Timeline of Epigenetic Changes Induced by Ethanol (EtOH) in HepG2 and HL60

Sophia Sayeed Qureshi

This thesis is submitted in partial fulfilment of the requirements for the degree of Master of Science

Department of Biological Science

University of Bergen

November 2019

Acknowledgements

This project is a collaboration between the Department of Molecular Biology, Faculty of Mathematics and Natural Sciences, University of Bergen and Dr. Einar Martens Research Group for Biological Psychiatry, Department of Clinical Science, University of Bergen.

I would like to express my gratitude to various people that have, for better and for worse, been journeying with me in this thesis. First, I would like thank my supervisors; Professor Stephanie Le Hellard for letting me a part of the group and for her positive attitude, Anne-Kristin Stavrum for her never-ending patience and Anja Torsvik for her amazing knowledge in cell culture (and helping my cells to live) An honourable mention to Andrea, Carla and Jonelle, for helping me and occasionally making me laugh. Working with all of you has indeed been an immensely educational experience!

To all family and friends, who in one way or another shared their support, either morally, academically or technically, thank you. Lastly, I would like to thank myself for rising from numerous mental breakdowns & crying sessions throughout my academic career, and for writing this thesis! Sophia, you did good! 😉

Sophia Sayeed Qureshi,
2019 - Bergen

Table of contents

1. Abstract

Gene activity may be modified though epigenetics. These changes are dynamic and provide a mechanism to regulate gene expression without engraving changes into the genome. Evidence show take gene activity may adapt to environmental factors such as nutrition, behaviour, lifestyle, climate and toxins that have an impact on different cells and ultimately change their DNA. Although DNA methylation changes are reversible, several studies have shown that some of the changes remain for a long time, and some have even been shown to be transmitted trans-generationally. The occurrence of epigenetic modifications has been established, but the timescale regarding how fast these modifications take place are still unknown.

In this study, we aimed to determine an epigenetic timeline that would describe how fast epigenetic modifications occur, and how fast the epigenetic modifications are reversed. Ethanol (EtOH) was chosen as inducer as it is already well documented that exposure of alcohol, particularly EtOH, has an effect on the DNA methylation, and may affect other layers of epigenetic modification. The cells were subjected to two different EtOH concentrations; 20mM and 100mM, that corresponds to 0.08 grams alcohol/100 mL (0.8 ‰) and 4.6 grams alcohol/100 mL (4.6 ‰), respectively. Cell lines from hepatocytes (HepG2) and the immune system (HL60) were chosen since methylation changes in DNA due to alcohol are already well documented for both. Both cell lines were exposed to EtOH for 24 hours, and the timeline was set to one week where the time points for harvesting of cells were 0h (control), 0.5h, 1h, 3h, 6h, 12h, 24h, 48h, 96h and 192h. Since EtOH evaporates quite easily, the alcohol-based cell medium was changed every 8h. After 24h, the cells were switched back to growth media to recover. The methylation data was assessed genome-wide using the 850k EPIC array (Illumina).

Our main finding is that DNA methylation changes much faster in response to EtOH than what we expected. The first sample was collected 30 min after addition of EtOH, with intention of picking up the initiation of DNA methylation. However, our results indicate that the process of methylation has already started, postulating that the process begins even before 30 minutes. Subsequently, we discovered that DNA methylation at some positions are also reversed shortly after ending the exposure to EtOH. The first post-treated sample was collected after 48h i.e. 24h after the cells were exposed to EtOH. This time interval apparently gives the cells sufficient/enough time to recover and reverse some of the DNA methylation changes back to its original state without being picked up by the following time points. Other areas that displayed a gradual change in methylation or remained changed after treatment. Our findings are of importance as the timeline of DNA methylation in epigenetic changes has previously been unknown.

2. Abbreviations

5caC	5-formylcytosine 5-carboxylcytosine (5caC),
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytocine
5hmU	5-hydroxymethyluracil
5mC	5-methylcytosine
A	Adenine
AID	Activation induced deaminase
AM-AR	Active modification-active removal
AML	Acute myeloid leukaemia
APOBEC	Apolipoprotein B mRNA-editing enzyme complex
BER	Base excision pair
BME (β-ME)	2-mercaptoethanol
bp	Base pairs
C	Cytosine
CML	Chronic myeloid leukaemia
CpG	Dinucleotide segments of cytosine-guanine
CREB	cAMP response element-binding
DMEM	Dulbecco's Modified Eagles' Medium (DMEM)
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
EPIC	HumanMethylationEPIC
EtBr	Ethidium bromide
EtOH	Ethanol
EWAS	Epigenome-wide association study
FBS	Fetal bovine serum
FDR	False discovery rate
G	Guanine
GR	Glucocorticoid receptor
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HAD	Histone deacetyltransferase

HCl	Hydrochloric acid
IMDM	Iscove's Modified Dulbecco's Medium
K	Lysine
Limma	Linear Models for Microarray
mRNA	Messenger RNA
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
Oxi-mCs	Oxidised methylcytosines
PBS	Phosphate-buffered saline
PCA	Principal component analysis
Pen-Strep	Penicillin Streptomycin
PTM	Post-Translational Modification
PTSD	Post-traumatic stress disorder
QQ-plot	Quantile-quantile plot
R	Arginine
RNA	Ribonucleic acid
ROS	Reactive oxygen species
rRNA	Ribosomal RNA
RT	Room temperature
RNAse –	Ribonuclease
S	Serine
SAM	S-adenosyl methionine
T	Thymine (base)
T	Threonine (amino acid)
TDG	Thymine DNA glycosylase
TEMED	N, N, N', N'-tetramethylethane-1,2-diamine
TET	Ten-eleven translocation
tRNA	Transfer RNA
TFs	Transcription Factors
TFBSs	Transcription Factor Binding Sites
Tris	2-amino-2-hydroxymethyl-1,3-propanediol
TSS	Transcription Start Site
U	Uracil
Y	Tyrosine

3. List of Figures

4. List of Tables

5. Introduction

The term “genetics” encompasses a wide branch of biological sciences related to genes, heredity and variation in organisms, and overlaps with other areas such as agriculture, biotechnology and medicine (Winchester, 2019). The definition arose from the identification of genes, the fundamental unit responsible for hereditary information. Alterations such as mutations, deletions, insertions and translocations may all introduce changes in this information by affecting the gene activity and function. Whether geneticists work at a molecular, cellular, organismal, familial, populational or evolutionary level, genes are always central in their studies (Griffiths AJF, 2000). Genetics may therefore be defined as the study of genes at several levels; from the way they behave in the cell to how they are transmitted from one generation to another. Furthermore, the term may be applied to principles and analytical procedures related to genes (Griffiths AJF, 2000). In order to comprehend the significance of genes, their impact on chemical reactions occurring within the cell, and gene-environment interactions, it is essential to understand the chemical composition of genes i.e. deoxyribonucleic acid (DNA) (Winchester, 2019).

5.1 What is DNA?

DNA is the chemical name of the molecule that carries genetic information, and it is the hereditary material in almost all organisms (Watson et al., 2013). It consists of two strands that are twisted around each other, forming a spiral called a double helix. Each strand is composed of nucleotides where each monomeric unit is made up of a phosphate group and a deoxyribose sugar group, forming the backbone of the DNA, along with one of the chemical bases adenine (A), guanine (G), cytosine (C) and thymine (T) (Watson et al., 2013). The nucleotides are linked together by forming a covalent bond between the sugar-group of one nucleotide and the phosphate-group attached to the following nucleotide, thereby yielding an alternating sugar-phosphate backbone (Watson et al., 2013). The order of these nitrogenous bases determines the information available for coding, building and maintaining an organism. The bases can be divided into two categories; purines and pyrimidines. The purines include adenine and guanine, both derived from the double-ringed structure whereas pyrimidines include the remaining bases cytosine and thymine (see **Figure 1**) (Watson et al., 2013). With respect to both shape and bonding properties of the bases, adenine is complementary with thymine, and cytosine with guanine, and vice versa. The A-T base pairs form two hydrogen bonds whereas C-G form three, being slightly more thermodynamically stable. However, as both pairs always consist of one purine base and one pyrimidine base, the distance across the double helix remains constant (Bruce R. Korf, 2013). The structure of the double helix is stabilised by base pairing as well as base stacking (Watson et al., 2013).

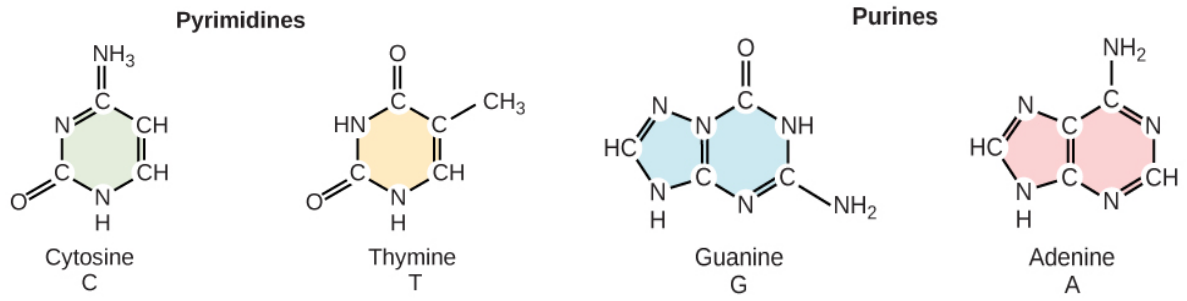


Figure 1 – Bases found in the DNA

Chemical structure of the nitrogenous bases adenine (A), cytosine (C), guanine (G) and thymine (T) that the DNA molecule consists of. Figure adapted from https://socratic.org/questions/what-are-the-four-bases-of-dna

5.2 DNAs association within the cell

Within the cell, the DNA molecule is associated with proteins and organised into structures called chromosomes. This organisation holds true for all types of cells. Eukaryotic organisms primarily store their DNA in the cell nucleus, and to a lesser extent in the mitochondria. However, the main DNA storage site for prokaryotic cells is the cytoplasm. Packaging of DNA into chromosomes serves several purposes. Firstly, a compact form of the DNA fits easily inside the cell and protects it from damage. Secondly, "naked" DNA is relatively unstable, whereas chromosomal DNA is greatly stable in cells. During mitotic cell division, only tightly packaged DNA has the ability to be transmitted to both daughter cells (Watson et al., 2013). Ultimately, the chromosomal structure confers an overall organisation to each molecule of the DNA, thereby regulating its accessibility as well as cellular processes that are linked to it (Watson et al., 2013).

In eukaryotic cells, most of the DNA is packed into nucleosomes, with each unit consisting of a core of eight histone proteins and DNA wrapped around them. Histones are positively charged proteins that have a close interaction with the negatively charged DNA, and play an essential role in the structural and functional transition between active and inactive chromatin states (Mariño-Ramírez, Kann, Shoemaker, & Landsman, 2005). The octameric core has a high degree of conservation, as it is well preserved by surrounding DNA. They consist of amino-terminal "tails" that contribute to stabilising the wrapping of DNA around the octamer. However, the tails are susceptible to post-translational modifications (PTM) e.g. acetylation, phosphorylation and methylation. It is important to specify that such modifications do not affect the DNA sequence per se, although they regulate the availability to the transcriptional machinery as well as the degree of compactness (how tightly or loosely DNA is packed around the octamer) (Alaskhar Alhamwe et al., 2018). The importance and impact of histone modifications will be discussed later.

The genome does not have a universal structure but is rather comprised of regions with varying chromosomal density. These regions are commonly divided into two groups; heterochromatin is characterised by a condensed structure with limited gene expression, whereas euchromatin consists of a more open structure and higher degree of gene expression. In humans, regions that surround the centromere and telomere on a chromosome are heterochromatic, while most of the chromosome arms are transcriptionally competent euchromatin (Grunau, Buard, Brun, & De Sario, 2006). Although heterochromatic regions display low levels of gene expression, they are of importance as their compact structure in centromeric regions are associated with elevated levels of DNA methylation, specific methylation patterns for histones, low recombination frequency and repression of transcription (Grunau et al., 2006).

Almost half of the genome consists of repetitive DNA that are often subclassified as satellite DNA and genome-wide repeats. Satellites DNA is a collective term for tandem repeats i.e. sequences consisting of one or more segments that are repeated right next to each other with lengths varying from ~ 5-50 bp (Vieira, Santini, Diniz, & Munhoz, 2016). These repetitive segments arise from difficulties in accurately duplicating the DNA and are more prone to mutation (Vieira et al., 2016; Watson et al., 2013). Genome-wide repeats are larger segments where each repeated unit ranges from >100 bp to >1kb. Genome-wide repeats are transposable elements i.e. sequences that have the ability to move from one region to another within the genome (Vieira et al., 2016). These segments can either be found as closely spaced clusters or dispersed as single copies throughout the genome (Biscotti, Olmo, & Heslop-Harrison, 2015; Vieira et al., 2016). However, these repetitive DNA fragments present technical challenges to analyses such as assembly programs and sequence alignments. Although some repeating units appear to be non-functional, others have contributed to the human evolution by creating variety and at times novel functions (Treangen & Salzberg, 2011).

5.3 Gene function

The basic tenet of molecular genetics revolves around how the genetic information is translated and expressed, also referred as the central dogma. DNA, as formerly mentioned, consists of the hereditary message and encodes ribonucleic acid (RNA) which further encodes the amino acid sequence of proteins. Previously, genes were defined as the fundamental unit responsible for hereditary information. However, in some cases, a gene also refers to a sequence of nucleotides that encodes a specific RNA and/or protein.

The first step of gene expression is transcription, a process where a particular DNA sequence of a gene is copied into messenger RNA (mRNA). There are some genes that are involved in basic cell maintenance and are expressed nearly ubiquitously to maintain constant levels of expression in all cells

and conditions (Eisenberg & Levanon, 2013). These are referred to as housekeeping genes and are often used for calibration in many biotechnical procedures and genomic studies. Conversely, other genes have strictly regulated expression with specific genes being turned on/off in particular cells at specific time points and in response to a given signal (Bruce R. Korf, 2013).

Gene expression is regulated by proteins called transcription factors (TFs) that bind to specific regions named transcription factor binding sites (TFBSs) along the DNA. This may activate or repress transcription, and can occur at different time points during the process (Karni, 2007). Some TFs have the ability to regulate multiple genes, while others can act both as activators and repressors (on the same gene region) depending on the signals they receive from their surroundings. These sites are mostly located in a region that is upstream (5'-end) with respect to the orientation of the DNA sequence and are referred to as promoters. The length can vary from 100 to 2000 bp upstream of the transcription start site (TSS). In most cases the promoter regions are found within ~ 100 bp of TSS and are characterised by having a base sequence consisting of T and A bases called the TATA box (Karni, 2007). Some genes may have multiple alternative promoters situated at different sites within the gene that respond to regulatory factors in other cellular tissues. These factors may occur adjacent to the promoter or located thousands of base pairs away. The latter are known as enhancers and function regardless of whether they are located upstream or downstream within the gene. Many transcription factors, either being activators or repressors of transcription that bind to the promoter or regulatory sequences, are regulated by specific signalling molecules called ligands. These molecules change the conformation in order to activate or inactivate the TFs. In some cases, the presence of one or several additional proteins is also required for a reaction to occur or to facilitate the transition. These "helper" proteins are called cofactors but are often sub-grouped as coactivator and corepressor with respect to their properties. **Figure 2** display the necessary segments within the gene that are required for initiation of transcription. Exons are DNA fragments that code for amino acids, ultimately yielding a protein while introns are non-coding segments. To obtain a functional mature mRNA, the introns must be removed before the remaining exons can be spliced together.

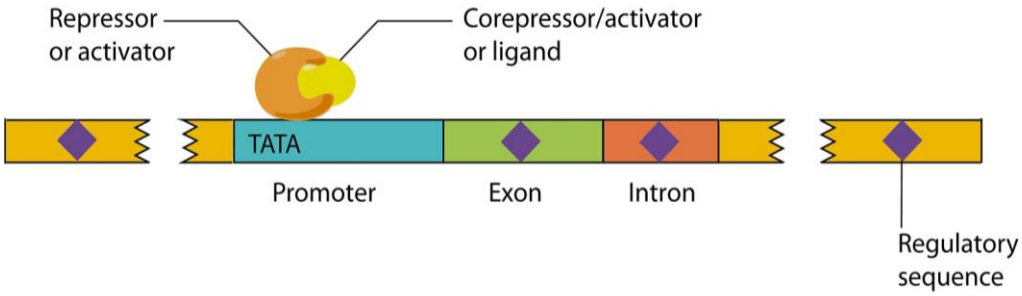


Figure 2 – Gene transcription

Necessary components that are needed to regulate gene expression. The process is initiated by the binding RNA polymerase in the promoter region and is controlled by transcription factors (activator or repressor). Some of them require the presence of cofactors such as ligands. Regulatory sequences are alternative promoters, and exons and introns are the coding and non-coding genetic regions, respectively. Figure from Human Genetics and Genomics (2013), 4th edition. Korf & Irons

5.4 Transcription of DNA into mature mRNA

The transcription process is initiated by the attachment of RNA polymerase, an enzyme that binds to the promoter region and catalyses the production of complimentary RNA (Clancy, 2008). In eukaryotic organisms, the enzyme can be divided into three types, designated as type I, II and III. Most gene transcriptions for mRNA and therein proteins are transcribed by RNA polymerase type II, whereas type I and III transcribe ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs), respectively (Clancy, 2008). The transcribed strand is an exact replicate of the DNA sequence, with exception of thymine that is replaced with uracil (U). Additionally, a 7-methyl guanine residue is attached to the 5'-end of the RNA molecule in the beginning of transcription whereas ~100-200 adenine bases are added at the 3'-end prior to termination (Clancy, 2008). These components stabilise the molecule and facilitate its export into the cytoplasm (Clancy, 2008).

During transcription the introns are spliced away, resulting in exon-segments that are ligated together and yields the mature mRNA. However, not necessarily all exons are spliced together, or in the same chronological order. This process, called both alternative splicing and differential splicing, allows individual genes to produce multiple mRNAs with varying exons (Matlin, Clark, & Smith, 2005). Due to the complexity of this process, the spliced mRNA is vulnerable to disruption by mutations. However, it can also be a point of control for gene expression. Under the impact of control molecules found in specific cells, certain exons may or may not be introduced in the mRNA due to differential splicing (Matlin et al., 2005). This leads to the possibility of generating multiple distinct proteins from the same gene, thereby substantially increasing the diversity of proteins being encoded by the genome.

Particular exons may conform to certain functional protein domains, resulting in production of several proteins with a range of functions from the same gene (Matlin et al., 2005). Occasionally, alternative splicing pathways may be determined through the presence or absence of a single regulator. More frequently, splicing pathways are determined through a combination of various factors. This establishes the concept of "cellular codes", consisting of specific combinations of regulatory factors (Matlin et al., 2005).

5.5 Translation of mature mRNA into protein

Once the mature mRNA is transcribed in the cell's nucleus, it is exported to the cytoplasm for translation, commonly known as protein synthesis. As the name indicates, during this process the mRNA sequence is read and converted into the amino acid sequence of a protein. The translation is performed by a protein-RNA complex called ribosome, which initiates the process by binding itself to the 7-methyl guanine cap at the 5'-end of the mRNA and scanning along until the initiation codon AUG (methionine) is recognised (Hershey, Sonenberg, & Mathews, 2012). However, for the process to continue, tRNA molecules as well as aminoacyl-tRNA enzymes are required (Lodish, 2013). The genetic code consists of 20 amino acids, and each of them are recognised by its own tRNA molecule and tRNA synthase.

Although, modifications may occur, in turn yielding alternated amino acids and tRNA molecules. A study from Pereira et al. (2018) reported that more than 80 tRNA have been reported, and an average of 13 modified bases are expected per tRNA molecule (Pereira et al., 2018). A tRNA molecule consisting of the anticodon that corresponds with the codon on mRNA are base-paired by the enzyme that codes for that particular amino acid, forming an aminoacyl-tRNA complex (Lodish, 2013).

This process is repeated, thereby building a polypeptide chain consisting of amino acid residues that will ultimately become a protein. Translation is terminated when a stop codon is reached (UAA, UAG or UGA).

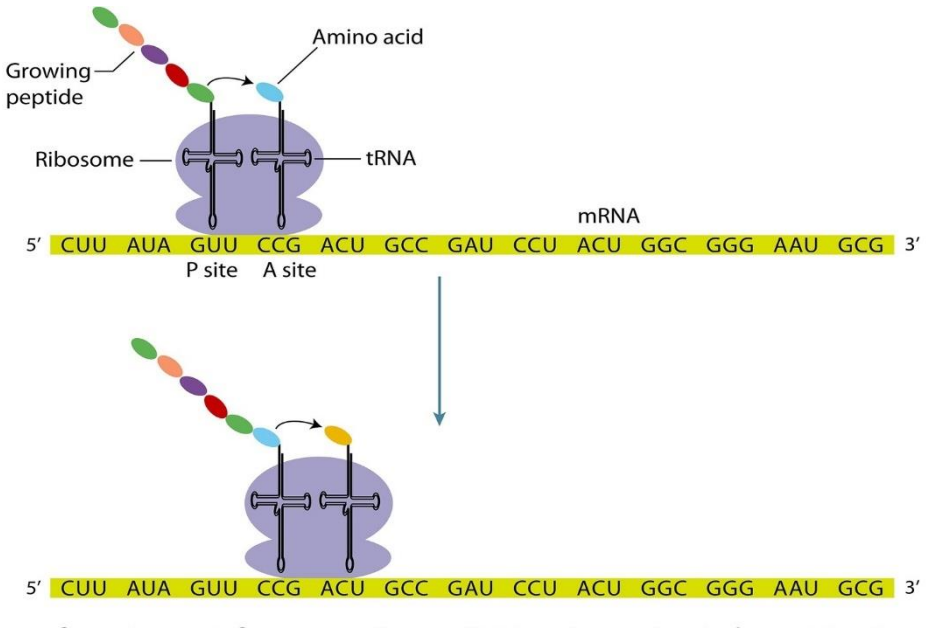


Figure 3 – Protein translation

Formation of polypeptide consisting of amino acid residues from mature mRNA. The process takes place on the ribosomal machinery, which binds to the RNA stand. Specific amino acyl tRNA molecules are bound to mRNA through base-pair complementarity between the codon on mRNA and its anticodon on tRNA, yielding the elongating polypeptide. The process is terminated once a stop codon is read. Figure from Human Genetics and Genomics (2013), 4th edition. Korf & Irons

5.6 Epigenetics

Genes can be temporarily activated or silenced through epigenetics. However, there are situations in which one or multiple genes are permanently altered. The term epigenetics generally refers to chemical modification to DNA or its associated histone proteins that further affects the compaction of the chromatin structure and readability of the genomic regions without altering the base sequence in the DNA (Rivera & Bennett, 2010). These changes are heritable and can be stably transmitted through multiple cell divisions (Rivera & Bennett, 2010).

The changes are initially the consequence of gene expression of specific transcription factors. They regulate cellular differentiation as well as control the expression of other genes that encode transcription factors and proteins involved in cell-cell communication (Rivera & Bennett, 2010). However, recent findings suggest that epigenetic modifications can also be induced by environmental factors at different time points, being permanent or temporary (Sadakierska-Chudy, Kostrzewa, & Filip, 2015). Previously, the classical epigenetic modifications consisted of DNA methylation and PTMs of histone proteins. Gene silencing of one of the two X-chromosomes in females and genomic imprinting are both examples of permanent changes that occur through methylation.

However, proteins such as DNA methyltransferase (DNMT), histone acetyltransferase (HAT) and deacetyltransferase (HDA) and protein complexes with epigenetic modifications are now included under the umbrella definition of epigenetics (Rivera & Bennett, 2010). Additionally, emerging evidence suggests that several classes of functional non-coding RNAs must be added to the definition as they regulate gene regulation at a transcriptional and post-transcriptional level (Garber, 2019; Rivera & Bennett, 2010).

5.6.1 DNA Methylation

DNA methylation is a biochemical modification where a methyl (CH_3) group is added to the 5' position of the cytosine ring, thus forming 5-methylcytosine (5mC) (see **Figure 4**). This mechanism primarily occurs in regions near the promoter where cytosine is followed by a guanine residue (CpG), but the modification can also be found at different CpG sites throughout the genome. Regions that have a high occurrence of the mentioned dinucleotides are referred as CpG island (Joseph, Strand, & Vezina, 2018). Many of these islands function as promoters and/or enhancers for initiation of transcriptional expression and are generally unmethylated (Joseph et al., 2018). In epigenetic studies these methylated islands are used as markers that represent observed changes within an organism (Moore, Le, & Fan, 2013).

The addition of methyl group to cytosine residues is a process that is highly conserved (Joseph et al., 2018), and is heritable by somatic cells after cell division (Bollati & Baccarelli, 2010). DNMT, which is a group of enzymes, bind to the DNA and "flips" out cytosine from the double helix before attaching a methyl group from S-adenosyl methionine (SAM) to C5 position (Joseph et al., 2018; Sadakierska-Chudy et al., 2015). DNMT1 preferentially methylates at hemi-methylated DNA sites and preserves the methylation patterns during cell replication. In contrast, DNMT3A and DNMT3B prefer un-methylated CpG sites and perform *de novo* methylation in gametogenesis and early development (Sadakierska-Chudy et al., 2015). **Figure 4** illustrates the reaction mechanism for methylation of cytosine to 5-methylcytosine with the presence of SAM and DNMT.

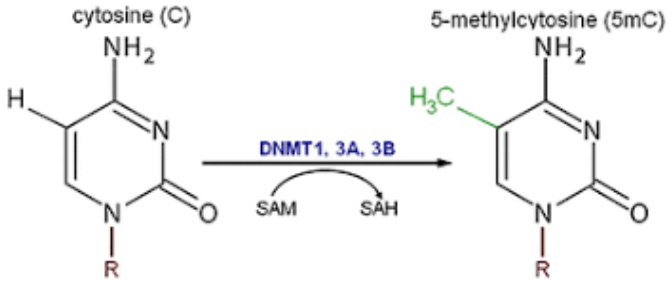


Figure 4 – Methylation of cytosine to 5-methylcytosine

Schematic diagram of cytosine being methylated at C5 position with the presence of S-adenosyl methionine (SAM) and DNA methyltransferase (DNMT) to 5-methylcytosine. Figure from Sadakierska-Chudy et al. (2015)

Gene silencing is primarily associated with hypermethylation of CpG sites at promoter region, while hypomethylation causes elevated levels of gene expression (Sadakierska-Chudy et al., 2015). CG-rich sequences are loosely bound to the histone octamer compared to CG-poor sequences, as more energy is required to bend them into the loops to wrap them around, making them more accessible. The level of DNA methylation of cytosine thereby affects transcriptional activities, and a deregulation may therefore have an impact on development, cellular function or diseases (Sadakierska-Chudy et al., 2015). DNA methylation is a reversible modification, thus implying that DNA can also be demethylated. Both active and passive demethylation pathways have been proposed.

5.6.2 Active DNA demethylation

Cytosine is known to exist in one of two functional states, unmethylated or methylated. However, recent studies have discovered and identified several variants of cytosine (see **Figure 5**) (Sadakierska-Chudy et al., 2015). These oxidised methylcytosines (oxi-mCs) are achieved through step-wise oxidation of 5-methylcytosine to 5-hydroxymethylcytocine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), respectively, by Ten-eleven translocation (TET) enzymes. These derivatives are all intermediates that are formed in the reversal of the methylation process and serve as stable epigenetic modifications that exert distinctive regulatory roles (An, Rao, & Ko, 2017).

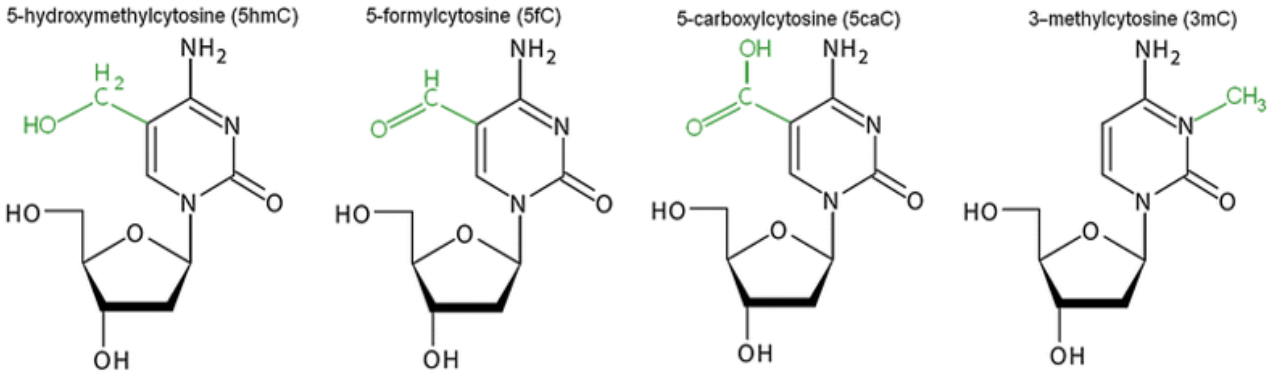


Figure 5 – Cytosine derivatives found in DNA

Chemical structures of the currently discovered cytosine derivatives. Figure from Sadakierska-Chudy et al. (2015)

Active DNA demethylation can be achieved through oxidation or deamination of the cytosine derivatives. Via the oxidation pathway, the oxidised methylcytosines 5fC and 5caC are directly excised by the DNA repair enzyme thymine DNA glycosylase (TDG), creating an abasic site. Subsequently, the damaged base is repaired in a step-wise process by base excision repair (BER) that ultimately yields an unmethylated cytosine (An et al., 2017; Wu & Zhang, 2017).

Since TET enzymes appear to have a certain preference for their substrates, 5hmC is in most cases not committed to the demethylation pathways due to its restrained conformation within the active site (An et al., 2017). This ultimately leads to a decrease in catalytic efficiency as 5hmC is less prone to further oxidation to 5fC and 5caC, it is considered a stable mark for epigenetic modifications (An et al., 2017). The remaining pathway is deamination, a biochemical reaction where an amino group (NH_3) group is removed from a molecule. In DNA, deamination has primarily been associated with the conversion of cytosine to uracil (Bochtler, Kolano, & Xu, 2017). However, 5mC and 5hmC can also be deaminated and yield thymine and 5-hydroxymethyluracil (5hmU), respectively, by enzymes from the activation induced deaminase (AID) and apolipoprotein B mRNA-editing enzyme complex (APOBEC) family (Sadakierska-Chudy et al., 2015). The formed intermediates are subjected to various DNA glycosylase enzymes that restores cytosine through the BER pathway (Bochtler et al., 2017; Sadakierska-Chudy et al., 2015).

As one can observe from the mentioned pathways, the process of active DNA demethylation can be regulated at multiple levels. The enzymatic reactions are directly dependent on substrate availability and in some cases the presence of cofactors. Furthermore, all genes can be regulated at transcriptional, post-transcriptional and post-translational levels (Wu & Zhang, 2017). Factors that target the demethylation machinery to particular genomic regions may also regulate the process (An et al., 2017). Since both TET and TDG are essential components of active DNA demethylation, it may be postulated that these proteins and their catalytic products are potential key regulators of processes such as embryonic development, stem cell function and cell lineage differentiation (An et al., 2017). Additionally, emerging evidence suggests that active DNA demethylation is linked to DNA repair as TDG-mediated BER is part of the active modification-active removal (AM-AR) process.

5.6.3 Passive DNA demethylation

In similarity with active DNA demethylation, this process allows 5mC to be reversed into its unmodified state. This occurs when there is a reduction in activity or absence of DNMTs that leads to a dilution of 5mC as the cell divisions (DNA replication) occur. In vitro, it was observed that lack of DNMT1 in both murine and human cells lead to death of differentiated cells within a few generations (Bochtler et al., 2017). Additionally, emerging evidence indicates that deficiency of DNMT1 causes a delay in the embryonal development, ultimately leading to death in mid-gestation (Bochtler et al., 2017). **Figure 6** display a schematic overview of passive and active DNA demethylation (oxidation pathway).

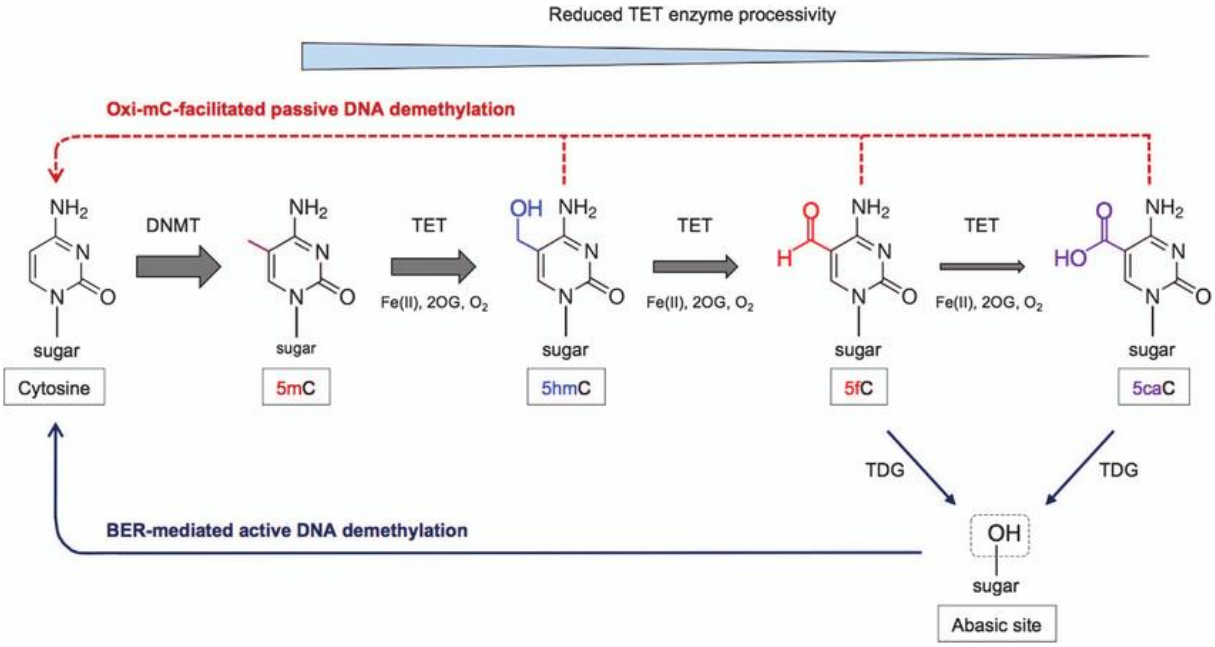


Figure 6 – Active and passive DNA demethylation in eukaryotes

Schematic overview of active and passive DNA demethylation. In active DNA demethylation the oxidised products 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) are first subjected to thymine DNA glycosylase (TDG) that creates an abasic site before being restored to cytosine with DNA glycosylase (TDG) and base excision repair (BER). In passive DNA demethylation there is a replication-dependent dilution of 5mC as multiple cell divisions occur. Figure from An et al. (2017)

5.6.4 Histone variants

As previously mentioned, histones are small, positively charged proteins that the DNA is wrapped around. Its octameric core is well preserved from alterations but consists of tails that "stick" out and are susceptible to a variety of PTMs. They are composed of various residues of positively charged amino acids, mainly arginine ® and lysine (K), that facilitate binding with the negatively charged DNA. The most common modifications consist of acetylation of arginine, acetylation and methylation of lysine and phosphorylation of serine (S), threonine (T) and tyrosine (Y) residues (Bannister & Kouzarides, 2011). Modifications of these tails will affect the degree of compactness formed between the DNA strand and the histone proteins, and further affect the transcriptional machinery as well. For instance, acetylation or phosphorylation of histone tails reduces the overall positive charge, subsequently reducing the affinity (Brehove et al., 2015; Higashi et al., 2007).

According to a study from Vogler et al (2010), it appears that the C-terminal tail of the H2A, a subunit of the octameric core, provides stabilisation to the nucleosomal core particle and interacts with components that control dynamics and conformation of the chromatin (Vogler et al., 2010). Although they are highly conserved, there are greatly similar variants of histones. In humans, there are currently 57 histone variants that are encoded by 94 genes, and it is the incorporation of these variants that affects

the chromosomal structure and genetic expression (Sadakierska-Chudy et al., 2015). They have emerged as important factors in regulating the chromatin states and DNA repair in response to genotoxic treatment, hence suggesting that they may serve as prognostic indicators of cancers for future treatments (Sadakierska-Chudy et al., 2015).

5.7 Environmental epigenetics

It has been discovered that epigenomic changes are dynamic and can regulate which genes are turned on and off, whereas the genome and its sequence is fixed. These changes allow organisms to adapt to changes introduced in an environment without engraving that particular change into the genome (Drouet, 2019). Emerging evidence suggests that genes take notice of the environment in ways that affects ones health and behaviour (Drouet, 2019). Several factors such as nutrition, behaviour, lifestyle, climate and toxins are examples of environmental cues that can have an impact on different cells and ultimately their DNA (Tiffon, 2018).

5.7.1 Epigenetics after exposure of environmental chemicals and toxins

Identification of chemical substances and effects of exposure are essential for risk assessment and prevention, and have been categorised according to their capability to induce mutations and alter the DNA sequence (Baccarelli & Bollati, 2009). However, studies have reported epigenetic modifications that have changed the genomic function after exogenous influence, thereby indicating that some chemical substances yield both genetic and epigenetic modifications.

Metals are a group of chemical substances that have an established association with DNA methylation (Baccarelli & Bollati, 2009). Several studies have reported that metals increase the production of reactive oxygen species (ROS) and damage the DNA. Oxidative DNA damage can affect DNMT by interfering with their ability to interact with DNA, thereby resulting in altered methylation. Metals such as cadmium and arsenic appeared to be linked to ROS and yielded global hypomethylation whereas nickel was reported to hyper-methylate regions that would lead to inactivation of gene expression (Baccarelli & Bollati, 2009). Nickel was also shown to affect histone modification by reducing acetylation and increasing both demethylation and ubiquitination on several histone proteins (Baccarelli & Bollati, 2009). It was reported that in human lung cells acetylation was lost on histone H2A, H2B, H3 and H4 whereas the ubiquitination of H2A and H2B were increased (Baccarelli & Bollati, 2009). Cigarette smoking is an example of chemical substance that alters DNA methylation. Studies have reported that smoking modulates the content of DNMT1 at a transcript and protein level (Lee & Pausova, 2013). Nicotine, which is a stimulant added in cigarettes, effects the gene expression by activating the nicotinic acetylcholine receptors.

This interaction increases the level of intracellular calcium and leads to activation of cAMP response element-binding (CREB) protein, an essential transcription factor for many genes (Lee & Pausova, 2013). Studies have also reported that cigarette smoke may alter DNA methylation through regulation of DNA-binding factors as well as hypoxia (Lee & Pausova, 2013).

Alcohol is another example of chemical substance that affects the epigenetic machinery. However, it stands out from other substances as it is widely accepted in society, consumed by "everyone" and is easily purchased (Bruna Brands, 1998). It has been postulated that exposure at different ages may perturb the chromatin function and have a direct impact on neuron plasticity, thereby being highly correlated with learning and memory function (Dobs & Ali, 2019). Pandey et al. (2015) reported that adolescents exposed to intermittent alcohol consumption had increased activity of nuclear histone deacetylase (HDAC) and expression of HDAC2, and decreased global histone acetylation that persisted in adulthood (Pandey, Sakharkar, Tang, & Zhang, 2015). Additionally, they suggested that remodelling of chromatin due to increased amygdaloid HDAC may be engaged in the process alcohol preference and dependence (Pandey et al., 2015). Their claims are straightened by Sakharkar et al. (2015) that reported an increase in hippocampal HDAC after intermittent ethanol exposure, ultimately reducing CREB, neuroplasticity and nerve growth factor (Dobs & Ali, 2019). Downregulation of CREB in hippocampus is further linked to behavioural disorders such as anxiety and depression (Dobs & Ali, 2019). Emerging evidence has indicated that consumption of alcohol may alter gene expression through epigenetic processes that further appear to be passed on to offspring. Both hypermethylation and hypomethylation due to alcohol has been observed. Sarkar et al (2016) reported that exposure to alcohol led to epigenetic changes that affected both immune and endocrine-neuronal related genes for more than three generations (Sarkar, 2016).

5.7.2 Nutritional epigenetics

The proverbial "you are what you eat" is a colloquially notion used to imply that in order to be fit one needs to have a healthy diet. In today's society it is considered common knowledge to have at least a vague idea of which foods are supposedly health-promoting and which aliments may damage or have a negative impact on our body. However, what if certain food items could affect DNA and the genetic machinery? Nutrition is one of the most studied and better understood factors in epigenetics, with several associations observed between adverse prenatal nutritional conditions, postnatal health and risk of diseases (Tiffon, 2018). Nutriepigenomics is a newly developed field that examines connections between diet and epigenetic marks that may affect gene expression.

Many studies have conveyed that nutrition consumed in early life may induce long-term changes in DNA methylation that further impact health and age-related diseases throughout life. However, some epigenetic modifications can with the passage of time become permanent. In 2007, Lumey et al. reported a link between epigenetics and diet by conducting a study about the consequences of hunger famine in The Netherlands during 1944-45. The collected data suggested that women who experienced it first-hand and became pregnant experienced epigenetic changes that were further passed on to next generation (Lumey et al., 2007). Studies of children born during the hunger famine indicated that they suffered pathologies such as diabetes, cardiovascular diseases, metabolic disorders and obesity to a greater extent. Additionally, it was reported that these children were shorter than average, and subsequently had children who also were shorter than average (Drouet, 2019; Lumey et al., 2007). It is therefore postulated that foetuses may have the ability to adapt to limited supply of nutrients (Tiffon, 2018).

Another example of nutritional epigenetics is the "Agouti model" discovered in mice. *Agouti* is a gene that contributes to the coat colour. If the *agouti* gene was subjected to little or no methylation, then it was active in all cells and the coat colour was yellow (Morgan, Sutherland, Martin, & Whitelaw, 1999). Studies reported that these mice were more susceptible to develop several diseases such as diabetes, cancer and obesity. However, if the *agouti* gene was subjected to hypermethylation then the mouse had brown fur with no health problems (Morgan et al., 1999). Between these two extremes, the methylation on the *agouti* gene could vary and further affect the gene activity accordingly. Mice with yellow fur were fed B-vitamins but were not cured. However, they yielded offspring that were carriers of the *agouti* gene but were healthy (Morgan et al., 1999). Additionally, agouti mice that were not subjected to vitamin B yielded offspring that remained sick. This was observed for several generations (Morgan et al., 1999).

5.7.3 Behavioural epigenetics

Identical twins have identical genes, yet as individuals they may be quite different with respect to personality, interests, behaviour, health and even in appearances. Behavioural epigenetics refers to study of how different signals from the surrounding environment can trigger biological changes and affect the behaviour in both human and animals (Powledge, 2011). For instance, maternal effects have shown to influence development of defensive responses to threat in various organisms (Weaver et al., 2004). In rodents, pups that were licked, which equivalates to caressing in humans, by their mother were calmer and more resistant to stress (Weaver et al., 2004). In terms of epigenetic modifications, these pups had altered histone acylation and transcription factor (NGFI-A) binding to the glucocorticoid receptor (GR) that is further involved in reducing the stress concentration. These modifications were observed in hippocampus and persisted into adulthood (Weaver et al., 2004).

Pups that were neglected by their mother were reported to be more susceptible to stress and anxiety. Several findings have established the importance of both DNA methylation and histone modification for learning and remembering.

Post-traumatic stress disorder (PTSD) is a psychiatric disorder that is triggered by a terrifying event, either by experiencing it first-hand or witnessing it. An increasing body of evidence suggests that PTSD emerges from interactions between genetic and environmental factors (Zannas, Provençal, & Binder, 2015). Experiments performed on mice showed that they displayed fear for location where they previously had been subjected to electric shocks (Powledge, 2011). The DNA methylation levels were elevated in the hippocampus, which is the region where memories are saved. Inhibition of DNMT prevents the hippocampus from being methylated and prevents formation of memories and "remembrance". Acetylation of histones was also reported (Powledge, 2011).

5.8 Profile of DNA methylation with microarray technology

DNA microarray, also known as DNA chips, is a collection of microscopic DNA spots that are attached to solid surface, normally a chip (Guigó, 2013). Each spot consists of several thousands of copies of a particular DNA sequence that corresponds to a short section of a gene, known as probes (Guigó, 2013). The microarray chip that was chosen for this study is the HumanMethylationEPIC (EPIC) BeadChip from Illumina that can interrogate more than 850 000 methylation sites in CpGs at single-nucleotide resolution across the genome (Illumina, 2015; Pidsley et al., 2016).

Prior to hybridisation to the microarray, the genomic DNA is bisulphite converted. The DNA sample is subjected to sodium bisulphite that converts unmethylated cytosines to uracil through deamination whilst leaving the methylated residues unchanged. During subsequent amplification, uracil is converted to thymine (Masser et al., 2018). The methylated cytosine residues are protected from the conversion and are read as cytosine. Probes on the microchip are designed to detect unconverted cytosine and converted thymine at a CG-site.

There are two different probes with different fluorescent colours attached to the microarray that report the methylation status. **Figure 7** displays the major steps of microarray analysis.

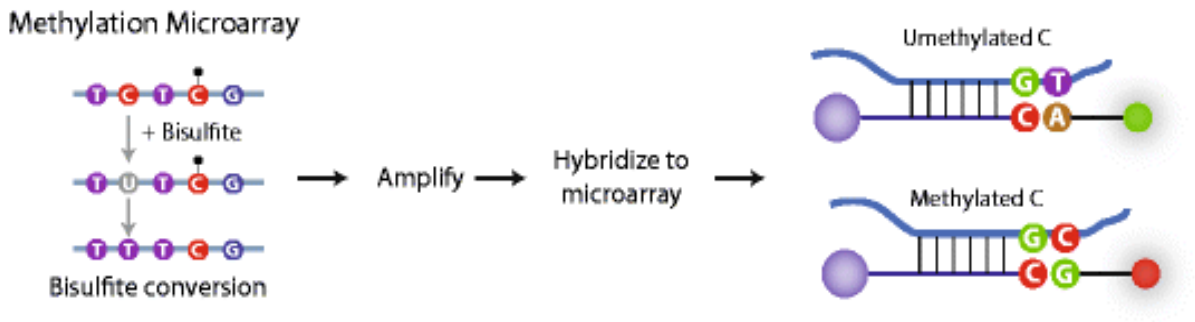


Figure 7 – Bisulphite sequencing & microarray analysis for quantification of methylated cytosines

Unmethylated cytosine are converted to uracil before subsequently copied to thymine. The probes are designed to detect both the unconverted cytosine and converted thymine residues by having different reporter colours. Figure adapted from Masser et al. (2018)

The EPIC chip uses bead technology where the probes are attached to beads. In short, beads are mixed in solution and hybridised with their complementary DNA sequence. The mixture is poured on a chip that contains millions of wells. Each bead will fit into one well, and the excess solution is washed away prior to scanning (Illumina, 2015). The chip employs a combination of two types of assay for coverage of array. Infinium I assay consists of two bead types per CpG locus; one probe that binds to the methylated states and the other to unmethylated states (Illumina, 2015). Infinium II assay uses one bead type per locus. Through single base extension, a labelled base (guanine or adenine) is added to the probe so that it is complementary to the methylated cytosine or the converted thymine from unmethylated cytosine (Illumina, 2015). Both assays emit red fluorescence for unmethylated cytosine and green for methylated cytosine.

The methylation levels are usually described by a β-value or an M-value. The β-value is calculated ratio of the methylated probe of interest divided by the overall measured intensity i.e. sum of the methylated and unmethylated probe intensity (Du et al., 2010). This method ranges from 0 to 1 and yields the methylation percentage (Du et al., 2010). The M-value method takes the log2 ratio of the measured intensity of the methylated probe versus the unmethylated probe. This method is recommended as it is more statistically valid for differential analysis of methylation levels (Du et al., 2010).

6. Aim of the thesis

Epigenetic modifications have been discovered to be dynamic and regulate gene expression without engraving changes into the genome. Emerging evidence suggests that genes take notice of the environmental factors such as nutrition, behaviour, lifestyle, climate and toxins that can have an impact on different cells, thus ultimately changing the DNA sequence. Although epigenetic modifications are considered to be temporary, several studies have discovered that these changes have affected the genomic function, yielding a permanent alteration.

A number of studies have reported and established the occurrence of epigenetic changes within the lifespan of one individual organism, but the timeframe regarding how fast these modifications occur are yet to be determined. This thesis aims to shed light on the epigenetic timeline that describes:

1. How fast DNA methylation in epigenetic modifications occurs.
2. How fast the DNA methylation in epigenetic modifications is reversed.

7. Materials

Table 1 – Chemicals used

Chemicals	Abbreviation/Formula	Supplier	Catalogue #
2-amino-2-hydroxymethyl-1,3-propanediol	Tris	Sigma-Aldrich®	77-86-1
2-mercaptoethanol	β-ME, BME	MERCK-Schuchardt	805740
Agarose (SeaKem LE agarose)		MedProbe AS	9012-36-6
Dimethyl sulfoxide	DMSO	Sigma-Aldrich®	472301
DNA Gel Loading Dye (6X)	-	BioNordika AS	60-00-4
GeneRuler 1 kb DNA ladder	-	Thermo Fisher Scientific	SM0311F
Ethanol	EtOH	Antibac AS	600068
Hydrochloric acid	HCl	Sigma-Aldrich®	H1758
N, N, N', N'-tetramethylethane-1,2-diamine	TEMED	Bio-Rad	9700106
Trypan blue solution	-	Sigma-Aldrich®	T8154

Table 2 – Cell lines

Name	Description	Supplier	Catalogue #
HepG2	Human Caucasian hepatocyte carcinoma	European Collection of Authenticated Cell Cultures (ECACC)	85011430
HL60	Human Caucasian promyelocytic leukaemia	American Type Culture Collection (ATCC)	CCL-240

Table 3 – Reagents for cell culture

Chemicals	Supplier	Catalogue #
Dulbecco's Modified Eagles' Medium (DMEM) 1X + GlutaMAX™ (containing 4.5 g/L D-Glucose and 110 mg/L Sodium Pyruvate)	Thermo Fisher Scientific	10569010
Dulbecco's Modified Eagles' Medium (containing 4.5 g/L D-Glucose), No Phenol Red	Thermo Fisher Scientific	31053028
Dulbecco's Phosphate Buffered Saline (DPBS)	Sigma-Aldrich®	D8537
Fetal Bovine Serum (FBS)	Thermo Fisher Scientific	10270106
Iscove's Modified Dulbecco's Medium (IMDM)	Sigma-Aldrich®	I6529
Iscove's Modified Dulbecco's Medium (IMDM), No Phenol Red, HEPES	Thermo Fisher Scientific	31053028
100x Penicillin Streptomycin (Pen-Strep)	Thermo Fisher Scientific	15140122
Trypsin-EDTA 1x (0.05%)	Thermo Fisher Scientific	25300054

Table 4 – Commercial kits and reagents

Name	Description	Supplier	Catalogue #
AllPrep® DNA/RNA/Protein Mini Kit (50)	Purification of DNA and RNA	Qiagen	80204
Amicon® Ultra-0.5 Centrifugal Filter Devices	Up-concentration of DNA	Merck Millipore	C82301
Universal Mycoplasma Detection Kit	Detection of myco-plasma contamination in cell cultures	American Type Culture Collection (ATCC)	30-1012K
Vybrant™ MTT Cell Viability Assay	Cell viability	Thermo Fisher Scientific	V13154

Table 5 – Equipment and software

Instrument	Description	Software	Supplier
2720 Thermal Cycler	PCR		Applied Biosystems
7900HT Sequence Detection System, ABI PRISM	Real-time PCR	ABI 7900HT Sequence Detection System Version 2.4	Thermo Fisher Scientific
Automated Cell Counter	Counting of viable cells		Millipore
Centrifuge 5415 D	Centrifugation		Sigma-Aldrich®
Centrifuge 5810 R	Centrifugation		Sigma-Aldrich®
ChemiDoc™ MP Imaging System	Imaging of agarose gel	Quantity One	Bio-Rad
Hoefer™ HE33 Mini Submarine	Agarose gel electrophoresis		Roche Life Science
Megafuge 1.0 centrifuge	Centrifugation		Heraeus instruments
MWG Sirius HT luminometer	Cell viability	Gen5™ Version 2.06	BioTek
NanoDrop® ND-1000 Spectrophotometer	Measurement of DNA and RNA concentration	Nanodrop 1000 3.8.1	Thermo Fisher Scientific
Nikon TMS Inverted Microscope	Observation of cell lines		Nikon
Steri-Cycle CO_2 Incubator	Incubation of cells		Thermo Fisher Scientific
	Results: statistical analysis and gene lists	R (RStudio)	R Foundation for Statistical Computing

7.1 Genes

The obtained candidate genes were compared against the published article "*Alcohol and DNA Methylation: An Epigenome-wide Association Study in Blood and Normal Breast Tissue*" from Wilson et al. (2019). consisting of a list of genes that they had discovered were related to consumption of alcohol (or affected by the consumption of alcohol). Information regarding the genes mentioned in this thesis has been obtained from Gene Cards – The Human Gene Database (https://www.genecards.org/).

8. Methods

8.1 Cell culture

8.1.1 – Cultivation of cells

The HepG2 cell line was grown in complete cell culture medium composed of Dulbecco's Modified Eagles' Medium (DMEM) supplemented with; 10% FBS and 1% Pen-Strep. The cells were cultured in T75 cell flasks. The HL60 cell line was grown in Iscove's Modified Dulbecco's Medium (IMDM) with same supplements as mentioned for the HepG2 cell line and were cultured in T25 cell culture flasks. Both cell lines were incubated in a humidified incubator at 37°C with 5% CO_2. All cell work was conducted under aseptic conditions in a laminar-flow hood.

8.1.2 – Passage of cells

8.1.2.1 – Sub-culturing of adherent cells

The HepG2 cell line consists of adherent cells i.e. they attach and grow on the surface of the culture vessel. For passaging, the growth medium was removed, and the cells were washed with 1x PBS, pH 7.4. PBS was removed and 0.05% 1x Trypsin-EDTA was added (till the surface was barely covered; ~2 ml for T75 cell flasks) and incubated at 37°C until the cells were detached from the vessel. The trypsin was deactivated by addition of fresh growth medium (~10 ml) that was preheated to 37°C, before the cell suspension was pipetted up and down to obtain a homogenous suspension. The cells were split in a sub-cultivation ratio of 1:2-1:6. They were split approximately every third day.

8.1.2.2 – Sub-culturing of suspension cells

The HL60 cell line consists of suspension cells i.e. they float freely in the solution (3 dimensional). They were passaged by splitting in a sub-cultivation ratio of 1:2-1:6 approximately every third day.

8.1.3 – Cell counting

The number of cells was determined by using a cell counting chamber called a haemocytometer. Prior to counting, both the glass haemocytometer and its coverslip were moistened and affixed before 10 µL of uniformly distributed cell culture was injected in the two chambers.

The space where the cells are counted have a known dept of 0.01mm and an area of $1mm^2$, giving a volume of $0.1mm^3$ (or 0.1µL). By using a microscope (10x objective), each set of the 16 squares were counted. The average from the four sets were calculated and multiplied with 10 000 to give the number of cells per mL. Cells were always passaged with a known cell concentration.

8.1.4 – Thawing frozen cells

The HepG2 was s newly purchased cell line and stored on dry ice (~ -80°C) upon arrival. The cells were thawed in a water bath at 37°C before being diluted with prewarmed growth medium. The cells were spun down by centrifugation at 1000 rpm for 7 min to remove DMSO remnants before the formed pellet was resuspended in 2 mL fresh growth medium. The cells were transferred to a T25 flask and incubated in a final volume of 10 mL at 37°C with 5% CO_2.

8.1.5 – Cryopreserving cells

An aliquot of HepG2 cells were cryopreserved to maintain a stock cell line for future use. It is recommended to make an aliquot with low passage number since cells may change phenotypically and morphologically upon passaging. The growth medium was removed, and the cells were washed with 1x PBS, pH 7.4. After centrifugation, the buffer was removed and 0.05% 1x Trypsin-EDTA was added and the cells were incubated till they detached. 3-5 mL fresh growth medium was added before the suspension was collected into a tube. The cells were centrifuged at 1000 rpm for 7 min, before being resuspended in freezing medium consisting of fresh growth medium with 10% FBS, 1% Pen-Strep and 5% DMSO. The mixture was aliquoted and transferred into CryoPure tubes and stored at -20°C for 2h before being placed at -80°C. The freezing stocks were later transferred to and stored in liquid nitrogen for long-term storage.

8.1.6 – Mycoplasma testing

Prior to the experiment, both cell lines were subjected to mycoplasma testing to detect potential contaminants in cell culture. The kit used was the Universal Mycoplasma Detection Kit from ATCC (Manassas, Virginia, US) and the testing was performed according to the manufacturer's protocol. Briefly explained, HepG2 cells were harvested by removing the cell media and washed twice with PBS, before the cells were scraped and collected into an Eppendorf tube. HL60 cells were spun down by centrifugation at 1000 rpm for 7 min, washed with PBS and centrifuged again to obtain the cell pellet. For both cell lines, 5×10^4 cells/mL were collected and centrifuged at 13 000 rpm for 3 min at 4°C.

The supernatant was discarded, and the cells were resuspended in Lysis Buffer (50 µL). The mixture was vortexed and incubated 15 min at 37°C. The samples were further heated up at 95°C for 10 min before being centrifuged at 13 000 rpm for 3 min at 4°C and transferred into new tubes. Duplicates of test samples, positive and negative controls with PCR + Primers Mix were prepared for PCR. The tubes were placed in Veriti 96 Well Thermal Cycler from Thermo Fisher Scientific with following parameters (Table 1):

Table 1 – Thermal program settings for PCR

	Temperature (°C)	**Time (s)**	**Cycles**
Denaturation	94	90	
Amplification;	94	30	
Touchdown for PCR parameters	70 → 60.5*	30	20
	72	45	
Amplification;	94	30	
Continuation of cycles at a	60	30	12
constant annealing temperature	72	45	
Final Elongation	72	240	
Hold	4	-	

* Temperature decreases 0.5°C after each cycle

A 3% agarose gel was prepared; 2 drops of ethidium bromide (EtBr) was added to 50 mL agarose-mixture. The mixture was poured into a tray with a comb and let to solidify. A DNA loading dye was added to all samples, before the PCR products and a DNA ladder were loaded on the gel and run at 100 V till the tracking dye had migrated about $^{3}/_{4}$ the length of the gel.

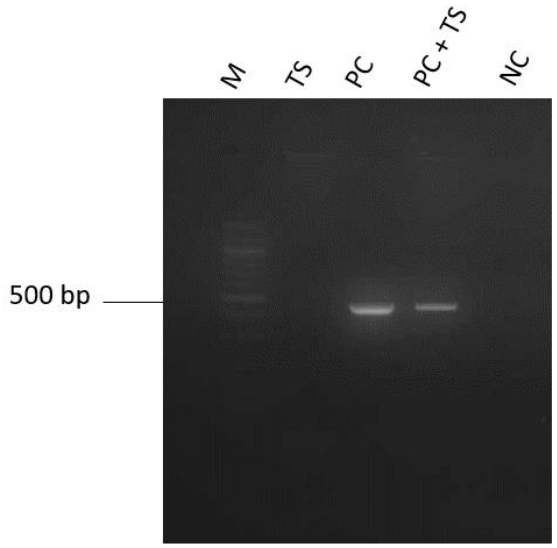


Figure 8 – Mycoplasma testing for HepG2 and HL60

Both HepG2 and HL60 cell lines were subjected to mycoplasma testing for detection of potential contaminants and yielded negative results. This figure displays the results for the HL60 cell line. Lane 1 is a 100 bp DNA ladder (M) that displays an expected band at 500 bp, lane 2 is the test sample (TS) that consists of HL60 cells, lane 3 is 2.5pg of positive control (PC); arginini chromosomal DNA that displays an expected band at ~ 500 bp, lane 4 is TS with PC and also displays an expected band at ~ 500 bp, and lane 5 is the negative control (NC)

8.2 Ethanol evaporation test and viability assays

8.2.1 – Measurement of ethanol evaporation

Since ethanol (EtOH) will evaporate over time, we did a measurement of ethanol concentration after incubation at regular cell culture conditions. The growth medium was exposed to concentrations of 20mM and 100mM ethanol and incubated for 6h and 24h at 37°C and 5% CO_2. The tests were run in duplicates. The EtOH-level in the samples was measured at 340 nm on a Cobas8000 and the analysis was performed by the Department of Laboratory Medicine and Pathology at Haukeland University Hospital, Bergen, Norway.

8.2.2 – MTT assay

Cell proliferation and viability were determined with Vybrant™ MTT Cell Viability Assay from Thermo Fisher Scientific. This colorimetric assay measures the reduction of a yellow tetrazolium component (MTT) into formazan, a purple coloured insoluble product produced by the mitochondria of viable cells, and evaluates the cell viability by measuring the growth rate of cells by displaying a linear relationship between the cellular activity and absorbance at 570 nm (Mahajan et al., 2012).

Prior to performance of the MTT assay, both HepG2 and HL60 were subjected to blank phenol-free cell medium with the same composition as DMEM and IMDM, respectively. The growth medium was switched ~24h before the assay was started.

Cells were seeded in six replicates in a 96-well plate with densities of 10 000 and 25 000 cells/well. HepG2 were plated the day before to let cells attach to the surface. For HepG2, the medium was removed and replaced with fresh medium with EtOH that had at final concentration of 20mM and 100mM (in 100 µL) whereas the HL60 cells were centrifuged, resuspended and distributed with respect to the final total volume and set well density. For both cell lines, two types of controls were included; the first type consisted of cells with EtOH-free growth medium and the other consisted of growth medium without any cells present (background control). The plates were incubated for 24h at 37°C. 10 µL MTT stock solution (12 mM) was added to each well before being incubated for 4h at 37°C. 100 µL SDS-HCl solution was added and mixed thoroughly and further incubated for 18h at 37°C. The latter incubation was recommended in a humid environment and the plates were packed with wet paper towels. The absorbance was read at 570nm and 690nm on a MWG Sirius HT luminometer from BioTek (Winooski, Vermont, US). From the obtained data, the reference values (background control) were subtracted and the cell viability was calculated by dividing the average absorption value for alcohol-exposed cells against alcohol-free control.

8.2.3 – Trypan blue staining assay

Trypan blue is an assay that measures cell viability by staining dead cells. The principle of the assay is that viable cells have an impermeable cell membrane that the dye cannot penetrate (Fang & Trewyn, 2012). However, dead cells have compromised membranes and the dye can pass through the porous barrier, enter the cytoplasm and bind to intracellular protein. (Fang & Trewyn, 2012) These interactions render the cells a blue colour and allows a direct identification and proportion of viable and dead cells.

Equal parts of cell suspension and 0.4% trypan blue solution were added and mixed to obtain 1:2 dilution before being incubated for 2 minutes at room temperature (RT). 20 µl of the suspension was added to a haemocytometer, cells were counted (as described in section 8.1.3), and the percentage of viable cells was calculated by **Formula 1:**

$$Cell\ viability = \frac{viable\ cells}{Counted\ cells\ (viable\ and\ dead)} \times 100\% \quad (1)$$

8.3 Design of experiment and treatment

Prior to conducting the experiment, the experimental parameters were determined. The choice of cell lines was driven by the fact that methylation changes in DNA due to alcohol is already well documented for both blood and liver cells. The cells were subjected to were two different ethanol concentrations; 20mM and 100mM. The first concentration (20mM) corresponds to 0.08 grams alcohol/100 mL (0.8 ‰), which in many countries is set as the standard value for legal limit for driving under the influence of alcohol whereas the highest concentration (100mM) corresponds to 4.6 grams alcohol/100 mL (4.6 ‰) that is a toxic level of alcohol in blood and can be life-threatening in most cases (Dasgupta, 2017). It was important to choose a cell concentration that would keep cells at exponential growth during the experimental time-frame, as well as leave a sufficient number of cells after treatment for further analysis. With respect to the confluency and cell density for normal and stress-free growth, most cells have a required density between 1×10^5 and 1×10^6 cells/mL. The cell density was initially tested at 2.5×10^5 and 5×10^5 cells/mL Exposure to EtOH yielded lower values of viable cells for both concentrations. Additionally, material was lost during the different experimental steps and the obtained yield was neither satisfying nor sufficient for further analysis. Therefore, the cell density was increased to 1×10^6 cells/mL, and the total volume for each well/flask was augmented to 4 mL. Results from MTT and trypan blue assays showed that the cells can tolerate the chosen EtOH concentrations and therefore remained unchanged.

For all treatments in both cell lines, the total volume for each sample was 4 mL of cell culture with a density of 1×10^6 cells/mL that were exposed to 20mM and 100mM ethanol for 24h. Since ethanol evaporates rather easily, the alcohol-based cell medium was changed every 8h. After 24h, the cells were switched back to regular growth media. The time points for harvesting the cells were 0h (control), 0.5h, 1h, 3h, 6h, 12h, 24h, 48h, 96h and 192h. The time points after ethanol exposure had technical duplicates, while the timeline of ethanol-exposed cell had one unit (flask/well) per time point.

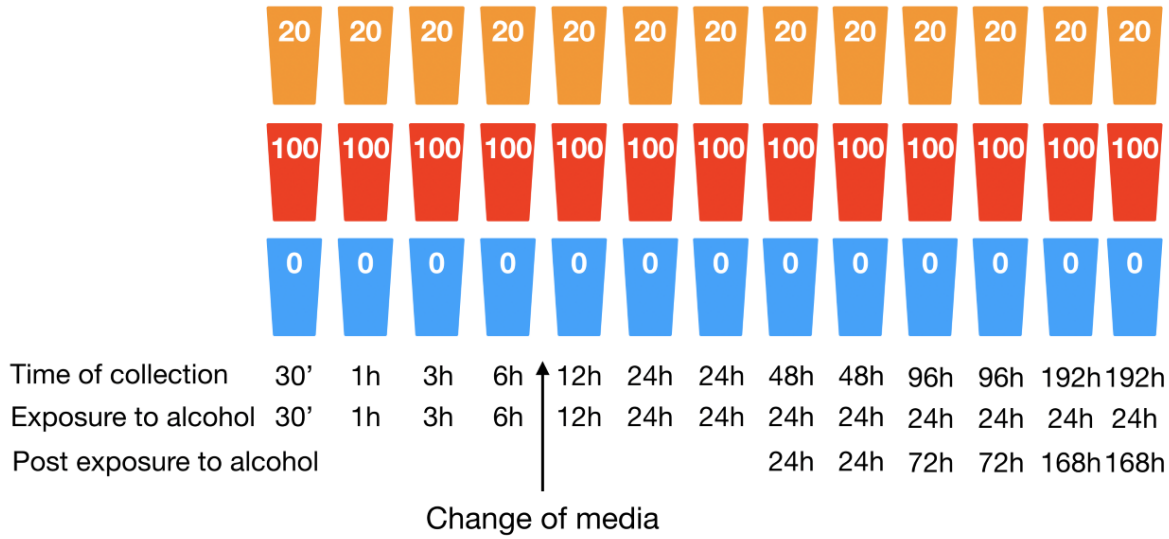


Figure 9 – Overview of samples

Schematic overview of total amount of samples for each cell line

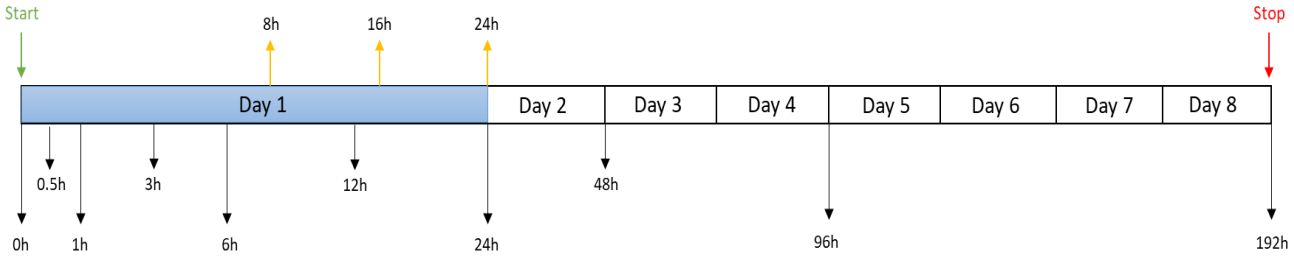


Figure 10 – Timeline of harvested cells

Schematic diagram of time points for harvesting cells (black arrows) and growth media changes (yellow arrows)

For HepG2, cells were seeded in a T25 flask (1×10^6 cells/mL) for each time point and were plated approximately 12-15h prior to EtOH exposure to let the cells adhere to the culture flask. At the starting point, the growth medium in all samples were substituted with alcohol-based cell medium, with exception of control samples. The control and treated samples were run subsequently but were at all time incubated in separate incubators. The different treatments were performed in several rounds. At each time point, the corresponding sample was retrieved.

The growth medium was removed, and the cellswere washed twice with 5 mL PBS. Approximately 1 mL PBS was added, and the cells were scraped from the surface and transferred to Eppendorf tubes and placed on ice. The samples were centrifuged at 5 000 rpm (3000 g) for 5 min at 4°C. The supernatant was discarded, and the cells were resuspended with PBS (2 mL) and recentrifuged. The supernatant was once again discarded and the cell pellet stored at -80°C.

After 96h, the remaining samples were trypsinised and split to a cell density of 1×10^6 cells/mL to keep them in exponential growth for the remaining test period.

For HL60, cells were seeded in 6-well plates with one well for each time point/treatment. Prior to exposure, 4 mL of cell culture with a density of 1×10^6 cells/mL were pipetted into each well, before being centrifuged at 1000 rpm for 5 min. The supernatant was discarded, and the cell pellet was resuspended with the alcohol-based cell medium. At each media substitution, the samples were collected in 5mL Eppendorf tubes, centrifuged and resuspended with fresh alcohol-based cell medium and pipetted back into their corresponding well. Upon harvesting, each sample was collected into two Eppendorf tubes and immediately placed on ice. The samples were centrifuged at 5 000 rpm (3000 g) for 5 min at 4°C. Supernatant was discarded and the sample was resuspended with PBS (2 mL) and recentrifuged. The cell pellet was stored at -80°C.

8.4 Purification and extraction of total RNA and genomic DNA

Prior to purification and extraction of nucleic acid, the collected samples from all cell treatments were randomised with respect to treatment and timepoint, yielding three batches for each cell line. Purification and extraction of RNA and DNA was performed using the AllPrep® DNA/RNA/Protein Mini Kit from Qiagen (Hilden, Germany) and its standard protocol for animal cells. In short, the cells were disrupted with RLT buffer with 1% 2-mercaptoethanol (β-ME) (600 µL), transferred to an AllPrep DNA spin column and centrifuged at 10 000 rpm (8000 g) for 1 min at 4°C to collect the DNA. The spin columns were stored at 4°C for later DNA isolation. One volume of 70% EtOH was added to the flow-through, and the mixture was passed through a RNeasy spin column. The column was subjected to Buffer RW1 and Buffer RPE, respectively, and was centrifuged at 10 000 rpm (8000 g) at 4°C after each step. The RNeasy spin column was placed in a new collection tube, eluted with RNase-free water (50 µL) and centrifuged at 10 000 rpm (8000 g) for 2 min at 4°C to elute RNA. The AllPrep DNA spin column was subjected to Buffer AW1 and Buffer AW2, respectively, and was centrifuged at 10 000 rpm (8000 g) at 4°C after each step.

The AllPrep DNA spin column was placed in a new collection tube, and DNA was eluted with Buffer EB (100 µL) and centrifuged at 10 000 rpm (8000 g) for 2 min at 4°C. While working, the samples were at all time stored on ice. Both RNA and DNA concentrations were measured with NanoDrop® ND-1000 Spectrophotometer from Thermo Fisher Scientific. Prior to measurement, the spectrophotometer was calibrated with RNase-free water for RNA and Buffer EB for DNA. The samples were stored at -80°C.

8.5 Up-concentration of DNA

The methylation analysis required 15 µL of each sample with a final DNA concentration of 60 ng/µL. Some of the samples that were extracted did not meet the set criteria and had to be up-concentrated. In short, the samples were pipetted into 30K membrane Amicon® Ultra-0.5 Centrifugal Filter tubes from Merck Milipore (Burlington, Massachusetts, US) and centrifuged at 14 000 g for 15 min at RT. The Amicon® Ultra Filter was separated and inversely placed into a new column, further centrifuged for 2 min at 1000 g before the "new" concentration was measured on NanoDrop®.

8.6 Agarose gel electrophoresis

Agarose gel electrophoresis was performed to verify that the DNA remained double-stranded after extraction. A 1% agarose gel (50 mL 1x TAE buffer, 0.5g agarose and 50 µg/mL EtBr) was used to separate DNA fragments and detection of nucleic acids. DNA samples were diluted with DNA loading dye from BioNordika AS (Oslo, Norway) to a final concentration of 1x. DNA ladder (5 µl) from Thermo Fisher Scientific (Waltham, Massachusetts, US) was loaded onto the gel and used as standard marker. The gel was placed into Hoefer™ HE33 Mini Submarine from Roche Life Science (Penzberg, Germany) and was covered with 1x TAE buffer. An additional drop of EtBr was added to each side of the chamber for more clear and visible gels. The electrophoresis was run at 100 V till the tracking dye had migrated about $^{3}/_{4}$ the length of the gel. The agarose gels were visualised using ChemiDoc™ MP Imaging System from Bio-Rad (Hercules, California, US).

8.7 Methylation typing

The DNA-extracted samples from HL60 and HepG2 were shipped to the Institute of Human Genetics at the University Hospital of Bonn (Germany) where they were assayed on Infinium® Human MethylationEPIC BeadChip from Illumina, covering more than 850 000 CpG sites.

The data was received in an IDAT format. The pre-processing* and analyses of the datasets were performed in R, an open-source and free program for statistical computing and graphics (Team, 2013). R consists of packages that includes codes, data, documentation and information that can be uploaded for the required analysis. For this project, the open source software for bioinformatics, Bioconductor, was downloaded and provided a variety of necessary software that are required for analysis of the microarray and comprehension of the genomic data (W. Huber, 2015).

* *The pre-processing was performed by a bioinformatician in our research group.*

Table 7 is listing all the packages that were used for statistical analyses.

Table 7 – Packages in R that were used for statistical genomic analyses

Package	Description	Authors
Limma – Linear Models for Microarray (version 3.9)	Data analysis, linear models and differential expression for microarray data	Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK
Splines	Functions for working with regression splines	
qqman (version 0	Creates qq- and Manhattan plots (for GWAS)	Turner S (2015)
IlluminaHumanMethylation EPICanno.ilm10b2.hg19 (version 0.6.0)	Provides genetic elements and maps probes to genes	Hansen K.D (2016)

8.7.1 Principal component analysis (PCA)

To evaluate potential covariates or unwanted variation on methylation, principal component analysis (PCA) was run. This technique takes multi-variation into account and tries to identify the linear components of a set of given variables (Unwin, 2013). For both cell lines the covariates array, batch effect, concentration, slide, time points and treatment were tested for. "Array" represents the plate that the samples were pipetted on, "Slide" is the specific lane on the array, "Batch Effect" is representing the laboratory conditions that may have occurred during extraction and isolation of DNA, "Time Points" represent the distribution of samples with respect to the time points at which they were collected and "Treatment" represents the distribution of samples with respect to their treated concentration. The tests performed were on treatment against control, and the obtained results were further used to make quantile-quantile (Q-Q plots) for better visualisation of the effect of correcting for the different covariates. Based on PCA- and Q-Q plots, it was decided if the cell lines needed corrections, and which variables they had to be corrected for.

8.7.2 Factor analysis

Factor analysis is a statistical method that reduces the correlation between each pair of variables (called R-matrix) down to its underlying dimensions by observing which variables appear to cluster together, thereby describing variability among observed, correlated variables with respect to potentially lower number of unobserved and/or unexplainable variables (called factors) (Unwin, 2013). Based on PCA- and Q-Q plots from the previous section, factor analysis was performed for further investigation of variable relationships.

Following models were made:

Model	Description
$y = 1 + model$	Baseline model
$y = 1 + model + error_{(i)}$	Model with included errors/variables that must be accounted for

Where y is the methylation values, 1 is the set intercept, model is the treated samples divided by control samples and error is the subset of variables or components that the dataset must be corrected. The matrices were made from normalised methylation expression where the probes are in the rows whereas the necessary information of the samples are in the columns.

8.7.3 Spline analysis

In general, spline is a function that is piecewise made of polynomials i.e. adding lines or curve together to make a continuous curve. Spline modelling has become an established tool for analysis of statistical regression and are regularly used in clinical research for building explanatory models (Perperoglou, Sauerbrei, Abrahamowicz, & Schmid, 2019). In modern biostatistics, several methodological developments use spline to model smooth functions of interest such as modelling time series, modelling of time-dependent effects in survival analysis and frequency distributions to mention some. Prior to the spline analysis, the samples were sorted based on collected time points, having a chronological order. For the post-exposed sample that consisted of two replicates, the samples where either merged or one sample was randomly removed in order to have one sample per time point.

8.7.4 Profile searches

A profile search consists of a profile that is made of numbers that yields different types of curves (s-curve, linear, exponential, polynomial etc) and aims to mimic different types of behaviour observed for genes. With respect to the number of samples collected for each time point, profiles with the same number of constants were plotted. Since one did not expect changes in the control samples, they had a fixed value of 1.

Once a profile was created it was run against the matrix that consisted of normalised methylation values, and CpGs with a response curve correlated with the search profile were detected. **Figure 6.2** displays one of the created profiles for this analysis

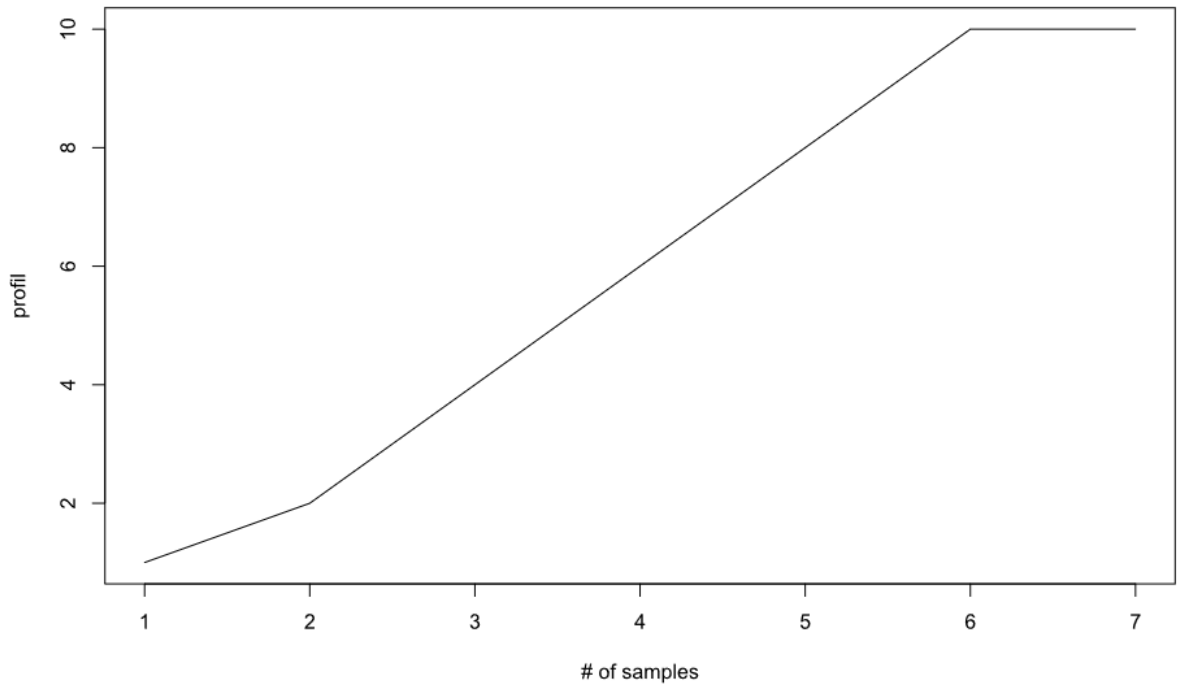


Figure 11 – Profile search
This profile is created based on a gradual change with eventually a lasting effect and the chronological order corresponds to time points.

8.7.5 Linear models for microarray

Linear models for microarray (limma) is a package in R that is commonly used for analysis of gene expression data by finding differences between two or several groups of samples (M. E. Ritchie, 2015). Limma is initiated by fitting a linear model with multiple linear regression and takes into account the potential variable(s) of interest (covariates) that need to be corrected to run the analysis, thereby improving the estimation of variance (M. E. Ritchie, 2015). As previously mentioned, limma can identify features that are different between two or several groups by using a contrast matrix, thereby highlighting the differences. Additionally, differentially methylated probes that were identified by limma were further mapped to probes by using the EPIC annotation library; a register that consists of all genes that are associated with or correspond to the probes found on the EPIC array. In supplementary to differentially methylated regions these probes also consist of hypersensitive, enhancer and non-coding sites. The output retrieved from limma analysis provides information on probes that map to corresponding genes.

9. Results

Previously conducted studies have established the occurrence of epigenetic changes within the lifespan of one individual organism, but the timeframe regarding how fast these modifications occur are yet to be determined. The focus of this master project has been to establish a timeline that describes how fast DNA methylation in epigenetic modifications occurs, and how fast it is reversed. Since epigenetic changes due to exposure of alcohol are well documented notably on DNA methylation level, we chose exposure to alcohol as a model for this experiment.

9.1 Determination of design & experimental parameters

9.1.1 Evaporation of ethanol (EtOH)

Ethanol is a substance that evaporates quickly. A study from Eysseric et al. (1997) tried to maintain a constant concentration of ethanol in a long-term cell culture, and reported that more than 70% had evaporated within 72 hours (Eysseric et al., 1997). The chosen concentrations for ethanol were 20mM and 100mM, corresponding to a blood alcohol content of 0.8‰ and 4.6‰, respectively. **Table 8** displays the obtained ethanol concentration values per mille for control, 20mM and 100mM after 6h and 24h, respectively. After 6h, the measured concentration for EtOH was 0.92‰ in control, 0.66‰ for the 20mM-treated samples and 2.65‰ for 100mM-treated samples. After 24h, the measured concentration for EtOH was 0.72‰, 0.53‰ and 0.96‰, respectively. Based on the yielded results, it was confirmed that EtOH evaporates easily. Additionally, it was discovered that both control samples contained EtOH, suggesting that some of the solution condenses into neighbouring wells.

Based on the obtained results, the experiment was set up. Samples from different treatments were not plated together nor placed in the same incubator. In fact, each treatment was run separately. The loss of EtOH due to evaporation was overcome by switching the alcohol-based growth medium every 8h. The cell medium for control samples was also changed to maintain an identical procedure of all treatments.

***Table 8 – Concentration values of ethanol present in control, 20mM and 100mM after incubation of 6h and 24h at* 37°C**

Incubation	Sample	1 (‰)	2 (‰)	Average (‰)
After 6h	Control	0.88	0.95	0.92
	20mM	0.64	0.68	0.66
	100mM	2.59	2.71	2.65
After 24h	Control	0.71	0.72	0.72
	20mM	0.51	0.54	0.53
	100mM	0.94	0.97	0.96

9.1.2 Cell viability and proliferation

The cell viability and proliferation were determined with MTT Assay and trypan blue to verify if the chosen cell lines could withstand the chosen EtOH concentrations. The cells were treated with the highest ethanol concentration for 24h. The viability of cells that were grown without the presence of EtOH were defined as 100%. **Figure 12** displays the results from MTT assay for HepG2 cells that were plated with cell densities of 10 000 and 25 000 cells per well, having a percentage viability of 92% and 103%, respectively, whereas **Figure 13** displays the results for HL60 cells with estimated cell densities of 100 000 and 50 000 cells per well and a percentage viability of 67% and 35%, respectively.

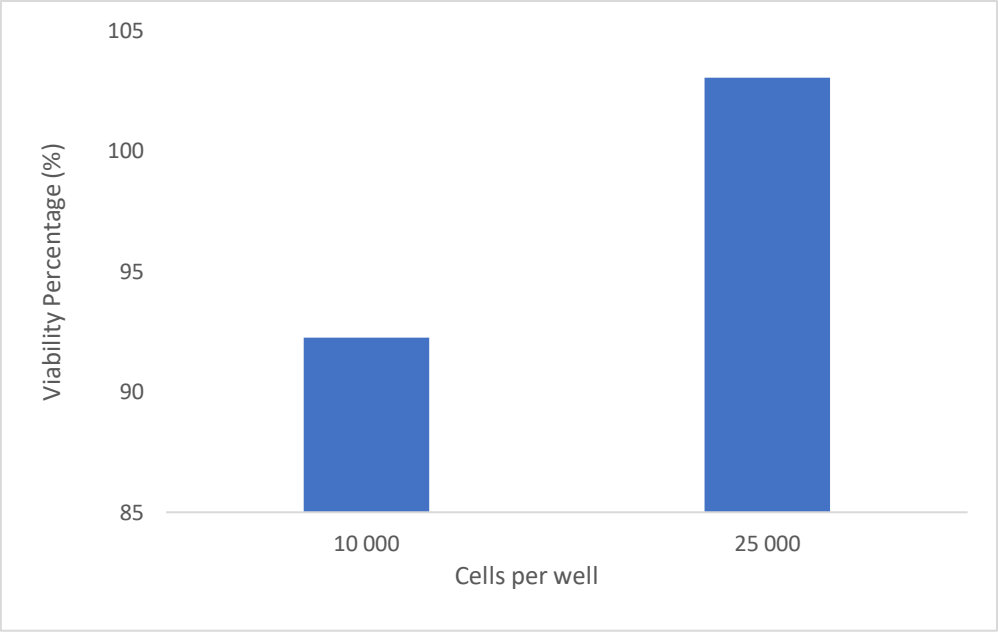


Figure 12 – Viability percentage of HepG2 cells incubated with 100mM EtOH for 24h

HepG2 cells were incubated with cell densities 10 000 and 25 000 cells per well (96-well plate) and exposed to ethanol (EtOH) for 24h at 37°C. The viability of cells grown in the absence of EtOH was set as 100%, and the viability percentage for 10 000 and 25 000 cells per well were calculated to be 92% and 103%, respectively.

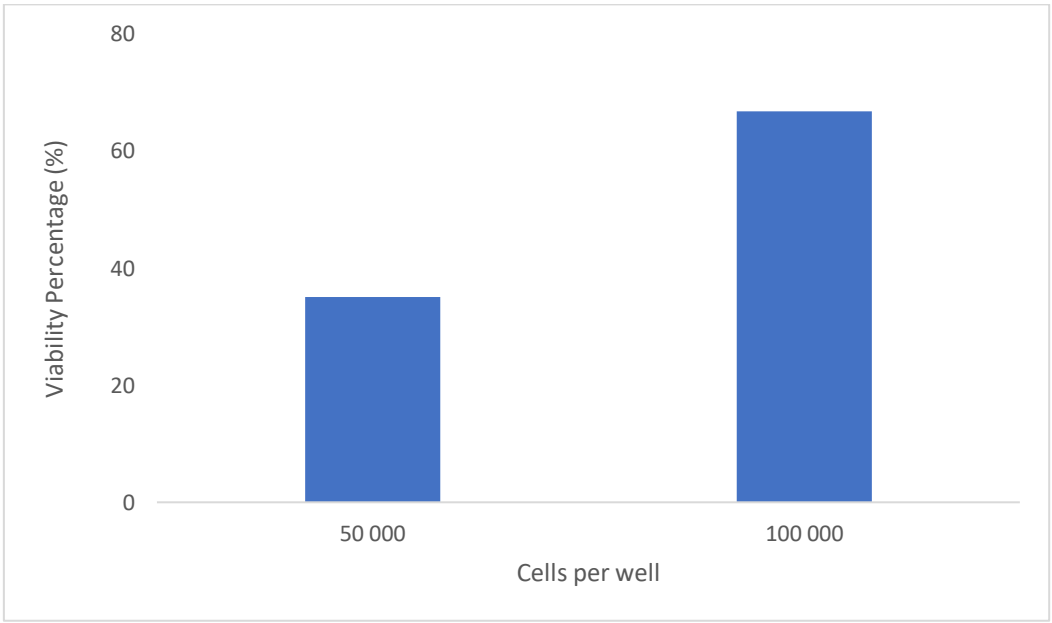


Figure 13 – Viability percentage of HL60 cells incubated with 100mM EtOH for 24h

HL60 cells were incubated with cell densities of 1 x 10^6 and 5 x 10^5cells/mL and exposed to ethanol (EtOH) for 24h at 37°C. The viability of cells grown in the absence of EtOH was set as 100%, and the viability percentage for 100 000 and 50 000 cells per well were calculated to be 67% and 35%, respectively.

The cell viability was also determined with trypan blue solution. As for MTT assay, the cells were treated with the highest ethanol concentration for 24h before being subjected to the dye. The cell viability was ~ 100% cell for HepG2 cells with cell densities 10 000 and 25 000 cells per well. For HL60, the cells that had a cell density of 1×10^6 cells/mL had a cell viability of ~ 88% for those that were EtOH-exposed whereas 5×10^5 cells/mL had a cell viability of ~ 132% respectively. The cell viability for HL60 with 5×10^5 cells/mL was increased due to loss of cells in the control samples during the experiment.

The cells were subjected to 20mM and 100mM, and the samples were run according to the set timeline. Once all the samples from both cell lines were collected, the samples were randomised, and subjected to DNA and RNA extraction. Samples from each cell line were divided into three batches. All samples were either diluted or up-concentrated, yielding a final concentration of 60 ng/µL. The samples were once again randomised with respect to time point at which they were collected, the exposed treatment and which batch they belonged to. The samples were sent to the genomics facility in Bonn (Germany) for whole genome DNA methylation typing. The results were received 1,5 months later.

9.2 Processing of data and performance of methylation analysis

9.2.1 Principal component analysis (PCA)

Potential covariates or unwanted variation on methylation was evaluated with principal component analysis (PCA) and Q-Q plots. PCA is used to find potential patterns and/or regularities by itself, reducing the dimensionality of high-dimensional data sets while preserving the original structure. Each treatment was assigned a colour to observe any type of trend or pattern. **Figure 14** is illustrating one of the PCA-plots where "Treatment" was chosen as the covariate and displays the distribution of samples from the HepG2 cell lines with respect to the concentration they were treated with and the time point at which they were collected. It is observed that most of the samples are clustered in the middle with some few outliers. The data upmost to the left is the control sample for treatment 20mM, whereas the two samples to the right are both from the 100mM treatment but are representing different time points (192h and 96h, respectively). Samples that consist of "Rep2" are those that had duplicates

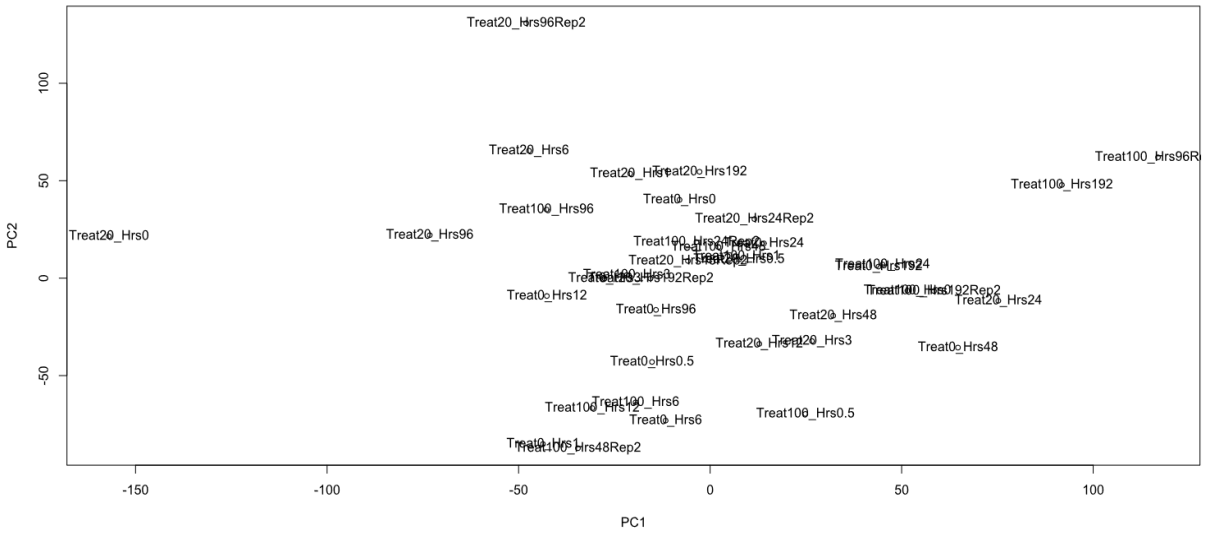


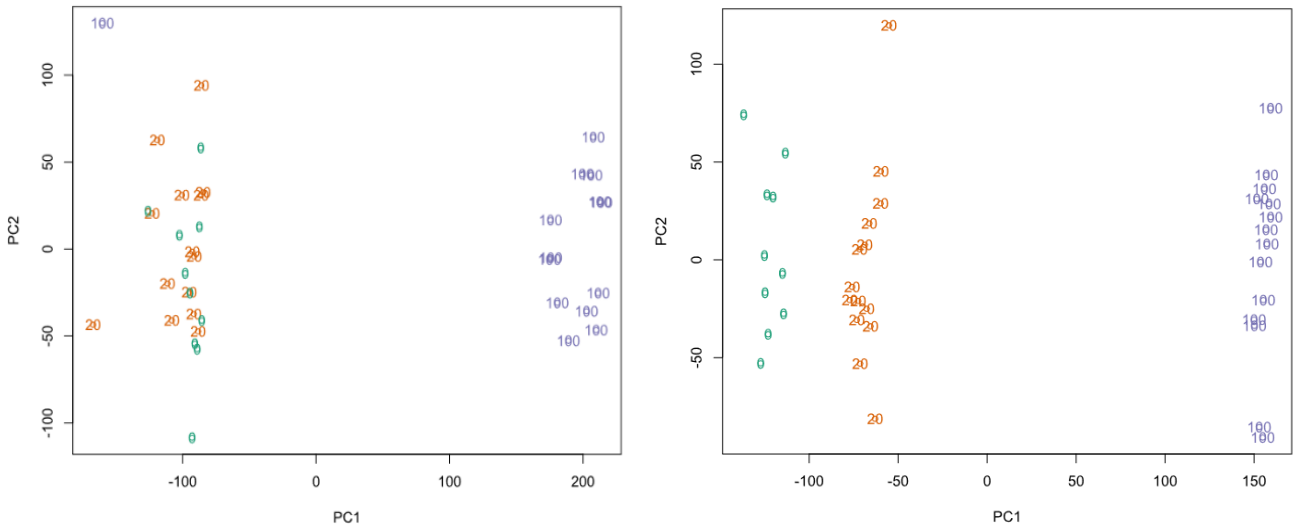
Figure 14 – Principal Component Analysis (PCA) for HepG2

This PCA-plot displays the distribution of samples from the HepG2 cell line with respect to their treated concentration and the time point at which they were collected. One can observe that the samples are mostly clustered in the middle with some few outliers.

In **Figure 15**, the samples appear to be clustered in two groups. Further inspection of the sample distribution suggested that they were according to the time point at which the experiments took place. Samples that are located to the right are mostly samples that were subjected to 100mM EtOH-treatment and performed in the first round whereas the remaining group is representing the control and 20mM EtOH-treated samples that were performed later/in the second round. There was approximately a time span of three weeks between the two rounds. It may be postulated that the observed differences are due to the concentrations the cell line was subjected to rather than the time points at which each sample was collected. Within the clustering of the control and 20mM EtOH-treaded samples, it can be observed that the control samples mostly have a higher PCA1-value and are located towards right whereas the 20mM EtOH-treaded samples have a lower value and are located towards left. This may suggest that the alcohol concentration is the dominating factor that influences the results. However, one sample for the 100mM EtOH-treated (48h replicate 2) HL60 cell line was missing and rerun subsequently with control and 20mM EtOH-treaded samples. As observed from **Figure 15 (a),** that particular sample is an outlier in the upper-left corner with a substantial low PCA1-value compared to the other 100mM EtOH-treated samples, thereby implying that long-term culturing and higher passage of cells is the dominating factor that affects the methylation values.

Based on these discoveries, it was decided that HL60 samples needed to be corrected for the time point at which the experiment took place. A new covariate “Easter” was created and categorised the samples according to the time point they were collected i.e. before and after the Easter Holidays.

Figure 15 (a) more specifically displays this distribution where the samples to the right are mostly 100mM treated cells that where treated first, and control and 20mM EtOH-treated samples that were performed later. In **Figure 15 (b),** the dataset has been corrected with respect to the round in which the samples were collected in. Once the dataset is corrected for "Easter", the outlier from **Figure 15 (a)** is grouped together with the 100mM treated samples in **Figure 15 (b).** Furthermore, the collected samples are now clustered into three groups, thereby indicating that the methylation values are affected/influenced by concentration.



(a) *Not corrected* ***(b)*** *Corrected*

Figure 15 – Principal Component Analysis (PCA) for HL60 non-corrected versus corrected

This PCA-plot displays the distribution of samples from the HL60 cell line with respect to exposed treatment. The clustering of samples into two groups are due to the time point at which the experiment took place and represents a separation/time span of three weeks difference.

Figure 16 and **Figure 17** display the distribution of samples for the original and corrected dataset, respectively. The samples are categorised as "Under Treatment" and "After treatment", representing samples that were collected while being subjected to ethanol treatment and those that were recovering, respectively. In **Figure 16,** we can as expected observe that the samples are divided into two cluster. However, within each cluster, there is a prominent contrast between "Under Treatment" and "After Treatment". **Figure 17** displays a similar effect where most of the "Under Treatment" samples have higher PCA1 and PCA2 values compared to "After Treatment".

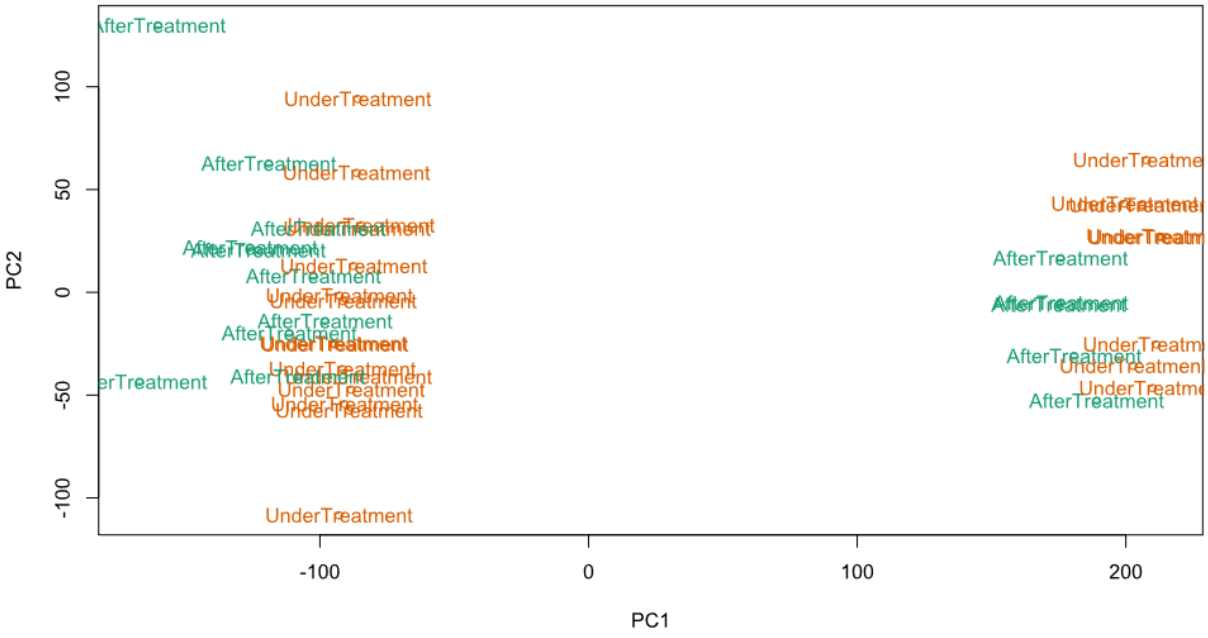


Figure 16 – Principal Component Analysis (PCA) for non-corrected HL60 cells based on "Treatment"

This PCA-plot displays the distribution of samples from the HL60 cell line with respect to their treatment status i.e. if they are under the exposure of EtOH or if they are recovering from the treatment (post-treated cells). The clustering of samples into two groups are due to the time point at which the experiment took place and represents a separation/time span of three weeks difference.

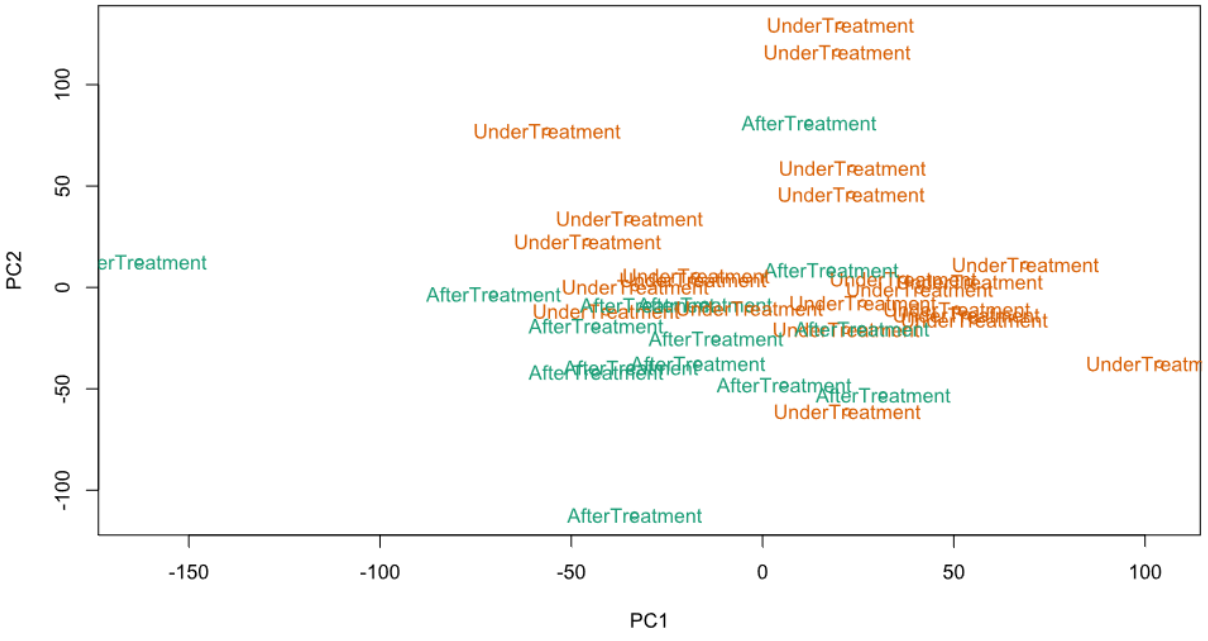
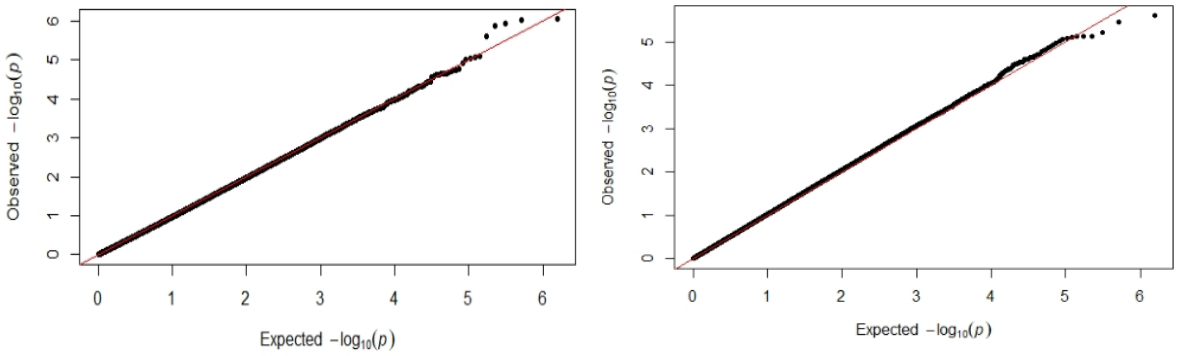


Figure 17 – Principal Component Analysis (PCA) for corrected HL60 cells based on "Treatment"

This PCA-plot displays the distribution of samples for the corrected HL60 cell line with respect to their treatment status i.e. if they are under the exposure of EtOH or if they are recovering from the treatment (post-treated cells).

9.2.2 Quantile-quantile (Q-Q) plot

After PCA, Q-Q plots were made to inspect if the distribution was as expected. For both cell lines and treated concentrations, the subgroups were plotted. The plots were first made without taking in account the presence of variance in the samples, before one of the covariates was included. The analysis was followed by creating plots where two or several covariates were included. **Figure 18** displays Q-Q plots for HepG2 and HL60.



(a) *HepG2* ***(b)*** *HL60*

Figure 18 – Q-Q plots for 100mM AfterTreatment (AT) against UnderTreatment 1 (UT1) for HepG2 (a) and HL60 (b) cells with correction of covariates "Array", "Slide" and "Batch effect"

These Q-Q plots display the distribution of Aftertreatment (AT) against UnderTreatment1 (UT1) for 100mM treatment for HepG2 and HL60, respectively. In both cases a linear trend till an expected value of 5 can be observed. From that point samples vacate from having a linear relationship between the predicted and observed value. The HepG2 Q-Q plot (a) consists of data points that vacate positively from the linear trend and negatively for HL60 (b).

With respect to the findings from the PCA- and Q-Q plots, it was decided that HepG2 should not undergo any corrections on the data set as the Q-Q plot for HepG2 without any corrections displayed the best linear relationship. Additionally, the correlation coefficient for the covariates were low, suggesting that there are no relationships between the variables. The HL60 samples needed to be corrected for the time point at which samples for a certain treatment was collected. Subsequently, the HL60 samples needed to be corrected for the covariate "Slide" as the correlation coefficient was 0.988 with PC1 and indicated a positive correlation between the variables. The remaining variables for HL60 had correlation values ~ 0.

9.2.3 Spline analysis

In spline analysis, one sample is enough per time point. For time points with replicated samples, one of them was randomly chosen. The analysis was first restricted to samples that were under treatment i.e 24h, then run for the whole timeline. However, nonsignificant results were obtained for both of them.

The analysis was repeated, and for samples with replicates the other copy was chosen, still yielding nonsignificant hits, but with different genes and statistical values. Some genes from the previous analysis were present, but the order was different with respect to their statistical values. When comparing the graphs that displayed the methylation values over time for the top candidate genes, it was discovered that there was substantial variation between some of replicates even though they were performed simultaneously. Replicated samples were therefore added together and the average was calculated. Merging of the replicated samples was thought useful to reduce some of the observed variability between the samples. The tests consisting of the average value for the replicated samples were run and yielded better results as well as better methylation plots. Additionally, when running the spline analyses, it was observed that the merged samples yielded slightly different p-values and order of candidate genes each time the procedure was repeated, thereby suggesting that there might be underlying variation in the data set. Nonetheless, the hits were still nonsignificant. Therefore, the spline analysis was dismissed as both merged and the randomly selected samples had too much variation between them. The analysis appears to be sensitive to outliers since a large impact at one time point was enough to rank a top of the hit list. **Figure 19** displays spline analysis for HL60 for the whole timeline where one of the samples was randomly selected, and the mentioned trend was observed.

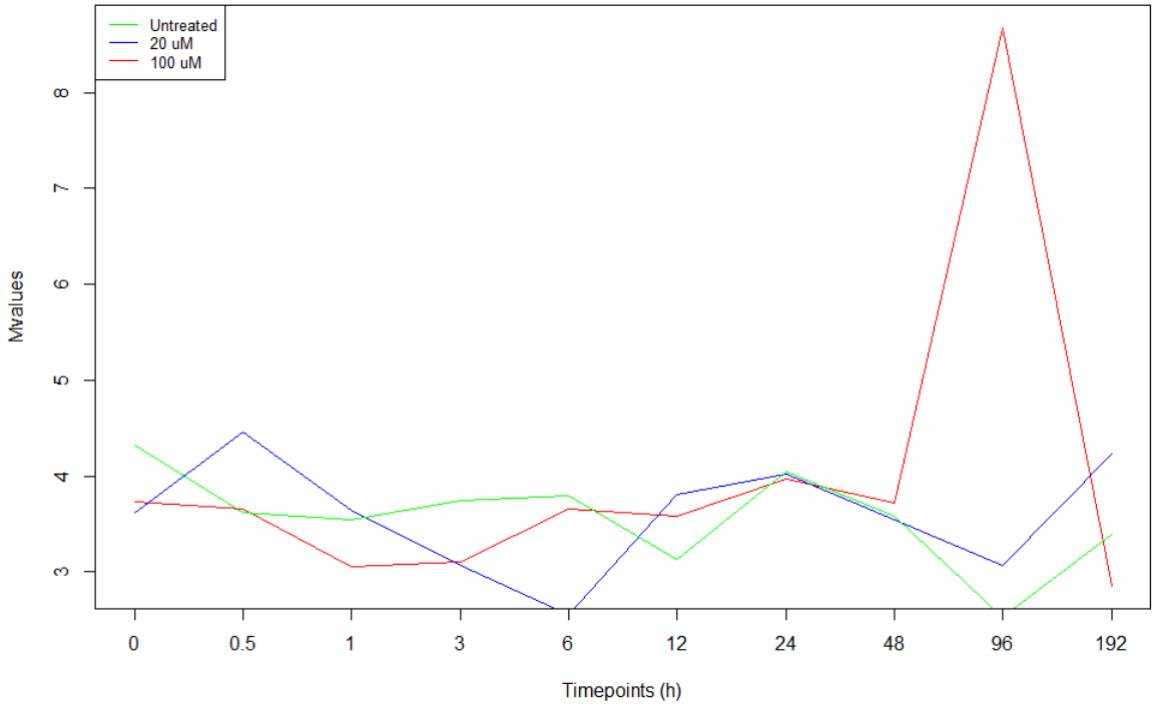


Figure 19 – Spline analysis for HL60 for the whole timeline

This function is illustrating the methylation over time a potential gene (FEZ2). However, it was observed that all spline analyses consisted of one or several outliers that would heavily affect the obtained results.

9.2.4 Factor analysis with linear models for microarray (limma)

Based on PCA- and Q-Q plots, factor analysis was performed for further investigation of variable relationships. For this analysis, two model matrices were created; the first matrix represents the null model and consists only of the covariate whereas the remaining model consists of the covariate as well as the variable(s) of interest. "Treatment" was created and consisted of treated samples divided by control samples with respect to the treatment they were subjected to and the time point at which they were collected. HepG2 did not undergo any corrections and needed only one model for analysis. The created models for factorial analysis are listed in **Table 9**.

Table 9 – Model matrices that were made for factor analysis of HepG2 and HL60 cell lines:

Cell line	Model	Description
HepG2	mod = model.matrix(~1 + Treatment)	Baseline model
HL60	mod = model.matrix(~1 + Treatment)	Baseline model
	mod = model.matrix(~1 + Treatment + factor(Slide) + Easter)	With covariates

Initially, one sample was intended per time point as we had plan to use the spline method for time-course analysis. The number of replicates were compensated on expense of having several time points at which the cells were collected to create an epigenetic timeline. Ten time points were set; 7 samples were to be collected during the exposure and 3 would be collected as the cells recovered. As there are no current studies that display the time series of samples after ended treatment, replicates were added for the post-treated time points, including 24h. The control treatment consisted of one sample for every time point. Factor analysis was performed for the whole timeline, and the contrast for treatment 20mM and 100mM against control were run, respectively.

To determine how fast the epigenetic changes occurred, the analysis was first restricted to the exposure period i.e. 24h. The changes were analysed with the different models and did not yield any significant hits. The time points were then clustered into subgroups; 1) 0.5h, 1h and 3h and 2) 6h, 12h and 24h, 1 and 2 are under treatment, 3) 48h, 96h, and 192h is post-exposed. By clustering them into subgroups, the different times acted as replicates for each other. The different combinations of subgroups were tried but yielded few significant hits. Additionally, the visualisation of methylation over time for the top gene candidates displayed variations between the time points. In each combination, there was at least one or two time points that were outliers, and their position affected the obtained results. Factorial analysis was further performed on the early collected samples against those that were collected at the end of exposure i.e. subgroup 1 versus subgroup 2. This combination did not yield any significant hits.

Finally, since there was two replicates for the 24h timepoint and since there was only a difference of half an hour between the first two collected samples, another factor analysis was performed where the first two time points 0.5h and 1h were merged and run against the last exposed time point. Sixty probes were significant ($P < 0.05$), but neither of obtained genes were alcohol related.

Table 10 displays the different values and genes that were retrieved for top10 hit for 20mM HL60 where the first two collected time points were clustered together (called Early) and run against the replicates for the last exposed time point (called Late). "Probes" are referring to the particular probe that was matched, "Difference" is the measured value between the compared subgroups, "AveExpr" is the average $\log_2$-expression for the probe of interest over all arrays, the moderated F-statistics value "F-value" that tells if the distribution of subgroups are significantly different, p-value and adjusted p-value indicates how significant the hits are and "Gene" displays the name of the gene that is related to that specific probe.

Table 10 – Top 10 probes and their average logarithmic expression, F-statistics, p-values and corresponding genes for Late (24h) vs Early (0.5h & 1h) HL60 cells that were treated with 20mM EtOH

Probes	Difference (Late vs Early)	Average Expression (AveExpr)	F-value (F)	P-Value (P.Value)	Adjusted P-value (adj.P.Value)	Gene
cg06169065	- 6.27	3.32	322.74	3.31E-11	1.1E-05	-
cg23264443	- 7.48	4.66	318.65	3.62E-11	1.41E-05	*PLIN3*
cg23623270	3.30	- 3.05	276.13	9.62E-11	2.1E-05	*SLC25A39*
cg27134827	4.09	- 2.32	269.68	1.13E-10	2.21E-05	*PPT2*
cg10006408	6.57	- 3.40	228.08	3.52E-10	5.49E-05	*WDR89*
cg08166863	- 10.18	7.17	211.27	5.89E-10	7.66E-05	*GNAS, GNASAS*
cg06979118	4.34	- 2.56	199.13	8.76E-10	9.77E-05	*SHANK2*
cg14270857	3.00	- 1.80	166.80	2.85E-09	0.000262	*HPSE2*
cg21077245	- 4.61	3.95	165.33	3.02E-09	0.000262	*ZFYVE21*
cg06551854	3.84	- 3.41	127.47	1.65E-08	0.001291	-

While performing the statistical analysis for the HepG2 cell line, extreme variations between the different time points were observed in both spline and factor analysis. Furthermore, we discovered that the obtained hits for HepG2 consisted of a broad spectrum of genes where most of them stress related genes. Due to background noises and observed variations, we decided to focus on HL60 cell line.

9.2.5 Profile searches

To determine the change of DNA methylation in epigenetic modifications, profiles that mimicked the different types of possible changes in DNA methylation were created. To determine how fast the epigenetic changes occurred, profile searches for samples under the presence of EtOH were chosen, restricting the timeline to the exposure (24h).

Briefly they consist of profile 1: gradual change (increase or decrease) in methylation with lasting effect, profile 2: fast change with lasting effect and profile 3: momentarily change with lasting effect. **Figure 20-22** are illustrating the different profiles for the EtOH-exposed cells (24h), and for the whole timeline.

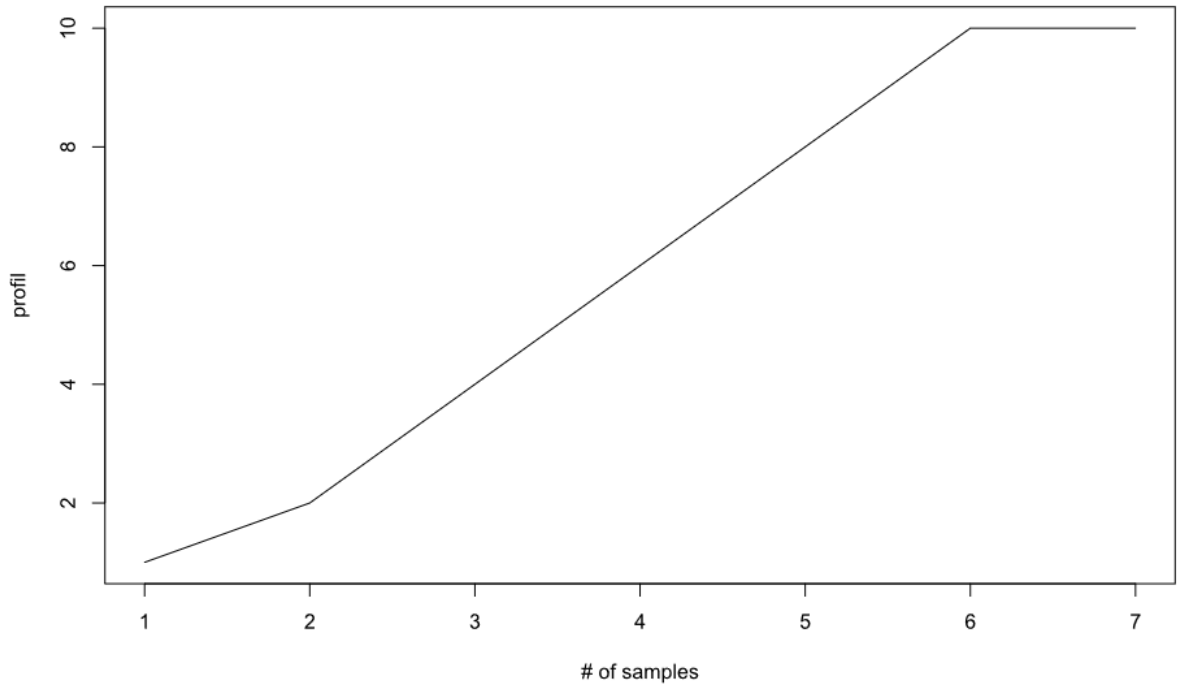


Figure 20 – Profile search #1 for 24h

Profile search #1 was created based on a gradual change with eventually a lasting effect. The chronological order corresponds to the time points 0h, 0.5h, 1h, 3h, 6h, 12h and 24h, With respect to correlation the change could be an increase or a decrease in methylation values.

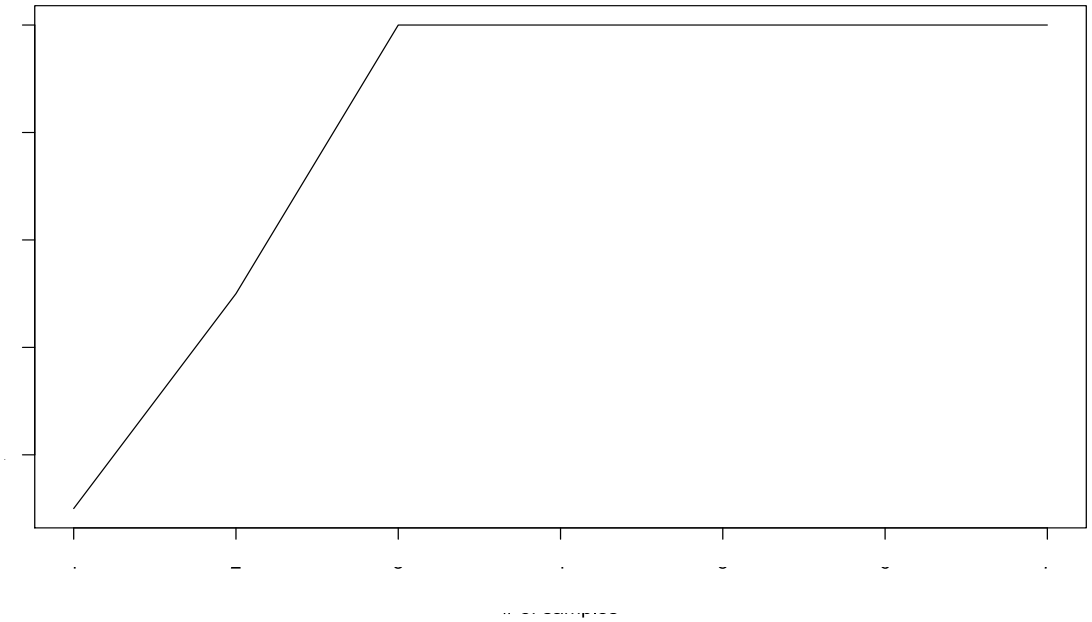


Figure 21 – Profile search #2 for 24h

Profile search #2 was created based on fast change with eventually a lasting effect. The chronological order corresponds to the time points 0h, 0.5h, 1h, 3h, 6h, 12h and 24h, With respect to correlation the change could be an increase or a decrease in methylation values.

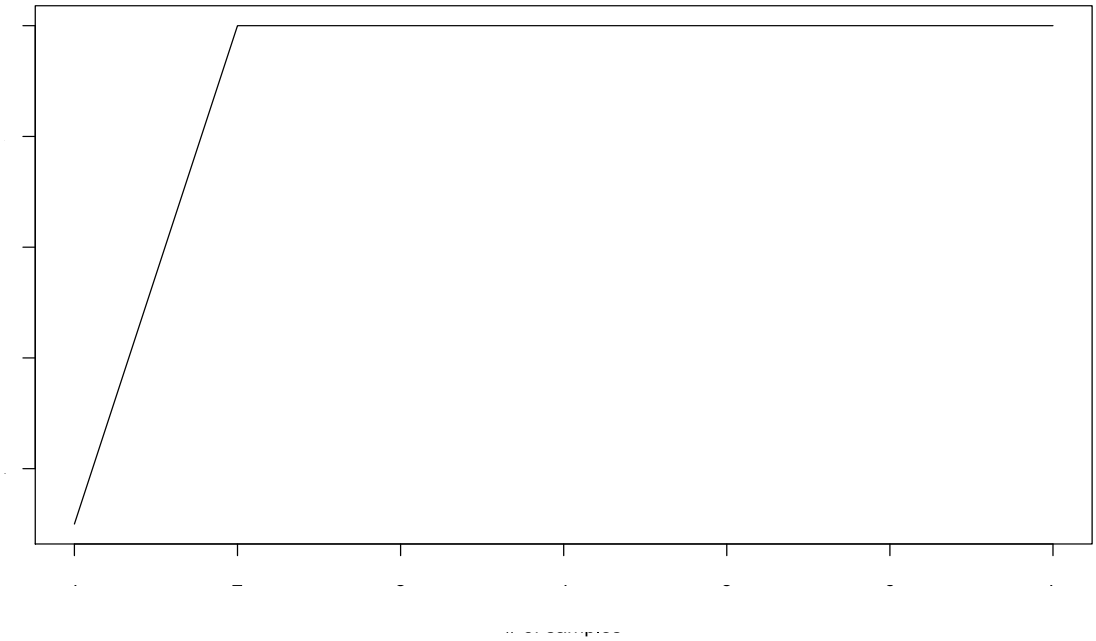


Figure 22 – Profile search #3 for 24h

Profile search #3 was created based on fast change with eventually a lasting effect. The chronological order corresponds to the time points 0h, 0.5h, 1h, 3h, 6h, 12h and 24h, With respect to correlation the change could be an increase or a decrease in methylation values.

Each profile was correlated with the methylation values for each probe, before the p-value and false discovery rate (FDR) were calculated. FDR tries to identify as many significant comparisons as while maintaining a low rate for false positive (Type I error) for multiple variable testing. P-value is the calculated probability of finding the observed results when the null hypothesis (H_0) is true, and is used to determine the significance of the obtained results. The probes were sorted with respect to most significant hits for both FDR and p-value. The annotation package for Illumina's EPIC methylation array was run on probes and yielded genes that consisted of that particular sequence. With respect to the FDR and adjusted p-values one obtained for the profile searches, the most significant hits of obtained candidate genes were compared against published article "*Alcohol and DNA Methylation: An Epigenome-wide Association Study in Blood and Normal Breast Tissue*" from Wilson et al. (2019), consisting of a list of genes that they had discovered were related to consumption of alcohol (or affected by the consumption of alcohol). The significant p-value for overlap with Wilson et al (2019) was 0.05 ($P < 0.05$).

Profile 1 represented a gradual change in methylation and yielded an FDR-value of ~1. The overlap p-value with Wilson et al. was 0.102, yielding candidate genes *TRAPPC9* and *ERRFI1* with correlation values -0.76 and 0.71, respectively. The second profile search represented a fast change in methylation with lasting effect, also yielding an FDR-value of ~1. The p-value for overlap was calculated to be 0.0023 followed by candidate genes *ACTN1, ERRFI1, ETV6* and *RNU-19P.* It was discovered that the last two genes share the same probe and have the ability to form a "partner gene" (al., 2018). The last profile represented momentarily change in methylation with lasting effect. However, this profile did not yield any hits. With respect to the overlap value as well as candidate genes, the second profile was chosen for further analysis.

Table 11 displays the calculated FDR, p-value for overlap between the significant genes and Wilson et al. (2019) for the different profile searches, and **Figures 23-25** display the methylation over time for 24h for the candidate genes.

Table 11 – Obtained false discovery rate (FDR), p-value for overlap and overlap genes for the different profile searches for 24h:

Profile Search (#)	FDR-value (FDR)	P-Value (P.Value)	Gene	Correlation
1	~ 1	0.102	TRAPPC9	- 0.76
			ERRFI1	0.72
2	~ 1	0.002	ACTN1	- 0.73
			EFFFI1	0.73
			ETV6	0.69
			RNU6-19P	0.69
3	~ 1	-	-	-

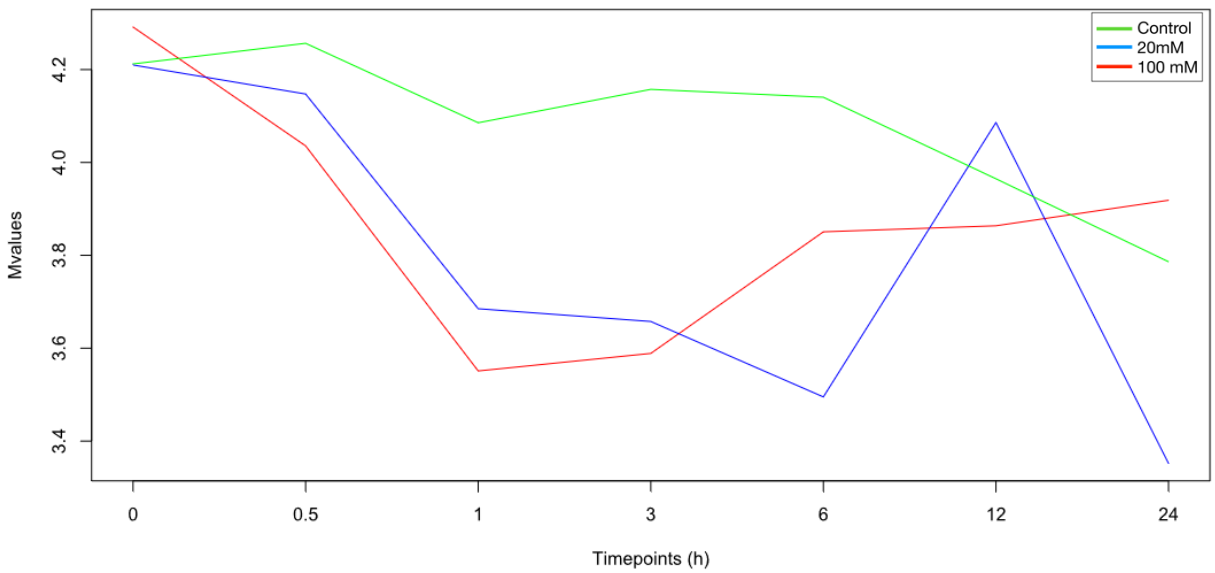


Figure 23 – Methylation over time (24h) for ACTN1 in HL60

This function is illustrating the methylation over time for ACTN1, one of the candidate gene that were discovered while searching for genes that experienced a fast change with lasting effect. The green line displays the control samples, blue is the 20mM EtOH-treated samples and red represents 100mM EtOH-treated samples. ACTN1 was negatively correlated to the profile search, i.e. there is demethylation of CpG in the ACTN1 following exposure to EtOH.

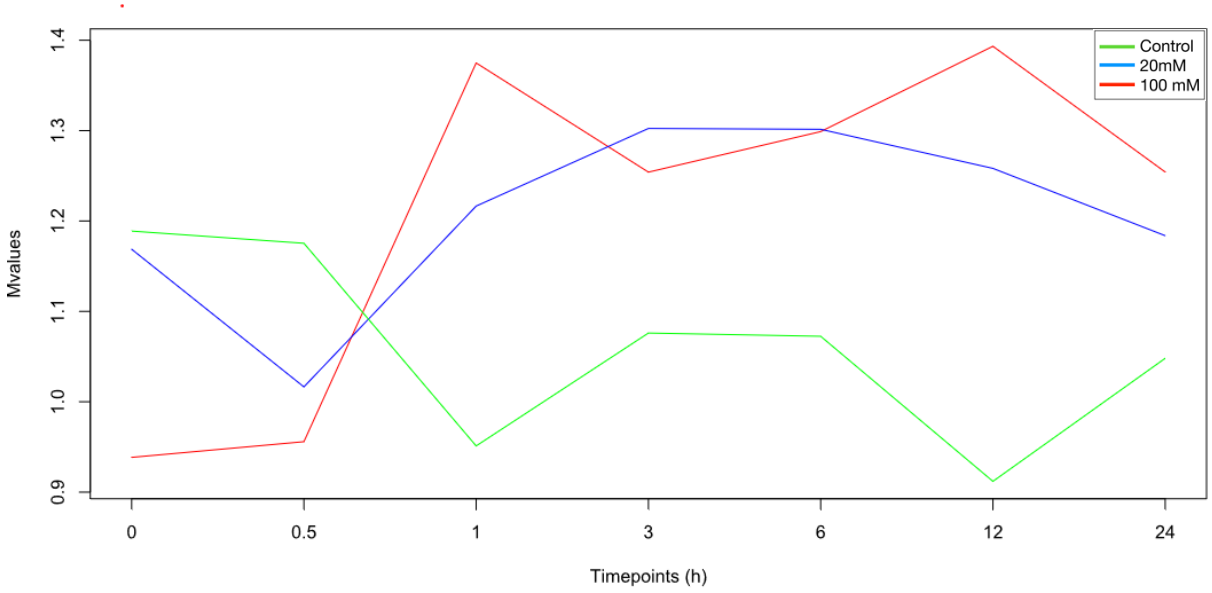


Figure 24 – Methylation over time (24h) for ERRFI1 in HL60

This function is illustrating the methylation over time for ERRFI1, one of the candidate gene that were discovered while searching for genes that experienced a fast change with lasting effect. The green line displays the control samples, blue is the 20mM EtOH-treated samples and red represents 100mM EtOH-treated samples. ERRFI1 was positively correlated to the profile search.

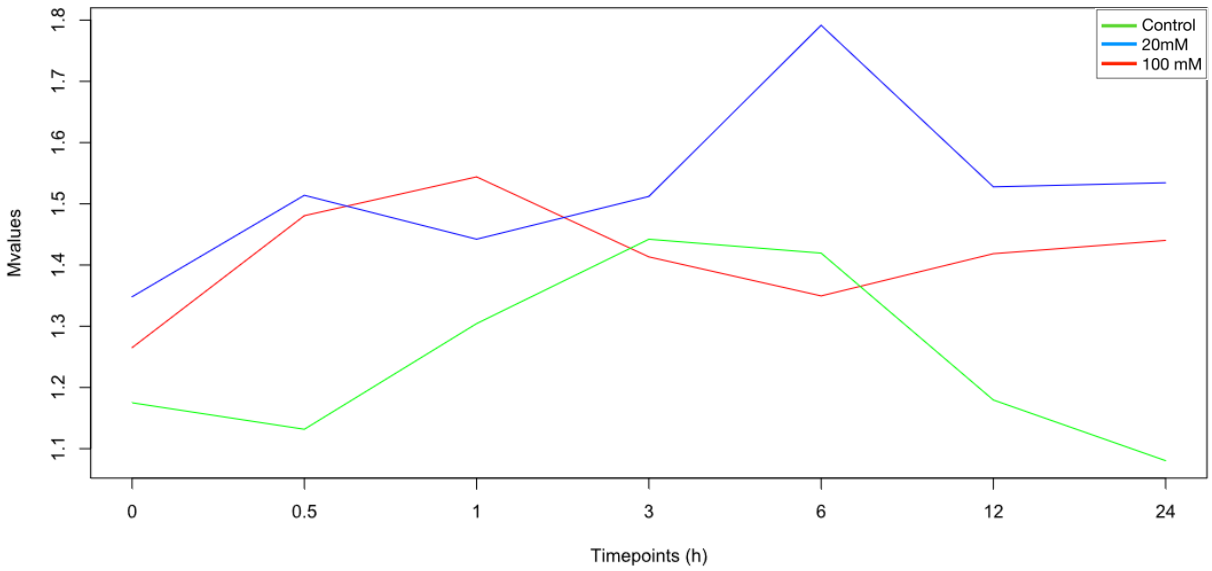


Figure 25 – Methylation over time (24h) for ETV6;RNU6-19P in HL60

This function is illustrating the methylation over time for ETV6;RNU6-19P, two of the candidate genes that were discovered while searching for genes that experienced a fast change with lasting effect. Both genes have the same probe and appear to form a "partner gene". The green line displays the control samples, blue is the 20mM EtOH-treated samples and red represents 100mM EtOH-treated samples. ETV6;RNU6-19P was positively correlated to the profile search.

The candidate genes retrieved from the first part of profile search were run for the whole timeline to determine how fast the epigenetic changes were reversed. For samples that consisted of replicates, one copy was randomly selected. From the previous part, *ACTN1* appeared to have a trend where the methylation occurs fast as it was ranked highest within that alternative with respect to the FDR and adjusted p-value. However, for the whole timeline, that particular probe and its corresponding gene were found at the 1158th position, and had a correlation value of – 0.61. *ERRFI1* appeared to have a trend where the change in methylation occurs gradually. The gene was found at the 42175th position and had a correlation value of 0.46 with the profile. The last candidate consists of genes *ETV6* and *RNU6-19P* that shared the same probe. They also appeared to have a trend where the change occurs gradually. The gene was found at the 48199th position and had a correlation value of - 0.45. **Figures 26-28** display the methylation over time (whole timeline) for the candidate genes.

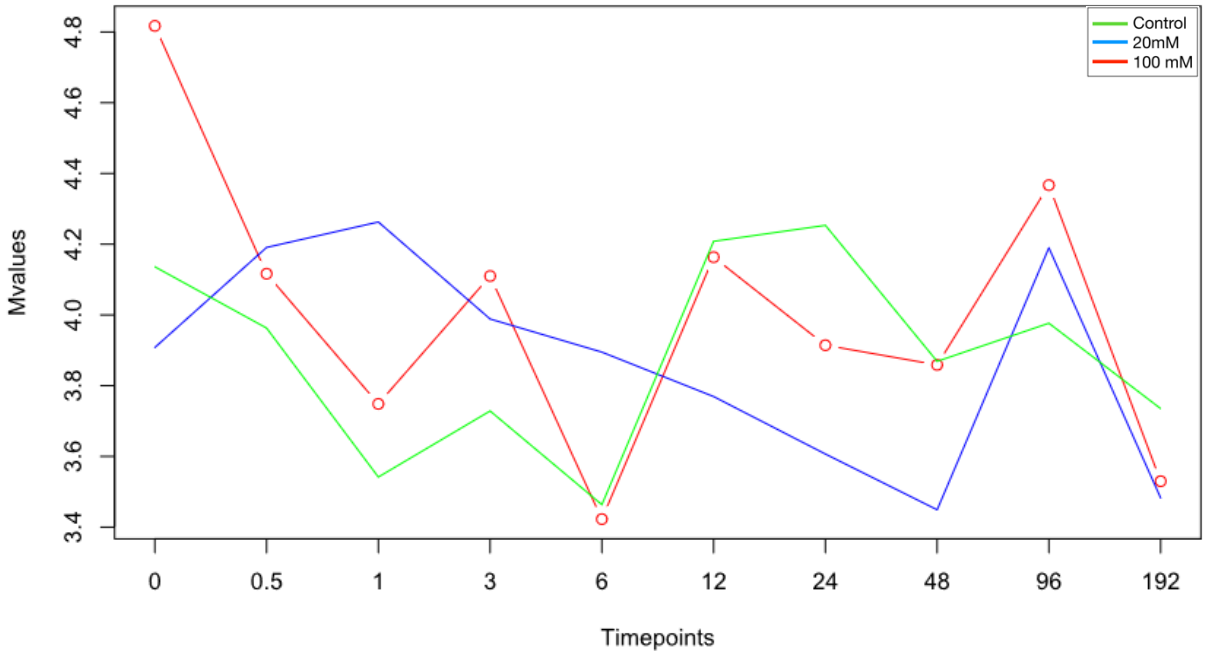


Figure 26 – Methylation over time (whole timeline) for ACTN1 in HL60

This function is illustrating the methylation for ACTN1 for the whole timeline. The green line displays the control samples, blue is the 20mM EtOH-treated samples and red represents 100mM EtOH-treated samples. ACTN1 was negatively correlated to the profile search.

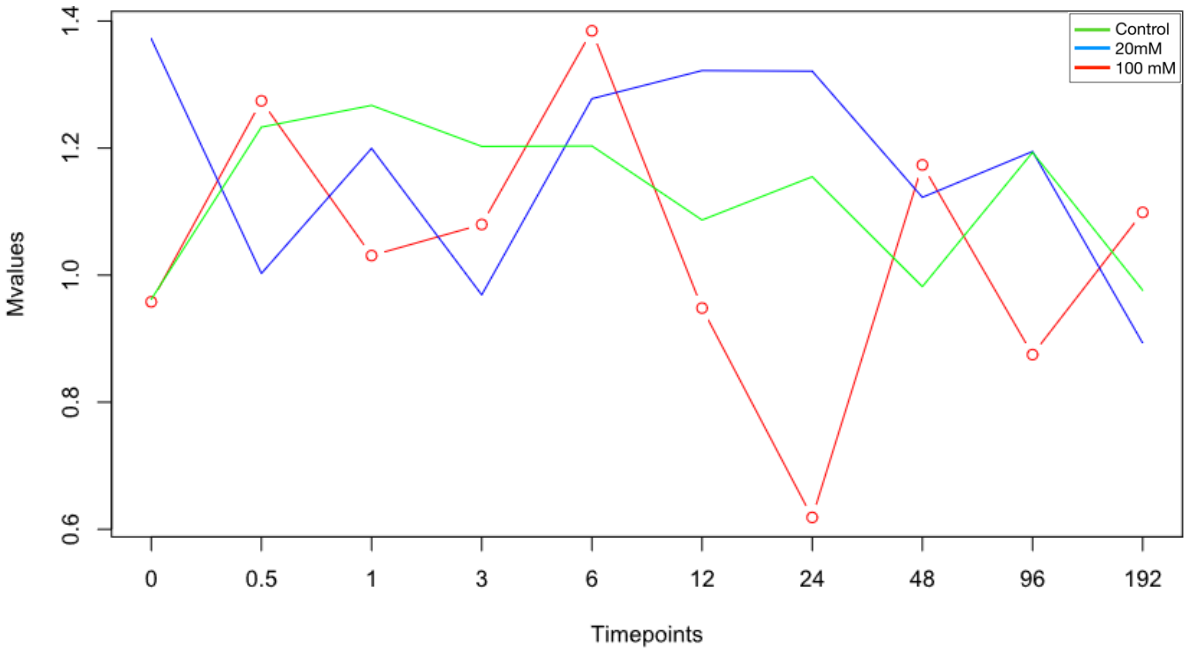


Figure 27 – Methylation over time (whole timeline) for ERRFI1 in HL60

This function is illustrating the methylation for ERRFI1 for the whole timeline. The green line displays the control samples, blue is the 20mM EtOH-treated samples and red represents 100mM EtOH-treated samples. ERRFI1 was positively correlated to the profile search.

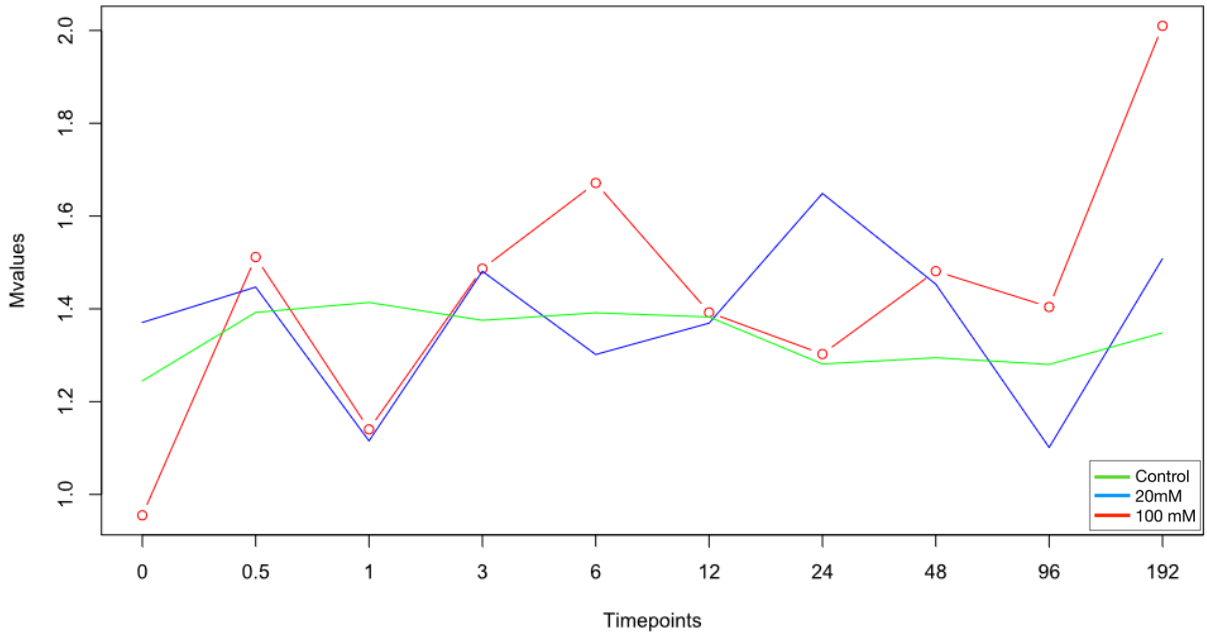


Figure 28 – Methylation over time (whole timeline) for ETV6;RNU6-19P in HL60

This function is illustrating the methylation for ETV6;RNU6-19P for the whole timeline. The green line displays the control samples, blue is the 20mM EtOH-treated samples and red represents 100mM EtOH-treated samples. ERRFI1 was negatively correlated to the profile search.

In the second part of the profile search, all time points were included to determine how fast the epigenetic changes were reversed. For samples that consisted of replicates, one copy was randomly selected. From the previous part, the second profile of fast change with lasting effect was chosen, and further constructed with respect the post-treated samples. The profiles consist of with fast change (increase or decrease) in methylation with lasting effect, gradual change and unchanged. **Figure 29-32** are illustrating the different profiles for for the whole timeline.

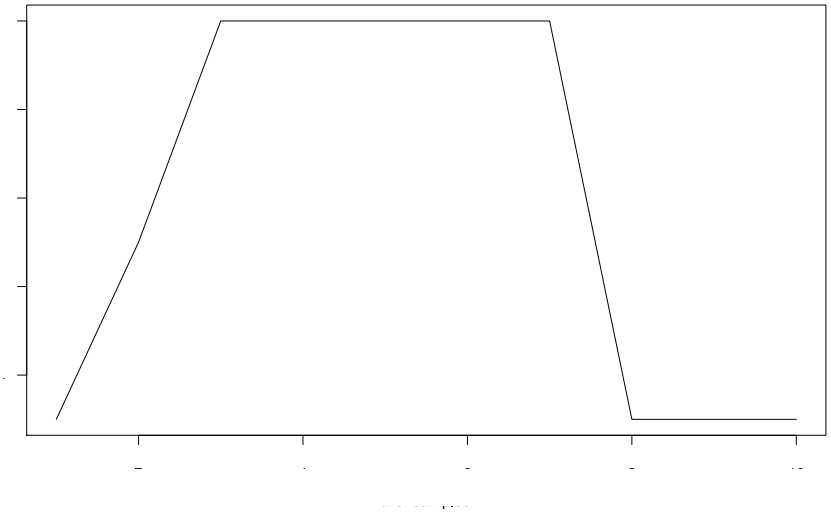


Figure 29 – Profile search #2 for the whole timeline with fast change

From the first part of profile searches (24h), profile search #2 was chosen for further analysis. The first alternative is built based on fast decrease in signal once EtOH is removed. The chronological order corresponds to the time points 0h, 0.5h, 1h, 3h, 6h, 12h, 24h, 48h, 96h and 192h, respectively. With respect to correlation the change could be increase or a decrease in methylation values.

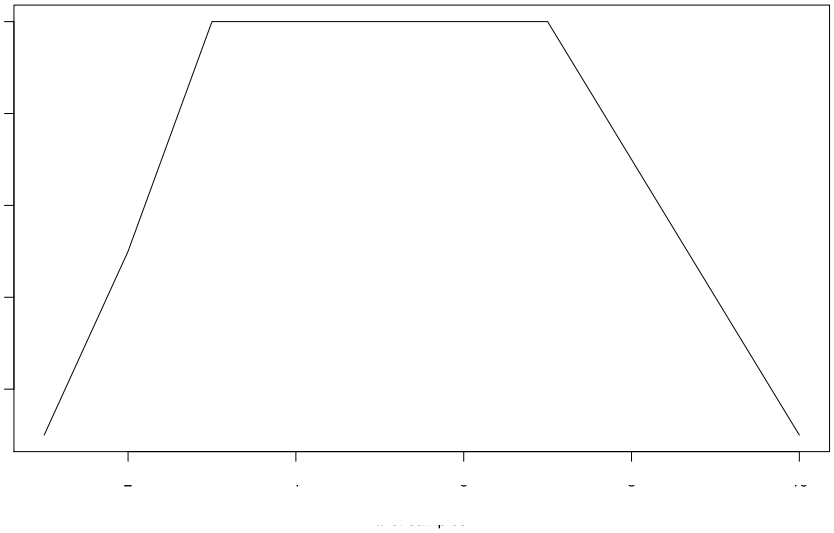


Figure 30 – Profile search #2 for the whole timeline with gradual change

From the first part of profile searches (24h), profile search #2 was chosen for further analysis. The second alternative is built based on gradual decrease in signal once EtOH is removed. The chronological order corresponds to the time points 0h, 0.5h, 1h, 3h, 6h, 12h, 24h, 48h, 96h and 192h, respectively. With respect to correlation the change could be increase or a decrease in methylation values.

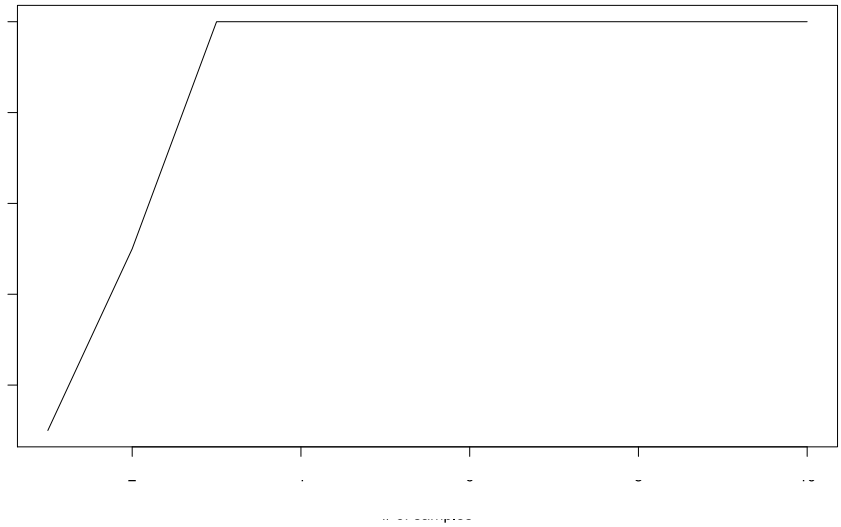


Figure 31 – Profile search #2 for the whole timeline with unchanged signal

From the first part of profile searches (24h), profile search #2 was chosen for further analysis. The last alternative is built based on that the signal remains unchanged once EtOH is removed. The chronological order corresponds to the time points 0h, 0.5h, 1h, 3h, 6h, 12h, 24h, 48h, 96h and 192h, respectively.

The first alternative represented a fast change in methylation once the ethanol was removed with lasting effect. The FDR was set to 0.10 and did not yield any significant hits nor any potential candidate genes. The second alternative represented a gradual change in methylation and yielded 1019 observations. The p-value foroverlap with Wilson et al. (2019) was 0.345 and yielded candidate gene *ANKRD11* with a correlation value of - 0.75. The last alternative remained unchanged signal once ethanol was removed and yielded three observations below FDR of 10%. The overlapped p-value was 0.435 and yielded *ZFHX3* as candidate gene. The correlation value was 0.71. **Table 12** display the number of observations, p-value for overlap between the significant genes and Wilson et al. (2019) for the different profile searches, and **Figures 32** and **33** display the methylation over time (whole timeline) for the candidate genes.

Table 12 – Obtained observations, p-value for overlap and overlap genes for the different profile searches for the whole timeline:

Alternative	Observations (FDR = 10%)	P-Value (P.Value)	Gene	Correlation
1 – Fast change	238	-	-	-
2 – Gradual change	1019	0.345	*ANKRD11*	- 0.75
3 – Unchanged	3	0.435	*ZFHX3*	0.71

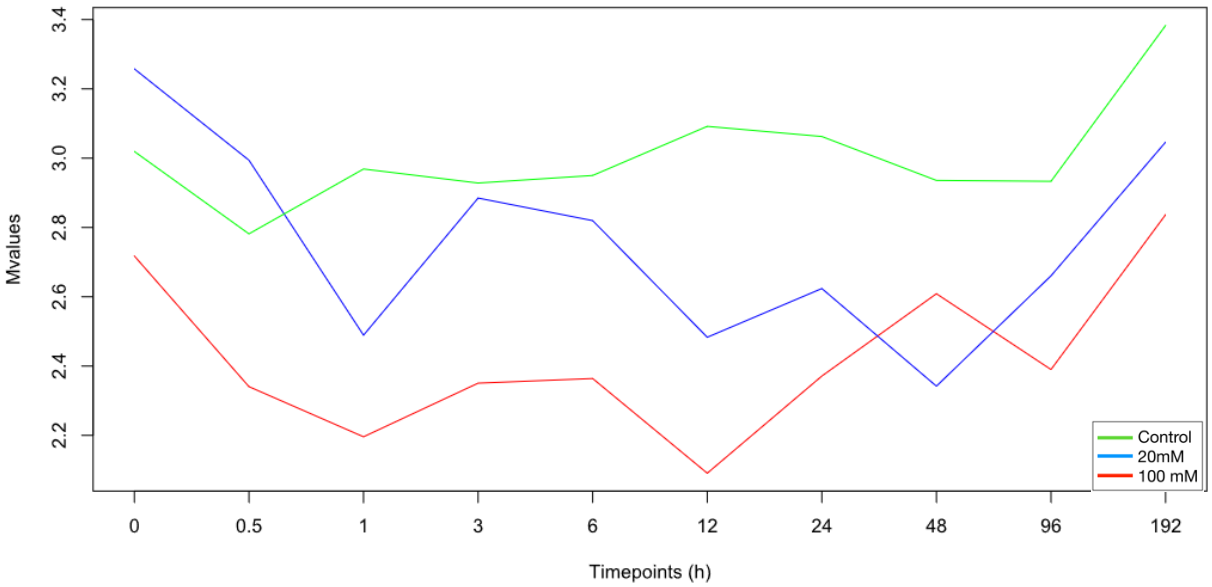


Figure 32 – Methylation over time (whole timeline) for ANKRD11 in HL60

This function is illustrating the methylation over time for ANKRD11, one of the candidate gene that were discovered while searching for genes that experienced a fast change with lasting effect once ethanol was removed. The green line displays the control samples, blue is the 20mM EtOH-treated samples and red represents 100mM EtOH-treated samples. ANKD11 was negatively correlated to the profile search.

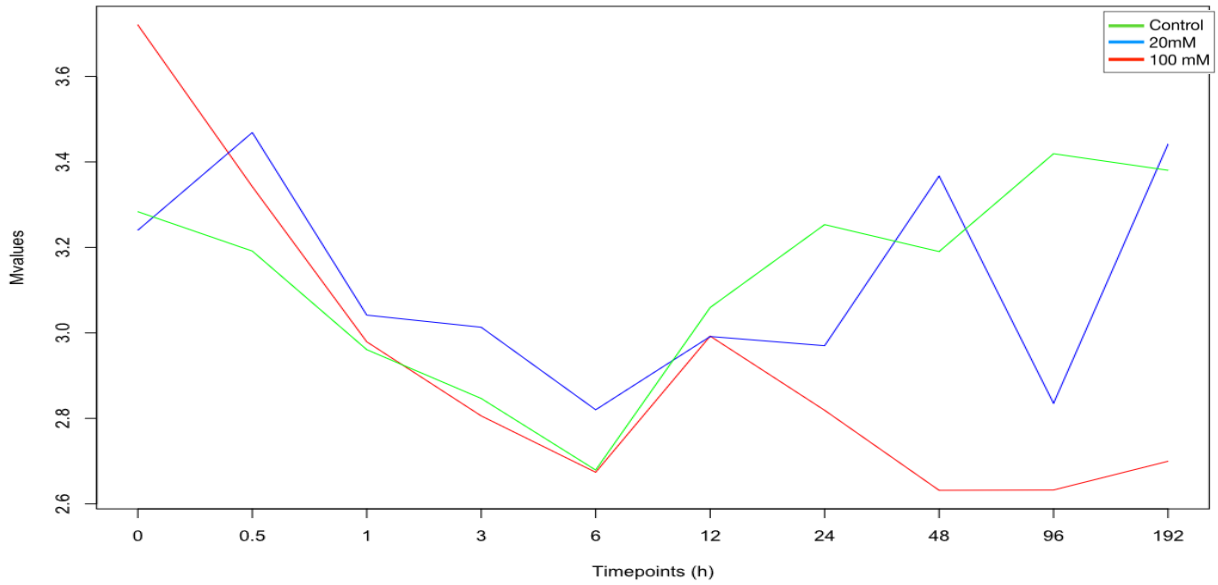


Figure 33 – Methylation over time (whole timeline) for ZFHX3 in HL60

This function is illustrating the methylation over time for ZFHX3, one of the candidate gene that were discovered while searching for genes that remained unchanged when ethanol was removed. The green line displays the control samples, blue is the 20mM EtOH-treated samples and red represents 100mM EtOH-treated samples. ZFHX3 was positively correlated to the profile search.

10. Discussion

Epigenetics has the potential to explain various biological phenomena. Numerous investigations have explored the relationship between environmental factors and epigenetics, discovering several components that modify epigenetic marks. Some changes have been identified as temporary, whereas other are permanent and have subsequently shown to be transmitted trans-generationally. The occurrence of these changes has been established, but the timescale regarding how fast these modifications take place are still unknown.

This project was initially designed after a literature search for finding how dynamic methylation changes are, namely how fast changes are induced and how fast they are removed. Since we could not find such information in the literature our aim was to determine DNA methylation timeline under a well characterised environmental exposure. As mentioned previously, there are several factors that can affect the epigenome. In our project, we chose EtOH as inducer as it is already well documented that exposure of alcohol, particularly EtOH, has an effect on the DNA methylation as well affects other layers of epigenetic modification.

10.1 Main findings

Based on our findings, it was discovered that changes in DNA methylation have a fast response. In the conducted profile searches for exposed samples, genes that displayed a trend with fast change with lasting effect yielded the most significant results. Our profile searches considered both positively and negatively correlated methylation values, representing both DNA methylation and DNA demethylation. In the mentioned search, we assumed that the maximal effect of methylation would be achieved after one hour of exposure. The first sample was collected 30 min after addition of EtOH, with intention of picking up the initiation of DNA methylation, but our results indicate that the process of methylation has already started, postulating that the process begins even before 30 minutes. Based on profile searches, the DNA methylation is not only occurring rapidly but is also reversed rapidly for some genes. However, the first post-treated sample was collected after 48h i.e. 24h after the cells had been exposed to EtOH. This time interval gives the cells sufficient time to recover and reverse the changes back to its original state without being picked up by the following time points.

Our findings are of interest since no study has previously documented how fast the DNA methylation is affected by its environment. Studies have always assumed that DNA methylation is a stable process, especially since the addition of methyl group to cytosine is a highly conserved mechanism (Joseph et al., 2018). The results give us reason to believe that there are regions within the DNA that are versatile, where some regions appear to be more susceptible to modification than others.

Profile search for 24h identified DNA methylation changes for 3 probes (cg25476208, cg03335216 and cg06753604) which are found in or near candidate genes *ACTN1, ERRFI1, ETV6* and *RNU6-19P*, respectively. These genes have previously been associated with DNA methylation changes in an epigenome-wide association study (EWAS) by Wilson et al. (2019). *ACTN1* represents actinins, a major group of cytoskeletal proteins that primarily mediate muscular function as well as regulating cytokinesis, cell addition and cell migration (Yang et al., 2019). The gene of interest is one of four possible isoforms found in mammals, and are universally expressed in most cell types and tissues (Yang et al., 2019). Several studies have independently reported the association of *ACTN1* with smoking (Santos et al., 2012; Takahashi et al., 2018; Wehby, Prater, Ryckman, Kummet, & Murray, 2015). Baker et al. (2017) conducted analyses on rodents, and reported altered gene expression of *ACTN1* in several of the brain compartments (hippocampus, amygdala, ventral tegmental etc) when they were exposed to acute EtOH (Baker et al., 2017). Although these findings were discovered in the human brain, they can also be applied to blood cells (HL60). It was formerly mentioned that the HL60 cell line is originated from a female patient with acute myeloid leukaemia (AML). A study from Yang et al. (2019) reported that high levels of isoforms *ACTN1* and *ACTN3* were associated with unfavourable prognosis in AML. One can therefore postulate that *ACTN1* is already overexpressed in the given cell line prior to treatment, but is even more expressed under the presence of EtOH (Yang et al., 2019). *ERRFI1* represents a protein found in the cell's cytoplasm that is upregulated with cell growth. This gene was also found in Baker et al. (2017) study conducted on rodents, and was differentially expressed after exposure to EtOH, stress and a combination of both (Baker et al., 2017; Bönsch et al., 2006) *ETV6* represents a protein that functions as a transcription factor by repressing gene activity, and is primarily associated with regulation of blood cell development and embryonic development, whereas *RNU6-19P* is a pseudogene. During the analysis, it was discovered that *ETV6* and *RNU6-19P* shared the same probe, indicating the formation of a fusion gene i.e. a hybrid gene composed of the two separate genes. This partnered gene has also been discovered by Trivedi et al. (2018) in a study on chronic myelomonocytic leukaemia (CML) (Trivedi et al., 2018)

For *ACTN1* the overall methylation signal for control and treatments were diminished from the start (0h) to the end point (24h). Due to fast decrease in M-value with respect to the time points, we can postulate that this gene segment is demethylated through active demethylation. However, the positive M-values indicate that more than half of the cells that consist of that probe are methylated. For *ERRFI1* and ETV6;*RNU6-19P* the methylation values for both treatments were increased whereas their controls had diminished methylation values. Nonetheless, the M-value for both genes was positive. Common for all three genes is the presence of variation between the different time point, due to background noise in the dataset. We suspected that the variation between the different time points are due to mechanical stress that are inflicted on cells during the experiment. It is therefore challenging to determine if the different genes undergo demethylation, and what type of demethylation pathway they may undergo.

Based on current knowledge, we have assumed that any observed decrease is due to active demethylation as passive methylation is reported to be observed with dilution i.e. passage of several cell divisions. Since the whole experiment only lasted a week, we assume that the cells must have divided no more than 4-5 times. Unbiased of the variation, these graphs do give an indication as to how fast the epigenetic modifications are occurring, and one can therefore deduce that these alterations take place fast and adapt to its environment/surroundings within ~1h. The exact time will vary with respect to the position of a certain gene within the DNA and the processes it is involved in.

In the second part of the profile search, all time points were included with focus on the post-exposed data to discover if the changes were reversed once EtOH was removed. The obtained results suggest that the changes are mostly reversed gradually. With respect to the false discovery rate and adjusted p-values *ANKRD11* was obtained as candidate gene. According to its profile, the overall methylation value increases. *ANKRD11* represents the protein ankyrin repeat domain 11, and interacts with other proteins such as HDA to control the gene activity. Multiple mutations in *ANKRD11* have been found to cause Kashin-Beck disease (KBD), a condition characterised by unusual facial features and intellectual disability. Emerging evidence suggest that *ANKRD11* is involved in enhancing the activity of p53, a well-known tumour suppressor gene involved in controlling and regulating cell division and cell apoptosis.

10.2 Discussion of performed analyses

Factor analysis was performed for the whole timeline. Since there already was an observed difference between the 100mM-treated and control samples with respect to the time point at which the experiments took place, one could not know if the yielded results were due to the exposure of EtOH or the prolonged cell culturing of the control. Even though the dataset for HL60 was corrected prior to factor analysis, there was a still a possibility that the contrast was due to one or several underlying factors that somehow were influenced by when the experiment took place and escaped the correction. This might further implicate that the yielded results are due to the time when the treatment was performed rather than actual differences between the 100mM EtOH-treated and control samples. The following analyses were therefore primarily performed with the 20mM EtOH treatment. However, one did not get any significant hits for potential genes that were affected by the treatment.

The time points were further clustered into subgroups, acting as replicates for each other. The different combinations of subgroups were tried but yielded few significant hits and visualisation of the top gene candidates displayed variation between the time points. It was later noticed that the clustering into subgroups was not ideal with respect to the interval between the set time points.

The first subgroup had a time interval of 2.5h between the first and last collected samples whereas the last subgroup had a time interval of six days, yielding an uneven balance of time.

10.3 Methodological considerations

10.3.1 Evaporation of EtOH and growth medium

Prior to treatment, the experimental parameters were determined. EtOH was chosen as substance in this experiment due to its well documented effects. However, EtOH evaporates quickly. Before applying the substance, the rate of evaporation was measured. On a 6-well plate, duplicates of samples consisting of 20mM, control and 100mM were plated. Samples that consisted of EtOH had a diminished solution whereas both control samples that in theory should have had 0 ‰ had an average value of 0.92‰ and 0.72‰ after 6h and 24h, respectively. However, these observations can be explained by the laws of Thermodynamics. Since all the samples are under the same atmosphere, there is a net flux of molecules that can motion from a region with high concentration to one of lower concentration. When collecting the samples, it was observed that the alcohol-containing samples, especially 100mM, had decreased volume than start whereas the control samples had final volumes that were slightly more than initial volumes. For samples that were collected after 24h, ~ 0.3 mL appeared to have evaporated. Boyle's law states that the pressure of a gas tends to increase as the volume decreases given that the temperature remains constant. Ethanol has a boiling point of 78.2 ℃, and will also start of evaporate at a much lower temperature as compared to water (in or case cell culture medium). Therefore, the 100mM samples that contain most EtOH will experience a higher loss of volume compared to samples that consisted of a lower concentration, which indeed was observed during the experimental set-up. Samples could have been sealed to avoid evaporation, however such conditions would have hindered the flow/source of CO_2 and further decreased the pH that ultimately would have affected the cell viability. Studies have shown that this decrease could have been overcome by adding a non-carbonate buffer in the cell culture medium, but only for experiments that had a total duration of less than 24h (Eysseric et al., 1997). Based on obtained results, it was decided that samples from different treatment should not be plated together nor be placed in the same incubator. The loss of EtOH due to evaporation was overcome by switching the cell culture medium every 8h. The cell medium for control samples were also changed to maintain an identical procedure of all treatments

10.3.2 Cells

When setting up the experiment, it was observed that cells were lost due to multiple centrifugations and change of containers, yielding poor DNA concentrations. This was especially observed for HL60 cell line as they are suspension cells, and was solved by doubling the total volume as well as the cell concentration. Ideally, it is not recommended to have cell density above 1×10^6 cells/mL, but due to the loss during the experiment as well as the exposure to EtOH one assumed that the cell density would in practical terms be lower than the given value. Additionally, this cell density concentration was needed with respect to extraction and isolation of DNA and RNA for further analysis. During the course of the practical experiment, it was noticed that the HL60 cell line had a reduced growth rate. The cells were observed under the microscope but did not show any signs to cell enlargement or any form of disfiguration. Furthermore, the cells were subjected to MTT assay and trypan blue solution, yielding a cell viability and proliferation of ~ 100%. The cells were also checked for contamination of mycoplasma and yielded negative results. Both growth medium and its composition as well as cell density was followed according to its protocol as well as other literature regarding the cell culturing of HL60. Since growth rate was slow during the treatment and the growth medium had not been discoloured, the HL60 cells were not passaged in that time period. One might therefore postulate that the reduced growth rate of cells might have affected the methylation results and to some extent caused the observed variation between the replicated samples in both factor and spline analysis.

While performing the statistical analysis for the HepG2 cell line, extreme variations between the different time points were observed. Furthermore, we discovered that the obtained hits for HepG2 consisted of a broad spectrum of genes; however, most of them were stress related genes. These may be genes that may indirectly linked to exposure to EtOH but have not yet been associated. HepG2 cells are adherent cells. In order to avoid trypsinisation and additional stress of cells, and affect the ongoing treatment, HepG2 cells were scraped from the surface. We suspect that scraping induced mechanical stress, yielding epigenetic changes and variating methylation values. In fact, several studies have reported that even influence of neighbouring cells can induce stress in adherent cells (Djukelic, Wixforth, & Westerhausen, 2017). Due to background noises and observed variations, we decided to focus on HL60 cell line. It should be mentioned that HL60 and HepG2 are both cell lines that originate from cancer. Imbalance of methylation is characteristic of cancer, and methylation of genes in these cell lines can therefore be unstable and yield variance (Hentze, Høgdall, & Høgdall, 2019).

10.3.3 Evaluation of covariates

During the evaluation the potential covariates, we discovered an unwanted variation for the HL60 cell line, clustering the samples into two groups. Further inspection of the distribution of the samples showed that they were grouped according to the time point at which the experiments took place. Control and 20mM EtOH-treaded samples were clustered within one region whereas the 100mM EtOH-treated samples were clustered farther apart. It may be postulated that the observed differences might have been due to the different concentrations the cell line was treated with rather than the time points at which each sample was collected.

It was observed that within the clustering of the control and 20mM EtOH-treaded samples, the control samples mostly have a higher PCA1-value and are located towards right whereas the 20mM EtOH-treaded have a lower value and are located towards left within the clustered region, thereby suggesting that the concentration of treatment is the dominating factor that influences the results. However, one sample for the 100mM EtOH-treated (48h replicate 2) HL60 cell line was missing and rerun within the same period time as the control and 20mM EtOH treaded samples. That sample is an outliers with a substantial low PCA1-value compared to the other 100mM EtOH-treaded samples, thus suggesting that long-term culturing and higher passage of cells is the dominating factor that impacts/affects the methylation values. Expectedly, most cells tend to reach a point where they undergo replicative senescence, a state where they are no longer able to divide normally and is often characterised by loss of differentiation potential, cell enlargement and growth arrest, ultimately affecting the DNA methylation pattern. A study from Bork et al (2010) discovered changes in DNA methylation upon long-term culturing of human mesenchymal stromal cells (MSC). The cells were harvested at early and late passages, and analyses unveiled that several CpG sites displayed a highly significant degree of differential methylation between the passages (Bork et al., 2010). The study further elucidated that only specific CpG sites were differently methylated upon long-term culturing where some of them were hypermethylated and others were hypomethylated. Additionally, the methylation changes were continuously acquired with every passage (or upon/by every passage). This claim is strengthened by other studies that have also reported decreased DNA methylation levels upon long-term cell culturing (Bork et al., 2010)

As mentioned previously, analyses of the different PCA-plots displayed that HL60 samples needed to be corrected with respect to the time point at which the experiment took place and were collected as well as slide. Although the decisions were based upon significant hits, by correcting for the mentioned covariates one might also face consequences such as over-correction i.e correcting the dataset multiple times for covariates that are somehow correlated, removal or overshadowing of potential results/significant hits.

11. Conclusion

Based on our findings, we may conclude that changes for DNA methylation have a fast response when exposed to EtOH. In the conducted profile searches for exposed samples, genes that displayed a trend with fast change with lasting effect yielded the most significant results. The first sample for measuring DNA methylation was collected 30 min after addition of EtOH with the intention of picking up the initiation of DNA methylation, but our results indicate that the process of methylation has already started, postulating that the process starts at an earlier stage. The reversion is also fast occurring, although slower

In conclusion, in order to identify the timeline of epigenetic modifications and changes in gene expression, the experiment should be repeated with more careful selection of sampling time points. Specifically, more time points should be added within the first 30 minutes of exposure as well as within the first 24 hours after ending the exposure.

12. Future aspects & work

After performing the experiment, we now have an idea of how fast the epigenetic changes are occurring. The first thing would be to change the timeline for harvested cells. Based on previous results, we should include more timepoints within a shorter time interval. For an improved timeline, the samples could be collected 5 min, 10 min, 15 min, 20 min, 30 min, 45 min, 1h and 2h during the presence of ethanol and after exposure. Also, since the changes are occurring rapidly, the exposure period should be reduced (from 24h to 12h for instance) We observed that already after 6h a quantity of EtOH had evaporated from the different samples. The alcohol-containing cell medium should therefore be switched more often, for instance every 3h. Initially, replicates were not included. To achieve high quality and credibility of results and limit variation between the time points, replicates must be included. This also gives us opportunities to try out other statistical methods without having to cluster them into subgroups. For performance of factor analysis, time points were clustered so that they could act as replicates for each other. However, some time points ended having a short time interval between collected time points whereas other had days in between yielding an uneven distribution.

Cells that originate from cancerous cell lines have reported to have an imbalance of methylation levels and may have an impact along with the EtOH exposure. One should consider to use fresh cell lines instead such as monocytes. Additionally, if using cells from adherent cell lines, another method that is "gentler" than scraping should be used for harvesting of cells. For future experiments, Accutase® may be used. This cell detachment solution is documented to be less toxic than trypsin, but mimics its action and is just as effective (Innovative Cell Technologies).

13. References

al., T. e. (2018). Chronic myelomonocytic leukemia with t (12;15) involving dup (1) identified by cytogenetic and molecular cytogenetic approach : A novel favorable-risk translocation. *Journal of Blood Disorders ant Treatment*.

Alaskhar Alhamwe, B., Khalaila, R., Wolf, J., von Bülow, V., Harb, H., Alhamdan, F., . . . Potaczek, D. P. (2018). Histone modifications and their role in epigenetics of atopy and allergic diseases. *Allergy, asthma, and clinical immunology : official journal of the Canadian Society of Allergy and Clinical Immunology, 14*, 39-39. doi:10.1186/s13223-018-0259-4

An, J., Rao, A., & Ko, M. (2017). TET family dioxygenases and DNA demethylation in stem cells and cancers. *Experimental & Molecular Medicine, 49*(4), e323-e323. doi:10.1038/emm.2017.5

Baccarelli, A., & Bollati, V. (2009). Epigenetics and environmental chemicals. *Current opinion in pediatrics, 21*(2), 243-251. doi:10.1097/mop.0b013e32832925cc

Baker, J. A., Li, J., Zhou, D., Yang, M., Cook, M. N., Jones, B. C., . . . Lu, L. (2017). Analyses of differentially expressed genes after exposure to acute stress, acute ethanol, or a combination of both in mice. *Alcohol (Fayetteville, N.Y.), 58*, 139-151. doi:10.1016/j.alcohol.2016.08.008

Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell research, 21*(3), 381-395. doi:10.1038/cr.2011.22

Biscotti, M. A., Olmo, E., & Heslop-Harrison, J. S. (2015). Repetitive DNA in eukaryotic genomes. *Chromosome Research, 23*(3), 415-420. doi:10.1007/s10577-015-9499-z

Bochtler, M., Kolano, A., & Xu, G.-L. (2017). DNA demethylation pathways: Additional players and regulators. *BioEssays, 39*(1), e201600178. doi:10.1002/bies.201600178

Bollati, V., & Baccarelli, A. (2010). Environmental epigenetics. *Heredity, 105*(1), 105-112. doi:10.1038/hdy.2010.2

Bork, S., Pfister, S., Witt, H., Horn, P., Korn, B., Ho, A. D., & Wagner, W. (2010). DNA methylation pattern changes upon long-term culture and aging of human mesenchymal stromal cells. *Aging cell, 9*(1), 54-63. doi:10.1111/j.1474-9726.2009.00535.x

Brehove, M., Wang, T., North, J., Luo, Y., Dreher, S. J., Shimko, J. C., . . . Poirier, M. G. (2015). Histone core phosphorylation regulates DNA accessibility. *The Journal of biological chemistry, 290*(37), 22612-22621. doi:10.1074/jbc.M115.661363

Bruce R. Korf, M. B. I. (2013). *Human Genetics and Genomics* (4th ed.): Wiley-Blackwell.

Bruna Brands, B. S., Joan Marshman. (1998). Straight Fact About Drugs and Drug Abuse. (3rd edition), 60.

Bönsch, D., Lenz, B., Fiszer, R., Frieling, H., Kornhuber, J., & Bleich, S. (2006). Lowered DNA methyltransferase (DNMT-3b) mRNA expression is associated with genomic DNA hypermethylation in patients with chronic alcoholism. *Journal of Neural Transmission, 113*(9), 1299-1304. doi:10.1007/s00702-005-0413-2

Clancy, S. (2008). DNA transcription *Nature Education, 1(1):6*.

Dasgupta, A. (2017). 1 - Alcohol a double-edged sword: Health benefits with moderate consumption but a health hazard with excess alcohol intake. In A. Dasgupta (Ed.), *Alcohol, Drugs, Genes and the Clinical Laboratory* (pp. 1-21): Academic Press.

Djukelic, M., Wixforth, A., & Westerhausen, C. (2017). Influence of neighboring adherent cells on laminar flow induced shear stress in vitro—A systematic study. *Biomicrofluidics, 11*(2), 024115. doi:10.1063/1.4979295

Dobs, Y. E., & Ali, M. M. (2019). The epigenetic modulation of alcohol/ethanol and cannabis exposure/co-exposure during different stages. *Open Biology, 9*(1), 180115. doi:doi:10.1098/rsob.180115

Drouet, E. (2019). Epigenetics: How the environment influences our genes. Retrieved from https://www.encyclopedie-environnement.org/en/health/epigenetics-how-the-environment-influences-our-genes/#_ftn12

Du, P., Zhang, X., Huang, C.-C., Jafari, N., Kibbe, W. A., Hou, L., & Lin, S. M. (2010). Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics, 11*(1), 587. doi:10.1186/1471-2105-11-587

Eisenberg, E., & Levanon, E. Y. (2013). Human housekeeping genes, revisited. *Trends in Genetics, 29*(10), 569-574. doi:https://doi.org/10.1016/j.tig.2013.05.010

Eysseric, H., Gonthier, B., Soubeyran, A., Bessard, G., Saxod, R., & Barret, L. (1997). There is no simple method to maintain a constant ethanol concentration in long-term cell culture: Keys to a solution applied to the survey of astrocytic ethanol absorption. *Alcohol, 14*(2), 111-115. doi:https://doi.org/10.1016/S0741-8329(96)00112-7

Fang, I. J., & Trewyn, B. G. (2012). Chapter three - Application of Mesoporous Silica Nanoparticles in Intracellular Delivery of Molecules and Proteins. In N. Düzgüneş (Ed.), *Methods in Enzymology* (Vol. 508, pp. 41-59): Academic Press.

Garber, K. (2019). Epigenetics comes to RNA. *Science, 365*(6448), 16. doi:10.1126/science.365.6448.16

Griffiths AJF, M. J., Suzuki DT, et al. (2000). *An Introduction to Genetic Analysis* (7th edition ed.). New York: W. H. Freeman.

Grunau, C., Buard, J., Brun, M.-E., & De Sario, A. (2006). Mapping of the juxtacentromeric heterochromatin-euchromatin frontier of human chromosome 21. *Genome research, 16*(10), 1198-1207. doi:10.1101/gr.5440306

Guigó, R. (2013). Chapter 2 - The Coding and the Non-coding Transcriptome. In A. J. M. Walhout, M. Vidal, & J. Dekker (Eds.), *Handbook of Systems Biology* (pp. 27-41). San Diego: Academic Press.

Hentze, J. L., Høgdall, C. K., & Høgdall, E. V. (2019). Methylation and ovarian cancer: Can DNA methylation be of diagnostic use? *Molecular and clinical oncology, 10*(3), 323-330. doi:10.3892/mco.2019.1800

Hershey, J. W. B., Sonenberg, N., & Mathews, M. B. (2012). Principles of translational control: an overview. *Cold Spring Harbor perspectives in biology, 4*(12), a011528. doi:10.1101/cshperspect.a011528

Higashi, T., Matsunaga, S., Isobe, K., Morimoto, A., Shimada, T., Kataoka, S., . . . Fukui, K. (2007). Histone H2A mobility is regulated by its tails and acetylation of core histone tails. *Biochemical and Biophysical Research Communications, 357*(3), 627-632. doi:https://doi.org/10.1016/j.bbrc.2007.03.203

Illumina. (2015). Illumina Methylation BeadChips Achieve Breadth of Coverage Using 2 Infinium® Chemistries. *Technical Note: Epigenetic Analysis.*

Innovative Cell Technologies, I. Accutase®. Retrieved from http://www.accutase.com/accutase.html

Joseph, D. B., Strand, D. W., & Vezina, C. M. (2018). DNA methylation in development and disease: an overview for prostate researchers. *American journal of clinical and experimental urology, 6*(6), 197-218. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/30697577

Karni, S. (2007). Analysis of Biological Networks : Transcriptional Networks-Promoter Sequence Analysis. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6334199/

Lee, K. W. K., & Pausova, Z. (2013). Cigarette smoking and DNA methylation. *Frontiers in genetics, 4*, 132-132. doi:10.3389/fgene.2013.00132

Lodish, H. (2013). *Molecular cell biology* (7th ed. ed.). New York: Freeman.

Lumey, L., Stein, A. D., Kahn, H. S., van der Pal-de Bruin, K. M., Blauw, G., Zybert, P. A., & Susser, E. S. (2007). Cohort Profile: The Dutch Hunger Winter Families Study. *International Journal of Epidemiology, 36*(6), 1196-1204. doi:10.1093/ije/dym126

M. E. Ritchie, B. P., D. Wu, Y. Hu, C. W. Law, W. Shi, G. K. Smyth. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research, 43*.

Mahajan, S. D., Law, W.-C., Aalinkeel, R., Reynolds, J., Nair, B. B., Yong, K.-T., . . . Schwartz, S. A. (2012). Chapter three - Nanoparticle-Mediated Targeted Delivery of Antiretrovirals to the Brain. In N. Düzgüneş (Ed.), *Methods in Enzymology* (Vol. 509, pp. 41-60): Academic Press.

Mariño-Ramírez, L., Kann, M. G., Shoemaker, B. A., & Landsman, D. (2005). Histone structure and nucleosome stability. *Expert Review of Proteomics, 2*(5), 719-729. doi:10.1586/14789450.2.5.719

Masser, D. R., Hadad, N., Porter, H., Stout, M. B., Unnikrishnan, A., Stanford, D. R., & Freeman, W. M. (2018). Analysis of DNA modifications in aging research. *GeroScience, 40*(1), 11-29. doi:10.1007/s11357-018-0005-3

Matlin, A. J., Clark, F., & Smith, C. W. J. (2005). Understanding alternative splicing: towards a cellular code. *Nature Reviews Molecular Cell Biology, 6*(5), 386-398. doi:10.1038/nrm1645

Moore, L. D., Le, T., & Fan, G. (2013). DNA methylation and its basic function. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology, 38*(1), 23-38. doi:10.1038/npp.2012.112

Morgan, H. D., Sutherland, H. G. E., Martin, D. I. K., & Whitelaw, E. (1999). Epigenetic inheritance at the agouti locus in the mouse. *Nature Genetics, 23*(3), 314-318. doi:10.1038/15490

Pandey, S. C., Sakharkar, A. J., Tang, L., & Zhang, H. (2015). Potential role of adolescent alcohol exposure-induced amygdaloid histone modifications in anxiety and alcohol intake during adulthood. *Neurobiology of Disease, 82*, 607-619. doi:https://doi.org/10.1016/j.nbd.2015.03.019

Pereira, M., Francisco, S., Varanda, A. S., Santos, M., Santos, M. A. S., & Soares, A. R. (2018). Impact of tRNA Modifications and tRNA-Modifying Enzymes on Proteostasis and Human Disease. *International journal of molecular sciences, 19*(12), 3738. doi:10.3390/ijms19123738

Perperoglou, A., Sauerbrei, W., Abrahamowicz, M., & Schmid, M. (2019). A review of spline function procedures in R. *BMC medical research methodology, 19*(1), 46-46. doi:10.1186/s12874-019-0666-3

Pidsley, R., Zotenko, E., Peters, T. J., Lawrence, M. G., Risbridger, G. P., Molloy, P., . . . Clark, S. J. (2016). Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biology, 17*(1), 208. doi:10.1186/s13059-016-1066-1

Powledge, T. M. (2011). Behavioral Epigenetics: How Nurture Shapes Nature. *BioScience, 61*(8), 588-592. doi:10.1525/bio.2011.61.8.4

Rivera, R. M., & Bennett, L. B. (2010). Epigenetics in humans: an overview. *Current Opinion in Endocrinology, Diabetes and Obesity, 17*(6), 493-499. doi:10.1097/MED.0b013e3283404f4b

Sadakierska-Chudy, A., Kostrzewa, R. M., & Filip, M. (2015). A comprehensive view of the epigenetic landscape part I: DNA methylation, passive and active DNA demethylation pathways and histone variants. *Neurotoxicity research, 27*(1), 84-97. doi:10.1007/s12640-014-9497-5

Santos, V., Chatkin, J., Bau, C., Paixão-Côrtes, V., Sun, y., Zamel, N., & Siminovitch, K. (2012). Glutamate and Synaptic Plasticity Systems and Smoking Behavior: Results from a Genetic Association Study. *PloS one, 7*, e38666. doi:10.1371/journal.pone.0038666

Sarkar, D. K. (2016). Male germline transmits fetal alcohol epigenetic marks for multiple generations: a review. *Addiction Biology, 21*(1), 23-34. doi:10.1111/adb.12186

Takahashi, K., Pavlidis, S., Ng Kee Kwong, F., Hoda, U., Rossios, C., Sun, K., . . . Chung, K. F. (2018). Sputum proteomics and airway cell transcripts of current and ex-smokers with severe asthma in U-BIOPRED: an exploratory analysis. *European Respiratory Journal, 51*(5), 1702173. doi:10.1183/13993003.02173-2017

Team, R. C. (2013). R: A Language and Environment for Statistical Computing. Retrieved from http://www.R-project.org

Tiffon, C. (2018). The Impact of Nutrition and Environmental Epigenetics on Human Health and Disease. *International journal of molecular sciences, 19*(11), 3425. doi:10.3390/ijms19113425

Treangen, T. J., & Salzberg, S. L. (2011). Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nature reviews. Genetics, 13*(1), 36-46. doi:10.1038/nrg3117

Trivedi, P. J., Patel, D. V., Patel, D., Kazi, M. M., Varma, P., Ladani, D., . . . Patel, P. S. (2018). Chronic myelomonocytic leukemia with t (12;15) involving dup (1) identified by cytogenetic and molecular cytogenetic approach : A novel favorable-risk translocation. *Journal of Blood Disorders ant Treatment*.

Unwin, A. (2013). Discovering Statistics Using R by Andy Field, Jeremy Miles, Zoë Field. *International Statistical Review, 81*(1), 169-170. Retrieved from https://EconPapers.repec.org/RePEc:bla:istatr:v:81:y:2013:i:1:p:169-170

Vieira, M. L. C., Santini, L., Diniz, A. L., & Munhoz, C. d. F. (2016). Microsatellite markers: what they mean and why they are so useful. *Genetics and molecular biology, 39*(3), 312-328. doi:10.1590/1678-4685-GMB-2016-0027

Vogler, C., Huber, C., Waldmann, T., Ettig, R., Braun, L., Izzo, A., . . . Schneider, R. (2010). Histone H2A C-terminus regulates chromatin dynamics, remodeling, and histone H1 binding. *PLoS genetics, 6*(12), e1001234-e1001234. doi:10.1371/journal.pgen.1001234

W. Huber, V. J. C., R. Gentleman, S. Anders, M. Carlson, B. S. Carvalho, H. C. Bravo, S. Davis, L. Gatto, T. Girke, R. Gottardo, F. Hahne, K. D. Hansen, R. A. Irizarry, M. Lawrence, M. I. Love, J. MacDonald, V. Obenchain, A. K. Ole, H. Page, A. Reyes, P. Shannon, G. K. Smyth, D. Tenenbaum, L. Waldron, M. Morgan. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods, 12*, 115-121. Retrieved from http://www.nature.com/nmeth/journal/v12/n2/full/nmeth.3252.html

Watson, J. D., Baker, T. A., Bell, S. P., Gann, A., Levine, M., & Losick, R. (2013). *Molecular Biology of the Gene*: Pearson Education.

Weaver, I. C. G., Cervoni, N., Champagne, F. A., D'Alessio, A. C., Sharma, S., Seckl, J. R., . . . Meaney, M. J. (2004). Epigenetic programming by maternal behavior. *Nature Neuroscience, 7*(8), 847-854. doi:10.1038/nn1276

Wehby, G. L., Prater, K. N., Ryckman, K. K., Kummet, C., & Murray, J. C. (2015). Candidate gene study for smoking, alcohol use, and body weight in a sample of pregnant women. *The journal of maternal-fetal & neonatal medicine : the official journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstetricians, 28*(7), 804-811. doi:10.3109/14767058.2014.932768

Winchester, A. M. (2019). Genetics. Retrieved from https://www.britannica.com/science/genetics

Wu, X., & Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. *Nature Reviews Genetics, 18*, 517. doi:10.1038/nrg.2017.33

Yang, X., Pang, Y., Zhang, J., Shi, J., Zhang, X., Zhang, G., . . . Fu, L. (2019). High Expression Levels of ACTN1 and ACTN3 Indicate Unfavorable Prognosis in Acute Myeloid Leukemia. *Journal of Cancer, 10*(18), 4286-4292. doi:10.7150/jca.31766

Zannas, A. S., Provençal, N., & Binder, E. B. (2015). Epigenetics of Posttraumatic Stress Disorder: Current Evidence, Challenges, and Future Directions. *Biological Psychiatry, 78*(5), 327-335. doi:https://doi.org/10.1016/j.biopsych.2015.04.003