

https://dspace.org/

The impact of 1,25(OH)2D3 on genome-wide methylation

Nyamupangedengu, Kudakwashe

2021

https://hdl.handle.net/10539/33031

Downloaded from a DSpace repository. https://dspace.org/



The Impact of $1\alpha,25(OH)_2D_3$ on genome-wide methylation

by

Kudakwashe Nyamupangedengu

1043152

Dissertation

Submitted in fulfilment of the requirements for the degree

Master of Science

in

Molecular and Cell Biology

in the Faculty of Science, University of the Witwatersrand, Johannesburg, South Africa

Supervisor: Dr Vanessa Meyer

Co-supervisor: Dr Nikki Gentle

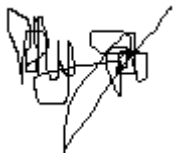
July 2021

Declaration

I, Kudakwashe Nyamupangedengu (1043152), am a student registered for the degree of Master of Science in the academic year 2021.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise and acknowledged. In this context, I understand that the use of editing services is considered aided work and must be declared.
- I have not submitted this work before for any other degree or examination at this or any other University.
- The information used in the Dissertation HAS NOT been obtained by me while employed by, or working under the aegis of, any person or organisation other than the University.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.



_____ _____

Kudakwashe Nyamupangedengu 30 July 2021

For Mbuya Mai J na Simbarashe Chekwa

Through me you live forever.

Acknowledgements

To Almighty, Ever-living God. Thank you for this opportunity. At the point where I couldn't do it anymore, you took over. You carried me on your shoulders. Completion of this masters is testament to your unbelievable love and power. We live by faith and not be sight – but today all I see is your glory.

To Dr V. I have said the words "thank you" to you countless times. I will say thank you forevermore. You trusted me with this project and believed in me. You invested in me: your time, your patience, your effort and your MONEY. You were with me through all my experimental failures: the failed extractions, the contaminations, the degraded samples and through a PANDEMIC. You've taught me so much. You were always gentle with me when you didn't have to and even gentler when I needed you to be. You always found time for me even through major moments in your own life. Not to be dramatic, but I see God in you. This project is result of your hard work. I just want to make you as proud to be my supervisor as I am proud to be your student. T.H.A.N.K Y.O.U.

To Dr G. I see a lot of me in you or rather a lot of what I aspire to be. You are a force of nature. You raise the level of every room you walk in. You speak your mind with conviction and you get things done. I admire you so much. You once said to me: "This degree is going to teach you more about life than about anything else." I have been telling myself that every day. There are so many things that I didn't know I could handle before embarking of this degree, but there is always a way – I have learnt that from you. I could not have done this without you. Thank you especially, for being with me when Dr V was away. I was never alone because of you. This project is also the result of your hard work. I just want to make you proud too.

To my family, all our prayers got us here. To my parents, you are the rock on which ALL this is built on. I am the product of all your sacrifices, all your love, all your ambition. I definitely see God in you. I don't want any of it, if I can't have it with you. To Mai A, Taku and Sekuru A – your support was always with me. You carried

me through the worst times. To my little sister, who carries my heart with her everywhere she goes. Your love is the air I breathe. It has been, is and always will be for you that I do this for.

To all FunGenics lab members. Especially Daniesha and Suheena. You've been there since the beginning. I cherish y'all. To Darisia, Matthew and Duncan for always being willing to help me. I couldn't have done this without you.

To my friends, thank you for always checking on me and not allowing me to go crazy. To Vuyisa and Thomas especially for always believing in me and telling me how great I am even when I didn't feel it. To Kamo for always putting me back together when I fell apart and loving me in ways I didn't know existed. To the moon and all the way back.

A special acknowledgement to the Johnsons. You love me for free. You were always there to tell me how much life sucked when it did and did it with a plate of meat-filled food. You have become my home. I love y'all so much. I could not have done this without you. Thank you.

Finally, to myself. Every day that the sun rose, you woke up and tried your best and because of that, no matter the outcome, we won. I'm proud of you boy.

Table of Contents

Table of Contents 7
List of Tables 10
List of Figures 11
List of Abbreviations 14
Chapter 1: Literature overview 18
1.1 Vitamin D 18
1.1.1 Metabolism 18
1.2 The Vitamin D Receptor 21
1.2.1 VDR Structure 22
1.2.2 *VDR* Gene structure 24
1.3 Serum 25(OH)D_3 levels 24
1.3.1 Defining 25(OH)D_3 Status 24
1.3.2 Serum 25(OH)D_3 Levels and Disease 26
1.4 Epigenetics 29
1.4.1 DNA Methylation 29
1.4.2 The Methylome 31
1.4.3 Methylation Functions 33
1.4.4 DNA Methylation and Disease 36
1.4.5 DNA Methylation and Vitamin D 39
1.4.6 Tissue specificity of DNA methylation 40
1.4.7 Methods for analysing DNA methylation 40
1.5 Hypothesis 42
1.6 Aim 42
1.7 Objectives 42
Chapter 2: Methods and Materials 44
2.1 Cell Culture and Treatment 44
2.2 DNA Extraction 46
2.3 DNA Quantification 47
2.4 DNA Integrity Assessment 48
2.5 Reduced Representation Bisulphite Sequencing 49
2.5.1 Enzymatic Digestion 51

2.5.2 Paired-ends Preparation 51
2.5.3 Adapter Ligation 51
2.5.4 Library Quantification 52
2.5.5 Bisulphite Sequencing 53
2.2 Bioinformatic Analysis 55
2.2.1 MethylKit 57
2.2.2 eDMR 60
2.2.3 Genomic Distribution Analysis 62
2.2.4 Methylomic Distribution Analysis 62
2.2.5 Enriched Differential Methylation Analysis 63
2.2.6 Pathway Analysis 65
Chapter 3: Results 66
3.1 Quality control 66
3.1.1 High-molecular weight gDNA was extracted 66
3.1.2 The pooled RRBS library fragment profile closely matched that of the standard human genome 68
3.1.3 The bisulphite conversion of the sample DNA was efficient and sequenced reads mapped successfully to the reference genome 69
3.2. Differential Methylation Analysis 72
3.2.1 1α,25(OH)$_2$D$_3$ induce differential methylation in regions of the genome associated with gene expression control 72
3.2.2 Minimal technical variability in the methylomes of treated and untreated replicates in both THP-1 and HEK293 cells, respectively 77
3.2.3 1α,25(OH)$_2$D$_3$ induced differential methylation in both THP-1 and HEK293 cells 78
3.2.4 Differential methylation occurred most in open sea CpG sites and intergenic regions. 84
3.2.5 DMC distribution in promoter regions showed enrichment of nervous system and developmental biological processes 89
Chapter 4: Discussion 92
4.1 Quality control 92
4.1.1 High molecular weight gDNA was extracted 92
4.1.2 The pooled RRBS library fragment profile closely matched that of the standard human genome 93
4.1.3 The bisulphite conversion of the sample DNA was efficient and sequenced reads mapped successfully to the reference genome 93
4.2 Differential Methylation Analysis 95

4.2.1 The top differentially methylated sites were in regions with unknown functions and the top differentially methylated regions were in epigenetically-linked DNA. 95

4.2.2 Minimal technical variability in the methylomes of treated and untreated replicates in both THP-1 and HEK293 cells, respectively 96

4.2.3 1α,25$(OH)_2D_3$ induced differential methylation in both THP-1 and HEK293 cells 97

4.2.4 Differential methylation occurred most in intergenic regions and open sea CpG sites. 100

4.2.5 DMC distribution in promoter regions showed the tissue-specific effects of 1α,25$(OH)_2D_3$ treatment on differential methylation 103

Chapter 5: Conclusion 105

References 107

Appendices 126

Appendix A: GABPA Binding Motifs 126

List of Tables

Table 1. 1: Association between serum 25(OH)D_3 levels and disease. 27
Table 1. 2: Population-specific association between serum 25(OH)D_3 levels and disease. 28
Table 1. 3: The relationship between DNA methylation and disease. 38
Table 1. 4: Approaches and methods for investigating DNA methylation.....41

Table 2. 1: qPCR conditions for sample library quantification. 53
Table 2. 2: Thermocycler conditions for bisulphite conversion of library pools. 54
Table 2. 3: Enrichment PCR thermocycler conditions. 54

Table 3. 1: Summary of sample data following bisulfite conversion, sequencing, and alignment.....70
Table 3. 2: Genomic location of 1α,25$(OH)_2D_3$-induced differentially methylated sites in THP-1 cells as identified by methylKit. 73
Table 3. 3: Genomic location of 1α,25$(OH)_2D_3$-induced differentially methylated regions in THP-1 cells as identified by eDMR. 75
Table 3. 4: The number of DMCs detected with different combinations of q-value and methylation difference cut-off. 80

List of Figures

Figure 1. 1: The conversion of vitamin D pro-hormones to their active metabolites through hydroxylation, and the conversion of these active metabolites into waste products through oxidation 20
Figure 1. 2: Schematic representation of the domain structure of the VDR protein. The DNA binding domain is found in the N-terminal, while the ligand binding (E) and dimerization (E1) domains are found in the C-terminal. The complete protein is 427 amino acids long. The largest region is the hinge region (D), which spans 153 amino acids 22
Figure 1. 3: Crystal structure of VDR. The VDR ligand-binding domain is found in the C-terminal. This domain has 12 helices (red) and a 3-layer sheet (yellow). Upon activation, the ligand is tightly bound in the domain 23
Figure 1. 4: Structural organization of the vitamin D receptor gene. The VDR gene has a large regulatory region and a smaller coding region. There are eight coding exons (2-9), indicated by grey bars as well as six non-coding exons (1a-1f) indicated by black bars. The primary promoter is in the middle of the regulatory region. 24
Figure 1. 5: Our current understanding of the human methylome. CGIs (green) are found upstream of the transcription start site (black) and are usually 200-1 000 bases long. CpG shores (brown) directly flank CGIs. CpG shelves and the open sea (blue), found 2 000+ bases upstream of CGIs, are intergenic. 33

Figure 2. 1: Flow diagram showing step wise RRBS procedure. 50
Figure 2. 2: Flowchart showing bioinformatic pipeline rationale. Steps 1-4 were done through a service provider. 55
Figure 2. 3: Flowchart of methylKit features used for data analysis and how they are related to each other. The function (underlined) used for each feature are also included at each step. 59
Figure 2. 4: Workflow of eDMR. Features of the package and functions (underlined) used for each feature to group CpGs into regions and determine differential methylation in those regions.....61

Figure 3. 1: Agarose gel showing intact gDNA. This representative 1% Agarose gel shows distinct gDNA bands extracted from untreated THP-1 (lane 1-2), 1α,25$(OH)_2D_3$-treated THP-1 (lane 3-4), untreated HEK293 (lane 5-6), 1α,25$(OH)_2D_3$-treated HEK293 (lane 7-8and 1α,25$(OH)_2D_3$-treated MCF-7 (lane 9-10) (MWM: molecular weight marker). 67

Figure 3. 2: Fragment analysis confirmed gDNA integrity. The representative electropherogram generated by means of a Fragment Analyzer™ after shipment of samples and prior to sequencing shows a single peak indicating intact gDNA with a molecular size of 22 681 bp.....67
Figure 3. 3: Apart from two small fragments at 224 and 254 bp, the fragments generated after *Msp*1 cleavage of gDNA matched that of the human genome. The electropherogram of RRBS library shows the size and number of fragments after *Msp*1 cleavage. The different sized peaks correspond to the *Msp*1 restriction sites and are species-specific. The peaks visible at 202, 269, 342, 427 and 565 bp correspond to the standard human genome, whereas the peaks at 224 and 254 bp (red) do not. 68
Figure 3. 4: The majority of the CpG sites in all samples showed high (>95%) and low (<10%) levels of methylation with few sites showing intermediate levels of methylation, with no observed PCR bias. Percentage distribution showing a frequency of reads with different levels of methylation for THP-1(A), HEK293(B) and MCF-7(C) cells. The frequency of CpGs with a $\log_{10}$(coverage)>2 was low indicating no PCR bias for THP-1(D), HEK293(E) or MCF-7(F) cells.....72
Figure 3. 5: There was a strong positive correlation between the methylomes of control and treated replicates in THP-1 (A) and HEK293 (B) samples. Scatter plots of percentage methylation values for each replicate in the sample. Numbers on the upper right corner represent pairwise Pearson's correlation scores. The histograms represent the methylation distribution of CpG sites for each replicate.....77
Figure 3. 6: Methylation and demethylation events occurred in THP-1 and HEK293 samples following treatment with 1α,25$(OH)_2D_3$. Volcano plots show differentially methylated cytosines in THP-1 (A) and HEK293 (B) cells and differentially methylated regions in THP-1 (C) and HEK293 (D) cells. Blue dots indicate hypomethylated cytosines/regions and red dots indicate hypermethylated cytosines/regions. Grey dots indicate cytosines/regions excluded by the q-value (0.010) and methylation difference (15%) cut-offs. 82
Figure 3. 7: Total differential methylation percentage per chromosome was higher in HEK293 than in THP-1 cells following 1α,25$(OH)_2D_3$ supplementation. The bar graphs show the percentage of total hypermethylation (purple) and hypomethylation (green) identified by both methylKit and eDMR in THP-1 (A) and HEK293 (B) cells. 83
Figure 3. 8: 1α,25$(OH)_2D_3$ treatment primarily induced differential methylation in open sea CpG sites and CGIs. The bar graphs show the distribution of differentially methylated CpG sites in THP-1 (A) and HEK293 (B) cells and differentially methylated regions in THP-1 (C) and HEK293 (D) cells, across the methylome.....86

Figure 3. 9: There was a co-occurrence of differential methylation in the methylome. Differentially methylated CpG sites, identified in THP-1 (A) and HEK293 (B) cells, located in CpG shores and shelves occurred in regions which housed other CGIs.....87
Figure 3. 10: The distribution of 1α,25$(OH)_2D_3$-induced differential methylation in genomic features was consistent across the chromosomes. The bar graphs show the distribution of DMCs (in THP-1 (A) and HEK293 (B) cells and DMRs (in THP-1 (C) and HEK293 (D) cells), for different genomic features including promoter regions (pink), introns (purple), intergenic regions (orange) and exons (green). 88
Figure 3. 11: Dotplot showing the biological processes whose gene promoter is most enriched for 1α,25$(OH)_2D_3$-induced differential methylation identified by methylKit (THP-1: A, HEK293: B) and eDMR (THP-1: C, HEK293: D). Circles represents the biological processes with genes enriched for 1α,25$(OH)_2D_3$-induced differential methylation. The colour of the circle represents the adjusted p-value. The size of the circle represents the number of genes in the set which are linked to that biological process. The x-axis represents the ratio of genes associated with a biological process over the total number of genes. 91

List of Abbreviations

AF	-	Activation Function
BED	-	Browser Extensible Format
BER	-	Base Excision Repair
BMI	-	Body Mass Index
BS	-	Bisulphite Sequencing
CGI	-	CpG Island
CI	-	Confidence Interval
CTCF	-	CCCTC-binding factor
DBD	-	DNA Binding Domain
DM	-	Differential Methylation
DMCs	-	Differentially Methylated Cytosines
DMRs	-	Differentially Methylated Regions
DNA	-	Deoxyribose Nucleic Acid
DNMT	-	DNA Methyltransferase
EDTA	-	Ethylenediaminetetra-acetic acid
ENCODE	-	Encyclopedia of DNA Elements
FDR	-	False Discovery Rate
GO	-	Gene Ontology
GSEA	-	Gene Set Enrichment Analysis
HEK	-	Human Embryonic Kidney Cells

LBD	-	Ligand Binding Domain
MBD	-	methyl-CpG-binding-domain
MCF	-	Michigan Cancer Foundation-7
OR	-	Odds Ratio
ORA	-	Over-Representation Analysis
PCR	-	Polymerase Chain Reaction
RDA	-	Recommended Daily Allowance
RNA	-	Ribonucleic Acid
RRBS	-	Reduced Representation Bisulphite Sequencing
RXR	-	Retinoid X Receptor
SAM	-	*S*-adenosyl methionine
SBP	-	Systolic Blood Pressure
SLIM	-	Sliding Linear Model
TB	-	Tuberculosis
TC	-	Tissue Culture
TET	-	Ten Eleven Translocation
TSS	-	Transcription Start Site
UV	-	Ultraviolet
VDBP	-	Vitamin D Binding Protein
VDR	-	Vitamin D Receptor
VDRE	-	Vitamin D Response Element

Abstract

Vitamin D is a fat-soluble hormone attained through UV exposure and diet. 1α,25$(OH)_2D_3$, the biologically active form of vitamin D, bound to its nuclear receptor, the vitamin D receptor, regulates gene expression and has been suggested to influence gene regulatory mechanisms such as the epigenome. As an epigenetic mark, DNA methylation functions to maintain genomic stability and regulate gene expression. Both vitamin D deficiency and aberrant DNA methylation marks is often associated with increased disease risk, suggesting a link between the two features. This study therefore aimed to investigate the impact of 1α,25$(OH)_2D_3$ on genome-wide methylation. To this end, genomic DNA was extracted from the THP-1 cells – a highly 1α,25$(OH)_2D_3$ responsive monocytic cell line, MCF-7 breast cancer cells – a cell line known to show weak responsiveness to 1α,25$(OH)_2D_3$, and HEK293 cells – an embryonic kidney cell line with highly active methylation machinery. The cells were supplemented with/without 10 nM 1α,25$(OH)_2D_3$ and subjected to reduced representation bisulfite sequencing. Read alignment was performed using Bismark aligner and differential methylation analysis was performed using MethylKit (to identify differentially methylated cytosines; DMCs) and eDMR (to identify differentially methylated regions; DMRs). Pathway analysis was conducted using clusterProfiler. 1α,25$(OH)_2D_3$ induced differential methylation ($q < 0.01$, at $> 15\%$ difference) in both THP-1 (19 619 DMCs; 52% hypermethylated, 1 744 DMRs; 50.2% hypermethylated) and HEK293 (78 476 DMCs; 53% hypermethylated, 12 352 DMRs; 49.9% hypermethylated) cells. In both THP-1 and HEK293 cells, differential methylation was most concentrated in intergenic and open sea regions and least concentrated in exons. 1α,25$(OH)_2D_3$ induced both methylation and demethylation events that were enriched in pathways specific to monocytic and embryonic cell types, as well as specific to the gene expression requirements of the monocytic and embryonic cell state. MCF-7 samples were found to have been contaminated as *Msp*1 digestion peak profiles did not correspond with human samples and was thus excluded from downstream analyses. Taken together, results from this study suggest that

$1\alpha,25(OH)_2D_3$ is plays an active role in altering methylation patters in a cell-type specific manner.

Chapter 1: Literature overview

1.1 Vitamin D

Vitamin D (a term collectively referring to vitamin D_2, vitamin D_3 and their metabolites) is a fat-soluble steroid hormone that is involved in the regulation of many physiological processes including calcium and phosphate homeostasis, cell differentiation and proliferation, and immune system regulation (Holick, 2003; Lin and White, 2004). The two forms of vitamin D found in the human body are vitamin D_3 and vitamin D_2. The chemical structure of vitamin D_2 differs from that of vitamin D_3 because of a methyl group at C23 and a double bond between C22 and C23 on the vitamin D_2 side chain (Feldman *et al.*, 2008). Although they are structurally different, both vitamin D_2 and D_3 are converted into a similar active metabolite. Vitamin D_2 and D_3 can be described as pro-hormones (Baeke *et al.*, 2010) because they are synthesized or converted to their biologically active forms inside the body.

1.1.1 Metabolism

The discovery of vitamin D started with McCollum *et al.* (1922), who first identified an anti-rachitic substance in cod liver oil and named it vitamin D. Soon after, Steenbock (1924) discovered that irradiation activated a previously inactive anti-rachitic substance in rats. The anti-rachitic substance was vitamin D, and this was the first research linking radiation exposure to vitamin D production. Today we know that the sun is the primary source of vitamin D_3, while vitamin D_2 is plant-derived and accumulates through dietary intake (Feldman *et al.*, 2008).

The metabolism of vitamin D_2 and D_3 is a multistep process, which occurs in multiple locations prior to synthesis of metabolically active 1,25-dihydroxyvitamin D ($1\alpha,25(OH)_2D_3$; Figure. 1.1). Upon UV exposure, 7-dihydrocholestrol is photolytically

cleaved into pre-vitamin D_3 in the skin. Thermodynamically unstable pre-vitamin D_3 is isomerized to vitamin D_3 in a spontaneous reaction where double bonds are rearranged (Holick, 2003; Wacker and Holick, 2013). Vitamin D_3 is then carried by vitamin D binding protein to the liver, where the first hydroxylation reaction occurs. This reaction is carried out by cytochrome P450 enzymes: CYP24A1 all around the body and CYP2R1 in the liver and testes (Bikle, 2014). This step is vital because it produces 25-hydroxyvitamin D (25(OH)D), the main circulating form of vitamin D (Baeke *et al.*, 2010).

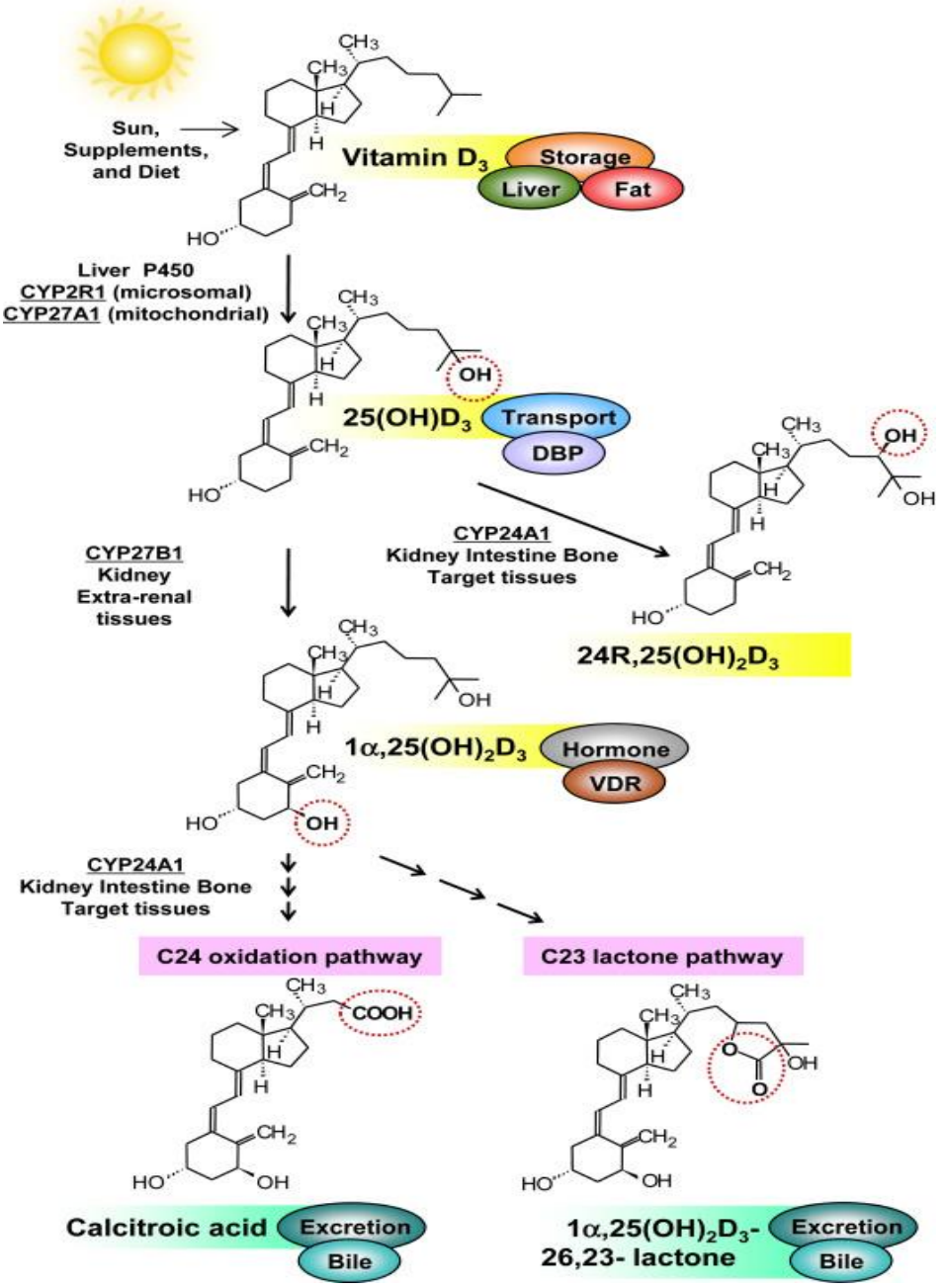


Figure 1. 1: The conversion of vitamin D pro-hormones to their active metabolites through hydroxylation, and the conversion of these active metabolites into waste products through oxidation (Taken from Jones *et al.*, 2018).

25(OH)D is then bound by vitamin D binding protein (VDBP) and transported around the body via the blood. 25(OH)D is transported to the kidneys where it is converted to 1α,25$(OH)_2$D. CYP27B1 is the only enzyme capable of this conversion because of its 25(OH)D 1α-hydroxylase activity (Bikle, 2014). CYP27B1 is found in vitamin D target tissues such as monocytes, macrophages, and epithelial cells from the skin, breast and intestine (Bikle, 2014). CYP27B1 differs from hepatic 25-hydroxylases because it is tightly regulated by parathyroid hormone, fibroblast growth factor and 1α,25$(OH)_2D_3$ itself (Bikle, 2014). VDBPs have a higher affinity for 25$(OH)D_3$ than 25$(OH)D_2$ (Armas *et al.*, 2004) and therefore mostly 25$(OH)D_3$ is hydroxylated into 1α,25$(OH)_2D_3$ by CYP27B1. 25$(OH)D_2$ is rapidly catabolised by CYP24A1, a 24-hydroxylase (Bikle, 2014). CYP24A1 is also responsible for catabolisation of 1α,25$(OH)_2$D. 1α,25$(OH)_2$D binding is preferred to 25(OH)D, however both are hydroxylated into 1,24,25$(OH)_3$D and 24,25$(OH)_2$D, respectively, by CYP24A1 (Sakaki *et al.*, 2000). 24-hydroxylation of 1α,25$(OH)_3$D leads to the eventual formation of calcitroic acid, which is excreted in the bile (Jones *et al.*, 2018). Preventing the accumulation of 1α,25$(OH)_2$D and 25(OH)D to toxic levels is the primary function of CYP24A1 and therefore the *CYP24A1* gene is strongly induced by 1α,25$(OH)_2$D (Bikle, 2014).

1.2 The Vitamin D Receptor

The vitamin D receptor (VDR) is a classical transcription factor from the nuclear receptor family (Mangelsdorf *et al.*, 1995). Members of the nuclear receptor superfamily share a common, structurally conserved DNA binding domain (DBD) in their variable N-terminals, and a ligand-binding domain (LBD) in their semi-conserved C-terminals (Mangelsdorf *et al.*, 1995). The nuclear receptor superfamily members differ from other DNA binding proteins as they have two highly conserved zinc fingers in the DBD (Berg, 1989). VDR distinguishes itself from other nuclear receptor family members in that it can be activated by nanomolar concentrations of its steroid ligand,

$1\alpha,25(OH)_2D_3$. VDR shares this trait with only a few other nuclear receptors, such as progesterone and estradiol receptors. Other members of the nuclear receptor family are activated by micromolar ligand concentrations.

1.2.1 VDR Structure

Nuclear receptors have a ligand-independent activation function domain (AF-1) in the N-terminal and a ligand-dependent activation function domain (AF-2) in the C-terminal (Figure. 1.2). VDR was originally believed to not contain an AF-1 (Sone *et al.*, 1991), however it has since been shown that not only does VDR have this domain, but this AF-1 domain is important in the repressive action of VDR (Alimirah *et al.*, 2010). The C-terminal has 12 helices, which form the dynamic ligand-binding domain (Rochel *et al.*, 2000; Figure. 1.3). The AF-2 domain undergoes a conformational change as a result of $1\alpha,25(OH)_2D_3$ binding to the LBD, such that the 12^{th} helix becomes tightly packed into the body of the LBD (Rochel *et al.*, 2000). VDR function occurs through two protein interacting surfaces (Smith and O'Malley, 2004). One of the surfaces is for dimerization with a partner for specific DNA binding, and the other is for binding secondary co-regulatory proteins for gene regulation.

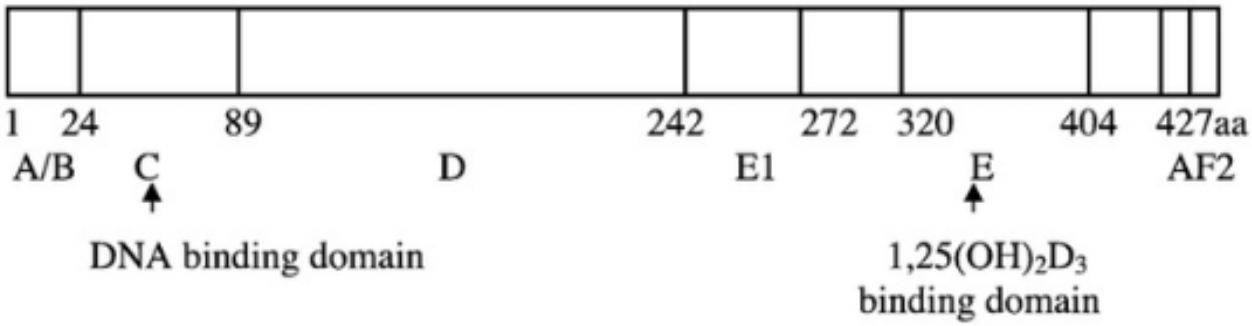


Figure 1. 2: Schematic representation of the domain structure of the VDR protein. The DNA binding domain is found in the N-terminal, while the ligand binding (E) and dimerization (E1) domains are found in the C-terminal. The complete protein is 427 amino acids long. The largest region is the hinge region (D), which spans 153 amino acids (Taken from Ramasamy, 2008).

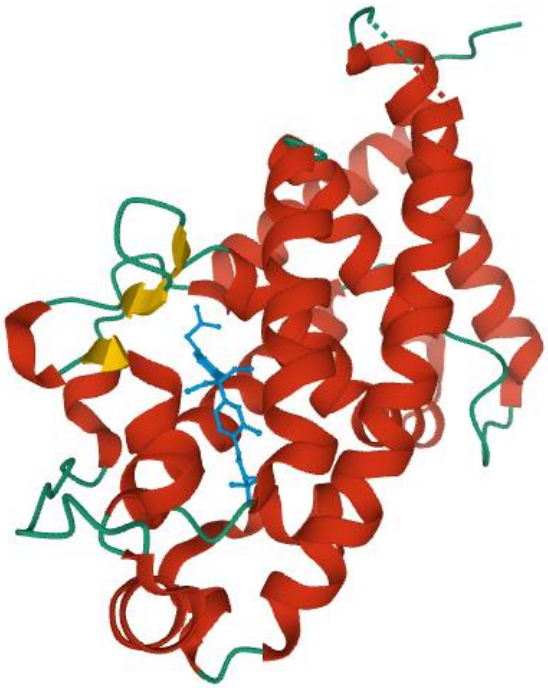


Figure 1. 3: Crystal structure of VDR. The VDR ligand-binding domain is found in the C-terminal. This domain has 12 helices (red) and a 3-layer sheet (yellow). Upon activation, the ligand is tightly bound in the domain (Taken from (Shaffer and Gewirth, 2002).

VDR binds to specific DNA sequences called Vitamin D Response Elements (VDREs). VDREs either contain a direct repeat of two hexanucleotide half-elements spaced by three nucleotides (DR3), or two half-element repeats spaced by six nucleotides (Haussler *et al.*, 1998). VDR can bind to VDREs alone, or as a heterodimer with other transcription factors (Seuter *et al.*, 2014). The optimal VDRE was found to be a direct repeat of two six-base half-elements, with VDR binding on the 5' side of the element and the RXR binding at the 3' end (Colnot *et al.*, 1995). The 1α,25$(OH)_2D_3$-activated VDR-RXR complex recruits co-modulators, regulating gene expression by bringing the transcription start sites near VDR-RXR bound enhancers through DNA looping. Over 22 000 VDREs have been identified in the human genome (Tuoresmäki *et al.*, 2014), in addition to 229 genes which have been shown to be differentially expressed in

response to vitamin D (Ramagopalan *et al.*, 2010). This supports the biological significance of vitamin D signalling through VDR. Additionally, VDR is ubiquitously expressed in the body suggesting that 1α,25(OH)$_2$D$_3$ can alter gene expression in a variety of tissues.

1.2.2 *VDR* Gene structure

VDR mRNA is transcribed from the negative strand of chromosome 12 (Chr12, 47 841 537 – 47 943 048). *VDR* spans 101 512 base pairs, including the exons, introns and regulatory region. The coding region is minor compared to the non-coding region and consists of eight coding exons. 63% of *VDR* is the 5' regulatory region, which consists of the primary promoter and six non-coding exons (1a-1f; Figure 1.4).

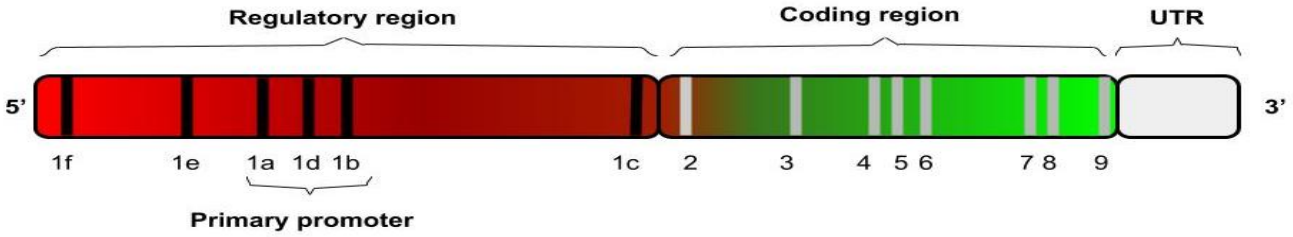


Figure 1. 4: Structural organization of the vitamin D receptor gene. The VDR gene has a large regulatory region and a smaller coding region. There are eight coding exons (2-9), indicated by grey bars as well as six non-coding exons (1a-1f) indicated by black bars. The primary promoter is in the middle of the regulatory region.

1.3 Serum 25(OH)D$_3$ levels

1.3.1 Defining 25(OH)D$_3$ Status

Serum 25(OH)D$_3$ has a longer half-life (approximately three weeks) than 1α,25(OH)$_2$D$_3$ (4-15 hours) and is therefore used to measure vitamin D levels in the body (Mailhot, 2015). The levels of 25(OH)D$_3$ used to define deficiency and sufficiency are

controversial, however < 50 nmol/L (Dawson-Hughes *et al.*, 2005) is generally regarded as the international standard for deficiency. Sun exposure is the biggest source of UV. The skin is equipped with mechanisms for protection against UV radiation (Slominski and Postlethwaite, 2015). Melanin-formation is one such mechanism. Melanin absorbs UV rays (Costin and Hearing, 2007) thus competing with 7-dihydrocholestrol for UVB (Armas *et al.*, 2007). Skin pigmentation may therefore influence $25(OH)D_3$ status (Bonilla *et al.*, 2014; Kauffman *et al.*, 2006; Meyer and Bornman, 2018; Xiang *et al.*, 2015). This is supported by studies that show $25(OH)D_3$ status and response to supplementation differ between ethnic groups (Alzaman et al., 2016; Gopal-Kothandapani et al., 2019; Gutiérrez et al., 2011; Parva et al., 2018; Zhang et al., 2019). Furthermore, vitamin D supplementation, according to recommended daily allowance (RDA) guidelines, does not result in uniform $25(OH)D_3$ status in children from different ethnic groups (Newton *et al.*, 2019). The lack of uniform response can be attributed to genetic variation between populations (Newton *et al.*, 2019).

$25(OH)D_3$ status is established through a combination of physiological, genetic, environmental and lifestyle factors (Fetahu et al., 2014). Therefore, the $25(OH)D_3$ levels used to determine sufficiency and deficiency for any given population should account for these factors. Furthermore, vitamin D supplementation should be adjusted according to how these various factors vary between populations (Herrmann et al., 2017). The complex nature with which $25(OH)D_3$ status is established is a major obstacle to the accurate assessment of the effect of $25(OH)D_3$ status on disease risk (Papadimitriou, 2017). It is possible that what is considered deficient in one population, is sufficient in another. Moreover, it is difficult to discern whether $25(OH)D_3$ deficiency is the cause or result of disease (Aydin et al., 2014). Studies aimed at assessing the effect of $1\alpha,25(OH)_2D_3$ on cellular function at a molecular level are thus essential to our understanding of the role of vitamin D in health.

1.3.2 Serum 25(OH)D_3 Levels and Disease

The scourge of rickets in the 19th century brought attention to the importance of a vitamin D adequate lifestyle. This lifestyle mostly involved abundant exposure to the sun and increased intake of vitamin D enriched milk. For many years, 25(OH)D_3 deficiency was associated exclusively with bone-deforming disease due to its classical role in calcium and phosphorus absorption. With the discovery that VDR is found ubiquitously in the body and that 1α,25$(OH)_2D_3$ has non-classical physiological roles, 25(OH)D_3 deficiency has been associated with many diseases including cancer, cardiovascular disease, and neurological disorders (Table 1.1). In addition, insulin resistance, body fat percentage, high BMI levels and Type 2 diabetes mellitus have also been negatively correlated with 25(OH)D_3 levels (Lim *et al.*, 2013; Vimaleswaran *et al.*, 2013).

The association of serum 25(OH)D_3 levels with chronic disease is especially important in an African context. Data on the 25(OH)D_3 status of African populations is scarce, however several studies have found 25(OH)D_3 deficiency to be prevalent in African populations, (Green *et al.*, 2015; Meddeb *et al.*, 2005) including South Africa (Durazo-Arvizu *et al.*, 2014; Naidoo *et al.*, 2018). This is important because it suggests that African populations may be more vulnerable to chronic diseases due to 25(OH)D_3 deficiency. In South Africa, where infectious and chronic diseases such as tuberculosis, hypertension and diabetes are prevalent (Peltzer and Phaswana-Mafuya, 2013; Pheiffer *et al.*, 2018), it is especially important to investigate the mechanisms by which 1α,25$(OH)_2D_3$ may increase the risk of disease. The importance of investigating these mechanisms is supported by the inconsistent association of 25(OH)D_3 levels with specific diseases when comparing populations from diverse geographical locations.

Table 1. 1: Association between serum 25(OH)D_3 levels and disease.

Disease	Prostate cancer	Colorectal cancer	Ischemic heart disease	Depression	Blood pressure
Sample population	Finnish (Finnish)	American Finnish (European)	Danish (European)	Netherlands (European)	African (South Africa)
Result	Prostate cancer risk was found to be inversely proportional to 25(OH)D_3. *OR = 6.3; (P<0.01)*	Vitamin intake associated with 50% lower risk of colorectal cancer. OR = 0.50; *P<0.02*	Low 25(OH)D_3 increased risk of heart disease. 95% CI; *P<0.01*	25(OH)D_3 was lower in individuals with minor and major depression and depression severity was associated with decreased 25(OH)D_3. 95% CI; *P<0.001.*	Women with low 25(OH)D levels were ~2 times more likely to have high SBP (95% CI: 3.23;1.05) and there was an inverse association between SBP and 25(OH)D (*P=0.04*)
Reference	(Ahonen *et al.*, 2000)	(Gorham *et al.*, 2005)	(Brøndum-Jacobsen *et al.*, 2012)	(Hoogendijk *et al.*, 2008)	(Kruger *et al.*, 2013)

Table 1. 2: Population-specific association between serum 25(OH)D_3 levels and disease.

Disease	Tuberculosis					
Demographic	• African (South Africa) • Median age of 8 months	• African (Botswana) • Under 2 years old	• African (Egypt) • Average age 36 (control) and 35 (patients) years.	• Africa (Guinea-Bissau) • Combined average age of 37.3 years. • 61% males, 39% females	• Asia (Indonesia) • Age range of 18-50 years	• Asia (India) • Females 31.6%, males 68.1%
Finding	• Low 25(OH)D_3 associated with latent TB (95% CI; p = 0.046).	• No significant difference in 25(OH)D status between children under 2 years old with and without active TB.	• Vitamin D deficiency prevalent in patients compared to control individuals (P<0.0001)	• Vitamin D deficiency more common and severe in control individuals compared to patients.	• No significant difference in distributions of serum 25(OH)D_3 between patients and healthy controls	• Lower levels of serum 25(OH)D_3 in HIV negative *(p = 0.002)* and positive *(p = 0.012)* TB patients compared to healthy controls.
Reference	(Gupta *et al.*, 2016)	(Ludmir *et al.*, 2016)	(Mahmoud and Ali, 2014)	(Wejse *et al.*, 2007)	(Grange *et al.*, 1985)	(Gupta *et al.*, 2018)

1.4 Epigenetics

Watson and Crick were the first to propose that the DNA structure has properties which suggest possible modification (Watson and Crick, 1953), and unknowingly, were the first to acknowledge what we now know as epigenetics. The concept of epigenetics is understood as mechanisms of heritable chemical and/or structural modifications to the genome without altering the primary gene sequence (Crabtree, 2020). These modifications occur anywhere in the genome and change in response to the environment. These modifications are important because they regulate the transcription of genes and thus affect all cellular processes including cell signalling, embryogenesis, development and aging (Crabtree, 2020). DNA methylation, histone modifications, the 3D organisation of the genome, RNA modification, and non-coding RNA are all forms of epigenetic modifications.

1.4.1 DNA Methylation

DNA methylation is the most studied epigenetic modification. DNA methylation involves the enzymatic addition of a methyl group to the 5' carbon of the cytosine ring to form 5-methylcytosine (5-mC; Moore *et al.*, 2013). DNA methylation was first identified in the early 20th century (Avery *et al.*, 1944). DNA methylation is performed by DNA methyltransferases (DMNTs), which transfers a methyl group from *S*-adenosyl methionine (SAM) to carbon-5 of cytosine in specific DNA sequences (Moore *et al.*, 2013). In mammals, methylation commonly occurs at a cytosine nucleotide followed by a guanine nucleotide (Moore *et al.*, 2013). This is known as a CpG site. The mammalian genome has a GC content of 42% (Lander *et al.*, 2001). From that percentage, the percentage of CpG dinucleotides should be 4.41% (21% cytosine * 21% guanine). However this percentage is closer to 1% (Stevens *et al.*, 2013). Early researchers believed that this was due to 5-mC spontaneously deaminating to thymine over time (Bird, 1980). We now know that mammals have evolved complex

mechanisms to maintain DNA integrity. One such mechanism is base excision repair (BER) (Krokan *et al.*, 2002). BER reduces the mutagenic threat of cytosine deamination. Addition of a methyl group increases the stability of the cytosine-guanine base pair (Klump and Löffler, 2009). The carbon-carbon bond between the methyl group and cytosine has a high bond strength making 5-mC biologically stable (Klungland and Robertson, 2017). Therefore, it is unlikely that spontaneous deamination is responsible for the low CpG dinucleotide percentage.

DMNT1 is referred to as the "maintenance DNMT" (Moore *et al.*, 2013) because it maintains the DNA methylation patterns from one cellular generation to the next. DNMT1 binds to the newly-synthesized strand of DNA, where it methylates the daughter strand according to the patterns on the parental strand (Hermann *et al.*, 2004). DNMT2 has sequence and structural homology with the other DNMT enzymes (Dong *et al.*, 2001). DNMT2 was found to have weak DNA methylation activity (Jeltsch *et al.*, 2006) and instead found to be an RNA methyltransferase (Goll *et al.*, 2006). It has been theorized that DNA methyltransferases could have evolved from a prokaryotic RNA methyltransferase ancestor similar to DNMT2 (Jeltsch *et al.*, 2006). However, phylogenetic analysis has instead shown that DNMTs are not closely related to other RNA methyltransferases and evolved separately from ancestral prokaryotic DNA methyltransferases (Jeltsch *et al.*, 2017). DMNT3s are known as *de novo* DMNTs (Moore *et al.*, 2013). DNMT3a and DNMT3b can establish new methylation patterns on unmodified DNA (Okano *et al.*, 1999). This *de novo* methylation mostly occurs during embryonic development (Iurlaro *et al.*, 2017) and in response to cellular signalling and environmental cues. DNMT1, DNMT3a and DNMT3b work in unison in order to establish and maintain methylation patterns in an organism from as early as embryonic development (Kim *et al.*, 2002).

The process of DNA demethylation is less well-understood. This is mostly due to the presence of numerous pathways that can result in demethylation. The high stability of 5-mC means that the direct removal of the methyl group is not thermodynamically

favourable. It is known that there are rapid and dramatic demethylation events during the reprogramming of the paternal genome after fertilization (Mayer et al., 2000; Reik et al., 2001). Similarly, demethylation is a crucial process during the development of female and male primordial germ cells (Smith and Meissner, 2013).. This is an example of passive demethylation or dilution of the 5-mC signal, which occurs through the formation of de novo unmethylated DNA during cell proliferation. It is important to note that while passive demethylation explains reduced global methylation, it does not include the modification or active removal of the 5-mC epigenetic marker.

Active demethylation occurs when 5-mC is oxidised, removed and replaced through DNA repair mechanisms. Here, 5-mC is oxidised to 5-hydroxymethylcytosine (5-hmC), which is then oxidised to 5-formylcytosine (5-fC), then to 5-carboxylcytosine (5-caC) along a pathway catalysed by Ten-Eleven Translocation (TET) enzymes (Wu and Zhang, 2017). Thymine DNA glycosylase (TDG) recognizes and excises 5-fC and 5-caC, followed by the insertion of an unmodified cytosine through DNA repair (Wu and Zhang, 2017). Active demethylation is important for many biological processes. Significantly, the maternal and paternal genomes in zygotes undergo epigenetic reprogramming that includes demethylation (Lee *et al.*, 2014). While the maternal genome is demethylated through passive dilution of 5-mC, the paternal genome is demethylated by both passive 5-mC dilution and active TET-TDR demethylation events (Wu and Zhang, 2017).

1.4.2 The Methylome

CpG Islands (CGIs) are 200-1 000 base pair stretches of DNA with a high concentration of CpG nucleotides, GC content of over 50% and an observed/expected CpG ratio of over 60% (Figure. 1.5; Sandoval *et al.*, 2011). Promoter regions and primary exons are often found in these CGIs (Illingworth and Bird, 2009; Saxonov *et al.*, 2006). 75% of CpG sites in mammalian somatic cells are methylated (Tost and Gut, 2007), suggesting that methylation is the favoured state for CpG dinucleotides.

This makes sense considering the stability of 5-mC. However, CGIs are largely unmethylated in mammalian genomes (Cooper *et al.*, 1983; Illingworth and Bird, 2009).

There are many plausible causes for this. Promoter CGIs may be unmethylated to allow for gene expression. Expression induced by transcription factor binding at these promoter regions during development (Morgan *et al.*, 2005) may lead to protection from *de novo* methylation. CXXC is a Zn-chelating DNA binding domain, which discriminately binds to unmethylated CpG dinucleotides (Voo *et al.*, 2000). Cfp1, a CXXC domain containing protein, recruits histone 3 lysine 4 methyltransferases (Thomson *et al.*, 2010). H3K4me3 is a common marker for transcription start sites (Guenther et al., 2007a) and is associated with an increase in expression. Mechanistically, the methylation state of promoter CGIs can be linked to gene expression. Furthermore, H3K9me3 interferes with DNA methyltransferase activity through inhibition of chromatin binding (Ooi *et al.*, 2007). Aberrant methylation of CGIs is a common feature in tumours (Feltus *et al.*, 2003; Sproul and Meehan, 2013; Ueki *et al.*, 2002).

Because DNA methylation (and lack thereof) is a common marker for gene regulation, aberrant CGI methylation is crucial to cancer progression by inactivating tumour suppressor genes or up-regulating oncogenes (Jones and Baylin, 2002). CpG shores are regions directly flanking CGIs (up to 2 000 base pairs from CGI) (Sandoval *et al.*, 2011). CpG shores are differentially methylated in a tissue-specific manner (Irizarry *et al.*, 2009). Furthermore, this methylation is conserved and correlates with tissue-specific gene expression (Irizarry *et al.*, 2009). CpG shelves are regions 2 000 to 4 000 base pairs away from CGIs and these regions are intergenic (Sandoval *et al.*, 2011). The open sea refers to isolated CpG sites in the genome (Sandoval *et al.*, 2011).

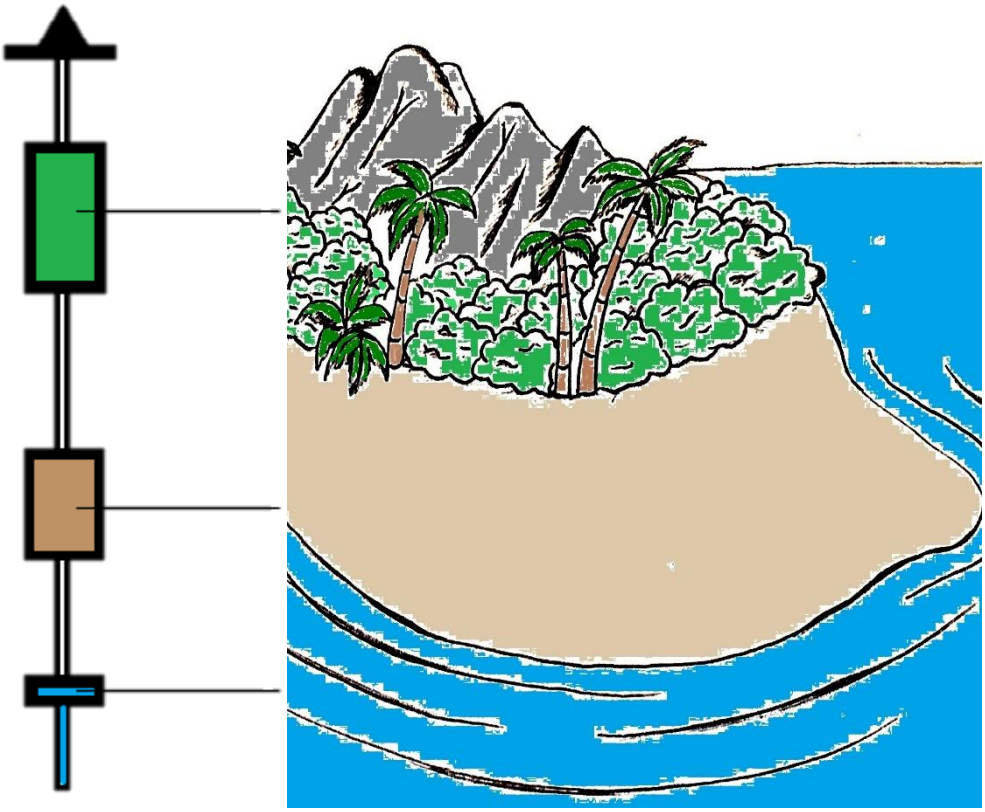


Figure 1. 5: Our current understanding of the human methylome. CGIs (green) are found upstream of the transcription start site (black) and are usually 200-1 000 bases long. CpG shores (brown) directly flank CGIs. CpG shelves and the open sea (blue), found 2 000+ bases upstream of CGIs, are intergenic.

1.4.3 Methylation Functions

DNA methylation is an epigenetic modification with multiple biological functions (Tirado-Magallanes *et al.*, 2016). Transcription factor binding sites are often GC rich and hence most promoters and transcription start sites are contained in CGIs (Saxonov *et al.*, 2006). DNA methylation therefore serves to modulate the interaction between promoter sequences and the transcriptional machinery. Although DNA methylation in

promoter regions is often associated with gene silencing, existing research has not determined the sequential timing of these two events. Previous studies have shown that methylation of a single cytosine at the centre of a transcription factor binding site not only prevented binding of some transcription factors, but also prevented transcription initiation by those transcription factors (Watt and Molloy, 1988). Actively expressed genes have promoters in nucleosome-deprived regions (Kelly *et al.*, 2010), and *de novo* methylation by DNMTs is dependent on the structural organization and methylation state of histones within a nucleosome (Ooi *et al.*, 2007). This suggests that in some cases, gene silencing precedes methylation. In this model, DNA methylation cements silencing events, thus securing the inheritance of epigenetic changes (You *et al.*, 2011).

Methylated DNA is also recognized by methyl-CpG-binding-domain (MBD) containing proteins. These methyl-CpG binding proteins (MeCP) can recruit histone deacetylases (Nan *et al.*, 1998) to form compact and closed chromatin. Histone acetylation is a common feature of active gene expression, whereas histone methylation mostly leads to gene silencing (Rice and Allis, 2001). MeCPs can bind and recruit histone deacetylases in favour of H3 lysine 9 (H3K9) methylation (Fuks *et al.*, 2003). Additionally, methylation binding proteins can bind to polycomb group proteins (Matarazzo *et al.*, 2007), which target H3 lysine 27 (H3K27) for methylation (Wiles and Selker, 2017). Histone modifications are well characterized as marks of gene silencing and H3K9/27 specifically leads to the spread of heterochromatin (Hyun *et al.*, 2017). This emphasizes how DNA methylation regulates gene expression through interaction with many different proteins, and how DNA methylation can not only initiate and but fortify gene silencing marks.

DNA methylation can also silence gene expression by physically blocking transcription factor binding, however, this only applies to methylation-sensitive transcription factors. Some transcription factors not only bind to methylated CpGs, but prefer binding to these methylated regions (Yin *et al.*, 2017). This highlights the complexity

of the gene regulation mechanisms of DNA methylation and shows that DNA methylation in promoter regions does not always denote gene silencing. Interestingly, DNA methylation of intragenic CGIs may lead to silencing. A great example, which also highlights a protective function of DNA methylation, is that methylation is one of the mechanisms through which repetitive, non-coding DNA sequences are repressed (Arand *et al.*, 2012; Walsh *et al.*, 1998).

The methylation state of CpG shelves and shores in the methylome is less documented in comparison to CGIs. A study explored the relationship between CpG shore methylation and high gene expression (Edgar *et al.*, 2014). The study identified lowly methylated CGIs that are flanked by highly methylated CpG shores (Edgar *et al.*, 2014). These regions, called ravines, are distinguishable from standard CGIs because they are found near constitutively expressed genes (Edgar *et al.*, 2014). This suggests a link between CpG shore methylation and gene expression levels. CpG shore methylation is an area of increased interest in disease. Methylation changes in colon cancer were identified in CpG shores, and not in promoter or CGI regions as expected (Irizarry *et al.*, 2009). The location of these methylation changes in colon cancer overlaps with the location of tissue-differentiating methylation changes in normal tissue, which suggests that CpG shore methylation plays a role in tissue-differentiation (Irizarry *et al.*, 2009) as well as disease progression. Furthermore, in colorectal cancer stem cells, CpG shore, shelf and open sea hypomethylation can be used to determine the stability of microsatellites (Visone *et al.*, 2019).

Methylation can also occur in the gene body. The function of gene body methylation is a vast area of interest in the field of epigenetics. Gene body methylation has been correlated with an increase in gene expression (Kulis *et al.*, 2012; Varley *et al.*, 2013; Yang *et al.*, 2014). Gene body methylation has also been found to prevent intragenic transcription by repressing intragenic promoter activity (Maunakea *et al.*, 2010). As research into gene-body methylation has grown, it has been shown that the relationship between gene-body methylation and gene expression levels is not linear nor one-

dimensional, but rather complex and dynamic. In plants and invertebrates, intragenic methylation does not correlate with transposon methylation (Zemach *et al.*, 2010). Furthermore, the relationship between intragenic methylation and gene expression levels in *Arabidopsis* follows a bell-shape where genes expressed at a high or low levels are poorly methylated, and genes expressed moderately are highly methylated (Zilberman *et al.*, 2007). In humans, the relationship between intragenic methylation and gene expression differs according to the tissue type. Gene body methylation correlated positively with gene expression in proliferative tissues, and did not differ between active and inactive genes in less proliferative tissues (Aran *et al.*, 2011). There may be other functions for gene body methylation. For example, studies have shown that exons have a higher methylation state than introns (Laurent *et al.*, 2010), suggesting that methylation may play a role in splicing.

Enhancers are short regions of DNA, found anywhere in the genome, that increase the likelihood of a gene being expressed (Shlyueva *et al.*, 2014). Enhancers associate with DNA looping proteins, which can manipulate the three-dimensional space in order to bring transcription up-regulators in close proximity with certain genes (Petrascheck *et al.*, 2005). The random nature of enhancer locations means that CpG-containing enhancers can be methylated regardless of their position in the genome. Enhancers can be hypermethylated, reducing chromatin accessibility and transcription factor binding (Thurman *et al.*, 2012), or hypomethylated, resulting in upregulated gene expression and accessible chromatin (Shull *et al.*, 2016) as seen in some diseases (Aran *et al.*, 2011; Li *et al.*, 2019). CpG rich enhancers also have different regulatory characteristics in comparison with their CpG poor counterparts (Bell and Vertino, 2017). DNA methylation can therefore also regulate gene expression through enhancer interactions with the gene regulating machinery.

1.4.4 DNA Methylation and Disease

Because DNA methylation events commonly occur in CGIs, which house transcription start sites and promoters, the relationship between DNA methylation and disease is through gene expression. Aberrant methylation leads to either silencing or over-expression of genes, which in turn leads to the disease state (Table 1.3). Additionally, DNA methylation machinery such as DNMTs and TETs are targets for mutation in disease (Feinberg *et al.*, 2016). This highlights the importance of the relationship between DNA methylation and disease. DNA methylation changes can be used as hallmarks for identifying disease progression.

Table 1. 3: The relationship between DNA methylation and disease.

Disease	Rheumatoid Arthritis	Diabetes	Tuberculosis	
Epigenetic mechanism	Differential methylation in the CpG associated with HLA-DBQR2.	Increased DNA methylation of CpG sites of the Insulin promoter.	Hypermethylation in genomic loci of infected macrophages.	DNA hypermethylation of the IL-2/STAT5, TNF/NF-ϰB, and IFN-γ signaling pathways in TB patients compared to asymptomatic patients.
Effect	Changes in basal expression, expression in response to antigen and alternative splicing of HLA-DBQR2.	Reduced expression of Insulin. Insulin regulates the blood sugar levels.	Gene expression modulation which changes immune response and chromatin reorganization.	Decreased immune responsiveness and decreased mitogen-induced upregulation of IFN-γ, TNF, IL-6, CXCL9, CXCL10, and IL-1β production.
Reference	(Liu *et al.*, 2013)	(Yang *et al.*, 2011)	(Sharma *et al.*, 2016)	(DiNardo *et al.*, 2020)

1.4.5 DNA Methylation and Vitamin D

DNA methylation also has a relationship with vitamin D. The vitamin D pathway itself is epigenetically regulated. Genes involved in vitamin D metabolism and function such as *CYP2R1*, *CYP24A1, CYP27B1* and *VDR*, have CGIs in their promoter regions and thus expression can be regulated through DNA methylation (Karolchik *et al.*, 2014). Methylation of these genes influences the response to vitamin D supplementation (Chen *et al.*, 2020). Furthermore, DNA methylation-induced silencing of vitamin D-target genes could result in the disruption of $1\alpha,25(OH)_2D_3$ function in the body. This is of importance with regards to disease. For example, $1\alpha,25(OH)_2D_3$ enhances innate immunity through expression of antimicrobial peptides, increasing resistance to *Mycobacterium tuberculosis* (Selvaraj *et al.*, 2015). TB patients often have low levels of $1\alpha,25(OH)_2D_3$ and $25(OH)D_3$ (Venturini *et al.*, 2014; Wang *et al.*, 2018), as well as differential CpG site methylation in vitamin D genes (Wang *et al.*, 2018). This suggests that the methylation status of vitamin D pathway genes is of particular importance in the infection of *M. tuberculosis*. Vitamin D can also influence the methylation of genes not related to the vitamin D pathway. Vitamin D supplementation altered DNA methylation in mothers and infants in genes involved in cell migration, cellular membrane function, cadherin signalling and immune function (Anderson *et al.*, 2018). Vitamin D status during development also influences DNA methylation across the genome (Xue *et al.*, 2018). Vitamin D deficiency has further been associated with differential methylation of immune-related genes responsible for the development and maturation of the thyroid axis (Zhu et al., 2013). In support, vitamin D supplementation was associated with an increase in global leucocyte methylation (Zhu et al., 2016). There is increasing evidence suggesting that vitamin D largely contributes to genome methylation changes, however, the exact mechanisms remain unknown.

1.4.6 Tissue specificity of DNA methylation

DNA methylation plays an important role in the regulation of biological processes. Although DNA methylation patterns are established during embryonic development, these patterns can change due to age, nutrition and environmental factors (Dmitrijeva et al., 2018; Lin et al., 2020; Slieker et al., 2018; Wang et al., 2020). The changes to DNA methylation occur in tissue-specific manner (Ghosh et al., 2010; Lokk et al., 2014). Interestingly, tissue-specific differentially methylated regions are not uniformly associated with gene expression. Tissue-specific differential methylation can classically lead to gene silencing of certain genes however, differential methylation of the same DNA elements, in a different tissue, can lead to increased gene expression. This is supported by a study which found that differential methylation of Spalt-Like Transcription Factor (Sall1) differs between brain and retina tissue and that this correlated to the difference in *Sall1* expression (Wan et al., 2015). This suggests that the gene expression effect of DNA methylation is tissue specific. Additionally, tissue-specific differentially methylated regions are often associated with tissue specific regulatory elements. This is highlighted by the finding that up to 37% of tissue-specific DNA methylation patterns are conserved across species (Zhou et al., 2017).

1.4.7 Methods for analysing DNA methylation

There is a myriad of methods for quantifying DNA methylation level. Following detection or conversion, treatment can be coupled to various detection methods including PCR based methods and sequencing. Table 1.4 describes the sequencing-based methods commonly used for whole genome approaches. Each approach has its advantages and disadvantages. Notably, Reduced Representation Bisulphite Sequencing (RRBS) reduces the cost for whole genome methylation analysis by specifically quantifying methylation at CpG-rich regions in the genome where differential methylation is most likely to take place (Eckhardt et al., 2006).

Table 1. 4: Approaches and methods for investigating DNA methylation

Approach	Method	Rationale	Available bioinformatic tools
Restriction-enzyme based	Methylation-sensitive restriction enzyme sequencing	Following restriction enzyme digestion, DNA fragments are selected and sequenced. Locations of unmethylated cytosines are detected (Barros-Silva et al., 2018).	MeDUSA for normalizing methylation calls (Wilson and Beck, 2016)
Antibody based	MeDIP-seq	Methylated DNA is isolated by methylcytosine-recognizing antibodies and sequenced (Yong et al., 2016).	
Bisulphite-conversion based	Whole genome bisulphite sequencing	Bisulphite treatment converts unmethylated cytosines to uracil whilst methylated cytosines remain intact. Treated DNA is then sequenced to find location of methylated cytosines (Barros-Silva et al., 2018).	methylKit and eDMR make use of logistic regression to determine differential methylation. BSmooth and Biseq use a smoothing approach. DSS, MOABS, RADMeth, methysig use a beta binomial approach.
	Reduced representation bisulphite sequencing	In addition to bisulphite sequencing, CpG-rich regions are selected following cleavage with *Msp*1, a methylation insensitive restriction enzyme (Yong et al., 2016).	

1.5 Hypothesis

It is evident that vitamin D regulates gene expression ubiquitously in the body. This gene regulation occurs through the interaction of VDR with gene regulators. The vitamin D pathway itself is regulated by DNA methylation, and it is possible that $1\alpha,25(OH)_2D_3$ and VDR use DNA methylation as a tool for gene regulation. This complex relationship between vitamin D and methylation may significantly influence cellular function and thus disease susceptibility. Here, we hypothesize that $1\alpha,25(OH)_2D_3$ influences CpG methylation at a genomic scale, thereby facilitating changes in gene expression in a tissue-specific manner.

1.6 Aim

The aim of the study was to assess the impact of $1\alpha,25(OH)_2D_3$ on genome-wide methylation in the highly $1\alpha,25(OH)_2D_3$-responsive THP-1 monocytic cell line, as well as the HEK293 embryonic kidney cell line.

1.7 Objectives

1. Establish monocytic THP-1 and embryonic HEK-293 cell cultures to be supplemented with or without $1\alpha,25(OH)_2D_3$.
2. Extract and quantify high quality, intact genomic DNA from cells supplemented with or without $1\alpha,25\ (OH)_2D_3$.
3. Quantify genome-wide changes in DNA methylation by reduced representation bisulphite sequencing.
4. Establish a bioinformatics pipeline for identifying differentially methylated sites throughout the genome.
5. Identify the positions of differentially methylated sites relative to the position of gene and gene regulatory elements.

6. Identify tissue-specific biological pathways significantly affected by 1α,25$(OH)_2D_3$ induced methylation changes.

Chapter 2: Methods and Materials

This chapter will detail the experimental methods conducted to assess the impact of $1\alpha,25(OH)_2D_3$ on genome-wide methylation. The human cell lines: THP-1, HEK293 and MCF-7 were chosen as models for this study. The cell lines were cultured and treated with $1\alpha,25(OH)_2D_3$. DNA, extracted from the cell lines, was quantified and its quality assessed. The DNA was then packaged and shipped to a service provider for RRBS. The exact RRBS methods conducted by the service provider were not made available for the study, however a brief overview of the experimental processes will be described. Base calling, trimming and alignment data was received back from the service provider and was subsequently used to identify differentially methylated cytosines. Comparisons of differentially methylated cytosines across treated and untreated samples were used to assess the impact of $1\alpha,25(OH)_2D_3$ on genome-wide methylation.

2.1 Cell Culture and Treatment

THP-1 cells (passage 18, Cellonex, Johannesburg, South Africa), derived from the blood of a one-year-old boy diagnosed with acute monocytic leukemia (Tsuchiya et al., 1980).) were grown in T25 culture flasks (TPP, Trasadingen, Switzerland) in RPMI-1640 medium (Sigma-Aldrich, St. Louis, Missouri) at a concentration of 250 000 cells/ml media. This cell concentration is optimum for growth as the cells remain in close contact. To allow cells to synthesize proteins, remain free from bacterial contamination, and grow optimally, tissue culture (TC) media was supplemented with 1% (w/v) L-glutamine (Sigma-Aldrich, St. Louis, Missouri), 1% (v/v) Penicillin/streptomycin (Sigma-Aldrich, St. Louis, Missouri) and 20% foetal bovine serum (PAN-Biotech, Aidenbach, Germany). Once confluent, cells were supplemented with either 10 nM $1\alpha,25(OH)_2D_3$ (Sigma-Aldrich, St. Louis, Missouri) or vehicle control (10 nM ethanol; Sigma-Aldrich, St. Louis, Missouri) and incubated for 24 hours.

Michigan Cancer Foundation-7 (MCF-7; passage 31; a gift from Prof Rob Veale) breast cancer cells and Human embryonic kidney 293 cells (HEK293; passage 12; Fox Chase Cancer Center, Philadelphia, PA) were cultured in Dulbecco's Modified Eagle Medium (DMEM; Biowest, Riverside, Missouri) supplemented with 5% (v/v) FBS and 1% (v/v) Penicillin-Streptomycin mix. Once cells reached about 80-90% confluency they were treated with 10 nM $1,25(OH)_2D_3$ or vehicle control for 18 h. Following treatment, cells were detached by incubation in 2 ml trypsin (Sigma, St. Louis, Missouri) at 37 °C for 5 min. The cell suspension was centrifuged at 1 000 x*g* and the supernatant was discarded. Cell pellets were re-suspended in 1 ml phosphate buffered saline (PBS; 0.01 M phosphate buffer, 0.0027 M potassium chloride and 0.137 M sodium chloride, pH 7.4; Sigma, St. Louis, Missouri). The establishment and maintenance of MCF-7 and HEK293 cultures was done by fellow MSc candidates, ChaJinsi Gaza and Daniesha Govender, respectively.

All centrifugation steps were conducted at 500 x*g* for 5 minutes. All wash steps were conducted by resuspending cells in Phosphate-Buffered Saline (PBS) buffer followed by centrifugation. The osmolarity (280 – 315 mOSm/kg) and ion concentration matches that of the body and therefore prevents cell rupturing and shrivelling caused by osmosis (Martin et al., 2006). All cultures were kept at 37 °C, 5% CO_2 and 80% humidity. All cultures were washed and counted immediately after harvesting and viability assessed. Cell counting and viability assessment were done on the Countess II FL Automated Cell Counter (Invitrogen, Carlsbad, California) using a Trypan blue (Sigma, St. Louis, Missouri) assay. The Trypan blue assay is based on the principle that viable cells have intact cell membranes that exclude dyes like trypan blue, whereas non-viable cells take up the dye (Strober, 2001). Thus, the cells that took up trypan blue were considered non-viable, while viable cells remained clear. Two independent experiments of $1\alpha,25(OH)_2D_3$-treatment, DNA extraction and DNA quality control were performed on all three cell lines.

2.2 DNA Extraction

Genomic DNA (gDNA) was extracted from cells ($1x10^6$ cells) treated with or without $1\alpha,25(OH)_2D_3$ with the Blood DNA Extraction Mini Kit (Qiagen, Hilden, Germany), in triplicate, according to the manufacturer's instructions. All extraction solutions were kept at room temperature and all subsequent steps and centrifugations were performed at room temperature. The suspension cells were centrifuged for 5 minutes at 500 x*g* and then resuspended in 200 µL PBS. 20 µL Proteinase K (Qiagen, Hilden, Germany) was added to the sample. Proteinase K is a broad spectrum serine protease, stabilized by calcium ions, which digests native proteins (Barrett et al., 2012). Proteinase K digests cellular membrane proteins to facilitate cell lysis. To remove RNA contamination, 4 µL RNase A stock solution (100 mg/ml) was added to the sample. RNase A specifically cleaves single-stranded RNA at cytosine and uracil residues (Raines, 1998). 200 µL of buffer AL (Qiagen, Hilden, Germany) was added to the sample and mixed by pulse-vortexing for 15 seconds. Lysis buffers contain compounds which denature proteins and break down the cell membrane. The samples were then incubated at 56 °C for 10 minutes on a heating block. 200 µL of molecular grade ethanol (Sigma, St. Louis, Missouri) was added to the sample and mixed by pulse-vortexing for 15 seconds. Molecular grade ethanol is ultrapure and purifies polar molecules, such as nucleic acids, from non-polar cellular debris. The mixture (containing sample, proteinase K, ethanol and buffer AL) was carefully added to a spin column in a 2ml collection tube (Qiagen, Hilden, Germany) and centrifuged at 6 000 x*g* for 1 minute. Column-based nucleic acid extraction methods are more convenient in comparison to other methods and use a matrix-containing column which binds to DNA specifically, excluding all other cellular components. The column was placed in a new collection tube. 500 µL of wash buffer AW1 (Qiagen, Hilden, Germany) was added directly to the column and centrifuged at 6 000 x*g* for 1 minute. The column was placed in a new collection tube. 500µL of wash buffer AW2 (Qiagen, Hilden, Germany) was added directly to the column and centrifuged at 10 000 x*g* for 3

minutes. Washing is important to ensure that the detergents, proteases and nucleases used during cell lysis are removed. These reagents can detrimentally interfere with downstream experiments. The column was placed in a new column (Zymo Research, Irvine, California) and centrifuged again to remove residual wash buffer AW2. The supernatant and collection tube were discarded. The column was then placed in a sterile 1.5 ml microcentrifuge tube. 15 µL of elution buffer AE (Qiagen, Hilden, Germany) was added to the column and incubated at room temperature for 5 minutes. The volume of 15 µL was enough to cover the surface area of the column thus allowing the DNA to fully dissociate from the column into the elution buffer, and therefore yielding a higher concentration. The column was then centrifuged 6 000 x*g* for 1 minute. The extracted DNA was quantified by fluorescence, and its integrity assessed by electrophoresis.

2.3 DNA Quantification

Assay working solution was prepared as follows: Qubit® dsDNA HS Reagent (Invitrogen, Carlsbad, California) was diluted with Qubit® dsDNA HS Buffer (Invitrogen, Carlsbad, California) at a ratio of 1:200. The final volume in each tube was 200 µL. The standard assays were prepared by mixing 10 µL of each Qubit® standard (Invitrogen, Carlsbad, California) and 190 µL of assay working solution. The sample assays were prepared by mixing 1 µL of each DNA sample and 199 µL of working solution. All tubes were incubated for a minimum of 2 minutes prior to fluorometric quantification with a Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, California). Qubit makes use of fluorescent dyes to quantify DNA. DNA can also be quantified using spectrophotometry. The advantages of using spectrophotometry are that it does not require reagents and can also measure the purity of the DNA. However, because nucleic acids absorb light at 260 nm, spectrophotometry is unable to distinguish between amino acids, RNA, DNA or free nucleotides. Qubit fluorometry was therefore a better option for accurate quantification of DNA. Spectrophotometry

was still used to assess DNA purity and the presence of protein contamination (high 260/280 ratio).

2.4 DNA Integrity Assessment

For the initial assessment of gDNA integrity, DNA was allowed to electrophorese on a 1% (w/v) agarose gel. In agarose gel electrophoresis, size and charge are used to separate molecules. An electric field influences the migration of the molecules. DNA, RNA and all negatively charged molecules migrate towards the positive pole. Large molecules migrate slower than small molecules through the gel (Sambrook and Russell, 2001). The gel was electrophoresed in TBE buffer consisting of 0.89 M Boric acid (Merck, Modderfontein, South Africa), 0.89 M Tris (Sigma Aldrich, St Louis, Missouri) and 0.5 M Ethylenediamintetra-acetic acid (EDTA) disodium salt (Honeywell Fluka Analytical, Hampton, NH) at a pH of 8. TBE buffer protects DNA from hydrolysis during electrophoresis by maintaining a slightly alkaline pH (Stellwagen et al., 2000). EDTA chelates magnesium ions which are needed for activating hydrolases (Finnegan and Percival, 2015). 10 mg/ml Ethidium Bromide (9 µL per 150 ml gel, Sigma Aldrich, St Louis, Missouri) was added to the gel before setting to ensure visualisation of the DNA. Ethidium bromide is a UV fluorescent dye which intercalates into the DNA. After ethidium bromide has been incorporated into DNA, DNA can be visualized under a UV light. For DNA to be deemed intact, a single distinct band should be seen under UV (Mayjonade et al., 2016). A smear on the gel would represent degraded DNA. Electrophoresis was conducted at 100 V and 500 mA for 60 minutes. A minimum of 500 ng of DNA was loaded onto the gel for each sample. Gel imaging was performed using a Gel Doc™ XR system (Bio-rad, Johannesburg, South Africa) in conjunction with Quantity One® 1-D analysis software (Bio-rad, Johannesburg, South Africa).

To confirm the quality of DNA following shipment and prior to sequencing, the samples were assessed with the Fragment Analyzer and the DNF-488 High Sensitivity genomic DNA Analysis Kit (Outsourced to Diagenode, Belgium, EU).

2.5 Reduced Representation Bisulphite Sequencing

RRBS was outsourced to Diagenode (Belgium, EU). All subsequent library preparation and sequencing experiments were conducted externally according to Diagenode's established protocols. Briefly, sample library preparation for the RRBS experiment was performed using the Premium RRBS Kit (Diagenode, Seraing, Belgium). The protocol (Figure 2.1) involved these main steps: enzymatic digestion, ends preparation, adaptor ligation and size selection, sample pooling, bisulphite conversion, and PCR amplification and clean up. The samples were kept at 8 °C in between the various steps.

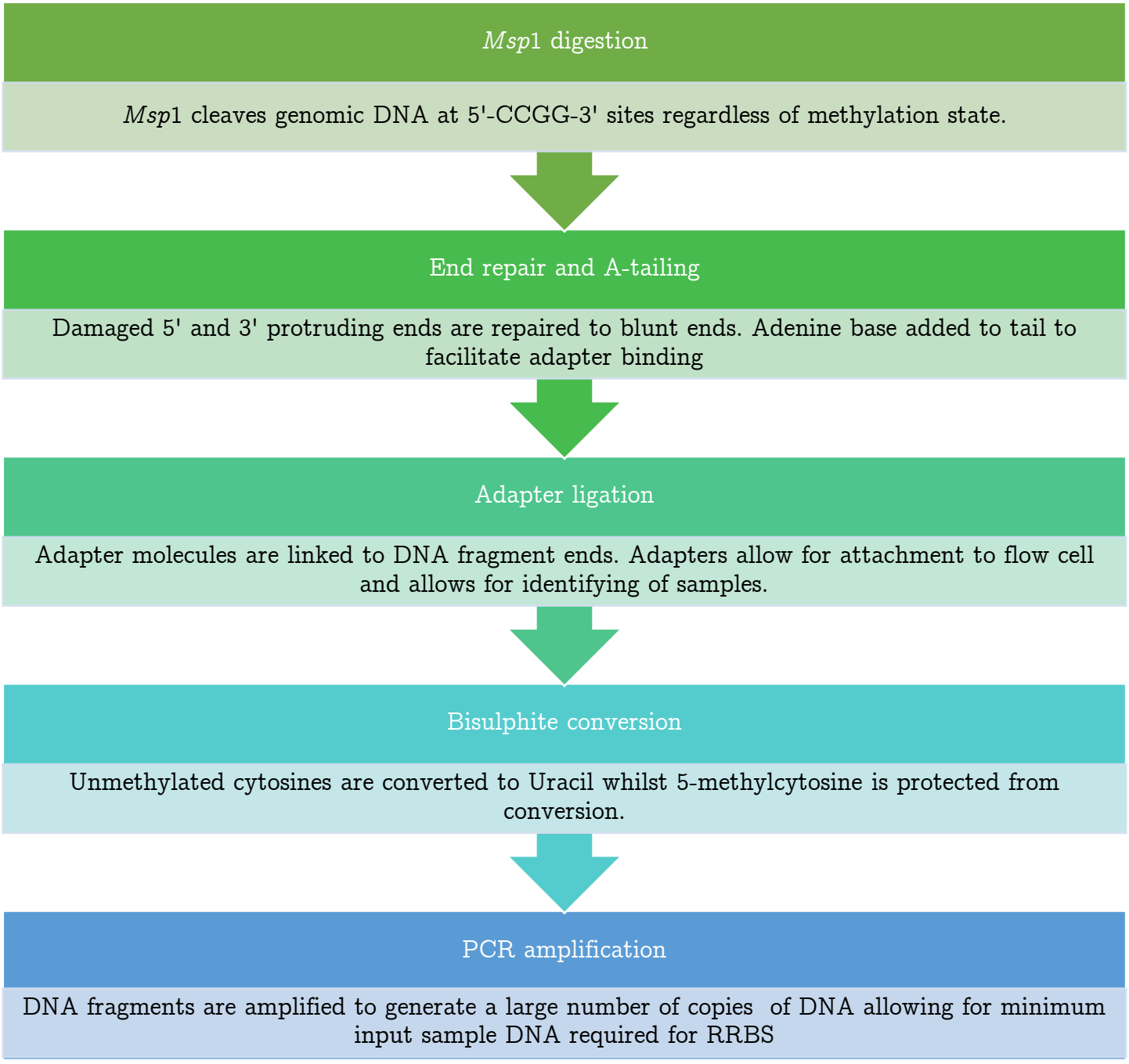


Figure 2. 1: Flow diagram showing step wise RRBS procedure.

2.5.1 Enzymatic Digestion

To ensure each fragment starts and ends with a CpG site, a digestion mixture of enzyme buffer and *Msp*I restriction enzyme was prepared according to manufacturer's instructions. *Msp*I recognizes the sequence 5'-CCGG-3' and cleaves between the cytosine dinucleotides. DNA and water were added to the digestion mixture and placed in a 96-well plate. The reaction was incubated in a thermocycler for 12 hours at 37 °C.

2.5.2 Paired-ends Preparation

To repair the damaged 5' and 3' ends as well as prepare for adapter ligation, an ends preparation mixture containing ends preparation enzyme, dNTP mix, unmethylated spike-in control and methylated spike in control was prepared. *Msp*1 cleavage produces sticky ends and therefore filling the 3' ends of fragments is required. The spike-in controls were added to monitor the efficiency of bisulphite conversion in later steps. The ends preparation mix was added to each sample. The reaction was then incubated in a thermocycler for a total of 20 minutes at 30 °C, 37 °C and 75 °C.

2.5.3 Adapter Ligation

The adapter ligation step was carried out immediately after completion of the ends preparation. Adapters were added into each well. Sequence adapters allow for the fragments to bind to adapters in the flow cell during Illumina sequencing reactions. A ligation mix consisting of ligase buffer and ligase was then prepared and added to each sample in the 96-well plate. The reaction was incubated in a thermocycler for 20 minutes at 25 °C and 10 minutes at 65 °C. For size selection, room temperature AMPure XP Beads were added to each well. A 15-minute incubation period ensured

DNA binding to the beads. DNA-bound beads formed a pellet and the supernatant containing the prior reagents was discarded. The pellet was then washed with 80% ethanol for a short period of time. This was important to ensure that the DNA remained bound to the beads. The pellet was resuspended in resuspension buffer by vortexing. The DNA was eluted from the beads and placed on a new 96-well plate.

2.5.4 Library Quantification

DNA was quantified using quantitative polymerase chain reaction (qPCR, Table 2.1). qPCR is technique which quantifies target DNA during amplification by PCR (i.e, in real time) (Logan et al., 2009). PCR products can be detected using DNA intercalating fluorescent dyes or sequence-specific DNA probes with a fluorescent reporter (Logan et al., 2009). All libraries were quantified in duplicate and the average Ct was used to calculate volumes required for pooling. A quantification mix was prepared with 2X qPCR master mix, primer mix and water. The sample libraries were diluted 1:3 with quantification mix and nuclease-free water. Dilution maximises the accuracy of a multichannel pipette. The diluted library was added to a 96-well plate along with quantification mix to a final volume of 10 µL. The samples were then pooled together to reduce handling time. During qPCR, Ct values are used to determine samples with similar concentrations. These samples are then pooled together. Pool concentrations can be used to calculate the volume of each sample DNA required. The pooled samples were bound to AMPure XP Beads and washed with 80% ethanol. The sample-bound beads formed a pellet, which was subsequently air-dryed. The pellet was resuspended in resuspension buffer and the samples were eluted from the beads.

Table 2. 1: qPCR conditions for sample library quantification.

Cycle Step	Temperature	Time	Cycles
Initial Denaturation	98°C	3 minutes	1
Denaturation	95°C	15 seconds	
Annealing	60°C	30 seconds	25
Extension	72°C	30 seconds	
Hold	4°C	∞	

2.5.5 Bisulphite Sequencing

Bisulphite treatment involves converting unmethylated cytosines into thymines through deamination of the 5' carbon (Chatterjee et al., 2012). Methylated cytosines are protected from this deamination and thus remain unchanged. In preparation for Bisulphite sequencing (BS), BS solubilization buffer and dilution buffer were added to a tube of BS conversion reagent. This was mixed at room temperature for 10 minutes through vortexing. The BS reaction buffer was then added and further mixed for 1 minute. The BS conversion reagent was added to each library pool in a 96-well plate. After brief centrifugation, the library pools were placed in a thermocycler to undergo bisulphite conversion (Table 2.5). Bisulphite-converted library pools were loaded into the BS spin columns containing BS binding buffer and centrifuged at full speed. The columns were washed with BS wash buffer and centrifugation repeated. BS desulphonation buffer was added to the column and incubated at room temperature for 20 minutes and centrifuged as before. The column was then washed twice with BS wash buffer as before. The bisulphite-converted library pools were then eluted from the column with BS elution buffer.

For enrichment PCR (Table 2.6), the amplification mix was prepared as before and added to bisulphite converted library pools. The enriched DNA was purified and cleaned using AMPure XP Beads. After a 15 minute incubation at room temperature, the DNA-bound beads were washed with 80% ethanol. The beads were left to air-dry for 10 minutes. The DNA was eluted from beads in resuspension buffer and transferred to a clean 1.5 ml tube. For quality control, the concentrations of the samples were determined using a Qubit High Sensitivity fluorescence-based assay (Life Technologies, Bengaluru, India). The BioAnalyzer profile of each library was determined on a High Sensitive Chip (Agilent, Santa Clara, Carlifonia).

Table 2. 2: Thermocycler conditions for bisulphite conversion of library pools.

Cycle	Temperature	Time	Cycles
Denaturation	95°C	1 minute	20
Conversion	60°C	10 minutes	
Hold	4°C	∞	

Table 2. 3: Enrichment PCR thermocycler conditions.

Cycle step	Temperature	Time	Cycles
Initial Denaturation	95°C	5 minutes	1
Denaturation	98°C	20 seconds	
Annealing	60°C	15 seconds	Optimal Ct value -1
Extension	72°C	45 seconds	
Final extension	72°C	7 minutes	
Hold	4°C	∞	

2.2 Bioinformatic Analysis

Following the sequencing procedure conducted by a service provider, raw sequence data quality was assessed. Phred scores describe the quality of base calling throughout the sequence procedure. Bases with a Phred score higher than the threshold are considered to be of satisfactory quality. Other quality measures assessed include: GC content, sequence length distribution, sequence duplication levels, overrepresentation and adapter content. The adapters from reads which were of satisfactory quality were trimmed using TrimGalore software (version 0.6.6). Parameters used during trimming include: maximum trimming error rate, minimum adapter overlap and minimum required sequence length for both reads before pair removed. The trimmed sequences were aligned to a reference genome to locate methylated cytosines. Bowtie 2 and Bismark (version 0.23.0) were used to extract methylation call information.

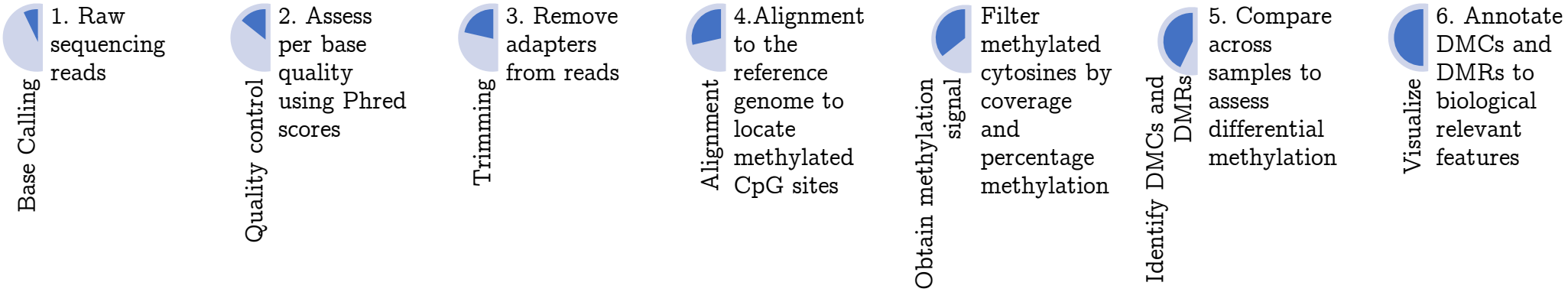


Figure 2. 2: Flowchart showing bioinformatic pipeline rationale. Steps 1-4 were done through a service provider.

Identifying differential methylation (DM) is an especially crucial step in studying epigenetic changes as quantified by sequencing methods. DM analysis involves comparing the methylation status of CpG sites in samples from two conditions of interest, in replicates. In order to identify differentially methylated cytosines (DMCs) or -regions (DMRs), one has to consider spatial correlation, sequencing depth and biological variation (Shafi *et al.*, 2018).

Spatial correlation means that CpG sites in close proximity usually have similar methylation levels. Spatial correlation reduces the risk of false positives as DMRs are not defined in isolation but by also considering the methylation levels of neighbouring regions. This allows for a more informative picture about the epigenetic differences between two samples, and simultaneously allows for the estimation of the methylation status of neighbouring CpGs. Additionally, spatial correlation allows increased accuracy even with reduced sequencing depth. Sequencing depth must be carefully considered when identifying DMCs and DMRs. Sequencing depth explains how many times, on average, each part of the genome has been sequenced. Sequencing depth determines the certainty of the methylation scores and accounts for variability introduced into the samples during sequencing. There is variation in the methylation of CpGs between individual cells and therefore there is natural variation within a sample. This biological variation can lead to false positive results when unaccounted for.

Shafi *et al.* (2018) comprehensively compared 22 methods for identifying DMRs using bisulphite sequencing data. The review focused on the primary features, advantages, and limitations of each method. Most importantly, the survey compared the core statistics which are used to determine DMRs. Using this survey as a guideline, the methods which were used in our study were MethylKit and eDMR.

2.2.1 MethylKit

MethylKit, first described in 2012, uses logistic regression for modelling methylation (Figure 2.1; Akalin *et al.*, 2012). Logistic regression predicts the value of 'y' using 'x', in a cause-and-effect manner (Dytham, 2011). MethylKit therefore models the proportion of methylated and unmethylated regions between control and treated samples (Akalin *et al.*, 2012), and transforms these proportions into logits. CpGs can only be classified as methylated and unmethylated such that the mentioned proportions will near 0 and 1. Logit transformation is powerful because it stretches out the values near 0 and 1 (Dytham, 2011) for more in-depth and accurate estimation of probabilities (p-values). DM, which is the relationship between treatment (cause) and methylation (effect), is calculated using the logit transformations. The equation for calculating methylation is:

$$\log\left({}^{P_i}/_{1-P_i}\right) = \beta_0 + \beta_1 + T_i$$

where P_i is the methylation proportion for sample i = 1 – number of biological replicates, β_1 is the log ratio between the treatment and control groups and T_i is the treatment for sample *i*. Independent tests for the methylation of each region test against the null hypothesis that $\beta_1 = 0$ (Akalin *et al.*, 2012), and if the null hypothesis is rejected, then that region is differentially methylated. Thus, all methylation ratios above 50% will have a positive value indicating a treatment-induced increase in methylation all ratios below 50% will have a negative value indicating a treatment-induced decrease in methylation, while those ratios with a logit for 50% of 0 will denote no treatment-induced change in methylation.

The logistic regression model assumes that variance in y is the same for every value of x (Dytham, 2011), and this is the case if x is obtained without any error. In methylation terms, methylKit assumes that variation in the treated sample is the same for every CpG used for comparison in the control sample, and vice versa. Ignoring

biological variation is a significant disadvantage of methylKit, as methylation in an organism differs by both cell and cell type (Zhang *et al.*, 2013).

The total sequence coverage summed up for all replicates contributes to the relevance of the sample size in each individual group (Akalin *et al.*, 2012). Coverage is thus an important factor to consider in order to improve the power of methylKit to detect DMRs (Akalin *et al.*, 2012). MethylKit corrects the calculated differential methylation p-values into q-values using a sliding linear model (SLIM) (Akalin *et al.*, 2012). The resulting q-value denotes the minimum positive false discovery rate that can occur for a p-value point (Wang *et al.*, 2011). SLIM was developed as a more robust method for estimating π_0, the only unknown and pre-estimated parameter used in the q-value formula (Wang *et al.*, 2011). A more accurate π_0 estimation increases the accuracy of the q-value calculation. SLIM correction for multiple tests is an important and advantageous feature of methylKit.

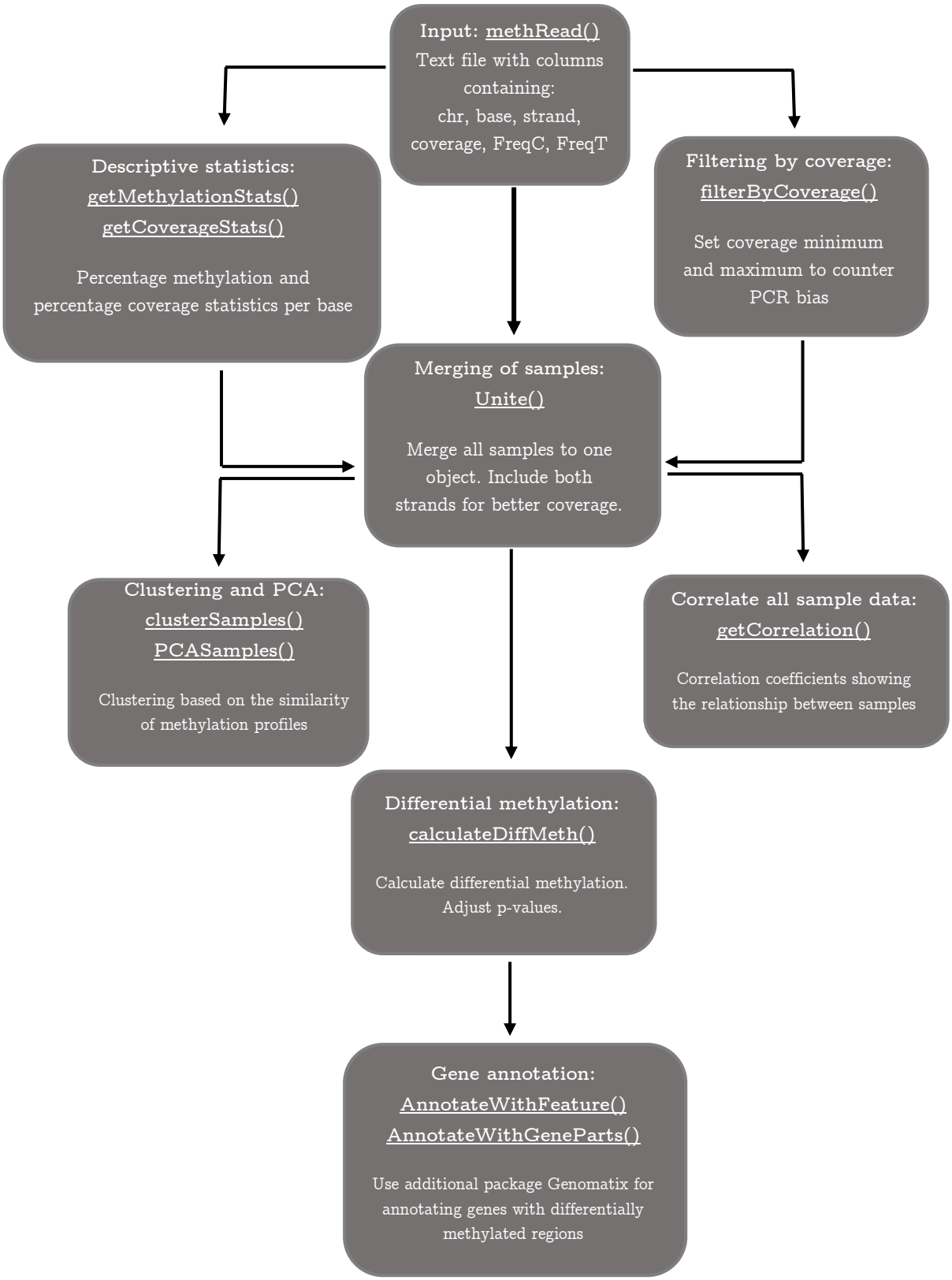


Figure 2. 3: Flowchart of methylKit features used for data analysis and how they are related to each other. The function (underlined) used for each feature are also included at each step.

2.2.2 eDMR

Empirical based differentially methylated regions (eDMR) is a method which was developed as an extension of methylKit in 2013 (Li *et al.*, 2013). The rationale of eDMR is to mathematically determine the optimum cut-off distance that separates two different methylated regions from each other (Figure 2.2; Li *et al.*, 2013). The gaps between adjacent CpGs across the genome are fitted to a bimodal normal distribution. The two modes represent regional and boundary CpGs (Li *et al.*, 2013). There is an imbalance in regional and boundary CpGs and hence the distance distribution of both populations. In order to address this, the cost of calculating a separation point (between two regions) is characterized by a weighted, combined probability function (Li *et al.*, 2013). This function evaluates and optimizes any given separation point (Li *et al.*, 2013). The initial fitting of the gaps between CpG regions to a bimodal normal distribution is heavily dependent on good sequencing coverage. In this way, eDMR is advantageous as it takes coverage into account. Once the CpG boundaries are determined, further filters are applied to each boundary. These filters include the number of CpGs and DMCs in that region, as well as the minimum mean methylation difference of that region between the treated and control samples (Li *et al.*, 2013). A limitation of eDMR is that it does not consider the biological variation in a sample and this can introduce bias in the filtering of methylated regions.

In eDMR, the methylation data is binned into segments using an adjustable parameter (for example, the number of bases). The autocorrelation, and hence refinement, of these segments is calculated and performed based on both the methylation levels and DMC p-values of that bin (Li *et al.*, 2013). The p-values of the DMCs in each region are combined using the dependency adjustment of the Stouffer-Liptak test for significance. This test is most applicable in this case, because it does not depend on the assumption that the distribution of p-values in each region is normal (which is rarely the case) (Kechris *et al.*, 2010). The combined p-values are also corrected for multiple comparisons using the false discovery rate (FDR) (Li *et al.*, 2013).



Figure 2. 4: Workflow of eDMR. Features of the package and functions (underlined) used for each feature to group CpGs into regions and determine differential methylation in those

2.2.3 Genomic Distribution Analysis

The location of differentially methylated cytosines, in the context of the genome, was analysed to provide insight into the biological relevance of 1α,25$(OH)_2D_3$-induced differential methylation. The R (version 4.0.4) package Genomation (version 1.20) was used to determine the genomic distribution of 1α,25$(OH)_2D_3$-induced differential methylation. There are many options for annotating differential methylation (Cavalcante and Sartor, 2017; Zhu et al., 2010). However, Genomation was developed along with methylKit for differential methylation annotation purposes (Akalin et al., 2012) and thus was appropriate to annotate DMCs identified by methylKit and eDMR. A file containing information on human (Hg38) protein-coding and non-protein-coding genes, taken from the NCBI RNA reference sequences (RefSeq) collection (Pruitt *et al.*, 2014), was downloaded from UCSC genome browser. The file was downloaded as a Browser Extensible Data (BED) file. The BED format is advantageous because it contains important information separated into columns. Furthermore, the information contained in files of this format is stored consistently, with *chromosome*, *start* and *end* co-ordinates always stored in the first three columns. This increases the ease of comparison with other BED files. The RefSeq file was loaded onto Genomation and the information in the file was used to obtain the locations of promoter, exonic, intronic and intergenic regions. The number of 1α,25$(OH)_2D_3$-induced DMCs/DMRs in each aforementioned genomic region was quantified and the genomic distribution of 1α,25$(OH)_2D_3$-induced differential methylation was visualized using the ggplot (version 3.2.3) tool of the tidyverse (1.3.0) R package.

2.2.4 Methylomic Distribution Analysis

The methylome describes the distribution of CpG dinucleotide-rich regions in the genome. The methylome is often correlated with genomic features (Illingworth and Bird, 2009; Saxonov et al., 2006). The methylome provides insight into possible

epigenetic regulation of genomic features. The methylome contains CGIs, CpG shores, CpG shelves and open sea CpG sites. A BED file containing information on CGIs was obtained from UCSC genome browser. Most genome databases contain CGI information however, UCSC assigns each CGI with a number. This was advantageous because subsequent CpG shores and shelves could be associated with their proximal CGI. CpG shores are regions 2 000 bp upstream and downstream of the CGI. A Bash (version 5.0.17(1)-release) script was developed which assigned the region spanning 2 000 bp before the CGI start co-ordinate as an upstream CpG shore. The script also assigned each CpG shore a number corresponding to its parent CGI. Similarly, downstream CpG shores were assigned as regions spanning 2 000 bp after the CGI end co-ordinate. CpG shelves, regions spanning 2 000 bp upstream and downstream of CpG shores, were assigned using the same process. The CpG sites which weren't assigned to either CGIs, CpG shores or CpG shelves were assigned as open Sea CpG sites. The script ensured each CpG site was assigned to only one methylomic region by prioritizing CGIs followed by CpG shores and CpG shelves respectively. However, CpG sites which were located in more than one methylomic region were also quantified for further analysis. The number of 1α,25$(OH)_2D_3$-induced DMCs/DMRs in each aforementioned methylomic region were quantified and then visualized using the ggplot (version 3.2.3) tool of the tidyverse (1.3.0) R package.

2.2.5 Enriched Differential Methylation Analysis

The functional elements enriched with differential methylation provide insight into 1α,25$(OH)_2D_3$-induced gene regulation. In order to determine these enriched regions, the human genome was first divided into regions. The human genome assembly (Hg38) was downloaded from UCSC genome browser. Ensembl and NCBI databases also have the same human genome assembly, however they use a 1-based co-ordinate system whereas UCSC uses a 0-based co-ordinate system. The human genome assembly from UCSC was therefore preferred after considering downstream analysis. The co-ordinates

of the methylated and unmethylated cytosines were obtained in 0-based format and therefore the format of UCSC co-ordinates matched those of the methylated and unmethylated cytosines. The makewindows tool from the Bedtools (version 2.30.0) collection was used to tile the genome into 1 000 bp regions. The size of 1 000 bp is the preferred tile size for differential methylation analysis (Akalin et al., 2012; Smith et al., 2020). A Bash (version 5.0.17(1)-release) script was developed to find the regions most enriched for differential methylation. In the script, Bedtools intersect tool was used to assign each 1α,25$(OH)_2D_3$-induced DMC/DMR to a genomic tile by intersecting the file containing DMCs/DMRs with the file containing genomic tiles. The intersect tool has a function which counts the number of intersections between two files. The number of DMCs/DMRs in each genomic tile could therefore be quantified. The genomic tiles were sorted by the most enriched for differential methylation (containing the most DMCs/DMRs).

Promoter regions are sites of transcriptional initiation and often targeted by epigenetic regulation. In order to define promoter regions, transcription start site (TSS) information was obtained from Ensembl database. The Ensembl database was chosen as the source of TSS information because NCBI Gene IDs (formerly Entrez IDs) can be assigned to each TSS. Each enriched promoter region could therefore be associated with a gene. Promoter regions were defined as the region 500 bp upstream and downstream of the TSS. Promoter-proximal elements can be found up to 200 bp upstream of the transcription initiation site (Griffiths et al., 2000). Additionally, the transcription initiation complex is large and is comprised of up to 70 polynucletiodes (Lodish et al., 2000) thus binding can extend downstream of the TSS. The marker of 500 bp upstream and downstream of the TSS was therefore chosen generously to account for all possible promoter elements as well as transcription initiation complex binding after the TSS. A Bash (version 5.0.17(1)-release) script was therefore developed to identify promoter regions enriched for 1α,25$(OH)_2D_3$-induced differential methylation. The script used Bedtools intersect tool to intersect the promoter region file with the file containing differential methylation enriched genomic tiles. Promoter regions most enriched for 1α,25$(OH)_2D_3$-induced differential methylation were thus

identified. The NCBI Gene ID of each enriched promoter region and its associated number of DMCs/DMRs were stored in a separate file and used as input for further analysis.

2.2.6 Pathway Analysis

Identifying the pathways associated with enriched genes can provide insight into the phenotypic effect of 1α,25$(OH)_2D_3$-induced differential methylation. clusterProfiler (version 3.16.1) was used to perform over-representation analysis (ORA) on methylation-enriched genes. ORA (Boyle et al., 2004) was used to identify biological processes and pathways which are enriched in a set of differentially methylated genes. The p-value was calculated using a hypergeometric test and corrected for multiple testing. The p-value was calculated by the equation (Boyle et al., 2004):

$$p = 1 - \sum_{i=0}^{k-1} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}$$

The total number of genes is N, and the number of genes within N that are annotated to the genes of interest is M. The size of the gene list of interest is n, and k is the number of genes within that list that are annotated to the pathway-associated gene set. Gene Set Enrichment Analysis (GSEA) can also be used to find enriched pathways. GSEA is advantageous because it can detect small differences in related genes (Subramanian et al., 2005). ORA was preferred to GSEA because GSEA is more appropriate for enrichment testing of particular pathways and not for whole genome analysis. A file containing a gene list of 1α,25$(OH)_2D_3$-induced differentially methylation genes was used as clusterProfiler input. The gene list consisted of NCBI Gene IDs and the number of DMCs/DMRs associated with each gene. The gene list was ranked by most enriched genes (genes associated with the highest number of DMCs/DMRs). The gene ontology (GO) terms associated with enriched genes were identified and visualized with ClusterProfiler (version 3.16.1).

Chapter 3: Results

The results depicted in section 3.1.1 below were obtained after DNA extraction, quantification and quality assessment by the candidate. 10 wells were made available by the service provider for 10 samples to undergo RRBS. Untreated and 1α,25(OH)$_2$D$_3$-treated THP-1 and HEK293 samples were sent in duplicate. The final 2 wells were used to analyse 1α,25(OH)$_2$D$_3$-treated MCF-7 samples. Methylation call data compiled from raw sequence data aligned to the human genome, was used to perform bioinformatic analysis. Briefly, methylKit and eDMR were used to identify differential methylation between treated and untreated samples from both THP-1 and MCF-7 cell lines. The differentially methylated sites and regions were annotated to genomic features. The biological pathways in which differential methylation was enriched, were identified using clusterProfiler.

3.1 Quality control

3.1.1 High-molecular weight gDNA was extracted

To assess the quality and integrity of the extracted gDNA, gDNA from each sample was electrophoresed on an agarose gel. Extracted DNA was intact as illustrated by the single high molecular weight band on the gel (Figure 3.1). Extracted DNA was relatively pure and free from RNA and protein contamination with 260/280 spectrophotometry ratios averaging between 1.9 and 1.95. To confirm quality after shipment, DNA integrity for each sample was further analysed using a Fragment Analyzer™ prior to RRBS library preparation. For all samples, single peaks at approximately 22 000 bp were present (Figure 3.2.)

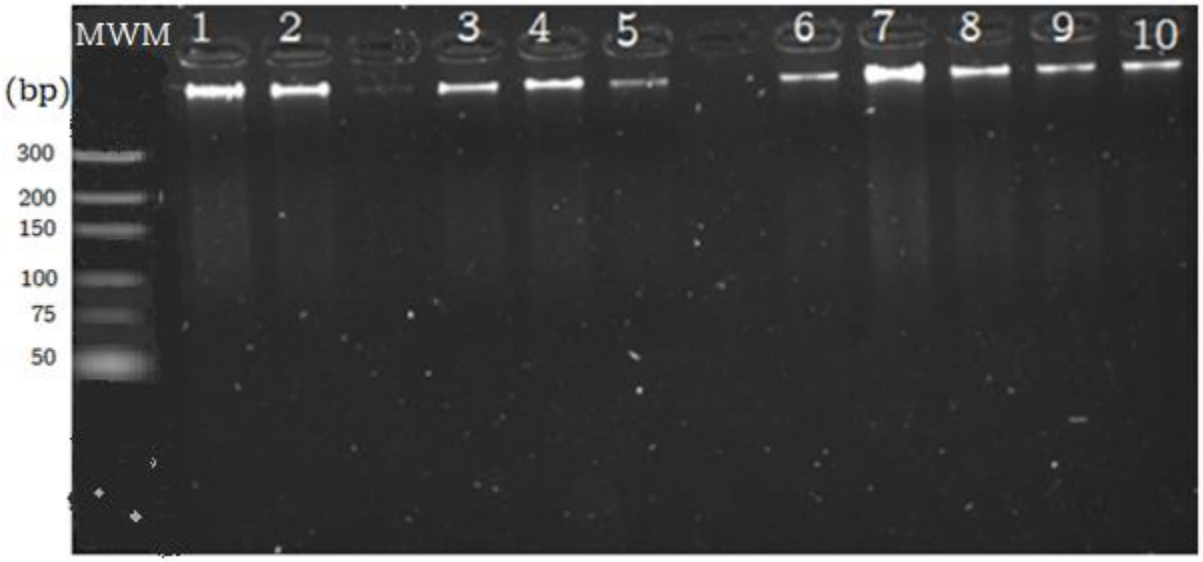


Figure 3. 1: Agarose gel showing intact gDNA. This representative 1% Agarose gel shows distinct gDNA bands extracted from untreated THP-1 (lane 1-2), 1α,25(OH)$_2$D$_3$-treated THP-1 (lane 3-4), untreated HEK293 (lane 5-6), 1α,25(OH)$_2$D$_3$-treated HEK293 (lane 7-8and 1α,25(OH)$_2$D$_3$-treated MCF-7 (lane 9-10) (MWM: molecular weight marker).

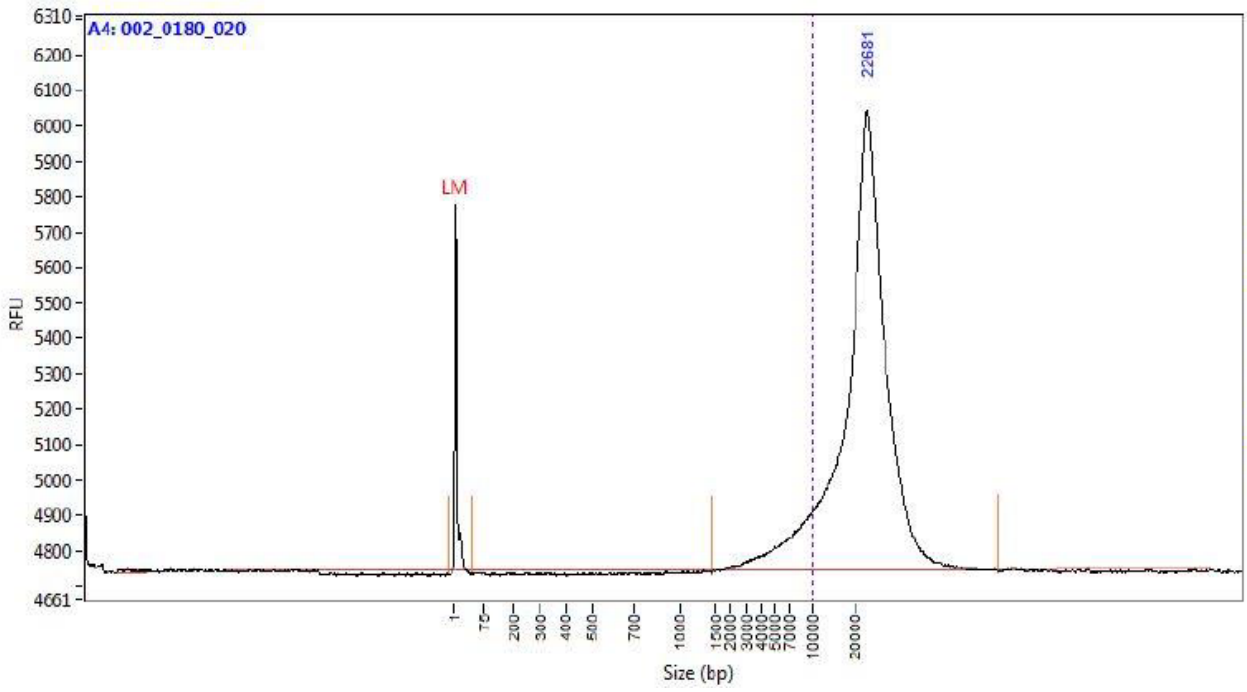


Figure 3. 2: Fragment analysis confirmed gDNA integrity. The representative electropherogram generated by means of a Fragment Analyzer™ after shipment of samples and prior to sequencing shows a single peak indicating intact gDNA with a molecular size of 22 681 bp.

3.1.2 The pooled RRBS library fragment profile closely matched that of the standard human genome

The gDNA sample libraries were pooled, prior to sequencing, to reduce handling time and reagent consumption. To assess the integrity of the pooled libraries following *Msp*1 digestion, the sample was subjected to fragment analysis (Figure 3.3). The fragment profile obtained mostly matched what was expected for *Msp*I digestion of a human genome, however, there were two small and unexpected peaks (red) at around 224 and 254 bp showing additional cleavage by *Msp*1.

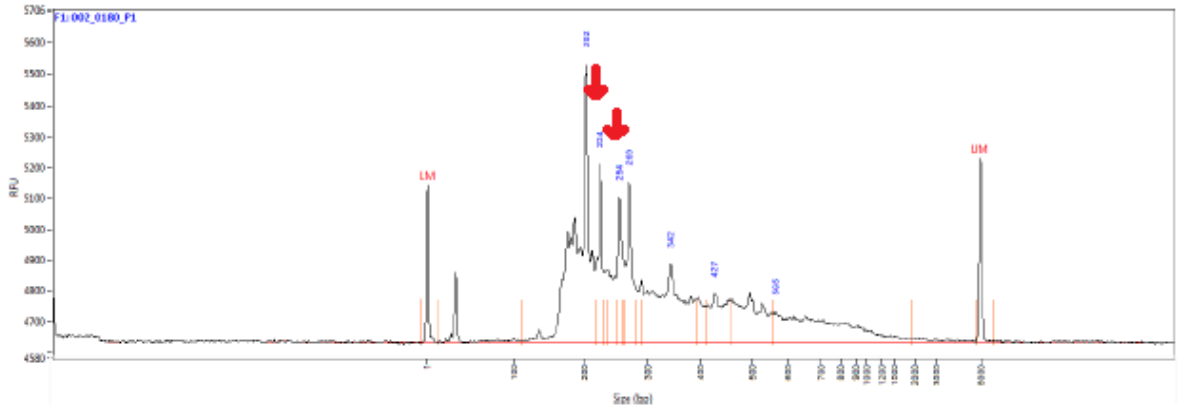


Figure 3. 3: Apart from two small fragments at 224 and 254 bp, the fragments generated after *Msp*1 cleavage of gDNA matched that of the human genome. The electropherogram of RRBS library shows the size and number of fragments after *Msp*1 cleavage. The different sized peaks correspond to the *Msp*1 restriction sites and are species-specific. The peaks visible at 202, 269, 342, 427 and 565 bp correspond to the standard human genome, whereas the peaks at 224 and 254 bp (red) do not.

3.1.3 The bisulphite conversion of the sample DNA was efficient and sequenced reads mapped successfully to the reference genome

Methylated and unmethylated spike-in control DNA was included in the bisulfite sequencing experiment to assess bisulfite conversion efficiency (Table 3.1). The conversion rate of methylated spike-in control DNA was 1.5-2% and that of unmethylated spike-in control DNA was 99.5-99.8%. To perform genome-wide analysis of vitamin-D induced differential methylation, RRBS was performed on THP-1, HEK293 and MCF-7 cell lines (n = 2), treated with and without $1\alpha,25(OH)_2D_3$. Depending on sample, between 27-49 million NovaSeq (maximum length of 2 x 150 bp) reads were generated for each sample. In the THP-1 and HEK293 samples, 63-67% of these reads mapped uniquely to the reference genome only once, whereas for MCF-7 samples, less than 1% of the reads did the same. For the THP-1 and HEK293 samples, the number of CpG sites detected were in the range of 6-7.2 million and of those, 2.2-4.4 million sites were covered by 10 or more reads. The number of CpG sites detected in MCF-7 samples was considerably lower at 483 048-505 447. Furthermore, only 43 158-47 752 of these CpG sites had a coverage of more than 10 reads. The coverage ranged between 10-21 reads. To assess the quality of methylation detected, the distributions of the methylation percentage and CpG site coverage for each sample were plotted (Figure 3.4). Over 8 x 10^5 reads were either hypomethylated (0%) or hypermethylated (100%), with less than 1 x 10^5 reads showing an intermediate methylated state. The average coverage for the THP-1 (13.30) and HEK293 (17.38) samples was over the 10 read mark. Given the poor sequencing and alignment quality obtained for the MCF-7 samples, these samples were excluded from further analysis.

Table 3. 1: Summary of sample data following bisulfite conversion, sequencing, and alignment.

Sample		Total read pairs	Uniquely aligned	Mapping efficiency	CpGs detected	CpGs coverage > 10	Average coverage	Conversion rate of methylated spike-in control (%)	Conversion rate of unmethylated spike-in control (%)
THP-1	Control R1[a]	33 040 038	21 305 457	64.48	6 636 404	2 744 219	12.03	1.71	99.77
	Control R2[b]	50 167 857	32 136 583	64.06	7 091 716	3 775 275	17.03	1.71	99.64
	Treated R1	27 671 160	17 522 205	63.32	6 173 512	2 263 395	10.74	1.58	99.74
	Treated R2	36 945 551	23 562 694	63.78	6 669 087	3 028 836	13.38	1.76	99.68
HEK	Control R1	61 475 746	41 383 667	67.32	7 217 181	4 359 074	21.60	1.70	99.68
	Control R2	43 852 197	29 703 176	67.73	6 809 601	3 642 591	16.38	1.62	99.69
	Treated R1	33 773 591	22 867 401	67.71	6 600 088	3 037 777	13.18	1.57	99.72
	Treated R2	47 042 039	31 845 764	67.70	6 769 943	3 694 402	18.37	1.72	99.76
MCF-7[c]	Treated R1	45 271 734	403 356	0.89	483 048	43 158	3.30	1.60	99.59
	Treated R2	52 543 357	419 239	0.80	504 447	47 752	3.50	2.00	99.67

[a] Replicate 1

[b] Replicate 2

[c] Control replicates excluded as only 10 sequencing slots were available. Control MCF-7bisulfite sequencing data is publicly available.

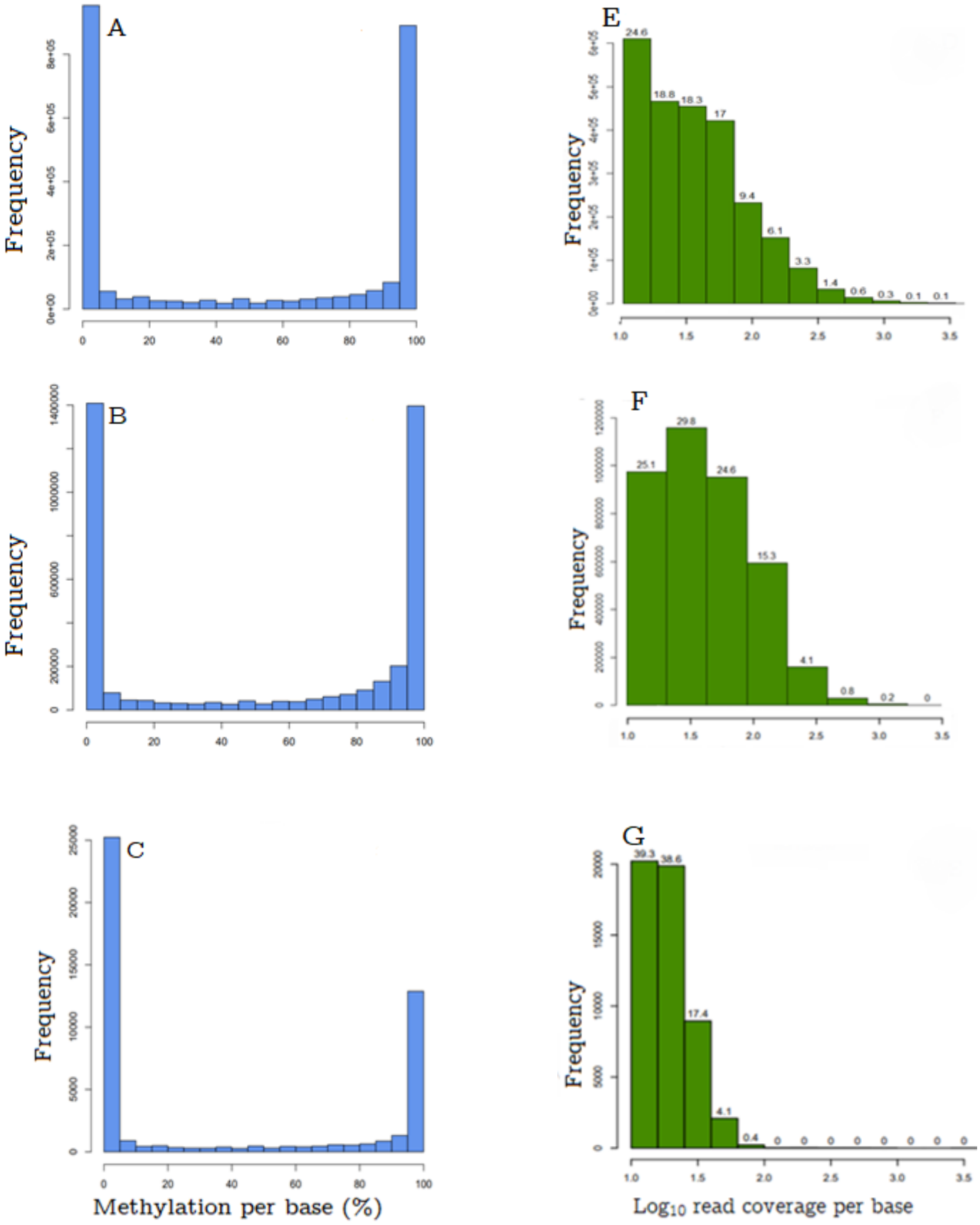

A
Frequency
0e+00
2e+05
4e+05
6e+05
8e+05
0
20
40
60
80
100
B
Frequency
0
200000
600000
1000000
1400000
0
20
40
60
80
100
C
Frequency
0
5000
10000
15000
20000
25000
0
20
40
60
80
100
Methylation per base (%)
E
Frequency
0e+00
1e+05
2e+05
3e+05
4e+05
5e+05
6e+05
24.6
18.8
18.3
17
9.4
6.1
3.3
1.4
0.6
0.3
0.1
0.1
1.0
1.5
2.0
2.5
3.0
3.5
F
Frequency
0
200000
400000
600000
800000
1000000
1200000
25.1
29.8
24.6
15.3
4.1
0.8
0.2
0
1.0
1.5
2.0
2.5
3.0
3.5
G
Frequency
0
5000
10000
15000
20000
39.3
38.6
17.4
4.1
0.4
0
0
0
0
0
0
0
1.0
1.5
2.0
2.5
3.0
3.5
Log10 read coverage per base

Figure 3. 4: The majority of the CpG sites in all samples showed high (>95%) and low (<10%) levels of methylation with few sites showing intermediate levels of methylation, with no observed PCR bias. Percentage distribution showing a frequency of reads with different levels of methylation for THP-1(A), HEK293(B) and MCF-7(C) cells. The frequency of CpGs with a $\log_{10}$(coverage)>2 was low indicating no PCR bias for THP-1(D), HEK293(E) or MCF-7(F) cells.

3.2. Differential Methylation Analysis

3.2.1 1α,25$(OH)_2D_3$ induce differential methylation in regions of the genome associated with gene expression control

To determine whether methylKit and eDMR were in agreement on where 1α,25$(OH)_2D_3$ treatment induced the largest change in methylation, the top sites (methylKit) and regions (eDMR) were identified on the UCSC genome browser. This provided insight into whether the location of differential methylation was influenced by the method of detection and increased confidence in the differential methylation detected by both methods. The UCSC genome browser has tracks which allow the user to identify exactly what element is located at a particular site or in a particular region (Table 3.3 and 3.4). methylKit-identified differential methylation did not match eDMR-identified differential methylation. Half of the most differentially methylated sites identified by methylKit were in genomic sites with unknown functions. eDMR identified many epigenetic-linked DNA regions as most differentially methylated.

Table 3. 2: Genomic location of $1\alpha,25(OH)_2D_3$-induced differentially methylated sites in THP-1 cells as identified by methylKit.

Co-ordinate[d]	Annotated Feature	Nearby gene[e]	Function of gene	Methylation (%)	q-value
Chr4:6,008,437	Protein coding open reading frame	*JAKMIP1*	Janus kinase and microtubule interacting protein links cytokine signalling from membrane receptors to signal transducers and STAT TFs.	82.60	3.96×10^{-6}
Chr8:48,910,751	Unknown feature	*SNAI2*	Snail family transcriptional repressor is involved in epithelial-mesenchymal transitions and has antiapoptotic activity.	82.46	3.26×10^{-9}
Chr19:33,663,362	Unknown feature	*CHST8*	Carbohydrate sulfotransferase 8 catalyses the transfer of sulphate to position 4 of non-reducing N-acetylgalactosamine (GalNAc) residues in both N-glycans and O-glycans.	80.76	3.14×10^{-7}
Chr4:6,008,489	Protein coding open reading frame	*JAKMIP1*	Janus kinase and microtubule interacting protein links cytokine signalling from membrane receptors to signal transducers and STAT TFs.	78.26	1.56×10^{5}
Chr8:72,729,611	Protein coding	*KCNB2*	Voltage-gated potassium ion channel regulating neurotransmitter release	77.71	7.67×10^{-7}

[d] Hypomethylated CpG sites in blue, hypermethylated CpG sites in red.
[e] Upstream or downstream gene in closest proximity to CpG site.

Chr17:13,044,059	Unknown feature	*ELAC2*	Catalyses the removal of the 3' trailer from precursor transfer RNAs	-90.32	8.10 x10^{-9}
Chr10:99,576,150	Unknown feature	*SLC25A28*	Mitochondrial iron transporter that mediates iron uptake	-88.89	2.57 x10^{-8}
Chr12:111,577,455	Protein coding	*Ataxin*	Involved in microsatellite-expansion disease	-81.48	8.80 x10^{-8}
Chr4:8,646,856	Unknown feature	CPZ	Enzyme displays carboxypeptidase activity towards substrates with basic C-terminal residues	-77.78	4.79 x10^{-7}
Chr16:13,299,570	Repeat	*U91319.1*	Long non-coding RNA	-75	8.32 x10^{-8}

Table 3. 3: Genomic location of 1α,25$(OH)_2D_3$-induced differentially methylated regions in THP-1 cells as identified by eDMR.

Co-ordinate	Annotated Feature	Nearby gene	Description	Methylation (%)	q-value
Chr6:128,882,979 - 128,883,104	CTCF-bound, Promoter-like signature	*Laminin*	Extracellular protein which mediates attachment, migration and tissue organization.	57.57	1.74 x10^{-7}
Chr20:59,571,930 - 59,571,955	Histone 3 lysine 27 acetylation	*PHACTR3*	Phosphatase and actin regulator.	50.48	4.23 x10^{-6}
Chr7:79,454,490 - 79,454,504	CTCF-bound, Proximal enhancer-like signature	*MAGI2* (antisense RNA)	Interacts with protein responsible for entatorubral and pallidoluysian atrophy.	48.29	7.83 x10^{-4}
Chr10:101,121,457 - 101,121,480	Non-coding RNA (TLX1 neighbour)	*TLX1*	NK-Like nuclear transcription factor.	46.07	2.40 x10^{-4}
Chr14:59,630,528 - 59,630,566	CTCF-bound, proximal enhancer-like signature	*Reticulon 1*	Involved in neuroendocrine secretion or in membrane trafficking in neuroendocrine cells.	43.65	7.83 x10^{-4}
Chr18:6,485,645 - 6,485,688	Repeat	*AP005202.2*	Short Interspersed Nuclear Element.	-53.07	5.37 x10^{-5}
Chr4:8,646,770 - 8,646,856	Histone 3 lysine 27 acetylation	*Carboxypeptidase Z*	Enzyme has carboxypeptidase activity towards substrates with basic C-terminal residues.	-51.63	5.96 x10^{-4}
Chr9:29,371,441 - 29,371,586	Repeat		Short Interspersed Nuclear Element.	-47.97	2.43 x10^{-6}
Chr16:73,377,892 - 73,377,905	Transcription Factor coding	*ZFHX3*	Zinc finger homeobox 3	-47.82	1.04 x10^{-4}

Chr2:131,646,906 - 131,647,027	Histone 3 lysine 4 trimethylation	*LINC01087*	Long intergenic non-protein coding RNA 1087	-46	5.54×10^{-5}

3.2.2 Minimal technical variability in the methylomes of treated and untreated replicates in both THP-1 and HEK293 cells, respectively

To assess the biological variability between the samples, Pearson correlation coefficients of the differentially methylated cytosines were calculated and used to construct scatter plots of DNA methylation between replicates in THP-1 and HEK293 samples (Figure 3.5). There was a strong positive correlation between control replicate 1 and 2, in the THP-1 and HEK293 cells (Pearson correlation coefficients of 0.97). Similarly, there was a strong positive correlation between treated replicate 1 and 2 in the THP-1 and HEK293 samples (Pearson correlation coefficients of 0.96 and 0.94 respectively). In both the THP-1 and HEK293 samples, there was a strong positive correlation between the control and treated replicates 1 and 2 (Pearson correlation coefficients of 0.94-0.97, respectively).

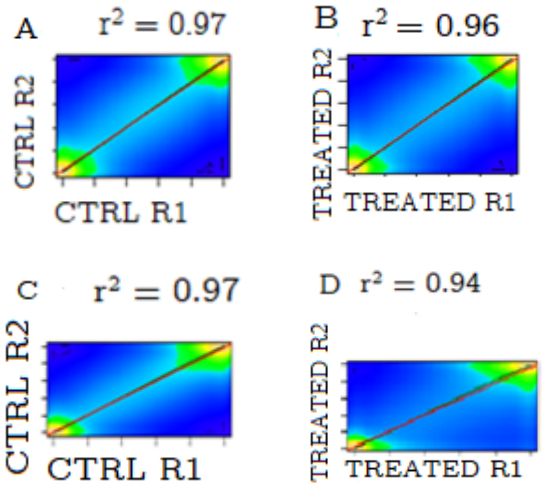


Figure 3. 5: There was a strong positive correlation between the methylomes of control and treated replicates in THP-1 (A) and HEK293 (B) samples. Scatter plots of percentage methylation values for each replicate in the sample. Numbers on the upper right

corner represent pairwise Pearson's correlation scores. The histograms represent the methylation distribution of CpG sites for each replicate.

3.2.3 1α,25(OH)$_2$D$_3$ induced differential methylation in both THP-1 and HEK293 cells

Each CpG site was assigned a methylation percentage following successful alignment with Bismark aligner. Methods from methylKit, a publicly available tool, were used to determine whether differentially methylated cytosines were present between 1α,25(OH)$_2$D$_3$-treated and untreated THP-1 and HEK293 cells. Data from both samples was first filtered to remove sites with coverage lower than 10X and coverage in the 99.9th percentile. This improved the quality of the differential methylation results; low coverage reduces the power of differential methylation scores and extremely high coverage indicates PCR bias (Aird *et al.*, 2011). The normalization of coverage was crucial because significance scores (P-values) calculated from methylKit rely on coverage (Akalin *et al.*, 2012). methylKit uses logistic regression to determine statistically significant methylation differences between treated and untreated samples. P-values, which represent the number of false positives in an experiment, were corrected to q-values for multiple testing. The most appropriate cut-off values for this, an exploratory study investigating the impact of 1α,25(OH)$_2$D$_3$ on genome-wide methylation, were chosen by looking at the total number of differentially methylated cytosines (DMCs) detectable with different q-value and methylation difference combinations (Table 3.2). DMCs of interest were chosen by applying cut-offs of a minimum methylation difference of 15% and a maximum q-value of 0.010. The same cut offs were applied in the determination of differentially methylated regions with eDMR.

1α,25(OH)$_2$D$_3$ induced four times more DMCs in HEK293 cells than in THP-1 cells, when the cut-offs were applied. In the THP-1 cells, 48% of the detected DMCs were

hypomethylated and 52% were hypermethylated. Similarly, in HEK293 cells, 47% of the detected DMCs were hypomethylated and 53% were hypermethylated. This distribution of hypo- and hypermethylated DMCs was similar at chromosome level for THP-1 cells (Figure 3.7). However, in HEK293 cells chromosome 15 and 21 have more hypermethylated DMCs. The majority of the CpG sites in both THP-1 (29 633) and HEK293 (41 298) samples were excluded by the q-value cut-off (Figure 3.6). In both THP-1 and HEK293 cells, there were both hypomethylated and hypermethylated CpG sites with a high methylation percentage. In the THP-1 cells, the hypermethylated CpG site with the highest q-value had a methylation difference percentage of 47% and the hypomethylated CpG site with the highest q-value had a methylation difference percentage of 56%. HEK293 cells had DMCs with very high q-values.

$1\alpha,25(OH)_2D_3$ induced seven times more DMRs in HEK293 cells than in THP-1 cells, when the cut-offs were applied. In the both cell lines, there were very similar numbers of hypomethylated (THP-1; 50.1%, HEK293; 49.8%) and hypermethylated (THP-1; 49.9%, HEK293; 50.2%) DMRs detected. At the chromosome level, the distribution of hypo- and hypermethylated DMRs identified with eDMR was the same as the distribution of DMCs identified by methylKit. DMRs identified in HEK293 cells, had higher methylation differences and higher q-values.

Table 3. 4: The number of DMCs detected with different combinations of q-value and methylation difference cut-off.

	THP-1			HEK293		
Methylation Difference	Q-value < 0.001	Q-value < 0.010	Q-value < 0.050	Q-value < 0.001	Q-value < 0.010	q-value < 0.050
5	5499	26189	91449	49114	117699	246082
10	5384	23821	69211	46493	99785	173778
15	5010	19619	48252	41015	78476	119774
20	4464	15210	32342	34319	58488	78690
25	3794	11337	30627	27019	40901	49573
30	3113	7977	12369	19625	26896	30041

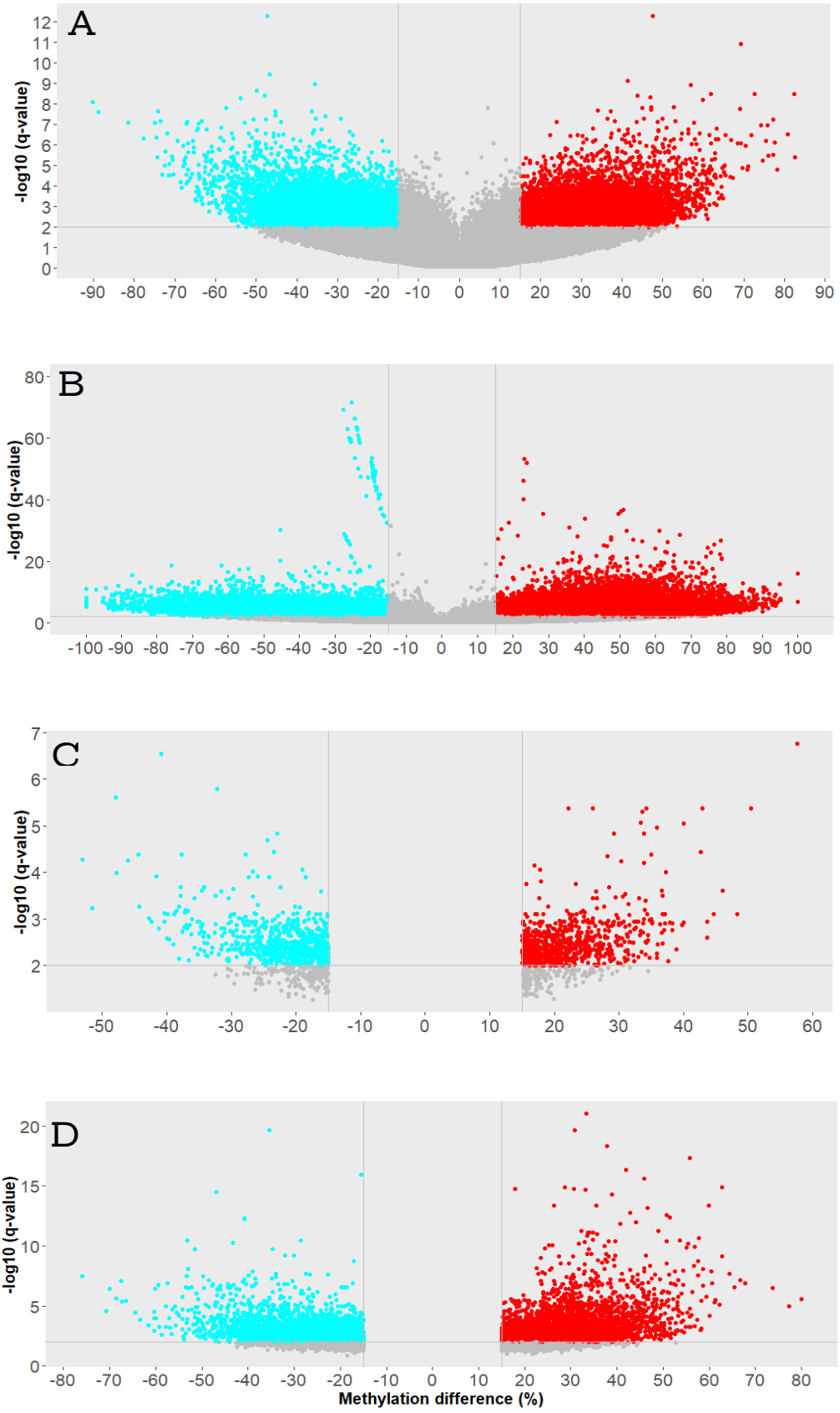

A
-log10 (q-value)
B
-log10 (q-value)
C
-log10 (q-value)
D
-log10 (q-value)
Methylation difference (%)

Figure 3. 6: Methylation and demethylation events occurred in THP-1 and HEK293 samples following treatment with $1\alpha,25(OH)_2D_3$. Volcano plots show differentially methylated cytosines in THP-1 (A) and HEK293 (B) cells and differentially methylated regions in THP-1 (C) and HEK293 (D) cells. Blue dots indicate hypomethylated cytosines/regions and red dots indicate hypermethylated cytosines/regions. Grey dots indicate cytosines/regions excluded by the q-value (0.010) and methylation difference (15%) cut-offs.

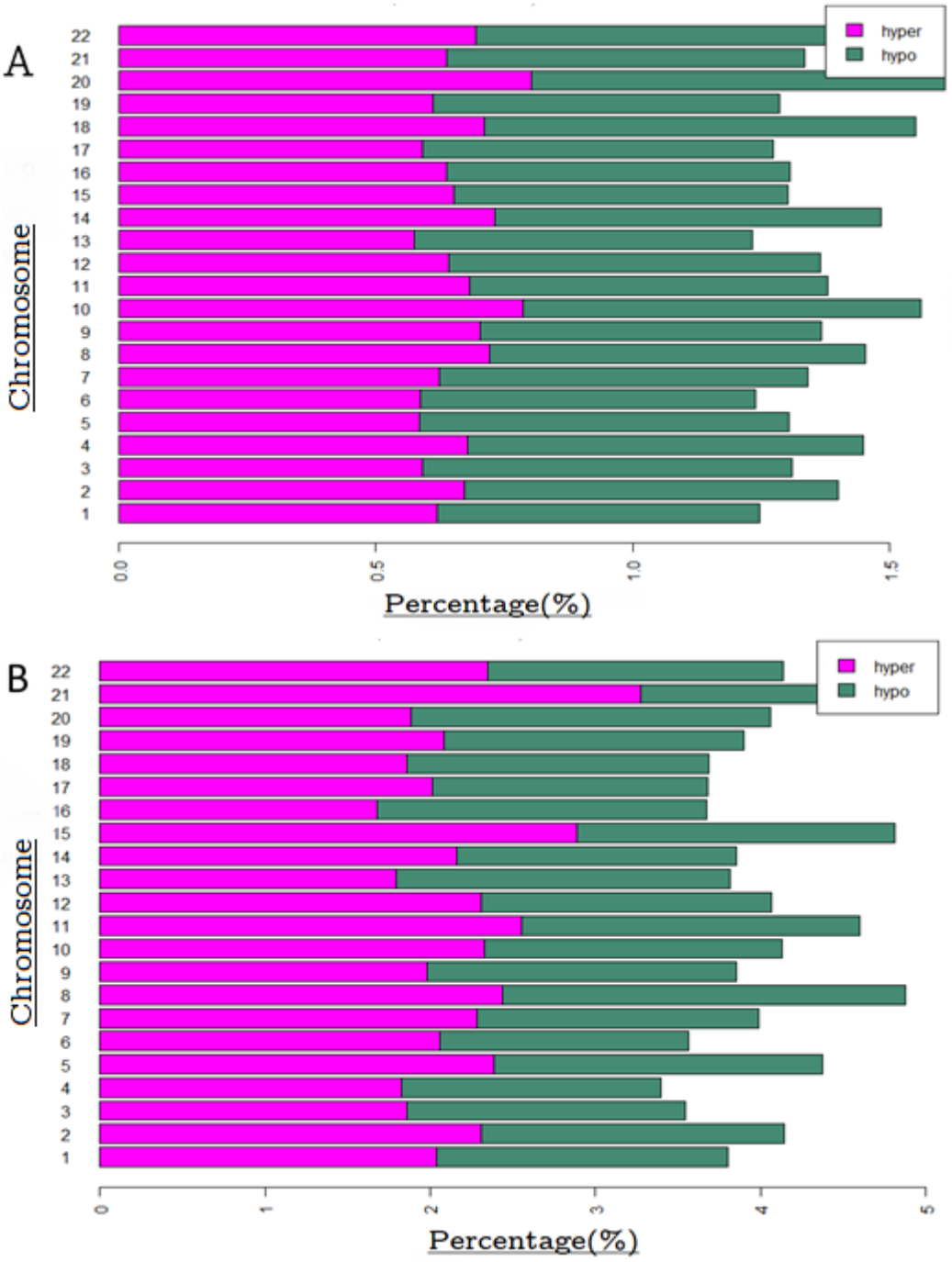


Figure 3. 7: Total differential methylation percentage per chromosome was higher in HEK293 than in THP-1 cells following $1\alpha,25(OH)_2D_3$ supplementation. The bar graphs show the percentage of total hypermethylation (purple) and hypomethylation (green) identified by both methylKit and eDMR in THP-1 (A) and HEK293 (B) cells.

3.2.4 Differential methylation occurred most in open sea CpG sites and intergenic regions.

The distribution of differential methylation across the methylome as well as biologically relevant genomic features was assessed using annotation data for the human genome version GRCh38 sourced from UCSC genome browser. A file containing the locations of experimentally verified CGIs was obtained and used to determine the locations of CpG shelves and CpG shores. The location where differential methylation occurred was compared to the location of methylome features and, following $1\alpha,25(OH)_2D_3$ treatment, the most differential methylation occurred in open sea CpG sites (54% for THP-1 cells, 48% for HEK293 cells; Figure 3.8). $1\alpha,25(OH)_2D_3$ induced substantial differential methylation in CGIs (26% in THP-1 cells, 35% in HEK293 cells), with the lowest level of differential methylation occurring in CpG shelves. Interestingly, in both THP-1 and HEK293 cells, some CpG shelf and CpG shore differential methylation occurred in methylome regions which housed other CGIs (Figure 3.9). DMRs, identified by eDMR, were also annotated to the methylome. In THP-1 cells, DMRs were most located in CGIs (34%) and least in CpG shelves (15%). In HEK293 cells, DMRs were most located in open seas (34%) and least in CpG shelves (25%).

To assess the distribution of differential methylation within genomic features, a file containing the genomic locations of promoters, introns and exons was obtained from UCSC genome browser and compared to the locations of differentially methylated CpG sites and regions (Figure 3.10). Following $1\alpha,25(OH)_2D_3$ treatment, the most differential methylation occurred in intergenic regions: THP-1 (35%) and HEK293 (32%). The least $1\alpha,25(OH)_2D_3$ induced differential methylation occurred in exons: THP-1 (9%) and HEK293 (8%). $1\alpha,25(OH)_2D_3$-induced differential methylation in promoter and intronic regions was intermediate: THP-1 (25% and 31% respectively) and HEK293 (30% respectively). The most exonic differential methylation, induced by

1α,25$(OH)_2D_3$, was at chromosome 19 (16% and 15%) in THP-1 and HEK293 cells, respectively.

DMRs, identified by eDMR, were also annotated to the genomic features. In both THP-1 and HEK293 cells, the most 1α,25$(OH)_2D_3$-induced differential methylation occurred equally in both intergenic (30.91 % THP-1, 32.39 HEK293) and intronic (30.16% THP-1, 32.69 HEK293) regions. The least 1α,25$(OH)_2D_3$-induced differential methylation occurred in exonic regions (9.69% in THP-1, 8.74% in HEK293). Interestingly, the distribution of DMRs identified by eDMR in genomic regions, varied more across chromosomes compared to the distribution of DMCs identified by eDMR.

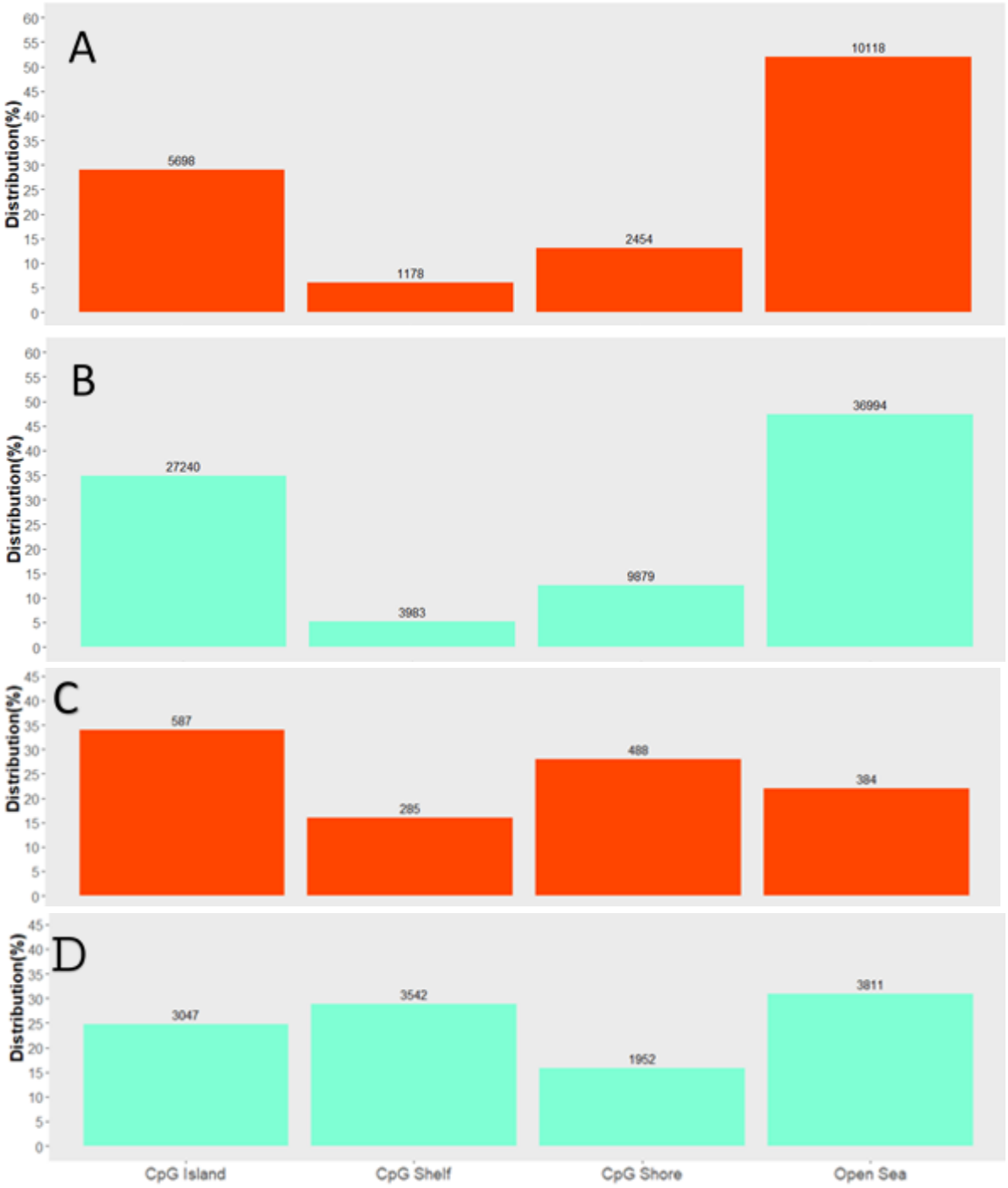


Figure 3. 8: $1\alpha,25(OH)_2D_3$ treatment primarily induced differential methylation in open sea CpG sites and CGIs. The bar graphs show the distribution of differentially methylated CpG sites in THP-1 (A) and HEK293 (B) cells and differentially methylated regions in THP-1 (C) and HEK293 (D) cells, across the methylome.

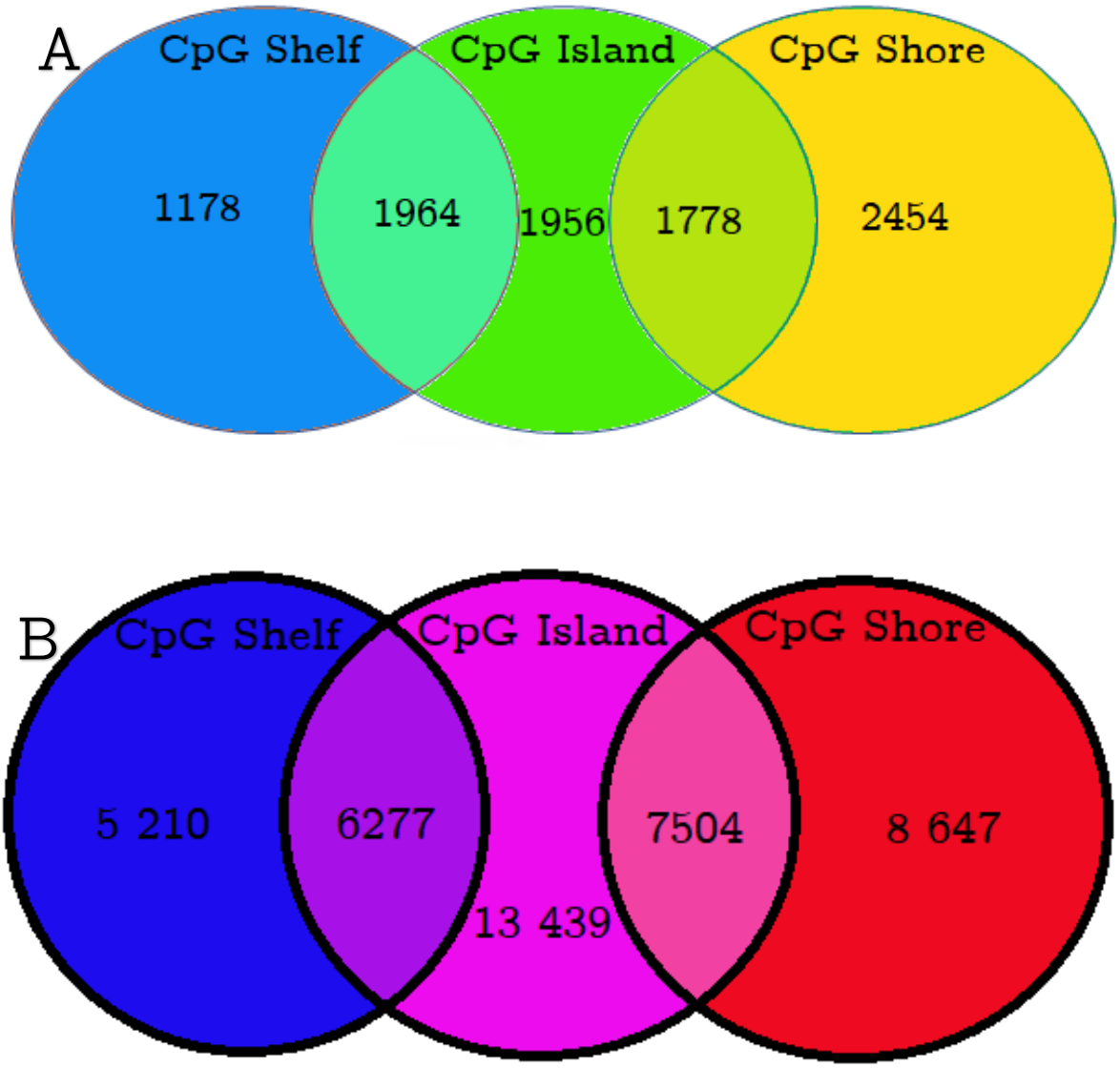


Figure 3. 9: There was a co-occurrence of differential methylation in the methylome. Differentially methylated CpG sites, identified in THP-1 (A) and HEK293 (B) cells, located in CpG shores and shelves occurred in regions which housed other CGIs.

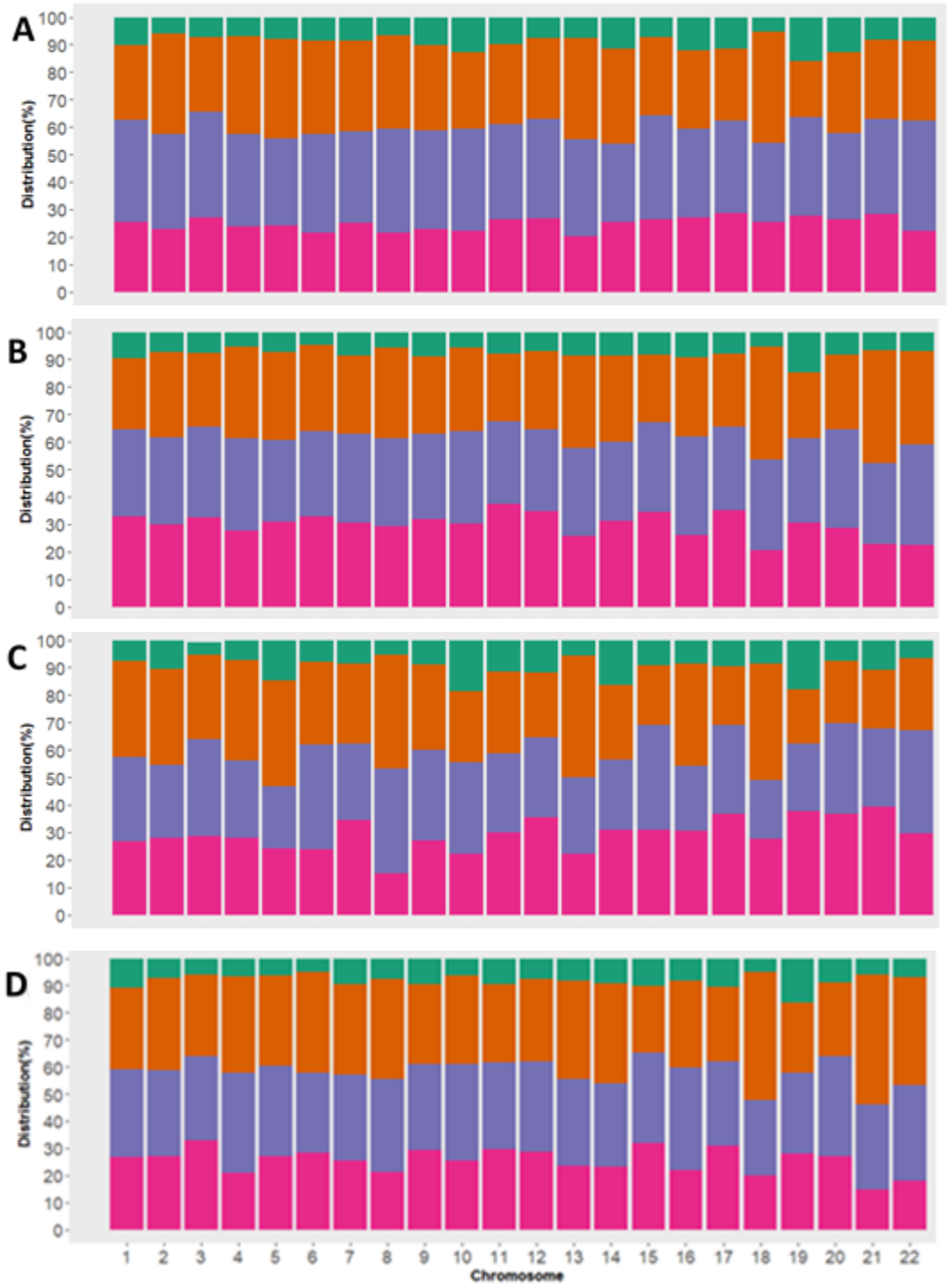
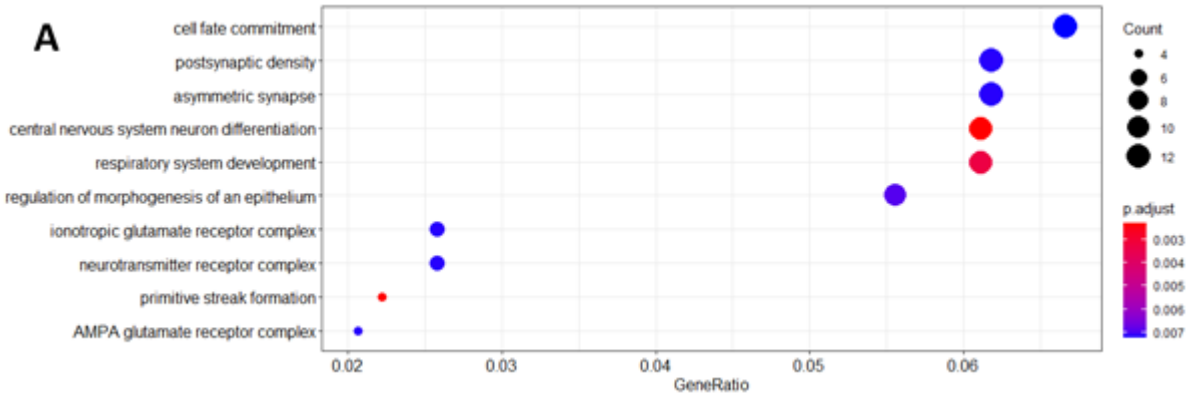
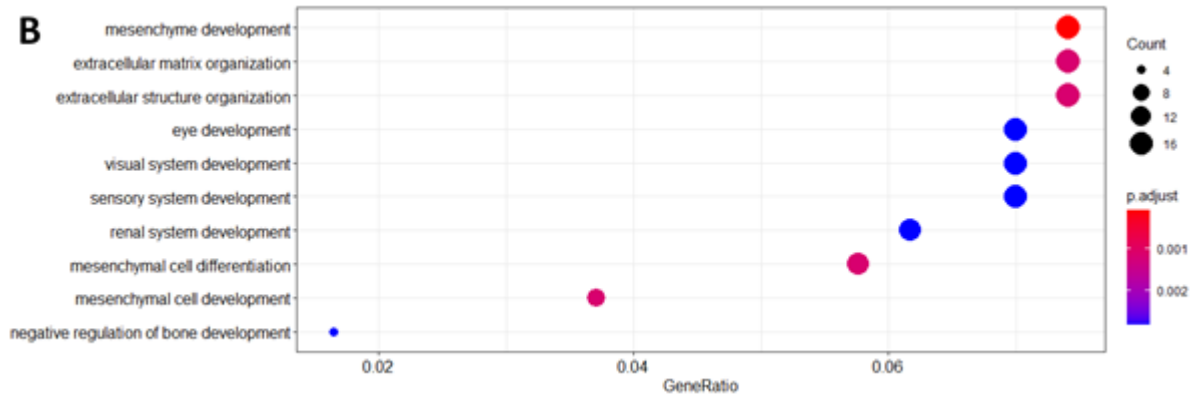


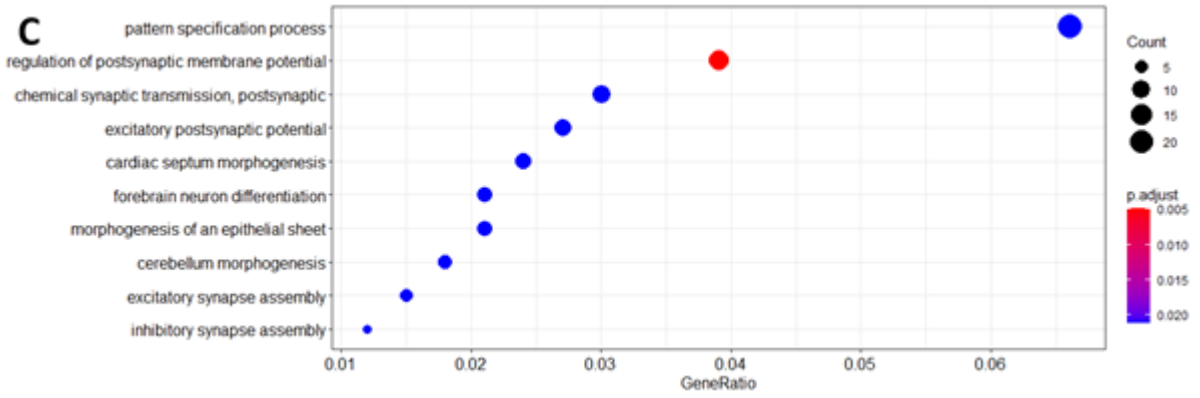
Figure 3. 10: The distribution of $1\alpha,25(OH)_2D_3$-induced differential methylation in genomic features was consistent across the chromosomes. The bar graphs show the distribution of DMCs (in THP-1 (A) and HEK293 (B) cells and DMRs (in THP-1 (C) and HEK293 (D) cells), for different genomic features including promoter regions (pink), introns (purple), intergenic regions (orange) and exons (green).

3.2.5 DMC distribution in promoter regions showed enrichment of nervous system and developmental biological processes

The genes targeted by 1α,25$(OH)_2D_3$-induced differential methylation often indicate which biological processes are influenced by 1α,25$(OH)_2D_3$ treatment. To explore which biological processes are enriched by 1α,25$(OH)_2D_3$ treatment, pathway analysis was conducted using clusterProfiler (Yu *et al.*, 2012). The genome was divided into adjacent 1 000 bp tiles and the regions with the highest number of DMCs (identified with methylKit) were intersected with promoter regions. Similarly, DMRs (identified by eDMR) were intersected with promoter regions. Promoter regions were created using the locations of transcription start sites from the Ensembl database. The gene IDs of the genes in these enriched promoter regions and the number of DMCs in these regions were used as gene list input for pathway analysis. In THP-1 cells, DMCs identified by methylKit, were enriched in the promoter regions of genes involved mostly (60%) in nervous-system related biological processes. In HEK293 cells, DMCs identified by methylKit, were enriched in the promoter regions of genes involved mostly (70%) in nervous system related biological processes. DMRs, identified by eDMR, were enriched in the promoter regions of genes involved in a wide range of biological processes.


A
cell fate commitment
postsynaptic density
asymmetric synapse
central nervous system neuron differentiation
respiratory system development
regulation of morphogenesis of an epithelium
ionotropic glutamate receptor complex
neurotransmitter receptor complex
primitive streak formation
AMPA glutamate receptor complex
0.02
0.03
0.04
0.05
0.06
GeneRatio
Count
4
6
8
10
12
p.adjust
0.003
0.004
0.005
0.006
0.007


B
mesenchyme development
extracellular matrix organization
extracellular structure organization
eye development
visual system development
sensory system development
renal system development
mesenchymal cell differentiation
mesenchymal cell development
negative regulation of bone development
0.02
0.04
0.06
GeneRatio
Count
4
8
12
16
p.adjust
0.001
0.002


C
pattern specification process
regulation of postsynaptic membrane potential
chemical synaptic transmission, postsynaptic
excitatory postsynaptic potential
cardiac septum morphogenesis
forebrain neuron differentiation
morphogenesis of an epithelial sheet
cerebellum morphogenesis
excitatory synapse assembly
inhibitory synapse assembly
0.01
0.02
0.03
0.04
0.05
0.06
GeneRatio
Count
5
10
15
20
p.adjust
0.005
0.010
0.015
0.020

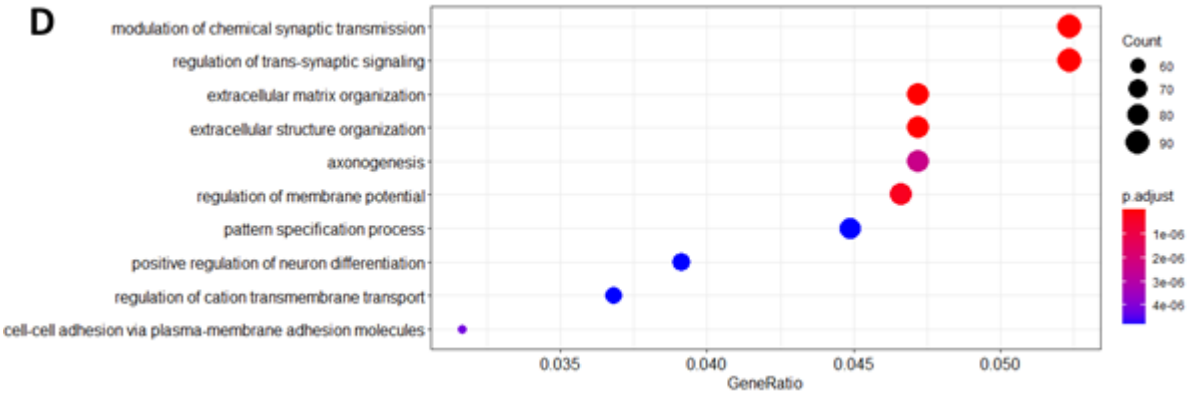

D
modulation of chemical synaptic transmission
regulation of trans-synaptic signaling
extracellular matrix organization
extracellular structure organization
axonogenesis
regulation of membrane potential
pattern specification process
positive regulation of neuron differentiation
regulation of cation transmembrane transport
cell-cell adhesion via plasma-membrane adhesion molecules
0.035
0.040
0.045
0.050
GeneRatio
Count
60
70
80
90
p.adjust
1e-06
2e-06
3e-06
4e-06

Figure 3. 11: Dotplot showing the biological processes whose gene promoter is most enriched for 1α,25$(OH)_2D_3$-induced differential methylation identified by methylKit (THP-1: A, HEK293: B) and eDMR (THP-1: C, HEK293: D). Circles represents the biological processes with genes enriched for 1α,25$(OH)_2D_3$-induced differential methylation. The colour of the circle represents the adjusted p-value. The size of the circle represents the number of genes in the set which are linked to that biological process. The x-axis represents the ratio of genes associated with a biological process over the total number of genes.

Chapter 4: Discussion

This study aimed to assess the impact of 1α,25$(OH)_2D_3$ on genome-wide methylation in various cell lines. RRBS was used to identify methylated cytosines because it specifically targets CpG-rich regions of the genome. This lowers the cost of sequencing and allows for methylation analysis in regions where differential methylation mostly occurs. methylKit was used to identify differentially methylated sites and eDMR was used to identify differentially methylated regions. 1α,25$(OH)_2D_3$ treatment induced methylation and demethylation events in both THP-1 and HEK293 cells. 1α,25$(OH)_2D_3$-induced differential methylation was targeted in intergenic regions and open sea CpG sites. Pathway analysis showed that 1α,25$(OH)_2D_3$-induced methylation changes were enriched in pathways specific to cell type and expression of genes required by that cell's state.

4.1 Quality control

4.1.1 High molecular weight gDNA was extracted

High molecular weight genomic DNA was extracted from HEK293, THP-1 and MCF-7 cells. The gDNA was intact and free of RNA and protein contamination. This was important for the study because degraded DNA and protein contaminants affect the quality of sequencing methods which use restriction enzyme digestion (Graham *et al.*, 2014). Protein contamination often negatively affects downstream sequencing by forming unmapped reads (Sangiovanni *et al.*, 2019). RNA contamination leads sequencing of RNA contaminants resulting in misinterpretation of results. Obtaining pure and intact genomic DNA ensured good quality libraries and confidence in sequencing results.

4.1.2 The pooled RRBS library fragment profile closely matched that of the standard human genome

*Msp*1 cleavage sites are species-specific and thus the number and size of fragments detected indicate the integrity and purity of the library pools. With the exception of two smaller fragments (224 and 254 bp), the size and number of fragments matched the standard human genome. This corroborated the spectrophotometry and gel electrophoresis results, suggesting that the gDNA obtained was mostly of good quality. However, the presence of the smaller fragments suggested the presence of foreign DNA in some of the samples. Since the DNA was free from RNA contamination, it is possible that another contaminate may have been present. According to (Razin, 1996), mycoplasma are a common cell culture contaminant. These bacteria do not possess a cell wall and are thus unaffected by common antibiotics (Razin, 1996). Moreover, their small size (0.15-0.3 µm size range) allows them to escape cell culture filtration (Razin, 1996). Taken together, these characteristics would have prevented elimination during cell culturing, and detection during DNA extraction, agarose gel electrophoresis or fragment analysis prior to restriction enzyme digestion. In our experiment, restriction enzyme digestion by *Msp*1 therefore acted as an unplanned quality control indicator for our pooled libraries.

4.1.3 The bisulphite conversion of the sample DNA was efficient and sequenced reads mapped successfully to the reference genome

The inclusion of unmethylated and methylated spike-in controls ensured confidence in the bisulphite conversion efficiency. A conversion rate of 1.5-2% and 99.5-99.8% for the methylated and unmethylated controls showed the bisulphite conversion

successfully and correctly converted unmethylated cytosines into thymines. Sequencing was successful in all three cell lines as there was a high output of reads following the experiment (27-49 million depending on the cell line). However, MCF-7 reads aligned very poorly (<1%) to the human reference genome compared to HEK293 and THP-1 reads (65-67%). Poor percentage alignment of reads corresponds with the *Msp*1 cleavage fragment profile, with both showing the presence of foreign DNA. As such, the MCF-7 samples were excluded from further analysis. This emphasised the importance of a clean and sterile work environment during cell culture and DNA extraction. This also highlighted the limitations of many contamination detection methods used in cell culture practise as they do not detect or prevent all possible contaminants. Frequent mycoplasma testing should be added to established decontamination practises during cell culture. This also emphasized the degree in which contamination during cell culturing and preparation can have detrimental effects on downstream experiments and analysis.

Nonetheless, the quality of methylation detected for THP-1 and HEK293 cells was high with the average coverage of >10 reads, with minimal PCR bias. Very high average coverage would have suggested PCR bias, as this is an experimental artifact resulting in some regions being amplified more than others due to the sequencing characteristics they possess (Warnecke *et al.*, 1997). Moreover, in both cell lines, a high percentage of cytosines were highly methylated or highly demethylated with very few intermediately methylated cytosines. This supports the bisulphite conversion efficiency percentages as it shows that methylated cytosines were not converted and unmethylated cytosines were fully converted (Kint *et al.*, 2018). Furthermore, the low levels of intermediately methylated cytosines show that CpG rich areas of the genome were successfully targeted by the RRBS. CpG methylation is often spatially correlated (Hansen et al., 2012) and therefore isolated CpG sites are more likely to be intermediately methylated.

4.2 Differential Methylation Analysis

4.2.1 The top differentially methylated sites were in regions with unknown functions and the top differentially methylated regions were in epigenetically-linked DNA.

Differential methylation identified by methylKit and eDMR was compared to determine the influence the method of analysis has on the differential methylation results obtained. methylKit uses logistic regression to identify differentially methylated cytosines and eDMR mathematically calculates the boundaries of proximal DMCs to identify differentially methylated regions (Li et al., 2013). Differential methylation was not identified in the same regions by the two methods. The sizes of the regions mapped by eDMR differed and were determined by the number of cytosines in the region and the coverage of those cytosines. The proximity of DMCs to each other heavily influenced how differentially methylated regions were identified. Furthermore, eDMR utilized information of pre-defined DMCs (Li et al., 2013) and therefore was influenced by the method used to identify those DMCs. methylKit was able identify differential methylation at CpG level. However, the CpG level differential methylation information did not provide insight into the function or effect of differential methylation.

methylKit and eDMR cannot be objectively compared because they ultimately measure different things. The interpretation of differential methylation is dependent on the output obtained from the method used to identify differential methylation. The biological question asked should therefore be the used to determine which method is best suited for differential methylation analysis. Additionally, the information obtained through differential methylation analysis alone is often not informative enough to conclusively answer a biological question. Answering a certain biological question requires contextualization to cell type, treatment, coverage, bisulphite conversion success rate and downstream analyses such as annotation and enrichment. For the purpose of the study, methylKit was the more appropriate method because we could

therefore assess the impact of 1α,25$(OH)_2D_3$ on genome-wide methylation by interrogating methylation changes at each individual CpG site.

4.2.2 Minimal technical variability in the methylomes of treated and untreated replicates in both THP-1 and HEK293 cells, respectively

To assess the technical variability within the samples, Pearson correlation coefficients of the differentially methylated cytosines were calculated and used to construct scatter plots of DNA methylation between replicates in THP-1 and HEK293 cells. The control-control and treated-treated comparison showed strong positive correlation indicating minimal variation within the replicates. This observation increases the confidence in the reproducibility of the experiment.

There was also a strong positive correlation between the control and treated replicates in both THP-1 and HEK293 cells suggesting that globally, 1α,25$(OH)_2D_3$ specific differences in methylation were low in both THP-1 and HEK293 cells. VDR is ubiquitously expressed in in the body (Bouillon *et al.*, 2008) and therefore 1α,25$(OH)_2D_3$-induced changes in methylation could occur in specific regions corresponding to the gene expression profiles of THP-1 and HEK293 cells specifically. It is possible that 1α,25$(OH)_2D_3$ induces methylation at some sites and demethylation at other sites, resulting in smaller observed global changes in methylation. This would explain the results obtained in this study. Additionally, VDR activity depends on 1α,25$(OH)_2D_3$ (Fetahu *et al.*, 2014), suggesting that 1α,25$(OH)_2D_3$ could regulate VDR-dependant methylation changes. The VDR itself is regulated at a pre-transcriptional, post-transcriptional and post-translational level (Saccone *et al.*, 2015). This strict VDR regulation could further result in methylation changes being of a transient nature.

4.2.3 1α,25(OH)$_2$D$_3$ induced differential methylation in both THP-1 and HEK293 cells

With the applied q-value and methylation difference cut-offs, the results obtained show that, on average, 1α,25(OH)$_2$D$_3$ induced differential methylation at four times more CpG sites and seven times more DMRs in HEK293 cells compared with THP-1 cells. Additionally, the percentage of hypomethylated and hypermethylated sites was similar in the two cell lines, suggesting that 1α,25(OH)$_2$D$_3$ is able to induce differential methylation regardless of cell type.

The results show that 1α,25(OH)$_2$D$_3$ treatment induced demethylation (47% of the DMCs and 49.8% of the DMRs) in HEK293 cells. Demethylation can occur passively through the dilution of the 5mC signal during DNA replication and cell proliferation (Smith and Meissner, 2013). 1α,25(OH)$_2$D$_3$ treatment has previous been found to reduce the expression and activity of DNMT1 and 3B in prostate cancer cells, as a mechanism for vitamin D resistance and cancer progression (Lai *et al.*, 2020). Although cancer cell behaviour is often specifically aberrant, an adoption of such a mechanism in HEK293 cells following 1α,25(OH)$_2$D$_3$ treatment would explain demethylation of CpG sites. However, 1α,25(OH)$_2$D$_3$ treatment was also found to induce VDR to bind to DNMT1 and 3B as a mechanism for gene repression through DNA methylation (Takeyama and Kato, 2011). This suggests that 1α,25(OH)$_2$D$_3$, with its receptor VDR, can influence the expression and activity of DNA methyltransferase enzymes but the effect of this influence differs depending on cell type and function. A more recent study however, found that 1α,25(OH)$_2$D$_3$ treatment led to the interaction of VDR with STAT3, and that the resulting transcription factor complex interacted with TET2 (Català-Moll *et al.*, 2020). Such 1α,25(OH)$_2$D$_3$-induced interactions in HEK293 would allow for targeted demethylation. In support, STAT3 is expressed and active in embryonic stem cells (Murphy *et al.*, 2005; Tai *et al.*, 2014) and would thus be available to interact with VDR following 1α,25(OH)$_2$D$_3$ treatment. From our study,

we cannot conclude whether demethylation was active through TET enzymes or passive though repression of DNMT 1 and 3B, however, considering that both forms of demethylation occur during embryonic development, 1α,25$(OH)_2D_3$-induced demethylation in HEK293 cells is most likely the result of a combination of the two.

The results also show that 1α,25$(OH)_2D_3$ treatment induced demethylation (48% of the DMCs and 50.1% of the DMRs) in THP-1 cells. DNA demethylation has been well studied in differentiating monocytes (Karpurapu et al., 2014; Mendes et al., 2021; Wallner et al., 2016). *Tet-2* expression is a common feature of DNA demethylation in differentiating monocytes (Klug et al., 2013). It is possible that DNA demethylation is necessary for initiating the maturation of elements required for monocytic lineages. Previous studies have differentiated monocytic cell lines to many different cell types using 1α,25$(OH)_2D_3$ (Piemonti et al., 2000; Zhang et al., 2012). The results obtained suggest that DNA demethylation is a crucial step in the process in which 1α,25$(OH)_2D_3$ differentiates THP-1 cells. Furthermore, Tet 2 activity in monocytes suggests that this DNA demethylation could be active.

There were four times less DMCs and seven times less DMRs identified in THP-1 cells than in HEK293 cells. This shows that 1α,25$(OH)_2D_3$ induced less global methylation changes in THP-1 cells compared to HEK293 cells. Many earlier studies have shown that 1α,25$(OH)_2D_3$ can differentiate the monocytic THP-1 cell line (Schwende *et al.*, 1996; Tóth *et al.*, 2021). THP-1 differentiation is driven by transcriptional changes (Gažová *et al.*, 2020). PU.1 is a "master transcription factor" involved in the differentiation of monocytes into macrophages (Bachurski and Trapnell, 2006; Rosa *et al.*, 2007). PU.1 regulates the expression of cell surface proteins such as immunoglobin receptors and integrins (Bachurski and Trapnell, 2006); all of which indicate differentiation into a macrophage-like state. 1α,25$(OH)_2D_3$ can modulate the binding of PU.1 with genomic sites (Seuter *et al.*, 2018). Although there is no evidence of direct binding between VDR and PU.1, overlap of PU.1 and VDR binding sites suggest possible co-location of the two transcription factors (Seuter *et al.*, 2018). PU.1 seems to be an important role player in the action of 1α,25$(OH)_2D_3$-induced monocyte-

macrophage differentiation. Interestingly, differential DNA methylation has been found to play a role in monocyte-macrophage differentiation. Gain of DNA methylation was located at transcription factor binding CpG sites including PU.1 factor and loss of DNA methylation was located at binding sites of transcription factors such as AP-1 (Dekkers *et al.*, 2019). Our study shows that there are differential methylation changes when THP-1 cells are treated with $1\alpha,25(OH)_2D_3$. These differential methylation changes possibly drive the transcriptional changes which occur during monocyte differentiation by influencing the binding of key transcription factors such as PU.1.

Interestingly, the distribution of total differential methylation differed between chromosomes in both cell lines. The distribution of total differential methylation in a chromosome is dependent on the landscape of that chromosome. Genes are non-randomly distributed within and between chromosomes and are often found in GC-rich regions of the chromosome (Saccone, 2008). GC density differs between the chromosomes and therefore the number of DMCs (and subsequent DMRs) would also differ. Genes of the same function cluster in space and not by proximity (Thévenin *et al.*, 2014). The results show that $1\alpha,25(OH)_2D_3$-induced differential methylation targets genes involved in different biological processes, the distribution of total differential methylation would therefore follow the pattern of the location of these target genes. Therefore, chromosomes with genes not targeted by $1\alpha,25(OH)_2D_3$-induced differential methylation would be associated with less total differential methylation and the chromosomes with $1\alpha,25(OH)_2D_3$ target genes, more total differential methylation. It is also important to note that RRBS targets CpG rich areas and thus excludes many regions of the genome. The results have also shown that differential methylation is not limited to CpG rich regions and can occur at isolated CpG sites. The distribution of total differential methylation noted in the study is therefore only a snapshot of the dynamic nature of epigenetic regulation, induced by $1\alpha,25(OH)_2D_3$, in monocytic-like and embryonic cells.

It should be noted that a fairly strict q-value cut-off of 0.010 was used for this analysis, excluding many CpG sites in both THP-1 and HEK293 cells (Table 3.2). It is likely that some of the excluded sites may be of great importance in the action of $1\alpha,25(OH)_2D_3$ and VDR. Early studies showed that methylation of a single CpG site in a transcription factor binding site can affect transcription (Watt and Molloy, 1988). Additionally, the methylation difference cut-off was set at 15%. Little is known on the minimum level of change in methylation that would have a significant effect on gene expression. Moreover, CpG methylation is correlated across the genome (Eckhardt *et al.*, 2006; Irizarry *et al.*, 2008) and thus differentially methylated CpG sites with a small level of methylation change could influence the methylation of surrounding CpG sites in the genome.

4.2.4 Differential methylation occurred most in intergenic regions and open sea CpG sites.

To determine the distribution of differential methylation, differentially methylated cytosines were annotated to genomic and methylome features. $1\alpha,25(OH)_2D_3$-induced differential methylation occurred most in intergenic regions in both THP-1 (35% DMCs, 30.91% DMRs) and HEK293 (32% DMCs, 32.39% DMRs) cells. Intergenic regions describe non-coding DNA found between genic regions in the genome. In support, a reduced representation bisulphite sequencing analysis of methylation in 82 cell lines, reported that DNA methylation was most variable in intergenic regions (ENCODE Project Consortium, 2012). Interestingly, transcription factor co-localization and binding also occurred in these regions (ENCODE Project Consortium, 2012). This suggests that intergenic regions are important for transcription initiation. Intergenic regions in THP-1 and HEK293 cells could therefore have been a target for $1\alpha,25(OH)_2D_3$-induced methylation mediated by VDR. Intergenic regions make up a large portion of the genome and thus their functions are not easy to characterize. This makes the purpose of this $1\alpha,25(OH)_2D_3$-induced targeting unclear. However,

1α,25$(OH)_2D_3$-activated VDR has been previously found to bind to enhancers in intergenic regions in order to regulate gene expression (Meyer *et al.*, 2010). Enhancer methylation plays a role in the growth of organs during embryonic development (Li *et al.*, 2018). Similarly, epigenetic regulation of enhancers is a crucial step in the mediation of monocyte differentiation by transcription factors (Pham *et al.*, 2012). The presence of enhancers in intergenic regions would make these regions favourable targets for 1α,25$(OH)_2D_3$-induced differential methylation in THP-1 and HEK293.

In terms of the methylome, differential methylation occurred most in open sea CpG sites in both THP-1 (54%) and HEK293 (48%). Additionally, in HEK293 cells, DMRs were also most located in open seas (34%) whereas, in THP-1 cells, DMRs were most located in CGIs (34%). This is an interesting finding because similar to intergenic regions, open sea CpG sites are abundant and currently poorly characterized. Such high levels of 1α,25$(OH)_2D_3$-induced differential methylation in these areas suggests that certain characteristics make intergenic regions and open sea CpG sites viable targets for gene regulation. Furthermore, in both THP-1 and HEK293 cells, DMCs in intergenic and open sea regions were both methylated and demethylated. Expression can be enhanced by demethylation in these cytosines, through enhanced transcription factor binding to alternate promoters, enhancers, insulators and other important regulatory elements found in these regions. Methylation of cytosines in this region might therefore serve to prevent expression by interfering with transcription factor binding to important regulatory elements. Overall, this would maintain genomic stability through strict regulation of expressed and repressed DNA.

In both THP-1 and HEK293 cells, CpG shore and CpG shelf differential methylation co-occurred with CGI methylation. This suggests that differential methylation is correlated across the methylome as supported by previous research (Eckhardt et al., 2006; Irizarry et al., 2008). This also suggests that differential methylation in a region can influence the expression of different genes depending on proximity to that region.

In both THP-1 and HEK293 cells, 1α,25$(OH)_2D_3$-induced differential methylation also occurred substantially in promoters (THP-1: 25% DMCs and 29.24% DMRs, HEK293:

30% DMCs and 26.18 DMRs). CpG island and promoter methylation are extensively studied events and this finding supports previous research (Deaton and Bird, 2011). Promoter methylation is known to regulate gene expression. There is no consensus on the mechanism of this regulation. Promoter regions can be methylated or demethylated, leading to silencing or expression, respectively (Calvanese *et al.*, 2012; Yang and Park, 2012). The results show that $1\alpha,25(OH)_2D_3$ induced both methylation and demethylation at promoter regions. It is known that $1\alpha,25(OH)_2D_3$ and VDR can up-regulate (Pike and Meyer, 2010) and down-regulate (Li *et al.*, 2015) gene expression depending on the target gene. This suggests that the methylation state of promoter regions alone does not determine the type of gene regulation which will occur. A recent review (Smith *et al.*, 2020), explored promoter methylation-mediated gene activation and its potential mechanisms including inhibition of transcription repressor binding, interaction with distal elements such as enhancers and insulators as well as expression through alternate promoters. It is likely that the interaction (or prevention thereof) between regulatory molecules and differentially methylated promoter regions determines the type of gene regulation which will occur. $1\alpha,25(OH)_2D_3$-induced differential methylation also occurred substantially in CpG islands (THP-1: 26%, HEK293: 35%). CpG islands are characterized by a transcriptionally permissive chromatin state (Tazi and Bird, 1990) and are mostly distributed in gene rich loci (Guenther et al., 2007b; Lander et al., 2001; Weber et al., 2005). Furthermore, CpG islands co-localize with promoters (Zhu *et al.*, 2008) and can even serve as alternate promoters (Sarda *et al.*, 2017). $1\alpha,25(OH)_2D_3$-induced differential methylation at these CpG island promoter regions can affect gene expression. Differential methylation at CpG island promoters is tissue-specific (Eckhardt *et al.*, 2006) and differs in germ lines compared to somatic cells (Shen *et al.*, 2007). CpG island promoter differential methylation induced by $1\alpha,25(OH)_2D_3$ in THP-1 cells can lead to differentiation through expression of genes required for the macrophage state. Similar, CpG island promoter differential methylation induced by $1\alpha,25(OH)_2D_3$ in HEK293 cells can drive development by changing the expression of genes. As seen with genes such as *PAX6*

and *OSR1*, which are associated with differential methylation at CpG island promoters during development (Illingworth *et al.*, 2008).

4.2.5 DMC distribution in promoter regions showed the tissue-specific effects of 1α,25$(OH)_2D_3$ treatment on differential methylation

Pathway analysis of genes with differential methylation enriched promoter regions was conducted with ClusterProfiler. The top enriched biological processes targeted by 1α,25$(OH)_2D_3$-induced differential methylation were nervous system-related in THP-1 cells (40%) and development-related in HEK293 cells (70%). Vitamin D is a key role player in brain function. VDR is widely expressed in the brain (Eyles *et al.*, 2005). 1α,25$(OH)_2D_3$ synthesizing enzymes are present in neurons as 1α,25$(OH)_2D_3$ influences replication, differentiation and survival in those cells (Fernandes de Abreu *et al.*, 2009). 1α,25$(OH)_2D_3$ also contributes to synaptic plasticity, neuroprotection and the dopaminergic signalling (Bivona *et al.*, 2019). In agreement, the results show that 1α,25$(OH)_2D_3$ induces differential methylation in the promoter regions of genes involved in these processes, albeit in monocytes. While the role of monocytes in the nervous system is not well known. Monocytes are recruited into the nervous system in many disorders (Zhao *et al.*, 2020). For example, in Alzheimer's disease, monocytes are recruited into the brain in response to inflammation. Macrophages derived from these monocytes play a crucial role in preventing plaque formation (Simard et al., 2006). The observation that in 1α,25$(OH)_2D_3$-induced differential methylation can target nervous system-related genes in THP-1 cells suggests that the presence and actions of 1α,25$(OH)_2D_3$ in partnership with monocyte recruitment contribute to nervous system functionality and protection.

The top enriched biological processes targeted by 1α,25$(OH)_2D_3$-induced differential methylation were development-related in HEK293 cells (70%). Methylation is a key feature of embryonic development as evident in the high expression of DNMT enzymes

(Kaneda *et al.*, 2004; Ueda *et al.*, 2006). Similarly, demethylation is also important in embryonic development as it allows for establishment of gene expression profiles required for cell differentiation (Bernstein and Bernstein, 2019). The role of $1\alpha,25(OH)_2D_3$ in embryonic development is less well known. However, $25\alpha(OH)D_3$ deficiency is associated with many pregnancy disorders such as preterm birth, gestational diabetes and foetal development defects (Wagner and Hollis, 2018). This suggests that $1\alpha,25(OH)_2D_3$ is a requirement for foetal development. This is supported by the results which show that $1\alpha,25(OH)_2D_3$-induced differential methylation in HEK293 cells targets genes involved in many developmental processes. It is important to note that over-representation analysis has its limitations. It does not take into account the effect of methylation and demethylation in enriched genes. Furthermore, enrichment is calculated as a comparison of genes within the gene set only (Boyle et al., 2004). The restriction would ignore the effect of differential methylation in many other genes involved in the same biological processes. In order to get the complete global physiological response picture, over-representation analysis would have to be conducted on genes whose full complement of regulatory regions is targeted by $1\alpha,25(OH)_2D_3$-induced differential methylation.

Chapter 5: Conclusion

In this study, we utilized RRBS to identify genome-wide differential methylation induced by 1α,25$(OH)_2D_3$ in THP-1 and HEK293 cells and explored the use of different data analysis methods of identifying differential methylation. 1α,25$(OH)_2D_3$ induced methylation and demethylation events in both THP-1 and HEK293 cells. Because 1α,25$(OH)_2D_3$ enacts by binding its receptor VDR, we postulate that VDR was in fact responsible for initiating methylation and demethylation events. VDR is able to bind many other regulatory molecules. Our previous work found that VDR binding partners such as GABPA, have binding motifs in the 5' and 3' untranslated regions of *DNMT* and *TET* genes (Appendix B). We postulate that VDR along with its binding partners increase the expression of DNA methylating and demethylating enzymes in response to 1α,25$(OH)_2D_3$ treatment. We also postulate that 1α,25$(OH)_2D_3$-activated VDR can bind to other molecules and direct DNA methylating and demethylating enzymes to specific target genes. Finally, we postulate that this VDR-mediated targeted differential methylation plays a role in 1α,25$(OH)_2D_3$-induced physiological responses seen in THP-1 and HEK293 cells such as differentiation.

The study showed that RRBS is a cost-effective method of investigating changes in methylation. Furthermore, the study demonstrated how pathway enrichment analysis could be coupled with differential methylation analysis to assess biological relevance of changes in DNA methylation. This approach to differential methylation analysis is limited however, because it only uses the location of differential methylation to identify potentially affected genes. The impact of differential methylation on gene expression are dynamic and are often the culmination of many molecular events. For example, silencer methylation (Govender et al., 2021) and promoter demethylation (Bogdarina et al., 2007) can potentially increase the expression of renin-angiotensin-aldosterone pathway genes. Future work can therefore include partnering this analysis approach with gene expression analysis to more accurately assess the biological impact of differential methylation, as influenced by 1α,25$(OH)_2D_3$. This would better explain the

link between differential methylation and changes in gene expression. Furthermore, such work would extend our understanding of how methylation of different DNA elements can lead to different biological effects depending on the gene.

Although we postulate an integral role for VDR we did not measure VDR levels or activity. This was a major limitation to our study as the role of VDR can therefore not be accurately described. Future work can include comparing $1\alpha,25(OH)_2D_3$-induced differential methylation in VDR-silenced cells. Future work can also additionally investigate VDR activity and expression in response to treatment and assess how that corresponds to the observed differential methylation. Increasing focus on VDR can also provide insight into potential mechanisms with which $1\alpha,25(OH)_2D_3$ causes methylation and demethylation events. VDR is a very versatile binding partner and depending on which molecules bind to VDR, this ability can potentially guide methylating and demethylating machinery to $1\alpha,25(OH)_2D_3$ targets.

The study also looked at the overall, global impact of $1\alpha,25(OH)_2D_3$ on methylation and thus did not separate methylation and demethylation events. The type of differential methylation which occurs determines the type of gene regulation which occurs as a result. With the knowledge that the relationship between DNA methylation and gene expression is not classical such that methylated DNA leads to gene suppression. Separating the analysis of methylation events and demethylation events will lead to a better understanding of the dynamic nature with which differential methylation influences gene expression.

References

Ahonen, M.H., Tenkanen, L., Teppo, L., Hakama, M., Tuohimaa, P., 2000. Prostate cancer risk and prediagnostic serum 25-hydroxyvitamin D levels (Finland). Cancer Causes Control CCC 11, 847–852.

Aird, D., Ross, M.G., Chen, W.-S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C., Gnirke, A., 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol. 12, R18. https://doi.org/10.1186/gb-2011-12-2-r18

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A., Mason, C.E., 2012. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol. 13, R87. https://doi.org/10.1186/gb-2012-13-10-r87

Alimirah, F., Vaishnav, A., McCormick, M., Echchgadda, I., Chatterjee, B., Mehta, R.G., Peng, X., 2010. Functionality of unliganded VDR in breast cancer cells: repressive action on CYP24 basal transcription. Mol. Cell. Biochem. 342, 143–150. https://doi.org/10.1007/s11010-010-0478-6

Alzaman, N.S., Dawson-Hughes, B., Nelson, J., D'Alessio, D., Pittas, A.G., 2016. Vitamin D status of black and white Americans and changes in vitamin D metabolites after varied doses of vitamin D supplementation12. Am. J. Clin. Nutr. 104, 205–214. https://doi.org/10.3945/ajcn.115.129478

Anderson, C.M., Gillespie, S.L., Thiele, D.K., Ralph, J.L., Ohm, J.E., 2018. Effects of Maternal Vitamin D Supplementation on the Maternal and Infant Epigenome. Breastfeed. Med. 13, 371–380. https://doi.org/10.1089/bfm.2017.0231

Aran, D., Toperoff, G., Rosenberg, M., Hellman, A., 2011. Replication timing-related and gene body-specific methylation of active human genes. Hum. Mol. Genet. 20, 670–680. https://doi.org/10.1093/hmg/ddq513

Arand, J., Spieler, D., Karius, T., Branco, M.R., Meilinger, D., Meissner, A., Jenuwein, T., Xu, G., Leonhardt, H., Wolf, V., Walter, J., 2012. In Vivo Control of CpG and Non-CpG DNA Methylation by DNA Methyltransferases. PLOS Genet. 8, e1002750. https://doi.org/10.1371/journal.pgen.1002750

Armas, L.A.G., Dowell, S., Akhter, M., Duthuluru, S., Huerter, C., Hollis, B.W., Lund, R., Heaney, R.P., 2007. Ultraviolet-B radiation increases serum 25-hydroxyvitamin D levels: the effect of UVB dose and skin color. J. Am. Acad. Dermatol. 57, 588–593. https://doi.org/10.1016/j.jaad.2007.03.004

Armas, L.A.G., Hollis, B.W., Heaney, R.P., 2004. Vitamin D2 is much less effective than vitamin D3 in humans. J. Clin. Endocrinol. Metab. 89, 5387–5391. https://doi.org/10.1210/jc.2004-0360

Avery, O.T., Macleod, C.M., McCarty, M., 1944. Studies on the chemical nature of inducing transformation of Pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. J. Exp. Med. 79, 137–158.

Aydin, F.N., Aydin, I., Agilli, M., 2014. Factors Affecting the Accuracy of Vitamin D Measurements. Clin. Med. Res. 12, 4. https://doi.org/10.3121/cmr.2014.1231

Bachurski, C.J., Trapnell, B.C., 2006. TRANSCRIPTION FACTORS | PU.1, in: Laurent, G.J., Shapiro, S.D. (Eds.), Encyclopedia of Respiratory Medicine. Academic Press, Oxford, pp. 274–278. https://doi.org/10.1016/B0-12-370879-6/00393-8

Baeke, F., Takiishi, T., Korf, H., Gysemans, C., Mathieu, C., 2010. Vitamin D: modulator of the immune system. Curr. Opin. Pharmacol. 10, 482–496. https://doi.org/10.1016/j.coph.2010.04.001

Barrett, A.J., Woessner, J.F., Rawlings, N.D., 2012. Handbook of Proteolytic Enzymes, Volume 1. Elsevier.

Barros-Silva, D., Marques, C.J., Henrique, R., Jerónimo, C., 2018. Profiling DNA Methylation Based on Next-Generation Sequencing Approaches: New Insights and Clinical Applications. Genes 9, 429. https://doi.org/10.3390/genes9090429

Bell, J.S.K., Vertino, P.M., 2017. Orphan CpG islands define a novel class of highly active enhancers. Epigenetics 12, 449–464. https://doi.org/10.1080/15592294.2017.1297910

Berg, J.M., 1989. DNA binding specificity of steroid receptors. Cell 57, 1065–1068.

Bernstein, C., Bernstein, H., 2019. Demethylation in Early Embryonic Development and Memory, DNA Methylation Mechanism. IntechOpen. https://doi.org/10.5772/intechopen.90306

Bikle, D.D., 2014. Vitamin D Metabolism, Mechanism of Action, and Clinical Applications. Chem. Biol. 21, 319–329. https://doi.org/10.1016/j.chembiol.2013.12.016

Bird, A.P., 1980. DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res. 8, 1499–1504. https://doi.org/10.1093/nar/8.7.1499

Bivona, G., Gambino, C.M., Iacolino, G., Ciaccio, M., 2019. Vitamin D and the nervous system. Neurol. Res. 41, 827–835. https://doi.org/10.1080/01616412.2019.1622872

Bogdarina, I., Welham, S., King, P.J., Burns, S.P., Clark, A.J.L., 2007. Epigenetic Modification of the Renin-Angiotensin System in the Fetal Programming of Hypertension. Circ. Res. 100, 520–526. https://doi.org/10.1161/01.RES.0000258855.60637.58

Bonilla, C., Ness, A.R., Wills, A.K., Lawlor, D.A., Lewis, S.J., Davey Smith, G., 2014. Skin pigmentation, sun exposure and vitamin D levels in children of the Avon Longitudinal Study of Parents and Children. BMC Public Health 14, 597. https://doi.org/10.1186/1471-2458-14-597

Bouillon, R., Carmeliet, G., Verlinden, L., van Etten, E., Verstuyf, A., Luderer, H.F., Lieben, L., Mathieu, C., Demay, M., 2008. Vitamin D and Human Health: Lessons from Vitamin D Receptor Null Mice. Endocr. Rev. 29, 726–776. https://doi.org/10.1210/er.2008-0004

Boyle, E.I., Weng, S., Gollub, J., Jin, H., Botstein, D., Cherry, J.M., Sherlock, G., 2004. GO::TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. Bioinformatics 20, 3710–3715. https://doi.org/10.1093/bioinformatics/bth456

Brøndum-Jacobsen, P., Benn, M., Jensen, G.B., Nordestgaard, B.G., 2012. 25-hydroxyvitamin d levels and risk of ischemic heart disease, myocardial infarction, and early death: population-based study and meta-analyses of 18 and 17 studies. Arterioscler. Thromb. Vasc. Biol. 32, 2794–2802. https://doi.org/10.1161/ATVBAHA.112.248039

Calvanese, V., Fernández, A.F., Urdinguio, R.G., Suárez-Alvarez, B., Mangas, C., Pérez-García, V., Bueno, C., Montes, R., Ramos-Mejía, V., Martínez-Camblor, P., Ferrero, C., Assenov, Y., Bock, C., Menendez, P., Carrera, A.C., Lopez-Larrea, C., Fraga, M.F., 2012. A promoter DNA demethylation landscape of human hematopoietic differentiation. Nucleic Acids Res. 40, 116–131. https://doi.org/10.1093/nar/gkr685

Cavalcante, R.G., Sartor, M.A., 2017. annotatr: genomic regions in context. Bioinformatics 33, 2381–2383. https://doi.org/10.1093/bioinformatics/btx183

Chatterjee, A., Rodger, E.J., Stockwell, P.A., Weeks, R.J., Morison, I.M., 2012. Technical Considerations for Reduced Representation Bisulfite Sequencing with Multiplexed Libraries. J. Biomed. Biotechnol. 2012, e741542. https://doi.org/10.1155/2012/741542

Chen, L., Dong, Y., Chen, J., Huang, Y., Zhu, H., 2020. Epigenetics Predicts Serum 25-Hydroxyvitamin D Response to Vitamin D3 Supplementation in African Americans. Mol. Nutr. Food Res. 64, 1900738. https://doi.org/10.1002/mnfr.201900738

Cooper, D.N., Taggart, M.H., Bird, A.P., 1983. Unmethylated domains in vertebrate DNA. Nucleic Acids Res. 11, 647–658. https://doi.org/10.1093/nar/11.3.647

Costin, G.-E., Hearing, V.J., 2007. Human skin pigmentation: melanocytes modulate skin color in response to stress. FASEB J. 21, 976–994. https://doi.org/10.1096/fj.06-6649rev

Crabtree, J.S., 2020. Chapter 3 - Fundamentals of epigenetics, in: Crabtree, J.S. (Ed.), Clinical Precision Medicine. Academic Press, pp. 27–37. https://doi.org/10.1016/B978-0-12-819834-6.00003-3

Dawson-Hughes, B., Heaney, R.P., Holick, M.F., Lips, P., Meunier, P.J., Vieth, R., 2005. Estimates of optimal vitamin D status. Osteoporos. Int. J. Establ. Result Coop. Eur. Found. Osteoporos. Natl. Osteoporos. Found. USA 16, 713–716. https://doi.org/10.1007/s00198-005-1867-7

Deaton, A.M., Bird, A., 2011. CpG islands and the regulation of transcription. Genes Dev. 25, 1010–1022. https://doi.org/10.1101/gad.2037511

Dekkers, K.F., Neele, A.E., Jukema, J.W., Heijmans, B.T., de Winther, M.P.J., 2019. Human monocyte-to-macrophage differentiation involves highly localized gain and loss of DNA methylation at transcription factor binding sites. Epigenetics Chromatin 12, 34. https://doi.org/10.1186/s13072-019-0279-4

DiNardo, A.R., Rajapakshe, K., Nishiguchi, T., Grimm, S.L., Mtetwa, G., Dlamini, Q., Kahari, J., Mahapatra, S., Kay, A., Maphalala, G., Mace, E.M., Makedonas, G., Cirillo, J.D., Netea, M.G., Crevel, R. van, Coarfa, C., Mandalakas, A.M., 2020. DNA hypermethylation during tuberculosis dampens host immune responsiveness. J. Clin. Invest. 130. https://doi.org/10.1172/JCI134622

Dmitrijeva, M., Ossowski, S., Serrano, L., Schaefer, M.H., 2018. Tissue-specific DNA methylation loss during ageing and carcinogenesis is linked to chromosome structure, replication timing and cell division rates. Nucleic Acids Res. 46, 7022–7039. https://doi.org/10.1093/nar/gky498

Dong, A., Yoder, J.A., Zhang, X., Zhou, L., Bestor, T.H., Cheng, X., 2001. Structure of human DNMT2, an enigmatic DNA methyltransferase homolog that displays denaturant-resistant binding to DNA. Nucleic Acids Res. 29, 439–448. https://doi.org/10.1093/nar/29.2.439

Durazo-Arvizu, R.A., Camacho, P., Bovet, P., Forrester, T., Lambert, E.V., Plange-Rhule, J., Hoofnagle, A.N., Aloia, J., Tayo, B., Dugas, L.R., Cooper, R.S., Luke, A., 2014. 25-Hydroxyvitamin D in African-origin populations at varying latitudes challenges the construct of a physiologic norm. Am. J. Clin. Nutr. 100, 908–914. https://doi.org/10.3945/ajcn.113.066605

Dytham, C., 2011. Choosing and using statistics: a biologist's guide, 3rd ed. ed. Wiley-Blackwell, Hoboken, NJ.

Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V.K., Attwood, J., Burger, M., Burton, J., Cox, T.V., Davies, R., Down, T.A., Haefliger, C., Horton, R., Howe, K., Jackson, D.K., Kunde, J., Koenig, C., Liddle, J., Niblett, D., Otto, T., Pettett, R., Seemann, S., Thompson, C., West, T., Rogers, J., Olek, A., Berlin, K., Beck, S., 2006. DNA methylation profiling of human chromosomes 6, 20 and 22. Nat. Genet. 38, 1378–1385. https://doi.org/10.1038/ng1909

Edgar, R., Tan, P.P.C., Portales-Casamar, E., Pavlidis, P., 2014. Meta-analysis of human methylomes reveals stably methylated sequences surrounding CpG islands associated with high gene expression. Epigenetics Chromatin 7, 28. https://doi.org/10.1186/1756-8935-7-28

ENCODE Project Consortium, 2012. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74. https://doi.org/10.1038/nature11247

Eyles, D.W., Smith, S., Kinobe, R., Hewison, M., McGrath, J.J., 2005. Distribution of the vitamin D receptor and 1 alpha-hydroxylase in human brain. J. Chem. Neuroanat. 29, 21–30. https://doi.org/10.1016/j.jchemneu.2004.08.006

Feinberg, A.P., Koldobskiy, M.A., Göndör, A., 2016. Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. Nat. Rev. Genet. 17, 284–299. https://doi.org/10.1038/nrg.2016.13

Feldman, D., Malloy, P.J., Krishnan, A.V., Balint, E., 2008. CHAPTER 13 - Vitamin D: Biology, Action, and Clinical Implications, in: Marcus, R., Feldman, D., Nelson, D.A., Rosen, C.J. (Eds.), Osteoporosis (Third Edition). Academic Press, San Diego, pp. 317–382. https://doi.org/10.1016/B978-012370544-0.50015-X

Feltus, F.A., Lee, E.K., Costello, J.F., Plass, C., Vertino, P.M., 2003. Predicting aberrant CpG island methylation. Proc. Natl. Acad. Sci. 100, 12253–12258. https://doi.org/10.1073/pnas.2037852100

Fernandes de Abreu, D.A., Eyles, D., Féron, F., 2009. Vitamin D, a neuro-immunomodulator: implications for neurodegenerative and autoimmune diseases. Psychoneuroendocrinology 34 Suppl 1, S265-277. https://doi.org/10.1016/j.psyneuen.2009.05.023

Fetahu, I.S., Höbaus, J., Kállay, E., 2014. Vitamin D and the epigenome. Front. Physiol. 5. https://doi.org/10.3389/fphys.2014.00164

Finnegan, S., Percival, S.L., 2015. EDTA: An Antimicrobial and Antibiofilm Agent for Use in Wound Care. Adv. Wound Care 4, 415–421. https://doi.org/10.1089/wound.2014.0577

Fuks, F., Hurd, P.J., Wolf, D., Nan, X., Bird, A.P., Kouzarides, T., 2003. The Methyl-CpG-binding Protein MeCP2 Links DNA Methylation to Histone Methylation. J. Biol. Chem. 278, 4035–4040. https://doi.org/10.1074/jbc.M210256200

Gažová, I., Lefevre, L., Bush, S.J., Clohisey, S., Arner, E., de Hoon, M., Severin, J., van Duin, L., Andersson, R., Lengeling, A., Hume, D.A., Summers, K.M., 2020. The Transcriptional Network That Controls Growth Arrest and Macrophage Differentiation in the Human Myeloid Leukemia Cell Line THP-1. Front. Cell Dev. Biol. 8. https://doi.org/10.3389/fcell.2020.00498

Ghosh, S., Yates, A.J., Fruhwald, M.C., Miecznikowski, J.C., Plass, C., Smiraglia, D.J., 2010. Tissue specific DNA methylation of CpG islands in normal human adult somatic tissues distinguishes neural from non-neural tissues. Epigenetics 5, 527–538. https://doi.org/10.4161/epi.5.6.12228

Goll, M.G., Kirpekar, F., Maggert, K.A., Yoder, J.A., Hsieh, C.-L., Zhang, X., Golic, K.G., Jacobsen, S.E., Bestor, T.H., 2006. Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science 311, 395–398. https://doi.org/10.1126/science.1120976

Gopal-Kothandapani, J.S., Evans, L.F., Walsh, J.S., Gossiel, F., Rigby, A.S., Eastell, R., Bishop, N.J., 2019. Effect of vitamin D supplementation on free and total vitamin D: A comparison of Asians and Caucasians. Clin. Endocrinol. (Oxf.) 90, 222–231. https://doi.org/10.1111/cen.13825

Gorham, E.D., Garland, C.F., Garland, F.C., Grant, W.B., Mohr, S.B., Lipkin, M., Newmark, H.L., Giovannucci, E., Wei, M., Holick, M.F., 2005. Vitamin D and prevention of colorectal cancer. J. Steroid Biochem. Mol. Biol., Vitamin D Workshop Meeting on VITAMIN D AND CANCER 97, 179–194. https://doi.org/10.1016/j.jsbmb.2005.06.018

Govender, D., Damjanovic, L., Gaza, C.A., Meyer, V., 2021. Vitamin D decreases silencer methylation to downregulate renin gene expression. Gene 786, 145623. https://doi.org/10.1016/j.gene.2021.145623

Graham, C., Boreham, D., Glenn, T., Lance, S., Martino, J., Manson, R., McArthur, A., Rogers, S., Wilson, J., Somers, C., 2014. Degraded DNA and next-generation sequencing: low quality DNA affects double digest restriction associated DNA sequencing. Genome 57, 390–390.

Grange, J.M., Davies, P.D., Brown, R.C., Woodhead, J.S., Kardjito, T., 1985. A study of vitamin D levels in Indonesian patients with untreated pulmonary tuberculosis. Tubercle 66, 187–191. https://doi.org/10.1016/0041-3879(85)90035-2

Green, R.J., Samy, G., Miqdady, M.S., El-Hodhod, M., Akinyinka, O.O., Saleh, G., Haddad, J., Alsaedi, S.A., Mersal, A.Y., Edris, A., Salah, M., 2015. Vitamin D deficiency and insufficiency in Africa and the Middle East, despite year-round sunny days. South Afr. Med. J. Suid-Afr. Tydskr. Vir Geneeskd. 105, 603–605.

Griffiths, A.J., Miller, J.H., Suzuki, D.T., Lewontin, R.C., Gelbart, W.M., 2000. Transcription: an overview of gene regulation in eukaryotes. Introd. Genet. Anal. 7th Ed.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., Young, R.A., 2007a. A chromatin landmark and transcription initiation at most promoters in human cells. Cell 130, 77–88. https://doi.org/10.1016/j.cell.2007.05.042

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., Young, R.A., 2007b. A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells. Cell 130, 77–88. https://doi.org/10.1016/j.cell.2007.05.042

Gupta, A., Montepiedra, G., Gupte, A., Zeldow, B., Jubulis, J., Detrick, B., Violari, A., Madhi, S., Bobat, R., Cotton, M., Mitchell, C., Spector, S., 2016. Low Vitamin-D Levels Combined with PKP3-SIGIRR-TMEM16J Host Variants Is Associated with Tuberculosis and Death in HIV-Infected and -Exposed Infants. PLoS ONE 11. https://doi.org/10.1371/journal.pone.0148649

Gupta, K., Mandal, D., Das, B., 2018. Serum and Bronchoalveolar Lavage Fluid 25(OH)Vitamin D3 Levels in HIV-1 and Tuberculosis: A Cross-Sectional Study from a Tertiary Care Center in North India. Curr. HIV Res. 16. https://doi.org/10.2174/1570162X16666180528112924

Gutiérrez, O.M., Farwell, W.R., Kermah, D., Taylor, E.N., 2011. Racial differences in the relationship between vitamin D, bone mineral density, and parathyroid hormone in the National Health and Nutrition Examination Survey. Osteoporos. Int. J. Establ. Result Coop. Eur. Found. Osteoporos. Natl. Osteoporos. Found. USA 22, 1745–1753. https://doi.org/10.1007/s00198-010-1383-2

Hansen, K.D., Langmead, B., Irizarry, R.A., 2012. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. Genome Biol. 13, R83. https://doi.org/10.1186/gb-2012-13-10-r83

Hermann, A., Goyal, R., Jeltsch, A., 2004. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. J. Biol. Chem. 279, 48350–48359. https://doi.org/10.1074/jbc.M403427200

Herrmann, M., Farrell, C.-J.L., Pusceddu, I., Fabregat-Cabello, N., Cavalier, E., 2017. Assessment of vitamin D status – a changing landscape. Clin. Chem. Lab. Med. CCLM 55, 3–26. https://doi.org/10.1515/cclm-2016-0264

Holick, M.F., 2003. Vitamin D: A millenium perspective. J. Cell. Biochem. 88, 296–307. https://doi.org/10.1002/jcb.10338

Hoogendijk, W.J.G., Lips, P., Dik, M.G., Deeg, D.J.H., Beekman, A.T.F., Penninx, B.W.J.H., 2008. Depression is associated with decreased 25-hydroxyvitamin D and increased parathyroid hormone levels in older adults. Arch. Gen. Psychiatry 65, 508–512. https://doi.org/10.1001/archpsyc.65.5.508

Hyun, K., Jeon, J., Park, K., Kim, J., 2017. Writing, erasing and reading histone lysine methylations. Exp. Mol. Med. 49, e324–e324. https://doi.org/10.1038/emm.2017.11

Illingworth, R., Kerr, A., DeSousa, D., Jørgensen, H., Ellis, P., Stalker, J., Jackson, D., Clee, C., Plumb, R., Rogers, J., Humphray, S., Cox, T., Langford, C., Bird,

A., 2008. A Novel CpG Island Set Identifies Tissue-Specific Methylation at Developmental Gene Loci. PLOS Biol. 6, e22. https://doi.org/10.1371/journal.pbio.0060022

Illingworth, R.S., Bird, A.P., 2009. CpG islands--'a rough guide'. FEBS Lett. 583, 1713–1720. https://doi.org/10.1016/j.febslet.2009.04.012

Irizarry, R.A., Ladd-Acosta, C., Carvalho, B., Wu, H., Brandenburg, S.A., Jeddeloh, J.A., Wen, B., Feinberg, A.P., 2008. Comprehensive high-throughput arrays for relative methylation (CHARM). Genome Res. 18, 780–790. https://doi.org/10.1101/gr.7301508

Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., Ji, H., Potash, J., Sabunciyan, S., Feinberg, A.P., 2009. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat. Genet. 41, 178–186. https://doi.org/10.1038/ng.298

Iurlaro, M., von Meyenn, F., Reik, W., 2017. DNA methylation homeostasis in human and mouse development. Curr. Opin. Genet. Dev. 43, 101–109. https://doi.org/10.1016/j.gde.2017.02.003

Jeltsch, A., Ehrenhofer-Murray, A., Jurkowski, T.P., Lyko, F., Reuter, G., Ankri, S., Nellen, W., Schaefer, M., Helm, M., 2017. Mechanism and biological role of Dnmt2 in Nucleic Acid Methylation. RNA Biol. 14, 1108–1123. https://doi.org/10.1080/15476286.2016.1191737

Jeltsch, A., Nellen, W., Lyko, F., 2006. Two substrates are better than one: dual specificities for Dnmt2 methyltransferases. Trends Biochem. Sci. 31, 306–308. https://doi.org/10.1016/j.tibs.2006.04.005

Jones, G., Prosser, D.E., Kaufmann, M., 2018. Chapter 5 - The Activating Enzymes of Vitamin D Metabolism (25- and 1α-Hydroxylases), in: Feldman, D. (Ed.), Vitamin D (Fourth Edition). Academic Press, pp. 57–79. https://doi.org/10.1016/B978-0-12-809965-0.00005-7

Jones, P.A., Baylin, S.B., 2002. The fundamental role of epigenetic events in cancer. Nat. Rev. Genet. 3, 415–428. https://doi.org/10.1038/nrg816

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., Sasaki, H., 2004. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature 429, 900–903. https://doi.org/10.1038/nature02633

Karolchik, D., Barber, G.P., Casper, J., Clawson, H., Cline, M.S., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M., Harte, R.A., Heitner, S., Hinrichs, A.S., Learned, K., Lee, B.T., Li, C.H., Raney, B.J., Rhead, B., Rosenbloom, K.R., Sloan, C.A., Speir, M.L., Zweig, A.S., Haussler, D., Kuhn, R.M., Kent, W.J., 2014. The UCSC Genome Browser database: 2014 update. Nucleic Acids Res. 42, D764–D770. https://doi.org/10.1093/nar/gkt1168

Karpurapu, M., Ranjan, R., Deng, J., Chung, S., Lee, Y.G., Xiao, L., Nirujogi, T.S., Jacobson, J.R., Park, G.Y., Christman, J.W., 2014. Krüppel Like Factor 4 Promoter Undergoes Active Demethylation during Monocyte/Macrophage Differentiation. PLOS ONE 9, e93362. https://doi.org/10.1371/journal.pone.0093362

Kauffman, G., Kneivel, D., Watschke, T.L., 2006. Effects of a Biostimulant on the Heat Tolerance Associated with Photosynthetic Capacity, Membrane Thermostability, and Polyphenol Production of Perennial Ryegrass | Digital Library. Crop Sci. 47, 261–267.

Kechris, K.J., Biehs, B., Kornberg, T.B., 2010. Generalizing moving averages for tiling arrays using combined p-value statistics. Stat. Appl. Genet. Mol. Biol. 9, Article29. https://doi.org/10.2202/1544-6115.1434

Kelly, T.K., Miranda, T.B., Liang, G., Berman, B.P., Lin, J.C., Tanay, A., Jones, P.A., 2010. H2A.Z Maintenance During Mitosis Reveals Nucleosome Shifting on Mitotically Silenced Genes. Mol. Cell 39, 901–911. https://doi.org/10.1016/j.molcel.2010.08.026

Kim, G.-D., Ni, J., Kelesoglu, N., Roberts, R.J., Pradhan, S., 2002. Co-operation and communication between the human maintenance and de novo DNA (cytosine-5) methyltransferases. EMBO J. 21, 4183–4195. https://doi.org/10.1093/emboj/cdf401

Kint, S., Spiegelaere, W.D., Kesel, J.D., Vandekerckhove, L., Criekinge, W.V., 2018. Evaluation of bisulfite kits for DNA methylation profiling in terms of DNA fragmentation and DNA recovery using digital PCR. PLOS ONE 13, e0199091. https://doi.org/10.1371/journal.pone.0199091

Klug, M., Schmidhofer, S., Gebhard, C., Andreesen, R., Rehli, M., 2013. 5-Hydroxymethylcytosine is an essential intermediate of active DNA demethylation processes in primary human monocytes. Genome Biol. 14, R46. https://doi.org/10.1186/gb-2013-14-5-r46

Klump, H.H., Löffler, R., 2009. Reversible Helix/Coil Transitions of Left-Handed Z-DNA Structures. Comparison of the Thermodynamic Properties of poly(dG) · poly(dC), poly[d(G-C)] · poly[d(G-C)], and poly(dG-m5dC) · poly(dG-m5dC). Biol. Chem. 366, 345–354. https://doi.org/10.1515/bchm3.1985.366.1.345

Klungland, A., Robertson, A.B., 2017. Oxidized C5-methyl cytosine bases in DNA: 5-Hydroxymethylcytosine; 5-formylcytosine; and 5-carboxycytosine. Free Radic. Biol. Med., Oxidative DNA Damage & Repair 107, 62–68. https://doi.org/10.1016/j.freeradbiomed.2016.11.038

Krokan, H.E., Drabløs, F., Slupphaug, G., 2002. Uracil in DNA--occurrence, consequences and repair. Oncogene 21, 8935–8948. https://doi.org/10.1038/sj.onc.1205996

Kruger, I.M., Kruger, M.C., Doak, C.M., Schutte, A.E., Huisman, H.W., Van Rooyen, J.M., Schutte, R., Malan, L., Malan, N.T., Fourie, C.M.T., Kruger, A., 2013. The Association of 25(OH)D with Blood Pressure, Pulse Pressure and Carotid-Radial Pulse Wave Velocity in African Women. PLoS ONE 8. https://doi.org/10.1371/journal.pone.0054554

Kulis, M., Heath, S., Bibikova, M., Queirós, A.C., Navarro, A., Clot, G., Martínez-Trillos, A., Castellano, G., Brun-Heath, I., Pinyol, M., Barberán-Soler, S., Papasaikas, P., Jares, P., Beà, S., Rico, D., Ecker, S., Rubio, M., Royo, R., Ho, V., Klotzle, B., Hernández, L., Conde, L., López-Guerra, M., Colomer, D., Villamor, N., Aymerich, M., Rozman, M., Bayes, M., Gut, M., Gelpí, J.L., Orozco, M., Fan, J.-B., Quesada, V., Puente, X.S., Pisano, D.G., Valencia, A.,

López-Guillermo, A., Gut, I., López-Otín, C., Campo, E., Martín-Subero, J.I., 2012. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. Nat. Genet. 44, 1236–1242. https://doi.org/10.1038/ng.2443

Lai, G.-R., Lee, Y.-F., Yan, S.-J., Ting, H.-J., 2020. Active vitamin D induces gene-specific hypomethylation in prostate cancer cells developing vitamin D resistance. Am. J. Physiol. Cell Physiol. 318, C836–C847. https://doi.org/10.1152/ajpcell.00522.2019

Lander, E.S., Lauren, L.M., Birren, B., Nusbaum, C., Zody, M.C., 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921. https://doi.org/10.1038/35057062

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C.T., Low, H.M., Kin Sung, K.W., Rigoutsos, I., Loring, J., Wei, C.-L., 2010. Dynamic changes in the human methylome during differentiation. Genome Res. 20, 320–331. https://doi.org/10.1101/gr.101907.109

Lee, H.J., Hore, T.A., Reik, W., 2014. Reprogramming the methylome: erasing memory and creating diversity. Cell Stem Cell 14, 710–719. https://doi.org/10.1016/j.stem.2014.05.008

Li, C., Fan, Y., Li, G., Xu, X., Duan, J., Li, R., Kang, X., Ma, X., Chen, X., Ke, Y., Yan, J., Lian, Y., Liu, P., Zhao, Y., Zhao, H., Chen, Y., Yu, Y., Liu, J., 2018. DNA methylation reprogramming of functional elements during mammalian embryonic development. Cell Discov. 4, 1–12. https://doi.org/10.1038/s41421-018-0039-9

Li, P., Marshall, L., Oh, G., Jakubowski, J.L., Groot, D., He, Y., Wang, T., Petronis, A., Labrie, V., 2019. Epigenetic dysregulation of enhancers in neurons is associated with Alzheimer's disease pathology and cognitive symptoms. Nat. Commun. 10, 2246. https://doi.org/10.1038/s41467-019-10101-7

Li, S., Garrett-Bakelman, F.E., Akalin, A., Zumbo, P., Levine, R., To, B.L., Lewis, I.D., Brown, A.L., D'Andrea, R.J., Melnick, A., Mason, C.E., 2013. An optimized algorithm for detecting and annotating regional differential methylation. BMC Bioinformatics 14 Suppl 5, S10. https://doi.org/10.1186/1471-2105-14-S5-S10

Li, Z., Jia, Z., Gao, Y., Xie, D., Wei, D., Cui, J., Mishra, L., Huang, S., Zhang, Y., Xie, K., 2015. Activation of Vitamin D Receptor Signaling Downregulates the Expression of Nuclear FOXM1 Protein and Suppresses Pancreatic Cancer Cell Stemesss. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 21, 844–853. https://doi.org/10.1158/1078-0432.CCR-14-2437

Lim, S., Kim, M.J., Choi, S.H., Shin, C.S., Park, K.S., Jang, H.C., Billings, L.K., Meigs, J.B., 2013. Association of vitamin D deficiency with incidence of type 2 diabetes in high-risk Asian subjects. Am. J. Clin. Nutr. 97, 524–530. https://doi.org/10.3945/ajcn.112.048496

Lin, P.-I., Shu, H., Mersha, T.B., 2020. Comparing DNA methylation profiles across different tissues associated with the diagnosis of pediatric asthma. Sci. Rep. 10, 151. https://doi.org/10.1038/s41598-019-56310-4

Lin, R., White, J.H., 2004. The pleiotropic actions of vitamin D. BioEssays News Rev. Mol. Cell. Dev. Biol. 26, 21–28. https://doi.org/10.1002/bies.10368

Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D., Darnell, J., 2000. RNA Polymerase II Transcription-Initiation Complex. Mol. Cell Biol. 4th Ed.

Logan, J.M.J., Edwards, K.J., Saunders, N.A., 2009. Real-time PCR: Current Technology and Applications. Caister Academic Press.

Lokk, K., Modhukur, V., Rajashekar, B., Märtens, K., Mägi, R., Kolde, R., Koltšina, M., Nilsson, T.K., Vilo, J., Salumets, A., Tõnisson, N., 2014. DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. Genome Biol. 15, r54. https://doi.org/10.1186/gb-2014-15-4-r54

Ludmir, J., Mazhani, L., Cary, M.S., Chakalisa, U.A., Pettifor, J.M., Molefi, M., Redwood, A., Stallings, V.A., Gross, R., Steenhoff, A.P., 2016. Vitamin D Status in Botswana Children under 2 Years Old with and without Active Tuberculosis. Am. J. Trop. Med. Hyg. 94, 971–974. https://doi.org/10.4269/ajtmh.15-0864

Mahmoud, A.A., Ali, A., 2014. Vitamin D receptor gene polymorphism and 25 hydroxy vitamin D levels in Egyptian patients with pulmonary tuberculosis. Egypt. J. Chest Dis. Tuberc. 63. https://doi.org/10.1016/j.ejcdt.2014.02.015

Mailhot, G., 2015. Chapter 11 - Vitamin D Bioavailability in Cystic Fibrosis, in: Watson, R.R. (Ed.), Diet and Exercise in Cystic Fibrosis. Academic Press, Boston, pp. 75–95. https://doi.org/10.1016/B978-0-12-800051-9.00011-0

Mangelsdorf, D.J., Thummel, C., Beato, M., Herrlich, P., Schütz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P., Evans, R.M., 1995. The nuclear receptor superfamily: the second decade. Cell 83, 835–839.

Martin, N.C., Pirie, A.A., Ford, L.V., Callaghan, C.L., McTurk, K., Lucy, D., Scrimger, D.G., 2006. The use of phosphate buffered saline for the recovery of cells and spermatozoa from swabs. Sci. Justice J. Forensic Sci. Soc. 46, 179–184. https://doi.org/10.1016/S1355-0306(06)71591-X

Matarazzo, M.R., De Bonis, M.L., Strazzullo, M., Cerase, A., Ferraro, M., Vastarelli, P., Ballestar, E., Esteller, M., Kudo, S., D'Esposito, M., 2007. Multiple binding of methyl-CpG and polycomb proteins in long-term gene silencing events. J. Cell. Physiol. 210, 711–719. https://doi.org/10.1002/jcp.20879

Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., D'Souza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y., Turecki, G., Delaney, A., Varhol, R., Thiessen, N., Shchors, K., Heine, V.M., Rowitch, D.H., Xing, X., Fiore, C., Schillebeeckx, M., Jones, S.J.M., Haussler, D., Marra, M.A., Hirst, M., Wang, T., Costello, J.F., 2010. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature 466, 253–257. https://doi.org/10.1038/nature09165

Mayer, W., Niveleau, A., Walter, J., Fundele, R., Haaf, T., 2000. Demethylation of the zygotic paternal genome. Nature 403, 501–502. https://doi.org/10.1038/35000656

Mayjonade, B., Gouzy, J., Donnadieu, C., Pouilly, N., Marande, W., Callot, C., Langlade, N., Muños, S., 2016. Extraction of high-molecular-weight genomic

DNA for long-read sequencing of single molecules. BioTechniques 61, 203–205. https://doi.org/10.2144/000114460

McCollum, E.V., Simmonds, N., Becker, J.E., Shipley, P.G., 1922. Studies on Experimental Rickets Xxi. an Experimental Demonstration of the Existence of a Vitamin Which Promotes Calcium Deposition. J. Biol. Chem. 53, 293–312.

Meddeb, N., Sahli, H., Chahed, M., Abdelmoula, J., Feki, M., Salah, H., Frini, S., Kaabachi, N., Belkahia, C., Mbazaa, R., Zouari, B., Sellami, S., 2005. Vitamin D deficiency in Tunisia. Osteoporos. Int. J. Establ. Result Coop. Eur. Found. Osteoporos. Natl. Osteoporos. Found. USA 16, 180–183. https://doi.org/10.1007/s00198-004-1658-6

Mendes, K., Schmidhofer, S., Minderjahn, J., Glatz, D., Kiesewetter, C., Raithel, J., Wimmer, J., Gebhard, C., Rehli, M., 2021. The epigenetic pioneer EGR2 initiates DNA demethylation in differentiating monocytes at both stable and transient binding sites. Nat. Commun. 12, 1556. https://doi.org/10.1038/s41467-021-21661-y

Meyer, M.B., Goetsch, P.D., Pike, J.W., 2010. A Downstream Intergenic Cluster of Regulatory Enhancers Contributes to the Induction of CYP24A1 Expression by 1α,25-Dihydroxyvitamin D3*. J. Biol. Chem. 285, 15599–15610. https://doi.org/10.1074/jbc.M110.119958

Meyer, V., Bornman, L., 2018. Cdx-2 polymorphism in the vitamin D receptor gene (VDR) marks VDR expression in monocyte/macrophages through VDR promoter methylation. Immunogenetics 70, 523–532. https://doi.org/10.1007/s00251-018-1063-5

Moore, L.D., Le, T., Fan, G., 2013. DNA Methylation and Its Basic Function. Neuropsychopharmacology 38, 23–38. https://doi.org/10.1038/npp.2012.112

Morgan, H.D., Santos, F., Green, K., Dean, W., Reik, W., 2005. Epigenetic reprogramming in mammals. Hum. Mol. Genet. 14 Spec No 1, R47-58. https://doi.org/10.1093/hmg/ddi114

Naidoo, Y., Moodley, J., Madurai, L., Naicker, T., 2018. Prevalence of Vitamin D deficiency in a multiracial female population in KwaZulu-Natal province, South Africa. South Afr. Fam. Pract. 1–6. https://doi.org/10.1080/20786190.2018.1541579

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., Bird, A., 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature 393, 386–389. https://doi.org/10.1038/30764

Newton, D.A., Baatz, J.E., Kindy, M.S., Gattoni-Celli, S., Shary, J.R., Hollis, B.W., Wagner, C.L., 2019. Vitamin D binding protein polymorphisms significantly impact vitamin D status in children. Pediatr. Res. 1–8. https://doi.org/10.1038/s41390-019-0322-y

Okano, M., Bell, D.W., Haber, D.A., Li, E., 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99, 247–257. https://doi.org/10.1016/s0092-8674(00)81656-6

Ooi, S.K.T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C.D., Cheng, X., Bestor, T.H., 2007. DNMT3L

connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448, 714–717. https://doi.org/10.1038/nature05987

Papadimitriou, D.T., 2017. The Big Vitamin D Mistake. J. Prev. Med. Pub. Health 50, 278–281. https://doi.org/10.3961/jpmph.16.111

Parva, N.R., Tadepalli, S., Singh, P., Qian, A., Joshi, R., Kandala, H., Nookala, V.K., Cheriyath, P., 2018. Prevalence of Vitamin D Deficiency and Associated Risk Factors in the US Population (2011-2012). Cureus 10. https://doi.org/10.7759/cureus.2741

Peltzer, K., Phaswana-Mafuya, N., 2013. Hypertension and associated factors in older adults in South Africa. Cardiovasc. J. Afr. 24, 66–72. https://doi.org/10.5830/CVJA-2013-002

Petrascheck, M., Escher, D., Mahmoudi, T., Verrijzer, C.P., Schaffner, W., Barberis, A., 2005. DNA looping induced by a transcriptional enhancer in vivo. Nucleic Acids Res. 33, 3743–3750. https://doi.org/10.1093/nar/gki689

Pham, T.-H., Benner, C., Lichtinger, M., Schwarzfischer, L., Hu, Y., Andreesen, R., Chen, W., Rehli, M., 2012. Dynamic epigenetic enhancer signatures reveal key transcription factors associated with monocytic differentiation states. Blood 119, e161–e171. https://doi.org/10.1182/blood-2012-01-402453

Pheiffer, C., Wyk, V.P., Joubert, J.D., Levitt, N., Nglazi, M.D., Bradshaw, D., 2018. The prevalence of type 2 diabetes in South Africa: a systematic review protocol. BMJ Open 8, e021029. https://doi.org/10.1136/bmjopen-2017-021029

Piemonti, L., Monti, P., Sironi, M., Fraticelli, P., Leone, B.E., Dal Cin, E., Allavena, P., Di Carlo, V., 2000. Vitamin D3 affects differentiation, maturation, and function of human monocyte-derived dendritic cells. J. Immunol. Baltim. Md 1950 164, 4443–4451. https://doi.org/10.4049/jimmunol.164.9.4443

Pike, J.W., Meyer, M.B., 2010. The Vitamin D Receptor: New Paradigms for the Regulation of Gene Expression by 1,25-Dihydroxyvitamin D3. Endocrinol. Metab. Clin. North Am. 39, 255–269. https://doi.org/10.1016/j.ecl.2010.02.007

Pruitt, K.D., Brown, G.R., Hiatt, S.M., Thibaud-Nissen, F., Astashyn, A., Ermolaeva, O., Farrell, C.M., Hart, J., Landrum, M.J., McGarvey, K.M., Murphy, M.R., O'Leary, N.A., Pujar, S., Rajput, B., Rangwala, S.H., Riddick, L.D., Shkeda, A., Sun, H., Tamez, P., Tully, R.E., Wallin, C., Webb, D., Weber, J., Wu, W., DiCuccio, M., Kitts, P., Maglott, D.R., Murphy, T.D., Ostell, J.M., 2014. RefSeq: an update on mammalian reference sequences. Nucleic Acids Res. 42, D756–D763. https://doi.org/10.1093/nar/gkt1114

Raines, R.T., 1998. Ribonuclease A. Chem. Rev. 98, 1045–1066. https://doi.org/10.1021/cr960427h

Ramasamy, I., 2008. Inherited disorders of calcium homeostasis. Clin. Chim. Acta Int. J. Clin. Chem. 394, 22–41. https://doi.org/10.1016/j.cca.2008.04.011

Razin, S., 1996. Mycoplasmas, in: Baron, S. (Ed.), Medical Microbiology. University of Texas Medical Branch at Galveston, Galveston (TX).

Reik, W., Dean, W., Walter, J., 2001. Epigenetic Reprogramming in Mammalian Development. Science 293, 1089–1093. https://doi.org/10.1126/science.1063443

Rice, J.C., Allis, C.D., 2001. Histone methylation versus histone acetylation: new insights into epigenetic regulation. Curr. Opin. Cell Biol. 13, 263–273. https://doi.org/10.1016/S0955-0674(00)00208-8

Rochel, N., Wurtz, J.M., Mitschler, A., Klaholz, B., Moras, D., 2000. The crystal structure of the nuclear receptor for vitamin D bound to its natural ligand. Mol. Cell 5, 173–179.

Rosa, A., Ballarino, M., Sorrentino, A., Sthandier, O., De Angelis, F.G., Marchioni, M., Masella, B., Guarini, A., Fatica, A., Peschle, C., Bozzoni, I., 2007. The interplay between the master transcription factor PU.1 and miR-424 regulates human monocyte/macrophage differentiation. Proc. Natl. Acad. Sci. U. S. A. 104, 19849–19854. https://doi.org/10.1073/pnas.0706963104

Saccone, D., Asani, F., Bornman, L., 2015. Regulation of the vitamin D receptor gene by environment, genetics and epigenetics. Gene 561, 171–180. https://doi.org/10.1016/j.gene.2015.02.024

Saccone, S., 2008. Gene Distribution in Human Chromosomes. https://doi.org/10.1002/9780470015902.a0005790.pub2

Sakaki, T., Sawada, N., Komai, K., Shiozawa, S., Yamada, S., Yamamoto, K., Ohyama, Y., Inouye, K., 2000. Dual metabolic pathway of 25-hydroxyvitamin D3 catalyzed by human CYP24. Eur. J. Biochem. 267, 6158–6165. https://doi.org/10.1046/j.1432-1327.2000.01680.x

Sambrook, J., Russell, D.W., 2001. Molecular Cloning: A Laboratory Manual. CSHL Press.

Sandoval, J., Heyn, H., Moran, S., Serra-Musach, J., Pujana, M.A., Bibikova, M., Esteller, M., 2011. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics 6, 692–702.

Sangiovanni, M., Granata, I., Thind, A.S., Guarracino, M.R., 2019. From trash to treasure: detecting unexpected contamination in unmapped NGS data. BMC Bioinformatics 20, 168. https://doi.org/10.1186/s12859-019-2684-x

Sarda, S., Das, A., Vinson, C., Hannenhalli, S., 2017. Distal CpG islands can serve as alternative promoters to transcribe genes with silenced proximal promoters. Genome Res. 27, 553–566. https://doi.org/10.1101/gr.212050.116

Saxonov, S., Berg, P., Brutlag, D.L., 2006. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc. Natl. Acad. Sci. U. S. A. 103, 1412–1417. https://doi.org/10.1073/pnas.0510310103

Schwende, H., Fitzke, E., Ambs, P., Dieter, P., 1996. Differences in the state of differentiation of THP-1 cells induced by phorbol ester and 1,25-dihydroxyvitamin D3. J. Leukoc. Biol. 59, 555–561.

Selvaraj, P., Harishankar, M., Afsal, K., 2015. Vitamin D: Immuno-modulation and tuberculosis treatment. Can. J. Physiol. Pharmacol. 93, 377–384. https://doi.org/10.1139/cjpp-2014-0386

Seuter, S., Neme, A., Carlberg, C., 2018. ETS transcription factor family member GABPA contributes to vitamin D receptor target gene regulation. J. Steroid Biochem. Mol. Biol., Proceedings of the 20th Vitamin D Workshop 177, 46–52. https://doi.org/10.1016/j.jsbmb.2017.08.006

Shaffer, P.L., Gewirth, D.T., 2002. Structural basis of VDR-DNA interactions on direct repeat response elements. EMBO J. 21, 2242–2252. https://doi.org/10.1093/emboj/21.9.2242

Shafi, A., Mitrea, C., Nguyen, T., Draghici, S., 2018. A survey of the approaches for identifying differential methylation using bisulfite sequencing data. Brief. Bioinform. 19, 737–753. https://doi.org/10.1093/bib/bbx013

Sharma, G., Sowpati, D.T., Singh, P., Khan, M.Z., Ganji, R., Upadhyay, S., Banerjee, S., Nandicoori, V.K., Khosla, S., 2016. Genome-wide non-CpG methylation of the host genome during M. tuberculosis infection. Sci. Rep. 6, 25006. https://doi.org/10.1038/srep25006

Shen, L., Kondo, Y., Guo, Y., Zhang, J., Zhang, L., Ahmed, S., Shu, J., Chen, X., Waterland, R.A., Issa, J.-P.J., 2007. Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. PLoS Genet. 3, 2023–2036. https://doi.org/10.1371/journal.pgen.0030181

Shlyueva, D., Stampfel, G., Stark, A., 2014. Transcriptional enhancers: from properties to genome-wide predictions. Nat. Rev. Genet. 15, 272–286. https://doi.org/10.1038/nrg3682

Shull, A.Y., Luo, J., Pei, L., Lee, E.-J., Liu, J., Choi, J., Awan, F.T., Shi, H., 2016. DNA Hypomethylation within B-Cell Enhancers and Super Enhancers Reveal a Dependency on Immune and Metabolic Mechanisms in Chronic Lymphocytic Leukemia. Blood 128, 1049–1049. https://doi.org/10.1182/blood.V128.22.1049.1049

Simard, A.R., Soulet, D., Gowing, G., Julien, J.-P., Rivest, S., 2006. Bone marrow-derived microglia play a critical role in restricting senile plaque formation in Alzheimer's disease. Neuron 49, 489–502. https://doi.org/10.1016/j.neuron.2006.01.022

Slieker, R.C., Relton, C.L., Gaunt, T.R., Slagboom, P.E., Heijmans, B.T., 2018. Age-related DNA methylation changes are tissue-specific with ELOVL2 promoter methylation as exception. Epigenetics Chromatin 11, 25. https://doi.org/10.1186/s13072-018-0191-3

Slominski, A., Postlethwaite, A.E., 2015. Skin Under the Sun: When Melanin Pigment Meets Vitamin D. Endocrinology 156, 1–4. https://doi.org/10.1210/en.2014-1918

Smith, C.L., O'Malley, B.W., 2004. Coregulator function: a key to understanding tissue specificity of selective receptor modulators. Endocr. Rev. 25, 45–71. https://doi.org/10.1210/er.2003-0023

Smith, J., Sen, S., Weeks, R.J., Eccles, M.R., Chatterjee, A., 2020. Promoter DNA Hypermethylation and Paradoxical Gene Activation. Trends Cancer 6, 392–406. https://doi.org/10.1016/j.trecan.2020.02.007

Smith, Z.D., Meissner, A., 2013. The simplest explanation: passive DNA demethylation in PGCs. EMBO J. 32, 318–321. https://doi.org/10.1038/emboj.2012.349

Sone, T., Kerner, S., Pike, J.W., 1991. Vitamin D receptor interaction with specific DNA. Association as a 1,25-dihyroxyvitamin D3-modulated heterodimer. J. Biol. Chem. 266, 23296–23305.

Sproul, D., Meehan, R.R., 2013. Genomic insights into cancer-associated aberrant CpG island hypermethylation. Brief. Funct. Genomics 12, 174–190. https://doi.org/10.1093/bfgp/els063

Steenbock, H., 1924. The Induction Of Growth Promoting And Calcifying Properties In A Ration By Exposure To Light. Science 60, 224–225. https://doi.org/10.1126/science.60.1549.224

Stellwagen, N.C., Gelfi, C., Righetti, P.G., 2000. DNA and buffers: the hidden danger of complex formation. Biopolymers 54, 137–142. https://doi.org/10.1002/1097-0282(200008)54:2<137::AID-BIP6>3.0.CO;2-I

Stevens, M., Cheng, J.B., Li, D., Xie, M., Hong, C., Maire, C.L., Ligon, K.L., Hirst, M., Marra, M.A., Costello, J.F., Wang, T., 2013. Estimating absolute methylation levels at single-CpG resolution from methylation enrichment and restriction enzyme sequencing methods. Genome Res. 23, 1541–1553. https://doi.org/10.1101/gr.152231.112

Strober, W., 2001. Trypan blue exclusion test of cell viability. Curr. Protoc. Immunol. Appendix 3, Appendix 3B. https://doi.org/10.1002/0471142735.ima03bs21

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P., 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. 102, 15545–15550. https://doi.org/10.1073/pnas.0506580102

Takeyama, K., Kato, S., 2011. The Vitamin D3 1alpha-Hydroxylase Gene and Its Regulation by Active Vitamin D3. Biosci. Biotechnol. Biochem. 75, 208–213. https://doi.org/10.1271/bbb.100684

Tazi, J., Bird, A., 1990. Alternative chromatin structure at CpG islands. Cell 60, 909–920. https://doi.org/10.1016/0092-8674(90)90339-G

Thévenin, A., Ein-Dor, L., Ozery-Flato, M., Shamir, R., 2014. Functional gene groups are concentrated within chromosomes, among chromosomes and in the nuclear space of the human genome. Nucleic Acids Res. 42, 9854–9861. https://doi.org/10.1093/nar/gku667

Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R.W., Deaton, A., Andrews, R., James, K.D., Turner, D.J., Illingworth, R., Bird, A., 2010. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature 464, 1082–1086. https://doi.org/10.1038/nature08924

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., Garg, K., John, S., Sandstrom, R., Bates, D., Boatman, L., Canfield, T.K., Diegel, M., Dunn, D., Ebersol, A.K., Frum, T., Giste, E., Johnson, A.K., Johnson, E.M., Kutyavin, T., Lajoie, B., Lee, B.-K., Lee, K., London, D., Lotakis, D., Neph, S., Neri, F., Nguyen, E.D., Qu, H., Reynolds, A.P., Roach, V., Safi, A., Sanchez, M.E., Sanyal, A., Shafer, A., Simon, J.M., Song, L., Vong, S., Weaver, M., Yan, Y., Zhang, Zhancheng, Zhang, Zhuzhu, Lenhard, B., Tewari, M., Dorschner, M.O., Hansen, R.S., Navas, P.A., Stamatoyannopoulos, G., Iyer, V.R., Lieb, J.D., Sunyaev, S.R., Akey, J.M., Sabo, P.J., Kaul, R., Furey, T.S., Dekker, J., Crawford, G.E., Stamatoyannopoulos, J.A., 2012. The accessible chromatin

landscape of the human genome. Nature 489, 75–82. https://doi.org/10.1038/nature11232

Tirado-Magallanes, R., Rebbani, K., Lim, R., Pradhan, S., Benoukraf, T., 2016. Whole genome DNA methylation: beyond genes silencing. Oncotarget 8, 5629–5637. https://doi.org/10.18632/oncotarget.13562

Tost, J., Gut, I.G., 2007. DNA methylation analysis by pyrosequencing. Nat. Protoc. 2, 2265–2275. https://doi.org/10.1038/nprot.2007.314

Tóth, E., Erdődi, F., Kiss, A., 2021. Myosin Phosphatase Is Implicated in the Control of THP-1 Monocyte to Macrophage Differentiation. Int. J. Mol. Sci. 22. https://doi.org/10.3390/ijms22052516

Tsuchiya, S., Yamabe, M., Yamaguchi, Y., Kobayashi, Y., Konno, T., Tada, K., 1980. Establishment and characterization of a human acute monocytic leukemia cell line (THP-1). Int. J. Cancer 26, 171–176. https://doi.org/10.1002/ijc.2910260208

Ueda, Y., Okano, M., Williams, C., Chen, T., Georgopoulos, K., Li, E., 2006. Roles for Dnmt3b in mammalian development: a mouse model for the ICF syndrome. Dev. Camb. Engl. 133, 1183–1192. https://doi.org/10.1242/dev.02293

Ueki, T., Walter, K.M., Skinner, H., Jaffee, E., Hruban, R.H., Goggins, M., 2002. Aberrant CpG island methylation in cancer cell lines arises in the primary cancers from which they were derived. Oncogene 21, 2114–2117. https://doi.org/10.1038/sj.onc.1205275

Varley, K.E., Gertz, J., Bowling, K.M., Parker, S.L., Reddy, T.E., Pauli-Behn, F., Cross, M.K., Williams, B.A., Stamatoyannopoulos, J.A., Crawford, G.E., Absher, D.M., Wold, B.J., Myers, R.M., 2013. Dynamic DNA methylation across diverse human cell lines and tissues. Genome Res. 23, 555–567. https://doi.org/10.1101/gr.147942.112

Venturini, E., Facchini, L., Martinez-Alier, N., Novelli, V., Galli, L., de Martino, M., Chiappini, E., 2014. Vitamin D and tuberculosis: a multicenter study in children. BMC Infect. Dis. 14, 652. https://doi.org/10.1186/s12879-014-0652-7

Vimaleswaran, K.S., Berry, D.J., Lu, C., Tikkanen, E., Pilz, S., Hiraki, L.T., Cooper, J.D., Dastani, Z., Li, R., Houston, D.K., Wood, A.R., Michaëlsson, K., Vandenput, L., Zgaga, L., Yerges-Armstrong, L.M., McCarthy, M.I., Dupuis, J., Kaakinen, M., Kleber, M.E., Jameson, K., Arden, N., Raitakari, O., Viikari, J., Lohman, K.K., Ferrucci, L., Melhus, H., Ingelsson, E., Byberg, L., Lind, L., Lorentzon, M., Salomaa, V., Campbell, H., Dunlop, M., Mitchell, B.D., Herzig, K.-H., Pouta, A., Hartikainen, A.-L., Genetic Investigation of Anthropometric Traits-GIANT Consortium, Streeten, E.A., Theodoratou, E., Jula, A., Wareham, N.J., Ohlsson, C., Frayling, T.M., Kritchevsky, S.B., Spector, T.D., Richards, J.B., Lehtimäki, T., Ouwehand, W.H., Kraft, P., Cooper, C., März, W., Power, C., Loos, R.J.F., Wang, T.J., Järvelin, M.-R., Whittaker, J.C., Hingorani, A.D., Hyppönen, E., 2013. Causal relationship between obesity and vitamin D status: bi-directional Mendelian randomization analysis of multiple cohorts. PLoS Med. 10, e1001383. https://doi.org/10.1371/journal.pmed.1001383

Visone, R., Bacalini, M.G., Di Franco, S., Ferracin, M., Colorito, M.L., Pagotto, S., Laprovitera, N., Licastro, D., Di Marco, M., Scavo, E., Bassi, C., Saccenti, E., Nicotra, A., Grzes, M., Garagnani, P., De Laurenzi, V., Valeri, N., Mariani-Costantini, R., Negrini, M., Stassi, G., Veronese, A., 2019. DNA methylation of shelf, shore and open sea CpG positions distinguish high microsatellite instability from low or stable microsatellite status colon cancer stem cells. Epigenomics 11, 587–604. https://doi.org/10.2217/epi-2018-0153

Voo, K., Carlone, D., Jacobsen, B., Flodin, A., Skalnik, D., 2000. Cloning of a Mammalian Transcriptional Activator That Binds Unmethylated CpG Motifs and Shares a CXXC Domain with DNA Methyltransferase, Human Trithorax, and Methyl-CpG Binding Domain Protein 1. Mol. Cell. Biol. 20, 2108–21. https://doi.org/10.1128/MCB.20.6.2108-2121.2000

Wagner, C.L., Hollis, B.W., 2018. The Implications of Vitamin D Status During Pregnancy on Mother and her Developing Child. Front. Endocrinol. 9. https://doi.org/10.3389/fendo.2018.00500

Wallner, S., Schröder, C., Leitão, E., Berulava, T., Haak, C., Beißer, D., Rahmann, S., Richter, A.S., Manke, T., Bönisch, U., Arrigoni, L., Fröhler, S., Klironomos, F., Chen, W., Rajewsky, N., Müller, F., Ebert, P., Lengauer, T., Barann, M., Rosenstiel, P., Gasparoni, G., Nordström, K., Walter, J., Brors, B., Zipprich, G., Felder, B., Klein-Hitpass, L., Attenberger, C., Schmitz, G., Horsthemke, B., 2016. Epigenetic dynamics of monocyte-to-macrophage differentiation. Epigenetics Chromatin 9, 33. https://doi.org/10.1186/s13072-016-0079-z

Walsh, C.P., Chaillet, J.R., Bestor, T.H., 1998. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat. Genet. 20, 116–117. https://doi.org/10.1038/2413

Wan, J., Oliver, V.F., Wang, G., Zhu, H., Zack, D.J., Merbs, S.L., Qian, J., 2015. Characterization of tissue-specific differential DNA methylation suggests distinct modes of positive and negative gene expression regulation. BMC Genomics 16, 49. https://doi.org/10.1186/s12864-015-1271-4

Wang, H.-Q., Tuominen, L.K., Tsai, C.-J., 2011. SLIM: a sliding linear model for estimating the proportion of true null hypotheses in datasets with dependence structures. Bioinforma. Oxf. Engl. 27, 225–231. https://doi.org/10.1093/bioinformatics/btq650

Wang, K., Liu, S., Svoboda, L.K., Rygiel, C.A., Neier, K., Jones, T.R., Colacino, J.A., Dolinoy, D.C., Sartor, M.A., 2020. Tissue- and Sex-Specific DNA Methylation Changes in Mice Perinatally Exposed to Lead (Pb). Front. Genet. 0. https://doi.org/10.3389/fgene.2020.00840

Wang, M., Kong, W., He, B., Li, Z., Song, H., Shi, P., Wang, J., 2018. Vitamin D and the promoter methylation of its metabolic pathway genes in association with the risk and prognosis of tuberculosis. Clin. Epigenetics 10, 118. https://doi.org/10.1186/s13148-018-0552-6

Warnecke, P.M., Stirzaker, C., Melki, J.R., Millar, D.S., Paul, C.L., Clark, S.J., 1997. Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. Nucleic Acids Res. 25, 4422–4426. https://doi.org/10.1093/nar/25.21.4422

Watt, F., Molloy, P.L., 1988. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. Genes Dev. 2, 1136–1143.

Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., Schübeler, D., 2005. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat. Genet. 37, 853–862. https://doi.org/10.1038/ng1598

Wejse, C., Olesen, R., Rabna, P., Kaestel, P., Gustafson, P., Aaby, P., Andersen, P.L., Glerup, H., Sodemann, M., 2007. Serum 25-hydroxyvitamin D in a West African population of tuberculosis patients and unmatched healthy controls. Am. J. Clin. Nutr. 86, 1376–1383. https://doi.org/10.1093/ajcn/86.5.1376

Wiles, E.T., Selker, E.U., 2017. H3K27 methylation: a promiscuous repressive chromatin mark. Curr. Opin. Genet. Dev., Genome architecture and expression 43, 31–37. https://doi.org/10.1016/j.gde.2016.11.001

Wilson, G.A., Beck, S., 2016. Chapter 5 Computational Analysis and Integration of MeDIP-seq Methylome Data. Gener. Seq. - Adv. Appl. Chall. https://doi.org/10.5772/61207

Wu, X., Zhang, Y., 2017. TET-mediated active DNA demethylation: mechanism, function and beyond. Nat. Rev. Genet. 18, 517–534. https://doi.org/10.1038/nrg.2017.33

Xiang, F., Lucas, R., de Gruijl, F., Norval, M., 2015. A systematic review of the influence of skin pigmentation on changes in the concentrations of vitamin D and 25-hydroxyvitamin D in plasma/serum following experimental UV irradiation. Photochem. Photobiol. Sci. Off. J. Eur. Photochem. Assoc. Eur. Soc. Photobiol. 14, 2138–2146. https://doi.org/10.1039/c5pp00168d

Xue, J., Gharaibeh, R.Z., Pietryk, E.W., Brouwer, C., Tarantino, L.M., Valdar, W., Ideraabdullah, F.Y., 2018. Impact of vitamin D depletion during development on mouse sperm DNA methylation. Epigenetics 13, 959–974. https://doi.org/10.1080/15592294.2018.1526027

Yang, M., Park, J.Y., 2012. DNA Methylation in Promoter Region as Biomarkers in Prostate Cancer. Methods Mol. Biol. Clifton NJ 863, 67–109. https://doi.org/10.1007/978-1-61779-612-8_5

Yang, X., Han, H., De Carvalho, D.D., Lay, F.D., Jones, P.A., Liang, G., 2014. Gene Body Methylation Can Alter Gene Expression and Is a Therapeutic Target in Cancer. Cancer Cell 26, 577–590. https://doi.org/10.1016/j.ccr.2014.07.028

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., Nitta, K.R., Taipale, M., Popov, A., Ginno, P.A., Domcke, S., Yan, J., Schübeler, D., Vinson, C., Taipale, J., 2017. Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science 356, eaaj2239. https://doi.org/10.1126/science.aaj2239

Yong, W.-S., Hsu, F.-M., Chen, P.-Y., 2016. Profiling genome-wide DNA methylation. Epigenetics Chromatin 9, 26. https://doi.org/10.1186/s13072-016-0075-3

You, J.S., Kelly, T.K., De Carvalho, D.D., Taberlay, P.C., Liang, G., Jones, P.A., 2011. OCT4 establishes and maintains nucleosome-depleted regions that

provide additional layers of epigenetic regulation of its target genes. Proc. Natl. Acad. Sci. U. S. A. 108, 14497–14502. https://doi.org/10.1073/pnas.1111309108

Yu, G., Wang, L.-G., Han, Y., He, Q.-Y., 2012. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS J. Integr. Biol. 16, 284–287. https://doi.org/10.1089/omi.2011.0118

Zemach, A., McDaniel, I.E., Silva, P., Zilberman, D., 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328, 916–919. https://doi.org/10.1126/science.1186366

Zhang, Bo, Zhou, Y., Lin, N., Lowdon, R.F., Hong, C., Nagarajan, R.P., Cheng, J.B., Li, D., Stevens, M., Lee, H.J., Xing, X., Zhou, J., Sundaram, V., Elliott, G., Gu, J., Shi, T., Gascard, P., Sigaroudinia, M., Tlsty, T.D., Kadlecek, T., Weiss, A., O'Geen, H., Farnham, P.J., Maire, C.L., Ligon, K.L., Madden, P.A.F., Tam, A., Moore, R., Hirst, M., Marra, M.A., Zhang, Baoxue, Costello, J.F., Wang, T., 2013. Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm. Genome Res. 23, 1522–1540. https://doi.org/10.1101/gr.156539.113

Zhang, X., Tu, W., Manson, J.E., Tinker, L., Liu, S., Cauley, J.A., Qi, L., Mouton, C., Martin, L.W., Hou, L., Song, Y., 2019. Racial/Ethnic Differences in 25-Hydroxy Vitamin D and Parathyroid Hormone Levels and Cardiovascular Disease Risk Among Postmenopausal Women. J. Am. Heart Assoc.

Zhang, Y., Leung, D.Y.M., Richers, B.N., Liu, Y., Remigio, L.K., Riches, D.W., Goleva, E., 2012. Vitamin D Inhibits Monocyte/macrophage Pro-inflammatory Cytokine Production by Targeting Mitogen-Activated Protein Kinase Phosphatase 1. J. Immunol. Baltim. Md 1950 188, 2127–2135. https://doi.org/10.4049/jimmunol.1102412

Zhao, M., Tuo, H., Wang, S., Zhao, L., 2020. The Roles of Monocyte and Monocyte-Derived Macrophages in Common Brain Disorders. BioMed Res. Int. 2020, e9396021. https://doi.org/10.1155/2020/9396021

Zhou, J., Sears, R.L., Xing, X., Zhang, B., Li, D., Rockweiler, N.B., Jang, H.S., Choudhary, M.N.K., Lee, H.J., Lowdon, R.F., Arand, J., Tabers, B., Gu, C.C., Cicero, T.J., Wang, T., 2017. Tissue-specific DNA methylation is conserved across human, mouse, and rat, and driven by primary sequence conservation. BMC Genomics 18, 724. https://doi.org/10.1186/s12864-017-4115-6

Zhu, H., Bhagatwala, J., Huang, Y., Pollock, N.K., Parikh, S., Raed, A., Gutin, B., Harshfield, G.A., Dong, Y., 2016. Race/Ethnicity-Specific Association of Vitamin D and Global DNA Methylation: Cross-Sectional and Interventional Findings. PLOS ONE 11, e0152849. https://doi.org/10.1371/journal.pone.0152849

Zhu, H., Wang, X., Shi, H., Su, S., Harshfield, G.A., Gutin, B., Snieder, H., Dong, Y., 2013. A genome-wide methylation study of severe vitamin D deficiency in African American adolescents. J. Pediatr. 162, 1004-1009.e1. https://doi.org/10.1016/j.jpeds.2012.10.059

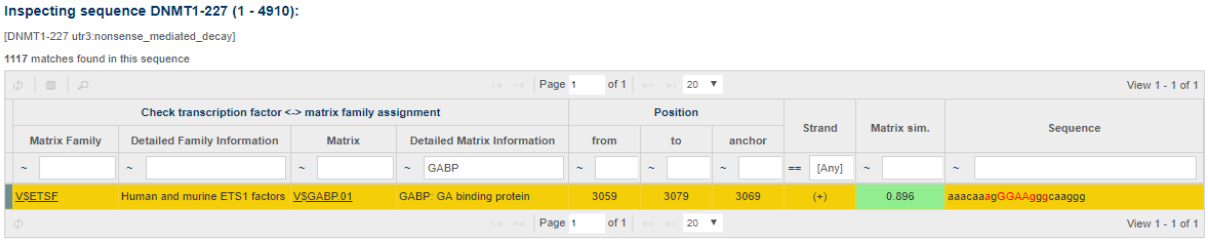
Zhu, J., He, F., Hu, S., Yu, J., 2008. On the nature of human housekeeping genes. Trends Genet. 24, 481–484. https://doi.org/10.1016/j.tig.2008.08.004

Zhu, L.J., Gazin, C., Lawson, N.D., Pagès, H., Lin, S.M., Lapointe, D.S., Green, M.R., 2010. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics 11, 237. https://doi.org/10.1186/1471-2105-11-237

Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T., Henikoff, S., 2007. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nat. Genet. 39, 61–69. https://doi.org/10.1038/ng1929

Appendices

Appendix A: GABPA Binding Motifs



Inspecting sequence DNMT1-227 (1 - 4910):

[DNMT1-227 utr3:nonsense_mediated_decay]

1117 matches found in this sequence

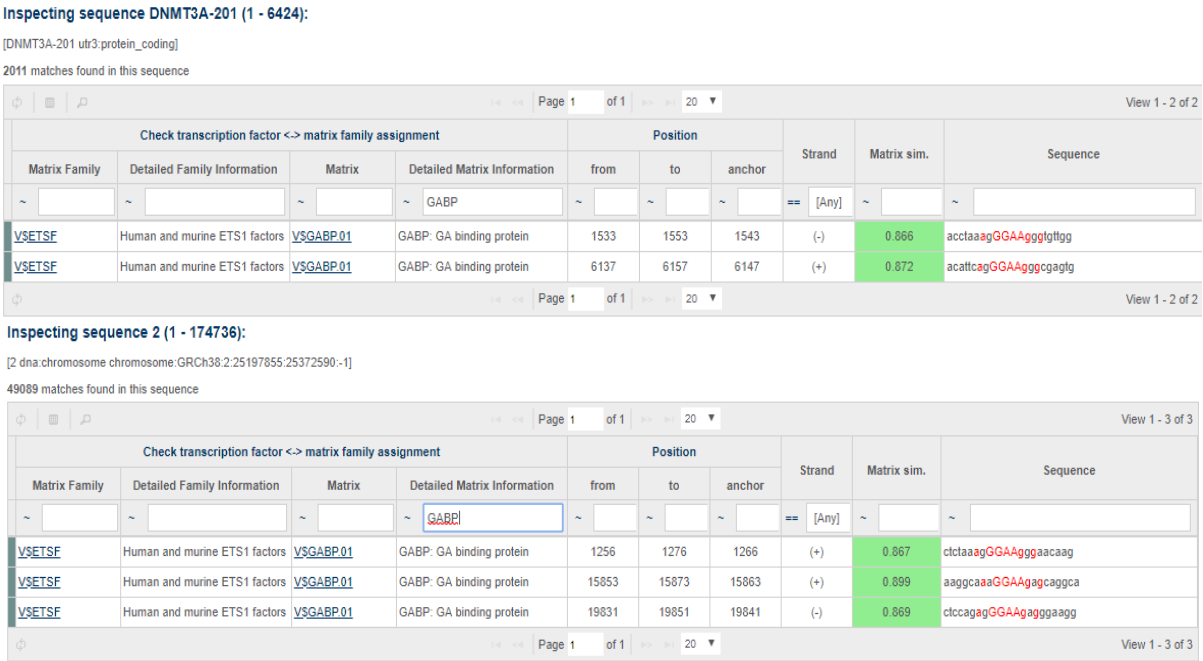
Page 1 of 1 — View 1 - 1 of 1

Check transcription factor <-> matrix family assignment				Position			Strand	Matrix sim.	Sequence
Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	from	to	anchor			
			GABP				[Any]		
V$ETSF	Human and murine ETS1 factors	V$GABP.01	GABP: GA binding protein	3059	3079	3069	(+)	0.896	aaacaaagGGAAgggcaaggg

Page 1 of 1 — View 1 - 1 of 1

Figure A. 1: A GABPA binding motif was identified in the 3' untranslated region of DNMT1. The matrix similarity was 89% with the core GABPA DNA binding site housed in the 20-nucleotide element.

Figure A. 2: Two GABPA binding motifs were found in the 3' UTR of DNMT3a. One was identified on the positive strand and one was identified on the



Inspecting sequence DNMT3A-201 (1 - 6424):

[DNMT3A-201 utr3:protein_coding]

2011 matches found in this sequence

Page 1 of 1 — View 1 - 2 of 2

Check transcription factor <-> matrix family assignment				Position			Strand	Matrix sim.	Sequence
Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	from	to	anchor			
			GABP				[Any]		
V$ETSF	Human and murine ETS1 factors	V$GABP.01	GABP: GA binding protein	1533	1553	1543	(-)	0.866	acctaaagGGAAgggtgttgg
V$ETSF	Human and murine ETS1 factors	V$GABP.01	GABP: GA binding protein	6137	6157	6147	(+)	0.872	acattcagGGAAgggcgagtg

Page 1 of 1 — View 1 - 2 of 2

Inspecting sequence 2 (1 - 174736):

[2 dna:chromosome chromosome:GRCh38:2:25197855:25372590:-1]

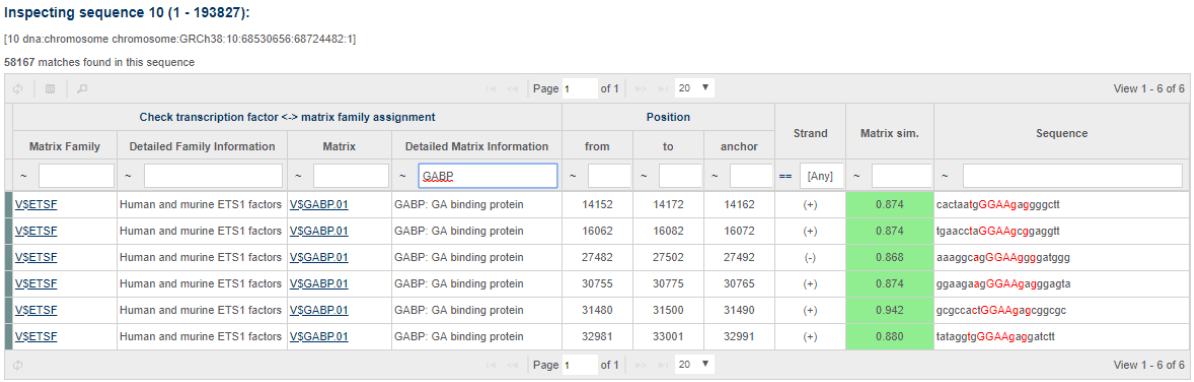
49089 matches found in this sequence

Page 1 of 1 — View 1 - 3 of 3

Check transcription factor <-> matrix family assignment				Position			Strand	Matrix sim.	Sequence
Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	from	to	anchor			
			GABP				[Any]		
V$ETSF	Human and murine ETS1 factors	V$GABP.01	GABP: GA binding protein	1256	1276	1266	(+)	0.867	ctctaaagGGAAgggaacaag
V$ETSF	Human and murine ETS1 factors	V$GABP.01	GABP: GA binding protein	15853	15873	15863	(+)	0.899	aaggcaaaGGAAgagcaggca
V$ETSF	Human and murine ETS1 factors	V$GABP.01	GABP: GA binding protein	19831	19851	19841	(-)	0.869	ctccagagGGAAgagggaagg

Page 1 of 1 — View 1 - 3 of 3

negative strand. The matrix similarity percentage was 87%. The core GABPA DNA binding site is located in the 20-nucleotide sequence. Additionally, three other GABPA binding sites were found in the 5’ flanking sequence.



Inspecting sequence 10 (1 - 193827):

[10 dna:chromosome chromosome:GRCh38:10:68530656:68724482:1]

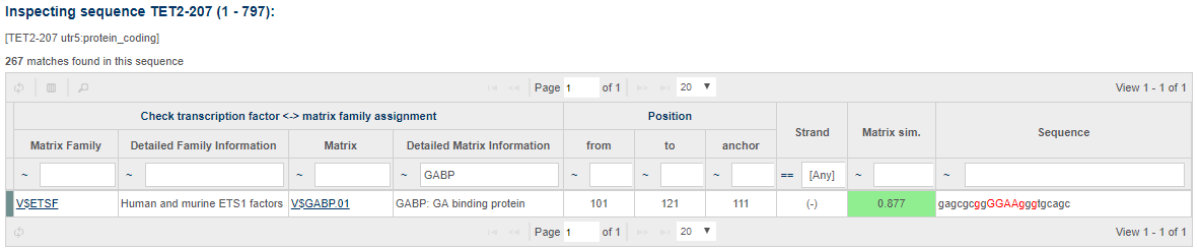
58167 matches found in this sequence

Page 1 of 1 — 20 — View 1 - 6 of 6

Check transcription factor <-> matrix family assignment				Position			Strand	Matrix sim.	Sequence
Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	from	to	anchor			
~	~	~	~ GABP	~	~	~	== [Any]	~	~
V$ETSF	Human and murine ETS1 factors	V$GABP.01	GABP: GA binding protein	14152	14172	14162	(+)	0.874	cactaatgGGAAgagggctt
V$ETSF	Human and murine ETS1 factors	V$GABP.01	GABP: GA binding protein	16062	16082	16072	(+)	0.874	tgaacctaGGAAgcggaggtt
V$ETSF	Human and murine ETS1 factors	V$GABP.01	GABP: GA binding protein	27482	27502	27492	(-)	0.868	aaaggcagGGAAggggatggg
V$ETSF	Human and murine ETS1 factors	V$GABP.01	GABP: GA binding protein	30755	30775	30765	(+)	0.874	ggaagaagGGAAgagggagta
V$ETSF	Human and murine ETS1 factors	V$GABP.01	GABP: GA binding protein	31480	31500	31490	(+)	0.942	gcgccactGGAAgagcggcgc
V$ETSF	Human and murine ETS1 factors	V$GABP.01	GABP: GA binding protein	32981	33001	32991	(+)	0.880	tataggtgGGAAgaggatctt

Page 1 of 1 — 20 — View 1 - 6 of 6

Figure A. 3: No GABPA DNA binding motifs were found in the genomic region, 5' UTR and 3'UTR of TET1. Five GABPA DNA binding motifs were found in the 5' flanking sequence.



Inspecting sequence TET2-207 (1 - 797):

[TET2-207 utr5:protein_coding]

267 matches found in this sequence

Page 1 of 1 — 20 — View 1 - 1 of 1

Check transcription factor <-> matrix family assignment				Position			Strand	Matrix sim.	Sequence
Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	from	to	anchor			
~	~	~	~ GABP	~	~	~	== [Any]	~	~
V$ETSF	Human and murine ETS1 factors	V$GABP.01	GABP: GA binding protein	101	121	111	(-)	0.877	gagcgcggGGAAgggtgcagc

Page 1 of 1 — 20 — View 1 - 1 of 1

Figure A. 4: A single GABPA DNA binding motif was found in the 5' UTR of TET2. The matrix similarity was 88%. The core GABPA binding site was located in the centre of the 20-nucleotide sequence.