

Review

# From Genotype to Phenotype: Through Chromatin

Julia Romanowska <sup>1,2</sup> and Anagha Joshi <sup>2,\*</sup>

<sup>1</sup> Department of Global Public Health and Primary Care, University of Bergen, 5018, Bergen, Norway; Julia.Romanowska@uib.no

<sup>2</sup> Computational Biology Unit, Department of Clinical Science, University of Bergen, 5021, Bergen, Norway

\* Correspondence: Anagha.Joshi@uib.no; Tel.: +47-55-58-54-35

Received: 19 December 2018; Accepted: 21 January 2019; Published: 23 January 2019

**Abstract:** Advances in sequencing technologies have enabled the exploration of the genetic basis for several clinical disorders by allowing identification of causal mutations in rare genetic diseases. Sequencing technology has also facilitated genome-wide association studies to gather single nucleotide polymorphisms in common diseases including cancer and diabetes. Sequencing has therefore become common in the clinic for both prognostics and diagnostics. The success in follow-up steps, i.e., mapping mutations to causal genes and therapeutic targets to further the development of novel therapies, has nevertheless been very limited. This is because most mutations associated with diseases lie in inter-genic regions including the so-called regulatory genome. Additionally, no genetic causes are apparent for many diseases including neurodegenerative disorders. A complementary approach is therefore gaining interest, namely to focus on *epigenetic* control of the disease to generate more complete functional genomic maps. To this end, several recent studies have generated large-scale epigenetic datasets in a disease context to form a link between genotype and phenotype. We focus DNA methylation and important histone marks, where recent advances have been made thanks to technology improvements, cost effectiveness, and large meta-scale epigenome consortia efforts. We summarize recent studies unravelling the mechanistic understanding of epigenetic processes in disease development and progression. Moreover, we show how methodology advancements enable causal relationships to be established, and we pinpoint the most important issues to be addressed by future research.

**Keywords:** epigenetics; chromatin modification; sequencing; regulatory genomics; disease variants

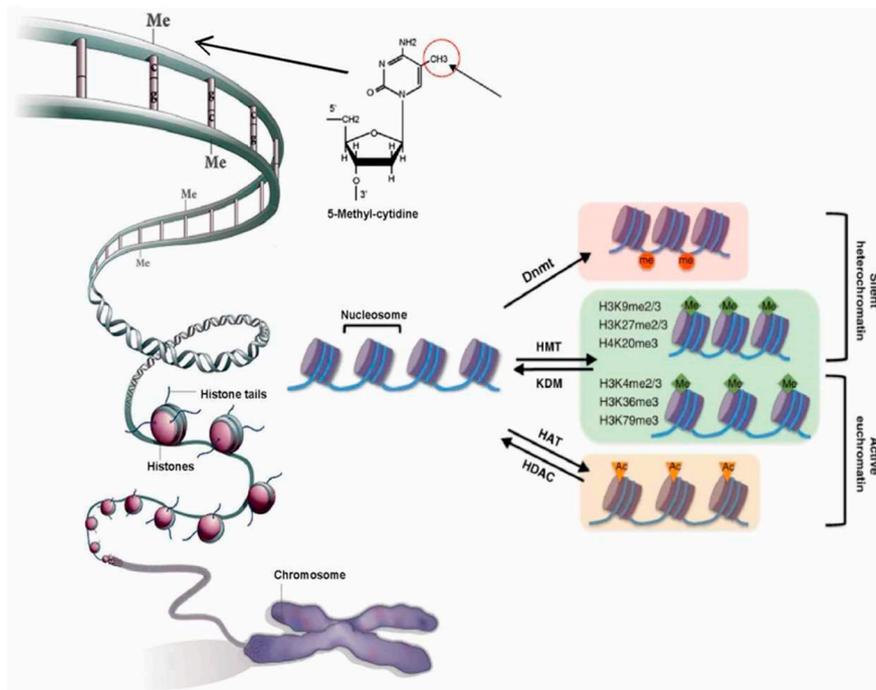
---

## 1. Introduction

### 1.1. Definition of Epigenetics

The human body consists of hundreds of different tissues and cell types, each with its characteristic well-defined function. For example, myosin is produced by muscle cells while hemoglobin is produced by red blood cells to facilitate cell type specific functions. Despite the diversity of functional molecules in an individual cell type, nearly all cell types in an organism contain the same genetic information or genome. To explain how this diversity of cell types can be achieved from a single cell or zygote, Conrad Waddington proposed the concept of “*epigenesis*” in 1956, where pluripotent cells have the “potential” to generate all other cell types of restricted potential, in which they gradually lose this “potential” during differentiation, famously depicted by the Waddington landscape [1]. This so-called potential was later associated with a physical phenomenon, the methylation of DNA [2], which is a methyl group added to position 5 on the cytosine ring. In mammals, it is mainly 5′–C–phosphate–G–3′ dinucleotide (CpG) that is subjected to methylation. Originally, methylation was found to act as a silencing mark. Accordingly, in embryonic stem cells, the majority of promoters have un-methylated DNA, and some of them

become methylated during differentiation, assisting the acquisition of their final cell identity [3]. Over the years, many other epigenetic and transcription control mechanisms responsible for establishing unique gene expression profiles characteristic for different cell and tissue types during embryonic development have been studied in detail [4,5]. Gene regulatory elements receive and execute transcriptional signals, dependent on their epigenetic state and chromatin accessibility, controlling the expression of key developmental factors [6]. Chromatin dynamics are regulated through two main mechanisms: methylation of DNA and post-translational modifications of histone tails [7] (Figure 1). Histone modifications include, among others, phosphorylation, acetylation, methylation, and ubiquitylation, with methylation at specific residues as one of the most important posttranslational modifications regulating nuclear function, including transcriptional regulation, epigenetic inheritance, and maintenance of genome integrity [8]. Recently, it has become evident that histone modifications act together and a term “histone code” was coined to refer to a scheme of gene control exhibited by the complex interactions of histone modifications [9,10]. Accordingly, specific functions can be associated to a group of histone modifications, such as H3K27ac and H3K4me1, and are associated with enhancer regions. Several reviews written over the years focus on state-of-the-art studies providing structure function associations of histone modifications and successive layers of chromatin structure in mammalian genomes [11–13].



**Figure 1.** Diagrammatic representation of epigenetic mechanisms namely DNA methylation and chromatin modifications [14].

## 1.2. Broadening the Definition of Epigenetics

Epigenetics are widely understood as any mechanism by which heritable changes in gene expression occur without changing the DNA sequence, but the precise definition has evolved over the years. Apart from the above mechanisms, the role of non-coding RNAs (ncRNAs) is becoming evident in epigenetic control (reviewed in [15]). In short, ncRNAs are transcribed from the genome sequence without producing a functional protein, are highly cell type specific and regulate epigenetic patterning by establishing epigenetic modifications (DNA methylation and chromatin modifications). For example, *Xist* is an ncRNA expressed from the X chromosome that silences the other X chromosome in females. Non-coding RNAs can function as a guide or tethers, and may be the molecules of choice for epigenetic regulation of DNA methylation [16]. Some authors therefore

now include ncRNAs in their definition of epigenetics. Nevertheless, we will stick to the classical definition and discuss only DNA methylation and chromatin modifications in this review.

### *1.3. Epigenetic Mechanisms Regulate Gene Expression Using Environmental Cues*

Epigenetic mechanisms are thought to act as a memory of a cell and might be the key process by which the environment interacts with the genome [17]. DNA methylation plays a crucial role during early development including active demethylation of paternal genome before the first cleavage and subsequent demethylation of maternal genome [18]. Furthermore, environmental factors also affect gene expression via epigenetic mechanisms during embryonic development, which can manifest into adulthood or even old age. Cigarette smoking is an environmental factor, associated with dose- and time-dependent changes in the DNA methylation signature, which manifests in gene and protein expression leading to an increased vulnerability to other forms of complex illnesses [19,20]. Harmful environmental factors need not be substances. Trauma and stress also influence gene expression through epigenetic mechanisms, and furthermore these epigenetic modifications can be passed over the generations [21].

## **2. Chromatin Modifications and the Genome Organization**

### *2.1. Chromatin's Structure Defines Its Function*

To understand epigenetic control mechanisms, we will begin with the structure of chromatin. DNA is wrapped around the core histone proteins, forming a structure named nucleosome (two copies of H2A, H2B, H3, H4, and 147 base pairs (bps) of DNA around them). This is further compacted, with the assistance of assembly and packaging related proteins, to form a higher-order chromatin structure [22], with two distinct chromatin states “euchromatin” and “heterochromatin” (Figure 1). A more open chromatin environment, euchromatin, is where the majority of active genes localize, while heterochromatin is characterized by a more compact environment where inactive genes, non-coding DNA and repeat elements reside [8]. Heterochromatin can be further separated into two groups, facultative and constitutive. Facultative heterochromatin includes regions that consist of genes that are highly differentially expressed during development. Constitutive heterochromatin on the other hand is gene poor, rich in repeat elements, mainly found in centromeres and telomeres, and silenced indefinitely [23]. These chromatin states are marked by distinct epigenetic factors [24] (Figure 1), and in euchromatin, the histone modification density correlates with the density of TF binding sites [25]. However, neither euchromatin nor heterochromatin is marked uniformly with epigenetic and transcriptional signals. Chromatin is further organized into so-called topologically associated domains (TADs), (first described by Dixon et al. (2012) [26]), regions spanning several hundred kilobases. Topologically associated domains are organized hierarchically and are highly enriched for insulating factor CCCTC-binding factor (*CTCF*) binding and histone marks at the boundaries [27]. Intra-chromosomal interactions are particularly enriched within TADs and accordingly genes within a TAD show highly correlated gene expression. The chromatin structure allows manifestation of genetic information in a cellular context, and mutations in chromatin organization genes lead to developmental pathologies [28,29]. Understanding of cell type specific 3D genome organization is therefore highly valuable in a disease context [30,31], where by disruption of TADs can result in chromatin interaction changes leading to mis-regulation of oncogenic or tumor suppressor genes [32].

### *2.2. Chromatin Structure is Dynamic and Marked by Histone Modifications*

The chromatin structure is organized with the help of DNA sequence and epigenetic modifications, including histone modifications, and a cross-talk between them is potentially facilitated through histone amino (N)-terminal tails interacting with neighboring nucleosomes [33]. Of various histone modifications, the most well-studied types are methylation, acetylation, phosphorylation, and ubiquitination [8]. Histone modifications influence chromatin mainly in two ways. The first mode of the modifications affects directly the structure of the chromatin over a long

or short distance by the recruitment of DNA binding proteins and chromatin remodelers affecting nucleosome location. Hence, nucleosome removal could open the chromatin and a possible transcription factor binding motif could be revealed, or otherwise, newly recruited nucleosomes could conceal a binding motif, hindering transcriptional machinery recruitment at the locus [34]. The second mode of histone modification is carried out by three sets of enzymes named “writer”, “reader”, and “eraser”, based on the function of each enzyme related to each histone modification. For example, COMPASS family members maintain H3K4me3 modification, while polycomb family members maintain H3K27me3 modification. Both activating (H3K4me3) and repressing (H3K27me3) modifications are indeed present simultaneously at promoters enriched for developmental genes and have a distinct sequence signature [35]. Histone modifications also work jointly with DNA methylation, for repression of gene loci [36].

### 3. Epigenetics in Disease Context

#### 3.1. Genome-Wide Studies Are Not Enough

Monogenic diseases are caused by the malfunctioning of only a single gene. For example, fragile x syndrome is caused by epigenetic changes in the *FMR1* gene. The silenced promoter of *FMR1* in disease shows heterochromatin markers, including DNA hypermethylation and histone deacetylation. This can be treated by pharmacological reactivation of gene transcription, particularly through the use of DNA demethylating agents or inhibitors of histone deacetylases [37]. Unfortunately, the vast majority of common diseases are not caused by mutations in a single gene, but rather by a large number of single nucleotide variations (SNPs) spread throughout the genome. These diseases are therefore called complex diseases. Complex diseases including cancer, diabetes, and neurodegenerative disorders such as Alzheimer’s and Parkinson’s disease are common and therefore form a global health burden. Though a large number of genetic variants have been identified (and will be identified) that increase the risk for these diseases, most explain only a small fraction of risk. Moreover, despite the fact that over 1000 genetic loci are associated with susceptibility to common diseases in human [38], only a handful of these loci have resulted in the identification of causal genes or pathways for potential therapeutic applications [39]. It is becoming clear that understanding of only genetic variation will not be sufficient to get a complete understanding of disease, and the role of epigenetic alterations in gene regulation is becoming evident in many diseases, including cancer. Understanding how a genotype influences human health and disease now requires characterization of the epigenome as well. For example, copy number aberrations of genes responsible for writing, reading, and removing H3K9 methylation were identified in medulloblastoma, demonstrating that defective control of the histone code contributes to the pathogenesis of medulloblastoma [40]. Large studies have therefore been designed to unravel epigenetic malfunctionalities in diverse diseases (Table 1). It is important to note that another major challenge in interpreting genome-wide data in a clinical context is the fact that the vast majority of genetic and epigenetic modifications lie in non-coding genomic regions, particularly [41] where the disease-associated variants in enhancers explain a greater proportion of the disease heritability [42].

**Table 1.** A collection of epigenetic studies (excluding DNA methylation) in disease context including the data type, number of samples, disease type, and publication reference.

Num.	Data Type	Disease	Available data	# of Samples	Reference
1	ATAC-seq	23 cancer types	Genotype, ATAC-seq, RNA-seq	410	[43]
2	ChIP-seq	Prostate cancer	H3K27ac, H3K4me3, H3K27me3	100	GSE120738
3	ChIP-seq	Breast cancer	H3K4me1, TFs	-	[44]
4	ChIP-seq	Adenocarcinoma	H3K27ac, H3K4me3, H3K4me1	94	[45]
5	ChIP-seq	Acute myeloid leukemia	H3K9me3	108	[46]
6	ChIP-seq	Glioma	Multiple	-	[47]
7	ChIP-on-chip	Acute myeloid leukemia	H3	73	[48]
8	ChIP-on-chip	Acute promyelocytic leukemia	H3, H3K9me3, H3K4me3	372	[49]

9	ChIP-seq	Acute myeloid leukemia	H3K9me2	16	[50]
10	ChIP-seq	Hepatocarcinoma	Multiple	5	[51]
11	ATAC-seq, ChIP-seq	Colorectal cancer	Multiple	4	[52]
12	FAIRE-seq, ChIP-seq	Ovarian cancer	H3K27ac, H3K4me1	5	[53]

### 3.1. Largescale Epigenetic Studies in Cancer

#### 3.1.1. Epigenetic Mechanisms Are Major Drivers in Cancer

The studies exploring mutational landscapes of cancer have highlighted frequent mutations in genes encoding chromatin-associated proteins. The exploration of functional mechanisms behind these mutations have improved our understanding of oncogenic mechanisms at different levels of chromatin organization and regulation (reviewed in Valencia et al. (2019) [54]). DNA methylation remains by far the most studied epigenetic mechanism in cancer where inactivation of tumor-suppressor genes occurs as a consequence of hypermethylation of the gene promoters. Numerous studies have identified a broad range of genes silenced by DNA methylation in different cancer types [55]. Importantly, different cancer subtypes show characteristic DNA methylation signatures [56], which can be translated in clinical medicine by using hypermethylated promoters as biomarkers. Human pluripotent stem cells were found to have more hypermethylated DNA than fibroblast cells [57]. Similarly, oncogenesis is thought to modify the cell state into a stem or progenitor epigenetic state. In cancer, mutations in key transcription factors lead to changes in DNA methylation, such that the number of genes with gene expression changes explained by DNA methylation are 10-fold higher than those explained by genetic mutations. Over 75% of DNA hypermethylated genes are marked by polycomb repressor components forming bivalent chromatin [58]. Wang et al. [59] pointed to one molecular mechanism to explain the role of *MLL3* mutations in cancer pathogenesis by examining changes in histone modification and gene expression after depletion of Polycomb or COMPASS family members. Next, they proposed a potential therapeutic strategy for cancers harboring COMPASS mutations which will allow resetting the epigenetically (Polycomb/COMPASS) balanced state of gene expression.

#### 3.1.2. Epigenetic Mechanisms in Hematopoietic Malignancies and Their Therapeutic Implications

Epigenetic changes in cancer are possibly reversible making them precious targets for cancer therapy. Indeed, DNA methylation biomarkers with diagnostic, prognostic, and predictive power are already in clinical trials or in a clinical setting [60]. DNA methyltransferase inhibitors have been approved for the treatment of several hematopoietic malignancies, including myelodysplastic syndromes, chronic myelomonocytic leukemia, and acute myelogenous leukemia (AML) [61]. Other epigenetic regulatory mechanisms also play a critical role in the pathogenesis of AML. Epigenome-wide analyses of histone H3 acetylation identified that epigenetic silencing of *PRDX2*, a growth suppressor, contributed to the malignant phenotype in AML [48]. A combination of the H3K9me3 signature with established clinical prognostic markers outperformed prognosis prediction based on clinical parameters alone in AML [46]. Epigenetic control is systematically studied in other hematopoietic malignancies as well. For example, the translocation t(15;17) forming a chimeric PML-RAR $\alpha$  transcription factor is the initiating event of acute promyelocytic leukemia. PML-RAR $\alpha$  regulates key cancer related genes and pathways by inducing a repressed chromatin at its target genes [49]. The PML-RAR $\alpha$  binding universally led to histone deacetylase (HDAC) recruitment, loss of histone H3 acetylation, and increased H3K9me3. Accordingly, several anticancer drugs acting as inhibitors of HDAC or bromodomain and extra-terminal proteins (*BET*) were designed, tested, and in clinical trial. The use of these inhibitors is not limited to hematopoietic malignancies. The HDAC inhibitors have been used in glioblastomas, where mutations in tumor suppressors such as *IDH1* induce epigenetic changes that drive the development of gliomas [47]. Both HDAC and BET inhibitors work synergistically, primarily by suppressing super-enhancers, the regulatory regions driving

cancer phenotype through epigenetic reprogramming. Indeed, adenocarcinoma super-enhancers classified according to their somatic alteration status display distinct epigenetic, transcriptional and pathway enrichments and are enriched in genetic risk SNPs associated with cancer predisposition [45].

### 3.1.3. Epigenetic Targets for Cancer Therapy

Unfortunately, the current cancer drugs targeting epigenetic mechanisms are unspecific and can often have serious side effects. Understanding other epigenetic changes in cancer is therefore highly urgent to open up avenues for new therapies. The pharmaceutical industry is therefore focused on identifying new compounds that target the reader, writer, and eraser mechanisms of histone modifications. To this end, functional genomics studies in disease are gaining pace. A recent large study generated ATAC-seq data, a proxy for mapping genome-wide open chromatin, in over 400 tumors across 23 cancer types from The Cancer Genome Atlas project [43]. The authors further identified enhancer–promoter interactions in different cancer types by integrating it with RNA-seq data and validated some of their predictions through CRISPR-Cas9 assays [43].

### 3.2. Largescale Epigenetic Studies in Other Diseases

The potential of epigenetic therapies for cancer treatment has influenced an increase in studies investigating epigenetic control across a wide range of other diseases. Such efforts have generated knowledge about the combinatorial effects of genetic mutations and epigenetics on the phenotype. For example, the interaction of genetic variants and DNA methylation of the interleukin-4 receptor gene increases the risk of asthma [62], and a genetic/epigenetic interaction in the reduced folate carrier (RFC1) gene locus influence fetal predisposition to autism [63]. The study of epigenetic mechanisms is highly relevant to some diseases. One of the major concerns of the aging world population today are neurodegenerative disorders. There is no cure for many of the neuropathies and the majority of the cases have no genetic basis. Many compounds function via epigenetic mechanisms, and epidrugs (discussed above) developed for cancer treatment have been submitted to clinical trials for the treatment of Alzheimer's and Parkinson's diseases [64]. For example, HDAC inhibitors change the epigenetic state and expression of *FXN* in the neurodegenerative disease Friedreich ataxia, making it highly effective in an in vitro disease model and also showing promising results in a patient study [65]. In summary, understanding of epigenomic landscape of neurodegenerative and other disorders will likely provide a possibility of early detection and intervention of pre-symptomatic pathological events. This will allow development and implementation of novel strategies or treatments to halt pathological progress. It is important to stress that it is the putative reversibility of epigenetic aberrations that enables pharmacological interventions (epidrugs) as potential novel candidates for successful treatments of multifactorial disorders [64].

## 4. Computational Approaches towards Epigenetic Data Analysis and Integration

### 4.1. Epigenetic Data Integration to Understand the “Epigenetic Code”

Several studies have connected specific combinations of histone modifications and DNA methylation to the presence or absence of transcriptional activity and genomic functional elements. For instance, H3K4me3 is highly enriched at the promoters of actively transcribed genes [25], H3K36me3 is found on the gene body of genes under transcription and high levels of H3K9me3 are associated with facultative heterochromatin [23]. ChIP sequencing technology has allowed to generate a genome-wide high-resolution map of the distribution and co-localization of histone marks. Large initiatives have focused on unravelling the human epigenetic landscape. The Roadmap Epigenomics consortium has collected 111 reference human epigenomes by profiling histone modification patterns, DNA accessibility, DNA methylation, and RNA expression to define global maps of regulatory elements, regulatory modules of coordinated activity, and their likely activators and repressors [41]. They further used a method based on Hidden Markov Models (HMMs) to derive a minimal informative set of epigenetic modifications for differentiating between cell types, tissues

and development stages, as well as between healthy and diseased cells. Increasingly, epigenetic data is generated in clinical settings, for a move towards precision medicine. For example, Polak et al. [66] were able to pinpoint differences in the mutational landscape between cancers based on their cell type of origin. In their work, a random forest based approach was used to predict mutation densities using 424 predictor variables. When gene expression is available, together with DNA methylation levels and genotypes, one could construct a network of interactions between these features, as introduced by Hou et al. [67]. Such an approach is useful in prognosis of various cancers. This was also demonstrated by Zhu et al. [68], who tested a kernel machine learning method on various omics data and clinical factors to predict prognosis in 14 cancer types. They found that the prognostic power of copy number and somatic mutations was quite low compared to expression profiles. Moreover, they demonstrated that incorporating omics data to predictions based on clinical variables can improve the results, as it may account for the absence of unknown or unmeasured clinical features.

#### 4.1.1. The Function of Epigenetic Modifications Still Remains Understudied

Sekhon et al. [69] integrated five different histone modification datasets to predict gene expression levels with the use of deep neural networks. Hlady et al. [51] performed integrative analysis of multiple epigenetic modifications in hepatic cancer to identify epigenetic driver loci, and further demonstrated that two loci, *COMT* and *FMO3*, increase apoptosis and decrease cell viability in a liver-derived cancer cell line. There is an effort to integrate more and more epigenetic phenomena in such studies, but the large number of histone modifications possible at histone tails increases the combinatorial complexity of the histone code. Furthermore, histone modifications or the histone status varies during development [70]. The histone code is therefore complex and dynamic. More importantly, the causal relationship between histone modifications and transcription activity has not yet been deciphered. For example, H3K4me1 is present at regulatory elements called enhancers, and is widely used to predict enhancer elements [71]. However, whether H3K4me1 controls or simply correlates with enhancer activity and function has remained unclear. Recent studies suggest that H3K4me1 might fine-tune, rather than tightly control, enhancer activity and function [72].

### 4.2. Linking Epigenetic Mechanisms to Phenotypes: Epigenetic Epidemiology

#### 4.2.1. More Data Equals More Challenges

The success of genome-wide association studies (GWAS) in identifying genetic loci associated with common diseases have facilitated exploration of epigenetic loci associated with diseases, also known as the epigenome-wide association studies (EWAS). Much focus in the EWAS-type analysis has been on genome-wide DNA methylation studies, where a statistical framework is developed to identify statistically significant association between the methylation level of each CpG site and the trait of interest (reviewed in References [73,74]). However, as the technologies constantly improve to make data from other epigenetic markers available, more and more researchers integrate this data, together with genetic information to improve predictions and risk assessment [75–78]. The integration of data from diverse sources is generally a daunting task. This challenge can be simplified with the help of new experimental methods such as assay for transposase-accessible chromatin using sequencing (ATAC-seq) allow for extracting information about different epigenetic phenomena from a single experiment [79]. Moreover, one can use existing databases that enable visualization of publicly available datasets, sometimes also giving the possibility to overlay user's data [80].

#### 4.2.2. New Data Integration Opportunities

The most widely applied method in epigenetic epidemiology is to use a regression model to check associations between variations in the data and the trait, as in standard epidemiology. This methodology is used by various studies where principal components (PCs) [81], level of methylation [82,83], or association score from EWAS analysis [84] are used to represent the variation. In order to facilitate the interpretation of the results from such an analysis, one typically uses bioinformatics databases to search for possible biological explanations for connections between the significant

genomic regions and the trait of interest. This can be done, for example, in the Cistrome database [85] that gathers published gene regulatory data, and enables interactive visual analysis. Another easy-to-use tool is HaploReg [86]. Although the output is less intuitive, the database provides rich information about possible regulatory functions of SNPs or genomic regions of interest. Having found a set of genes that contain differential epigenetic modifications allows to perform a gene enrichment analysis, for example, with the help of LAGO (<https://go.princeton.edu/cgi-bin/LAGO>), STRING [87] or Reactome [88]. Another interesting possibility is to infer disease–gene connections by accounting for associations between different types of data, as implemented in Hetionet [89]. This tool integrates around 30 different databases, creating a heterogeneous network from information such as expression data, differential gene regulation, GWAS gene-trait associations, drug banks, etc. The implementation of the database in a neo4j network service allows for a quick online querying and visually appealing output that can inform on hidden connections between, for example, influence of vitamin intake, genes, and a disease [90].

#### 4.2.3. Epigenome-Wide Association Studies Analyses Are Informative Only about an Association and Not Causality

In classical epidemiology Mendelian randomization (MR) is widely used to infer causality whenever a standard randomized trial is impossible to perform. It is based on an assumption that the underlying genotype is randomly assigned to each individual and is the cause of the measured exposure (e.g., body mass index (BMI)), not vice versa. This method has been recently adapted to DNA methylation data [91,92]. However, since DNA methylation can be both an inducer and the outcome of the disease, MR with epigenetic data needs to be used with caution [93]. Nevertheless, the remarkably simple idea behind the MR allows researchers to make very interesting claims, studying causality between the epigenetic marks and a wide range of outcomes, from blood lipid levels [94] through features such as physical aggression [95]. Used together with EWAS and GWAS analyses, MR gives us the possibility to propose biomarker loci or targets for therapies for patients [82].

Many more methods have been developed recently to infer causality from epigenetic data. For example, Howey et al. [96] fit a Bayesian network to the most significant findings from their linear regression modeling to show the directions of influence between DNA methylation and blood lipid levels. In another study, structural equation modeling (SEM) was used to search for the pathways by which the genetic variants lead to a disease [97]. With this method, one can establish significant interactions between all the different measurements (here, blood lipid levels, variant allele in the chosen SNP, and methylation levels on the nearby CpGs) and importantly, the model predicts the directionality of these interactions.

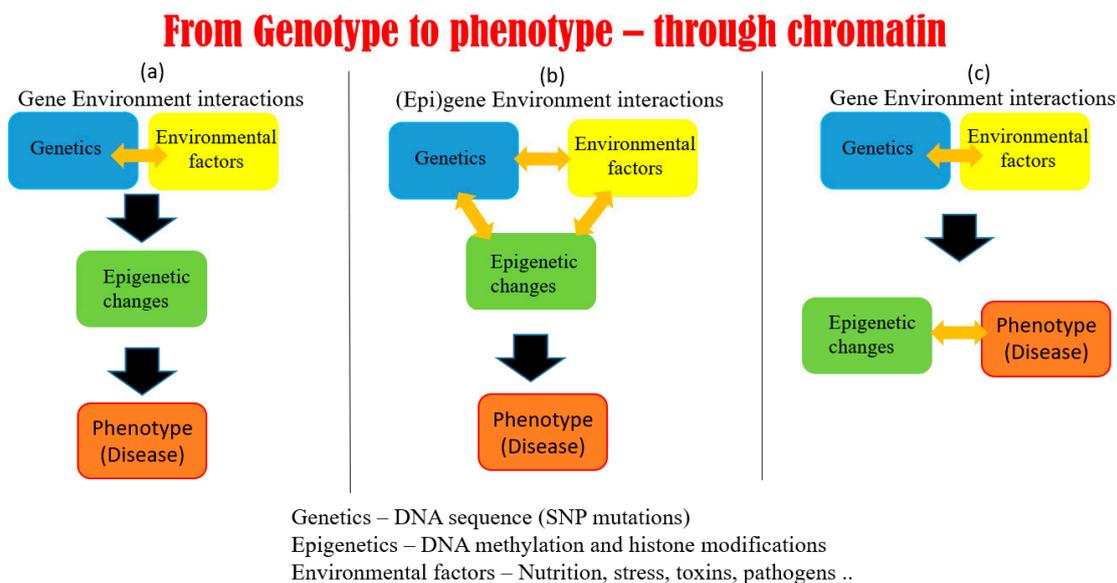
#### 4.2.4. Causality Inference from Translational Studies

To test whether a change in the levels of epigenetic modifications is the cause or consequence of a disease, one can conduct a translational study, following patients over a specific time. Such time-dependent information can then be used to check whether a certain locus displays epigenetic changes, e.g., DNA methylation, before or after a certain event; disease onset. Using this concept, a computational approach GATE [98] has been implemented as a two-layer model, where one layer categorizes the spatial characteristics of the chromatin, and the other layer focuses on transitions between different chromatin states. This allows to create a model of transitions between different epigenetic states of a cell. Another recent method, ChromTime, uses the raw signal from data generated by CHiP-seq and similar techniques to track temporal changes in the peaks [99]. It not only detects diminishing or appearing peaks, but also asymmetrical changes in peak shapes. The authors further demonstrate that ChromTime can be applied on ATAC-seq, CHiP-seq, and DNase-seq data to infer on gene expression levels and TF binding.

### 4.3. Combining Levels of Epigenetic Marks within Genomic Regions

One of the important shortcomings of many methods is that they consider each epigenetic locus independently of other loci to evaluate its significance for association with a certain trait. For example, the majority of studies focus on methylation level of one CpG at a time even when they integrate it with several other data sources. Recent studies summarized methylation level within a region [81,83], though this is not yet widely used despite the fact that changes in DNA methylation of only one CpG site would likely not lead to big changes in TF binding affinity to this site, unless it is followed by coordinated changes on neighboring CpG sites [100]. To this end, we developed a statistical framework that integrates DNA methylation and genetic information to identify statistically significant interactions between an SNP and methylation level within a group of neighboring CpGs [101]. The CpGs are grouped based on whether the CpGs are assigned to a promoter, enhancer or a gene body, to facilitate the downstream analysis for the biological interpretations.

The ultimate goal is to understand how genetic and epigenetic variations manifest in a phenotype under certain environmental conditions (Figure 2). To this end, an ideal computational approach would take into account the genotype and several epigenetic modifications at the same time, to explain a phenotype or perhaps a proxy such as transcriptomic data. There is already a huge amount of such data in the public domain, and tools and resources such as Omics Discovery Index web service (<https://www.omicsdi.org/>) to search for datasets. There is a need to take the advantage of this enormous amount of data, test ideas, and to develop tools to maximize the information extraction from the data.



**Figure 2.** The figure depicts three likely scenarios where epigenetics might fit with from the genotype to phenotype (gene expression) information flow: (a) epigenetic changes are downstream of gene environment interactions and determine the phenotype; (b) genome sequence, environment, and epigenetic modification work together to establish the phenotype; and (c) epigenetic landscape and phenotype are both determined and established by gene–environment interactions. SNP: single nucleotide variations

## 5. Conclusions

### 5.1. Possible Scenarios Linking Epigenetics, Genetics, and Phenotype

Hundreds of human cell types have a unique gene expression signature despite sharing the same genome sequence, largely due to tight control by epigenetic modifications of the non-coding genome in a cell type specific manner. Epigenetic aberrations are thought to result in complex diseases such as cancers. The vast majority of genetic variants found associated with common diseases by genome-

wide association studies are indeed located in the non-coding genome. Only in a very limited number of cases such as for Crohn's disease or rheumatoid arthritis, have these associations led to the successful identification of causal genes with a potential of being therapeutic targets. However, most disease-associated variants have no known biological context to disease, limiting their utility for prognosis or treatment. Human epidemiological studies provide evidence for prenatal and early postnatal environmental factors influencing adult risk of developing various chronic diseases, such as cancer, cardiovascular disease, diabetes, obesity, and behavioral disorders such as schizophrenia [17]. Some of these environmental factors can be linked directly to alterations of the epigenetic landscape that affect gene regulation and finally the disease. Though the association is proven in many cases, the chain of causality remains to be established. This leads to three possible scenarios of how epigenetic mechanisms control genes and influence disease occurrence (Figure 2). The first scenario is where environmental factors alter epigenetic modifications, which in turn alter the phenotype (Figure 2A). This scenario is supported by mouse experiments where maternal methyl-donor supplementation during pregnancy with folic acid, vitamin B12, choline, and betaine was shown to affect the phenotype of the Avy (viable yellow agouti) offspring by directly altering the epigenome [102]. The second possibility is that gene–environment interactions affect both epigenetic status and transcription read-out, as their correlation does not imply causality (Figure 2B). Indeed, as most of the epigenetic modifications are “lost” during the gametogenesis, this scenario is assumed to be true for many cases. Careful research has nevertheless identified that at least some epigenetic modifications are passed on to the next generation [103]. This leads to a third scenario where epigenetic modifications are not downstream of but work together with gene environment interactions to result in a phenotype (Figure 2C). The relative abundance of the three scenarios and the molecular mechanisms controlling them need to be understood. Over the coming years, research should be focused not only on identifying epigenetic phenomena affecting gene regulation to find epigenetic biomarkers for disease and environmental exposure, but also on establishing the causal relationship between the three components (gene–environment, epigenetics, and phenotype). Only by understanding causal relations can we develop new epigenetic interventions to truly revolutionize medicine to move towards preventive medicine.

### *5.2. New Approaches and Technologies Must Aim on Establishing a Causal Link between Epigenetics and Disease*

The most important challenge in precision medicine is thus to link genetic variation within the non-coding genome to candidate causal gene(s) or pathways for disease or other physiological phenotypes. It is urgent not only to identify the regulatory regions but also the spatial organization of DNA to understand how these regulatory regions interact to manifest into a phenotype. It is now accepted that a large number of possible regulatory interactions are potentially pathogenic and might be unique to tumors [43]. Although experimental techniques such as chromosome conformation capture (3C, 4C) [104] combined with next generation sequencing (Hi-C) show a great promise [105,106], their time and cost will limit the availability of comprehensive, experimentally verified 3D chromatin landscapes to a tiny fraction of the hundreds of different human cell types in the foreseeable future. The development of novel cost-effective high-throughput experimental methods is ongoing. Meanwhile, computational tools to predict enhancer–promoter interactions will be essential to model the effects of non-coding genetic variation on epigenetic modifications and downstream gene expression programs in human health and disease. Though a regulatory region is associated to its proximal promoter, the integration of known or putative enhancer promoter interactions in GWAS analysis has a potential to identify novel disease associated genes and pathways [107]. This will require a significant leap beyond studies which have only used correlations between epigenetic states of enhancers with promoter expression [108]. We have recently performed preliminary work to establish causality using regulatory information [109]. More computational approaches to systematically combine epigenetic information into causal network models are needed.

### *5.3. Epigenetic Studies and Therapies Have an Important Role in Shaping the Future of Medicine*

Finally, segregating patients based on different factors into more coherent groups for better treatment is the foundation of precision medicine, but many factors used to stratify patients have no known functional mechanisms. For example, sexual differences in cancer risk and survival are well studied, with males having an increased risk and poorer survival for most cancers [110]. The understanding of functional mechanisms behind these sex differences is gathering pace. For example, male breast cancer is rare, poorly characterized and resistant to hormonal treatment. An integrative epigenetic and transcriptomic analysis revealed a gender-selective and genomic location-specific hormone receptor action associated with survival in male breast cancer [39]. Epigenetics therefore has a big role to play in the foundations of the precision medicine.

**Author Contributions:** Conceptualization, A.J.; data curation, J.R. and A.J.; writing—original draft preparation, J.R. and A.J.; writing—review and editing, J.R. and A.J.; visualization, A.J.; funding acquisition, A.J.

**Funding:** AJ is supported by the Bergen Research Foundation Grant no. BFS2017TMT01. The APC was funded by open access fund of University of Bergen.

**Acknowledgments:** We thank Tom Michoel for useful feedback.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Nicoglou, A. Waddington's epigenetics or the pictorial meetings of development and genetics. *Hist. Philos. Life Sci.* **2018**, *40*, 61.
- Bird, A.P.; Wolffe, A.P. Methylation-induced repression—Belts, braces, and chromatin. *Cell* **1999**, *99*, 451–454.
- Mohn, F.; Weber, M.; Rebhan, M.; Roloff, T.C.; Richter, J.; Stadler, M.B.; Bibel, M.; Schübeler, D. Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol. Cell* **2008**, *30*, 755–766.
- Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **2002**, *16*, 6–21.
- Morgan, H.D.; Santos, F.; Green, K.; Dean, W.; Reik, W. Epigenetic reprogramming in mammals. *Hum. Mol. Genet.* **2005**, *14*, R47–R58.
- Reik, W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **2007**, *447*, 425–432.
- Turner, B.M. Defining an epigenetic code. *Nat. Cell Biol.* **2007**, *9*, 2–6.
- Bannister, A.J.; Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **2011**, *21*, 381–395.
- Jenuwein, T.; Allis, C.D. Translating the histone code. *Science* **2001**, *293*, 1074–1080.
- Strahl, B.D.; Allis, C.D. The language of covalent histone modifications. *Nature* **2000**, *403*, 41–45.
- Zhou, V.W.; Goren, A.; Bernstein, B.E. Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.* **2011**, *12*, 7–18.
- Zentner, G.E.; Henikoff, S. Regulation of nucleosome dynamics by histone modifications. *Nat. Struct. Mol. Biol.* **2013**, *20*, 259–266.
- Allis, C.D.; Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* **2016**, *17*, 487–500.
- Maleszewska, M.; Kaminska, B. Is Glioblastoma an Epigenetic Malignancy? *Cancers* **2013**, *5*, 1120–1139.
- Frias-Lasserre, D.; Villagra, C.A. The Importance of ncRNAs as Epigenetic Mechanisms in Phenotypic Variation and Organic Evolution. *Front. Microbiol.* **2017**, *8*, 2483.
- Lee, J.T. Lessons from X-chromosome inactivation: Long ncRNA as guides and tethers to the epigenome. *Genes Dev.* **2009**, *23*, 1831–1842.
- Jirtle, R.L.; Skinner, M.K. Environmental epigenomics and disease susceptibility. *Nat. Rev. Genet.* **2007**, *8*, 253–262.
- Okano, M.; Bell, D.W.; Haber, D.A.; Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **1999**, *99*, 247–257.
- Philibert, R.A.; Beach, S.R.H.; Brody, G.H. The DNA methylation signature of smoking: An archetype for the identification of biomarkers for behavioral illness. *Neb. Symp. Motiv.* **2014**, *61*, 109–127.
- Joubert, B.R.; Felix, J.F.; Yousefi, P.; Bakulski, K.M.; Just, A.C.; Breton, C.; Reese, S.E.; Markunas, C.A.; Richmond, R.C.; Xu, C.-J.; et al. DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. *Am. J. Hum. Genet.* **2016**, *98*, 680–696.

21. Keleher, M.R.; Zaidi, R.; Shah, S.; Oakley, M.E.; Pavlatos, C.; El Idrissi, S.; Xing, X.; Li, D.; Wang, T.; Cheverud, J.M. Maternal high-fat diet associated with altered gene expression, DNA methylation, and obesity risk in mouse offspring. *PLoS ONE* **2018**, *13*, e0192606.
22. Li, E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nat. Rev. Genet.* **2002**, *3*, 662–673.
23. Trojer, P.; Reinberg, D. Facultative heterochromatin: Is there a distinctive molecular signature? *Mol. Cell* **2007**, *28*, 1–13.
24. Fillion, G.J.; van Bommel, J.G.; Braunschweig, U.; Talhout, W.; Kind, J.; Ward, L.D.; Brugman, W.; de Castro, I.J.; Kerkhoven, R.M.; Bussemaker, H.J.; et al. Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in *Drosophila* Cells. *Cell* **2010**, *143*, 212–224.
25. Barski, A.; Cuddapah, S.; Cui, K.; Roh, T.-Y.; Schones, D.E.; Wang, Z.; Wei, G.; Chepelev, I.; Zhao, K. High-resolution profiling of histone methylations in the human genome. *Cell* **2007**, *129*, 823–837.
26. Dixon, J.R.; Selvaraj, S.; Yue, F.; Kim, A.; Li, Y.; Shen, Y.; Hu, M.; Liu, J.S.; Ren, B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **2012**, *485*, 376–380.
27. Filippova, D.; Patro, R.; Duggal, G.; Kingsford, C. Identification of alternative topological domains in chromatin. *Algorithms Mol. Biol. AMB* **2014**, *9*, 14.
28. Hnisz, D.; Weintraub, A.S.; Day, D.S.; Valton, A.-L.; Bak, R.O.; Li, C.H.; Goldmann, J.; Lajoie, B.R.; Fan, Z.P.; Sigova, A.A.; et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **2016**, *351*, 1454–1458.
29. Kaiser, V.B.; Taylor, M.S.; Semple, C.A. Mutational Biases Drive Elevated Rates of Substitution at Regulatory Sites across Cancer Types. *PLoS Genet.* **2016**, *12*, e1006207.
30. Mifsud, B.; Tavares-Cadete, F.; Young, A.N.; Sugar, R.; Schoenfelder, S.; Ferreira, L.; Wingett, S.W.; Andrews, S.; Grey, W.; Ewels, P.A.; et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* **2015**, *47*, 598–606.
31. Corces, M.R.; Corces, V.G. The three-dimensional cancer genome. *Curr. Opin. Genet. Dev.* **2016**, *36*, 1–7.
32. Achinger-Kawecka, J.; Clark, S.J. Disruption of the 3D cancer genome blueprint. *Epigenomics* **2017**, *9*, 47–55.
33. Kouzarides, T. SnapShot: Histone-modifying enzymes. *Cell* **2007**, *131*, 822.
34. Margueron, R.; Trojer, P.; Reinberg, D. The key to development: Interpreting the histone code? *Curr. Opin. Genet. Dev.* **2005**, *15*, 163–176.
35. Mantsoki, A.; Devailly, G.; Joshi, A. CpG island erosion, polycomb occupancy and sequence motif enrichment at bivalent promoters in mammalian embryonic stem cells. *Sci. Rep.* **2015**, *5*, 16791.
36. Bartke, T.; Vermeulen, M.; Xhemalce, B.; Robson, S.C.; Mann, M.; Kouzarides, T. Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell* **2010**, *143*, 470–484.
37. Tabolacci, E.; Chiurazzi, P. Epigenetics, fragile X syndrome and transcriptional therapy. *Am. J. Med. Genet. A.* **2013**, *161A*, 2797–2808.
38. Welter, D.; MacArthur, J.; Morales, J.; Burdett, T.; Hall, P.; Junkins, H.; Klemm, A.; Flicek, P.; Manolio, T.; Hindorf, L.; et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **2014**, *42*, D1001–D1006.
39. Visscher, P.M.; Brown, M.A.; McCarthy, M.I.; Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **2012**, *90*, 7–24.
40. Northcott, P.A.; Nakahara, Y.; Wu, X.; Feuk, L.; Ellison, D.W.; Croul, S.; Mack, S.; Kongkham, P.N.; Peacock, J.; Dubuc, A.; et al. Multiple recurrent genetic events converge on control of histone lysine methylation in medulloblastoma. *Nat. Genet.* **2009**, *41*, 465–472.
41. Roadmap Epigenomics Consortium; Kundaje, A.; Meuleman, W.; Ernst, J.; Bilenky, M.; Yen, A.; Heravi-Moussavi, A.; Kheradpour, P.; Zhang, Z.; Wang, J.; et al. Integrative analysis of 111 reference human epigenomes. *Nature* **2015**, *518*, 317–330.
42. Corradin, O.; Scacheri, P.C. Enhancer variants: Evaluating functions in common disease. *Genome Med.* **2014**, *6*, 85.
43. Corces, M.R.; Granja, J.M.; Shams, S.; Louie, B.H.; Seoane, J.A.; Zhou, W.; Silva, T.C.; Groeneveld, C.; Wong, C.K.; Cho, S.W.; et al. The chromatin accessibility landscape of primary human cancers. *Science* **2018**, *362*, eaav1898.
44. Severson, T.M.; Kim, Y.; Joosten, S.E.P.; Schuurman, K.; van der Groep, P.; Moelans, C.B.; Ter Hoeve, N.D.; Manson, Q.F.; Martens, J.W.; van Deurzen, C.H.M.; et al. Characterizing steroid hormone receptor chromatin binding landscapes in male and female breast cancer. *Nat. Commun.* **2018**, *9*, 482.

45. Ooi, W.F.; Xing, M.; Xu, C.; Yao, X.; Ramlee, M.K.; Lim, M.C.; Cao, F.; Lim, K.; Babu, D.; Poon, L.-F.; et al. Epigenomic profiling of primary gastric adenocarcinoma reveals super-enhancer heterogeneity. *Nat. Commun.* **2016**, *7*, 12983.
46. Müller-Tidow, C.; Klein, H.-U.; Hascher, A.; Isken, F.; Tickenbrock, L.; Thoennissen, N.; Agrawal-Singh, S.; Tschanter, P.; Disselhoff, C.; Wang, Y.; et al. Profiling of histone H3 lysine 9 trimethylation levels predicts transcription factor activity and survival in acute myeloid leukemia. *Blood* **2010**, *116*, 3564–3571.
47. Turcan, S.; Makarov, V.; Taranda, J.; Wang, Y.; Fabius, A.W.M.; Wu, W.; Zheng, Y.; El-Amine, N.; Haddock, S.; Nanjangud, G.; et al. Mutant-IDH1-dependent chromatin state reprogramming, reversibility, and persistence. *Nat. Genet.* **2018**, *50*, 62–72.
48. Agrawal-Singh, S.; Isken, F.; Agelopoulos, K.; Klein, H.-U.; Thoennissen, N.H.; Koehler, G.; Hascher, A.; Bäumer, N.; Berdel, W.E.; Thiede, C.; et al. Genome-wide analysis of histone H3 acetylation patterns in AML identifies PRDX2 as an epigenetically silenced tumor suppressor gene. *Blood* **2012**, *119*, 2346–2357.
49. Hoemme, C.; Peerzada, A.; Behre, G.; Wang, Y.; McClelland, M.; Nieselt, K.; Zschunke, M.; Disselhoff, C.; Agrawal, S.; Isken, F.; et al. Chromatin modifications induced by PML-RARalpha repress critical targets in leukemogenesis as analyzed by ChIP-Chip. *Blood* **2008**, *111*, 2887–2895.
50. Salzberg, A.C.; Harris-Becker, A.; Popova, E.Y.; Keasey, N.; Loughran, T.P.; Claxton, D.F.; Grigoryev, S.A. Genome-wide mapping of histone H3K9me2 in acute myeloid leukemia reveals large chromosomal domains associated with massive gene silencing and sites of genome instability. *PLoS ONE* **2017**, *12*, e0173723.
51. Hlady, R.A.; Sathyanarayan, A.; Thompson, J.J.; Zhou, D.; Wu, Q.; Pham, K.; Lee, J.H.; Liu, C.; Robertson, K.D. Integrating the Epigenome to Identify Novel Drivers of Hepatocellular Carcinoma. *Hepatol. Baltim. Md* **2018**.
52. Kelso, T.W.R.; Porter, D.K.; Amaral, M.L.; Shokhirev, M.N.; Benner, C.; Hargreaves, D.C. Chromatin accessibility underlies synthetic lethality of SWI/SNF subunits in ARID1A-mutant cancers. *eLife* **2017**, *6*, e30506.
53. Coetzee, S.G.; Shen, H.C.; Hazelett, D.J.; Lawrenson, K.; Kuchenbaecker, K.; Tyrer, J.; Rhie, S.K.; Levanon, K.; Karst, A.; Drapkin, R.; et al. Cell-type-specific enrichment of risk-associated regulatory elements at ovarian cancer susceptibility loci. *Hum. Mol. Genet.* **2015**, *24*, 3595–3607.
54. Valencia, A.M.; Kadoch, C. Chromatin regulatory mechanisms and therapeutic opportunities in cancer. *Nat. Cell Biol.* **2019**, doi:10.1038/s41556-018-0258-1.
55. Kulis, M.; Esteller, M. DNA Methylation and Cancer. *Adv. Genet.* **2010**, doi:10.1016/B978-0-12-380866-0.60002-2
56. Bernhart, S.H.; Kretzmer, H.; Holdt, L.M.; Jühling, F.; Ammerpohl, O.; Bergmann, A.K.; Northoff, B.H.; Doose, G.; Siebert, R.; Stadler, P.F.; et al. Changes of bivalent chromatin coincide with increased expression of developmental genes in cancer. *Sci. Rep.* **2016**, *6*, 37393.
57. Deng, J.; Shoemaker, R.; Xie, B.; Gore, A.; LeProust, E.M.; Antosiewicz-Bourget, J.; Egli, D.; Maherali, N.; Park, I.-H.; Yu, J.; et al. Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nat. Biotechnol.* **2009**, *27*, 353–360.
58. Easwaran, H.; Johnstone, S.E.; Van Neste, L.; Ohm, J.; Mosbrugger, T.; Wang, Q.; Aryee, M.J.; Joyce, P.; Ahuja, N.; Weisenberger, D.; et al. A DNA hypermethylation module for the stem/progenitor cell signature of cancer. *Genome Res.* **2012**, *22*, 837–849.
59. Wang, L.; Zhao, Z.; Ozark, P.A.; Fantini, D.; Marshall, S.A.; Rendleman, E.J.; Cozzolino, K.A.; Louis, N.; He, X.; Morgan, M.A.; et al. Resetting the epigenetic balance of Polycomb and COMPASS function at enhancers for cancer therapy. *Nat. Med.* **2018**, *24*, 758–769.
60. Mikeska, T.; Craig, J.M. DNA methylation biomarkers: Cancer and beyond. *Genes* **2014**, *5*, 821–864.
61. Da Costa, E.M.; McInnes, G.; Beaudry, A.; Raynal, N.J.-M. DNA Methylation-Targeted Drugs: *Cancer J.* **2017**, *23*, 270–276.
62. Soto-Ramírez, N.; Arshad, S.H.; Holloway, J.W.; Zhang, H.; Schauburger, E.; Ewart, S.; Patil, V.; Karmaus, W. The interaction of genetic variants and DNA methylation of the interleukin-4 receptor gene increase the risk of asthma at age 18 years. *Clin. Epigenet.* **2013**, *5*, 1.
63. James, S.J.; Melnyk, S.; Jernigan, S.; Pavliv, O.; Trusty, T.; Lehman, S.; Seidel, L.; Gaylor, D.W.; Cleves, M.A. A functional polymorphism in the reduced folate carrier gene and DNA hypomethylation in mothers of children with autism. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **2010**, *153B*, 1209–1220.
64. Teijido, O.; Cacabelos, R. Pharmacoeconomic Interventions as Novel Potential Treatments for Alzheimer's and Parkinson's Diseases. *Int. J. Mol. Sci.* **2018**, *19*, 3199.
65. Soragni, E.; Miao, W.; Iudicello, M.; Jacoby, D.; De Mercanti, S.; Clerico, M.; Longo, F.; Piga, A.; Ku, S.; Campau, E.; et al. Epigenetic therapy for Friedreich ataxia. *Ann. Neurol.* **2014**, *76*, 489–508.

66. Polak, P.; Karlic, R.; Koren, A.; Thurman, R.; Sandstrom, R.; Lawrence, M.S.; Reynolds, A.; Rynes, E.; Vlahovicek, K.; Stamatoyannopoulos, J.A.; et al. Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature* **2015**, *518*, 360–364.
67. Hou, X.; He, X.; Wang, K.; Hou, N.; Fu, J.; Jia, G.; Zuo, X.; Xiong, H.; Pang, M. Genome-Wide Network-Based Analysis of Colorectal Cancer Identifies Novel Prognostic Factors and an Integrative Prognostic Index. *Cell. Physiol. Biochem.* **2018**, *49*, 1703–1716.
68. Zhu, B.; Song, N.; Shen, R.; Arora, A.; Machiela, M.J.; Song, L.; Landi, M.T.; Ghosh, D.; Chatterjee, N.; Baladandayuthapani, V.; et al. Integrating Clinical and Multiple Omics Data for Prognostic Assessment across Human Cancers. *Sci. Rep.* **2017**, *7*, 16954.
69. Sekhon, A.; Singh, R.; Qi, Y. DeepDiff: DEEP-learning for predicting DIFFerential gene expression from histone modifications. *Bioinformatics* **2018**, *34*, i891–i900.
70. Wu, J.; Xu, J.; Liu, B.; Yao, G.; Wang, P.; Lin, Z.; Huang, B.; Wang, X.; Li, T.; Shi, S.; et al. Chromatin analysis in human early development reveals epigenetic transition during ZGA. *Nature* **2018**, *557*, 256–260.
71. Hon, G.C.; Hawkins, R.D.; Ren, B. Predictive chromatin signatures in the mammalian genome. *Hum. Mol. Genet.* **2009**, *18*, R195–R201.
72. Rada-Iglesias, A. Is H3K4me1 at enhancers correlative or causative? *Nat. Genet.* **2018**, *50*, 4–5.
73. Laird, P.W. Principles and challenges of genomewide DNA methylation analysis. *Nat. Rev. Genet.* **2010**, *11*, 191–203.
74. Rakyant, V.K.; Down, T.A.; Balding, D.J.; Beck, S. Epigenome-wide association studies for common human diseases. *Nat. Rev. Genet.* **2012**, *12*, 529–541.
75. Jiang, S.; Mortazavi, A. Integrating ChIP-seq with other functional genomics data. *Brief. Funct. Genom.* **2018**, *17*, 104–115.
76. Lappalainen, T.; Grealley, J.M. Associating cellular epigenetic models with human phenotypes. *Nat. Rev. Genet.* **2017**, *18*, 441–451.
77. Ritchie, M.D.; Holzinger, E.R.; Li, R.; Pendergrass, S.A.; Kim, D. Methods of integrating data to uncover genotype-phenotype interactions. *Nat. Rev. Genet.* **2015**, *16*, 85–97.
78. Auerbach, J.; Howey, R.; Jiang, L.; Justice, A.; Li, L.; Oualkacha, K.; Sayols-Baixeras, S.; Aslibekyan, S.W. Causal modeling in a multi-omic setting: Insights from GAW20. *BMC Genet.* **2018**, *19*, 74.
79. Buenrostro, J.D.; Giresi, P.G.; Zaba, L.C.; Chang, H.Y.; Greenleaf, W.J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **2013**, *10*, 1213–1218.
80. Devailly, G.; Mantsoki, A.; Joshi, A. Heat\*seq: An interactive web tool for high-throughput sequencing experiment comparison with public data. *Bioinforma. Oxf. Engl.* **2016**, *32*, 3354–3356.
81. Romanescu, R.G.; Espin-Garcia, O.; Ma, J.; Bull, S.B. Integrating epigenetic, genetic, and phenotypic data to uncover gene-region associations with triglycerides in the GOLDN study 06 Biological Sciences 0604 Genetics. *BMC Proc.* **2018**, *12*, 57.
82. Liang, L.; Willis-Owen, S.A.G.; Laprise, C.; Wong, K.C.C.; Davies, G.A.; Hudson, T.J.; Binia, A.; Hopkin, J.M.; Yang, I.V.; Grundberg, E.; et al. An epigenome-wide association study of total serum immunoglobulin e concentration. *Nature* **2015**, *520*, 670–674.
83. Wang, B.; DeStefano, A.L.; Lin, H. Integrative methylation score to identify epigenetic modifications associated with lipid changes resulting from fenofibrate treatment in families. *BMC Proc.* **2018**, *12*, 28.
84. Shah, S.; Bonder, M.J.; Marioni, R.E.; Zhu, Z.; McRae, A.F.; Zhernakova, A.; Harris, S.E.; Liewald, D.; Henders, A.K.; Mendelson, M.M.; et al. Improving Phenotypic Prediction by Combining Genetic and Epigenetic Associations. *Am. J. Hum. Genet.* **2015**, *97*, 75–85.
85. Zheng, R.; Wan, C.; Mei, S.; Qin, Q.; Wu, Q.; Sun, H.; Chen, C.-H.; Brown, M.; Zhang, X.; Meyer, C.A.; et al. Cistrome Data Browser: Expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Res.* **2018**, 1–7.
86. Ward, L.D.; Kellis, M. HaploReg v4: Systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res.* **2016**, *44*, D877–D881.
87. Szklarczyk, D.; Franceschini, A.; Wyder, S.; Forslund, K.; Heller, D.; Huerta-Cepas, J.; Simonovic, M.; Roth, A.; Santos, A.; Tsafou, K.P.; et al. STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **2015**, *43*, D447–D452.

88. Fabregat, A.; Jupe, S.; Matthews, L.; Sidiropoulos, K.; Gillespie, M.; Garapati, P.; Haw, R.; Jassal, B.; Korninger, F.; May, B.; et al. The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* **2018**, *46*, D649–D655.
89. Himmelstein, D.S.; Lizee, A.; Hessler, C.; Brueggeman, L.; Chen, S.L.; Hadley, D.; Green, A.; Khankhanian, P.; Baranzini, S.E. Systematic integration of biomedical knowledge prioritizes drugs for repurposing. *eLife* **2017**, *6*, e26726.
90. Haaland, Ø.A.; Lie, R.T.; Romanowska, J.; Gjerdevik, M.; Gjessing, H.; akon K.; Jugessur, A. A Genome-Wide Search for Gene-Environment Effects in Isolated Cleft Lip with or without Cleft Palate Triads Points to an Interaction between Maternal Periconceptional Vitamin Use and Variants in ESRRG. *Front. Genet.* **2018**, *9*, 1–16.
91. Lawlor, D.A.; Harbord, R.M.; Sterne, J.A.C.; Timpson, N.; Davey Smith, G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* **2008**, *27*, 1133–63.
92. Relton, C.L.; Davey Smith, G. Two-step epigenetic mendelian randomization: A strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int. J. Epidemiol.* **2012**, *41*, 161–176.
93. Latvala, A.; Ollikainen, M. Mendelian randomization in (epi)genetic epidemiology: An effective tool to be handled with care. *Genome Biol.* **2016**, *17*, 156.
94. Dekkers, K.F.; van Iterson, M.; Sliker, R.C.; Moed, M.H.; Bonder, M.J.; van Galen, M.; Mei, H.; Zhernakova, D.V.; van den Berg, L.H.; Deelen, J.; et al. Blood lipids influence DNA methylation in circulating cells. *Genome Biol.* **2016**, *17*, 138.
95. Cecil, C.A.M.; Walton, E.; Pingault, J.-B.; Provençal, N.; Pappa, I.; Vitaro, F.; Côté, S.; Szyf, M.; Tremblay, R.E.; Tiemeier, H.; et al. DRD4 methylation as a potential biomarker for physical aggression: An epigenome-wide, cross-tissue investigation. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **2018**, *177*, 746–764.
96. Howey, R.A.J.; Cordell, H.J. Application of Bayesian networks to GAW20 genetic and blood lipid data. *BMC Proc.* **2018**, *12*, 19.
97. Justice, A.E.; Howard, A.G.; Fernández-Rhodes, L.; Graff, M.; Tao, R.; North, K.E. Direct and indirect genetic effects on triglycerides through omics and correlated phenotypes. *BMC Proc.* **2018**, *12*, 22.
98. Yu, P.; Xiao, S.; Xin, X.; Song, C.-X.; Huang, W.; McDee, D.; Tanaka, T.; Wang, T.; He, C.; Zhong, S. Spatiotemporal clustering of the epigenome reveals rules of dynamic gene regulation. *Genome Res.* **2013**, *23*, 352–64.
99. Fizev, P.; Ernst, J. ChromTime: Modeling spatio-temporal dynamics of chromatin marks. *Genome Biol.* **2018**, *19*, 109.
100. Guo, S.; Diep, D.; Plongthongkum, N.; Fung, H.L.; Zhang, K.; Zhang, K. Identification of methylation haplotype blocks AIDS in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. *Nat. Genet.* **2017**, *49*, 635–642.
101. Romanowska, J.; Haaland, Ø.A.; Jugessur, A.; Gjerdevik, M.; Xu, Z.; Taylor, J.; Wilcox, A.J.; Jonassen, I.; Lie, R.T.; Håkon, K. Gjessing Integrating genome-wide methylation and genotype data to elucidate how region-wise methylation level might influence allele-defined relative risks. Submitted, 2019.
102. Dolinoy, D.C.; Weidman, J.R.; Waterland, R.A.; Jirtle, R.L. Maternal genistein alters coat color and protects Avy mouse offspring from obesity by modifying the fetal epigenome. *Environ. Health Perspect.* **2006**, *114*, 567–572.
103. Briffa, J.F.; Wlodek, M.E.; Moritz, K.M. Transgenerational programming of nephron deficits and hypertension. *Semin. Cell Dev. Biol.* **2018**, doi:10.1016/j.semcdb.2018.05.025.
104. Dekker, J.; Rippe, K.; Dekker, M.; Kleckner, N. Capturing chromosome conformation. *Science* **2002**, *295*, 1306–1311.
105. Lieberman-Aiden, E.; van Berkum, N.L.; Williams, L.; Imakaev, M.; Ragooczy, T.; Telling, A.; Amit, I.; Lajoie, B.R.; Sabo, P.J.; Dorschner, M.O.; et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **2009**, *326*, 289–293.
106. Fraser, J.; Williamson, I.; Bickmore, W.A.; Dostie, J. An Overview of Genome Organization and How We Got There: From FISH to Hi-C. *Microbiol. Mol. Biol. Rev. MMBR* **2015**, *79*, 347–372.
107. Wu, C.; Pan, W. Integration of Enhancer-Promoter Interactions with GWAS Summary Results Identifies Novel Schizophrenia-Associated Genes and Pathways. *Genetics* **2018**, *209*, 699–709.
108. Thurman, R.E.; Rynes, E.; Humbert, R.; Vierstra, J.; Maurano, M.T.; Haugen, E.; Sheffield, N.C.; Stergachis, A.B.; Wang, H.; Vernot, B.; et al. The accessible chromatin landscape of the human genome. *Nature* **2012**, *489*, 75–82.

109. Vipin, D.; Wang, L.; Devailly, G.; Michoel, T.; Joshi, A. Causal Transcription Regulatory Network Inference Using Enhancer Activity as a Causal Anchor. *Int. J. Mol. Sci.* **2018**, *19*, 3609.
110. Radkiewicz, C.; Johansson, A.L.V.; Dickman, P.W.; Lambe, M.; Edgren, G. Sex differences in cancer risk and survival: A Swedish cohort study. *Eur. J. Cancer* **2017**, *84*, 130–140.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).