.3.2. diarrhea or vomiting? ☐No ☐Yes
24.3.3. runny or stuffy nose? ☐No ☐Yes
24.3.4. severe headaches? ☐No ☐Yes
24.3.5. breathlessness? ☐No ☐Yes

24.4. How soon after eating this food did/do you get the first sympoms?

Less than half an hour	½ - 1 hour	1-2 hours	2-4 hours	More than 4 hours
☐	☐	☐	☐	☐

24.5. How old were you when you **first** had this attack? _____years

24.6. How old were you when you **last** had this attack? _____years

Smoking habits

25. Do you smoke? (this applies even if you only smoke the odd cigarette/cigar or pipe every week) ☐No ☐Yes

26. Did you smoke previously?..... ☐No ☐Yes

If NO to question 25 and 26 go to question 27, if YES:

26.1. How much do or did you smoke? *(give an average)*

Cigarettes/day	Cigars/week	Pkts pipe tobacco/week

26.2. How old were you when you started smoking? _____years

26.3. For how long have you smoked? (applies to both smokers and ex-smokers) _____years

26.4. If you are an ex- smoker, how old were you when you stopped smoking? _____years

27. Do you use moist snuff, nicotine patches, or other products containing nicotine? ☐No ☐Yes

28. Did you use moist snuff, nicotine patches, or other products containing nicotine previously? ☐No ☐Yes

If NO to question 27 and 28 go to question 30, if YES:

29. What kind of nicotine-containing product do /did you use?

29.1. Moist Snuff ☐No ☐Yes

If you use/have used moist snuff:

29.1.1. How old were you when you started using moist snuff? _____years

29.1.2. For how long have you been using moist snuff? (applies to both current users and past users) _____years

29.1.3. If you did use moist snuff previously, how old were you when you stopped using it? _____years

29.2. Nicotine patches/ gum /tablets ☐No ☐Yes

If you have been using nicotine patches/gum/tablets:

29.2.1. For how long have you used nicotine patches/gum/tablets: _____months

Childhood and family

30. What term best describes **the place you lived most of the time before the age of 5 years**? (tick one box only)

Farm with livestock	Farm without livestock	Village in rural area	Small town	Suburb of city	Inner city
☐	☐	☐	☐	☐	☐

30.1. What term best describes the place **your father** lived as a child? (tick one box only)

Farm with livestock	Farm without livestock	Village in rural area	Small town	Suburb of city	Inner city	Don't know
☐	☐	☐	☐	☐	☐	☐

30.2. What term best describes the place **your mother** lived as a child? (tick one box only)

Farm with livestock	Farm without livestock	Village in rural area	Small town	Suburb of city	Inner city	Don't know
☐	☐	☐	☐	☐	☐	☐

30.3. What term best describes the place **your grandparents'** lived as a child? (tick one box for each grandparent)

	Farm	Village in rural area	Small town	Inner city	Don't know
Father's father	☐	☐	☐	☐	☐
Father's mother	☐	☐	☐	☐	☐
Mother's father	☐	☐	☐	☐	☐
Mother's mother	☐	☐	☐	☐	☐

31. How many persons, including yourself, lived in your home when you were 5 years old (where you lived most of the time)? (number)……

32. Did you have a serious respiratory infection before the age of five years?....☐No ☐ Yes ☐ Don't know

33. Did your father ever smoke regularly during your childhood? ……☐No☐ Yes ☐ Don't know

34. Did your mother ever smoke regularly during your childhood? ……☐No ☐ Yes ☐ Don't know

If NO / DON'T KNOW go to question 35, if YES:

34.1. Did your mother smoke when she was pregnant with you? ☐No ☐ Yes ☐ Don't know

35. Did other people (other than parents) smoke regularly at home during your childhood?..... ☐No ☐ Yes ☐ Don't know

36. How often did you take cod liver oil when you were a child? (tick one box only)

Never	Rarely	Every week	Daily
☐	☐	☐	☐

37. How often did you eat fresh fruits and berries when you were a child? (tick one box only)

Never	Rarely	Every week	Almost daily	Almost daily in the autumn season
☐	☐	☐	☐	☐

38. How often did you eat potatoes or vegetables that **you or your family had cultivated** when you were a child? (tick one box only)

Never	Rarely	Almost weekly in the growing season	Almost daily in the growing season
☐	☐	☐	☐

39. Was there a cat in your home?

39.1. During your first year of life ☐No ☐ Yes ☐ Don't know

39.2. When you were aged 1 to 4 years ☐No ☐ Yes ☐ Don't know

39.3. When you were aged 5- 15 years ☐No ☐ Yes ☐ Don't know

40. Was there a dog in your home?

40.1. During your first year of life ☐ No ☐ Yes ☐ Don't know

40.2. When you were aged 1 to 4 years ☐ No ☐ Yes ☐ Don't know

40.3. When you were aged 5- 15 years ☐ No ☐ Yes ☐ Don't know

41. What was the highest level of education your mother has/had? (tick one box only)

Primary school (up to the minimum school leaving age)..... ☐

Secondary school / technical school (past the minimum age)…..... ☐

College or university..... ☐

42. What was the highest level of education your father has/had? (tick one box only)

Primary school (up to the minimum school leaving age) ..... ☐

Secondary school / technical school (past the minimum age)..... ☐

College or university..... ☐

43. Did your biological parents ever suffer from any of the following:

	Mother (tick box if YES)	Father (tick box if YES)
Asthma	☐	☐
Chronich bronchitis, emphysema and/or COPD	☐	☐
Heart disease	☐	☐
Hypertension	☐	☐
Stroke	☐	☐
Diabetes	☐	☐
Cancer	☐	☐

44. Do you have any biological children? ☐No ☐ Yes

If NONE go to question 45, if you have (had) biological children:

44.1. How many children? NUMBER_____

44.2. Please write the years when your biological children were born, and tick "YES" if they have had any of the following:

	Year of birth	Girl/ boy	Asthma before 10 years	Asthma after 10 years	Hayfever/ Rhinitis	Atopic eczema/ skin allergies
Child 1						
Child 2						
Child 3						
Child 4						
Child 5						
Child 6						

Education and occupation

45. Please mark the educational level which best describes your level: *(tick one box only)*)

Primary school ..... ☐

Secondary school/technical school ..... ☐

College or University ..... ☐

46. Which is your current or most recent work or occupation?

Employed	Self- employed	Homemaker	Student	Unemployed	Other
☐	☐	☐	☐	☐	☐

47. Do you currently have /have you ever had paid work?..... ☐No ☐Yes

*Please **do not** include occupations of shorter duration than three months.*
*Please **do** include part time jobs of 20 or more hours per week.*
If NO go to question 54, if YES:

48. Which is your current or most recent work or occupation? *(please use capital letters)*

.....

48.1. How many years have you worked / did you work in this occupation? ..... years

49. Does being at your current workplace ever cause breathing problems (chest tightness, wheezing, coughing)? ☐No ☐Yes

50. In your current job, are you regularly exposed to vapours, gas, dust or fumes? ☐No ☐Yes

51. Have you ever changed job because the job affected your breathing?..... ☐No ☐Yes

52. Have you ever changed job because of hay fever or nasal symptom?..... ☐No ☐Yes

53. Have you ever changed job because of eczema or skin disease?..... ☐No ☐Yes

In-door environment

54. Do you keep a cat? ☐No ☐Yes

If NO go to 55,if YES:

54.1. Is your cat (are your cats) allowed inside the house? ☐No ☐Yes

54.2. Is your cat (are your cats) allowed in the bedroom? ☐No ☐Yes

55. Do you keep a dog? ☐No ☐Yes

If NO go to question 56, if YES:

55.1. Is your dog (are your dogs) allowed inside the house? ☐No ☐Yes

55.2. Is your dog (are your dogs) allowed in your bedroom? ☐No ☐Yes

56. In which type of accommodation do you live? *(tick one box only)*

Detached house ☐

Semidetached or terraced house ☐

Apartment ☐

Other ☐

57. When did you move to your current home?..... Year_____

58. Have you ever moved house because of breathing problems?..... ☐No ☐Yes

59. When was your present home built?..... Year _____

60. Does tobacco smoking take place in your present home? *(tick one box only)*

Yes, every day	Yes, frequently 1-4 times/week	Yes, sometimes 1-3 times/month	No, never
☐	☐	☐	☐

61. Have any of the following been identified in your home in **the last 12 months:**

61.1. Water leakage or water damage indoors in walls, floor or ceilings?..... ☐No ☐Yes

61.2. Bubbles or yellow discoloration on plastic floor covering, or black discoloration of parquet floor? ..... ☐No ☐Yes

61.3. Visible mould growth indoors on walls, floor or ceilings..... ☐No ☐Yes

62. Have you seen any signs of damp, water leakage or mould in your home at any time **in the last 10 years**? ..... ☐No ☐Yes

63. Have you noticed the odour of mould or mildew (not from food) in your home at any time **in the last 12 months**?..... ☐No ☐Yes

General health

64. Have you had a course of antibiotics in **the last 12 months**?..... ☐No ☐Yes
(*i.e. Apocillin, Azitromax, Imacillin) LIST the three most commonly used antibiotics in your country*

64.1. If YES, how many courses of antibiotics..... (number) _____

65. Have you had a course of antibiotics in **the last 14 days**?..... ☐No ☐Yes

66. Does your gum bleed when you brush your teeth? *(tick one box only)*

Always	Often	Sometimes	Rarely	Never
☐	☐	☐	☐	☐

67. How often do you usually brush your teeth? *(tick one box only)*

2 times/day or more	Once daily	Less than daily
☐	☐	☐

68. How frequently do you exercise? *(give an average, tick one box only)*

Never	Less than once a week	Once a week	2-3 times a week	Almost every day
☐	☐	☐	☐	☐

If you do such exercise as frequently as one or more times a week:

68.1. How hard do you push yourself? *(tick one box only)*

I take it easy without breaking into a sweat or losing my breath ..... ☐

I push myself so hard that I lose my breath and break into a sweat..... ☐

I push myself to near-exhaustion..... ☐

68.2. How long does each session last? *(give an average, tick one box only)*

Less than 15 minutes..... ☐

16-30 minutes ..... ☐

30 minutes to 1 hour..... ☐

More than 1 hour ..... ☐

Sleep and daytime symptoms

69. How often has it occurred **in the last months** *(circle one number for each question):*

1: *Never or almost never*	**2**: *Less than once a week*	**3**: *Once or twice a week*	**4**: *3- 5 nights/days a week*	**5**: *Almost every day or night*

69.1. ... that you snore loudly and disturbingly?.....	1	2	3	4	5
69.2 ... that you have heartburn or belching when you have gone to bed?	1	2	3	4	5
69.3. ... that you have difficulty in getting to sleep at night?...	1	2	3	4	5
69.4. ... that you wake up repeatedly during the night?…….	1	2	3	4	5
69.5. ... that you perspire heavily during the night? ……..	1	2	3	4	5
69.6. ... that you feel drowsy in the daytime?	1	2	3	4	5
69.7 ... that you wake up too early and have difficulty In getting to sleep again?.....	1	2	3	4	5

70. How long time do you usually sleep per night? _____Hours _____Minutes

Other diseases

71. Has a doctor or health professional ever told you that you have?

71.1. Diabetes? ☐No ☐Yes

If NO go to question 71.2, if YES:

71.1.1. How old were you when you were diagnosed with diabetes? _____years

71.1.2. What treatment are you currently using for diabetes? *(tick one box only)*

Insulin	Tablets	Both insulin and tablets	Only diet
☐	☐	☐	☐

71.1.3. Which type of diabetes do/did you have:

☐ Type 1 ☐ Type 2 ☐ Only in pregnancy ☐ Don't know

71.2. Psoriasis? ☐No ☐Yes

If NO go to question 71.3, if YES::

71.2.1. How old were you when you were diagnosed with psoriasis? _____years

71.3. Bechterew's disease? ☐No ☐Yes

If NO go to question 71.4, if YES:

71.3.1. How old were you when you were diagnosed with Bechterew's disease? _____years

71.4. Rheumatiod arthritis? ☐No ☐Yes

If NO go to question 71.5, if YES:

71.4.1. How old were you when you were diagnosed with rheumatoid arthritis? _____years

71.5. Ulcerous Colitis? ☐No ☐Yes

If NO go to question 71.6, if YES:

71.5.1. How old were you when the disease started? ..... _____years

71.6. Crohn's disease? ☐No ☐Yes

If NO go to question 71.7, if YES:

71.6.1. How old were you when the disease started? ..... _____years

71.7. Sleep apnea? ☐No ☐Yes

If NO go to question 71.8, if YES:

71.7.1. How old were you when you were diagnosed with sleep apnea? _____years

71.7.2. What treatment are you currently using for sleep apnea? (more than one box may apply)

CPAP	Oral appliance (bite splint)	Other
☐	☐	☐

71.8. Hypertension (high blood pressure)? ☐No ☐Yes

If NO go to question 71.9, if YES:

71.8.1. How old were you when you were diagnosed with hypertension (high blood pressure)? _____years

71.8.2. Are you currently taking any medication for hypertension (high blood pressure)? ☐No ☐Yes

71.9. Heart infarction or angina pectoris? ☐No ☐Yes

If NO go to question 72, if YES:

71.9.1. Have you ever been treated in hospital because of heart infarction or angina pectoris? ☐No ☐Yes

If NO go to question 72, if YES:

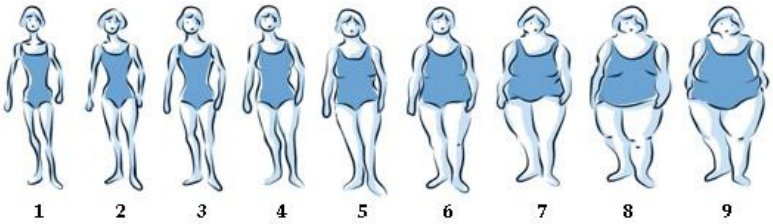
71.9.2. How old were you when you were treated in hospital (for the first time) for heart infarction or angina pectoris? _____years

Body shape

72. Gender: ☐Man ☐Woman

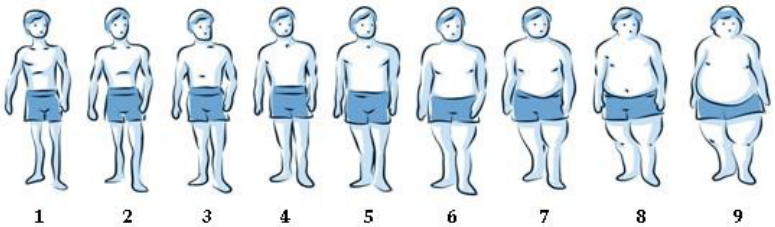
What picture best describes your body shape at each age
(tick one box only for each age/ period you have reached)

72.1. WOMEN



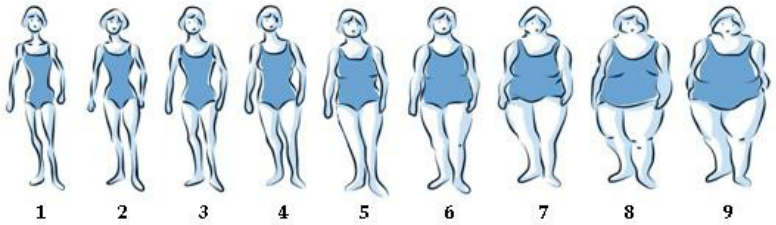
Current	☐	☐	☐	☐	☐	☐	☐	☐	☐
Age 8 years	☐	☐	☐	☐	☐	☐	☐	☐	☐
At first menstruation	☐	☐	☐	☐	☐	☐	☐	☐	☐
Age 20	☐	☐	☐	☐	☐	☐	☐	☐	☐
Age 30	☐	☐	☐	☐	☐	☐	☐	☐	☐
Age 45	☐	☐	☐	☐	☐	☐	☐	☐	☐

72.2. MEN



Current	☐	☐	☐	☐	☐	☐	☐	☐	☐
Age 8 years	☐	☐	☐	☐	☐	☐	☐	☐	☐
At voice break	☐	☐	☐	☐	☐	☐	☐	☐	☐
Age 20	☐	☐	☐	☐	☐	☐	☐	☐	☐
Age 30	☐	☐	☐	☐	☐	☐	☐	☐	☐
Age 45	☐	☐	☐	☐	☐	☐	☐	☐	☐

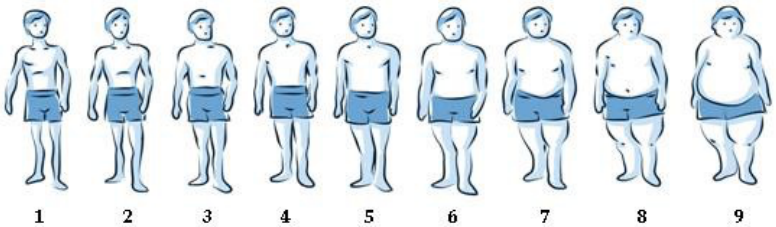
73. What picture best describes the body shape of your biological mother at



Don't know

Age 30	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Age 45	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

74. What picture best the body shape of your biological father at



Don't know

Age 30	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Age 45	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Food and drinks

75. How many days each week do you usually eat/ drink the following:

	Never	Rarely	1 day a week	2 days a week	3 days a week	4 days a week	5 days a week	6 days a week	7 days a week
Meat or sausage	☐	☐	☐	☐	☐	☐	☐	☐	☐
Fish	☐	☐	☐	☐	☐	☐	☐	☐	☐
Vegetables	☐	☐	☐	☐	☐	☐	☐	☐	☐
Raw vegetables, salad, vegetable juice	☐	☐	☐	☐	☐	☐	☐	☐	☐
Potatoes	☐	☐	☐	☐	☐	☐	☐	☐	☐
Potatoes or vegetables you or your family have cultivated yourselves	☐	☐	☐	☐	☐	☐	☐	☐	☐
Olive oil	☐	☐	☐	☐	☐	☐	☐	☐	☐
Citrus fruit or citrus fruit juice	☐	☐	☐	☐	☐	☐	☐	☐	☐
Any fruit (except citrus fruit)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Milk (not including milk you have in tea or coffe)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Soft drinks	☐	☐	☐	☐	☒	☐	☐	☐	☐
Dark (not white) bread	☐	☐	☐	☐	☐	☐	☐	☐	☐
Food heated in plastic container in microwave	☐	☐	☐	☐	☐	☐	☐	☐	☐
Unpasteurized milk?	☐	☐	☐	☐	☐	☐	☐	☐	☐
Beer or wine	☐	☐	☐	☐	☐	☐	☐	☐	☐
Naturally fermented foods	☐	☐	☐	☐	☐	☐	☐	☐	☐

76. To collect data on outdoor exposures in places you have lived, we would like to ask for your address history. Some countries provide address information through registries, others do not.

Which country do you live in?

If you live in NORWAY, SWEDEN, DENMARK or SWITZERLAND:

Your country provides address history through registries.
Thank you for participation in this survey.

If you live in AUSTRALIA, ICELAND, SPAIN, or ESTONIA:

76.1. We would like to know where you have lived since January 1990.
Please give the address, including postcode, of all homes you have lived in **for at least one year since 1990, starting with your current address**

House number	**Street name**	**City**	**Postcode**	**Moved in**	**Lived there until (YEAR)**
					current

Norwegian consent form
To be signed before submitting the *postal* questionnaire

Respondent number

Project title	Project number
The Lung Health Investigation's Generation study	

Project leader

Department/hospital

Participation in the study is voluntary. If you want to participate, you have to sign this consent form. If you agree to participate, you can at any time and without giving a reason, withdraw your consent. Further, this will not have any consequences for your future contact with the health care system.

If you want to withdraw, or have any questions about the study, you can contact the project leader.

I would like to participate in this study

Name in capitals

Date	Signed
____/____/ 20 ____	

Thank you for your help!

Participation in the study is voluntary. If you want to participate, you have to sign this consent form by ticking 'yes' at the bottom of this page. If you agree to participate, you can at any time and without giving a reason, withdraw your consent. Further, this will not have any consequences for your future contact with the health care system.

If you want to withdraw, or have any questions about the study, you can contact the project leader.

I would like to participate in this study:

Appendix 3. Questionnaire from the RHINE study

Kompetansesenter for klinisk forskning
HAUKELAND SYKEHUS

Institutt for indremedisin
Seksjon for lungemedisin

Ernst Omenaas

Airways symptoms

1. Have you had wheezing or whistling in your chest at any time **in the last 12 months?** ☐ No ☐ Yes

   If NO go to question 2, if YES:

   1.1 Have you been at all breathless when the wheezing noise was present? ☐ No ☐ Yes

   1.2 Have you had this wheezing or whistling when you did not have a cold? ☐ No ☐ Yes

2. Have you woken up with a feeling of tightness in your chest at any time **in the last 12 months?** ☐ No ☐ Yes

3. Have you been woken by an attack of shortness of breath at any time **in the last 12 months?** ☐ No ☐ Yes

4. Have you been woken by an attack of coughing at any time **in the last 12 months?** ☐ No ☐ Yes

5. Have you had an attack of asthma **in the last 12 months?** ☐ No ☐ Yes

6. Are you currently taking any medicine *(including inhalers, aerosols or tablets)* for asthma? ☐ No ☐ Yes

7. Do you have any nasal allergies including hay fever? ☐ No ☐ Yes

8. Do you have any nasal allergies including hay fever? ...../...../.....

9. What is today's date? ...../...../.....

10. Are you male or female ☐ Male ☐ Female

11. How tall are you? .....cm

12. How much do you weigh?.....kg

13. In recent years, have you been troubled by a protracted cough? ☐ No ☐ Yes

14. Do you usually bring up phlegm or do you have phlegm in your lungs which you have difficulty bringing up? ☐ No ☐ Yes

    If NO go to question 18, if YES:

15. Do you bring up phlegm in this way almost every day for **at least three months every year?** ☐ No ☐ Yes

    If NO go to question 18, if YES:

16. Have you had periods of this kind for at least two years in a row? NO YES

    If NO go to question 18, if YES:

17. How old were you when these problems began?.....years

Smoking habits

18. Are you a smoker *(this applies even if you only smoke the odd cigarette/cigar or pipe every week)?* ☐ No ☐ Yes

19. Are you an ex-smoker?

If NO to question 18 and 19 go to question 20, if YES:

19.1 Smoke/smoked ..... cigarettes/day

..... cigars/week

..... pkts pipe tobacco/week

How old were you when you started smoking? ..... (age)

Smoked for..... years *(applies to both smokers and ex-smokers)*

Stopped smoking in.....(year)

Upper and lower airways

20. Do you have or have you ever had asthma? ☐ No ☐ Yes

If NO go to question 24, if YES:

21. Have you ever had asthma diagnosed by a doctor? ☐ No ☐ Yes

22. How old were you when you first experienced asthma symptoms?.....years

23. In which year did you last experience asthma symptoms? 19...../ 20.....

24. Has a doctor ever told that you have COPD (BOLD) ☐ No ☐ Yes

25. Have you ever had wheezing or whistling in your chest? ☐ No ☐ Yes

25.1 If "Yes", how old were you when you first noticed wheezing or whistling in your chest? .....years

25.2. If "Yes", when was the last year you noticed wheezing and whistling in your chest? 19...../ 20.....

26. Have you ever experienced nasal symptoms such as nasal congestion, rhinorrhoea (runny nose) and/or sneezing attacks without having a cold? ☐ No ☐ Yes

If NO go to question 25, if YES:

26.1 How old were you when you experienced them for the first time? .....years

26.2 Have you had these kind of nasal symptoms in the last 12 months? ☐ No ☐ Yes

26.3 At which time of the year are your nasal symptoms worst?

Spring	Summer	Autumn	Winter	Always	Don't know
☐	☐	☐	☐	☐	☐

27. Has your nose been blocked **for more than 12 weeks during the last 12 months?** ☐ No ☐ Yes

28. Have you had pain or pressure around the forehead, nose or eyes **for more than 12 weeks during the last 12 months?** ☐ No ☐ Yes

29. Have you had discoloured nasal discharge (snot) or discoloured mucus in the throat **for more than 12 weeks during the last 12 months?** ☐ No ☐ Yes

30. Has your sense of smell been reduced or absent **for more than 12 weeks during the last 12 months?** ☐ No ☐ Yes

In-door and out-door environment

31. In which type of accommodation do you live?

Detached house	Semidetached or terraced house	Apartment	Other
☐	☐	☐	☐

32 When did you move to your current home? 19 .....

33. How many hours per day do you spend in your home most days? Approx. .....hours/day

34. Does tobacco smoking take place in your present home?

Yes every day	Yes, frequently 1-4 times/week	Yes, sometimes 1-3 times/month	No never
☐	☐	☐	☐

35. Have any of the following been identified in your home during **the past 12 months:**

35.1 *Water leakage or water damage indoors in walls, floor or ceilings ☐ No ☐ Yes

35.2 *Bubbles or yellow discoloration on plastic floor covering, or black discoloration of parquet floor ☐ No ☐ Yes

35.3 *Visible mould growth indoors on walls, floor or ceilings. ☐ No ☐ Yes

36. Have you seen any signs of damp, water leakage or mould in your home at any time during the past X years? ☐ No ☐ Yes

37. Have you seen any signs of damp, water leakage or mould in your workplace at any time during the past X years? ☐ No ☐ Yes

38. Is your bedroom window towards a nearby street (<20 m)?

☐ No
☐ Yes a street with little traffic
☐ Yes a street with moderate traffic
☐ Yes a street with much traffic

39. Can you in your bedroom hear traffic noise?

- ☐ Not at all
- ☐ A little
- ☐ Much
- ☐ Very much

40. How much time do you usually spend walking or travelling along streets with busy traffic a typical weekday? Approx..... minutes/day

Marital status

41. What is your marital status? *(more than one alternative may be true)*

- ☐ 1.Single
- ☐ 2 Currently married
- ☐ 3 Cohabitating
- ☐ 4 Separated or divorced
- ☐ 5 Widowed
- ☐ 6 Do not wish to answer

Marital status

42. Please mark the educational level which best describes your level:

- ☐ 1) Primary school
- ☐ 2) Lower or upper secondary school, or technical school
- ☐ 3) College or university

Occupation and work

43. Are you currently working? ☐ No ☐ Yes

44.. Which is your current or most recent work or occupation?

…….....…….....

How many years have you worked or did you work in this occupation? ..... years

45. We assume that your work ability, when it was as best, was 100 percent. How would you rate your current work ability, expressed in percent? ..... %

46. Have you ever changed job because the job affected your breathing? ☐ No ☐ Yes

46.1 If "Yes", in which years? .....

46.2 If "Yes", from which occupation/job did you change? *(could be several)* .....

47. Have you ever changed job because of hayfever or nasal symptom ☐ No ☐ Yes

47.1 If Yes, in which years? .....

47.2 If "Yes", from which occupation/job did you change? *(could be several)* .....

48 Have you ever changed job because of other health problems/diseases? ☐ No ☐ Yes

48.1 If Yes, in which years? .....

48.2 If "Yes", which occupation/job did you change from? *(could be several)* .....

49. Have you ever worked as a painter? ☐ No ☐ Yes

If "Yes", between which years? .....

50. Have you ever worked as a cleaner? ☐ No ☐ Yes

If "Yes", between which years? .....

51. Have you been reporting any days of sick leave during the last 12 months? ☐ No ☐ Yes

51.1 If yes, how many days have you been on sick leave?

☐ 1 – 7 days ☐ 8-30 days ☐ 31 days – 90 days ☐ More than three months

52. Have you been reporting any days of sick leave because of breathing problems during the last 12 months? ☐ No ☐ Yes

52.1 If yes, how many days have you been on sick leave for breathing problems?

1 – 7 days	8-30 days	31 days – 90 days	More than three months
☐	☐	☐	☐

Childhood and family

53. What term best describes the place you lived most of the time when you were under the age of five years?

☐ Farm with livestock
☐ farm without livestock
☐ village in rural area
☐ small town
☐ suburb of city
☐ inner city

54. When you were a child, which of the following were regularly used for heating?

Open wood	Coke or coal fire	Paraffin	Electricity	Gas or oil fired boiler
☐	☐	☐	☐	☐

55. Did you have a serious respiratory infection before the age of five years? ☐ Yes ☐ No ☐ Don't know

56.1. Did your father ever smoke regularly during your childhood? ☐ Yes ☐ No ☐ Don't know

56.2 Did your mother ever smoke regularly during your childhood? ☐ Yes ☐ No ☐ Don't know

56.3 Did other people (other than parents) smoke regularly at home during your childhood? ☐ Yes ☐ No ☐ Don't know

57. When you were a child, how often did you eat fresh fruits?

Never	Rarely	Every week	Almost daily	Almost daily in the autumn season
☐	☐	☐	☐	☐

58. Did your biological parents ever suffer from any of the following:

	Mother (yes)	Father (yes)
Asthma	☐	☐
Chronich bronchitis, emphysema and/or COPD	☐	☐
Heart disease	☐	☐
Hypertension	☐	☐
Stroke	☐	☐
Diabetes	☐	☐
Cancer	☐	☐

59. Do you have children (including grown-up children)? □ No □ Yes

If yes, how many? .....children

Please write the years when your children were born, and tick "yes" if they have had any of the following:

Child	Birth year of child (year)	Asthma before 10 year (yes)	Asthma after 10 years (yes)	Hayfever/ rhinitis (yes)	Atopic eczema/Skin allergies (yes)
1					
2					
3					
4					
5					
6					
7					

Sleep and daytime symptoms

The numbers mean
1: Never or almost never
2: Less than once a week
3: once or twice a week
4: 3- 5 nights/days a week
5: Almost every day or night

How often has it occurred in the last months:

60. that You snore loudly and disturbingly?	1	2	3	4	5
61. that You have heartburn or belching when you have gone to bed?	1	2	3	4	5
62. that You have difficulty in getting to sleep at night?	1	2	3	4	5
63. that You wake up repeatedly during the night?	1	2	3	4	5
64. that You perspire heavily during the night?	1	2	3	4	5
65. that You feel drowsy in the daytime?	1	2	3	4	5
66. that You wake up too early and have difficulty in getting to sleep again?	1	2	3	4	5

67. Have you ever had sleep apnoea diagnosed by a doctor? ☐ No ☐ Yes

If "No" go to question 69, if "Yes":

67.1 What year did you get the diagnosis of sleep apnoea? Year .....

67.2 If you are currently treated for sleep apnoea, what treatment do you have?

☐ CPAP

☐ Oral appliance (bite splint)

☐ Previous surgery in the throat or nose

☐ Others

68 How long time do you usually sleep per night?

I usually sleep .....hours and.....minutes.

Other diseases

69. Have ever had hypertension (high blood pressure) diagnosed by a doctor? ☐ No ☐ Yes

If yes:

69.1 When did you get the diagnosis hypertension (high blood pressure)? Year .....

69.2 Are you currently taking any medication for hypertension (high blood pressure)? ☐ No ☐ Yes

70. Have you ever had stroke? ☐ No ☐ Yes

70.1 If you have had stroke, in which year was it? Year .....

71. Have you ever been treated in hospital because of heart infarction or angina pectoris? ☐ No ☐ Yes

If yes:

71.1 When were you treated (for the first time) at a hospital because of heart infarction or angina pectoris? Year .....

72. Have you ever had diabetes diagnosed by a doctor? ☐ No ☐ Yes

If yes:

72.1 What year did you get the diagnosis diabetes? Year: .....

72.2 What treatment are you currently using for diabetes?

- ☐ Insulin
- ☐ Tablets
- ☐ Both insulin and tablets
- ☐ Only diet

73. Do you have or have you ever had ulcerative collitis? ☐ No ☐ Yes

73.1 If yes: how old were you when the disease started? ..... years

74. Do you have or have you ever had Crohn's disease? ☐ No ☐ Yes

74.1 If, yes, how old were you when the disease started? ..... years

General health

75 Does your gum bleed when you brush your teeth?

- ☐ Always
- ☐ Often
- ☐ Sometimes
- ☐ Rarely
- ☐ Never

76 How often do you usually brush your teeth?

- ☐ 2 times/day or more
- ☐ Once daily
- ☐ Less than daily

77. How frequently do you exercise? (Give an average)

Never	Less than once a week	Once a week	2-3 times a week	Almost every day
☐	☐	☐	☐	☐

77.1. If you do such exercise as frequently as once or more times a week: How hard do you push yourself? *(Give an average)*

- ☐ I take it easy without breaking into a sweat or losing my breath /
- ☐ I push myself so hard that I lose my breath and break into a sweat /
- ☐ I push myself to near-exhaustion

77.2. How long does each session last? (Give an average)

Less than 15 minutes	6-30 minutes	30 minutes to 1 hour	More than 1 hour
☐	☐	☐	☐

78. Body silhouettes

Information and contact conscent

In case we need to get in touch with you again please write your telephone number below

Telephone number: Daytime .....

Evening .....

THANK YOU FOR YOUR HELP

10. Papers I-III

Paper I

I

Environmental Epigenetics, 2019, 1–10

doi: 10.1093/eep/dvz023
Research article

RESEARCH ARTICLE

Epigenome-wide association of father's smoking with offspring DNA methylation: a hypothesis-generating study

G.T. Mørkve Knudsen[1,2,*,†], F.I. Rezwan[3,†], A. Johannessen[2,4], S.M. Skulstad[2], R.J. Bertelsen[1], F.G. Real[1], S. Krauss-Etschmann[5,6], V. Patil[7], D. Jarvis[8], S.H. Arshad[9,10], J.W. Holloway[3,‡] and C. Svanes[2,4,‡]

[1]Department of Clinical Science, University of Bergen, N-5021 Bergen, Norway; [2]Department of Occupational Medicine, Haukeland University Hospital, N-5021 Bergen, Norway; [3]Human Genetics and Genomic Medicine, Human Development and Health, Faculty of Medicine, University of Southampton, Southampton SO16 6YD, UK; [4]Department of Global Public Health and Primary Care, Centre for International Health, University of Bergen, N-5018 Bergen, Norway; [5]Division of Experimental Asthma Research, Research Center Borstel, 23845 Borstel, Germany; [6]German Center for Lung Research (DZL) and Institute of Experimental Medicine, Christian-Albrechts University of Kiel, 24118 Kiel, Germany; [7]David Hide Asthma and Allergy Research Centre, St. Mary's Hospital, Isle of Wight PO30 5TG, UK; [8]Faculty of Medicine, National Heart & Lung Institute, Imperial College, London SW3 6LY, UK; [9]Clinical and Experimental Sciences, University of Southampton, Southampton General Hospital, Southampton SO16 6YD, UK; [10]NIHR Respiratory Biomedical Research Unit, University Hospital Southampton, Southampton SO16 6YD, UK

*Correspondence address. Haukanesvegen 260, N-5650 Tysse, Norway; Tel: þ47 977 98 147; E-mail: Gerd.Knudsen@uib.no and torilmknudsen@gmail.com

[†]Equal first authors.
[‡]Equal last authors.
Managing Editor: Moshe Szyf

Abstract

Epidemiological studies suggest that father's smoking might influence their future children's health, but few studies have addressed whether paternal line effects might be related to altered DNA methylation patterns in the offspring. To investigate a potential association between fathers' smoking exposures and offspring DNA methylation using epigenome-wide association studies. We used data from 193 males and females (11–34 years) participating in two population-based cohorts. DNA methylation was quantified in whole blood using Illumina Infinium MethylationEPIC Beadchip. Comb-p was used to analyse differentially methylated regions (DMRs). Robust multivariate linear models, adjusted for personal/maternal smoking and cell-type proportion, were used to analyse offspring differentially associated probes (DMPs) related to paternal smoking. In sensitivity analyses, we adjusted for socio-economic position and clustering by family. Adjustment for inflation was based on estimation of the empirical null distribution in BACON. Enrichment and pathway analyses were performed on genes annotated

Received 18 February 2019; revised 22 August 2019; accepted 4 November 2019

to cytosine-phosphate-guanine (CpG) sites using the gometh function in missMethyl. We identified six significant DMRs (Sidak-corrected *P* values: 0.0006–0.0173), associated with paternal smoking, annotated to genes involved in innate and adaptive immunity, fatty acid synthesis, development and function of neuronal systems and cellular processes. DMP analysis identified 33 CpGs [false discovery rate (FDR) < 0.05]. Following adjustment for genomic control (k ¼ 1.462), no DMPs remained epigenome-wide significant (FDR < 0.05). This hypothesis-generating study found that fathers' smoking was associated with differential methylation in their adolescent and adult offspring. Future studies are needed to explore the intriguing hypothesis that fathers' exposures might persistently modify their future offspring's epigenome.

Key words: EWAS; population cohorts; paternal smoking exposure; offspring DNA methylation

Introduction

It has been increasingly acknowledged that environmental conditions during *in utero* development and early life may contribute to later onset health and disease. Evolving evidence suggests that paternal line exposures can also affect offspring health (1–6). In particular, recent epidemiological reports have demonstrated that fathers' smoking is associated with an increased asthma risk and adiposity (7, 8) in their children.

Efforts in identifying biochemical mechanisms underlying such altered phenotypes have suggested epigenetic regulatory systems as a possible mechanistic link between environmental exposures and disease risk (9). Epigenetic processes propagate regulatory information through mitosis essential for normal cell tissue function and development (10). However, the epigenome also displays a high degree of structural adaption, and is determined by the combined response to both environmental and genetic factors (11). The plasticity of these systems is important as they affect gene transcriptional activity and lead to long-lasting phenotypic changes in a disease-related manner that may also persist through meiosis, i.e. between generations.

There is clear evidence for altered epigenetic programming in response to tobacco smoke exposure, and several genome-wide studies have identified associations between personal smoking and changes in DNA methylation at single cytosine-phosphate-guanine (CpG) sites in whole blood or isolated peripheral blood mononuclear cells (12–15). Methylation differences in cord blood of offspring born to smoking mothers have also been reported (16–18), and such differences have been shown to persist until adulthood (19, 20). However, to our knowledge, evidence for a persistent methylation effect in offspring due to paternal tobacco use has yet to be demonstrated.

As DNA methylation can be stably propagated through mitotic and possibly meiotic cell divisions (10, 11), it seems theoretically plausible that offspring DNA methylation might be persistently influenced by paternal smoking exposure. We hereby present a hypothesis-generating analysis of a relatively small number of persons, with the aim to investigate the association between paternal smoking and genomic methylation patterns in offspring, and to explore potential biological impact of methylated regions and annotated genes.

Results

Characteristics of the study populations are presented in Table 1. There was an equal gender distribution in both cohorts, with mean age of 26 and 44 years for RHINESSA and European Community Respiratory Health Survey (ECRHS), respectively. A substantial proportion of the subjects had fathers that smoked during their childhood (66%), for RHINESSA participants this was due to enrichment of samples from persons with smoking fathers for DNA methylation.

Differentially Methylated Region Analysis

Analysis of differentially methylated regions (DMRs) using comb-p identified six significant DMRs (Sidak-corrected *P* values: 0.0006–0.0173) (Table 2). Among these DMRs, spanning between 3 and 5 DNA methylation sites, five were mapped to known genes. Two of the annotated genes were related to innate immune system pathways (*ATP6V1E1*, *C2*), whereas one

Table 1: characteristics of study participants by cohort, RHINESSA (*n* ¼ 95), and ECRHS2 (*n* ¼ 100)

Descriptive variables	RHINESSA *N* ¼ 95	ECRHS *N* ¼ 100	*P*-value[a]
Sex, *n* (%)			
Male	46 (48)	44 (44)	0.63
Female	49 (52)	56 (56)	
Age, mean 6 SD	26 6 7.5	44 6 6.2	<0.001
Range	11–45	31–54	
Education, *n* (%)			
Primary	5 (5)	10 (10)	0.52
Secondary	33 (35)	37 (37)	
College/university	51 (54)	53 (53)	
Smoke status, *n* (%)			
Never	68 (72)	41 (41)	<0.001
Ex	13 (14)	29 (29)	
Current	14 (15)	30 (30)	
Pack years, median (range)	2 (0–23)	8 (1–37)	<0.001
Childhood smoke exposure, *n* (%)			
Father smoked[b]	66 (69)	63 (63)	0.67
Mother smoked	31 (33)	31 (31)	0.56
Father and mother smoked	31 (33)	24 (24)	0.44
No parent smoked	25 (26)	28 (28)	0.44
Father education, *n* (%)[c]			
Primary	10 (11)	46 (46)	<0.001
Secondary	38 (40)	22 (22)	
College/university	45 (47)	25 (25)	
Mother education, *n* (%)[c]			
Primary	11 (12)	62 (62)	
Secondary	30 (32)	24 (24)	
College/university	43 (45)	7 (7)	
Father age childbirth, mean 6SD[d] range	31 6 5.8 20–54	32 6 6.5 20–58	0.69

[a]Chi square test for categorical variables; *t*-test for continuous (norm. distributed); Wald test for continuous (non-norm. distributed).
[b]RHINESSA sample included 23 persons with father smoking starting <age 15 years, 43 with father smoking starting >15 years and smoking for at least 4 years before conception of offspring, and 29 with non-smoking fathers/mothers.
[c]Missing RHINESSA; Educ. 6 (6%); father educ. 2 (2%); mother educ. 6 (6%); ECRHS; father educ./mother educ. 7 (7%).
[d]Father's age in ECRHS obtained from registry data.

Table 2: statistically significant DMRs (Sidak $P < 0.05$) as associated with father's smoking

Location	No. probes	Slk[a] *P*-value	Sidak[a] *P*-value	Ref gene name and feature	CpG feature
Chr22:18111277-18111521	4	6.01E-07	0.0019	ATP6V1E1 Intron, 5⁰UTR, cds	Island
Chr6:31865522-31865866	5	2.49E-06	0.0055	C2 TSS, intron, exon, 5⁰UTR	Shore
Chr2:80752765-80752967	4	1.69E-06	0.0006	CTNNA2 intron	NA
Chr16:89180587-89180843	3	5.83E-06	0.0173	ACSF3 intron, cds, nc_intron, nc_exon, nc_intron	NA
Chr1:182669050-182669315	3	6.67E-07	0.0019	LINCO1688 intergenic	NA
Chr7:158766826-158767135	3	5.24E-06	0.0129	WDR60 intergenic	Island

[a]Both Slk, uncorrected Stouffer-Liptak-Kechris *P* values, and Sidak *P* values corrected for multiple testing are reported.
5⁰UTR , 5 prime untranslated region; cds, coding sequence; TSS, transcription start site; nc_intron, non-coding intron, nc_exon, non-coding exon.

Table 3: characteristics of DMRs

Genes annotated to DMRs	Putative gene function	Related pathways
ATP6V1E1 (ATPase H þ transporting V1 subunit E1)	Encodes component of vacuolar ATPase (V-ATPase) that mediates acidification of intracellular compartments in eukaryotic cells necessary for variety of intracellular processes (32, 66, 67)	Innate immune system Synaptic vesicle cycle
C2 (complement C2)	Serum glycoprotein part of pathway of the complement system responsible for regulating immune responses (33, 68)	Innate immune system Complement pathway
CTNNA2 (catenin alpha 2)	Involved in regulating cell–cell adhesion and differentiation in the nervous system. Essential for proper regulation of cortical neuronal migration and neurite growth (34, 69)	Blood–brain barrier and immune cell transmigration Sertoli–sertoli cell junction dynamics
ACSF3 (acyl-CoA synthetase family member 3)	Catalyzes initial reaction in mitochondrial fatty acid synthesis (70)	Regulation of lipid metabolism by peroxisome proliferator-activated receptor alpha (PPARalpha) Fatty acid biosynthesis
Linc01688 (long intergenic non-protein coding RNA 1688)	Unknown	
WDR60 (Wd repeat domain 60)	Encodes a member of the WD repeat protein family. Involved in variety of cellular processes including cell cycle progression, signal transduction, apoptosis, and gene regulation (71)	Organelle biogenesis and maintenance Intraflagellar transport

DMR was involved in lipid metabolism regulation and fatty acid biosynthesis (*ACSF3*). One DMR overlapped with the catenin alpha 2 gene (*CTNNA2*), which are related to development of the nervous system. One DMR mapped to the WD repeat domain 60 gene (*WDR60*), which regulates a variety of cellular processes including cell cycle progression, signal transduction, and gene regulation (Table 3).

Differentially Mediated Probe Analysis

Epigenome-wide association between father's smoking and offspring DNA methylation at a single probe level identified 33 CpGs that passed epigenome-wide significance at a FDR rate $P < 0.05$ (Fig. 1). However, the EWAS exhibited a genomic inflation factor (lambda) of 1.462 (Supplementary Fig. S1). After applying correction for genomic inflation using the BACON method, epigenome-wide association between father's smoking and offspring DNA methylation identified 37 significantly differentially methylated CpG sites (inflation-adjusted *P*-value <0.0001) (Supplementary Figs S2 and S3). After subsequent filtering of data and removal of CpG sites having SNPs within the region of 650 bp of the CpG, and with minor allele frequency 2:0.05, we retained 32 differentially mediated probes (DMPs) with differential methylation between exposure groups for enrichment analysis (Supplementary Table S1). The top 10 DMPs are presented in Table 4. Among these, four were related to innate and adaptive immunity and various immune cell subsets (*BCAS1*, *MFGE8*, *UNC93B1*, and *RALB*) (21–24). Another DMP (*DLGAP1*) was related to neuronal systems and behavioural disorders (25).

Enrichment Analysis

Enrichment analysis of the 32 DMPs (Supplementary Table S1) using Enrichr for transcription factor-binding sites identified by the Encyclopedia of DNA elements (ENCODE) and Epigenomic roadmap project did not identify significant enrichment in regulatory regions (Supplementary Tables S2–S4 and Figs S4–S6).

Analyses using ontologies defined in the KEGG (Kyoto Encyclopedia of Genes and Genomes) and GO (Gene Ontology) databases retrieved pathways and terms, and although not statistically significant, results from top 10 KEGG pathways showed enrichment of addiction behaviours (nicotine addiction). Summary statistics of top 10 GO and KEGG enrichment results are shown in Tables 5 and 6, respectively.

Sensitivity Analyses

To address the issue of relatedness among some of the participants (siblings in RHINESSA, $n = 44$), we performed linear mixed

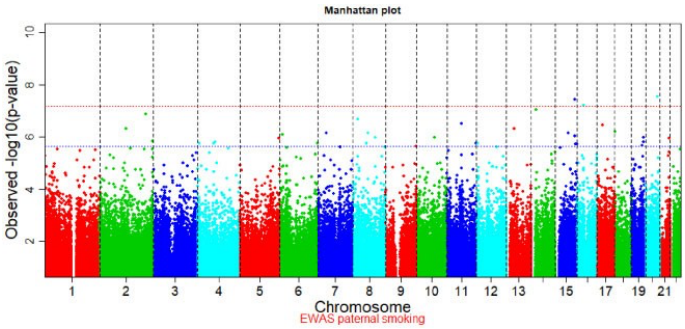


Figure 1: Manhattan plot for paternal smoking EWAS (before adjusted for genomic inflation). In the plot, the vertical axis indicates (−log10 transformed) observed *P* values, and the horizontal axis indicates chromosome positions with the points indicating individual CpG. Red line: Bonferroni threshold and blue line: Multiple testing correction threshold (FDR < 0.05)

Table 4: differentially methylated probe analysis (corrected *P*-value <0.00001)

PROBEID	BETA	SE	*P*-value	Adj *P*-value	CHR	MAPINFO	Gene
cg05019203	−0.018	0.003	2.83E-08	4.40E-06	20	52612962	BCAS1
cg25727029	0.013	0.002	3.56E-08	5.16E-06	15	89482453	MFGE8
cg00626693	−0.014	0.003	6.27E-08	7.64E-06	16	30622810	ZNF689
cg19754387	0.006	0.001	1.33E-07	1.29E-05	2	208576057	CCNYL1
cg24534854	−0.013	0.003	2.09E-07	1.76E-05	8	22582613	PEBP4
cg20272935	0.024	0.005	3.02E-07	2.27E-05	11	67765720	UNC93B1
cg04164584	−0.010	0.002	3.44E-07	2.49E-05	17	27235821	PHF12
cg06876354	0.017	0.003	4.65E-07	3.07E-05	2	121020189	RALB
cg25012097	−0.012	0.002	4.74E-07	3.11E-05	13	39263863	FREM2
cg07217718	0.025	0.005	6.17E-07	3.73E-05	18	3585484	DLGAP1

PROBEID, probe identifiers; BETA, estimates; SE, standard error; Adj *P*-value, *P*-value adjusted by multiple test correction; CHR, chromosome; MAPINFO, position of the CpGs in the chromosome; Gene, UCSC RefGene.

Table 5: top 10 enriched pathways in GO molecular function, biological processes, and cell compartment identified using genes CpGs (threshold: inflation-adjusted *P*-value <0.0001)

Ontology and term[a]		ID	CpGs in tern	Meth CpGs	*P*-value
MF	Selenomethionine adenosyltransferase activity	GO:0098601	1	1	<0.001
MF	Methionine adenosyltransferase activity	GO: 0004478	2	1	0.001
MF	Extracellularly glutamate-gated chloride channel activity	GO:0008068	1	1	0.002
BP	Regulation of exocyst assembly	GO:0001928	1	1	0.002
BP	Regulation of exocyst localization	GO:0060178	1	1	0.002
CC	Excitatory synapse	GO:0060076	48	2	0.002
BP	S-adenosylmethionine biosynthetic process	GO:0006556	3	1	0.002
BP	Sequestering of neurotransmitter	GO:0042137	2	1	0.003
BP	Synaptic vesicle lumen acidification	GO:0097401	2	1	0.003

[a]Ontology: BP, biological process; CC, cell compartment; MF, molecular function; ID, GO identifier; CpG in term, number of CpGs in GO term; Meth.CpGs, number of significant CpGs.

models on the 32 significant (inflation-adjusted *P*-value <0.0001) CpG sites, where family ID was included as random effect. All 32 CpGs were sustained in these analyses (Supplementary Table S5).

To account for potential confounding by social class, we conducted a sensitivity analysis adjusting for paternal socio-economic background by adding education as a proxy for socio-economic status to the regression model. Methylation at all the selected CpG sites (inflation-adjusted *P*-value <0.0001) was still associated with paternal smoking in this analysis (Supplementary Table S6).

Replication Analysis

Due to the amount of missing CpG sites between the EPIC and the 450 K microarray, we could not pursue replication of the significant DMRs identified in the DMR analysis. We undertook replication of the selected CpG sites (inflation-adjusted *P*-value <0.0001) in a subsample from Isle of Wight (IoW) with available data from cord blood DNA samples (N ¼ 159, study characteristics presented in Supplementary Table S7). However, due to different methylation array platforms, and because some CpGs were discarded by pre-processing, only 13 out of the 32 CpGs

Table 6: top 10 enriched pathways in KEGG using genes CpGs (threshold: inflation-adjusted *P*-value <0.0001)

KEGG	Pathway	ID	CpGs in path	Meth. CpGs	*P*-value
KEGG	ECM–receptor interaction	path:hsa04512	86	2	0.006
KEGG	Glutamatergic synapse	path:hsa04724	114	2	0.011
KEGG	Nicotine addiction	path:hsa05033	40	1	0.047
KEGG	Cysteine and methionine metabolism	path:hsa00270	48	1	0.049
KEGG	Biosynthesis of amino acids	path:hsa01230	74	1	0.063
KEGG	Synaptic vesicle cycle	path:hsa04721	78	1	0.093
KEGG	Pancreatic cancer	path:hsa05212	75	1	0.095
KEGG	Colorectal cancer	path:hsa05210	86	1	0.104
KEGG	Retrograde endocannabinoid signalling	path:hsa04723	141	1	0.149
KEGG	Cytokine–cytokine receptor interaction	path:hsa04060	289	1	0.167

Pathway, KEGG pathway; ID, pathway identifier; CpG in path, number of CpGs in pathway; Meth.CpGs, number of significant CpGs.

identified in the ECRHS/RHINESSA cohort were available for replication in the IoW cohort (Supplementary Table S7).

Discussion

In the present study, we have measured epigenome-wide CpG site-specific DNA methylation in adolescent and adult offspring and identified six significant DMRs (Sidak-corrected *P* values 0.0006–0.0173) related to father's smoking. To our knowledge, this is the first study suggesting persisting effects of paternal smoking on offspring DNA methylation. Although previous genome-wide associations of maternal smoking suggest that associations with DNA methylation changes in offspring tend to weaken with increasing age of the offspring (26), and our study subjects will have accumulated a range of exposures influencing DNA methylation, it is remarkable that we were able to detect methylation differences associated with paternal smoking in persons aged 11–54 years.

Of the six statistically significant DMRs identified, one region overlapped with intron 11 within the catenin alpha-2 (*CTNNA2*) gene. *CTNNA2* has previously been shown to be differentially methylated in relation to smoking (18, 27, 28). It is expressed across the central nervous system and suggested involved in behavioural dysfunction and addiction (29). Although it did not harbour a CpG island, which would have provided additional support for a regulatory role for this region, DNA methylation at intronic sequences outside CpG islands may also be of functional important (30). Two DMRs (*ATP6V1E* and *WDR60*), co-localized with CpG islands, and the region within ATP6V1E1 covered parts of the 5 prime untranslated region (5^0UTR) and the coding sequence of the gene. One DMR, annotated to the *C2* gene on chromosome 6, was located to a CpG island shore (regions within 2000 bp of a CpG island), and overlapped with the transcription start site (TSS) as well as the 5^0UTR and exon 1 of *C2*. Although this indicates regulatory functions of the DMRs, they consist of CpGs of only nominal significance and differential methylation could reflect irregular spacing of probes and should be interpreted with caution as they may introduce false-positive results.

When exploring the biological impact of annotated genes, there were similar patterns in the DMR and DMP analyses, although the identified DMPs did not remain significant at epigenome-wide levels of significance. Two of the significant DMRs (*ATP6V1E1* and *C2*) and four of the top DMPs (*BCAS1*, *MFGE8*, *UNC93B1*, and *RALB*) were annotated to genes related to innate and adaptive immunity and to different immune cell subsets (21–24, 31, 32). Furthermore, one DMR (*CTNNA2*) and one DMP (*DLGAP1*) mapped to genes involved in function and development of neuronal systems (25, 33), and to behavioural dysfunction (29, 34, 35).

Except *CTNNA2* (18), none of our significant DMRs or top DMPs are previously reported in epigenome-wide studies of the effect of maternal smoking (16, 17, 36–39), or current or lifetime personal smoking exposure (12–15, 40–42). This is also in agreement with Joubert *et al.* who demonstrated that the CpGs differentially methylated in relation to maternal smoking were not associated with paternal smoking (43). Given the differences in gamete development in males and females, it seems biologically plausible that exposure effects through the maternal and paternal line may differ and induce epigenetic modifications at different loci. Further, it seems plausible that effects transmitted across generations may differ from those of personal smoking. To investigate whether the DMP-specific DNA methylation differences were driven by relatedness among participants, we conducted a sensitivity analysis accounting for family. All the top DMPs remained suggesting that our findings were not due to residual confounding by genetic or family-related environmental factors.

There is increasing evidence of shared pathophysiology between nicotine dependence and neuropsychiatric disorders (44), and smoking has been reported to modify genes that predispose to addictive behaviours (27, 45). In previous literature, maternal smoking during pregnancy has been associated with adverse neurodevelopmental outcome (46) and behavioural alterations in offspring (20, 47). Enrichment analysis of the top 32 differentially methylated probes (adj. $P < 0.00001$) identified GO terms and KEGG pathways involved in developmental and regulatory processes of the brain and the central nervous system and nicotine addiction, suggesting that paternal smoking may also induce aberrant methylation in genes related to neurodevelopment. However, as the identified CpGs did not remain significant epigenome-wide after adjustment for inflation, results from the KEGG and GO enrichment analysis should be interpreted with caution and may not be valid.

When we explored the biological and regulatory role of differentially methylated loci by investigating ENCODE and Epigenomic roadmaps annotated regulatory domains, we found no significant enrichment for histone modification signatures and transcription factor sites among our significant CpG sites. Whether the detected methylation differences can introduce functional changes at the gene transcriptional level needs further investigation.

The present study cannot differentiate whether the observed association of father's smoking with offspring DNA methylation may be due to second-hand smoke exposure during the

gestational period and/or childhood (post-conception) or due to altered sperm DNA methylation patterns transmitted to the offspring (pre-conception). A pre-conception effect is suggested by previous studies showing that the strongest effect of father's smoking on offspring phenotype was observed when smoking occurred before conception and particularly at an early age (7, 8, 48). However, further studies with detailed information about exposure onset in large samples will be required to address this.

The identified DMPs associated with father's smoking showed relatively small effect estimates, with top 10 CpG beta values relative to offspring of smoking and non-smoking fathers ranging from -0.02 to 0.03. This is in line with previous findings where DNA methylation differences associated with environmental exposures are characterized by small changes on the scale of 2–10% (30, 49). However, previous studies have demonstrated that even small changes can impact transcriptional activity and be consistent in different populations and across age groups (17, 49). Although associations with *in utero* maternal smoking have shown higher estimates, ranging from −0.28 to 0.18 (16, 18, 26), we would expect DNA methylation changes related to paternal exposures to be subtler when compared to direct effects from placenta–foetus interactions. Further, smaller effect estimates could be expected considering that we analyzed associations of father's smoking with DNA methylation in adolescents and adults. The fact that we found epigenomic regions (DMRs) associated with paternal smoking, adds functional relevance to our discoveries, as it implies differential methylation in regions that may affect regulation of transcription. These regional changes are also more robust as they are less prone to SNP effects and risk of false-positive findings as compared to site-by-site analysis, and they improve the specificity and potentially functional relevance of our findings (50).

A main limitation of our study is the relatively small study population. The present study was underpowered to allow stratification by offspring's sex or age, hence we did not address potential variability of effect estimated by gender or in different age groups. On the other hand, the study participants come from population-based cohorts, which is a strength of the study and to some degree allows for generalization of the results. In thorough analyses, we have accounted appropriately for the study design with two linked cohorts and family members. Further, we had information on personal smoking as well as smoking in both parents and have been able to account for main confounding factors (potentially associated with both the exposure and the outcome) in the analyses. However, rest confounding from included and unknown factors may still be present.

We have not been able to verify our findings in an independent cohort. We pursued replication in a sample from the IoW third-generation study, however, replication of significant DMRs proved not be possible as different methylation platforms were used in the two cohorts (Illumina 450K in IoW and Illumina EPIC Beadchip in RHINESSA/ECRHS) and a large number of sites were missing in the replication analysis. Few other cohorts have reliable and extensive information on father's smoking, while personal or maternal smoking are often well documented. Thus, the novel findings of DMRs related to father's tobacco smoking in our analyses, should be considered hypothesis generating and be interpreted with caution.

Conclusion

In conclusion, this hypothesis generating EWAS study is the first to report associations between paternal smoking and DNA methylation characteristics in adult and adolescent offspring. It is notable that differential methylation was detectable in this age group. Our results are intriguing as they indicate that fathers' exposures might persistently modify their future offspring's epigenome. This emphasizes the necessity to focus on male-line exposures in relation to phenotypic variation in their children, and further research to replicate our findings and explore potential mechanisms.

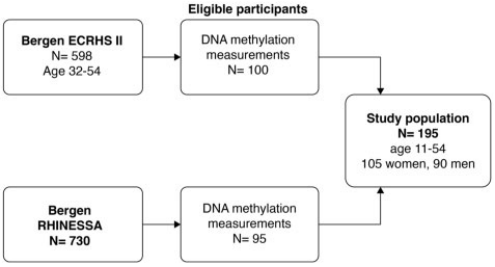


Figure 2: flowchart of study population. Offspring originate from two linked study populations with standardized and harmonized protocols: the ECRHS and the RHINESSA

Methods

Study Population

This study included data from 195 males and females aged 11–54 years participating in two linked population-based cohorts (Fig. 2).

The ECRHS conducted a study of population-based random samples of adult women and men aged 20–44 years in 1990–94 and followed up participants with clinical investigations in 2002–04 and 2012–14. The present analysis included 100 participants from the Bergen study centre with available DNA methylation data from DNA collected in ECRHS II. Information on father's year of birth was obtained from the Norwegian National Registry.

The Respiratory Health in Northern Europe, Spain and Australia study (RHINESSA) (www.rhinessa.net) investigated the offspring of ECRHS study participants in 10 study centres, following standardized protocols harmonized with the ECRHS protocols. The present analysis included 95 participants from the Bergen study centre in which DNA methylation was measured.

For the present analysis, offspring from the two cohorts were merged and analysed together. Information on smoking and other variables were obtained through interviews. Unless otherwise stated, definitions are similar in the two cohorts.

The study was approved by the Regional Committee for Medical and Health Research Ethics in Western Norway (RHINESSA: 2012/2017; ECRHS: 2010/759), and each participant gave written informed consent prior to participation.

Smoking Exposure and Covariates

In the RHINESSA cohort, information on fathers' smoking habits was collected from longitudinal data given by the fathers themselves as participants in the ECRHS II study, responding to the question: (i) *Have you ever smoked for as long as a year?*. In the ECRHS cohort, information on father's smoking was reported by the ECRHS participants and based on the question: *Did your*

father ever smoke regularly during your childhood? Father's smoking was categorized as a binary variable, as having smoked or not during offspring's childhood. In the present analysis paternal smoking was not defined in more detail as information regarding age of smoking onset was only available for RHINESSA participants.

Information on mothers smoking was reported by participants based on the question: *Did your mother ever smoke regularly during your childhood, or while pregnant with you? with the answering* categories 'no' (n ¼ 128), 'yes' (n ¼ 62), or 'don't know' (n ¼ 5) Maternal smoking was dichotomized as either having smoked ('yes') or never smoked ('no') during offspring's childhood, whereas 'don't' know' replies were excluded from further analyses.

Personal smoking was classified as current, ex or never smoking, based on the questions: i. *Have you ever smoked for as long as a year?* (ii) *If yes How old were you when you started smoking?* (iii) *Have you stopped or cut down smoking?* (iv) *How old were you when you stopped or cut down smoking?* Number of pack years was calculated based on the number of years smoked and the average number of daily cigarettes.

Parental educational attainment was categorized in as lower (primary school), intermediate (secondary school) and higher education (college or university). Personal education level was defined the same way in RHINESSA and categorized in three levels based on reported age when education was completed in ECRHS.

Methylation Measurements and Quality Control

DNA was extracted from whole blood using a standard salting out procedure (51). Samples were processed with the Illumina MethylationEPIC Beadchip microarray, which assesses methylation at > 850 000 CpGs. Methylation measurements were performed by the Oxford Genomics Centre (Oxford, UK) using the EZ 96-DNA methylation kit (Zymo Research, CA, USA), following the manufacturer's standard protocol, with multiple identical control samples assigned to each bisulphite conversion batch to assess assay variability. Samples were randomly distributed on microarrays to control against batch effects. The CPACOR pipeline (52) was used to pre-process and normalize the methylation data. We removed probes with CpG loci located on sex chromosomes and probes located at 0 distance to known SNPs. We applied Illumina background correction to all intensity values. Any intensity values having detection P values >¼10^{-16} were set as missing data. Samples with call rate <98% were excluded. After pre-processing, 765 082 sites remained for subsequent analysis. A quantile normalization was applied using limma on intensity values separately based on six different probe-type categories (Type-I M red, Type-I U red, Type-I M green, Type-I U green, Type-II red, and Type-II green). Beta values were then calculated from these normalized intensity values. ComBat was used to correct for batch effects (53).

Statistical Analyses

For identification of DMRs, composed of multiple signals across individual CpG positions, we used Comb-p (54) (Python version 2.7). This method identifies regions enriched for low P values based on the probe location and unadjusted P values from the site-specific CpG analysis. For each region the comb-p algorithm adjusts the CpG P values for auto-correction between probes by using the Stouffer-Liptak-Kechris (slk) correction, followed by multiple testing adjustment using a one-step Sidak correction method (54). Regions containing at least two probes and having a Sidak-corrected P-value <0.05 were considered statistically significant.

Robust multivariate linear regression models were used to analyse the association of offspring differentially associated probes (DMPs) adjusted for paternal and offspring age, offspring gender, as well as personal and maternal smoking status. Educational level was added in sensitivity analyses to account for socioeconomic status. Cell proportions (CD8T, CD4T, NK, B Cells, Monocytes, Granulocytes) were estimated using the minfi package (55) (R version 3.4.2), and cell composition coefficients were derived using the Houseman method (56). These were additively included in the model. Multiple hypothesis testing was accounted for by controlling the false discovery rate (FDR), using Benjamini and Hochberg's algorithm (57). CpGs with FDR-corrected P-value <0.1 were considered statistically significant and normalized methylation betas were used as outcome measurements. In order to address possible inflation of our test statistics by systematic biases, a Bayesian method based on estimation of the empirical null distribution was applied using the R/Bioconductor package BACON (58), and P values were estimated.

Some of the study participants originated from the same family. To account for this, we performed linear mixed model analysis on the top CpGs including family IDs as random effect.

For CpG annotation, we used the UCSC Genome browser annotation provided by Illumina in the array manifest together with SNIPPER (version 1.2, http://csg.sph.umich.edu/boehnke/snipper/) to annotate the nearest gene within 10 Mb of each CpG.

To investigate the regulatory context of the top differentially methylated probes (inflation-adjusted P-value <0.00001), we performed enrichment analysis in annotated regulatory elements (TF Chip seq and histone modification signatures) from the ENCODE (59), as well as the Epigenomics roadmap (60) using Enrichr (61).

Pathway analysis was conducted using KEGG (62), and GO databases (63) using gometh function in the missMethyl package (52).

Replication in Isle of Wight Cohort

To pursue replication of findings, we used the IoW third-generation study which since 2010 has enrolled children born to second-generation parents—the original Birth cohort. Extensive descriptions of the IoW multigenerational cohort design and objectives have been published elsewhere (64, 65). Father's smoking information given by the fathers themselves, and DNA methylation measurements using the Illumina Infinium HumanMethylation450 Beadchip array in cord blood DNA available for 159 subjects were included in the present analysis.

Availability of Data and Material

The data that support the findings of this study are available from Bergen study centre of RHINESSA and ECRHS generational population studies, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are, however, available from the authors upon reasonable request and with permission of RHINESSA and ECRHS.

Ethics Approval and Consent to Participate

The study was approved by the Regional Committee for Medical and Health Research Ethics in Western Norway

(RHINESSA: 2012/2017; ECRHS: 2010/759), and each participant gave written informed consent prior to participation.

Consent for Publication

Not applicable.

Acknowledgements

The authors thank Professor Stephanie London for her contribution to interpretation of early analyses and the reviewer's for their helpful comments on the manuscript.

Funding

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 633212, and from the Western Norway Regional Health Authorities strategic investment grant no. 912011.

Supplementary Data

Supplementary data are available at *EnvEpig* online.

Conflict of interest statement. None declared.

References

1. Bygren LO, Kaati G, Edvinsson S. Longevity determined by paternal ancestors' nutrition during their slow growth period. *Acta Biotheoret* 2001;49:53–9.
2. Kaati G, Bygren LO, Edvinsson S. Cardiovascular and diabetes mortality determined by nutrition during parents' and grandparents' slow growth period. *Eur J Hum Genet* 2002;10:682–8.
3. Pembrey ME, Bygren LO, Kaati G, Edvinsson S, Northstone K, Sjo¨ stro¨ m M, Golding J. Sex-specific, male-line transgenerational responses in humans. *Eur J Hum Genet* 2006;14:159–66.
4. Carone BR, Fauquier L, Habib N, Shea JM, Hart CE, Li R, Bock C, Li C, Gu H, Zamore PD, Meissner A, Weng Z, Hofmann HA, Friedman N, Rando OJ. Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell* 2010;143:1084–96.
5. Ng SF, Lin RC, Laybutt DR, Barres R, Owens JA, Morris MJ. Chronic high-fat diet in fathers programs beta-cell dysfunction in female rat offspring. *Nature* 2010;467:963–6.
6. Fullston T, Ohlsson Teague EMC, Palmer NO, DeBlasio MJ, Mitchell M, Corbett M, Print CG, Owens JA, Lane M. Paternal obesity initiates metabolic disturbances in two generations of mice with incomplete penetrance to the F2 generation and alters the transcriptional profile of testis and sperm microRNA content. *FASEB J* 2013;27:4226–43.
7. Northstone K, Golding J, Davey Smith G, Miller LL, Pembrey M. Prepubertal start of father's smoking and increased body fat in his sons: further characterisation of paternal transgenerational responses. *Eur J Hum Genet* 2014;22:1382–6.
8. Accordini SC, Johannessen A, Portas L, Benediktsdo´ ttir B, Bertelsen R. A three-generation study on the association of tobacco smoking with asthma. *Int J Epidemiol* 2018;47:1106–17.
9. Bakulski KM, Fallin MD. Epigenetic epidemiology: promises for public health research. *Environ Mol Mutagen* 2014;55: 171–83.
10. Bird A. Perceptions of epigenetics. *Nature* 2007;447:396–8.
11. Feinberg AP. Phenotypic plasticity and the epigenetics of human disease. *Nature* 2007;447:433–40.
12. Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am J Hum Genet* 2011;88:450–7.
13. Zeilinger S, Ku¨ hnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, Weidinger S, Lattka E, Adamski J, Peters A, Strauch K, Waldenberger M, Illig T. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS One* 2013;8:e63812.
14. Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, Belvisi MG, Brown R, Vineis P, Flanagan JM. Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. *Hum Mol Genet* 2013;22:843–51.
15. Philibert RA, Beach SR, Brody GH. Demethylation of the aryl hydrocarbon receptor repressor as a biomarker for nascent smokers. *Epigenetics* 2012;7:1331–8.
16. Joubert BR, Ha˚ berg SE, Nilsen RM, Wang X, Vollset SE, Murphy SK, Huang Z, Hoyo C, Midttun Ø, Cupul-Uicab LA, Ueland PM, Wu MC, Nystad W, Bell DA, Peddada SD, London SJ. 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. *Environ Health Perspect* 2012;120:1425–31.
17. Murphy SK, Adigun A, Huang Z, Overcash F, Wang F, Jirtle RL, Schildkraut JM, Murtha AP, Iversen ES, Hoyo C. Gender-specific methylation differences in relation to prenatal exposure to cigarette smoke. *Gene* 2012;494:36–43.
18. Richmond RC, Simpkin AJ, Woodward G, Gaunt TR, Lyttleton O, McArdle WL, Ring SM, Smith ADAC, Timpson NJ, Tilling K, Davey Smith G, Relton CL. Prenatal exposure to maternal smoking and offspring DNA methylation across the lifecourse: findings from the Avon Longitudinal Study of Parents and Children (ALSPAC). *Hum Mol Genet* 2015;24:2201–17.
19. Tehranifar P, Wu H-C, McDonald JA, Jasmine F, Santella RM, Gurvich I, Flom JD, Terry MB. Maternal cigarette smoking during pregnancy and offspring DNA methylation in midlife. *Epigenetics* 2018;13:129–34.
20. Wiklund P, Karhunen V, Richmond R, Rodriguez A, De Silva M, Wielscher M *et al*. DNA methylation links prenatal smoking exposure to later life health outcomes in offspring. *Clin Epigenetics* 2019;11:97.
21. Entrez Gene: Breast Carcinoma Amplified Sequence 1. http://www.genecards.org (8 August 2019, date last accessed).
22. Kudo M, Khalifeh Soltani SMA, Sakuma SA, McKleroy W, Lee T-H, Woodruff PG, Lee JW, Huang K, Chen C, Arjomandi M, Huang X, Atabai K. Mfge8 suppresses airway hyperresponsiveness in asthma by regulating smooth muscle contraction. *Proc Natl Acad Sci USA* 2013;110:660–5.
23. Entrez Gene: Unc-93 Homolog B1, TLR Signaling Regulator. http://www.genecards.org (8 August 2019, date last accessed).
24. Entrez Gene: RAS Like Proto-oncogene B. http://www.genecards.org (8 August 2019, date last accessed).
25. Entrez Gene: DLG Associated Protein 1. http://www.genecards.org (8 August 2019, date last accessed).
26. Joubert BR, Felix JF, Yousefi P, Bakulski KM, Just AC, Breton C, Reese SE, Markunas CA, Richmond RC, Xu C-J, Ku¨ pers LK, Oh SS, Hoyo C, Gruzieva O, So¨ derha¨ ll C, Salas LA, Ba¨ız N, Zhang H, Lepeule J, Ruiz C, Ligthart S, Wang T, Taylor JA, Duijts L, Sharp GC, Jankipersadsing SA, Nilsen RM, Vaez A, Fallin MD, Hu D, Litonjua AA, Fuemmeler BF, Huen K, Kere J, Kull I, Munthe-Kaas MC, Gehring U, Bustamante M, Saurel-Coubizolles MJ, Quraishi BM, Ren J, Tost J, Gonzalez JR, Peters

MJ, Ha˚berg SE, Xu Z, van Meurs JB, Gaunt TR, Kerkhof M, Corpeleijn E, Feinberg AP, Eng C, Baccarelli AA, Benjamin Neelon SE, Bradman A, Merid SK, Bergstro¨m A, Herceg Z, Hernandez-Vargas H, Brunekreef B, Pinart M, Heude B, Ewart S, Yao J, Lemonnier N, Franco OH, Wu MC, Hofman A, McArdle W, Van der Vlies P, Falahi F, Gillman MW, Barcellos LF, Kumar A, Wickman M, Guerra S, Charles M-A, Holloway J, Auffray C, Tiemeier HW, Smith GD, Postma D, Hivert M-F, Eskenazi B, Vrijheid M, Arshad H, Anto´ JM, Dehghan A, Karmaus W, Annesi-Maesano I, Sunyer J, Ghantous A, Pershagen G, Holland N, Murphy SK, DeMeo DL, Burchard EG, Ladd-Acosta C, Snieder H, Nystad W, Koppelman GH, Relton CL, Jaddoe VWV, Wilcox A, Mele´n E, London SJ. DNA methylation in newborns and maternal smoking in pregnancy: genome-wide consortium meta-analysis. *Am J Hum Genet* 2016;98:680–96.

27. Liu M, Fan R, Liu X, Cheng F, Wang J. Pathways and networks-based analysis of candidate genes associated with nicotine addiction. *PLoS One* 2015;10:e0127438.
28. Uhl GR, Liu Q-R, Drgon T, Johnson C, Walther D, Rose JE, David SP, Niaura R, Lerman C. Molecular genetics of successful smoking cessation: convergent genome-wide association study results. *Arch Gen Psychiatry* 2008;65:683–93.
29. Mexal S, Berger R, Pearce L, Barton A, Logel J, Adams CE, Ross RG, Freedman R, Leonard S. Regulation of a novel alphaN-catenin splice variant in schizophrenic smokers. *Am J Med Genet* 2008;147b:759–68.
30. Mill J, Heijmans BT. From promises to practical strategies in epigenetic epidemiology. *Nat Rev Genet* 2013;14:585.
31. Entrez Gene: ATPase HþTransporting V1 Subunit E1. http://www.genecards.org (8 August 2019, date last accessed).
32. Entrez-Gene: Complement C2. http://www.genecards.org (8 August 2019, date last accessed).
33. Entrez-Gene: Catenin Alpha 2. http://www.genecards.org (8 August 2019, date last accessed).
34. Fan Z, Qian Y, Lu Q, Wang Y, Chang S, Yang L. DLGAP1 and NMDA receptor-associated postsynaptic density protein genes influence executive function in attention deficit hyperactivity disorder. *Brain Behav* 2018;8:e00914.
35. Mattheisen M, Samuels JF, Wang Y, Greenberg BD, Fyer AJ, McCracken JT, Geller DA, Murphy DL, Knowles JA, Grados MA, Riddle MA, Rasmussen SA, McLaughlin NC, Nurmi EL, Askland KD, Qin H-D, Cullen BA, Piacentini J, Pauls DL, Bienvenu OJ, Stewart SE, Liang K-Y, Goes FS, Maher B, Pulver AE, Shugart YY, Valle D, Lange C, Nestadt G. Genome-wide association study in obsessive-compulsive disorder: results from the OCGAS. *Mol Psychiatry* 2015;20:337–44.
36. de Vocht F, Simpkin AJ, Richmond RC, Relton C, Tilling K. Assessment of offspring DNA methylation across the life-course associated with prenatal maternal smoking using Bayesian mixture modelling. *Int J Ent Res Public Health* 2015;12:14461–76.
37. Markunas CA, Xu Z, Harlid S, Wade PA, Lie RT, Taylor JA, Wilcox AJ. Identification of DNA methylation changes in newborns related to maternal smoking during pregnancy. *Environ Health Perspect* 2014;122:1147–53.
38. Kupers LK, Xu X, Jankipersadsing SA, Vaez A, la Bastide-van Gemert S, Scholtens S *et al*. DNA methylation mediates the effect of maternal smoking during pregnancy on birthweight of the offspring. *Int J Epidemiol* 2015;44:1224–37.
39. Wiklund P, Karhunen V, Richmond RC, Parmar P, Rodriguez A, De Silva M, Wielscher M, Rezwan FI, Richardson TG, Veijola J, Herzig K-H, Holloway JW, Relton CL, Sebert S, Ja¨rvelin M-R. DNA methylation links prenatal smoking exposure to later life health outcomes in offspring. *Clin Epigenet* 2019;11:97.
40. Wan ES, Qiu W, Baccarelli A, Carey VJ, Bacherman H, Rennard SI, Agusti A, Anderson W, Lomas DA, DeMeo DL. Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome. *Hum Mol Genet* 2012;21:3073–82.
41. Zhang Y, Yang R, Burwinkel B, Breitling LP, Brenner H. F2RL3 methylation as a biomarker of current and lifetime smoking exposures. *Environ Health Perspect* 2014;122:131–7.
42. Monick MM, Beach SRH, Plume J, Sears R, Gerrard M, Brody GH, Philibert RA. Coordinated changes in AHRR methylation in lymphoblasts and pulmonary macrophages from smokers. *Am J Med Genet* 2012;159b:141–51.
43. Joubert BR, Haberg SE, Bell DA, Nilsen RM, Vollset SE, Midttun O *et al*. Maternal smoking and DNA methylation in newborns: in utero effect or epigenetic inheritance? *Cancer Epidemiol* 2014;23:1007–17.
44. Loukola A, Hallfors J, Korhonen T, Kaprio J. Genetics and smoking. *Curr Addict Rep* 2014;1:75–82.
45. Wang J, Li MD. Common and unique biological pathways associated with smoking initiation/progression, nicotine dependence, and smoking cessation. *Neuropsychopharmacology* 2010;35:702–19.
46. Kiechl-Kohlendorfer U, Ralser E, Pupp Peglow U, Reiter G, Griesmaier E, Trawo¨ger R. Smoking in pregnancy: a risk factor for adverse neurodevelopmental outcome in preterm infants? *Acta Paediatr* 2010;99:1016–9.
47. Toledo-Rodriguez M, Lotfipour S, Leonard G, Perron M, Richer L, Veillette S, Pausova Z, Paus T. Maternal smoking during pregnancy is associated with epigenetic modifications of the brain-derived neurotrophic factor-6 exon in adolescent offspring. *Am J Med Genet* 2010;153b:1350–4.
48. Svanes C, Koplin J, Skulstad SM, Johannessen A, Bertelsen RJ, Benediktsdottir B, Bra˚ba¨ck L, Elie Carsin A, Dharmage S, Dratva J, Forsberg B, Gislason T, Heinrich J, Holm M, Janson C, Jarvis D, Jo¨gi R, Krauss-Etschmann S, Lindberg E, Macsali F, Malinovschi A, Modig L, Norba¨ck D, Omenaas E, Waatevik Saure E, Sigsgaard T, Skorge TD, Svanes Ø, Tore´n K, Torres C, Schlu¨nssen V, Gomez Real F. Father's environment before conception and asthma risk in his children: a multi-generation analysis of the Respiratory Health In Northern Europe study. *Int J Epidemiol* 2017;46:235–45.
49. Breton CV, Marsit CJ, Faustman E, Nadeau K, Goodrich JM, Dolinoy DC, Herbstman J, Holland N, LaSalle JM, Schmidt R, Yousefi P, Perera F, Joubert BR, Wiemels J, Taylor M, Yang IV, Chen R, Hew KM, Freeland DMH, Miller R, Murphy SK. Small-magnitude effect sizes in epigenetic end points are important in children's environmental health studies: the Children's Environmental Health and Disease Prevention Research Center's Epigenetics Working Group. *Environ Health Perspect* 2017;125:511–26.
50. Ward LD, Kellis M. Interpreting noncoding genetic variation in complex traits and human disease. *Nat Biotechnol* 2012;30:1095–106.
51. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 1988;16:1215.
52. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics* 2016;32:286–8.
53. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007;8:118–27.

54. Pedersen BS, Schwartz DA, Yang IV, Kechris KJ. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated *P*-values. *Bioinformatics* 2012;28:2986–8.
55. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 2014;30:1363–9.
56. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 2012;13:86.
57. Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc Ser B Methodol* 1995;57:289–300.
58. van Iterson M, van Zwet EW, Heijmans BT. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol* 2017;18:19.
59. Consortium EP. An integrated Encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57–74.
60. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR, Farnham PJ, Hirst M, Lander ES, Mikkelsen TS, Thomson JA. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* 2010;28:1045–8.
61. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 2016;44:W90–7.
62. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 2017;45:D353–d61.
63. Consortium GO. Gene Ontology Consortium: going forward. *Nucleic Acids Research* 2015;43:D1049–56.
64. Arshad SH, Karmaus W, Zhang H, Holloway JW. Multigenerational cohorts in patients with asthma and allergy. *J Allergy Clin Immunol* 2017;139:415–21.
65. Felix JF, Joubert BR, Baccarelli AA, Sharp GC, Almqvist C, Annesi-Maesano I, Arshad H, Ba¨ız N, Bakermans-Kranenburg MJ, Bakulski KM, Binder EB, Bouchard L, Breton CV, Brunekreef B, Brunst KJ, Burchard EG, Bustamante M, Chatzi L, Cheng Munthe-Kaas M, Corpeleijn E, Czamara D, Dabelea D, Davey Smith G, De Boever P, Duijts L, Dwyer T, Eng C, Eskenazi B, Everson TM, Falahi F, Fallin MD, Farchi S, Fernandez MF, Gao L, Gaunt TR, Ghantous A, Gillman MW, Gonseth S, Grote V, Gruzieva O, Ha˚ berg SE, Herceg Z, Hivert M-F, Holland N, Holloway JW, Hoyo C, Hu D, Huang R-C, Huen K, Ja¨ rvelin M-R, Jima DD, Just AC, Karagas MR, Karlsson R, Karmaus W, Kechris KJ, Kere J, Kogevinas M, Koletzko B, Koppelman GH, Ku¨ pers LK, Ladd-Acosta C, Lahti J, Lambrechts N, Langie SAS, Lie RT, Liu AH, Magnus MC, Magnus P, Maguire RL, Marsit CJ, McArdle W, Mele´n E, Melton P, Murphy SK, Nawrot TS, Nistico` L, Nohr EA, Nordlund B, Nystad W, Oh SS, Oken E, Page CM, Perron P, Pershagen G, Pizzi C, Plusquin M, Raikkonen K, Reese SE, Reischl E, Richiardi L, Ring S, Roy RP, Rzehak P, Schoeters G, Schwartz DA, Sebert S, Snieder H, Sørensen TI, Starling AP, Sunyer J, Taylor JA, Tiemeier H, Ullemar V, Vafeiadi M, Van Ijzendoorn MH, Vonk JM, Vriens A, Vrijheid M, Wang P, Wiemels JL, Wilcox AJ, Wright RJ, Xu C-J, Xu Z, Yang IV, Yousefi P, Zhang H, Zhang W, Zhao S, Agha G, Relton CL, Jaddoe VWV, London SJ. Cohort profile: Pregnancy and Childhood Epigenetics (PACE) Consortium. *Int J Epidemiol* 2018;47:22–3u.
66. Baud V, Mears AJ, Lamour V, Scamps C, Duncan AM, McDermid HE, Lipinski M. The E subunit of vacuolar H(+)-ATPase localizes close to the centromere on human chromosome 22. *Human Molecular Genetics* 1994;3:335–9.
67. van Hille B, Vanek M, Richener H, Green JR, Bilbe G. Cloning and tissue distribution of subunits C, D, and E of the human vacuolar H(þ)–ATPase. *Biochem Biophys Res Commun* 1993;197:15–21.
68. Ishii Y, Zhu ZB, Macon KJ, Volanakis JE. Structure of the human C2 gene. *J Immunol* (Baltimore, Md : 1950). 1993;151:170–4.
69. Schaffer AE, Breuss MW, Caglayan AO, Al-Sanaa N, Al-Abdulwahed HY, Kaymakcalan H *et al*. Biallelic loss of human CTNNA2, encoding alphaN-catenin, leads to ARP2/3 complex overactivity and disordered cortical neuronal migration. *Nature Genetics* 2018;50:1093–101.
70. www.genecards.org, "Entrez-Gene: Acyl-CoA synthesase Family member 3" Retrieved 2019-08-08.
71. www.genecards.org, "Gene-Entrez: Wd Repeat Domain 60" Retrieved 2019-08-08.

Supplementary Table S1: Differentially methylated DMPs

PROBEID	Beta	SE	P-value	Adjusted P-v	CHR	MAPINFO	Gene	CpG Islands	Relation to CpG Island	SNP ID	SNP_DISTANCE	SNP MAF
cg00626693	−0,01372851	0,002537177	6,27E−08	7,64E−06	16	30622810	ZNF689	chr16:30621000-30621826	S_Shore			
cg01079199	−0,01823898	0,00374623	1,12E−06	5,66E−05	5	175644411	LOC643201			rs569031559	1	0,0002
cg01798114	0,02388316	0,004813461	6,99E−07	4,07E−05	8	68644499	CPA6			rs184091018	2	0,0002
cg04164584	−0,009861003	0,001934469	3,44E−07	2,49E−05	17	27235821	PHF12					
cg04523826	0,02491398	0,005276232	2,34E−06	9,41E−05	12	92587036	BTG1			rs75244279	48	0,017572
cg05019203	−0,01750032	0,003152337	2,83E−08	4,40E−06	20	52612962	BCAS1					
cg05235344	0,01499294	0,003039008	8,08E−07	4,50E−05	6	11810356	ADTRP			rs553022132;rs566450556	0;41	0.000200;0.000200
cg05800723	−0,009612961	0,001996907	1,48E−06	6,85E−05	4	79103126	FRAS1	chr4:79102900-79103261	Island	rs550010703	37	0
cg06876354	0,01746758	0,003465755	4,65E−07	3,07E−05	2	121020189	RALB			rs143634079	0	0,000399
cg07217718	0,02545433	0,0051054	6,17E−07	3,73E−05	18	3585484	DLGAP1			rs573174367	1	0,0002
cg07530264	0,02937474	0,006235608	2,47E−06	9,78E−05	8	144965497	EPPK1	chr8:144966638-144966880	N_Shore	rs569098231;rs11782138;rs558007321	0;1;26	0.000200;0.011182;0.000200
cg09108969	−0,01300435	0,002724283	1,81E−06	7,88E−05	15	102094562	PCSK6			rs561859890	16	0,001198
cg09696115	0,02119014	0,004271342	7,01E−07	4,08E−05	15	59841613	FAM81A			rs547318521;rs540911708	41;2	0.001398;0.000399
cg10284656	−0,01361793	0,002846137	1,71E−06	7,59E−05	8	61086196	CA8	chr8:61086125-61086327	Island			
cg12404462	−0,02240805	0,004647295	1,42E−06	6,67E−05	2	239490640				rs186011166	29	0,0002
cg13209192	−0,01388532	0,002904336	1,75E−06	7,69E−05	12	5352389	NTF3			rs573941541;rs192795704;rs183633905;rs575911112;rs139597544	44;37;23;2;1	0.000200;0.000599;0.000200;0.000200;0.006589
cg13971124	−0,01245127	0,002602005	1,71E−06	7,57E−05	11	133808270	IGSF9B			rs539784721	31	0,000799
cg17190420	−0,01263732	0,002643739	1,75E−06	7,71E−05	6	170545114	LOC154449	chr6:170546535-170547104	N_Shore	rs369731030;rs534634087	1;15	0.000599;0.000200
cg17557522	−0,01083295	0,002296594	2,39E−06	9,58E−05	7	100090170	NYAP1	chr7:100091180-100091598	N_Shore	rs186500512;rs4727456;rs537212512	1;11;37	0.000399;0.044129;0.000200
cg19423196	0,03963108	0,008112568	1,03E−06	5,34E−05	10	82049429	MAT1A			rs575292837;rs181353304;rs141011130	8;9;13	0.000200;0.000200;0.000998
cg19754387	0,005632547	0,001067865	1,33E−07	1,29E−05	2	208576057	CCNYL	chr2:208576001-208577106	Island	rs549186130	45	0,0002
cg20272935	0,02376082	0,004638743	3,02E−07	2,27E−05	11	67765720	UNC93B1	chr11:67764918-67765256	S_Shore	rs547262516;rs565837193;rs539277800;rs539072978;rs569826127	26;10;9;2;1	0.000200;0.000200;0.000200;0.001198;0.000399
cg20615832	0,04494551	0,009398729	1,73E−06	7,65E−05	4	74719172	PF4V1	chr4:74719087-74719339	Island	rs141256311	25	0,00023
cg21545602	−0,01375361	0,0029097	2,28E−06	9,26E−05	9	139566747	EGFL7	chr9:139564999-139565286	S_Shore	rs375782632;rs78568191;rs377086652;rs201487346;rs201787878;rs77612807	0;1;22;27;29;31	0.000200;0.003395;0.001198;0.000200;0.001000;0.033946
cg21719704	0,03984661	0,008160612	1,05E−06	5,38E−05	19	55948588	SHISA7	chr19:55951800-55952103	N_Shelf	rs189561755	21	0,001198
cg22856279	0,0522964	0,010732216	1,10E−06	5,58E−05	21	44830477	SIK1B	chr21:44830845-44831152	N_Shore	rs565975952;rs374121752	1;47	0.000799;0.001398
cg23530133	0,01877393	0,003959268	2,12E−06	8,80E−05	19	49945298	SLC17A7	chr19:49944408-49944656	S_Shore			
cg24317857	0,0320016	0,006446162	6,89E−07	4,03E−05	7	38711703	FAM183BP			rs141581803;rs117943535	15;31	0.000200;0.007987
cg24534854	−0,01327091	0,00255638	2,09E−07	1,76E−05	8	22582613	PEBP4			rs552431023	0	0,0002
cg25012097	−0,01162767	0,002308697	4,74E−07	3,11E−05	13	39263863	FREM2	chr13:39261236-39262299	S_Shore	rs140101984;rs369187680	17;1	0.001398;0.000200
cg25727029	0,0130095	0,002360503	3,56E−08	5,16E−06	15	89482453	MFGE8			rs114173492	14	0,007788
cg25903220	0,01264042	0,002586482	1,02E−06	5,30E−05	8	101532284	ANKRD46					

PROBEID = probe identifiers; BETA = estimates; SE = standard error; Adjusted P-value = P-value adjusted by multiple test correction; CHR = chromosome; MAPINFO = position of the CpGs in the chromosome; Gene = UCSC RefGene; CpG Islands = Nearest UCSC CpG island names; Relation to CpG Island = relation to the nearest UCSC CpG islands; SNP ID = SNP identifiers within the 50 bp of the probe; SNP DISTANCE = Distance of the SNP; SNP MAF = SNP minor allele frequency

Suplementary Table S2: ENCODE annotated histone modification signatures

Term	P-value	Adjusted P-value	z-score	Combined Score	Genes
H3K27me3_skeletal muscle myoblast_hg19	1,24E−04	0,0509351644200152	3,54838709677419	31,9291858041441	EGFL7;FRAS1;BTG1;PEBP4;DLGAP1;CA8;PCSK6;FAM81A;SHISA7;FREM2;PF4V1
H3K27me3_osteoblast_hg19	0,00257386363009652	0,530215907799884	2,90322580645161	17,3100401157256	EGFL7;FRAS1;RALB;LOC643201;DLGAP1;CA8;PCSK6;FREM2;PF4V1
H3K27me3_endothelial cell of umbilical vein_hg19	0,00325822246856827	0,447462552350043	2,2817375431391	13,0665377196303	FRAS1;LOC643201;NTF3;SLC17A7;DLGAP1;MAT1A;IGSF9B;CA8;PCSK6;FAM81A;SHISA7;FREM2
H3K27me3_BJ_hg19	0,0034793447966563	0,358372514055599	2,77820651335082	15,7271805888886	NTF3;SLC17A7;IGSF9B;CA8;PCSK6;FAM81A;SHISA7;FREM2;PF4V1
H3K27me3_fibroblast of lung_hg19	0,00449056102431885	0,370022228403873	2,32528643301061	12,5699813952273	EGFL7;FRAS1;NTF3;EPPK1;SLC17A7;DLGAP1;CA8;FAM81A;SHISA7;FREM2;PF4V1
H3K9me3_CD14-positive monocyte_hg19	0,00953191088963985	0,654524547755269	2,58064516129032	12,0080259828062	EGFL7;FRAS1;NTF3;DLGAP1;IGSF9B;PCSK6;FAM81A;FREM2
H3K27me3_kidney epithelial cell_hg19	0,0171890792211282	1	2,32805156634219	9,4599933669613	LOC643201;SLC17A7;DLGAP1;MAT1A;CA8;FAM81A;SHISA7;FREM2
H3K27me3_cardiac mesoderm_hg19	0,0328173204442117	1	1,85125190910353	6,32535537676559	UNC93B1;LOC643201;NTF3;SLC17A7;IGSF9B;CA8;PCSK6;FAM81A;SHISA7;PF4V1
H3K27me3_bronchial epithelial cell_hg19	0,0368814264417684	1	2,1691301787983	7,1582319797456	NTF3;SLC17A7;DLGAP1;CA8;FAM81A;SHISA7;FREM2
H3K27me3_CD14-positive monocyte_hg19	0,0725213857993744	1	1,68547216629992	4,4224662309432	FRAS1;NYAP1;NTF3;DLGAP1;MAT1A;IGSF9B;CA8;FREM2;PF4V1

Term = Enrichment terms; P-value = P-value from hypergeometric test; Adjusted P-value = adjusted p-value using the Benjamini-Hochberg method; z-score = measuing deviation from an expected rank; Combined Score = combination of the p-value and z-score calculated by multiplying the two scores, where P-value is natural log converted; Genes = genes observed in the enriched term

Suplementary Table S3: ENCODE annotated transription factor sites

Term	P-value	Adjusted P-value	z-score	Combined Score	Genes
EP300_MCF-7_hg19	0,00247204267985733	1	11,0599078341013	66,3894245996823	BTG1;MAT1A;FREM2
ESR1_T47D_hg19	0,0275809481119422	1	4,55407969639468	16,3520153187787	MAT1A;FREM2;BCAS1
EZH2_keratinocyte_hg19	0,0304526995974828	1	2,25806451612903	7,88421433037727	RALB;LOC643201;DLGAP1;MAT1A;CA8;PCSK6;PF4V1
RNF2_K562_hg19	0,0304526995974828	1	2,25806451612903	7,88421433037727	EGFL7;FRAS1;RALB;BTG1;ANKRD46;SLC17A7;CA8
ZNF217_MCF-7_hg19	0,0305152097972264	1	2,4782123828012	8,64779654981532	FRAS1;BTG1;EPPK1;MAT1A;MFGE8;FREM2
ZEB1_HepG2_hg19	0,0341902068110115	1	6,86341798215511	23,1696364205555	MAT1A;CA8
TCF7L2_HepG2_hg19	0,0521615560257644	1	3,53190487402872	10,4311615141031	FRAS1;MAT1A;BCAS1
EZH2_fibroblast of dermis_hg19	0,0832613775871922	1	1,93548387096774	4,81116869370144	ANKRD46;UNC93B1;NTF3;CA8;PCSK6;SHISA7
CBX2_K562_hg19	0,0832613775871922	1	1,93548387096774	4,81116869370144	FRAS1;RALB;PEBP4;LOC154449;SLC17A7;SHISA7
TCF7L2_MCF-7_hg19	0,0832613775871922	1	1,93548387096774	4,81116869370144	FRAS1;BTG1;ANKRD46;ZNF689;FREM2;BCAS1

Term = Enrichment terms; P-value = P-value from hypergeometric test; Adjusted P-value = adjusted p-value using the Benjamini-Hochberg method; z-score = measuing deviation from an expected rank; Combined Score = combination of the p-value and z-score calculated by multiplying the two scores, where P-value is natural log converted; Genes = genes observed in the enriched term

Suplementary Table S4: Epigenomic roadmap annotated histone modification marks

Term	P-value	Adjusted P-value	z-score	Combined Score	Genes
H3K27me3 Penis Foreskin Fibroblast Primary Cells	0,00296418600860966	1	3,12995168137091	18,2199270488896	BTG1;NTF3;EPPK1;DLGAP1;CA8;FAM81A;FREM2;BCAS1
H3K27me3 Colonic Mucosa	0,00398353312721357	0,7628465938614	2,36086965853239	13,0451886471749	EGFL7;FRAS1;PEBP4;NTF3;EPPK1;MAT1A;IGSF9B;CA8;MFGE8;FAM81A;FREM2
H3K27me3 Rectal Smooth Muscle	0,00990381067966931	1	2,8208176341 3995	13,0176098613181	EGFL7;PEBP4;LOC643201;DLGAP1;CA8;FAM81A;FREM2
H3K27me3 Stomach Smooth Muscle	0,017320698145 1786	1	3,24854627554169	13,1756263783488	EGFL7;FRAS1;CPA6;FREM2;PF4V1
H3K27me3 Mobilized CD34 Primary Cells	0,028611650683 7016	1	1,73858870062226	6,17884214781405	FRAS1;PEBP4;LOC643201;NTF3;LOC154449;DLGAP1;MAT1A;CA8;FAM81A;SHISA7;FREM2;PF4V1
H3K4me1 H1 Derived Mesenchymal Stem Cells	0,035148759531 2799	1	1,54320193826163	5,16689618684938	ANKRD46;UNC93B1;MAT1A;PCSK6;SHISA7;FRAS1;LOC643201;NTF3;SLC17A7;DLGAP1;IGSF9B;CA8;FAM81A;FREM2;PF4V1
H3K27me3 IMR90	0,048112238297 5393	1	2,2234162791 1285	6,74633125146416	FRAS1;LOC643201;DLGAP1;CA8;FAM81A;FREM2
H3K27me3 Brain Cingulate Gyrus	0,054315959078 8866	1	2,76893257649176	8,06572667737932	EGFL7;FRAS1;MAT1A;PCSK6
H3K27me3 Adipose Derived Mesenchymal Stem Cell Cultured Cells	0,062094550642 9752	1	2,30414746543778	6,40344941293531	EGFL7;PEBP4;LOC154449;FAM81A;PF4V1
H3K27me3 Bone Marrow Derived Mesenchymal Stem Cell Cultured Cells	0,067765522628557	1	2,57036370646446	6,91864751624606	FRAS1;LOC643201;DLGAP1;PF4V1

Term = Enrichment terms; P-value = P-value from hypergeometric test; Adjusted P-value = adjusted p-value using the Benjamini-Hochberg method; z-score = measuing deviation from an expected rank; Combined Score = combination of the p-value and z-score calculated by multiplying the two scores, where P-value is natural log converted; Genes = genes observed in the enriched term

Suplementary Table S5: Sensitivity analysis with family id included as random effect in mixed linear regression model

PROBEID	BETA	SE	P-value	Adjusted P-val	BETA (mixed-model)	SE (mixed-model)	P-value (mixed-m	Adjusted P-value (mixed-model)
cg05019203	–0,01750032	0,003152337	2,83E–08	0,013632939	0,0268	0,00444	1,01E–08	3,23E–07
cg25727029	0,0130095	0,002360503	3,56E–08	0,013632939	0,029	0,005	3,46E–08	5,54E–07
cg00626693	–0,01372851	0,002537177	6,27E–08	0,015999501	–0,0145	0,00257	7,07E–08	7,54E–07
cg19754387	0,005632547	0,001067865	1,33E–07	0,020373034	0,0123	0,00224	1,39E–07	1,11E–06
cg24534854	–0,01327091	0,00255638	2,09E–07	0,026648193	0,0247	0,00469	4,44E–07	2,84E–06
cg20272935	0,02376082	0,004638743	3,02E–07	0,032932653	–0,0135	0,0027	1,39E–06	6,81E–06
cg04164584	–0,009861003	0,001934469	3,44E–07	0,032932653	–0,0138	0,00276	1,49E–06	6,81E–06
cg06876354	0,01746758	0,003465755	4,65E–07	0,036305051	–0,0136	0,00277	2,14E–06	8,56E–06
cg25012097	–0,01162767	0,002308697	4,74E–07	0,036305051	–0,012	0,00249	3,50E–06	1,24E–05
cg07217718	0,02545433	0,0051054	6,17E–07	0,038350721	0,0312	0,00655	4,24E–06	1,29E–05
cg24317857	0,0320016	0,006446162	6,89E–07	0,038350721	–0,00931	0,00196	4,43E–06	1,29E–05
cg01798114	0,02388316	0,004813461	6,99E–07	0,038350721	0,049	0,0104	5,19E–06	1,29E–05
cg09696115	0,02119014	0,004271342	7,01E–07	0,038350721	0,0422	0,00896	5,25E–06	1,29E–05
cg05235344	0,01499294	0,003039008	8,08E–07	0,04096965	–0,0165	0,00352	6,11E–06	1,40E–05
cg25903220	0,01264042	0,002586482	1,02E–06	0,04096965	0,00583	0,00125	6,77E–06	1,44E–05
cg19423196	0,03963108	0,008112568	1,03E–06	0,04096965	0,0249	0,00543	9,20E–06	1,78E–05
cg21719704	0,03984661	0,008160612	1,05E–06	0,04096965	0,0199	0,00435	9,48E–06	1,78E–05
cg22856279	0,0522964	0,010732216	1,10E–06	0,04096965	0,0114	0,0025	1,09E–05	1,94E–05
cg01079199	–0,01823898	0,00374623	1,12E–06	0,04096965	–0,0158	0,00366	2,87E–05	4,83E–05
cg12404462	–0,02240805	0,004647295	1,42E–06	0,043594702	–0,0118	0,00277	3,17E–05	5,07E–05
cg05800723	–0,009612961	0,001996907	1,48E–06	0,043594702	0,028	0,00663	4,02E–05	6,13E–05
cg13971124	–0,01245127	0,002602005	1,71E–06	0,043594702	–0,0124	0,00296	4,41E–05	6,41E–05
cg10284656	–0,01361793	0,002846137	1,71E–06	0,043594702	0,0153	0,0037	5,38E–05	7,49E–05
cg20615832	0,04494551	0,009398729	1,73E–06	0,043594702	0,0182	0,00444	6,98E–05	8,95E–05
cg13209192	–0,01388532	0,002904336	1,75E–06	0,043594702	–0,00936	0,00229	6,99E–05	8,95E–05
cg17190420	–0,01263732	0,002643739	1,75E–06	0,043594702	–0,0214	0,00527	7,72E–05	9,49E–05
cg09108969	–0,01300435	0,002724283	1,81E–06	0,043594702	–0,0104	0,00257	8,01E–05	9,49E–05
cg23530133	0,01877393	0,003959268	2,12E–06	0,049151385	–0,0102	0,00253	8,42E–05	9,62E–05
cg21545602	–0,01375361	0,0029097	2,28E–06	0,050431784	–0,0125	0,0031	8,79E–05	9,70E–05
cg04523826	0,02491398	0,005276232	2,34E–06	0,050431784	0,0396	0,00992	1,00E–04	0,000106666666666667
cg17557522	–0,01083295	0,002296594	2,39E–06	0,050431784	0,0358	0,009	1,05E–04	0,000108387096774194
cg07530264	0,02937474	0,006235608	2,47E–06	0,050431784	0,0116	0,00322	3,99E–04	0,000399

PROBEID = probe identifiers; BETA = estimates; SE = standard error; Adjusted P-value = P-value adjusted by multiple test correction; mixed-model= estimate/standard error/ p-value/adjusted p-value after adressing samples from the family in the mixed model

Supplementary Table S6 : Sensitivity analysis with socioeconomic status added in the regression model

PROBEID	Beta	SE	P-value	Adjusted P-C	HR	MAPINFO	Gene	CpG Islands	Relation to CpG Island	SNP ID	SNP_DISTANCE	SNP MAF
cg00719303	-0,01758039	0,003168136	2,91E-08	7,10E-07	20	5391296	BCAS1					
cg02727029	0,01265803	0,002312691	4,44E-08	7,10E-07	15	89482453				rs11473492	14	0,007788
cg00636993	-0,01341825	0,002521135	1,02E-07	1,09E-06	16	3062281 0	ZNF689;ZNF689;ZNF689;ZNF689	chr16:3062 1000-3062 1826	S_Shore			
cg19754387	0,00542396 4	0,001042561	1,97E-07	1,57E-06	2	2085760 57	CCNYL1;CCNYL1	chr2:2085760 01-2085771 06	Island	rs9491 86130	45	0,0032
cg02353444	0,01548261 4	0,003093547	5,69E-07	3,19E-06	6	11810286				rs9530 2013 2;rs9644 9056	0,41	0,00203 5 0,00320
cg08763547	0,01728696 9	0,003478794	6,62E-07	3,19E-06	2	1212 0189	RALB			rs1436 34079	0	0,003199
cg12404462	-0,02246 4231	0,004571033	8,71E-07	3,19E-06	2	2394606 40				rs1860 11166	29	0,0032
cg03272935	0,02369 65778	0,004825499	9,37E-07	3,19E-06	11	6776572 0	UNC93B1	chr11:6776491 9-6776525 6	S_Shore	rs5470 2631 6;rs9639 7193;rs3927 7800;rs6390 7978;rs9918 2617 27	26 1,0 9,2 1	0,00203 0,0 ,00320 0,0 ,00320 0,0 ,00118 9,0 ,00399
cg04317857	0,03127567	0,006387 68	9,78E-07	3,19E-06	7	3971170 3				rs1415 8180 3;rs1179 4535	15 3 1	0,00320 0,0 ,00798 1
cg13209192	-0,01346 0013	0,002756 45	9,97E-07	3,19E-06	12	535 2589				rs0739 4154 1;rs1927 8570 4;rs1 8363 3905;rs7591 1112;rs1 3699 7544	44 3,7 23 ,2 1	0,00200 0,0 ,00399 9,0 ,00320 0,0 ,00320 0,0 ,00458 9
cg25903220	0,01 3008049	0,002675491	1,17E-06	3,21E-06	8	1015 32284	ANKRD46					
cg25012097	-0,01110464	0,002387163	1,20E-06	3,21E-06	13	3926 3863	FREM2	chr13:3926 1236-3926 2299	S_Shore	rs1401 0184 4;rs2691 87880	17 5	0,00139 8,0 ,00200
cg17572622	-0,01079 3039	0,002326899	1,40E-06	3,28E-06	7	1009001 76	NYAP1	chr7:1009011 80-1009015 98	N_Shore	rs1895 0391 2;rs4727 4564;rs6372 1252	1,11 3 7	0,00399 9,0 ,04412 9,0 ,00320 0
cg10179199	-0,01884 3813	0,003908682	1,43E-06	3,28E-06	5	1756444 11				rs6690 3159	1	0,0032
cg08496115	0,02936 56264	0,00436 2018	1,58E-06	3,36E-06	15	5984160 3				rs5479 1816 2;rs5409 11708	41 2	0,00139 8,0 ,00399
cg05800723	-0,00991 0213	0,002093219	2,25E-06	4,49E-06	4	791 03126	FRAS1;FRAS1	chr4:7910 2900-7910 3261	Island	rs6500 1070 3	37	0
cg17 604 20	-0,01291 3693	0,002718641	2,40E-06	4,52E-06	6	1705451 4		chr6:1 7054453 5-1705 471 04	N_Shore	rs3697 3102;rs9346 34087	1,15	0,00599 9,0 ,00320 0
cg04164984	-0,00945 5153	0,002018817	2,82E-06	5,01E-06	17	7235821	PHF12					
cg17 197 04	0,03 850333 3	0,006240732	2,98E-06	5,02E-06	19	5984 8568	SHISA7	chr19:5956 1800-5956 2103	N_Shelf	rs1895 617 55	21	0,001198
cg20615822	0,04 370194	0,008434363	3,53E-06	5,41E-06	4	747 10172	PF4V1	chr4:747 1008 5-747 1933 9	Island	rs1412 9311	25	0,00203
cg13971124	-0,01020 2698	0,002392 46	3,55E-06	5,41E-06	11	1330 82 70	IGSF9B			rs3397 84721	31	0,00799
cg02894956	-0,01373 0256	0,002984845	4,48E-06	6,52E-06	8	6109 6196		chr8:6109 6125-6109 8327	Island			
cg23956279	0,04861 1371	0,010 739423	6,00E-06	8,35E-06	21	4483 0477		chr21:4483 0345-4483 1152	N_Shore	rs9595 7952;rs3741 21752	1,47	0,00799 9,0 ,00139 8
cg02177518	0,02318 2715	0,005 166934	7,23E-06	9,84E-06	18	3195494	DLGAP1;DLGAP1;DLGAP1;DLGAP1;DLGAP1;DLGAP1;DLGAP1;DLGAP1;DLGAP1;DLGAP1			rs5731 74367	1	0,0032
cg09 089659	-0,01238 5079	0,002774936	8,07E-06	1,03E-05	15	1020 34562				rs9618 59890	16	0,001198
cg23303133	0,01 775176 6	0,003997337	8,99E-06	1,10E-05	19	4994 6268	SLC17A7	chr19:4994 4408-4994 6656	S_Shore			
cg07788114	0,02301 4821	0,005 278528	1,31E-05	1,55E-05	8	6864 469	CPA6			rs1840 91018	2	0,0032
cg25302654	0,02799 6079	0,006518523	1,75E-05	2,00E-05	8	1449654 97		chr8:1 4496 6638-1 4496 8880	N_Shore	rs6909 8231;rs1178 2138;rs5880 7321	0;1,26	0,00200 0,0 ,01118 2,0 ,00320 0
cg04534804	-0,01936 7235	0,002 835044	2,08E-05	2,39E-05	8	2354 6913	PEBP4			rs6524 31023	0	0,0032
cg04523826	0,02 459655 5	0,005 851109	2,66E-05	2,84E-05	12	9258 7036				rs7534 2279	48	0,01 7572
cg21546822	-0,01238 1496	0,003 107048	3,08E-05	3,46E-05	9	1395 6747	EGFL7;EGFL7;EGFL7;EGFL7;EGFL7;EGFL7	chr9:1 3956 4899-1395 6528 6	S_Shore	rs3737 8032;rs7858 1911;rs3770 8692;rs2014 8734 6;rs2017 8787 8;rs	0;1,22 ,27 2 9,3 1	0,00200 0,0 ,00339 5,0 ,01118 9,0 ,00203 0,0 ,01003 0,0 ,03394 8
cg19423196	0,03 784708 3	0,009587267	7,89E-05	8,77E-05	10	820 49429	MATN4;MAT1A			rs5752 9283;rs1835 3043;rs1 4011130	8,9,13	0,00203 0,0 ,00320 0,0 ,00398

PROBEID = probe identifiers; BETA = estimates; SE = standard error; Adjusted P-value = P-value adjusted by multiple test correction; CHR = chromosome; MAPINFO = position of the CpG's in the chromosome; Gene = UCSC RefGene; CpG Islands = Nearest UCSC CpG island names; Relation to CpG Island = relation to the nearest UCSC CpG islands; SNP ID = SNP identifiers within the 50 bp of the probe; SNP DISTANCE = Distance of the SNP; SNP MAF = SNP minor allele frequency

Suplementary Table S7: Replication of common significant CpGs in loW cohort

Descriptive characteristics	loW, N=159
Sex n(%)	
Male	82 (52)
Female	77 (48)
Childhood smoke exposure n(%)	
Father smoked	32 (20)
Mother smoked	30 (19)
Father and mother smoked	18 (11)
No parent smoked	38 (24)
Father age at childbirth mean ± SD	26 ± 4.4

PROBEID	Beta	SE	P-value	Adjusted P-va	CHR	MAPINFO	Gene	Relation to	SNP ID	SNP_DISTANCE	SNP MAF
cg05800723	−0,006546462	0,002247443	0,003581475	0,046559175	4	79103126	FRAS1;FRAS1	Island	rs550010703	37	0
cg17190420	−0,002188506	0,002017256	0,277969311	0,999613284	6	170545114		N_Shore	rs369731030;rs534634087	1;15	0.000599;0.000200
cg20272935	0,005844752	0,006401257	0,361209296	0,999613284	11	67765720	UNC93B1	S_Shore	rs547262516;rs565837193;rs539277800;rs539072978;rs569826127	26;10;9;2;1	0.000200;0.000200;0.000200;0.001198;0.000399
cg12404462	−0,00287863	0,004774651	0,546576155	0,999613284	2	239490640			rs186011166	29	0,0002
cg19754387	0,000733648	0,001307717	0,574787819	0,999613284	2	208576057	CCNYL1;CCNYL1	Island	rs549186130	45	0,0002
cg13971124	0,001077702	0,002224357	0,628030724	0,999613284	11	133808270	IGSF9B		rs539784721	31	0,000799
cg25727029	−0,001708086	0,003657751	0,640516222	0,999613284	15	89482453			rs114173492	14	0,007788
cg19423196	0,003487528	0,007603777	0,646480283	0,999613284	10	82049429	MAT1A;MAT1A		rs575292837;rs181353304;rs141011130	8;9;13	0.000200;0.000200;0.000998
cg20615832	0,003516692	0,009971038	0,724320327	0,999613284	4	74719172	PF4V1	Island	rs141256311	25	0,00023
cg05235344	0,000469354	0,003073342	0,878620715	0,999613284	6	11810356	RP3-413H6.2		rs553022132;rs566450556	0;41	0.000200;0.000200
cg04164584	−0,000193384	0,002035979	0,924327921	0,999613284	17	27235821	PHF12				
cg13209192	0,000123529	0,002366765	0,958374896	0,999613284	12	5352389	RP11-319E16.1		rs573941541;rs192795704;rs183633905;rs575911112;rs139597544	44;37;23;2;1	0.000200;0.000599;0.000200;0.000200;0.006589
cg09696115	1,60E−06	0,00330522	0,999613284	0,999613284	15	59841613			rs547318521;rs540911708	41;2	0.001398;0.000399

PROBEID = probe identifiers; BETA = estimates; SE = standard error; Adjusted P-value = P-value adjusted by multiple test correction; CHR = chromosome; MAPINFO = position of the CpGs in the chromosome; Gene = UCSC RefGene; CpG Islands = Nearest UCSC CpG island names; Relation to CpG Island = relation to the nearest UCSC CpG islands; SNP ID = SNP identifiers within the 50 bp of the probe; SNP DISTANCE = Distance of the SNP; SNP MAF = SNP minor allele frequency

Supplementary material

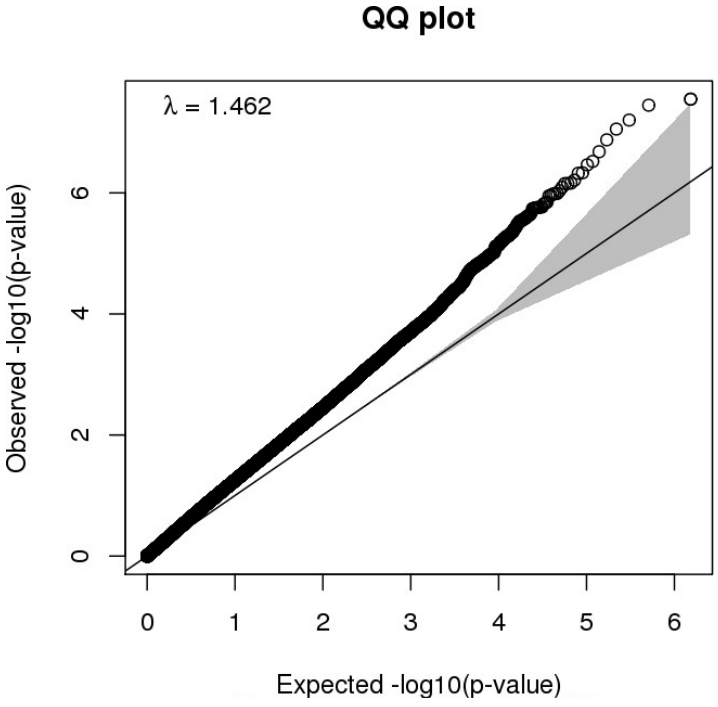
Epigenome-wide association of father's smoking with offspring DNA methylation - A hypothesis-generating study

G.T. Mørkve Knudsen[1,2*], F.I. Rezwan[3*], A. Johannessen[4,2], S.M. Skulstad[2], R.J. Bertelsen[1], F.G. Real[1], S. Krauss-Etschmann[5], V. Patil[6], D. Jarvis[7], S. H. Arshad[8,9], J. W. Holloway[3], C. Svanes[4, 2**]**

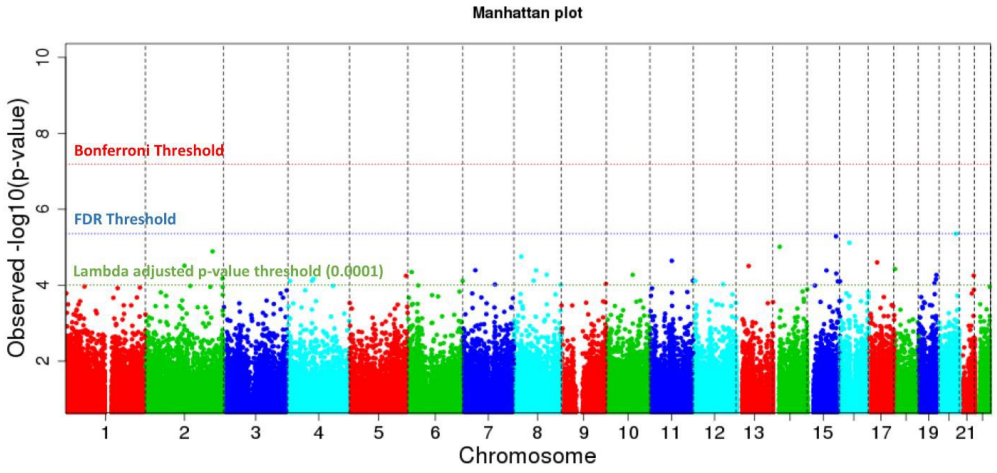
** Equal first authors ** Equal last authors*

[1]Department of Clinical Science, University of Bergen – Bergen (Norway), [2]Department of Occupational Medicine, Haukeland University Hospital, Bergen (Norway), [3]Human Genetics and Genomic Medicine, Human Development and Health, Faculty of Medicine, University of Southampton – Southampton (United Kingdom), [4]Centre for International Health, Department of Global Public Health and Primary Care, University of Bergen – Bergen (Norway), [5]Research Center Borstel, Borstel, Germany. Member of the German Center for Lung Research (DZL) and Institute of Experimental Medicine, Christian-Albrechts-University of Kiel, Kiel, Germany,

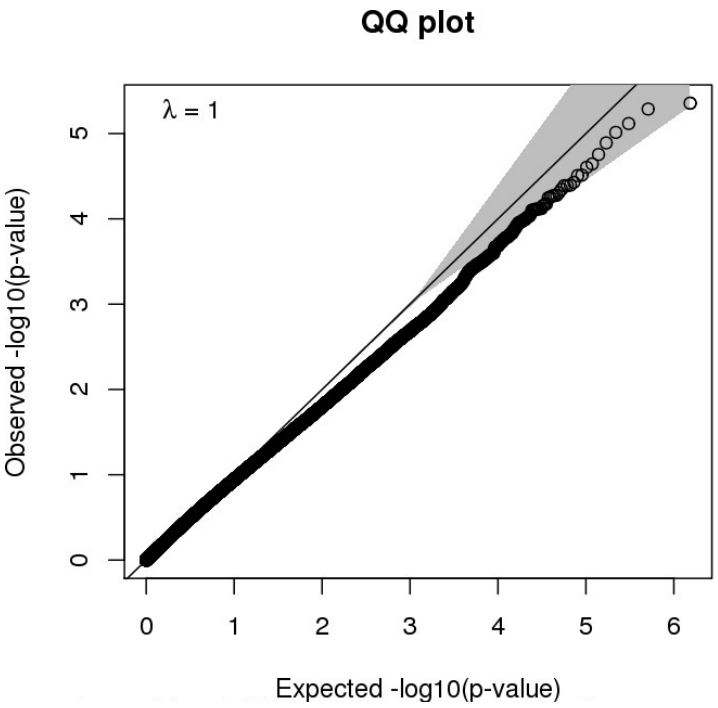
[6]Faculty of Medicine, National Heart & Lung Institute, Imperial College – London (United Kingdom), [7]David Hide Asthma and Allergy Research Centre, St. Mary's Hospital, Isle of Wight, UK,[8]Clinical and Experimental Sciences, University of Southampton, Southampton General Hospital, Southampton, UK. [9]NIHR Respiratory Biomedical Research Unit, University Hospital Southampton, Southampton, UK.



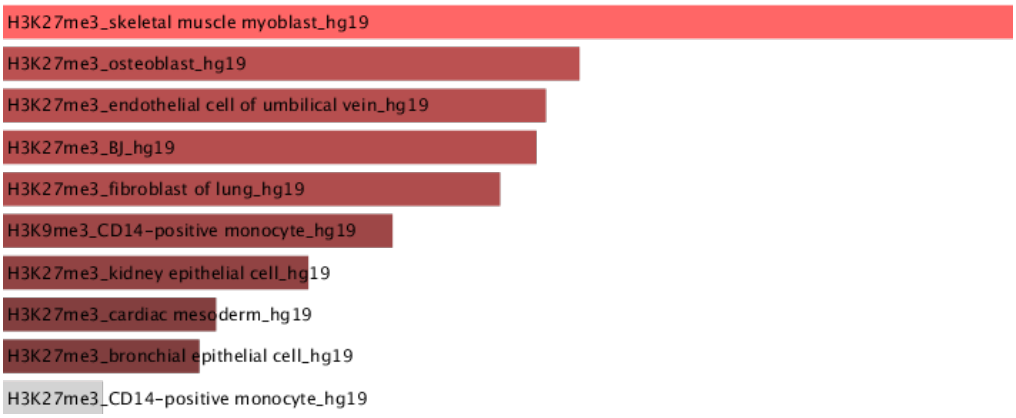
Supplementary Figure S1: **Quantile-quantile (QQ)plot for paternal smoking EWAS before correcting for inflation.** In the plot, the horizontal axis and the vertical axis indicates (−log10 transformed) expected P-values and observed P-values respectively. Here, the lambda (λ) represents genomic inflation factor.



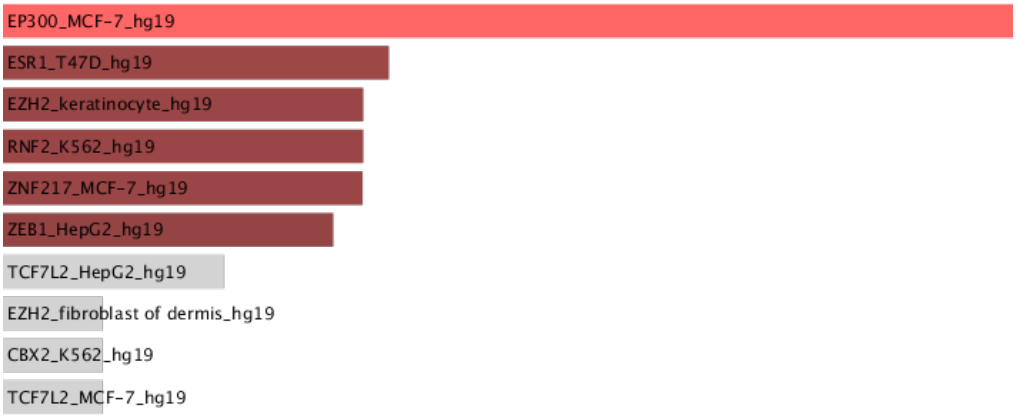
Supplementary Figure S2: Manhattan plot for paternal smoking EWAS after correcting for inflation. Here lambda (λ) represents genomic inflation factor. In the plot, the vertical axis indicates (-log10 transformed) observed P-values, and the horizontal axis indicates chromosome positions with the points indicating individual CpG. Red line: Bonferroni threshold; blue line: Multiple testing correction threshold (FDR < 0.05); and green line: inflation adjusted P-value threshold (< 0.0001).



Supplementary Figure S3: Quantile-quantile (QQ) plot for paternal smoking EWAS after correcting for inflation. Here lambda (λ) represents genomic inflation factor. In the plot, the horizontal axis and the vertical axis indicates (−log10 transformed) expected P-values and observed P-values respectively. Here, the lambda (λ) represents genomic inflation factor.

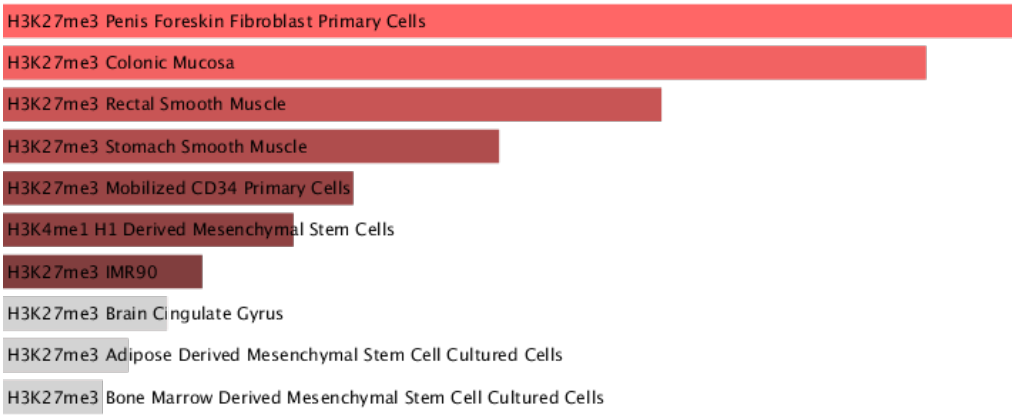


Supplementary Figure S4: Enrichment of ENCODE histone modification marks. Top 10 enriched terms in ENCODE histone modification marks using genes CpGs (threshold: inflation adjusted P-value < 0.0001). The length of the bar represents the significance of that specific gene-set or term. In addition, the brighter the colour, the more significant that term is.



Supplementary Figure S5: Enrichment of ENCODE transcription factor ChIP seq marks. Top 10 enriched terms in ENCODE transcription factor ChIP seq marks using genes CpGs (threshold: inflation adjusted P-

value < 0.0001). The length of the bar represents the significance of that specific gene-set or term. In addition, the brighter the colour, the more significant that term is.



Supplementary Figure S6: Enrichment of Epigenomics Roadmap histone modification and ChIP seq marks. Top 10 enriched terms in Epigenomics Roadmap histone modification and ChIP seq marks using genes CpGs (threshold: inflation adjusted P-value < 0.0001). The length of the bar represents the significance of that specific gene-set or term. In addition, the brighter the color, the more significant that term is.

Paper II

II

PLOS ONE

RESEARCH ARTICLE

Parents' smoking onset before conception as related to body mass index and fat mass in adult offspring: Findings from the RHINESSA generation study

Gerd Toril Mørkve Knudsen[1,2]*, Shyamali Dharmage[3], Christer Janson[4], Michael J. Abramson[5], Brynd´ıs Benediktsdo´ttir[6,7], Andrei Malinovschi[8], Svein Magne Skulstad[2], Randi Jacobsen Bertelsen[1,9], Francisco Gomez Real[1], Vivi Schlu¨nssen[10,11], Nils Oskar Jõgi[1,2,12], Jose´ Luis Sa´nchez-Ramos[13], Mathias Holm[14], Judith Garcia-Aymerich[15,16,17], Bertil Forsberg[18], Cecilie Svanes[2,19‡], Ane Johannessen[19‡]

1 Department of Clinical Science, University of Bergen, Bergen, Norway, **2** Department of Occupational Medicine, Haukeland University Hospital, Bergen, Norway, **3** School of Population and Global Health, The University of Melbourne, Carlton, Australia, **4** Department of Medical Sciences: Respiratory, Allergy and Sleep Research, Uppsala University, Uppsala, Sweden, **5** School of Public Health & Preventive Medicine, Monash University, Melbourne, Australia, **6** Faculty of Medicine, University of Iceland, Reykjavik, Iceland, **7** Department of Sleep Medicine, Landspitali, Reykjavik, Iceland, **8** Department of Medical Sciences: Clinical Physiology, Uppsala University, Uppsala, Sweden, **9** Oral Health Center of Expertise in Western Norway, Hordaland, Bergen, Norway, **10** Department of Public Health, Work, Environment and Health, Danish Ramazzini Centre, Aarhus University Denmark, Aarhus, Denmark, **11** National Research Centre for the Working Environment, Copenhagen, Denmark, **12** Lung Clinic, Tartu University Hospital, Tartu, Estonia, **13** Department of Nursing, Huelva University, Huelva, Spain, **14** Occupational and Environmental Medicine, School of Public Health and Community Medicine, Institute of Medicine, University of Gothenburg, Gothenburg, Sweden, **15** ISGlobal, Barcelona, Spain, **16** Universitat Pompeu Fabra (UPF), Barcelona, Spain, **17** CIBER Epidemiolog´ıa y Salud Pu´blica (CIBERESP), Madrid, Spain, **18** Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden, **19** Centre for International Health, Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway

‡ These authors are joint senior authors on this work.
* gerd.knudsen@uib.no

Citation: Knudsen GTM, Dharmage S, Janson C, Abramson MJ, Benediktsdo´ttir B, Malinovschi A, et al. (2020) Parents' smoking onset before conception as related to body mass index and fat mass in adult offspring: Findings from the RHINESSA generation study. PLoS ONE 15(7): e0235632. https://doi.org/10.1371/journal.pone.0235632

Editor: Seana Gall, University of Tasmania, AUSTRALIA

Received: March 17, 2020

Accepted: June 18, 2020

Published: July 6, 2020

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0235632

Data Availability Statement: Due to Norwegian ethical and legal restrictions the data underlying

Abstract

Emerging evidence suggests that parents' preconception exposures may influence offspring health. We aimed to investigate maternal and paternal smoking onset in specific time windows in relation to offspring body mass index (BMI) and fat mass index (FMI). We investigated fathers (n = 2111) and mothers (n = 2569) aged 39–65 years, of the population based RHINE and ECRHS studies, and their offspring aged 18–49 years (n = 6487, mean age 29.6 years) who participated in the RHINESSA study. BMI was calculated from self-reported height and weight, and FMI was estimated from bioelectrical impedance measures in a subsample. Associations with parental smoking were analysed with generalized linear regression adjusting for parental education and clustering by study centre and family. Interactions between offspring sex were analysed, as was mediation by parental pack years, parental BMI, offspring smoking and offspring birthweight. Fathers' smoking onset before conception of the offspring (onset 15 years) was associated with higher BMI in the offspring when adult (β 0.551, 95%CI: 0.174–0.929, p = 0.004). Mothers' preconception and postnatal smoking onset was associated with higher offspring BMI (onset <15 years:

this study are available upon request to qualified researchers. Requests for data access can be directed to Haukeland University Hospital, 5021 Bergen, Norway. Att. Head of Department. Dept. of Occupational Medicine, Marit Grønning; email: postmottak@helse-bergen.no; phone: +47 55975000. Org, nr. 983 974 724.

Funding: Co-ordination of the RHINESSA study has received funding from the Research Council of Norway (Grants No. 274767, 214123, 228174, 230827 and 273838), ERC StG project BRuSH #804199, the European Union's Horizon 2020 research and innovation program under grant agreement No. 633212 (the ALEC Study WP2), the Bergen Medical Research Foundation, and the Western Norwegian Regional Health Authorities (Grants No. 912011, 911892 and 911631). Study centres have further received local funding from the following: Bergen: the above grants for study establishment and co-ordination, and, in addition, World University Network (RDF and Sustainability grant), Norwegian Labour Inspection, and the Norwegian Asthma and Allergy Association. Albacete and Huelva: SEPAR. Fondo de Investigacio´n Sanitaria (FIS PS09). Gøteborg, Umeå and Uppsala: the Swedish Lung Foundation, the Swedish Asthma and Allergy Association. Reykjavik: Iceland University. Melbourne: NHMRC, Melbourne University, Tartu: the Estonian Research Council (Grant No. PUT562). Århus: The Danish Wood Foundation (Grant No. 444508795), the Danish Working Environment Authority (Grant No. 20150067134). The RHINE study received funding by Norwegian Research Council, Norwegian Asthma and Allergy Association, Danish Lung Association, Swedish Heart and Lung Foundation, Vårdal Foundation for Health Care Science and Allergy Research, Swedish Asthma and Allergy Association, Swedish Lung Foundation, Icelandic Research Council, and Estonian Science Foundation. The co-ordination of ECRHS was supported by European Union's Horizon 2020 research and innovation program under grant agreement No. 633212 (the ALEC Study), the European Commission frameworks 5 and 7 (ECRHS I and II) and the Medical Research Council (ECRHS III).

Competing interests: The authors have declared that no competing interests exist.

β1.161, 95%CI 0.378–1.944; onset 15 years: β0.720, 95%CI 0.293–1.147; onset after offspring birth: β2.257, 95%CI 1.220–3.294). However, mediation analysis indicated that these effects were fully mediated by parents' postnatal pack years, and partially mediated by parents' BMI and offspring smoking. Regarding FMI, sons of smoking fathers also had higher fat mass (onset <15 years β1.604, 95%CI 0.269–2.939; onset 15 years β2.590, 95%CI 0.544–4.636; and onset after birth β2.736, 95%CI 0.621–4.851). There was no association between maternal smoking and offspring fat mass. We found that parents' smoking before conception was associated with higher BMI in offspring when they reached adulthood, but that these effects were mediated through parents' pack years, suggesting that cumulative smoking exposure during offspring's childhood may elicit long lasting effects on offspring BMI.

Background

Maternal smoking during pregnancy plays a significant role in increased risk of obesity and metabolic disorders in the offspring [1–4]. Nicotine and other tobacco constituents cross the placenta, and impair foetal growth [5, 6], which together with determinants such as low birthweight and subsequent rapid postnatal weight gain have been associated with risk of adiposity later in life [4]. Several epidemiological studies also report independent effects of paternal smoking (during pregnancy or postnatal life) associated with greater offspring BMI, body fat distribution and increased risk of overweight in children [7–12]. However, obesity is a complex multifactorial condition with a wide range of determinants, which besides environmental factors, also include behavioural and genetic components.

Recent evidence suggests that the germline cells of the parents might have critical exposure-sensitive periods for triggering epigenetic responses that can affect subsequent offspring's metabolic health and risk of becoming obese [13–15], thus suggesting an epigenetic basis of variation in BMI levels and fat mass. Observations from the Överkalix and ALSPAC cohorts showed that excess food supply and smoking during mid-childhood and pre-pubertal years were associated with metabolic and cardiovascular health, and risk of becoming obese in subsequent generation(s) [16–19]. These findings remain to be successfully replicated, and there exists a possibility of residual confounding due to unmeasured family factors, especially due to the social patterning and inequalities related to smoking behaviour [20, 21]. However, other epidemiological studies have reported adverse offspring outcomes related to paternal exposures in pre-puberty/puberty. Analyses of the RHINESSA, RHINE and ECRHS cohorts found that asthma was more common in offspring with fathers who were obese in puberty [22], as well as in offspring with fathers who smoked in adolescent years [23, 24].

With regard to sex-specific patterns, some studies report no sex differences in offspring BMI in relation to parental smoking [9, 25–27]. Other epidemiological [7, 8, 28] and experimental studies [29–32] indicate more pronounced effects among female offspring. In contrast, the ALSPAC study, reported associations between paternal smoking and increased risk of obesity to be significant only in the sons [16, 19]. Whether sexual dimorphism may be involved in parental transmission of smoking effects on offspring BMI, thus needs further investigation.

The aims of the present study were firstly, to investigate parental smoking onset in specific time windows (onset before 15 years; from age 15 and before conception; after offspring birth) in relation to offspring BMI and, in a subsample, fat mass. Secondly, we aimed to explore whether effects of preconception and early life parental smoking on offspring overweight was

modified by sex of the offspring, and mediated by parental pack years of smoking, parental BMI, offspring smoking and, in a subsample, offspring birthweight.

Methods

Study design and population

We investigated onset of parental smoking in relation to adult offspring BMI, using information from two generations. Data concerning the parent population were obtained from the population-based studies Respiratory Health in Northern Europe study (RHINE, www.rhine.nu) and the European Community Respiratory Health Survey (ECRHS, www.ecrhs.org). Information regarding their offspring were collected in the RHINESSA study (www.rhinessa.net). Medical research committees in each study centre approved the study protocols according to national legislation, and each participant gave written informed consent prior to participation (S1 File).

Parent population

The parent sample comprised subjects originating from the ECRHS postal survey in 1990–94. The participants from seven Northern European study centres (Reykjavik in Iceland, Bergen in Norway, Umea, Uppsala and Gothenburg in Sweden, Aarhus in Denmark, and Tartu in Estonia) were followed up in the RHINE questionnaire study, 10 and 20 years after this baseline survey. At each study wave, postal questionnaire information was collected on lifestyle habits, sociocultural factors, and environmental factors such as childhood and adult exposure to tobacco smoke. A sub-sample was invited for clinical investigation and interview in the ECRHS follow-up studies after 10 and 20 years. For parents in two Spanish centres (Albacete and Huelva) and one Australian centre (Melbourne), information from ECRHS was harmonized with the RHINE data. The questionnaire forms used in ECRHS and RHINE can be found at http://www.ecrhs.org/Quests/ECRHSIImainquestionnaire.pdf and http://rhine.nu/pdf/rhine%20Norway.pdf/ http://rhine.nu/pdf/ECRHS%20II%20Norway.pdf.

A flowchart of the study population is provided in Fig 1.

Offspring population

The RHINESSA study (www.rhinessa.net) includes adult offspring (> 18 years) of parents from seven RHINE study centres in Denmark, Iceland, Norway, Sweden and Estonia, and two Spanish (Huelva and Albacete) and one Australian (Melbourne) ECRHS centres. The offspring answered web-based and/or postal questionnaires in 2013–2015, which were harmonized with the RHINE protocols. Sub-samples of offspring who had parents with available clinical information, were invited for clinical investigation and interview, following standardized protocols harmonized with the ECRHS protocols. The questionnaire form used in the RHINESSA can be found at https://helse-bergen.no/seksjon/RHINESSA/Documents/RHINESSA%20Screening%20questionnaires%20adult%20offspring.pdf.

Exposure: Parental smoking

Parental smoking onset was defined from the questions: *i. "Are you a smoker?" ii. "Are you an ex-smoker?" iii. "If yes "How old were you when you started smoking?"* iiii. "*Smoked for . . . years." iv. "Stopped smoking in [year]"*. Ever-smokers were categorised according to age at smoking initiation (<15 years/ 15 years), and whether smoking started before conception (2 years before offspring birth year) or after the offspring was born (1 year after offspring birth year). Thus, we constructed a four-level exposure variable with the mutually exclusive

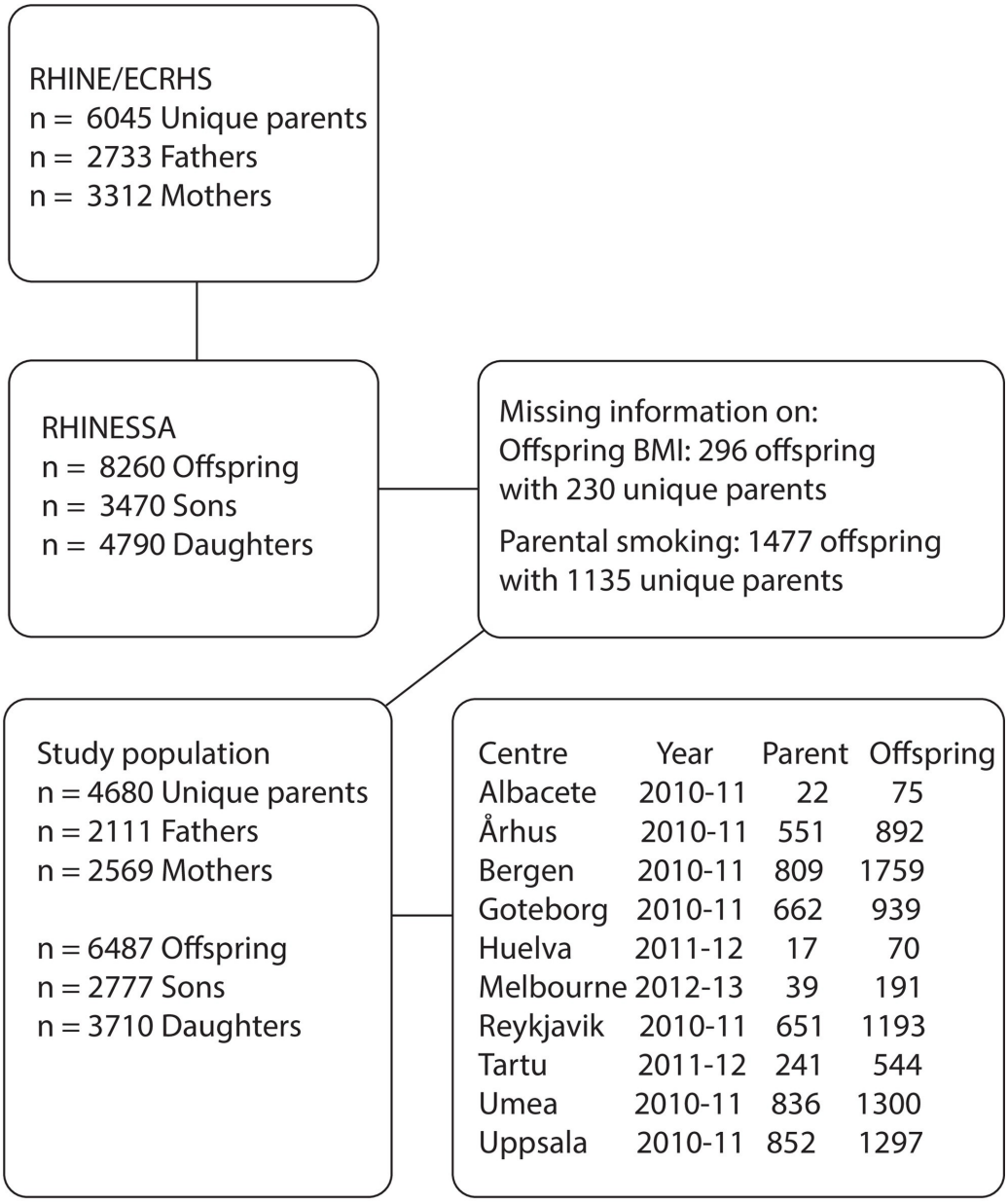


Fig 1. Flow chart of study population. Overview of eligible unique RHINE/ECRHS parents and their RHINESSA offspring, and number excluded due to missing information on offspring's BMI and parental smoking habits.

https://doi.org/10.1371/journal.pone.0235632.g001

categories: never smoked, started smoking before age 15 years, started smoking between age 15 years and conception (preconception), and started smoking after offspring birth (postnatal). Parent-offspring pairs for which parents started smoking during the two-year interval around pregnancy and conception (up to 15 months before conception and up to 1 year after birth of the child) were excluded from the analysis (n = 92).

Outcomes: Offspring body mass index and fat mass index

Body mass index (BMI) was calculated from self-reported height and weight [weight (kg)/ height $(m)^2$]. Body composition and fat mass were estimated from bioelectrical impedance analysis measured using Bodystat 1500 MDD (https://www.bodystat.com/medical/). Fat mass index (FMI) was calculated as fat mass (kg)/height $(m)^2$.

Potential confounders and mediators

Parental/offspring education was used as a proxy for socioeconomic status and categorised as lower (primary school), intermediate (secondary school) or higher (college or university). Parental pack years pre-conception/ from birth until age 18 years were calculated by multiplying the number of 20-packs of cigarettes smoked per day by the number of years the person had smoked up to 2 years before offspring birth year/ up to the offspring's eighteenth birth year. Parental BMI was calculated from self-reported height and weight at RHINE III. Offspring smoking was defined as ever smoking (current/ex-smokers) or never smoking based on the questions *i. "Do you smoke?"* ii. *"Did you smoke previously?"*. Offspring birthweight were obtained from national registry data for a subsample of 813 mother-offspring pairs.

Statistical analysis

Maternal and paternal lines were analysed separately. Generalized linear regressions were used to analyse the associations between parental smoking in specific time windows and offspring BMI (and FMI in a subsample of 240), with adjustment for parental education. Two-dimensional clustering accounted for study centre and family. We set the Heteroscedasticity Consistent Covariance Matrix (HCCM), to version HC1, which made a degree of freedom correction that inflated each residual by the factor $\sqrt{N=\delta N - K\Phi}$.

We tested for interactions between offspring sex and parental smoking onset on offspring BMI; the significance level for interaction effects was set to 0.05. We generated regression models and table/figure outputs by use of the 'jtool' package [33]. We considered other covariates, such as parental age, offspring education, the other parent's smoking habits, and BMI (data on the parent who did not participate in RHINE/ECRHS were obtained from the offspring themselves), to be included in the statistical model, as shown in S1 Fig. However, we did not find these factors likely to confound the relationship between parental smoking onset and offspring BMI, and therefore did not include them in the final models.

We constructed mediation models [34, 35] to investigate whether significant associations between parental smoking onset and offspring BMI were influenced by the following mediators: i. parental pack years, ii. parental BMI, iii. offspring smoking (never-smoked / ever smoked), and iv. offspring birthweight (only available for a subsample of offspring). To investigate whether effects differed by gender, we tested for effect modification by offspring sex. We conducted mediation analysis with the R package "Medflex"[36], embedded within the counterfactual framework, as this provided means to infer and interpret direct and indirect effect estimates in a nonlinear setting. Thus, the total effect of an exposure was decomposed into a natural direct effect (the part of the exposure effect not mediated by a given set of potential

mediators) and natural indirect effect (the part of the exposure effect mediated by a given set of potential mediators). We followed the imputation-based approach for expanding and imputing the data and fitted a working model for the outcome mean. We fitted separate natural effect models, specified with robust standard errors based on the sandwich estimator. We generated confidence interval plots to visualise the effect estimates and their uncertainty.

We performed all analyses using R version 3.5.2, downloaded at the Comprehensive R Archive Network (CRAN) at http://www.R-project.org/.

Results

Of unique fathers, 10% started smoking before age 15 years, 40% started smoking from age 15 years, and 2% started smoking after offspring birth. In the maternal line, 11% started smoking <15 years, 39% started smoking 15 years, and 3% started smoking after offspring birth. Fathers and mothers who started smoking prior to conception had higher current BMI and less education compared to never smoking parents (S1A and S1B Table). In both the paternal (n = 2111) and maternal (n = 2569) lines, daughters had higher education, lower current BMI, and higher FMI, and started smoking earlier compared to sons (Table 1A and 1B). In the maternal line, daughters had lower birthweight. Offspring of smoking parents had higher BMI, more frequently smoked themselves and had smoked more years, compared to offspring of never smoking parents. Sons with fathers who started smoking from age 15 but before conception also had higher FMI than sons with never smoking fathers.

Fathers' smoking onset and offspring BMI and FMI

In unadjusted analyses, father's preconception smoking, both starting before or from age 15 years, was associated with increased offspring BMI (Fig 2). There was no significant interaction between offspring sex and fathers' smoking onset with regard to offspring BMI (p = 0.395). With adjustment for father's education and offspring sex, father's smoking onset 15 years was significantly associated with increased BMI in their adult offspring (Table 2 and Fig 2). However, there was no association between postnatal smoking onset and offspring BMI.

In the subsample with data on FMI, father's preconception and postnatal smoking onset were associated with increased offspring FMI (Table 3 and Fig 3). There were significant differences between sons and daughters, and only sons of fathers' who started to smoke 15 years of age (interaction p = 0.014) or after birth (interaction p = 0.020) had significantly higher FMI compared to sons of never smoking fathers. This trend was not seen among daughters, however, analysis indicated that both sons and daughters of fathers who started to smoke before the age of 15 had higher fat mass (Table 3 and Figs 3 and 4).

Mothers' smoking onset and offspring BMI and FMI

Mother's smoking starting at all time points were associated with increased BMI in her offspring (Table 4 and Fig 5). There were no significant differences between sons and daughters, except that sons of mothers who started to smoke 15 years (interaction p = 0.010) had significantly higher BMI compared to sons of never smoking mothers. There was no such trend among daughters. There was no association with mothers' preconception and postnatal smoking onset and FMI in her offspring (S2 Table).

Mediation analyses of fathers' smoking onset and offspring BMI

For the association of father's smoking onset 15 years with offspring BMI, we analysed mediation by fathers' pack years of smoking, fathers' BMI, and offspring's smoking (Table 5

Table 1. A. Characteristics of 2111 fathers with 2939 sons and daughters. B. Characteristics of 2569 mothers with 3548 sons and daughters.

A

	Sons	**Daughters**	**P-value**
	N = 1255 (43)	N = 1684 (57)	
Paternal characteristics			
Age years, mean ± SD	55.1 ± 6.2	55.0 ± 6.0	p = 0.26
Range	39–65	39–65	
BMI kg/m^2, mean ± SD	26.9 ± 3.8	26.8 ± 3.7	p = 0.32
Range	16.5–53.3	16.8–53.7	
Educational level, n (%)			
Primary	186 (15)	267 (16)	p = 0.70
Secondary	466 (37)	617 (37)	
University/College	588 (47)	792 (47)	
Smoking status, n (%)			
Never smoked	616 (49)	783 (47)	p = 0.20
Preconception <15smoking onset	126 (10)	179 (11)	
Preconception 15 smoking onset	482 (38)	696 (41)	
Postconception smoking onset	31 (3)	26 (2)	
Years smoked, mean ± SD	12.0 ± 15.4	12.4 ± 15.0	p = 0.33
Range	0–59	0–52	
Packyears up to offspring age 18, median	17.4	16.7	p = 0.95
25th%, 75th%	8.0, 27.2	9.9, 25.0	
Packyears preconception years, median	7.0	7.0	p = 0.95
25th%, 75th%	3.8, 12.0	4.0, 11.7	
Age smoking onset, mean ± SD	17.6 ± 5.5	17.0 ± 4.5	p = 0.10
Range	6–53	7–50	
Offspring characteristics			
Age years, mean ± SD	29.5 ± 7.4	29.7 ± 7.3	p = 0.53
Range	18–49	18–50	
BMI kg/m^2, mean ± SD	25.1 ± 4.2	23.8 ± 4.8	p < 0.01
Range	15.8–52.5	14.3–67.2	
FMI fat mass kg/m^2, mean ± SD	4.7 ± 2.9	5.9 ± 2.4	p < 0.01
Range	1.1–11.7	2.5–14.4	
Educational level, n (%)			
Primary	41 (3)	40 (2)	p < 0.01
Secondary	567 (45)	550 (33)	
University/College	644 (51)	1089 (65)	
Smoking status, n (%)			
Never	886 (71)	1174 (70)	p = 0.43
Ever	363 (29)	503 (30)	
Years smoked, mean ± SD	9.4 ± 7.0	9.2 ± 7.0	p = 0.79
Range	0–36	0–33	
Age smoking onset, mean ± SD	16.9 ± 2.9	16.2 ± 2.7	p < 0.01
Range	9–28	10–30	

B

	Sons	**Daughters**	**p-value**
	N = 1522 (43)	N = 2026 (57)	
Maternal characteristics			
Age years, mean ± SD	54.3 ± 6.6	54.1 ± 6.4	p = 0.27

(*Continued*)

Table 1. (Continued)

Range	39–65	39–65	
BMI kg/m^2, mean ± SD	25.5 ± 4.3	25.7 ± 4.6	p = 0.19
Range	14.2–49.3	16.8–65.5	
Educational level, n (%)			
Primary	197 (13)	361 (18)	p < 0.01
Secondary	542 (36)	659 (33)	
University/College	773 (51)	999 (49)	
Smoking status, n (%)			
Never smoked	732 (48)	965 (48)	p = 0.42
Preconception <15smoking onset	154 (10)	232 (12)	
Preconception 15 smoking onset	594 (39)	780 (39)	
Postconception smoking onset	42 (3)	49 (2)	
Years smoked, mean ± SD	11.1 ± 14.3	11.2 ± 14.2	p = 0.79
Range	0–52	0–41	
Packyears up to offspring age 18, median	11.5	12.5	p = 0.45
25^{th0}%, 75^{th0}%	5.8, 18.8	6.0, 19.2	
Packyears preconception years, median	4.2	5.0	p = 0.01
25^{th0}%, 75^{th0}%	2.5, 7.0	3.0, 8.0	
Age smoking onset, mean ± SD	17.3 ± 4.3	17.0 ± 4.0	p = 0.22
Range	9–49	7–44	
Offspring characteristics			
Age years, mean ± SD	31.0 ± 7.8	30.9 ± 7.7	p = 0.49
Range	18–52	18–52	
Birthweight kg, mean ± SD	3.5 ± 0.6	3.4 ± 0.6	p < 0.01
Range	1.1–5.3	0.5–5.3	
BMI kg/m^2, mean ± SD	25.3 ± 3.9	23.8 ± 4.4	p < 0.01
Range	12.7–44.7	14.9–49.0	
FMI fat mass kg/m^2, mean ± SD	4.0 ± 1.7	7.3 ± 4.3	p <0.01
Range	1.0–6.6	3. 0–20.5	
Educational level, n (%)			
Primary	45 (3)	49 (2)	p < 0.01
Secondary	650 (43)	651 (32)	
University/College	826 (54)	1321 (65)	
Smoking status, n (%)			
Never	1023 (67)	1321 (65)	p = 0.32
Ever	493 (32)	699 (35)	
Years smoked, mean ± SD	9.4 ± 7.1	9.7 ± 7.2	p = 0.49
Range	0–37	0–35	
Age smoking onset, mean ± SD	16.6 ± 3.1	16.0 ± 2.7	p < 0.01
Range	7–32	10–36	

Test for sign differences between offspring sex; Wilcoxon Mann Whitney test for continuous variables, chi square and Kruskal Wallis test for categorical variables.
Missing paternal values: Age: NA = 37; BMI: NA = 34; Educational level: NA = 23; Packyears. NA = 836. Missing offspring values: Age: NA = 7, FMI: NA = 2812, Educational level: NA = 8; Smoking status: NA = 13; Years smoked: NA = 72; Age smoking onset: NA = 29.
Missing maternal values: Age: NA = 80; BMI: NA = 85; Educational level: NA = 17; Packyears: NA 868. Missing offspring values: Age: NA = 10; FMI: NA = 3440, Educational level: NA = 6; Smoking status: NA = 12; Years smoked: NA = 63; Age smoking onset: NA = 25. Birthweight only available in subsample n = 813 (335 males and 478 females)

https://doi.org/10.1371/journal.pone.0235632.t001

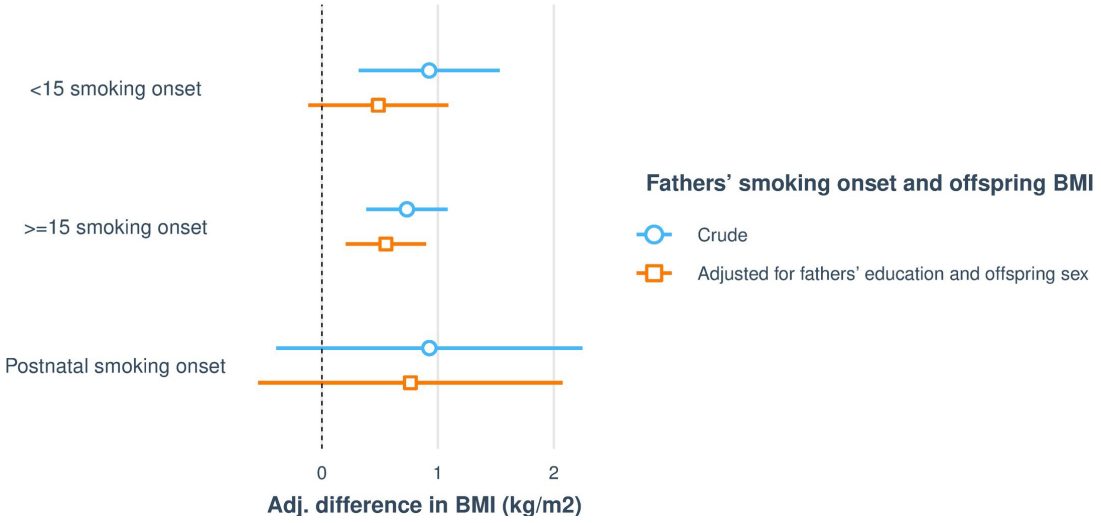


Fig 2. Visualising associations between fathers' smoking onset and offspring (n = 2916) BMI. The figure shows crude regressions and regressions adjusted for fathers' education and offspring sex. After adjustment for fathers' education, fathers' smoking onset 15 remains significantly associated with increased BMI in offspring.

https://doi.org/10.1371/journal.pone.0235632.g002

and S2 Fig). Mediation analysis by fathers' pack years up to offspring age 18 revealed indirect but no direct effect, thus suggesting full mediation of the observed association between fathers' smoking onset 15 years and offspring BMI by fathers' pack years. When restricting analysis to pack years in preconception years only, there was no mediation via fathers' accumulative smoking.

Mediation by fathers' BMI confirmed both a direct effect of fathers' smoking onset 15 years and an indirect effect via fathers' BMI, suggesting partial mediation by fathers' BMI.

Similarly, there was partial mediation of the association between fathers' smoking onset 15 years and offspring obesity by offspring smoking status with both a direct and an indirect effect.

None of the above observed effects were modified by offspring sex.

Table 2. Associations between fathers' smoking onset and offspring (n = 2916) BMI.

	Sons' and daughter's BMI		
Predictors (kg/m²)	***Adj. difference in BMI***	***95% CI***	***P***
Preconception smoking onset < 15 years of age n = 303	0.486	-0.196–1.169	0.162
Preconception smoking onset 15 years of age n = 1162	0.551	0.174–0.929	0.004
Postnatal smoking onset n = 57	0.763	-0.692–2.217	0.304

Estimates from generalized linear regression models with adjustment for offspring sex and fathers' education. Clustered by family id and study centre. P value significance level: .05,

.01, .001.

When adjusting for fathers' education, fathers' smoking onset 15 remains significantly associated with increased BMI in offspring.

https://doi.org/10.1371/journal.pone.0235632.t002

Table 3. Associations between fathers' smoking onset and offspring (n = 129) FMI.

	Sons' and daughter's FMI			
Predictors	***Adj. difference in FMI (fat mass kg/m²)***	***95% CI***	***P***	***Interaction sex P***
Preconception smoking onset < 15 years of age	1.604	0.269–2.939	0.019	0.982
Preconception smoking onset 15 years of age [a]	2.590	0.544–4.636	0.013	0.014
Postnatal smoking onset[b]	2.736	0.621–4.851	0.011	0.020

[a]moking onset 15: daughters β: -2.797, CI: (-5.023, -0.571)

[b] Postnatal smoking onset: daughters β: -3.041, CI: (-5.599, -0.483)

Estimates from generalized linear regression models with offspring sex as interaction term and adjustment for fathers' education.

Clustered by family id and study centre. P value significance level: .05,

.01, .001

https://doi.org/10.1371/journal.pone.0235632.t003

Mediation analyses of mothers' smoking onset and offspring BMI

With regard to the maternal line, there were significant associations of mother's smoking starting <15 years, 15 years, and postnatally, thus, we analysed mediation by mothers' pack years of smoking, mothers' BMI and offspring smoking for each of these associations.

Similarly to the mediation analyses in the paternal line, mediation analysis by mothers' pack years up to offspring's age 18 revealed presence of an indirect but no direct effect, suggesting full mediation of the observed association between mother's preconception smoking onset both before and from 15 years, and offspring BMI (onset <15 years: β: 1.059, p <0.001; onset 15 years β: 0.833, p <0.001; S3 Table and S3 Fig). There was partial mediation of mothers' pack years up to offspring's age 18 on mothers' postnatal smoking onset and offspring BMI where both indirect (β: 0.276, p = 0.001) and direct (β: 1.950, p <0.001) effects were

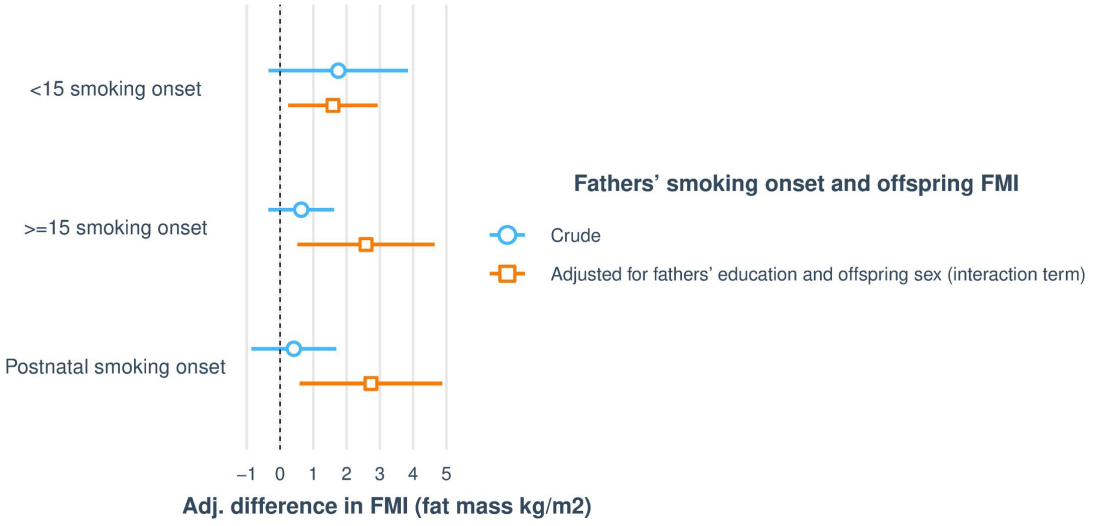


Fig 3. Visualising associations between fathers' smoking onset and offspring (n = 129) FMI. The figure shows crude regressions and regressions adjusted for fathers' education and offspring sex added as an interaction term. In fully adjusted model, fathers' smoking onset at all time points (<15, 15 and after birth) are significantly associated with increased FMI in offspring, but there are significant differences between offspring sex, and only sons of fathers who started to smoke 15 years of age (interaction p = 0.014) or after birth (interaction p = 0.020) had significantly higher FMI compared to sons of never smoking fathers.

https://doi.org/10.1371/journal.pone.0235632.g003

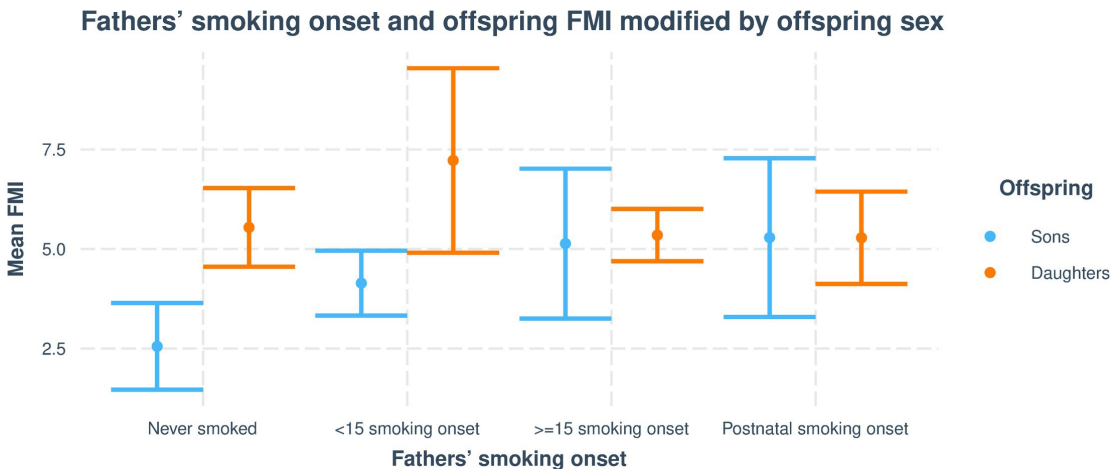


Fig 4. Visualising mean FMI differences in sons and daughters according to fathers' smoking onset. Interaction plot, depicting how offspring sex modify the associations between fathers' 15 and postnatal smoking onset and offspring's FMI.

https://doi.org/10.1371/journal.pone.0235632.g004

significant and pointed in the same direction. We did not find any direct or indirect effects via mothers' preconception pack years (S3 Table and S3 Fig).

Mediation by mothers' BMI confirmed partial mediation with presence of both a direct effect of mothers' preconception smoking onset before 15 years of age (β: 0.551, p = 0.026) as well as smoking onset after birth (β: 1.869, p <0.001), and an indirect effect via mothers' BMI (onset <15: β: 0.334, p <0.001; onset after birth: β: 0.320, p = 0.013). There was no evidence of direct or indirect effects via mothers' BMI in relation to mothers' preconception smoking onset 15 (S4 Table and S4 Fig).

There was indication of partial mediation by offspring's own smoking status, as both direct effects of smoking onset before 15 years of age (β: 0.841, p = 0.001) and smoking onset after birth (β: 2.090, p <0.001), as well as indirect effects via offspring's smoking were present (onset <15 years: β: 0.059, p = 0.016; onset 15 years β: 0.031, p = 0.019; onset after birth: β: 0.129, p = 0.013, S5 Table and S5 Fig).

Table 4. Associations between mothers' smoking onset and offspring (n = 3531) BMI.

	Sons' and daughter's BMI			
Predictors	***β-coef.***	***95% CI***	***P***	***Interaction sex P***
Preconception smoking onset < 15 years of age	1.161	0.378–1.944	0.004	0.338
Preconception smoking onset 15 years of age [a]	0.720	0.293–1.147	0.001	0.010
Postnatal smoking onset	2.257	1.220–3.294	<0.001	0.952

[a] Smoking onset 15: daughters β: -0.717, CI: (-1.264, -0.170)

Estimates from generalized linear regression with offspring sex as interaction term and adjustment for mothers' education. Clustered by family id and study centre. P value significance level: .05,

.01,

.001

https://doi.org/10.1371/journal.pone.0235632.t004

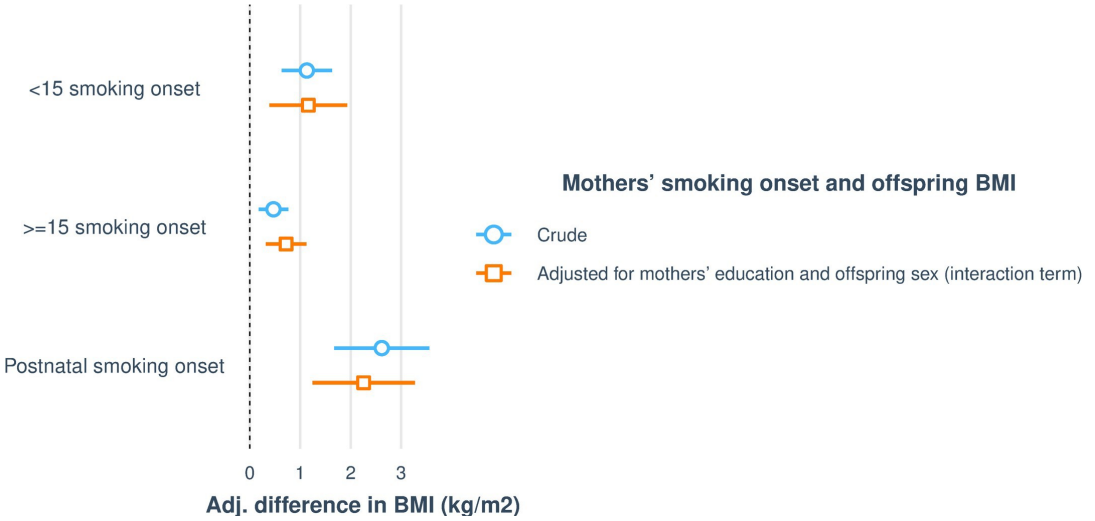


Fig 5. Visualising associations between mothers' smoking onset and offspring (n = 3531) BMI. The figure shows crude and adjusted regressions, with adjustment for mothers' education and offspring sex added as interaction term. In fully adjusted model, mothers' smoking onset at all time points (<15, 15 and after birth) are significantly associated with increased BMI in offspring, but with significant differences between offspring sex. Only sons of mothers who started to smoke 15 years (interaction p = 0.010) had significantly higher BMI compared to sons of never smoking mothers.

https://doi.org/10.1371/journal.pone.0235632.g005

In a subsample with birth weight data, there was no evidence of mediation by offspring birthweight as only a direct effect of mothers' smoking onset <15 years on offspring BMI were present (S6 Table and S6 Fig).

None of the above observed effects were modified by offspring sex.

Discussion

Father's smoking starting before conception was associated with higher BMI in his adult offspring. Bioimpedance measurements for a subsample also found that sons of smoking fathers, starting both before conception and during postnatal years, had higher fat mass, thus suggesting a consistent effect on sons' body composition. Mother's preconception and postnatal smoking onset was also associated with higher adult BMI in her offspring, but these associations were not supported by fat mass analysis in a subsample. Mediation analyses showed that the observed associations between parents' preconception smoking onset and offspring BMI were fully mediated via parents' postnatal pack years. Furthermore, parents BMI and offspring's own smoking status partially mediated the effects of parents' smoking onset on offspring BMI.

To our knowledge, this is the first study that has shown consistently higher BMI and fat mass levels in offspring of smoking fathers' where the offspring has reached adulthood. Our results further suggest that fathers' smoking may have more pronounced effects on their sons' fat mass when compared to daughters. A potential sex-specific effect on offspring's body composition supports previous reports of particularly paternal smoking trajectories to impact on sons' fat mass and risk of becoming obese [16, 19]. However, in contrast to findings in the ALSPAC study, where only fathers' smoking in mid-childhood and pre-pubertal years was associated with increased BMI and fat mass in the sons [19], our study indicate that father's

Table 5. Mediation of the observed association between fathers' 15 smoking onset and offspring BMI.

Causal mediation analysis father offspring				
Fathers' smoking onset	*Adj diff. BMI (kg/m²)*	*Std. error*	*z value*	*P value*
A) Mediation by fathers' packyears up to offspring age 18				
Preconception smoking onset 15				
Natural direct effect	0.240	0.318	0.756	0.450
Natural indirect effect	0.482	0.239	2.014	0.044
Total effect	0.722	0.237	3.047	0.002
Interaction by offspring sex: 0.209				
B) Mediation by fathers' preconception packyears				
Preconception smoking onset 15				
Natural direct effect	0.677	0.235	2.879	0.004
Natural indirect effect	- 0.092	0.130	- 0.708	0.479
Total effect	0.585	0.205	2.848	0.004
Interaction by offspring sex: 0.913				
C) Mediation by fathers' BMI				
Preconception smoking onset 15				
Natural direct effect	0.367	0.170	2.159	0.031
Natural indirect effect	0.214	0.053	4.058	< 0.001
Total effect	0.582	0.178	3.264	0.001
Interaction by offspring sex: 0.528				
D) Mediation by offspring smoking status				
Preconception smoking onset 15				
Natural direct effect	0.488	0.180	2.711	0.007
Natural indirect effect	0.080	0.028	2.900	0.004
Total effect Interaction by offspring sex: 0.134	0.568	0.177	3.215	0.001

Effect decomposition on the scale of the linear predictor with standard errors based on the sandwich estimator. Conditional on fathers' educational level and offspring sex. P value significance level:
.05,
.01,
.001

https://doi.org/10.1371/journal.pone.0235632.t005

preconception smoking starting both before and from age 15 years were associated with increased fat mass in his adult sons. This was also seen in sons where fathers started to smoke after birth. This may reflect the direct toxicogenic effects cigarette smoke exert on biological processes involved in metabolic health. Previous studies have found germ cells and elevated reactive oxygen species (ROS) to mediate metabolic phenotypes in offspring [29, 31, 37, 38]. Smoking has also been shown to induce both ROS overproduction as well as epigenetic changes to germ cells [29, 39], which adds biological plausibility of paternal smoking to be drivers of complex offspring phenotypes. Although increased adipose tissue does not necessarily translate into metabolic abnormalities, both BMI and FMI are regarded important determinants of metabolic health at the population level [40, 41], and childhood adiposity has been reported to be associated with increased risk of adult type 2 diabetes mellitus [42]. In a recent epigenome-wide association study, we found that adult offspring with smoking fathers had differential methylation in regions related to innate immune system pathways and fatty acid biosynthesis [43]. These are inflammatory signalling pathways and metabolic signals that have been linked to obesity [44]. However, whether the observed associations between increased BMI and FMI among offspring of smoking fathers relate to metabolic phenotypes needs

further investigation. Our study also indicated that parental smoking exposures transmit through the maternal line, as also mothers' pre- and postnatal smoking onset was related to higher BMI in her adult sons and daughters. However, offspring of smoking mothers did not have a higher fat mass. This may suggest that maternal and paternal smoking trajectories influence their offspring body composition and risk of obesity through different biological mechanisms and pathways.

Through independent mediation analyses, we sought to investigate how parental smoking onset may influence offspring BMI. By including parental pack years as a potential mediator, we aimed to disentangle the effect of parents' smoking onset, and specifically smoking onset before conception, from an accumulative and sustained smoking exposure during peri-and post- natal life. Our findings show that parents' smoking onset influence their offspring BMI via pack years smoked during childhood years, up to the offspring's age 18. This may very well reflect the importance of families' shared environment and the impact lifestyle-related factors, such as dietary habits and physical activity, exert on BMI levels and risk of obesity [45, 46]. This may also explain why both fathers' preconception as well as postnatal smoking onset was associated with increased fat mass in their sons, and why we did not find preconception pack years to mediate the association between parents' smoking onset and offspring BMI.

Furthermore, we found that parents' BMI, partially mediated the effect of pre- and postnatal smoking onset on offspring BMI. Although this may indicate a genetic contribution in body composition, we also found that offspring's smoking status partially mediated the effect of parents smoking onset on their adult BMI, where offspring who had or were smoking themselves, tended to have higher BMI in adulthood compared to offspring who had never smoked. As such, our results may reflect the influence of multiple pathways and the complex interplay between genetics, biology, behaviour, and environment, potentially involved in the aetiology of obesity [47, 48]. These multifactorial aspects may also explain why our results contrast from previous studies related to offspring asthma outcomes in the RHINESSA, RHINE and ECRHS cohorts, where the fathers' pubertal and adolescent years specifically have been shown to constitute an important time window for transmission of paternal lineage exposures [22–24].

Low birthweight due to growth restriction during pregnancy is one factor that has been thought to be on the causal pathway between maternal smoking and offspring's risk of obesity in later life [4]. We found no evidence that the association between mothers' smoking onset and offspring BMI was mediated via her sons' or daughters' birthweight. However, the present study was not able to distinguish true growth retarded newborns from those being born small due to genetic factors, thus a potential causal role of birthweight on overweight in subsequent years warrants further investigation.

A strength of the present study was that the study population originated from two linked inter-generational study cohorts that enables long-term investigation of exposures, across generations and in adult offspring. Further, we used multinational data following standardized protocols. The study also had clear limitations. The main outcome, offspring BMI, was based on self-reported height and weight which can possibly add bias to our estimates. However, we would expect this potential bias to be non-differential, since offspring of smoking and never smoking parents assessed their height and weight in the same manner. There is no reason to believe that offspring of smoking parents would report height and weight any differently than offspring of non-smoking parents. Moreover, studies assessing the validity of self-reported measurements of anthropometric characteristics, have showed that the correlation between self-reported and technician-measured BMI is high (0.92) [49]. Although BMI does not distinguish between lean and fat mass, it is commonly used to determine overweight in clinical research settings as it is closely related with body fat [50, 51]. In addition, we verified our findings in a sub-sample of sons and daughters with clinical data on fat mass. However, this sub-

sample was of limited size, and we did not have sufficient statistical power to conduct mediation analyses of the observed associations between fathers' smoking onset and offspring fat mass. With regard to smoking exposure, we had information only on the participating parent, and have thus not been able to account for a potential smoking exposure arising from the other parent in the household. Neither do we have detailed information about where the parents smoked (inside house/outside house/other places), thus we have not been able to address levels of cigarette smoke the offspring would have been exposed to. Furthermore, we excluded parent-offspring pairs with missing information on parental smoking (n = 1477), which consequently reduced our sample size. Some of the parental smoking onset categories were also limited in numbers, which potentially could influence the reliability of our results. A multitude of exposures and difference in genetic background exists in population studies, and as the offspring in the present study have reached adulthood, they have been exposed to a variety of environmental factors. However, to be regarded as potential confounders, they would per definition precede both the exposure (parental smoking onset in adolescent and early adult years) and outcome (adult offspring BMI) in time. Thus, this does rule out many factors that traditionally would be included in models assessing associations with BMI in adults. We investigated whether parents' adult BMI mediated the effect of parental smoking onset on offspring BMI. However, we did not have information on parents BMI in childhood and pre-adolescent years, which potentially can be of importance and a potential confounder as this would precede both the exposure and outcome in time. Moreover, we did not have information regarding adoption in the offspring population, and whether the participating parent was the biological parent. Thus, unmeasured factors may have impacted on our findings. We chose to use a mediation analysis embedded within the counterfactual framework due to its flexibility in handling non- linear parametric models. However, we have not been able to assess the robustness of our findings and investigated whether there are violations to the identification assumptions, especially with regard to all potential variables being independent and accounted for. This should be further investigated.

Conclusion

In this multicentre population-based study of two generations, we found that fathers beginning to smoke before conception was associated with higher BMI in their adult sons and daughters, and that father's smoking starting in any time window was associated with higher FMI in adult sons. In contrast, mothers' pre-as well as postnatal smoking onset was associated with higher offspring adult BMI, but not higher fat mass. Independent mediation analysis indicated that parents' pack years up to offspring's age 18, but not preconception pack years, fully mediated these effects. This may suggest that an accumulative smoking exposure during offspring's childhood may be needed in order to elicit long lasting effects on offspring BMI and risk of becoming obese. In addition, we found partial mediation by parents' BMI and offspring own smoking status, which may further reflect the importance of families' shared environment and the impact lifestyle-related factors, such as dietary habits and physical activity, exert on BMI levels and risk of obesity. As such, our results support the multifactorial aspects contributing to obesity.

Supporting information

S1 Fig. Directed Acyclic Graph (DAG). The figure presents covariates considered to be included in the statistical model.
(TIF)

S2 Fig. Visualising mediations of the association with fathers' 15 smoking onset on offspring BMI. Analyses reveal full mediation by fathers' pack years and partial mediation by fathers' BMI and offspring's own smoking status. There is no mediation via fathers' preconception accumulative smoking.
(TIF)

S3 Fig. Visualising mothers' pack years as mediator of the observed associations between mothers' smoking onset and offspring BMI.
(TIF)

S4 Fig. Visualising mothers' BMI as mediator of the observed associations between mothers' smoking onset and offspring BMI.
(TIF)

S5 Fig. Visualising offspring's smoking habits as mediator of the observed associations between mothers' smoking onset and offspring BMI.
(TIF)

S6 Fig. Visualising offspring's birthweight as mediator of the observed associations between mothers' smoking onset and offspring BMI.
(TIF)

S1 Table. A. Descriptive table of father offspring cohort grouped by fathers' smoking onset and stratified by offspring sex. B. Descriptive table of mother offspring cohort grouped by mothers' smoking onset and stratified by offspring sex. Parents who started smoking prior to conception have higher current BMI and less education compared to never smoking parents. Offspring of smoking parents have higher BMI, more frequently smoke themselves and have smoked more years, compared to offspring of never smoking parents. Sons with fathers who started smoking from age 15 but before conception also have higher FMI than sons with never smoking fathers.
(PDF)

S2 Table. Associations between mothers' smoking onset and offspring (n = 111) FMI. The figure shows regression model adjusted for mothers' education and offspring sex and reveals no association with mothers' preconception and postnatal smoking onset and FMI in her offspring.
(PDF)

S3 Table. Mothers' pack years as mediator of the observed associations between mothers' smoking onset and offspring BMI. The association between mothers' preconception smoking onset and offspring BMI is fully mediated by mothers' postnatal pack years, whereas mothers' postnatal smoking onset and offspring BMI is partially mediated by mothers' postnatal pack-years. There is no evidence of direct or indirect effects via mothers' preconception accumulative smoking in relation to mothers' smoking onset.
(PDF)

S4 Table. Mothers' BMI as mediator of the observed associations between mothers' smoking onset and offspring BMI. The association between mothers' preconception smoking onset before 15 years of age as well as smoking onset after birth and offspring BMI is partially mediated by mothers' BMI. There is no evidence of direct or indirect effects via mothers' BMI in relation to mothers' preconception smoking onset 15.
(PDF)

S5 Table. Offspring's smoking habits as mediator of the observed associations between mothers' smoking onset and offspring BMI. The association between mothers' preconception smoking onset before 15 years of age as well as smoking onset after birth and offspring BMI is partially mediated by offspring's own smoking status.
(PDF)

S6 Table. Offspring's birthweight as mediator of the observed associations between mothers' smoking onset and offspring BMI. In a subsample with birth weight data, there is no evidence of mediation by offspring birthweight.
(PDF)

S1 File. Table of ethic committee name and approval number for each study center.
(PDF)

Author Contributions

Conceptualization: Gerd Toril Mørkve Knudsen, Cecilie Svanes, Ane Johannessen.

Data curation: Shyamali Dharmage, Christer Janson, Michael J. Abramson, Brynd´ıs Benediktsdo´ttir, Andrei Malinovschi, Randi Jacobsen Bertelsen, Francisco Gomez Real, Vivi Schlünssen, Nils Oskar Jõgi, Jose´ Luis Sa´nchez-Ramos, Mathias Holm, Judith Garcia-Aymerich, Bertil Forsberg, Cecilie Svanes, Ane Johannessen.

Formal analysis: Gerd Toril Mørkve Knudsen, Cecilie Svanes, Ane Johannessen.

Funding acquisition: Shyamali Dharmage, Christer Janson, Brynd´ıs Benediktsdo´ttir, Randi Jacobsen Bertelsen, Francisco Gomez Real, Vivi Schlünssen, Jose´ Luis Sa´nchez-Ramos, Cecilie Svanes.

Investigation: Gerd Toril Mørkve Knudsen, Shyamali Dharmage.

Methodology: Gerd Toril Mørkve Knudsen, Shyamali Dharmage, Judith Garcia-Aymerich, Cecilie Svanes, Ane Johannessen.

Project administration: Cecilie Svanes, Ane Johannessen.

Supervision: Shyamali Dharmage, Svein Magne Skulstad, Cecilie Svanes, Ane Johannessen.

Visualization: Gerd Toril Mørkve Knudsen.

Writing – original draft: Gerd Toril Mørkve Knudsen, Cecilie Svanes, Ane Johannessen.

Writing – review & editing: Gerd Toril Mørkve Knudsen, Shyamali Dharmage, Christer Janson, Michael J. Abramson, Brynd´ıs Benediktsdo´ttir, Andrei Malinovschi, Svein Magne Skulstad, Randi Jacobsen Bertelsen, Francisco Gomez Real, Vivi Schlünssen, Nils Oskar Jõgi, Jose´ Luis Sa´nchez-Ramos, Mathias Holm, Judith Garcia-Aymerich, Bertil Forsberg, Cecilie Svanes, Ane Johannessen.

References

1. Riedel C, Schonberger K, Yang S, et al. Parental smoking and childhood obesity: higher effect estimates for maternal smoking in pregnancy compared with paternal smoking—a meta-analysis. International journal of epidemiology. 2014; 43(5):1593–606. Epub 2014/08/01. https://doi.org/10.1093/ije/dyu150 PMID: 25080528.
2. Ino T. Maternal smoking during pregnancy and offspring obesity: meta-analysis. Pediatrics international: official journal of the Japan Pediatric Society. 2010; 52(1):94–9. Epub 2009/04/30. https://doi.org/10.1111/j.1442-200X.2009.02883.x PMID: 19400912.

3. Ino T, Shibuya T, Saito K, et al. Relationship between body mass index of offspring and maternal smoking during pregnancy. International journal of obesity (2005). 2012; 36(4):554–8. Epub 2011/12/21. https://doi.org/10.1038/ijo.2011.255 PMID: 22184058.

4. Oken E, Levitan EB, Gillman MW. Maternal smoking during pregnancy and child overweight: systematic review and meta-analysis. International journal of obesity (2005). 2008; 32(2):201–10. Epub 2008/02/19. https://doi.org/10.1038/sj.ijo.0803760 PMID: 18278059; PubMed Central PMCID: PMC2586944.

5. Lisboa PC, de Oliveira E, de Moura EG. Obesity and endocrine dysfunction programmed by maternal smoking in pregnancy and lactation. Frontiers in physiology. 2012; 3:437. Epub 2012/11/28. https://doi.org/10.3389/fphys.2012.00437 PMID: 23181022; PubMed Central PMCID: PMC3500832.

6. Schneider S, Huy C, Schutz J, et al. Smoking cessation during pregnancy: a systematic literature review. Drug and alcohol review. 2010; 29(1):81–90. Epub 2010/01/19. https://doi.org/10.1111/j.1465-3362.2009.00098.x PMID: 20078687.

7. Durmus B, Heppe DH, Taal HR, et al. Parental smoking during pregnancy and total and abdominal fat distribution in school-age children: the Generation R Study. International journal of obesity (2005). 2014; 38(7):966–72. Epub 2014/01/23. https://doi.org/10.1038/ijo.2014.9 PMID: 24448598.

8. Leary SD, Smith GD, Rogers IS, et al. Smoking during pregnancy and offspring fat and lean mass in childhood. Obesity (Silver Spring, Md). 2006; 14(12):2284–93. Epub 2006/12/26. https://doi.org/10.1038/oby.2006.268 PMID: 17189557; PubMed Central PMCID: PMC1890311.

9. Kwok MK, Schooling CM, Lam TH, et al. Paternal smoking and childhood overweight: evidence from the Hong Kong "Children of 1997". Pediatrics. 2010; 126(1):e46–56. Epub 2010/07/01. https://doi.org/10.1542/peds.2009-2642 PMID: 20587672.

10. Hui LL, Nelson EA, Yu LM, et al. Risk factors for childhood overweight in 6- to 7-y-old Hong Kong children. International journal of obesity and related metabolic disorders: journal of the International Association for the Study of Obesity. 2003; 27(11):1411–8. Epub 2003/10/24. https://doi.org/10.1038/sj.ijo.0802423 PMID: 14574354.

11. von Kries R, Bolte G, Baghi L, et al. Parental smoking and childhood obesity—is maternal smoking in pregnancy the critical exposure? International journal of epidemiology. 2008; 37(1):210–6. Epub 2007/12/07. https://doi.org/10.1093/ije/dym239 PMID: 18056122.

12. Harris HR, Willett WC, Michels KB. Parental smoking during pregnancy and risk of overweight and obesity in the daughter. International journal of obesity (2005). 2013; 37(10):1356–63. Epub 2013/06/06. https://doi.org/10.1038/ijo.2013.101 PMID: 23736356; PubMed Central PMCID: PMC3795801.

13. Panzeri I, Pospisilik JA. Epigenetic control of variation and stochasticity in metabolic disease. Molecular metabolism. 2018. Epub 2018/06/18. https://doi.org/10.1016/j.molmet.2018.05.010 PMID: 29909200; PubMed Central PMCID: PMC6034039.

14. Soubry A. POHaD: why we should study future fathers. Environmental epigenetics. 2018; 4(2):dvy007. Epub 2018/05/08. https://doi.org/10.1093/eep/dvy007 PMID: 29732171; PubMed Central PMCID: PMC5920283.

15. Sales VM, Ferguson-Smith AC, Patti M-E. Epigenetic Mechanisms of Transmission of Metabolic Disease across Generations. Cell Metabolism. 2017; 25(3):559–71. https://doi.org/10.1016/j.cmet.2017.02.016 PMID: 28273478

16. Pembrey ME, Bygren LO, Kaati G, et al. Sex-specific, male-line transgenerational responses in humans. European journal of human genetics: EJHG. 2006; 14(2):159–66. Epub 2006/01/05. https://doi.org/10.1038/sj.ejhg.5201538 PMID: 16391557.

17. Bygren LO, Kaati G, Edvinsson S. Longevity determined by paternal ancestors' nutrition during their slow growth period. Acta biotheoretica. 2001; 49(1):53–9. Epub 2001/05/23. https://doi.org/10.1023/a:1010241825519 PMID: 11368478.

18. Kaati G, Bygren LO, Edvinsson S. Cardiovascular and diabetes mortality determined by nutrition during parents' and grandparents' slow growth period. European journal of human genetics: EJHG. 2002; 10(11):682–8. Epub 2002/10/31. https://doi.org/10.1038/sj.ejhg.5200859 PMID: 12404098.

19. Northstone K, Golding J, Davey Smith G, et al. Prepubertal start of father's smoking and increased body fat in his sons: further characterisation of paternal transgenerational responses. European journal of human genetics: EJHG. 2014; 22(12):1382–6. Epub 2014/04/03. https://doi.org/10.1038/ejhg.2014.31 PMID: 24690679; PubMed Central PMCID: PMC4085023.

20. Carslake D, Pinger PR, Romundstad P, et al. Early-Onset Paternal Smoking and Offspring Adiposity: Further Investigation of a Potential Intergenerational Effect Using the HUNT Study. PloS one. 2016; 11(12):e0166952. Epub 2016/12/03. https://doi.org/10.1371/journal.pone.0166952 PMID: 27911909; PubMed Central PMCID: PMC5135283.

21. Disney G, Gurrin L, Aitken Z, et al. Hierarchical Models for International Comparisons: Smoking, Disability, and Social Inequality in 21 European Countries. Epidemiology (Cambridge, Mass). 2020; 31(2):282–9. Epub 2019/12/24. https://doi.org/10.1097/ede.0000000000001154 PMID: 31868828.

22. Johannessen A, Lonnebotn M, Calciano L, et al. Being overweight in childhood, puberty, or early adulthood: Changing asthma risk in the next generation? The Journal of allergy and clinical immunology. 2020; 145(3):791-9.e4. Epub 2019/09/11. https://doi.org/10.1016/j.jaci.2019.08.030 PMID: 31505189.

23. Svanes C, Koplin J, Skulstad SM, et al. Father's environment before conception and asthma risk in his children: a multi-generation analysis of the Respiratory Health In Northern Europe study. International journal of epidemiology. 2017; 46(1):235–45. Epub 2016/08/28. https://doi.org/10.1093/ije/dyw151 PMID: 27565179.

24. Accordini S, Calciano L, Johannessen A, et al. A three-generation study on the association of tobacco smoking with asthma. International journal of epidemiology. 2018; 47(4):1106–17. Epub 2018/03/14. https://doi.org/10.1093/ije/dyy031 PMID: 29534228; PubMed Central PMCID: PMC6124624.

25. Yang S, Decker A, Kramer MS. Exposure to parental smoking and child growth and development: a cohort study. BMC pediatrics. 2013; 13:104. Epub 2013/07/12. https://doi.org/10.1186/1471-2431-13-104 PMID: 23842036; PubMed Central PMCID: PMC3717101.

26. Toschke AM, Koletzko B, Slikker W, et al. Childhood obesity is associated with maternal smoking in pregnancy. European journal of pediatrics. 2002; 161(8):445–8. Epub 2002/08/13. https://doi.org/10.1007/s00431-002-0983-z PMID: 12172830.

27. Toschke AM, Montgomery SM, Pfeiffer U, et al. Early intrauterine exposure to tobacco-inhaled products and obesity. American journal of epidemiology. 2003; 158(11):1068–74. Epub 2003/11/25. https://doi.org/10.1093/aje/kwg258 PMID: 14630602.

28. Vik T, Jacobsen G, Vatten L, et al. Pre- and post-natal growth in children of women who smoked in pregnancy. Early human development. 1996; 45(3):245–55. Epub 1996/07/19. https://doi.org/10.1016/0378-3782(96)01735-5 PMID: 8855398.

29. Lane M, McPherson NO, Fullston T, et al. Oxidative stress in mouse sperm impairs embryo development, fetal growth and alters adiposity and glucose regulation in female offspring. PLoS One. 2014; 9(7):e100832. Epub 2014/07/10. https://doi.org/10.1371/journal.pone.0100832 PMID: 25006800; PubMed Central PMCID: PMC4089912.

30. Fullston T, Ohlsson Teague EM, Palmer NO, et al. Paternal obesity initiates metabolic disturbances in two generations of mice with incomplete penetrance to the F2 generation and alters the transcriptional profile of testis and sperm microRNA content. FASEB journal: official publication of the Federation of American Societies for Experimental Biology. 2013; 27(10):4226–43. Epub 2013/07/13. https://doi.org/10.1096/fj.12-224048 PMID: 23845863.

31. Ng SF, Lin RC, Laybutt DR, et al. Chronic high-fat diet in fathers programs beta-cell dysfunction in female rat offspring. Nature. 2010; 467(7318):963–6. Epub 2010/10/22. https://doi.org/10.1038/nature09491 PMID: 20962845.

32. Kwong WY, Wild AE, Roberts P, et al. Maternal undernutrition during the preimplantation period of rat development causes blastocyst abnormalities and programming of postnatal hypertension. Development (Cambridge, England). 2000; 127(19):4195–202. Epub 2000/09/08. PMID: 10976051.

33. Long JA. jtools: Analysis and Presentation of Social Scientific Data. 2019.

34. MacKinnon DP, Luecken LJ. How and for whom? Mediation and moderation in health psychology. Health Psychol. 2008; 27(2s):S99–s100. Epub 2008/05/03. https://doi.org/10.1037/0278-6133.27.2(Suppl.).S99 PMID: 18377161; PubMed Central PMCID: PMC2821200.

35. Cohen J CP, West SG, Aiken LA. Applied multiple regression/correlation analysis for the behavioral sciences. 3. Hillsdale N, editor 2003.

36. Johan Steen TL, Beatrijs Moerkerke, Stijn Vansteelandt medflex: An R Package for Flexible Mediation Analysis using Natural Effect Models. Journal of Statistical Software. 2017; 76(11). doi: 10.18637.

37. Huypens P, Sass S, Wu M, et al. Epigenetic germline inheritance of diet-induced obesity and insulin resistance. Nature genetics. 2016; 48(5):497–9. Epub 2016/03/15. https://doi.org/10.1038/ng.3527 PMID: 26974008.

38. Ng SF, Lin RC, Maloney CA, et al. Paternal high-fat diet consumption induces common changes in the transcriptomes of retroperitoneal adipose and pancreatic islet tissues in female rat offspring. FASEB journal: official publication of the Federation of American Societies for Experimental Biology. 2014; 28(4):1830–41. Epub 2014/01/15. https://doi.org/10.1096/fj.13-244046 PMID: 24421403.

39. Marczylo EL, Amoako AA, Konje JC, et al. Smoking induces differential miRNA expression in human spermatozoa: a potential transgenerational epigenetic concern? Epigenetics. 2012; 7(5):432–9. Epub 2012/03/24. https://doi.org/10.4161/epi.19794 PMID: 22441141.

40. Goossens GH. The Metabolic Phenotype in Obesity: Fat Mass, Body Fat Distribution, and Adipose Tissue Function. Obes Facts. 2017; 10(3):207–15. Epub 2017/06/01. https://doi.org/10.1159/000471488 PMID: 28564650; PubMed Central PMCID: PMC5644968.

41. Peltz G, Aguirre MT, Sanderson M, et al. The role of fat mass index in determining obesity. Am J Hum Biol. 2010; 22(5):639–47. Epub 2010/08/26. https://doi.org/10.1002/ajhb.21056 PMID: 20737611; PubMed Central PMCID: PMC2929934.

42. Nianogo RA, Arah OA. Investigating the Role of Childhood Adiposity in the Development of Adult Type 2 Diabetes in a 64-year Follow-up Cohort: An Application of the Parametric G-formula Within an Agent-based Simulation Study. Epidemiology (Cambridge, Mass). 2019; 30 Suppl 2:S101–s9. Epub 2019/10/01. https://doi.org/10.1097/ede.0000000000001062 PMID: 31569159.

43. Morkve Knudsen GT, Rezwan FI, Johannessen A, et al. Epigenome-wide association of father's smoking with offspring DNA methylation: a hypothesis-generating study. Environmental epigenetics. 2019; 5 (4):dvz023. Epub 2019/12/13. https://doi.org/10.1093/eep/dvz023 PMID: 31827900; PubMed Central PMCID: PMC6896979.

44. Ringseis R, Eder K, Mooren FC, et al. Metabolic signals and innate immune activation in obesity and exercise. Exerc Immunol Rev. 2015; 21:58–68. Epub 2015/04/01. PMID: 25825956.

45. Ambrosini GL. Childhood dietary patterns and later obesity: a review of the evidence. Proc Nutr Soc. 2014; 73(1):137–46. Epub 2013/11/28. https://doi.org/10.1017/S0029665113003765 PMID: 24280165.

46. Brug J, van Stralen MM, Te Velde SJ, et al. Differences in weight status and energy-balance related behaviors among schoolchildren across Europe: the ENERGY-project. PloS one. 2012; 7(4):e34742. Epub 2012/05/05. https://doi.org/10.1371/journal.pone.0034742 PMID: 22558098; PubMed Central PMCID: PMC3338827.

47. Chooi YC, Ding C, Magkos F. The epidemiology of obesity. Metabolism: clinical and experimental. 2019; 92:6–10. Epub 2018/09/27. https://doi.org/10.1016/j.metabol.2018.09.005 PMID: 30253139.

48. Han TS, Hart CL, Haig C, et al. Contributions of maternal and paternal adiposity and smoking to adult offspring adiposity and cardiovascular risk: the Midspan Family Study. BMJ open. 2015; 5(11): e007682. https://doi.org/10.1136/bmjopen-2015-007682 PMID: 26525718

49. Tehard B, van Liere MJ, Com Nougue C, et al. Anthropometric measurements and body silhouette of women: validity and perception. J Am Diet Assoc. 2002; 102(12):1779–84. Epub 2002/12/19. https://doi.org/10.1016/s0002-8223(02)90381-0 PMID: 12487540; PubMed Central PMCID: PMC2020514.

50. Obesity: preventing and managing the global epidemic. Report of a WHO consultation. World Health Organ Tech Rep Ser. 2000; 894:i–xii, 1–253. Epub 2001/03/10. PMID: 11234459.

51. Stein CJ, Colditz GA. The epidemic of obesity. The Journal of clinical endocrinology and metabolism. 2004; 89(6):2522–5. Epub 2004/06/08. https://doi.org/10.1210/jc.2004-0288 PMID: 15181019.

S1 Resource. Ethic commi1ee name and approval number for each study center:

Australia, Melbourne: Monash University Human Research Ethics Committee Project # CF11/1818-2010001012

Belgium, South Antwerp and Antwerp City: Comité voor Medische Ethiek UZA/UA 11/41/288 – UA

Denmark, Aarhus: De Videnskabsetiske Komiteer for region Midtjylland. M-20110106

Estonia, Tartu: Research Ethics Committee of the University of Tartu (UT REC) 209T-17 and 225/M-24

France, Paris: Etude ECRHS III: promotion CHU de Grenoble. Ethical approval CPP Sud est V 4 mars 2011. Approval Ministry of Health AFSSAPS n°B110053-70.

France, Grenoble: Etude ECRHS III: promotion CHU de Grenoble. Ethical approval CPP Sud est V 4 mars 2011. Approval Ministry of Health AFSSAPS n°B110053-70.

France, Montpellier: Etude ECRHS III : promotion CHU de Grenoble. Ethical approval CPP Sud est V 4 mars 2011. Approval Ministry of Health AFSSAPS n°B110053-70.

France, Bordeaux: Etude ECRHS III : promotion CHU de Grenoble. Ethical approval CPP Sud est V 4 mars 2011. Approval Ministry of Health AFSSAPS n°B110053-70.

Germany, Hamburg: Ethikkommission der Bayerischen Landesärztekammer (Positive Votum: 10015

Germany, Erfurt: Ethikkommission der Bayerischen Landesärztekammer (Positive Votum: 10015

Iceland, Reykjavik: National Bioethics committee of Iceland VSN-11-121-S3

Norway, Bergen: Regional Ethics Committee West Norway 2010/759

Spain, Barcelona: Ethics Committee of the Parc de Salut Mar, Barcelona (Comité del etico d'investigacion clínica (CEIC)- Parc de Salut Mar, Barcelona (Approval number) 2009/3500/1

Spain, Galdakao: Ethics Committee of the Parc de Salut Mar, Barcelona (Comité etic d'investigacio clínica (CEIC)- Parc de Salut Mar, Barcelona (Approval num) 2009/3500/1

Spain, Albacete: Ethics Committee of the Parc de Salut Mar, Barcelona (Comité etic d'investigacio clínica (CEIC)- Parc de Salut Mar, Barcelona (Approval number) 2009/3500/1

Spain, Oviedo: Ethics Committee of the Parc de Salut Mar, Barcelona (Comité del etico d'investigacion clínica (CEIC)- Parc de Salut Mar, Barcelona (Approval number) 2009/3500/1

Spain, Huelva: Ethics Committee of the Parc de Salut Mar, Barcelona (Comité del etico d'investigacion clínica (CEIC)- Parc de Salut Mar, Barcelona (Approval number) 2009/3500/1

Sweden, Gothenburg: Regional Ethical Review Board in Uppsala. 2010/432

Sweden, Uppsala: Regional Ethical Review Board in Uppsala. And the number of the decision is 2010/432

Sweden, Umeaa: Regional Ethical Review Board in Uppsala. 2010/432

UK, Ipswich: NRES committee London-Stanmore REC Ref: 11/LO/0965

UK, Norwich: NRES committee London-Stanmore REC Ref: 11/LO/0965

S2 Table A: Descriptive table of father offspring cohort grouped by fathers' smoking onset and stratified by offspring sex, N=2939

	Never smoked		<15 smoking onset		≥ 15 smoking onset		Postnatal smoking onset	
	Sons'	Daughters'	Sons'	Daughters'	Sons'	Daughters'	Sons'	Daughters'
Father characteristics	N=616 (49)	N=783 (47)	N=126 (10)	N=179 (11)	N=482 (38)	N=696 (41)	N=31 (3)	N=26 (2)
Age, mean ± SD	53.4 ± 6.5	53.9 ± 6.0	**57.4 ± 5.2**	**55.3 ± 6.1**	**56.7 ± 5.6**	**56.0 ± 5.7**	54.8 ± 5.8	56.0 ± 5.0
Range	39 - 65	40 - 65	42 – 65	41 – 65	40 - 65	39 - 66	45 - 64	47 - 64
Bmi kg/m^2, mean ± SD	26.5 ± 3.6	26.3 ± 3.7	**27.7 ± 3.8**	**27.7 ± 3.9**	**27.2 ± 4.1**	**27.1 ± 3.7**	26.6 ± 4.4	26.6 ± 2.3
Range	17 - 44	18 - 54	20 - 39	19 - 39	19 - 53	17 - 43	19 - 37	23 - 32
Primary education, n (%)	58 (9)	79 (10)	**38 (30)**	**58 (32)**	**89 (19)**	**127 (18)**	**1 (3)**	**3 (12)**
Secondary education	216 (35)	266 (34)	**45 (36)**	**78 (44)**	**188 (39)**	**263 (38)**	**17 (55)**	**10 (39)**
University/College	338 (55)	437 (56)	**42 (33)**	**42 (24)**	**195 (41)**	**300 (43)**	**13 (42)**	**13 (50)**
Years smoked, mean ± SD	-	-	**28.5 ± 13.6**	**27.0 ± 13.6**	**23.0 ± 13.5**	**22.5 ± 12.7**	**14.0 ± 9.9**	**14.2 ± 10.9**
Range			0 - 59	0 - 52	0 - 52	0 - 49	0 - 38	0 - 34
Packyears, median; 25,75th%								
From birth to offspring's age 18	-	-	**23.2; 14.5, 30**	**21.5; 11.7, 30**	**16.5; 8.2, 26.5**	**16.5; 10, 24**	**2.6; 0.8, 10**	**2.6; 2.0, 12**
In preconception years	-	-	**10.2; 7.3, 16.8**	**9.8; 5.2, 15**	**6; 3, 10**	**6.5; 3.8, 10**	-	-
Age smoke onset, mean ± SD	-	-	**13.4 ± 1.2**	**13.2 ±1.3**	**17.5 ± 2.6**	**17.3 ± 2.5**	**36.4 ± 8.9**	**34.9 ± 10.1**
Range			6 - 14	7 - 14	15 - 33	15 - 30	20 - 53	21 - 50
Offspring characteristics								
Age, mean ± SD	28 ± 7.1	28.4 ± 6.8	**31.3 ± 7.6**	**30.8 ± 7.5**	**30.9 ± 7.3**	**30.7 ± 7.2**	**32.6 ± 7.4**	**33.7 ± 7.9**
Range	18 - 49	18 - 49	18 - 47	18 - 50	18 - 48	18 - 50	20 - 47	19 - 48
Bmi kg/m^2, mean ± SD	24.6 ± 3.8	23.4 ± 4.5	25.3 ± 4.2	**24.6 ± 5.5**	**25.6 ± 4.4**	**24.1 ± 4.9**	25.6 ± 5.2	24.1 ± 4.8
Range	16 - 53	15 - 67	17 - 41	15 - 53	16 - 53	14 - 52	19 - 39	19 - 38
FMI fat mass kg/m^2, mean ± SD	3.0 ± 1.7	5.7 ± 2.4	4.1 ± 0.8	7.7 ± 3.7	**5.6 ± 3.4**	5.7 ± 2.1	5.7 ± 2.4	5.4 ± 1.4
Range	1.1 – 6.6	2.5 – 13.7	3.4 – 5.1	3.9 – 13.8	1.2 – 11.7	3.0 – 14.4	4.0 – 7.4	3.5 – 7.6
Primary education, n (%)	24 (4)	19 (2)	3 (2)	**3 (2)**	13 (3)	17 (2)	1 (3)	1 (4)
Secondary education	280 (46)	251 (32)	65 (52)	**79 (44)**	214 (50)	212 (31)	8 (26)	8 (31)
University/College	312 (51)	511 (65)	58 (46)	**97 (54)**	252 (52)	464 (67)	22 (71)	17 (65)
Never smoked, n (%)	476 (77)	606 (77)	**78 (62)**	**104 (58)**	**313 (65)**	**448 (64)**	**19 (61)**	**16 (62)**
Ever smoked	134 (22)	174 (22)	**48 (38)**	**72 (40)**	**169 (35)**	**247 (36)**	**12 (39)**	**10 (39)**
Years smoked, mean ± SD	7.6 ± 6.2	8.3 ± 6.5	**10.6 ± 7.2**	**10.4 ± 6.8**	**10.4 ± 7.2**	**9.3 ± 6.6**	9.8 ± 7.5	**14.0 ± 7.9**
Range	0 – 27	0 - 33	0 – 29	1 – 30	0 – 36	0 – 33	2 - 25	3 - 24
Age smoke onset, mean ± SD	17.1 ± 3.1	16.4 ± 2.8	**16.1 ± 3.1**	**15.4 ± 2.3**	16.8 ± 2.8	16.5 ± 2.8	17.8 ± 3.3	16.5 ± 2.4
Range	12 – 28	10 – 28	9 – 25	12 – 27	11 – 25	10 – 30	13 - 25	14 - 21

Missing values: Paternal characteristics: Age: NA=37; BMI: NA=34, Education level=23; Packyears: NA=836. Offspring characteristics: Age: NA=7; FMI: NA=2812; Education level: NA=8; Smoking-status: NA= 13; Years smoked: NA=72; Age smoking onset: NA=29.

Sign group differences <0.05 calculated by Wilcoxon and chi square/Kruskal test, marked in bold.

S2 Table B: Descriptive table of mother offspring cohort grouped by mothers' smoking onset and stratified by offspring sex, N=3548

	Never smoked		<15 smoking onset		≥ 15 smoking onset		Postnatal smoking onset	
	Sons'	Daughters'	Sons'	Daughters'	Sons'	Daughters'	Sons'	Daughters'
Mother characteristics	N=732 (48)	N=965 (47)	N=154 (10)	N=232 (12)	N=594 (39)	N=780 (39)	N=42 (3)	N=49 (3)
Age, mean ± SD	53.6 ± 6.7	53.8 ± 6.7	52.7 ± 6.4	**52.6 ± 6.1**	**55.4 ± 6.2**	**56.8 ± 6.0**	**57.6 ± 5.7**	**56.4 ± 6.2**
Range	39 - 65	39 - 65	40 - 65	40 - 65	39 - 65	39 - 65	42 – 65	40 - 64
Bmi kg/m^2, mean ± SD	25.1 ± 4.1	25.4 ± 4.2	**26.8 ± 5.1**	**27.1 ± 5.9**	25.4 ± 4.1	25.7 ± 4.6	**27.3 ± 4.7**	26.4 ± 4.8
Range	16 - 45	17 - 52	18 - 49	17 - 66	14 - 45	17 - 60	17 - 36	19 - 40
Primary education, n (%)	73 (10)	128 (13)	**29 (19)**	**62 (27)**	**85 (14)**	**161 (21)**	**10 (24)**	**10 (20)**
Secondary education, n (%)	238 (33)	288 (30)	**61 (40)**	**88 (38)**	**225 (38)**	**256 (33)**	**18 (43)**	**27 (55)**
University/College, n (%)	417 (57)	545 (57)	**62 (40)**	**81 (35)**	**280 (47)**	**361 (46)**	**14 (33)**	**12 (25)**
Years smoked, mean ± SD	-	-	**23.4 ± 13.6**	**24.6 ± 13.4**	**21.0 ± 13.3**	**20.6 ± 12.7**	**17.5 ± 11.5**	**17.6 ± 12.5**
Range			0 - 49	0 - 51	0 - 52	0 - 49	0 - 40	0 – 43
Packyears, median; 25,75th%								
From birth to offspring's age 18	-	-	**14; 8.5, 22.9**	**16.1; 10.2,23.9**	**12.6; 6, 18.8**	**12.5; 6, 8.8**	**4.5; 2.1, 6.7**	**3; 1.5, 7.9**
In preconception years	-	-	**5.5; 3, 9**	**6; 4, 9.6**	**4; 2.4, 6.6**	**4.5; 2.5, 7.5**	-	-
Age smoke onset, mean ± SD	-	-	**13.3 ± 0.9**	**13.4 ± 0.9**	**17.4 ± 2.3**	**17.3 ± 2.2**	**29.5 ± 8.1**	**29.2 ± 7.4**
Range			9 - 14	7 - 14	15 - 30	15 - 30	20 - 49	15 - 44
Offspring characteristics								
Age, mean ± SD	30.1 ± 7.7	30.3 ± 7.8	31.2 ± 8.5	**31.1 ± 8.0**	**31.5 ± 7.4**	31.0 ± 7.4	**39.1 ± 7.6**	**38.0 ± 7.4**
Range	18 - 50	18 - 51	19 - 50	19 - 52	18 - 49	18 - 50	19 - 52	23 - 51
Birthweight, mean ± SD	3.6 ± 0.5	3.4 ± 0.6	3.4 ± 0.6	**3.1 ± 0.7**	3.5 ± 0.7	3.4 ± 0.6	3.4 ± 0.6	3.5 ± 0.5
Range	1.1-5.0	0.5-5.3	2.0-4.4	1.0 – 4.3	1.1-5.3	0.5-4.8	2.3-4.4	2.5-4.3
Bmi kg/m^2, mean ± SD	24.7 ± 3.6	23.6 ± 4.4	**26.1 ± 4.6**	**24.6 ± 4.6**	**25.6 ± 3.9**	23.7 ± 4.2	**27.3 ± 3.3**	**26.1 ± 5.3**
Range	15 - 43	15 - 49	17 - 43	16 - 41	13 - 43	15 - 44	18 - 34	18 - 37
FMI fat mass kg/m^2, mean ± SD	4.9 ± 2.3	8.2 ± 4.8	3.6 -	7.5 ± 4.0	4.1 ± 2.3	6.7 ± 3.8	4.7 -	10.1 ± 5.2
Range	2.6 – 9.6	3.2 – 20.4	-	4.1 – 16.5	1.0 – 8.7	3.0 – 20.5	-	5.4 – 15.4
Primary education, n(%)	19 (3)	23 (2)	**6 (4)**	**10 (4)**	17 (3)	13 (2)	3 (7)	3 (2)
Secondary education, n(%)	307 (42)	304 (32)	**80 (52)**	**88 (38)**	243 (41)	240 (31)	20 (48)	19 (39)
university/College, n(%)	405 (55)	638 (66)	**68 (44)**	**133 (57)**	334 (56)	524 (67)	19 (45)	26 (53)
Never smoked, n (%)	511 (70)	700 (73)	**92 (60)**	**130 (56)**	398 (67)	**477 (61)**	**22 (52)**	**14 (29)**
Ever smoked, n (%)	218 (30)	262 (27)	**62 (40)**	**101 (44)**	193 (33)	**301 (39)**	**20 (48)**	**35 (71)**
Years smoked, mean ± SD	8.6 ± 7.0	8.5 ± 7.4	10.4 ± 7.4	**11.2 ± 7.2**	9.5 ± 6.4	**9.8 ± 6.7**	**15.6 ± 9.7**	**13.7 ± 8.6**
Range	0 – 32	0 – 35	0 – 37	1 – 30	0 – 28	0 – 32	2 - 36	0 - 33
Age smoke onset, mean ± SD	17.0 ± 3.0	16.6 ± 2.9	**15.6 ± 2.9**	**15.2 ± 3.1**	16.7 ± 3.3	**15.9 ± 2.3**	16.1 ± 3.5	16.1 ± 2.0
Range	7 – 27	10 – 28	10 – 24	10 – 36	8 – 32	11 – 25	8 – 25	12 - 20

Missing values: Maternal characteristics: Age: NA=80; BMI: NA=85; Education level: NA=17; Packyears: NA=868. Offspring characteristics: Age: NA=10; Birthweight: NA=2735; FMI: NA=3440; Education level: NA=6; Smoking-status: NA=12; Years smoked: NA=63; Age smoking onset: NA=25. Sign group differences, <0.05 calculated by Wilcoxon and chi square/Kruskal test, marked in bold

S3 Table: Associations between mothers' smoking onset and offspring (n=111) FMI

	Sons' and daughter's FMI		
Predictors	*Adj. difference in FMI (kg/m²)*	*95% CI*	*P*
Preconception smoking onset < 15 years of age	- 0.315	- 3.171 - 2.541	0.829
Preconception smoking onset ≥ 15 years of age	- 1.029	- 2.810 - 0.752	0.257
Postnatal smoking onset	1.947	- 2.207 - 6.102	0.358

Estimates from generalized linear regression models with adjustment for offspring sex and mothers' education. Clustered by family id and study centre. P value significance level: *.05, **.01, ***.001

S4 Table: Mothers' pack years as mediator of the observed associations between mothers' smoking onset and offspring BMI

Causal mediation analysis mother offspring				
Mothers' smoking onset	*Adj diff. BMI (kg/m²)*	*Std. error*	*z value*	*P value*
A) Mediation by mothers' packyears up to offspring age 18				
Preconception smoking onset <15				
Natural direct effect	0.228	0.421	0.540	0.589
Natural indirect effect	1.059	0.253	4.193	< 0.000 ***
Total effect	1.287	0.349	3.692	< 0.000 ***
Interaction by offspring sex: 0.774				
Preconception smoking onset ≥15				
Natural direct effect	- 0.299	0.256	- 1.165	0.244
Natural indirect effect	0.833	0.199	4.178	< 0.000 ***
Total effect	0.534	0.193	2.776	0.006 **
Interaction by offspring sex : 0.542				
Postnatal smoking onset				
Natural direct effect	1.950	0.541	3.608	< 0.000 ***
Natural indirect effect	0.276	0.080	3.462	< 0.000 ***
Total effect	2.226	0.540	4.120	< 0.000 ***
Interaction by offspring sex: 0.743				
B) Mediation by mothers' preconception packyears				
Preconception smoking onset <15				
Natural direct effect	0.580	0.342	1.694	0.090
Natural indirect effect	0.291	0.181	1.607	0.108
Total effect	0.870	0.284	3.064	0.002 **
Interaction by offspring sex: 0.965				
Preconception smoking onset ≥15				
Natural direct effect	0.261	0.195	1.343	0.179
Natural indirect effect	0.191	0.120	1.594	0.111
Total effect	0.452	0.164	2.761	0.006 **
Interaction by offspring sex : 0.966				

Effect decomposition on the scale of the linear predictor with standard errors based on the sandwich estimator. Conditional on fathers' educational level and offspring sex.
P value significance level: *.05, **.01, ***.001

S5 Table: Mothers' BMI as mediator of the observed associations between mothers' smoking onset and offspring BMI

Causal mediation analysis mother offspring				
Mothers' smoking onset	*Adj diff. BMI (kg/m²)*	*Std. error*	*z value*	*P value*
Mediation by mothers' BMI				
Preconception smoking onset <15				
Natural direct effect	0.551	0.247	2.229	0.026 *
Natural indirect effect	0.334	0.074	4.483	< 0.000 ***
Total effect	0.884	0.258	3.436	< 0.000 ***
Interaction by offspring sex: 0.827				
Preconception smoking onset ≥15				
Natural direct effect	0.258	0.143	1.809	0.070
Natural indirect effect	0.023	0.039	0.596	0.551
Total effect	0.282	0.149	1.896	0.058
Interaction by offspring sex : 0.912				
Postnatal smoking onset				
Natural direct effect	1.869	0.495	3.774	< 0.000 ***
Natural indirect effect	0.320	0.128	2.496	0.013 *
Total effect	2.188	0.486	4.500	< 0.000 ***
Interaction by offspring sex: 0.354				

Effect decomposition on the scale of the linear predictor with standard errors based on the sandwich estimator. Conditional on fathers' educational level and offspring sex.
P value significance level: *.05, **.01, ***.001

S6 Table: Offspring's smoking habits as mediator of the observed associations between mothers' smoking onset and offspring BMI

Causal mediation analysis mother offspring				
Mothers' smoking onset	*Adj diff. BMI (kg/m²)*	*Std. error*	*z value*	*P value*
Mediation by sons' and daughters' ever/never smoking				
Preconception smoking onset <15				
Natural direct effect	0.841	0.254	3.310	< 0.000 ***
Natural indirect effect	0.059	0.024	2.411	0.016 *
Total effect	0.899	0.256	3.518	< 0.000 ***
Interaction by offspring sex: 0.767				
Preconception smoking onset ≥15				
Natural direct effect	0.278	0.147	1.892	0.059
Natural indirect effect	0.031	0.013	2.352	0.019 *
Total effect	0.309	0.147	2.108	0.035 *
Interaction by offspring sex : 0.335				
Postnatal smoking onset				
Natural direct effect	2.090	0.476	4.386	< 0.000 ***
Natural indirect effect	0.129	0.052	2.491	0.013 *
Total effect	2.219	0.477	4.652	< 0.000 ***
Interaction by offspring sex: 0.512				

Effect decomposition on the scale of the linear predictor with standard errors based on the sandwich estimator. Conditional on fathers' educational level and offspring sex.
P value significance level: *.05, **.01, ***.001

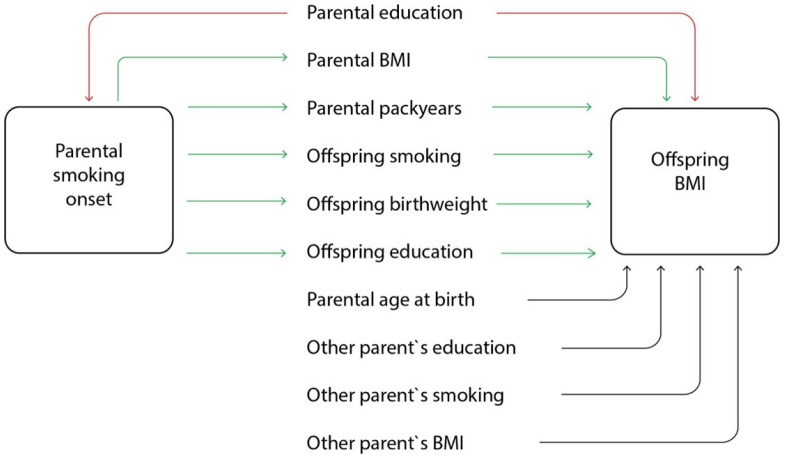
S7 Table: Offspring's birthweight as mediator of the observed associations between mothers' smoking onset and offspring BMI

Causal mediation analysis mother offspring				
Mothers' smoking onset	*Adj diff. BMI (kg/m²)*	*Std. error*	*z value*	*P value*
Mediation by sons' and daughters' birthweight				
Preconception smoking onset <15				
Natural direct effect	1.162	0.557	2.087	0.037 *
Natural indirect effect	- 0.021	0.062	- 0.338	0.736
Total effect	1.141	0.564	2.025	0.043 *
Interaction by offspring sex: 0.329				
Preconception smoking onset ≥15				
Natural direct effect	0.380	0.317	1.196	0.232
Natural indirect effect	- 0.008	0.025	- 0.328	0.743
Total effect	0.372	0.319	1.165	0.244
Interaction by offspring sex : 0.273				
Postnatal smoking onset				
Natural direct effect	0.932	0.820	1.136	0.256
Natural indirect effect	- 0.008	0.025	- 0.315	0.753
Total effect	0.924	0.819	1.128	0.259
Interaction by offspring sex: 0.362				

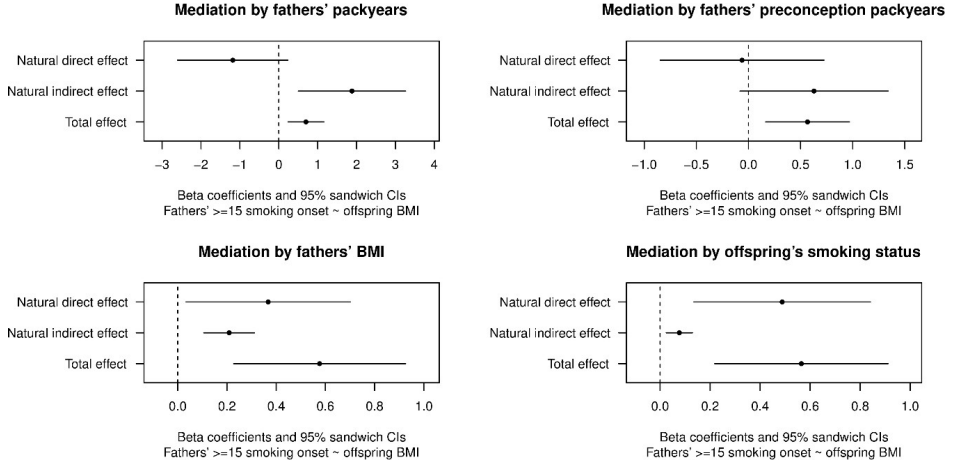
Effect decomposition on the scale of the linear predictor with standard errors based on the sandwich estimator. Conditional on fathers' educational level and offspring sex. Sub-sample with n = 813 offspring
P value significance level: *.05, **.01, ***.001

Supporting information

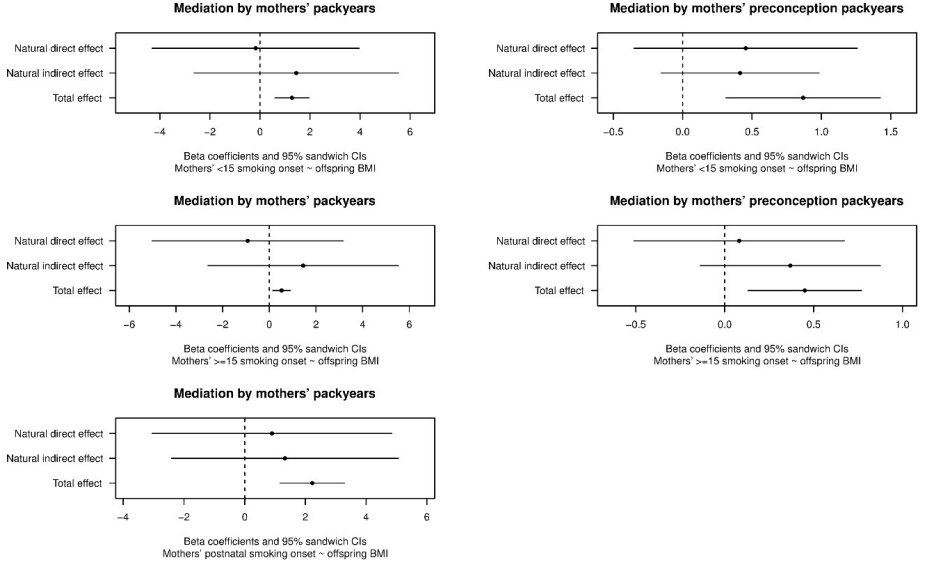
S1 Fig. Directed Acyclic Graph (DAG). The figure presents covariates considered to be included in the statistical model.



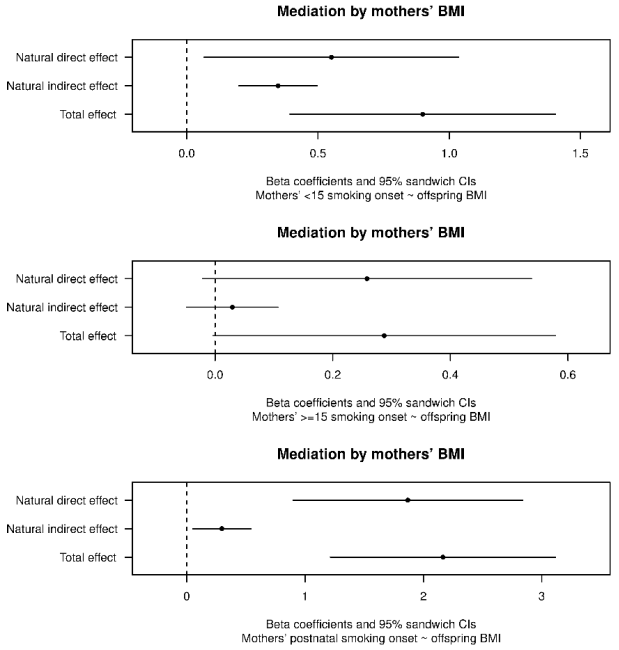
S2 Fig. Visualising mediations of the association with fathers' ≥15 smoking onset on offspring BMI. Analyses reveal full mediation by fathers' pack years and partial mediation by fathers' BMI and offspring's own smoking status. There is no mediation via fathers' preconception accumulative smoking



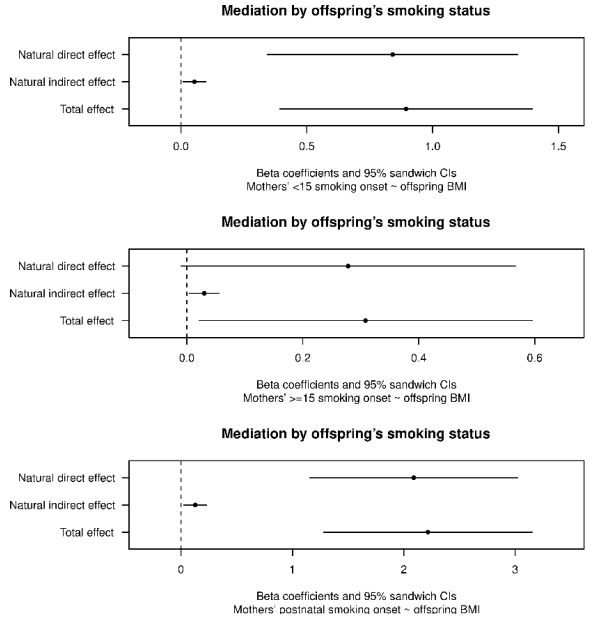
S3 Fig. Visualising mothers' pack years as mediator of the observed associations between mothers' smoking onset and offspring BMI



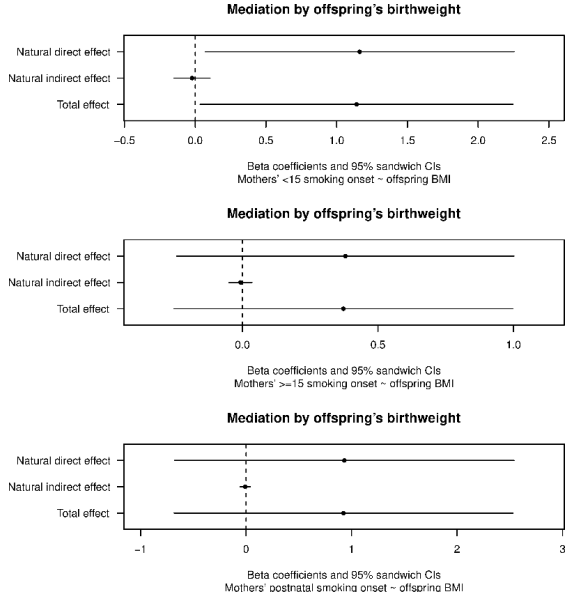
S4 Fig. Visualising mothers' BMI as mediator of the observed associations between mothers' smoking onset and offspring BMI



S5 Fig. Visualising offspring's smoking habits as mediator of the observed associations between mothers' smoking onset and offspring BMI



S6 Fig. Visualising offspring's birthweight as mediator of the observed associations between mothers' smoking onset and offspring BMI



Paper III

III

Fathers' preconception smoking and offspring DNA methylation: A two generation study

Fathers' preconception smoking and offspring DNA methylation: A two generation study

Negusse T Kitaba[1*], Gerd Toril Mørkve Knudsen[2,3*], Ane Johannessen[4,] Faisal I. Rezwan[5], Andrei Malinovschi[6], Anna Oudin[7], Bryndis Benediktsdottir[8,9], David Martino[10], Francisco Javier Callejas González[11], Leopoldo Palacios Gómez[12], Mathias Holm[13], Nils Oskar Jõgi[2,3], Shyamali Dharmage[14], Svein Magne Skulstad[3], Sarah H Watkins[15], Matthew Suderman[15], Francisco Gómez-Real[2,16], Vivi Schlünssen[17,18], Cecilie Svanes[3,4]#, John W. Holloway[1,19]#.

*Joint first authors

#Joint senior authors

[1]Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, UK

[2]Department of Clinical Sciences, University of Bergen, Norway

[3]Department of Occupational Medicine, Haukeland University Hospital, Bergen, Norway

[4]Centre for International Health, Department of Global Public Health and Primary Care, University of Bergen, Norway

[5] Department of Computer Science, Aberystwyth University, Aberystwyth, UK

[6]Department of Medical Sciences: Clinical Physiology, Uppsala University, Sweden

[7]Section of Sustainable Health, Department of Public Health and Clinical Medicine, Umeå University, Sweden

[8]Department of Allergy, Respiratory Medicine and Sleep, Landspitali University Hospital, Reykjavik, Iceland

[9]Faculty of Medicine, University of Iceland, Iceland

[10]Wal-yan Respiratory Research Centre, Telethon Kids Institute, Perth, Australia

[11]Department of Pulmonology, Albacete University Hospital Complex, Albacete, Spain

[12]El Torrejón Health Centre, Andalusian Health Service, Huelva, Spain

[13]Occupational and Environmental Medicine, School of Public Health and Community Medicine, Institute of Medicine, University of Gothenburg, Sweden

[14]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Australia

[15]University of Bristol, MRC Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, Bristol, United Kingdom

[16]Department of Gynaecology and Obstetrics, Haukeland University Hospital, Bergen, Norway

[17]Department of Public Health, Work, Environment and Health, Danish Ramazzini Centre, Aarhus University Denmark, Denmark

[18]National Research Center for the Working Environment, Copenhagen, Denmark

[19]NIHR Southampton Biomedical Research Center, University Hospitals Southampton, Southampton, UK

Correspondence:

Name: Prof John W. Holloway

Address: Human Development and Health, Faculty of Medicine, University of Southampton, Southampton SO16 6YD, UK

Tel: +44 23 8120 8758

Email: J.W.Holloway@soton.ac.uk

Funding:

Co-ordination of the RHINESSA study has received funding from the Research Council of Norway (Grants No. 274767, 214123, 228174, 230827 and 273838), ERC StG project BRuSH #804199, the European Union's Horizon 2020 research and innovation programme under

grant agreement No. 633212 (the ALEC Study), the Bergen Medical Research Foundation, and the Western Norwegian Regional Health Authorities (Grants No. 912011, 911892 and 911631). Study centres have further received local funding from the following: Bergen: the above grants for study establishment and co-ordination, and, in addition, World University Network (REF and Sustainability grants), Norwegian Labour Inspection, and the Norwegian Asthma and Allergy Association. Albacete and Huelva: Sociedad Española de Patología Respiratoria (SEPAR) Fondo de Investigación Sanitaria (FIS PS09). Gøteborg, Umeå and Uppsala: the Swedish Heart and Lung Foundation, the Swedish Asthma and Allergy Association. Reykjavik: Iceland University. Melbourne: National Health and Medical Research Council (NHMRC) of Australia (research grants 299901 and 1021275). Tartu: the Estonian Research Council (Grant No. PUT562). Århus: The Danish Wood Foundation (Grant No. 444508795), the Danish Working Environment Authority (Grant No. 20150067134), Aarhus University (PhD scholarship). ALSPAC funding: EPIC arrays age 15-17, John Templeton Foundation (60828) and EPIC arrays age 24, MRC (MC_UU_12013/2) & CLOSER (MRC and ESRC).

Word Count (Main Body): 3918

Author Contributions

Conceptualization: Cecilie Svanes, Gerd Toril Mørkve Knudsen, Faisal I Rezwan, Ane Johannessen, Negusse Kitaba, John W Holloway

Data curation: Gerd Toril Mørkve Knudsen, Bente Skottvoll, Ane Johannessen, Negusse Kitaba, John W Holloway

Formal analysis: Negusse Kitaba, Cecilie Svanes, John W Holloway

Methodology: Negusse Kitaba, Faisal I Rezwan, Cecilie Svanes, John W Holloway

Replication: Sarah Watkins, Matthew Suderman

Project administration: Cecilie Svanes

Writing – original draft: Negusse Kitaba, Gerd Toril Mørkve Knudsen, Cecilie Svanes, John W Holloway

Writing – review, editing and final approval: All authors

Running title: Paternal smoking and offspring DNA methylation

Abstract

Rationale: Experimental studies suggest that exposures may impact respiratory health across generations via epigenetic changes transmitted specifically through male germ cells. Studies in humans are however limited. We aim to identify epigenetic marks in offspring associated with father's preconception smoking.

Methods: We conducted epigenome-wide association studies (EWAS) in the RHINESSA cohort on father's any preconception smoking (N=875 offspring) and father's pubertal onset smoking <15 years (N=304), using Infinium MethylationEPIC Beadchip arrays, adjusting for offspring age, maternal smoking and personal smoking. EWAS of maternal and offspring personal smoking were performed for replication.

Results: Father's smoking commencing preconception was associated with methylation of blood DNA in offspring at two Cytosine-phosphate-Guanine sites (CpGs) (False Discovery Rate (FDR) <0.05) in *PRR5* and *CENPP*. Father's pubertal onset smoking was associated with 19 CpGs (FDR <0.05) mapped to 14 genes (*TLR9*, *DNTT*, *FAM53B*, *NCAPG2*, *PSTPIP2*, *MBIP*, *C2orf39*, *NTRK2*, *DNAJC14*, *CDO1*, *PRAP1*, *TPCN1*, *IRS1* and *CSF1R*). These differentially

methylated sites were hypermethylated and associated with promoter regions capable of gene silencing. Some of these sites were associated with offspring outcomes in this cohort including ever-asthma (NTRK2), ever-wheezing (DNAJC14, TPCN1), weight (FAM53B, NTRK2) and BMI (FAM53B, NTRK2) ($P < 0.05$). Pathway analysis showed enrichment for gene ontology pathways including regulation of gene expression, inflammation and innate immune responses.

Conclusion: Father's preconception smoking, particularly in puberty, is associated with offspring DNA methylation, providing evidence that epigenetic mechanisms may underly epidemiological observations that pubertal paternal smoking increases risk of offspring asthma, low lung function and obesity.

Key Words: Preconception, paternal effects, tobacco smoke, epigenetic, Epigenome-Wide Association Study, DNA methylation, RHINESSA

Take-home message

DNA methylation sites associated with asthma, wheezing and BMI have been identified in the offspring of fathers who smoke in early puberty. This provides evidence of molecular mechanisms underlying the observed cross-generational effects of smoking.

Introduction

There is growing consensus that perturbations of the epigenome through parental exposures even before their offspring are conceived may explain some of the variation in the heritability of health and disease not captured by Genome-Wide Association Studies (GWAS). The period of puberty in future parents, in particular fathers, may represent a critical window of physiological change and epigenetic reprogramming events, which may increase the individual's susceptibility for environmental exposures to be embodied in the developing gametes[1,2]. Animal and human studies have shown that prenatal as well as personal exposure to smoking are associated with epigenetic modifications that impact on sperm count and quality[3]. There is now growing interest in how epigenetic modifications, such as DNA methylation (DNAm), related to the parental *preconception* period may influence the health of the *next generation*[4].

Although smoking rates are generally declining, smoking commencing before the age of 15 is increasing[5,6]. Epidemiological studies have demonstrated that father's smoking in adolescent years may be a causal factor for poorer respiratory health in offspring. Both fathers' smoking initiation before age 15 and smoking duration before conception have been associated with more asthma and lower lung function in offspring[7–9]. Father's preconception smoking onset has also been associated with higher body fat mass in sons[10–13].

Epigenome-Wide Association Studies (EWAS) have identified extensive methylation biomarkers associated with personal smoking, all-cause mortality in current and former smokers, as well as mother's smoking during pregnancy[6,14–17]. While previous studies have identified DNA methylation signals in offspring blood[18] and cord blood[19] related to father's smoking, they have not specifically investigated the timing of exposure, partly because detailed smoking information from fathers is rarely available[20]. Methylation markers

144 associated with paternal preconception smoking, could have an important role in elucidating
145 long-term effects on the offspring epigenome, with the potential for developing efficient
146 intervention programs and improved public health.
147 This study aimed to investigate whether DNA methylation of DNA measured in offspring blood
148 is associated with fathers' smoking commencing before conception, and in particular, with
149 fathers' smoking starting in (pre)pubertal years (before age 15). We hypothesized that
150 epigenetic changes involving DNA methylation may explain the molecular mechanisms
151 underlying the association between fathers' smoking preconception and offspring health
152 observed in epidemiological studies. Additionally, we hypothesized that fathers' smoking in
153 the critical window of early puberty may have a more significant impact on the offspring
154 epigenome. In a two-generation cohort, we sought to identify the DNA methylation changes
155 in offspring blood associated with (1) father's smoking onset preconception compared with
156 never or later onset smoking, and (2) father's smoking onset before age 15 compared with
157 never smoking.
158
159

Methods

Study design and data

We used data and samples from offspring that participated in the RHINESSA study (www.rhinessa.net). Parent data, including detailed information on smoking habits, were retrieved from the population-based European Community Respiratory Health Survey (ECRHS, www.ecrhs.org) and/or the Respiratory Health in Northern Europe study (RHINE, www.rhine.nu) studies. This analysis comprised 875 offspring-parent pairs with complete information, from six study centres with available peripheral blood for offspring (Aarhus, Denmark; Albacete/Huelva, Spain; Bergen, Norway; Melbourne, Australia; Tartu, Estonia). All participants were of Caucasian ancestry. Medical research committees in each study centre approved the studies, and each participant gave written consent.

Father's smoking and age of starting/quitting was reported in interviews/questionnaires, and related to offspring's birth year, to define the categories: never smoked (N=547), any preconception smoking (N=328), preconception smoking with onset <15 years (pubertal smoking) (N=64) (cut point based on mean age of voice break 14.5 years, first nocturnal seminal emission 14.8 years). Personal smoking was classified as current, ex- or never smoking. Maternal smoking was defined by offspring's report on mothers' smoking during their childhood/pregnancy.

DNAm in offspring was measured using Illumina Infinium MethylationEPIC Beadchip arrays (Illumina, Inc. CA, USA) and data processed using an established pipeline as detailed in the online supplement. Following processing 726,661 CpGs were retained for analysis.

Statistical analysis

Two EWAS on preconception paternal smoking as exposure (any preconception smoking, prepuberty smoking) using robust regression were run with offspring blood DNA methylation

as outcome adjusting for offspring's sex, age, personal and mother's smoking, study center and cell-type proportions at significance level of false discovery rate (FDR) corrected p-value <0.05. Inflation from systematic biases was adjusted using BACON. Differentially methylated regions were detected using DMRCate and dmrff. In additional analyses, associations between fathers' any preconception smoking and offspring's DNA methylation were also stratified by offspring sex. Biological interpretation of significant dmCpGs is detailed in the supplementary methods.

We compared our EWAS results with findings from meta-analyses of EPIC array DNA methylation associated with personal smoking from four population-based cohorts[21], personal smoking-methylation effects from 16 cohorts using 450K arrays[16]; and the Pregnancy and Childhood Epigenetics Consortium (PACE) meta-analysis of mother's smoking in pregnancy on offspring cord blood methylation[22] to assess the shared count of dmCpG sites at (FDR<0.05) for the overlap between each EWAS.

Replication analysis

Replication was carried out in the ALSPAC (Avon Longitudinal Study of Parents and Children) cohort adjusted for predicted cell count proportions, batch effects (plate), maternal smoking during pregnancy, self-reported own smoking, age and sex using DNA methylation data from whole blood measured at age 15-17. A description of the ALSPAC cohort is provided in the supplementary methods. T-tests were used to compare the association of regression coefficient of RHINESSA's dmCpG sites at FDR <0.05 and the top 100 CpG sites with ALSPAC. Signed tests were used to test the direction of association.

Sensitivity analyses

To assess the effect of social class, father's education was used as a proxy for social class. In order to see the effect of CpGs changing with age, the correlation of methylation at dmCpGs

known to be associated with offspring age, known aging markers from RHINESSA EWAS, dmCpG sites for father smoking before age 15, and offspring age was assessed. To further investigate whether the identified dmCpGs were associated with respiratory outcomes and weight in the offspring, we conducted regression analysis between offspring's DNA methylation signals and offspring's own reports of ever-asthma, ever-wheeze, weight and BMI, while accounting for offspring sex

Results

The analysis included 875 RHINESSA participants (Table 1A), 457 males and 418 females, aged 7 to 50 years. Of these 328 had a father who had ever smoked of which 64 had started before age 15 years; 263 had a mother who had ever smoked, and 240 had smoked themselves. Characteristics are also given for the sub-sample of 304 offspring whose father either had started smoking before age 15 years, or never smoked (Table 1B).

Epigenome wide association analysis of preconception father's smoking

Epigenome-wide association between father's any preconception smoking and offspring DNA methylation identified two dmCpGs (inflation A=1.187); cg00870527 mapped to *PRR5* and cg08541349 mapped to *CENPP* (Table 2A, and supplementary table E1). The genome-wide distribution of associated dmCpGs is shown in Figure 1A. The comparison of methylation distribution between never- and ever-smoke exposed is shown in Figure 1C.

In sex-stratified analysis, in males (N=457) we identified four dmCpGs mapped to KCNJ1, GRAMD4 , TRIM2 and MYADML2. In females (N=418) there was one dmCpG mapped to

LEPROT1 (FDR <=0.05) (Supplementary Table E2). All sex-specific dmCpGs were hypomethylated.

To specifically determine the signature related to father's early onset smoking, we compared methylation differences between offspring of fathers who started to smoke <15 years (n=64) with offspring of never smoking fathers (n=240). We identified 55 dmCpGs at FDR <0.05 (Å=1.44) showing genome-wide significance. After adjusting for inflation using BACON, 19 dmCpGs showed significant association at FDR <0.05 with Å=1.29 (Table 2B, Figure 1B, and supplementary Table E3). These dmCpGs were mapped to 14 known genes and 5 intergenic regions. The genes include TLR9, DNTT, FAM35B, NCAPG2, MBIP, C2orf39, NTRK2, DNAJC14, CDO1, PRAP1, TPCN1, IRS1, PSTPIP2, and CF1R. All hits were hypermethylated in the exposed group. The comparison of methylation distribution between the never and smoke exposed is shown in Figure 1D.

The dmCpGs associated with father's preconception smoking were mainly located in open-sea genomic features and enriched for promoter regions (Table 2A). The dmCpGs associated with father's pubertal smoking were in open-sea genomic features and CpG island shores (flanking shore regions, <2 kb up-and downstream of CpG islands) and enriched for CpG islands and gene bodies (Table 2B).

Father's preconception smoking signatures as compared with signatures of personal and mother's smoking

To compare the effects of father's preconception and pubertal smoking on the offspring epigenome with that of other smoking exposures, the epigenome-wide effects of offspring's

255 own smoking as well as their mother's smoking during pregnancy and childhood were
256 assessed. We identified 33 dmCpGs related to personal smoking, and 14 dmCpGs associated
257 with mother's smoking (FDR<0.05) (Supplementary Tables E4 and E5, respectively).
258
259 To illustrate the distinct and shared genome-wide effects of personal, mother's, and father's
260 smoking on the offspring methylome, we generated a locus-by-locus genome comparison,
261 (Figure 2A). While there was similarity between the effects of personal smoking and mother's
262 smoking on chromosome 5, we observed distinct signatures for father's preconception
263 smoking on chromosome 22, and for mother's smoking exposure on chromosomes 7 and 15.
264
265 Comparing our EWAS results with findings from previous studies showed that 10 of the
266 dmCpGs we identified as related to maternal smoking, and 20 (14+6) and 19 (14+5) of the
267 dmCpGs identified as related to personal smoking, were present in the relevant meta-
268 analyses[16,2122] (Figures 2B and 2C). However, when we compared our top 100 dmCpGs for
269 father's any preconception smoking onset EWAS with mother's smoking, there was no
270 evidence for shared CpGs (Figure 2B). For father's pubertal smoking, only two CpG sites
271 (cg11380624 (*DNAJC14*), cg20728490 (*DNTT*)) were shared with analyses of personal smoking
272 by Joehanes et al.[21] and two sites (cg12053348 (intergenic), cg20728490 (*DNTT*)) with
273 Christiansen et al. [16], while 16 CpG sites were unique (Figure 2C).
274
275 **Enrichment of dmCpGs for related traits**
276 We investigated whether the significant dmCpGs associated with father's preconception
277 smoking onset overlapped with other traits, using the repository of published EWAS literature
278 in the EWAS atlas. The top 23 dmCpG sites for father's any preconception smoking (those

with p-value ≤9.86 x 10^{-06}, distinctly lower than the following sites) were enriched for traits that include Immunoglobulin E (IgE) level, muscle hypertrophy, maternal smoking, and birthweight (Figure 3A). dmCpGs (FDR<0.05) associated with father's pubertal smoking were enriched for traits such as autoimmune diseases, atopy, smoking, and puberty (Figure 3B). For comparison, maternal and personal smoking dmCpGs were enriched for shared traits including aging, birthweight, cognitive function, lung function, smoking and type 2 diabetes and cancers – whereas IgE level and atopy were specifically enriched in paternal smoking (Figure 3C and 3D).

Role of dmCpGs for father's pubertal smoking (smoking initiation < 15 years)

Given the stronger effects of father's pubertal smoking we further explored the biological relevance of these findings.

Transcription factor enrichment

We interrogated eFORGE TF for transcription factor enrichment in $CD4^+$ cells to determine the regulatory role of our 19 significant dmCpGs (FDR<0.05) related to father's pubertal smoking. We found significant enrichment of 27 transcription factor binding sites that overlapped with 7 of the dmCpGs (q-value<0.05) identified in our EWAS study (Supplementary Table E6).

EWAS atlas lookup

Of the 19 dmCpGs associated with father's pubertal smoking identified in our analysis, 11 were present in the EWAS atlas and correlated with gene expression in a variety of tissues in the EWAS atlas (Figure 4A) and overlapped with promoters (Figure 4B) (FDR <0.05). These were significantly associated with 9 other traits, including atopy and fractional exhaled nitric

oxide (cg23021329), smoking (cg20728490; cg16730908), BMI (cg03516318), Acute Lymphoblastic Leukemia (cg2240207), cancer (cg11380624), and Crohn's disease (cg10981514), (Supplementary Table E7).

Differentially methylated region (DMR) analysis

No DMRs were significantly associated with father's any preconception smoking using either DMRcate or dmrff. There were suggestive hits for father's pubertal smoking, such as DNTT at FDR= 0.084. All DMRs are listed in supplementary Table E8.

Pathway enrichment

To gain further insight into the functional roles of the dmCpGs, we used 14 genes that were mapped to dmCpGs associated with father's pubertal smoking to generate a protein-protein interaction network from the String database. The top 20 protein interactors were included with high confidence score cutoff 0.7 from protein-protein interaction data sources including experimentally validated protein physical complexes, curated databases and co-expressions. The network indicated that immune response related genes *TLR9*, *CSF1R*, *NTRK2*, *PSTPIP2*, *PTPN11* and *IL34* were well connected (Figure 5A) (p-value $<1.0X10^{-16}$). The molecular function enrichment analysis showed enrichment for gene expression, inflammatory response, innate immunity, and cytokine binding (Figure 5B). We also assessed enrichment of GO terms using gometh. The most significantly enriched biological process terms (FDR<0.05) include: Inactivation of MAPK activity involved in osmosensory signaling pathway (GO:0000173), negative regulation of interleukin-6 production (GO:0032715), regulation of mast cell chemotaxis (GO:0060753), regulation of neutrophil migration (GO:1902622) and insulin processing (GO:0030070) (Supplementary Table E9).

Replication of DNA methylation signatures associated with father's preconception smoking

The replication cohort in ALSPAC included 542 participants (female=280, male=262), of whom 86 had a father who started to smoke before the age of 15 and 456 had never-smoking fathers. There was no overlap of dmCpG sites significantly associated with father's smoking before age 15 between the two cohorts (FDR<0.05). However, of the 19 significant dmCpGs identified as related to father's pubertal smoking in RHINESSA, 11 showed nominal replication in ALSPAC ($p< 0.05$) with similar direction. The correlation of effects between studies is R=0.49. The binomial sign test showed the association to be significant at $p<0.05$. Expanding the comparison to the top 100 dmCpGs in RHINESSA, the correlation of effects between studies, R = 0.54, p-value = $3.04x10^{-05}$.

Sensitivity analyses

In order detect whether the associations identified were influenced by social class, we carried out regression analysis between paternal smoking associated dmCpGs as outcome and father's education as exposure. No association was found.

In order to see the effect of CpGs changing with age, we compared known aging-related CpG markers identified from Rhinessa EWAS and paternal smoking dmCpGs with offspring age. There was only weak correlation between paternal smoking dmCpGs and offspring age (maximum R =|0.2|, with 9 CpGs R = 0). In contrast, the age-related CpG markers showed a strong correlation with age (R >=|0.6|) (Supplementary Figure 1).

In order to determine whether paternal smoking dmCpGs were associated with offspring outcomes we ran logistic and linear regression on ever-asthma, ever-wheezing, weight and BMI. Some dmCpG sites showed association with ever-asthma (cg22402007: NTRK2), ever-

351 wheezing (cg11380624: DNAJC14, cg10981514: TPCN1), weight (cg12053348, cg03380960:
352 FAM53B, cg22402007: NTRK2[23]) and BMI (cg03380960: FAM53B, cg12053348, cg22402007:
353 NTRK2) at P<0.05 as shown in (Supplementary Table E10). The study power is shown in
354 Supplementary Table E11.
355

356 **Discussion**

357 To our knowledge, this is the first human study to investigate the potential epigenetic
358 mechanisms behind the impact of father's pubertal smoking on offspring. In this epigenome-
359 wide association study, using data from two generations of study participants, we found
360 differentially methylated CpG sites in offspring associated with father's preconception
361 smoking. Signatures related to father pubertal smoking (smoking initiation before age 15)
362 were much more pronounced than smoking starting at any time preconception. Sixteen of
363 our identified dmCpGs have not previously been reported to be associated with personal or
364 maternal smoking. We suggest these new smoking-associated methylation biomarkers may
365 be specific to smoking exposure of future fathers in early puberty. Several top dmCpGs were
366 enriched for promoter regions and overlapped with significant transcription factor sites that
367 correlated with gene expression in a variety of tissues. Besides unique sites identified for
368 father's preconception smoking onset, our study confirms previously reported DNA
369 methylation sites associated with personal and mother smoking, demonstrating the validity
370 of our cohort and analytical methods. The genes to which dmCpGs map are related to
371 regulation of innate immunity and inflammatory responses.
372
373 For father's any preconception smoking, we found two novel CpG sites that were not
374 previously linked with any previously investigated smoking phenotype. PRR5 (mapped with
375 cg008870527) is a component of the (mTOR) complex 2 which is upstream of major pathways
376 known to have a crucial role in metabolic regulation and is suggested to play a role in obesity
377 and the pathogenesis of insulin resistance[24]. CENPP (mapped with cg08541349), has been
378 associated with lung function, leucocyte count, BMI and type II hypersensitivity reaction in
379 GWAS studies[25]. In the male EWAS analysis, gene KCNJ1 is known to be associated with vital

capacity and linked with obesity. A population-based study of Hispanic children has shown association of GRAMD4 with IgE levels (relevant to asthma pathogenesis)[26]. TRIM2 is linked with low density lipoprotein measurement and total cholesterol, while MYADML2 is linked with vital capacity and BMI-adjusted waist-hip ratio. Of the female EWAS hits, LEPROTL1 has a role in lung function (FEV1/FVC ratio) and several cancers, and a regulatory effect on growth hormone action and glucose homeostasis[27].

For father's pubertal smoking, two of our 19 significant CpG sites, have previously been associated with personal smoking (cg20728490 in *DNTT* and cg16730908 in *PSTPIP2*), and they map to genes with important roles in innate immune responses to infections[28,29]. Upregulation of *PSTPIP2* has also been linked to neutrophilic airway inflammation and non-allergic asthma. When exploring the biological impact of other genes mapped to the dmCpGs uniquely associated with father's pubertal smoking, several were related to genes associated with innate immunity, allergic diseases, and asthma development, such as *TLR9*, *CSF1R*, *DNAJ14*, *NTRK2* and *TPCN1*[28–33]. We also identified CpGs and genes with links to obesity (*NTRK2*, *PSTPIP2*, *MBIP*)[25,35], and glucose and fat metabolism (*IRS1*). The differentially methylated CpGs were mainly located in open-sea genomic features, and enriched for promoter regions, CpG island and gene bodies. These findings suggest that the identified DNA methylation differences, even though of relatively small magnitude, have functional implications in terms of a regulatory role in specific gene expression. Pathway analysis and molecular function enrichment further found interconnection of immune response related genes, and enrichment for inflammatory response, innate immunity, and cytokine binding. When seeking replication of results in an independent sample in the ALSPAC, although no dmpCpGs overlapped in the two population cohorts, results showed that effect estimates

associated with fathers' preconception smoking were moderately correlated and with concordant directional effects.

Several mechanistic reports have demonstrated that the toxicogenic components in cigarette smoke impact on epigenetic germline inheritance and affect the offspring's metabolic health[36]. However, given this is the first study that investigated DNA methylation signatures in young and adult offspring in relation to a timing-specific exposure on father's smoking, there is limited published literature that is directly comparable to our findings. In a pilot study, we previously observed differentially methylated regions associated with father's ever smoking, among which annotated genes were related to innate and adaptive immunity and fatty acid synthesis[18]. Preconception paternal smoking has been shown to alter sperm DNA methylation[37], and independently increase asthma risk and reduce lung function in the offspring [9], especially if the smoking started before age 15[7,9]. The observed association between the dmCpG sites related to father's early onset smoking, and offspring asthma, wheezing and weight, suggests that epigenetic changes may lie on the casual pathway between paternal smoke exposures and offspring health outcomes.

Strikingly, the dmCpG sites we identified as related to fathers' preconception smoking (any preconception smoking as well as pubertal smoking), were quite unique and not the same as those previously reported or found in our data to be associated with mothers' or personal smoking. Reassuringly, our EWAS of mother's smoking and personal smoking, identified several of the dmCpG sites related to these exposures in other cohorts.

Available data for appropriate replication of our results is a major challenge. We found moderate correlation between RHINESSA and ALSPAC EWAS for paternal smoking before 15 years. Although the replication analysis found effect estimates to have concordant directions in several of the dmCpGs, we did not identify overlapping significant dmCpGs associated with fathers' preconception smoking in the replication cohort. The low sample size in both cohorts for paternal smoking before 15 might contribute to the lack of shared genome-wide significance. Even within the same population, using different platforms can cause difficulties with replication[38]. The similarity in the direction of association suggests a potential biological effect of early pre-puberty father's smoking, but further research is warranted in order to verify our novel results.

Although we accounted for personal and mother's smoking exposure in the analysis, we cannot disregard potential residual confounding related to maternal and personal smoking. Further, our analyses cannot fully disentangle effects of father's early onset smoking from effects of subsequent accumulating second hand smoke exposure. However, epidemiological analyses of various measures of father's smoking as related to offspring phenotype in over 20.000 father-offspring pairs found that effects of any other aspect of father's smoking was negligible as compared to that of starting smoking early[7]. We did not control for genetic variations at single nucleotide polymorphisms and cannot rule out that the differentially methylated CpG sites are affected by, or interact with, GWAS-associated genetic variants. However, a recent analysis of our study cohorts using highly advanced statistical probabilistic simulations demonstrated that unmeasured confounding had a limited impact on the effects of father's preconception smoking on offspring asthma[8]. This suggests that the identified

dmCpGs associated with father's preconception smoking, most likely are not driven by unmeasured confounding - by genetic factors or by lifestyle-related or environmental factors.

Self-reporting of smoking is another limitation of our study. However, based on validation studies there is an overall consensus that self-report provides a valid and reliable tool for assessing smoking behaviour in cohort studies. Furthermore, it is likely that error in father's reporting of smoking habits is independent of DNA methylation measured in the offspring, and that misclassification thus will have attenuated the observed results and that the underlying true results might be stronger[39,40].

We suggest that the observed association between father's preconception smoking and offspring DNA methylation marks could be caused by transmission through germline imprint of male sperm. Supported by previous mechanistic and epidemiological findings we also speculate that our novel results reflect that early adolescence may constitute a period of particular vulnerability for smoking exposure to modify the offspring's epigenome. A recent study demonstrated that preconception paternal cigarette exposure in mice from the onset of puberty until 2 days prior to mating modified the expression of miRNAs in spermatozoa and influenced the body weight of F1 progeny in early life[41]. As prepubertal years as well as the onset of puberty represents periods of epigenetic reprogramming events[42], we suggest early adolescence may be a critical time for tobacco-related exposures to interfere with germline epigenetic patterns. This is, however, most challenging to study in humans and multiple scientific approaches are needed to elucidate the molecular mechanisms underlying the current findings as well as previous epidemiological results.

Conclusion

We have identified dmCpG sites in offspring associated with father's onset of smoking before conception, with most pronounced effects when the father started to smoke already in early puberty (before the age of 15). The pattern differed from those of maternal smoking in pregnancy and of personal smoking, and we suggest these may be unique methylation signatures specific to father's early adolescent smoking. The genes to which the identified dmCpGs map, are related to asthma, IgE and regulation of innate immunity and inflammatory responses. Our study provide evidence for an epigenetic mechanism underlying the epidemiological findings of high risk of asthma, obesity and low lung function following father's early adolescent smoking. The functional links of hypermethylated genes suggest that particularly father's pubertal smoking can have cross-generational effects impacting on the long-term health in offspring. Smoking interventions in early adolescence may have implications for better public health, and potential benefits, not only for the exposed, but also for future offspring.

Table 1 A and B: General characteristics of study participants from the RHINESSA study with complete data on offspring DNA methylation and father's age of onset of tobacco smoking. A: for the full cohort of 875 offspring, and B: for the 304 offspring whose father started to smoke before age 15 years or never smoked.

	A: FULL Cohort			B: Start Smoking <15 years		
Characteristic	**No**, N = 547[1]	**Yes**, N = 328[1]	**p-value**[2]	**No**, N = 240[1]	**Yes**, N = 64[1]	**p-value**[3]
Sex			0.8			0.5
F	263 (48%)	155 (47%)		112 (47%)	33 (52%)	
M	284 (52%)	173 (53%)		128 (53%)	31 (48%)	
Study centre			<0.001			0.005
Albacete	24 (4.4%)	32 (9.8%)		7 (2.9%)	9 (14%)	
Arhus	34 (6.2%)	17 (5.2%)		14 (5.8%)	1 (1.6%)	
Bergen	320 (59%)	194 (59%)		174 (72%)	47 (73%)	
Huelva	17 (3.1%)	14 (4.3%)		5 (2.1%)	2 (3.1%)	
Melbourne	78 (14%)	14 (4.3%)		21 (8.8%)	1 (1.6%)	
Tartu	74 (14%)	57 (17%)		19 (7.9%)	4 (6.2%)	
Age	26 (8)	30 (8)	<0.001	26 (8)	27 (8)	0.2
Mother smoking	84 (15%)	179 (55%)	<0.001	30 (12%)	34 (53%)	<0.001
Offspring Smoking	121 (22%)	119 (36%)	<0.001	44 (18%)	25 (39%)	<0.001
B-cells	0.022 (0.019)	0.020 (0.016)	0.4	0.022 (0.016)	0.020 (0.015)	0.4
CD4-cells	0.030 (0.031)	0.032 (0.032)	0.3	0.026 (0.026)	0.026 (0.024)	>0.9
CD8-cells	0.13 (0.05)	0.12 (0.05)	0.5	0.13 (0.04)	0.12 (0.04)	0.7
NK-cells	0.07 (0.04)	0.07 (0.04)	0.8	0.07 (0.04)	0.07 (0.04)	0.4
Mononuclear cells	0.073 (0.022)	0.071 (0.020)	0.2	0.074 (0.020)	0.071 (0.020)	0.4
Neutrophil	0.67 (0.09)	0.68 (0.08)	0.3	0.68 (0.08)	0.68 (0.07)	0.8

[1] n (%); Mean (SD)
[2] Pearson's Chi-squared test; Wilcoxon rank sum test
[3] Pearson's Chi-squared test; Fisher's exact test; Wilcoxon rank sum test

*Including father never smoked and father started smoking after birth of the offspring

Table 2A and B. CpG sites associated with father's smoking at genome wide significance (FDR<0.05) **A:** for father's any preconception smoking, in the full cohort (N=875), and **B**: for father's smoking starting before age 15 years, in the subpopulation (N=304).

Fathers' smoking	CpG	Coefficient*	Average**	SD	Adj.P***	Gene	Location****
A: Any preconception smoking onset	cg00870527	-0.024	0.5	0.07	0.028	PRR5	N_Shelf
	cg08541349	-0.012	0.88	0.023	0.028	CENPP	OpenSea
B: Fathers' smoking onset before age 15	cg23021329	0.015	0.27	0.021	0.026	TLR9	S_Shore
	cg20728490	0.032	0.37	0.049	0.026	DNTT	OpenSea
	cg12053348	0.036	0.61	0.056	0.026	NA	OpenSea
	cg03380960	0.019	0.48	0.045	0.034	FAM53B	OpenSea
	cg26274304	0.018	0.36	0.027	0.037	NCAPG2	N_Shore
	cg16730908	0.021	0.39	0.032	0.037	PSTPIP2	S_Shore
	cg13904562	0.041	0.53	0.056	0.037	NA	OpenSea
	cg07508217	0.026	0.69	0.042	0.037	NA	OpenSea
	cg03516318	0.028	0.21	0.039	0.037	MBIP	OpenSea
	cg10883621	0.02	0.35	0.032	0.037	C2orf39	Island
	cg22402007	0.022	0.16	0.031	0.041	NTRK2	N_Shore
	cg11380624	0.024	0.27	0.036	0.041	DNAJC14	N_Shore
	cg15882605	0.025	0.44	0.051	0.041	NA	OpenSea
	cg03818156	0.017	0.9	0.028	0.041	NA	OpenSea
	cg13288863	0.02	0.79	0.049	0.048	CDO1	N_Shore
	cg03743584	0.018	0.3	0.025	0.048	PRAP1	OpenSea
	cg10981514	0.023	0.42	0.042	0.048	TPCN1	OpenSea
	cg06600694	0.005	0.06	0.008	0.048	IRS1	Island
	cg14700085	0.016	0.71	0.024	0.050	CSF1R	OpenSea

* Coefficient: Regression coefficient between father smoking/not smoking

** Average methylation across all samples

*** adj.P.Val: FDR adjusted p value

****N (north) Shelf: up to 2 kb outward from flanking shores; Open Sea: > 4 kb from CpG islands; N (north) and S (south) Shores: up to 2 kb from flanking CpG islands

Figure 1A and B. Manhattan plot for Genome-wide distribution of dmCpGs **A:** for father's any preconception smoking, and **B:** father's pubertal smoking starting before age 15. The red line shows genome-wide significance, the blue is the suggestive line. The y-axis represents -log10 of the p-value for each dmCpG (indicated by dots) showing the strength of association. The x-axis shows the position across autosomal chromosomes. The top dmCpGs on each chromosome were annotated to the closest gene.

Figure 1C and D. Comparison of methylation differences for **C**: for father's any preconception smoking, and **D**: for father's pubertal smoking starting before age 15.

Figure 2A. Circos plots showing genome-wide distribution across autosomal chromosomes of dmCpGs associated with **A:** personal smoking (in offspring), **B:** mother's smoking, **C:** father's any preconception smoking, and **D:** father's pubertal smoking starting before age 15. Each dot represents a CpG site; the radial line shows the -log10 p-value for each EWAS. Zoomed dots show significant sites in one of the EWAS; each zoomed dot colour shows a unique CpG site specific locus in all 4 EWASs.

Figure 2B and C. Venn diagram showing EWAS CpG top hits for personal smoking, mother's smoking (FDR<0.005), father's any preconception smoking (top 100 dmCpGs), and father's pubertal smoking starting before age 15 (FDR<0.05) in the RHINESSA cohort, which are shared with top hits from meta-analysis of **B:** mother smoking (blue oval) as reported by Joubert et al 2016, and **C:** personal cigarette smoking signature as reported by Christiansen et al 2021 (blue) and by Joehanes et al 2016 (green).

Figure 3: Traits associated with the CpG sites that in EWAS were identified to be differentially methylated according to **A:** father's any preconception smoking, **B:** father's pubertal smoking starting before age 15, **C:** Mother's smoking and **D:** personal smoking

*PPBAPDE: perinatal polychlorinated biphenyls and polychlorinated dibenzofurans exposure

Figure 4: Methylation effects on gene expression regulation across different tissue types for the CpG sites differently methylated according to father's pubertal smoking starting before age 15 years (FDR < 0.05). [Accessed on 20 June 2021]. Size of point represents -log10 p-value, colour scale shows CpG site correlation with expression; red to green represents increasing expression. In A) shape shows the tissue type, in B) shape shows genomic feature location.

Figure 5A and B. Interactome of dmCpGs associated with father's pubertal smoking starting before age 15 (FDR< 0.05). **A:** Network with high confidence score 0.7 and 20 top interactors. The interaction line colour shows dataset source: Red = experimentally determined, cyan = curated database, yellow-green = text mining. **B:** Functional enrichment for gene expression regulation, inflammatory response and innate immunity.

References

1. Skinner, M. K., Haque, C. G. B. M., Nilsson, E., Bhandari, R. & McCarrey, J. R. Environmentally Induced Transgenerational Epigenetic Reprogramming of Primordial Germ Cells and the Subsequent Germ Line. *PLoS One* **8**, (2013).
2. Senaldi, L. & Smith-raska, M. Evidence for germline non-genetic inheritance of human phenotypes and diseases. 1–12 (2020).
3. Hamad, M. F. *et al.* The status of global DNA methylation in the spermatozoa of smokers and non-smokers. *Reprod. Biomed. Online* **37**, 581–589 (2018).
4. Fernandez-Twinn, D. S., Constância, M. & Ozanne, S. E. Intergenerational epigenetic inheritance in models of developmental programming of adult disease. *Semin. Cell Dev. Biol.* **43**, 85–95 (2015).
5. Marcon, A. *et al.* Trends in smoking initiation in Europe over 40 years: A retrospective cohort study. *PLoS One* **13**, 1–14 (2018).
6. Pesce, G. *et al.* Time and age trends in smoking cessation in Europe. *PLoS One* **14**, 2000–2010 (2019).
7. Svanes, C. *et al.* Father's environment before conception and asthma risk in his children: A multi-generation analysis of the Respiratory Health In Northern Europe study. *Int. J. Epidemiol.* **46**, 235–245 (2017).
8. Accordini, S. *et al.* A three-generation study on the association of tobacco smoking with asthma. *Int. J. Epidemiol.* **47**, 1106–1117 (2018).
9. Accordini, S. *et al.* Prenatal and prepubertal exposures to tobacco smoke in men may cause lower lung function in future offspring: a three-generation study using a causal modelling approach. *Eur. Respir. J.* 2002791 (2021) doi:10.1183/13993003.02791-2020.
10. Northstone, K., Golding, J., Davey Smith, G., Miller, L. L. & Pembrey, M. Prepubertal start of father's smoking and increased body fat in his sons: further characterisation of paternal transgenerational responses. *Eur. J. Hum. Genet.* **22**, 1382–1386 (2014).
11. Golding, J. *et al.* Investigating possible trans/intergenerational associations with obesity in young adults using an exposome approach. *Front. Genet.* **10**, 1–11 (2019).
12. Knudsen, G. T. M. *et al.* Parents' smoking onset before conception as related to body mass index and fat mass in adult offspring: Findings from the RHINESSA generation study. *PLoS One* **15**, 1–21 (2020).
13. Golding, J. *et al.* Human transgenerational observations of regular smoking before puberty on fat mass in grandchildren and great-grandchildren. *Sci. Rep.* **12**, 1–8 (2022).
14. Breton, C. V. *et al.* Prenatal tobacco smoke exposure affects global and gene-specific DNA methylation. *Am. J. Respir. Crit. Care Med.* **180**, 462–467 (2009).
15. Morrow, J. D. *et al.* DNA Methylation is predictive of mortality in current and former smokers. *Am. J. Respir. Crit. Care Med.* **201**, 1099–1109 (2020).
16. Joehanes, R. *et al.* Epigenetic Signatures of Cigarette Smoking. *Circ. Cardiovasc. Genet.* **9**, 436–447 (2016).
17. Joubert, B. R. *et al.* DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. *Am. J. Hum. Genet.* **98**, 680–696 (2016).
18. Mørkve Knudsen, G. T. *et al.* Epigenome-wide association of father's smoking with offspring DNA methylation: A hypothesis-generating study. *Environ. Epigenetics* **5**, 1–10 (2019).

19. Wu, C. C. *et al.* Paternal tobacco smoke correlated to offspring asthma and prenatal epigenetic programming. *Front. Genet.* **10**, (2019).
20. Rutkowska, J., Lagisz, M., Bonduriansky, R. & Nakagawa, S. Mapping the past, present and future research landscape of paternal effects. *BMC Biol.* **18**, 1–24 (2020).
21. Christiansen, C. *et al.* Novel DNA methylation signatures of tobacco smoking with trans-ethnic effects. *Clin. Epigenetics* **13**, 1–13 (2021).
22. Joubert, B. R. *et al.* DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. *Am. J. Hum. Genet.* **98**, 680–696 (2016).
23. Metrustry, S. J. *et al.* Variants close to NTRK2 gene are associated with birth weight in female twins. *Twin Res. Hum. Genet.* **17**, 254–261 (2014).
24. Yuefeng, TanMartina, W. *et al.* Adipose tissue mTORC2 regulates ChREBP-driven de novo lipogenesis and hepatic glucose metabolism. *Nat. Commun.* | **16**, (2016).
25. Carvalho-Silva, D. *et al.* Open Targets Platform: New developments and updates two years on. *Nucleic Acids Res.* **47**, D1056–D1065 (2019).
26. Chen, W. *et al.* An epigenome-wide association study of total serum IgE in Hispanic children. *J. Allergy Clin. Immunol.* **140**, 571–577 (2017).
27. Touvier, T. *et al.* LEPROT and LEPROTL1 cooperatively decrease hepatic growth hormone action in mice. *J. Clin. Invest.* **119**, 3830–3838 (2009).
28. Haeryfar, S. M. M. *et al.* Terminal Deoxynucleotidyl Transferase Establishes and Broadens Anti-Viral CD8+ T Cell Immunodominance Hierarchies. **181**, 649–659 (2009).
29. Chitu, V. *et al.* Primed innate immunity leads to autoinflammatory disease in PSTPIP2-deficient cmo mice. *Blood* **114**, 2497–2505 (2009).
30. Murakami, Y. *et al.* TLR9–IL-2 axis exacerbates allergic asthma by preventing IL-17A hyperproduction. *Sci. Rep.* **10**, 1–17 (2020).
31. Boulakirba, S. *et al.* IL-34 and CSF-1 display an equivalent macrophage differentiation ability but a different polarization potential. *Sci. Rep.* **8**, 1–11 (2018).
32. Modena, B. D. *et al.* Gene expression correlated with severe asthma characteristics reveals heterogeneous mechanisms of severe disease. *Am. J. Respir. Crit. Care Med.* **195**, 1449–1463 (2017).
33. Shin, E. K. *et al.* Association between colony-stimulating factor 1 receptor gene polymorphisms and asthma risk. *Hum. Genet.* **128**, 293–302 (2010).
34. Caruso, M. *et al.* Increased interaction with insulin receptor substrate 1, a novel abnormality in insulin resistance and type 2 diabetes. *Diabetes* **63**, 1933–1947 (2014).
35. Kubota, N. *et al.* Differential hepatic distribution of insulin receptor substrates causes selective insulin resistance in diabetes and obesity. *Nat. Commun.* **7**, 1–16 (2016).
36. Ng, S. F. *et al.* Paternal high-fat diet consumption induces common changes in the transcriptomes of retroperitoneal adipose and pancreatic islet tissues in female rat offspring. *FASEB J.* **28**, 1830–1841 (2014).
37. Lakhani, C. M. Cigarette Smoking Significantly Alters Sperm DNA Methylation Patterns. *Physiol. Behav.* **176**, 139–148 (2019).
38. Watkins, S. H., Iles-Caven, Y., Pembrey, M., Golding, J. & Suderman, M. Grandmaternal smoking during pregnancy is associated with differential DNA methylation in peripheral blood of their grandchildren. *Eur. J. Hum. Genet.* (2022) doi:10.1038/s41431-022-01081-2.
39. Chiu, Y. L. *et al.* Validation of Self-reported Smoking with Urinary Cotinine Levels and Influence of Second-hand Smoke among Conscripts. *Sci. Rep.* **7**, 1–7 (2017).

40. Valladolid-López, M. D. C. *et al.* Evaluating the validity of self-reported smoking in Mexican adolescents. *BMJ Open* **5**, 1–8 (2015).
41. Hammer, B. *et al.* Preconceptional smoking alters spermatozoal miRNAs of murine fathers and affects offspring's body weight. *Int. J. Obes.* 1623–1627 (2021) doi:10.1038/s41366-021-00798-2.
42. Ly, L., Chan, D. & Trasler, J. M. Developmental windows of susceptibility for epigenetic inheritance through the male germline. *Semin. Cell Dev. Biol.* **43**, 96–105 (2015).

Figures

Fathers' preconception smoking and offspring DNA methylation: A two generation study

Negusse T Kitaba[1*], Gerd Toril Mørkve Knudsen[2,3*], Ane Johannessen[4], Faisal I. Rezwan[5], Andrei Malinovschi[6], Anna Oudin[7], Bryndis Benediktsdottir[8,9], David Martino[10], Francisco Javier Callejas González[11], Leopoldo Palacios Gómez[12], Mathias Holm[13], Nils Oskar Jõgi[2,3], Shyamali Dharmage[14], Svein Magne Skulstad[3], Sarah H Watkins[15], Matthew Suderman[15], Francisco Gómez-Real[2,16], Vivi Schlünssen[17,18], Cecilie Svanes[3,4]#, John W. Holloway[1,19]#.

*Joint first authors

#Joint senior authors

[1]Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, UK

[2]Department of Clinical Sciences, University of Bergen, Norway

[3]Department of Occupational Medicine, Haukeland University Hospital, Bergen, Norway

[4]Centre for International Health, Department of Global Public Health and Primary Care, University of Bergen, Norway

[5] Department of Computer Science, Aberystwyth University, Aberystwyth, UK

[6]Department of Medical Sciences: Clinical Physiology, Uppsala University, Sweden

[7]Section of Sustainable Health, Department of Public Health and Clinical Medicine, Umeå University, Sweden

[8]Department of Allergy, Respiratory Medicine and Sleep, Landspitali University Hospital, Reykjavik, Iceland

[9]Faculty of Medicine, University of Iceland, Iceland

[10]Wal-yan Respiratory Research Centre, Telethon Kids Institute, Perth, Australia

[11]Department of Pulmonology, Albacete University Hospital Complex, Albacete, Spain

[12]El Torrejón Health Centre, Andalusian Health Service, Huelva, Spain

[13]Occupational and Environmental Medicine, School of Public Health and Community Medicine, Institute of Medicine, University of Gothenburg, Sweden

[14]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Australia

[15]University of Bristol, MRC Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, Bristol, United Kingdom

[16]Department of Gynaecology and Obstetrics, Haukeland University Hospital, Bergen, Norway

[17]Department of Public Health, Work, Environment and Health, Danish Ramazzini Centre, Aarhus University Denmark, Denmark

[18]National Research Center for the Working Environment, Copenhagen, Denmark

[19]NIHR Southampton Biomedical Research Center, University Hospitals Southampton, Southampton, UK

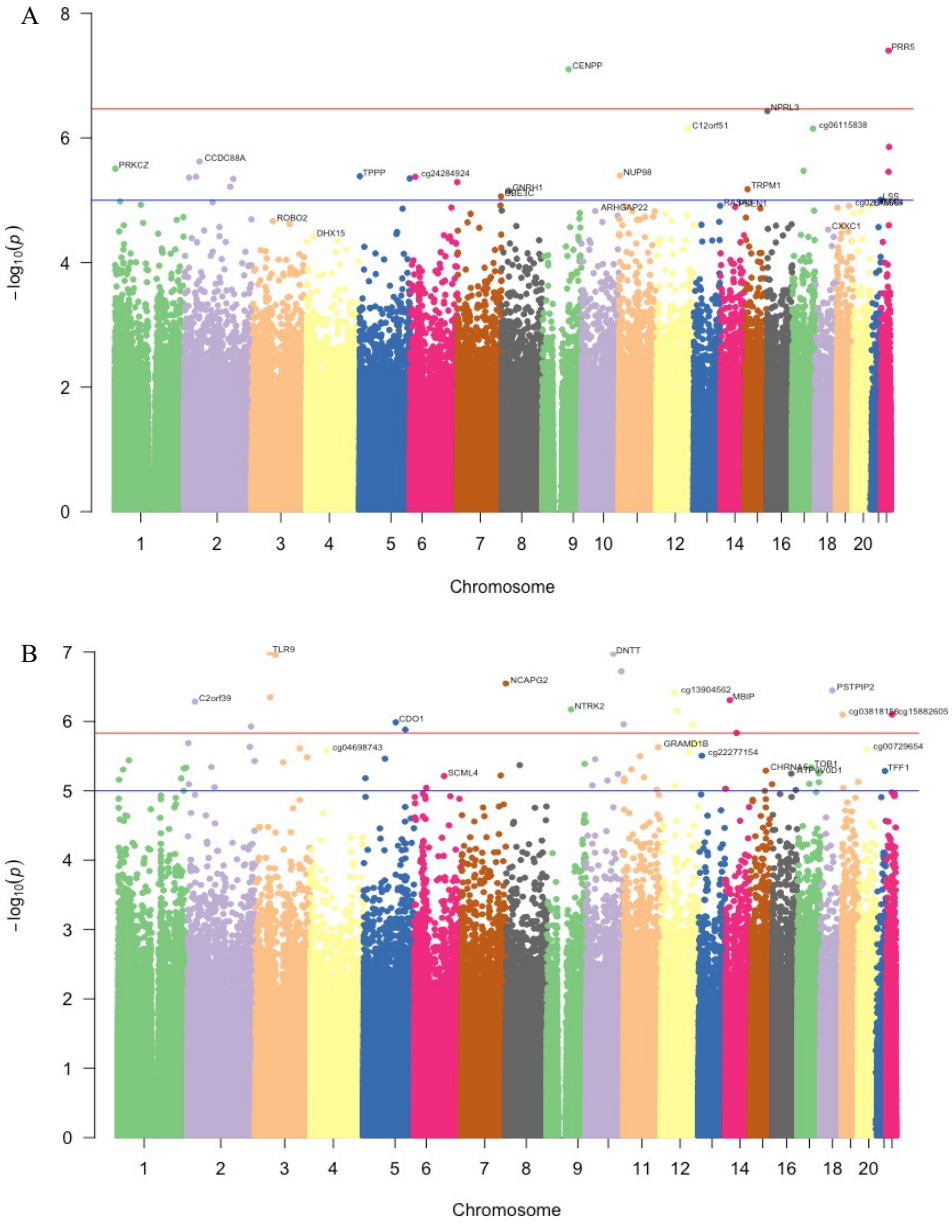


Figure 1A and B. Manhattan plot for Genome-wide distribution of dmCpGs **A:** for father's any preconception smoking, and **B:** father's pubertal smoking starting before age 15. The red line shows genome-wide significance, the blue is the suggestive line. The y-axis represents -log10 of the p-value for each dmCpG (indicated by dots) showing the strength of association. The x-axis shows the position across autosomal chromosomes. The top dmCpGs on each chromosome were annotated to the closest gene.

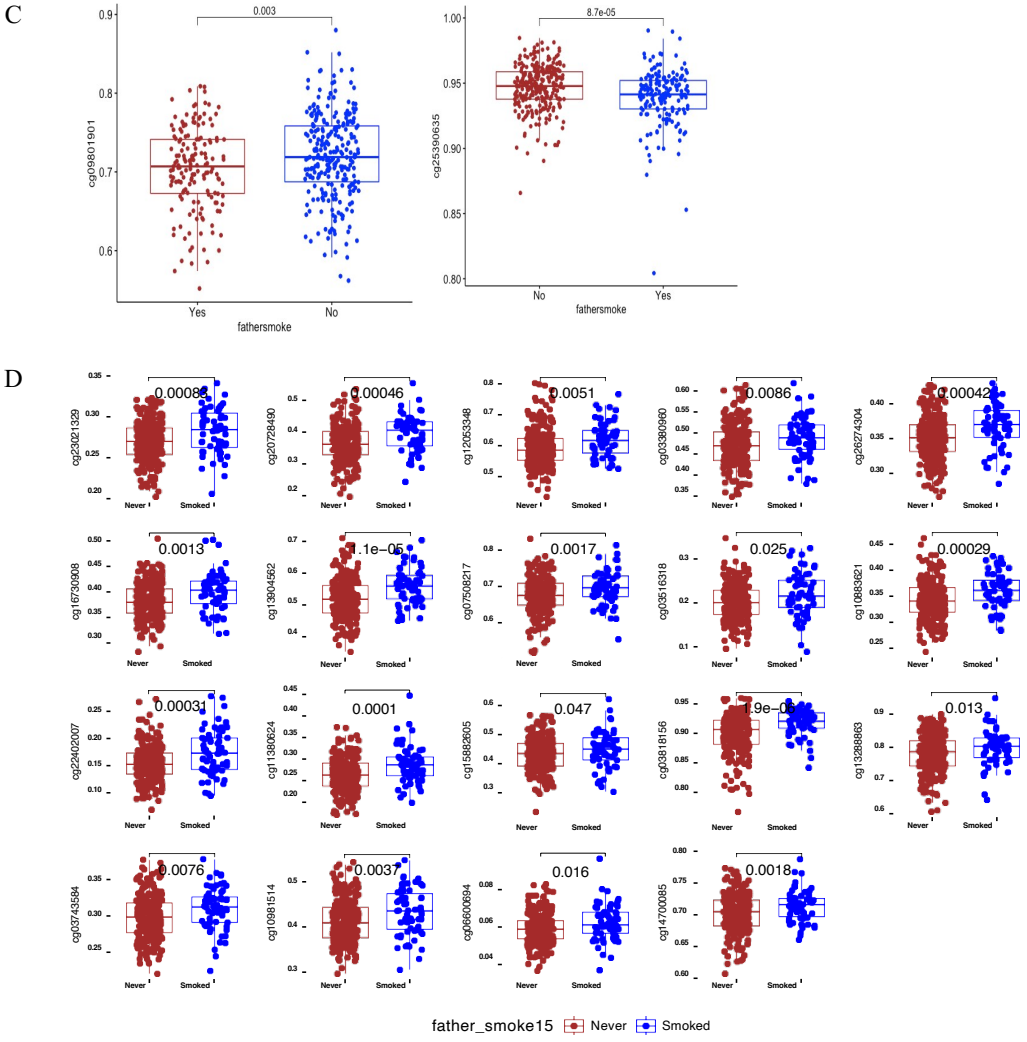


Figure 1C and D. Comparison of methylation differences for **C**: for father's any preconception smoking, and **D**: for father's pubertal smoking starting before age 15.

Figure 2A. Circos plots showing genome-wide distribution across autosomal chromosomes of dmCpGs associated with **A:** personal smoking (in offspring), **B:** mother's smoking, **C:** father's any preconception smoking, and **D:** father's pubertal smoking starting before age 15. Each dot represents a CpG site; the radial line shows the -log10 p-value for each EWAS. Zoomed dots show significant sites in one of the EWAS; each zoomed dot colour shows a unique CpG site specific locus in all 4 EWASs.

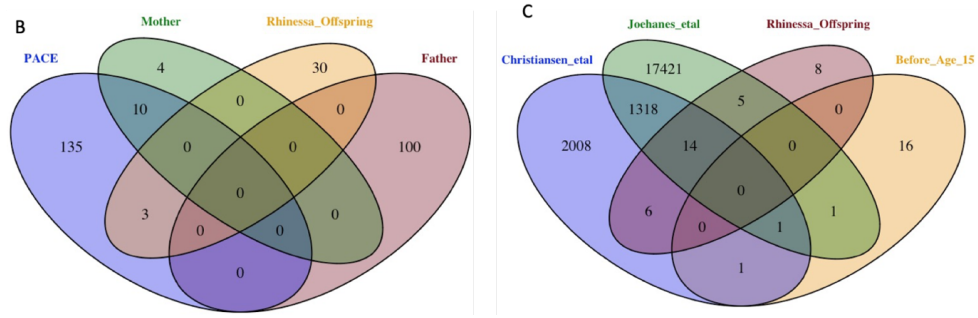


Figure 2B and C. Venn diagram showing EWAS CpG top hits for personal smoking, mother's smoking (FDR<0.005), father's any preconception smoking (top 100 dmCpGs), and father's pubertal smoking starting before age 15 (FDR<0.05) in the RHINESSA cohort, which are shared with top hits from meta-analysis of **B:** mother smoking (blue oval) as reported by Joubert et al 2016, and **C:** personal cigarette smoking signature as reported by Christiansen et al 2021 (blue) and by Joehanes et al 2016 (green).

Figure 3: Traits associated with the CpG sites that in EWAS were identified to be differentially methylated according to **A:** father's any preconception smoking, **B:** father's pubertal smoking starting before age 15, **C:** Mother's smoking and **D:** personal smoking

*PPBAPDE: perinatal polychlorinated biphenyls and polychlorinated dibenzofurans exposure

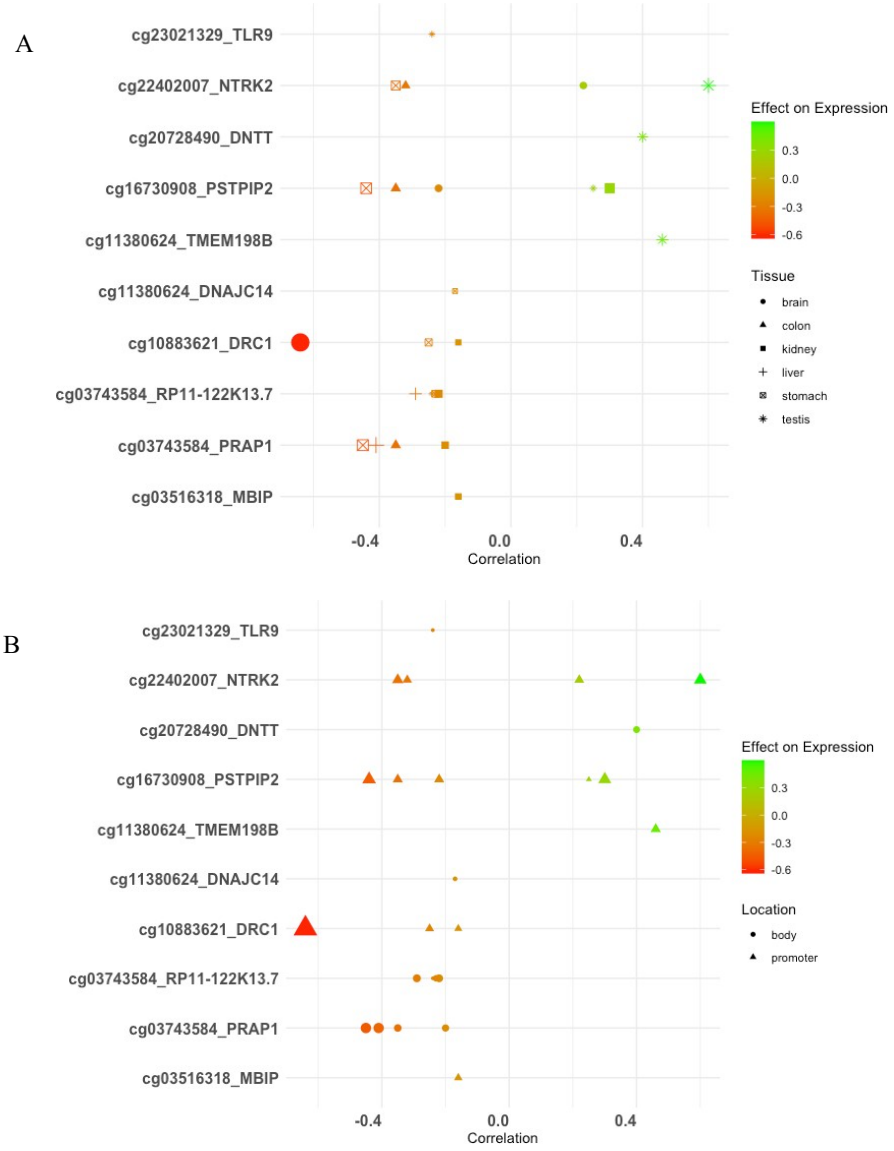
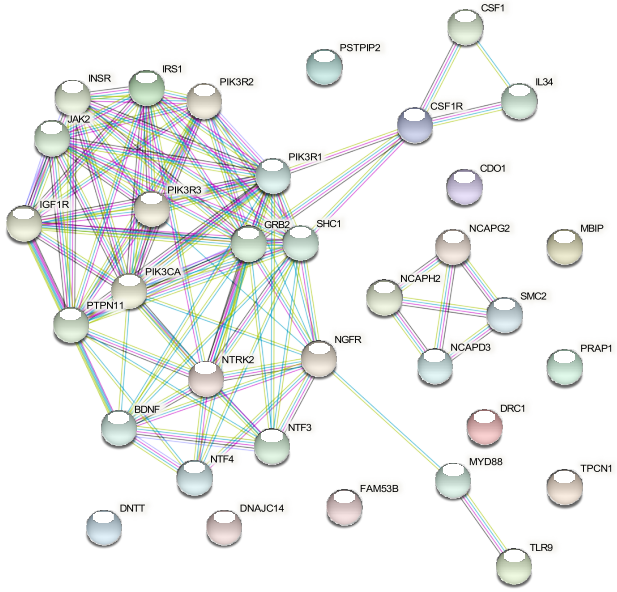


Figure 4: Methylation effects on gene expression regulation across different tissue types for the CpG sites differently methylated according to father's pubertal smoking starting before age 15 years (FDR < 0.05). [Accessed on 20 June 2021]. Size of point represents -log10 p-value, colour scale shows CpG site correlation with expression; red to green represents increasing expression. In A) shape shows the tissue type, in B) shape shows genomic feature location.

A



B

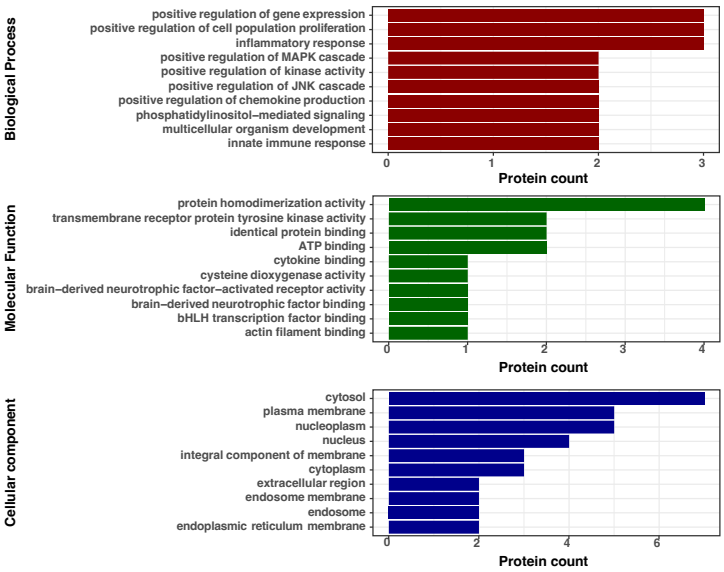


Figure 5A and B. Interactome of dmCpGs associated with father's pubertal smoking starting before age 15 (FDR< 0.05). **A:** Network with high confidence score 0.7 and 20 top interactors. The interaction line colour shows dataset source: Red = experimentally determined, cyan = curated database, yellow-green = text mining. **B:** Functional enrichment for gene expression regulation, inflammatory response and innate immunity.

Supplementary Data

Fathers' preconception smoking and offspring DNA methylation: A two generation study

Negusse T Kitaba[1*], Gerd Toril Mørkve Knudsen[2,3*], Ane Johannessen[4], Faisal I. Rezwan[5], Andrei Malinovschi[6], Anna Oudin[7], Bryndis Benediktsdottir[8,9], David Martino[10], Francisco Javier Callejas González[11], Leopoldo Palacios Gómez[12], Mathias Holm[13], Nils Oskar Jõgi[2,3], Shyamali Dharmage[14], Svein Magne Skulstad[3], Sarah H Watkins[15], Matthew Suderman[15], Francisco Gómez-Real[2,16], Vivi Schlünssen[17,18], Cecilie Svanes[3,4]#, John W. Holloway[1,19]#.
*Joint first authors
#Joint senior authors
[1]Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, UK
[2]Department of Clinical Sciences, University of Bergen, Norway
[3]Department of Occupational Medicine, Haukeland University Hospital, Bergen, Norway
[4]Centre for International Health, Department of Global Public Health and Primary Care, University of Bergen, Norway
[5] Department of Computer Science, Aberystwyth University, Aberystwyth, UK
[6]Department of Medical Sciences: Clinical Physiology, Uppsala University, Sweden
[7]Section of Sustainable Health, Department of Public Health and Clinical Medicine, Umeå University, Sweden
[8]Department of Allergy, Respiratory Medicine and Sleep, Landspitali University Hospital, Reykjavik, Iceland
[9]Faculty of Medicine, University of Iceland, Iceland
[10]Wal-yan Respiratory Research Centre, Telethon Kids Institute, Perth, Australia
[11]Department of Pulmonology, Albacete University Hospital Complex, Albacete, Spain
[12]El Torrejón Health Centre, Andalusian Health Service, Huelva, Spain
[13]Occupational and Environmental Medicine, School of Public Health and Community Medicine, Institute of Medicine, University of Gothenburg, Sweden
[14]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Australia
[15]University of Bristol, MRC Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, Bristol, United Kingdom
[16]Department of Gynaecology and Obstetrics, Haukeland University Hospital, Bergen, Norway
[17]Department of Public Health, Work, Environment and Health, Danish Ramazzini Centre, Aarhus University Denmark, Denmark
[18]National Research Center for the Working Environment, Copenhagen, Denmark
[19]NIHR Southampton Biomedical Research Center, University Hospitals Southampton, Southampton, UK

Supplementary Methods

Study design and data

Offspring were participants in the RHINESSA study (www.rhinessa.net). Parent data, including detailed information on smoking habits, were retrieved from the population-based European Community Respiratory Health Survey (ECRHS, www.ecrhs.org) and/or the Respiratory Health in Northern Europe study (RHINE, www.rhine.nu). Medical research committees in each study centre approved the studies, and each participant gave written consent. Father's smoking and age of starting/quitting was reported in interviews/questionnaires, and related to offspring's birth year, to define categories: never smoked (N=547), any preconception smoking (N=328), preconception smoking with onset <15 years (pubertal smoking) (N=64) (cut point based on mean age of voice break 14.5 years, first nocturnal seminal emission 14.8 years). Personal smoking was classified as current, ex- or never smoking. Maternal smoking was defined by offspring's report on mothers' smoking during their childhood/pregnancy.

Methylation profiling and processing

DNAm in offspring was measured in DNA extracted from peripheral blood, using a simple salting out procedure[1]. Bisulfite-conversion was undertaken using EZ 96-DNA methylation kits (Zymo Research, Irvine, CA, USA) at the Oxford Genomics Centre (Oxford, UK) and methylation assessed using Illumina Infinium MethylationEPIC Beadchip arrays (Illumina, Inc. CA, USA) with samples randomly distributed on microarrays to control against batch effects.

Data analysis was undertaken using Statistical Computing Program R, version 3.6.1[2]. Methylation intensity files were processed and quality was assessed using minfi[3] and Mefil[4]. Methylation distribution for outliers were assessed using density and multidimensional

scaling plots, methylated vs unmethylated ratio plot, sex mismatch and sex outliers, control probes and bisulphite conversion efficiency.

Normalization was carried out using BMIQ, which adjusts intra-sample the beta-values of type 2 design and type 1 probes[5]. To remove technical variation detected by champ.SVD function within the CHAMP package[6], ComBat from SVA was applied on plate and slides for batch effect correction[7]. Probes were excluded from analysis using the following criteria: detection p-value above 0.01 (n=24566 probes), probes associated with SNP, probes with a beadcount <3 in at least 5% of samples (n=1437), multiple locations, non-cg probes (n=2624), probes on the X or Y chromosomes (n= 16556) and cross-reactive probes on the EPIC array (n=43000)[8]. Cell-type proportion was estimated with EpiDISH (epigenetics Dissection of Intra-Sample Heterogeneity) [9]. Following processing, 726,661 CpGs were retained for analysis.

Statistical analysis

We ran two EWAS on preconception father's smoking as exposure (any preconception smoking, and prepuberty smoking) with DNA methylation as outcome. To identify differentially-methylated Cytosine-phosphate-Guanine (CpG) sites (dmCPG), robust multiple linear regression models were applied on beta values using limma[10] adjusting for offspring's sex, age, personal and mother's smoking, and cell-type proportions (B-cells, Natural killer cells, CD4 T-cells, CD8 T-cells, Monocyte, Neutrophils) at significance level of false discovery rate (FDR)[11] corrected p-value<0.05. Eosinophils were not included due to a very low estimate. Manhattan plots were generated using qqman[12] and circos plot with CMplot R package[13]. Inflation from systematic biases was adjusted using BACON[14]. Differentially methylated regions were detected using dmrff[15]. Transcription factor binding site prediction was performed using eFORGE TF[16]. Gene-disease association was identified using open

target[17]. Identified dmCpGs were compared against EWAS atlas for association with known biological traits[18]. To gain biological insight regarding the dmCpGs mapped to genes, gene interactors were identified using String[19] and enrichment was performed using UniprotR[20] and gometh[21].

We compared our EWAS results with findings from meta-analysis of EPIC DNA methylation associated with personal smoking from four population-based cohorts[22], personal smoking-methylation effects from 16 cohorts using 450K arrays[23]; and the Pregnancy and Childhood Epigenetics Consortium (PACE) meta-analysis of mother smoking on offspring cordblood methylation[24].

ALSPAC Cohort description

ALSPAC is a pre-birth cohort designed to determine the environmental and genetic factors that are associated with health and development of the study offspring (1-3). Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees (4). Consent for biological samples has been collected in accordance with the Human Tissue Act (2004). Informed consent for the use of data collected via questionnaires and clinics was obtained from participants following the recommendations of the ALSPAC Ethics and Law Committee at the time.

Pregnant women resident in Avon, UK, with expected dates of delivery 1st April 1991 to 31st December 1992 were invited to take part in the study. The initial number of pregnancies enrolled is 14,541 (for these at least one questionnaire has been returned or a “Children in Focus” clinic had been attended by 19/07/99). Of these initial pregnancies, there was a total

of 14,676 fetuses, resulting in 14,062 live births and 13,988 children who were alive at 1 year of age.

When the oldest children were approximately 7 years of age, an attempt was made to bolster the initial sample with eligible cases who had failed to join the study originally. As a result, when considering variables collected from the age of seven onwards (and potentially abstracted from obstetric notes) there are data available for more than the 14,541 pregnancies mentioned above. The number of new pregnancies not in the initial sample (known as Phase I enrolment) that are currently represented on the built files and reflecting enrolment status at the age of 24 is 913 (456, 262 and 195 recruited during Phases II, III and IV respectively), resulting in an additional 913 children being enrolled. The phases of enrolment are described in more detail in the cohort profile paper and its update (2, 3). The total sample size for analyses using any data collected after the age of seven is therefore 15,454 pregnancies, resulting in 15,589 fetuses. Of these 14,901 were alive at 1 year of age.

Please note that the study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool: http://www.bristol.ac.uk/alspac/researchers/our-data/

ALSPAC fully supports Wellcome and the RCUK policies on open access. The process for obtaining access to data is described on the study website: http://www.bristol.ac.uk/alspac/researchers/data-access/. The datasets for this study will not be made publicly available, as in order to preserve confidentiality of the participants it is important that the ALSPAC access rules are taken into account. The ALSPAC study website contains details of all the data that are available through a fully searchable data dictionary: http://www.bristol.ac.uk/alspac/researchers/our-data/.

1. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, Molloy L, Ness A, Ring S, Davey Smith G. Cohort Profile: The 'Children of the 90s'; the index offspring of The Avon Longitudinal Study of Parents and Children (ALSPAC). International Journal of Epidemiology 2013; 42: 111-127.
2. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, Henderson J, Macleod J, Molloy L, Ness A, Ring S. Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. International journal of epidemiology. 2013 Feb 1;42(1):97-110.
3. Northstone K, Lewcock M, Groom A, Boyd A, Macleod J, Timpson NJ, Wells N. The Avon Longitudinal Study of Parents and Children (ALSPAC): an updated on the enrolled sample of index children in 2019. Wellcome Open research 2019; 4:51 (https://doi.org/10.12688/wellcomeopenres.15132.1)
4. Birmingham, K. (2018). *Pioneering ethics in longitudinal studies: The early development of the ALSPAC Ethics & Law Committee*. Bristol: Policy Press

Acknowledgements

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses.

Funding of ALSPAC

The UK Medical Research Council and Wellcome Trust (Grant ref: 217065/Z/19/Z) and the University of Bristol currently provide core support for ALSPAC. A comprehensive list of

grants funding is available on the ALSPAC website

(http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf).

Supplementary Table E1: Top dmCpG sites associated with fathers' any preconception smoking onset (N=875)

CpG	Coefficient	Average	P- value	Adj.P	Chromosome	Location	Gene	Ref Gene
cg00870527	-0.02	0.50	3.97E-08	0.03	22	N_Shelf	PRR5	5'UTR
cg08541349	-0.01	0.88	7.92E-08	0.03	9	OpenSea	CENPP	Body
cg01000965	0.00	0.94	3.70E-07	0.09	16	OpenSea	NPRL3	Body
cg06115838	-0.09	0.53	7.09E-07	0.10	17	S_Shelf	cg06115838	
cg12819747	0.00	0.94	7.15E-07	0.10	12	Island	C12orf51	Body
cg25390635	-0.01	0.94	1.40E-06	0.17	22	Island	GRAMD4	TSS200
cg00061114	-0.02	0.76	2.38E-06	0.19	2	OpenSea	CCDC88A	Body
cg19646139	-0.01	0.93	3.12E-06	0.19	1	S_Shelf	PRKCZ	Body
cg26884359	-0.01	0.84	3.38E-06	0.19	17	N_Shelf	cg26884359	
cg22778120	-0.02	0.54	3.50E-06	0.19	22	N_Shelf	PRR5	5'UTR
cg06354075	-0.01	0.90	4.01E-06	0.19	11	OpenSea	NUP98	Body
cg05458764	-0.01	0.92	4.12E-06	0.19	5	Island	TPPP	Body
cg25817279	-0.01	0.83	4.19E-06	0.19	2	OpenSea	cg25817279	
cg24284924	0.00	0.95	4.21E-06	0.19	6	OpenSea	cg24284924	
cg06597248	-0.02	0.62	4.34E-06	0.19	2	OpenSea	cg06597248	
cg09543474	0.00	0.93	4.47E-06	0.19	5	OpenSea	CNOT6	Body
cg17931890	0.00	0.07	4.56E-06	0.19	2	Island	cg17931890	
cg17545662	-0.01	0.94	5.13E-06	0.21	6	OpenSea	C6orf70	Body
cg06868293	-0.01	0.92	6.07E-06	0.23	2	OpenSea	SCN2A	Body
cg10231096	-0.01	0.92	6.65E-06	0.24	15	OpenSea	TRPM1	Body
cg25710809	-0.01	0.79	7.08E-06	0.24	8	OpenSea	GNRH1	TSS1500
cg03657121	-0.01	0.82	8.70E-06	0.25	7	OpenSea	UBE3C	Body
cg09831081	-0.01	0.87	9.86E-06	0.25	21	N_Shore	LSS	Body

Supplementary Table E2: Sex-stratified dmCpGs (FDR<=0.05) associated with fathers' any preconception smoking in male (N=457) and female (N=418) offspring

Offspring sex	CpG	Coefficient	Average	P- value	Adj.P	Chr	Location	Gene	Ref Gene
Males	cg05193832	-0.01	0.94	5.16E-08	0.04	11	OpenSea	KCNJ1	5'UTR
	cg25390635	-0.01	0.94	1.05E-07	0.04	22	Island	GRAMD4	TSS200
	cg22905274	-0.01	0.04	2.55E-07	0.05	4	Island	TRIM2	Body
	cg02518394	-0.01	0.91	2.87E-07	0.05	17	OpenSea	MYADML2	TSS1500
Females	cg09801901	-0.02	0.71	3.18E-08	0.02	8	OpenSea	LEPROTL1	Body

Supplementary Table E3: dmCpGs associated with father's smoking onset before age 15 years (N=304, FDR<0.05)

CpG	Coefficient	Average	P-value	Adj.P	Chromosome	Location	Gene	Ref Gene
cg23021329	0.01	0.27	6.07E-08	0.03	3	S_Shore	TLR9	Body
cg20728490	0.03	0.37	6.30E-08	0.03	10	OpenSea	DNTT	5'UTR
cg12053348	0.04	0.61	6.55E-08	0.03	3	OpenSea	cg12053348	
cg03380960	0.02	0.48	1.11E-07	0.03	10	OpenSea	FAM53B	Body
cg26274304	0.02	0.36	1.66E-07	0.04	7	N_Shore	NCAPG2	5'UTR
cg16730908	0.02	0.39	2.08E-07	0.04	18	S_Shore	PSTPIP2	TSS1500
cg13904562	0.04	0.53	2.26E-07	0.04	12	OpenSea	cg13904562	
cg07508217	0.03	0.69	2.60E-07	0.04	3	OpenSea	cg07508217	
cg03516318	0.03	0.21	2.87E-07	0.04	14	OpenSea	MBIP	TSS1500
cg10883621	0.02	0.35	3.00E-07	0.04	2	Island	C2orf39	TSS200
cg22402007	0.02	0.16	3.88E-07	0.04	9	N_Shore	NTRK2	TSS1500
cg11380624	0.02	0.27	4.04E-07	0.04	12	N_Shore	DNAJC14	5'UTR
cg15882605	0.02	0.44	4.60E-07	0.04	22	OpenSea	cg15882605	
cg03818156	0.02	0.91	4.62E-07	0.04	19	OpenSea	cg03818156	
cg13288863	0.02	0.79	5.93E-07	0.05	5	N_Shore	CDO1	Body
cg03743584	0.02	0.30	6.33E-07	0.05	10	OpenSea	PRAP1	1stExon
cg10981514	0.02	0.42	6.49E-07	0.05	12	OpenSea	TPCN1	Body
cg06600694	0.00	0.06	6.82E-07	0.05	2	Island	IRS1	TSS200
cg14700085	0.02	0.71	7.56E-07	0.05	5	OpenSea	CSF1R	Body
cg25406294	0.02	0.78	8.42E-07	0.05	14	OpenSea	cg25406294	

Supplementary Table E4: dmCpGs associated with offspring's own smoking status (N=875, FDR<0.05)

CpG	Coefficient	Average	P-value	Adj.P	Chromosome	Location	Gene	Ref Gene
cg05575921	-0.04	0.83	2.68E-54	1.94E-48	5	N_Shore	AHRR	Body
cg21566642	-0.04	0.55	8.21E-29	2.97E-23	2	Island	cg21566642	
cg06644428	-0.03	0.12	5.35E-20	1.29E-14	2	Island	cg06644428	
cg01940273	-0.03	0.58	1.50E-16	2.72E-11	2	Island	cg01940273	
cg17739917	-0.03	0.39	4.85E-16	6.77E-11	17	S_Shelf	RARA	5'UTR
cg03636183	-0.03	0.63	5.61E-16	6.77E-11	19	N_Shore	F2RL3	Body
cg21911711	-0.02	0.84	7.79E-12	8.05E-07	19	N_Shore	F2RL3	TSS1500
cg25189904	-0.03	0.38	6.96E-11	6.30E-06	1	S_Shore	GNG12	TSS1500
cg12806681	-0.01	0.94	1.31E-10	1.06E-05	5	N_Shore	AHRR	Body
cg21161138	-0.02	0.73	1.60E-10	1.16E-05	5	OpenSea	AHRR	Body
cg09338374	0.02	0.49	9.29E-10	6.11E-05	22	S_Shelf	cg09338374	
cg24838345	-0.02	0.84	1.44E-09	8.13E-05	8	N_Shelf	MTSS1	Body
cg26703534	-0.02	0.68	1.56E-09	8.13E-05	5	S_Shelf	AHRR	Body
cg14753356	-0.022	0.41	1.57E-09	8.13E-05	6	OpenSea	cg14753356	
cg18110140	-0.024	0.44	1.95E-08	0.00094207	15	OpenSea	cg18110140	
cg26718213	0.035	0.22	3.50E-08	0.00158405	2	Island	SNED1	Body
cg05086879	-0.016	0.83	6.03E-08	0.00256768	22	OpenSea	MGAT3	5'UTR
cg21322436	-0.01	0.25	6.71E-08	0.00269941	7	N_Shore	CNTNAP2	TSS1500
cg25648203	-0.015	0.80	9.27E-08	0.00352967	5	OpenSea	AHRR	Body
cg20832643	-0.014	0.89	1.23E-07	0.00428032	6	OpenSea	TRAF3IP2-AS1	Body
cg09935388	-0.026	0.76	1.24E-07	0.00428032	1	Island	GFI1	Body
cg18096787	0.016	0.59	2.44E-07	0.00803438	21	OpenSea	cg18096787	
cg24090911	-0.018	0.75	3.31E-07	0.01041909	5	OpenSea	AHRR	Body
cg19859270	-0.006	0.95	4.17E-07	0.01258376	3	OpenSea	GPR15	1stExon
cg25159376	-0.002	0.021	5.53E-07	0.01600192	14	Island	KLHDC1	TSS200
cg17025708	-0.006	0.92	6.49E-07	0.01783236	10	OpenSea	VTI1A	Body
cg02978227	-0.006	0.94	6.65E-07	0.01783236	3	OpenSea	cg02978227	
cg26707709	0.017	0.11	1.28E-06	0.03257775	2	Island	SNED1	Body
cg23577033	-0.005	0.932	1.31E-06	0.03257775	10	S_Shelf	cg23577033	

cg25401612	-0.03	0.72	2.02E-06	0.04794569	12	OpenSea	cg25401612	
cg13525276	0.02	0.27	2.05E-06	0.04794569	14	OpenSea	TSHR	Body
cg17260354	0.014	0.45	2.13E-06	0.04819969	17	N_Shore	CDK3	3'UTR
cg06117824	-0.004	0.96	2.28E-06	0.04990539	1	S_Shelf	TMEM51	Body
cg25949550	-0.006	0.086	2.43E-06	0.05178064	7	S_Shore	CNTNAP2	Body
cg15342087	-0.006	0.914	2.74E-06	0.05448434	6	OpenSea		
cg22635676	0.046	0.24	2.79E-06	0.05448434	2	Island	SNED1	Body
cg21747070	-0.01	0.52	2.79E-06	0.05448434	5	N_Shore		

Supplementary Table E5: dmCpGs associated with mothers' smoking during offspring's childhood (N=875, FDR<0.05)

CpG	Coefficient	Average	P-value	Adj.P	Chromosome	Location	Gene	Ref Gene
cg19089201	0.03	0.71746013	1.47E-10	0.0001	7	Island	MYO1G	3'UTR
cg12803068	0.06	0.67728595	3.31E-10	0.0001	7	S_Shore	MYO1G	Body
cg05549655	0.02	0.18071069	6.91E-09	0.0015	15	Island	CYP1A1	TSS1500
cg14179389	-0.03	0.26117365	8.48E-09	0.0015	1	Island	GFI1	Body
cg04180046	0.03	0.45264404	1.73E-08	0.0025	7	Island	MYO1G	Body
cg25949550	-0.01	0.08628854	2.10E-08	0.0025	7	S_Shore	CNTNAP2	Body
cg13305373	0.01	0.91366976	5.71E-08	0.0059	17	OpenSea	RGS9	Body
cg22549041	0.03	0.32680072	1.63E-07	0.0147	15	Island	CYP1A1	TSS1500
cg05009104	0.03	0.69281994	2.53E-07	0.0180	7	S_Shore	MYO1G	Body
cg11924019	0.02	0.3723796	2.63E-07	0.0180	15	Island	CYP1A1	TSS1500
cg12101586	0.03	0.47187709	2.74E-07	0.0180	15	Island	CYP1A1	TSS1500
cg05777089	0.01	0.08624407	6.30E-07	0.0380	1	OpenSea	BRINP3	5'UTR
cg13570656	0.03	0.38212374	6.84E-07	0.0381	15	Island	CYP1A1	TSS1500
cg04785284	0.01	0.9528626	8.18E-07	0.0423	16	OpenSea	cg04785284	

Supplementary Table E6: Transcription factor Enrichment (q-value<0.05) for dmCpGs (FDR<0.05) associated with fathers' smoking onset before age 15 years

TF	Database	p-value	q-value	CpGs	Gene
V_GZF1_01	TRANSFAC	7.09E-06	0.00042543	cg22402007	NTRK2
HSF1_HSF_2	Taipale/SELEX	7.70E-06	0.00042543	cg12053348	N/A
HSF4_HSF_1	Taipale/SELEX	8.88E-06	0.00042543	cg12053348	N/A
HSF1_HSF_1	Taipale/SELEX	1.03E-05	0.00042543	cg12053348	N/A
MA0090.1-TEAD1	JASPAR	4.06E-05	0.00091734	cg10981514	TPCN1
V_TEF_01	TRANSFAC	4.06E-05	0.00091734	cg10981514	TPCN1
V_ELK1_03	TRANSFAC	4.60E-05	0.00091734	cg10981514	TPCN1
V_TAL1BETAITF2_01	TRANSFAC	4.67E-05	0.00091734	cg16730908	PSTPIP2
V_TAL1BETAE47_01	TRANSFAC	4.98E-05	0.00091734	cg16730908	PSTPIP2
V_ELK1_04	TRANSFAC	8.53E-05	0.00141511	cg10981514	TPCN1
V_HSF1_Q6	TRANSFAC	0.00014091	0.00212498	cg12053348	N/A
Ets1	UniProbe	0.00015508	0.00214386	cg10981514	TPCN1
MA0028.1-ELK1	JASPAR	0.00026585	0.00330968	cg10981514	TPCN1
V_ETS_Q4	TRANSFAC	0.00027932	0.00330968	cg15882605	N/A
Tcfe2a_secondary	UniProbe	0.00045637	0.00504704	cg16730908	PSTPIP2
ETV6_ETS_1	Taipale/SELEX	0.00071753	0.00743932	cg10981514	TPCN1
MA0079.2-SP1	JASPAR	0.00083946	0.00758437	cg11380624/ cg06600694	DNAJC14 / IRS1
V_ZBP89_Q4	TRANSFAC	0.00086685	0.00758437	cg11380624	DNAJC14
V_GADP_01	TRANSFAC	0.00086868	0.00758437	cg15882605	N/A
V_GABP_B	TRANSFAC	0.00114402	0.00948899	cg15882605	N/A
V_SP1_Q2_01	TRANSFAC	0.00197707	0.0156178	cg15882605	N/A
V_SP4_Q5	TRANSFAC	0.00257797	0.0194388	cg11380624/ cg06600694	DNAJC14/ IRS1
Zfp740_primary	UniProbe	0.0036371	0.0262327	cg06600694	IRS1

Supplementary Table E7: EWAS Atlas lookup for replication (accessed on 20 June 2021) for top dmCpGs.

CpG	Gene	Traits
cg00870527	PRR5	Gulf War Illness and Papuan ancestry proportions
cg08541349	CENPP	No report
cg01000965	NPRL3	No report
cg05193832	KCNJ1	No report
cg25390635	GRAMD4	serum immunoglobulin E (IgE) levels
cg22905274	TRIM2	No report
cg02518394	MYADML2	autism spectrum disorders (ASD)
cg09801901	LEPROTL1	No report
cg23021329	TLR9	atopy and fractional exhaled nitric oxide
cg20728490	DNTT	Smoking, aging
cg12053348	N/A	Aging
cg03380960	FAM53	Aging
cg26274304	NCAPG2	No report
cg16730908	PSTPIP2	Smoking
cg13904562	N/A	No report
cg07508217	N/A	No report
cg03516318	MBIP	BMI
cg10883621	C2orf39	B Acute Lymphoblastic Leukaemia with t(1;19)(q23;p13.3); E2A-PBX1 (TCF3-PBX1)
cg22402007	NTRK2	B Acute Lymphoblastic Leukaemia with t(12;21)(p13.2;q22.1); ETV6-RUNX1
cg11380624	DNAJC14	Cancer
cg15882605	N/A	No report
cg03818156	N/A	No report
cg13288863	CDO1	No report
cg03743584	PRAP1	AD
cg10981514	TPCN1	Crohn's disease
cg06600694	IRS1	No report
cg14700085	CSF1R	No report

Supplementary Table E8: Differentially Methylated Regions for father smoking onset before age 15 years.

Region	Site	Chr	Position	Gene	Estimate	P-value	dmr.start	dmr.end	dmr.z	dmr P.adj.
20599	635922	10	98064175	DNTT	0.03171911	1.04E-07	98064175	98064175	5.32642756	0.0841472
43571	649032	3	53461190		0.02527496	1.05E-07	53461190	53461190	5.32370777	0.08541576
6282	253453	10	126390659	FAM53B	0.00914814	0.00847902	126390317	126390659	5.25221053	0.12625419
6282	504848	10	126390317	FAM53B	0.0063052	0.02895291	126390317	126390659	5.25221053	0.12625419
24736	585899	12	56222881	DNAJC14	0.02359829	2.28E-07	56222881	56222881	5.17393919	0.19255909
27284	207465	14	61070289		0.02108314	2.39E-07	61070289	61070289	5.16542124	0.20153858
2688	141869		48690974		0.04041497	4.05E-07	48690974	48691036	5.15848642	0.20914664
2688	681241	12	48691036		0.02183902	7.20E-06	48690974	48691036	5.15848642	0.20914664
34971	273463	19	3073159		0.01729797	3.68E-07	3073159	3073159	5.08022606	0.31670631
25389	341664	12	113678406	TPCN1	0.02375583	4.26E-07	113678406	113678406	5.05092116	0.36937942
3443	273124	3	72536534		0.03521499	4.93E-07	72536534	72536534	5.0218192	0.43000027
705	350648	17	76130139	TMC8	0.02054525	4.64E-06	76129984	76130305	5.00220297	0.47615976
705	647240	17	76130305	TMC8	0.01594663	6.66E-06	76129984	76130305	5.00220297	0.47615976
705	405230	17	76129984	TMC8;TMC6	0.01500805	4.97E-05	76129984	76130305	5.00220297	0.47615976
47604	378684	5	115149472	CDO1	0.02026539	5.81E-07	115149472	115149472	4.98898749	0.50991255
9368	247050	3	159570903	SCHIP1	0.01794319	6.27E-07	159570903	159570903	4.97338472	0.55273525
324	705691	17	58499720	C17orf64	0.02077587	1.28E-05	58499679	58499720	4.92454708	0.71035553
324	504219	17	58499706	C17orf64	0.02006096	7.75E-05	58499679	58499720	4.92454708	0.71035553
324	30671	17	58499679	C17orf64	0.01173987	0.00023255	58499679	58499720	4.92454708	0.71035553
324	36291	17	58499700	C17orf64	0.0137547	0.00770448	58499679	58499720	4.92454708	0.71035553
7044	542568	14	36790242	MBIP;MBIP	0.02684703	8.04E-07	36790242	36790242	4.92331005	0.71486272
9085	389031	22	37562418		0.02544397	8.98E-07	37562418	37562418	4.90066973	0.80238468

Supplementary Table E9: Gene Ontology Terms from gometh for dmCpGs (FDR<0.05) associated with fathers' smoking onset before age 15years. BP: biological processing; MF: molecular function; CC cellular components; N: number of genes in GO term; De: number of differentially methylated genes; P.De: P-value for over-representation of GO term

GO_name	ONTOLOGY	TERM	N	DE	P.DE
GO:0000173	BP	inactivation of MAPK activity involved in osmosensory signaling pathway	1	1	0.001
GO:0032715	BP	negative regulation of interleukin-6 production	35	2	0.001
GO:0003912	MF	DNA nucleotidylexotransferase activity	1	1	0.001
GO:0060753	BP	regulation of mast cell chemotaxis	1	1	0.001
GO:1902622	BP	regulation of neutrophil migration	1	1	0.001
GO:0048709	BP	oligodendrocyte differentiation	24	2	0.002
GO:0004860	MF	protein kinase inhibitor activity	27	2	0.002
GO:0060175	MF	brain-derived neurotrophic factor-activated receptor activity	1	1	0.002
GO:0036019	CC	Endolysosome	1	1	0.002
GO:0044147	BP	negative regulation of development of symbiont involved in interaction with host	1	1	0.003
GO:0052362	BP	catabolism by host of symbiont protein	1	1	0.003
GO:1905597	BP	positive regulation of low-density lipoprotein particle receptor binding	1	1	0.003
GO:1905599	BP	positive regulation of low-density lipoprotein receptor activity	1	1	0.003
GO:0006304	BP	DNA modification	2	1	0.003

GO:0045356	BP	positive regulation of interferon-alpha biosynthetic process	4	1	0.003
GO:0010810	BP	regulation of cell-substrate adhesion	2	1	0.003
GO:0034163	BP	regulation of toll-like receptor 9 signaling pathway	2	1	0.004
GO:0006043	BP	glucosamine catabolic process	2	1	0.004
GO:0052405	BP	negative regulation by host of symbiont molecular function	3	1	0.004
GO:0019471	BP	4-hydroxyproline metabolic process	3	1	0.004
GO:0019834	MF	phospholipase A2 inhibitor activity	3	1	0.004
GO:0099183	BP	trans-synaptic signaling by BDNF, modulating synaptic transmission	3	1	0.004
GO:0044090	BP	positive regulation of vacuole organization	2	1	0.005
GO:0030070	BP	insulin processing	2	1	0.005
GO:0043121	MF	neurotrophin binding	2	1	0.005
GO:1990667	CC	PCSK9-AnxA2 complex	2	1	0.006
GO:0045322	MF	unmethylated CpG binding	3	1	0.006
GO:0032804	BP	negative regulation of low-density lipoprotein particle receptor catabolic process	2	1	0.006
GO:0071593	BP	lymphocyte aggregation	2	1	0.006
GO:0016671	MF	oxidoreductase activity, acting on a sulfur group of donors, disulfide as acceptor	3	1	0.006
GO:1905581	BP	positive regulation of low-density lipoprotein particle clearance	2	1	0.006

GO:1905602	BP	positive regulation of receptor-mediated endocytosis involved in cholesterol transport	2	1	0.006
GO:1901895	BP	negative regulation of calcium-transporting ATPase activity	3	1	0.007
GO:0042997	BP	negative regulation of Golgi to plasma membrane protein transport	3	1	0.007
GO:0060263	BP	regulation of respiratory burst	4	1	0.007
GO:0045454	BP	cell redox homeostasis	69	2	0.007
GO:0045359	BP	positive regulation of interferon-beta biosynthetic process	7	1	0.007
GO:0032741	BP	positive regulation of interleukin-18 production	5	1	0.008
GO:0032640	BP	tumor necrosis factor production	4	1	0.008
GO:0034123	BP	positive regulation of toll-like receptor signaling pathway	4	1	0.008
GO:0048403	MF	brain-derived neurotrophic factor binding	3	1	0.008
GO:0098772	MF	molecular function regulator	3	1	0.009
GO:0050871	BP	positive regulation of B cell activation	5	1	0.009
GO:0002639	BP	positive regulation of immunoglobulin production	6	1	0.010
GO:0031340	BP	positive regulation of vesicle fusion	6	1	0.011
GO:0010310	BP	regulation of hydrogen peroxide metabolic process	6	1	0.011
GO:0031547	BP	brain-derived neurotrophic factor receptor signaling pathway	5	1	0.012

GO:0008330	MF	protein tyrosine/threonine phosphatase activity	4	1	0.012
GO:0002237	BP	response to molecule of bacterial origin	8	1	0.012
GO:0055069	BP	zinc ion homeostasis	5	1	0.012
GO:0097066	BP	response to thyroid hormone	6	1	0.012
GO:0032717	BP	negative regulation of interleukin-8 production	9	1	0.013
GO:0008142	MF	oxysterol binding	5	1	0.013
GO:0042490	BP	mechanoreceptor differentiation	5	1	0.015
GO:0050707	BP	regulation of cytokine secretion	6	1	0.015
GO:0035095	BP	behavioral response to nicotine	7	1	0.015
GO:0035197	MF	siRNA binding	9	1	0.015
GO:0046548	BP	retinal rod cell development	8	1	0.015
GO:0004725	MF	protein tyrosine phosphatase activity	83	2	0.015
GO:0032460	BP	negative regulation of protein oligomerization	7	1	0.016
GO:0005149	MF	interleukin-1 receptor binding	9	1	0.016
GO:0042129	BP	regulation of T cell proliferation	10	1	0.016
GO:0010989	BP	negative regulation of low-density lipoprotein particle clearance	16	1	0.016
GO:0022417	BP	protein maturation by protein folding	9	1	0.017
GO:0035749	CC	myelin sheath adaxonal region	6	1	0.017
GO:0043304	BP	regulation of mast cell degranulation	8	1	0.017
GO:0036035	BP	osteoclast development	5	1	0.017
GO:0046486	BP	glycerolipid metabolic process	5	1	0.017

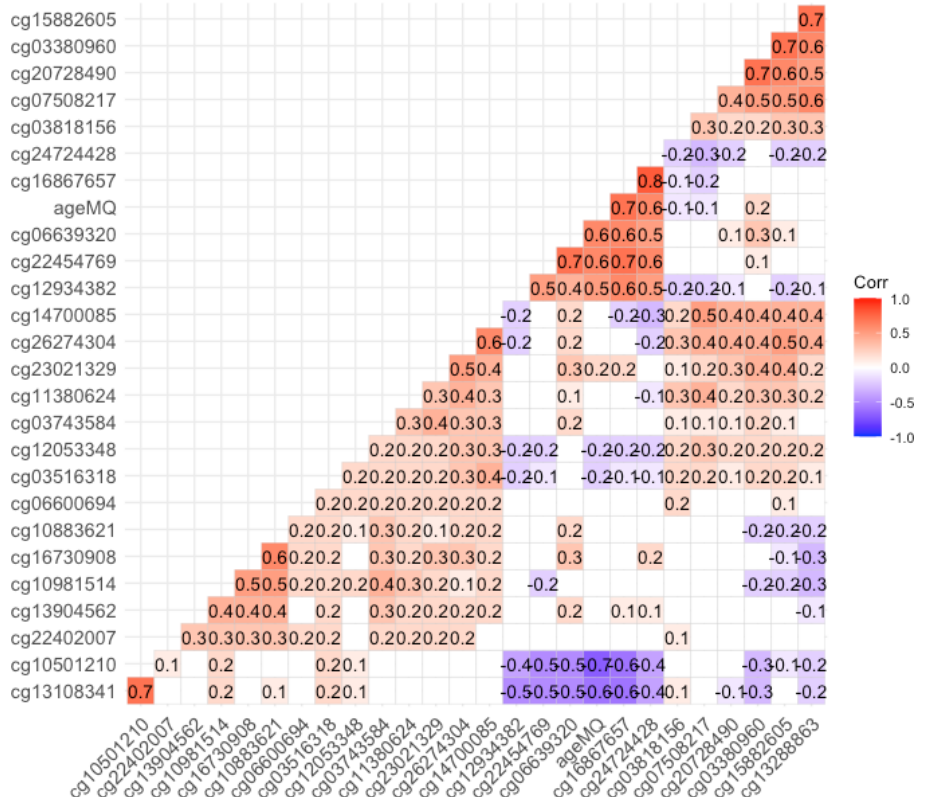
GO:0045078	BP	positive regulation of interferon-gamma biosynthetic process	12	1	0.017
GO:0007589	BP	body fluid secretion	9	1	0.018
GO:0001765	BP	membrane raft assembly	6	1	0.018
GO:0004713	MF	protein tyrosine kinase activity	78	2	0.018
GO:0070286	BP	axonemal dynein complex assembly	9	1	0.018
GO:0032725	BP	positive regulation of granulocyte macrophage colony-stimulating factor production	11	1	0.018
GO:0019887	MF	protein kinase regulator activity	7	1	0.019
GO:0015464	MF	acetylcholine receptor activity	10	1	0.019
GO:0005975	BP	carbohydrate metabolic process	118	2	0.019
GO:0097371	MF	MDM2/MDM4 family protein binding	8	1	0.019
GO:0022849	MF	glutamate-gated calcium ion channel activity	5	1	0.020
GO:1903997	BP	positive regulation of non-membrane spanning protein tyrosine kinase activity	7	1	0.020
GO:0006900	BP	vesicle budding from membrane	9	1	0.020
GO:0032009	CC	early phagosome	11	1	0.021
GO:0005892	CC	acetylcholine-gated channel complex	11	1	0.021
GO:0019227	BP	neuronal action potential propagation	9	1	0.021
GO:0043410	BP	positive regulation of MAPK cascade	95	2	0.022
GO:0007252	BP	I-kappaB phosphorylation	12	1	0.022
GO:0035335	BP	peptidyl-tyrosine dephosphorylation	99	2	0.022
GO:0000791	CC	Euchromatin	8	1	0.022

GO:0044354	CC	Macropinosome	8	1	0.022
GO:0005671	CC	Ada2/Gcn5/Ada3 transcription activator complex	15	1	0.023
GO:0004970	MF	ionotropic glutamate receptor activity	7	1	0.024
GO:0005858	CC	axonemal dynein complex	10	1	0.025
GO:0043220	CC	Schmidt-Lanterman incisure	11	1	0.025
GO:0016558	BP	protein import into peroxisome matrix	10	1	0.025
GO:0034162	BP	toll-like receptor 9 signaling pathway	16	1	0.025
GO:0045577	BP	regulation of B cell differentiation	11	1	0.025
GO:0004972	MF	NMDA glutamate receptor activity	7	1	0.025
GO:0006893	BP	Golgi to plasma membrane transport	13	1	0.026
GO:0045647	BP	negative regulation of erythrocyte differentiation	10	1	0.026
GO:0016580	CC	Sin3 complex	12	1	0.026
GO:0060285	BP	cilium-dependent cell motility	11	1	0.026
GO:0036020	CC	endolysosome membrane	13	1	0.027
GO:0045779	BP	negative regulation of bone resorption	12	1	0.027
GO:0021954	BP	central nervous system neuron development	13	1	0.027
GO:0097553	BP	calcium ion transmembrane import into cytosol	7	1	0.028
GO:0015276	MF	ligand-gated ion channel activity	13	1	0.028
GO:0051770	BP	positive regulation of nitric-oxide synthase biosynthetic process	14	1	0.028
GO:0030277	BP	maintenance of gastrointestinal epithelium	11	1	0.028

GO:0022848	MF	acetylcholine-gated cation-selective channel activity	16	1	0.028
GO:0005779	CC	integral component of peroxisomal membrane	16	1	0.029
GO:0043968	BP	histone H2A acetylation	13	1	0.030
GO:0098976	BP	excitatory chemical synaptic transmission	8	1	0.030
GO:0050765	BP	negative regulation of phagocytosis	16	1	0.030
GO:0033673	BP	negative regulation of kinase activity	13	1	0.031
GO:0048935	BP	peripheral nervous system neuron development	10	1	0.031
GO:0032722	BP	positive regulation of chemokine production	19	1	0.031
GO:0004143	MF	diacylglycerol kinase activity	11	1	0.031
GO:0035267	CC	NuA4 histone acetyltransferase complex	14	1	0.031
GO:0098691	CC	dopaminergic synapse	14	1	0.033
GO:0005548	MF	phospholipid transporter activity	14	1	0.033
GO:0034122	BP	negative regulation of toll-like receptor signaling pathway	17	1	0.033
GO:0006470	BP	protein dephosphorylation	127	2	0.033
GO:0032733	BP	positive regulation of interleukin-10 production	25	1	0.035
GO:0031982	CC	Vesicle	130	2	0.035
GO:0036150	BP	phosphatidylserine acyl-chain remodeling	21	1	0.035
GO:1902041	BP	regulation of extrinsic apoptotic signaling pathway via death domain receptors	17	1	0.035
GO:0001964	BP	startle response	13	1	0.035

GO:0003887	MF	DNA-directed DNA polymerase activity	21	1	0.035
GO:0046834	BP	lipid phosphorylation	12	1	0.035
GO:0007031	BP	peroxisome organization	19	1	0.036
GO:0034236	MF	protein kinase A catalytic subunit binding	13	1	0.036
GO:0030659	CC	cytoplasmic vesicle membrane	133	2	0.036
GO:0033198	BP	response to ATP	15	1	0.037
GO:0044548	MF	S100 protein binding	14	1	0.038
GO:0010243	BP	response to organonitrogen compound	16	1	0.038
GO:0043235	CC	receptor complex	127	2	0.039
GO:0090023	BP	positive regulation of neutrophil chemotaxis	24	1	0.039
GO:0017146	CC	NMDA selective glutamate receptor complex	11	1	0.039
GO:0045859	BP	regulation of protein kinase activity	20	1	0.039
GO:0001921	BP	positive regulation of receptor recycling	14	1	0.039
GO:0051019	MF	mitogen-activated protein kinase binding	18	1	0.040
GO:0007631	BP	feeding behavior	23	1	0.040
GO:0098641	MF	cadherin binding involved in cell-cell adhesion	17	1	0.042
GO:0018108	BP	peptidyl-tyrosine phosphorylation	120	2	0.042
GO:2000811	BP	negative regulation of anoikis	17	1	0.042
GO:0071375	BP	cellular response to peptide hormone stimulus	16	1	0.043
GO:0006259	BP	DNA metabolic process	23	1	0.043
GO:0010592	BP	positive regulation of lamellipodium assembly	20	1	0.043

GO:0019897	CC	extrinsic component of plasma membrane	21	1	0.044
GO:0032735	BP	positive regulation of interleukin-12 production	25	1	0.044
GO:0010039	BP	response to iron ion	18	1	0.044
GO:0060044	BP	negative regulation of cardiac muscle cell proliferation	21	1	0.044
GO:0003951	MF	NAD+ kinase activity	17	1	0.045
GO:0002755	BP	MyD88-dependent toll-like receptor signaling pathway	32	1	0.046
GO:0022011	BP	myelination in peripheral nervous system	17	1	0.047
GO:0070064	MF	proline-rich region binding	18	1	0.047
GO:0016575	BP	histone deacetylation	22	1	0.047
GO:0006298	BP	mismatch repair	22	1	0.047
GO:0030301	BP	cholesterol transport	21	1	0.048
GO:0035902	BP	response to immobilization stress	23	1	0.049
GO:0032728	BP	positive regulation of interferon-beta production	26	1	0.050
GO:0045453	BP	bone resorption	22	1	0.050
GO:0048011	BP	neurotrophin TRK receptor signaling pathway	20	1	0.050



Supplementary Figure 1: Correlations of dmCpGs related to fathers' smoking onset before age 15 years with aging marker CpGs from the RHINESSA Cohort (cg1686765: ELOVL2, cg24724428: ELOVL2, cg22454769: FHL2 and cg131083: DNAH9) and Offspring age. 9 dmCpGs showed correlation = 0 with age; aging methylation markers correlation >=|0.6|.

Supplementary Table E10: Associations between dmCpGs identified in relation to fathers' smoking onset before age 15years and phenotypic outcomes in the offspring (adjusted for offspring's sex).

Offspring outcomes	CpG site	Coefficient	P value	Gene
Ever asthma	cg22402007	10.7900	0.014	NTRK2
Ever wheezing	cg11380624	14.4008	0.000	DNAJC14
Ever wheezing	cg10981514	-8.3351	0.012	TPCN1
Weight	cg12053348	-51.0810	0.001	Intergenic
Weight	cg03380960	42.5750	0.023	FAM53B
Weight	cg22402007	-67.1670	0.013	NTRK2
BMI	cg23021329	0.0024	0.067	TLR9
BMI	cg12053348	-0.0013	0.004	Intergenic
BMI	cg03380960	0.0022	0.008	FAM53B
BMI	cg22402007	-0.0017	0.025	NTRK2
BMI	cg11380624	-0.0013	0.074	DNAJC14

Supplementary Table E11: Power calculations using R package pwrEWAS[28]

Samples	Target Delta			
	0.05	0.10	0.15	0.20
A: preconception father smoking: effect size=0.024, sample split=0.62, target dmCpGs=4, FDR=0.05, tissue type=Blood adult				
500	0.79	0.88	0.92	0.95
600	0.85	0.92	0.95	0.97
700	0.89	0.95	0.97	0.98
800	0.93	0.97	0.98	0.99
900	0.95	0.98	0.99	0.99
B: Start smoking before age <15; effect size=0.02, sample split=0.21, target dmCpGs=19, FDR=0.05, tissue type=Blood adult				
200	0.64	0.76	0.83	0.88
250	0.72	0.83	0.88	0.92
300	0.79	0.88	0.92	0.95
350	0.84	0.91	0.95	0.96
400	0.88	0.94	0.96	0.98

References

1. Miller, S. A., Dykes, D. D. & Polesky, H. F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16**, 1215 (1988).
2. R Core Team. R: A language and environment for statistical computing. *R Found. Stat. Comput. Vienna, Austria* https://www.R-project.org/ (2020).
3. Aryee, M. J. *et al*. Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).
4. Min, J. L., Hemani, G., Davey Smith, G., Relton, C. & Suderman, M. Meffil: efficient normalization and analysis of very large DNA methylation datasets. *Bioinformatics* **44**, 1–7 (2018).
5. Teschendorff, A. E. *et al.* A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* **29**, 189–196 (2013).
6. Tian, Y. *et al.* ChAMP: Updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics* **33**, 3982–3984 (2017).
7. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
8. Pidsley, R. *et al.* Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol.* **17**, (2016).
9. Zheng, S. C. *et al.* EpiDISH web server: Epigenetic dissection of intra-sample-heterogeneity with online GUI. *Bioinformatics* **36**, 1950–1951 (2020).
10. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
11. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).
12. Stephen D. Turner. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *TbioRxiv* **81**, 559–575 (2007).
13. LiLin-Yin. Package ‘ CMplot ’. 1–8 (2020).
14. van Iterson, M., van Zwet, E. W. & Heijmans, B. T. Controlling bias and inflation in epigenome- and transcriptome-wide association

studies using the empirical null distribution. *Genome Biol.* **18**, (2017).
15. Suderman, M. *et al.* Dmrff: Identifying differentially methylated regions efficiently with power and control. *bioRxiv* 1–26 (2018) doi:10.1101/508556.
16. Breeze, C. E. *et al.* eFORGE v2.0: updated analysis of cell type-specific signal in epigenomic data. *Bioinformatics* **35**, 1–3 (2019).
17. Carvalho-Silva, D. *et al.* Open Targets Platform: New developments and updates two years on. *Nucleic Acids Res.* **47**, D1056–D1065 (2019).
18. Li, M. *et al.* EWAS Atlas: A curated knowledgebase of epigenome-wide association studies. *Nucleic Acids Res.* **47**, D983–D988 (2019).
19. Szklarczyk, D. *et al.* STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
20. Soudy, M. *et al.* UniprotR: Retrieving and visualizing protein sequence and functional information from Universal Protein Resource (UniProt knowledgebase). *J. Proteomics* **213**, 103613 (2020).
21. Maksimovic, J., Oshlack, A. & Phipson, B. Gene set enrichment analysis for genome-wide DNA methylation data. *bioRxiv* 2020.08.24.265702 (2020).
22. Christiansen, C. *et al.* Novel DNA methylation signatures of tobacco smoking with trans-ethnic effects. *Clin. Epigenetics* **13**, 1–13 (2021).
23. Joehanes, R. *et al.* Epigenetic Signatures of Cigarette Smoking. *Circ. Cardiovasc. Genet.* **9**, 436–447 (2016).
24. Joubert, B. R. *et al.* DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. *Am. J. Hum. Genet.* **98**, 680–696 (2016).
25. Golding, G., Pembrey, P. & Jones, J. ALSPAC - The Avon Longitudinal Study of Parents and Children I. Study methodology. *Paediatr. Perinat. Epidemiol.* **15**, 74–87 (2001).
26. Boyd, A. *et al.* Cohort profile: The 'Children of the 90s'-The index offspring of the avon longitudinal study of parents and children. *Int. J. Epidemiol.* **42**, 111–127 (2013).
27. Fraser, A. *et al.* Cohort profile: The avon longitudinal study of parents and children: ALSPAC mothers cohort. *Int. J. Epidemiol.* **42**, 97–110 (2013).
28. Graw, S., Henn, R., Thompson, J. A. & Koestler, D. C. PwrEWAS: A user-friendly tool for comprehensive power estimation for epigenome wide association studies (EWAS). *BMC Bioinformatics* **20**, 1–11 (2019).

Graphic design: Communication Division, UiB / Print: Skipnes Kommunikasjon AS

uib.no

ISBN: 9788230862933 (print)
9788230861677 (PDF)