The Role of DNA Methylation: Linking Exposure to Phenotype

Dissertation zur Erlangung der naturwissenschaftlichen

Doktorwürde durch den Fachbereich I – Psychobiologie der

UNIVERSITÄT TRIER

Vorgelegt von M.Sc Fleur A. D. Leenen

Gutachter:

Prof. Dr. C. P. Muller

Prof. Dr. J. Meyer

Luxembourg, February 2017

This doctoral thesis has been performed at the Allergology – Immunology – Inflammation Research Unit, Department of Infection and Immunity, Luxembourg Institute of Health, Luxembourg and its predecessor the Institute of Immunology, Public Research Centre for Health (CRP-Santé) and the National Health Laboratory, Luxembourg

Under guidance of

Prof. Dr. Claude P. Muller, Department of Infection and Immunity, Luxembourg Institute of Health, Luxembourg; previously Institute of Immunology, Public Research Centre for Health (CRP-Santé) and the National Health Laboratory, Luxembourg; Department of Immunology, University of Trier, Germany

and

Prof. Dr. Jobst Meyer, Department of Neurobehavioural Genetics, University of Trier, Germany

Dissertationsort: **TRIER**

Acknowledgments

Firstly, I would like to thank to my promoter Prof. Dr. Claude P. Muller, Department of Infection and Immunity, Luxembourg Institute of Health (formerly Institute of Immunology, Public Research Centre for Health (CRP-Santé) and the National Health Laboratory) for giving me the opportunity to join his team, as well as for his scientific guidance over the last years. I would also like to thank my supervisor Dr. Jonathan D. Turner for supporting me through this PhD project, offering his expertise for both theoretical and practical issues and adding the required note of optimism during harder times.

I am obliged to many of my colleagues at the Luxembourg Institute of Health for their aid and support with the experiments and the data analysis. I would like to thank in particular Stephanie Schmitz, Axel Dubois, Sophie Kirschner, Jean-Philippe Bürckert, Martha Elwenspoek, Wibke Cramaro, Carole Weis, Oliver Hunewald, Gilles Iserentant, Dominique Revets, Dirk Brenner for their support, the stimulating conversations/discussions, the little jokes, the laughs, all contributing to an friendly and motivational atmosphere during my PhD. A special thanks goes to the IEE team (formerly known as 'Psycho's'). My co-PhD students Sophie Kirschner and Martha Elwenspoek, who shared frustration and joy of a PhD project and always offered a helping hand and advice. Having two very skilled technicians, Stephanie Schmitz and Sophie Mériaux, made a world of difference when setting up and performing experiments. Their knowledge and know-how was invaluable for the experimental part presented in this thesis. And last but not least, Dr Jonathan Turner, who had to manage the bunch of us. Being part of the Psycho's, their supportive, relaxed, yet driven group dynamics, enabled me to grow both as a researcher as well as a person. I am grateful for the fun times, the laughs and the support I got from you all!

For the financial support, I would like to thank the Fonds National de la Recherche, the Luxembourg Institute of Health and the Ministère de la Culture, de l'Enseignement Supérieur et de la Recherche from Luxembourg

I would not have achieved all of this without the never-ending support of my family. My parents' unconditional support and encouragement during my studies, the testing times of 2005, was invaluable for all my achievements. A big thank you to my siblings, Sylvie, John and Maxime, who definitely helped me relativize and are always there when needed. I want to thank my friends/roomies, in particular Axel Dubois, Vincent Phillippart, Alix Mathot, Charline Mathei, Beate Vieth, Iris Tokka, Nele Boon, Bence Bukor, Sven DeBreucker, Jo Robbelein, Jean-Pierre Nzamurambaho, FX Joyeux and Silvia Hbj. Your friendship, during good and bad times, has been invaluable and enriched my life. Last but not least, I want to thank Jean-Philippe Bürckert, for his care and love, being there for me, always being able to make me smile, and support during testing times.


LUXEMBOURG
INSTITUTE
OF HEALTH
RESEARCH DEDICATED TO LIFE


DFG
Deutsche
Forschungsgemeinschaft


aides à la
formation
recherche


Fonds National de la
Recherche Luxembourg

Contents

Acknowledgments ..... **iii**

Contents ..... **v**

List of Figures ..... **ix**

List of Tables ..... **xii**

Abbreviations ..... **xiii**

General Abstract ..... **xvi**

Chapter 1 General Introduction **1**

1.1 EPIGENETICS: AN OVERVIEW 2

1.1.1 What is Epigenetics? 2

1.1.2 Epigenetic Markers: Inheritance and Reprogramming 2

1.1.3 Epigenetic stimuli: the Importance of Type and Time of Exposure 3

1.1.4 The importance of Epigenetics: Genetics versus Epigenetics? 3

1.1.5 The Principal Mechanisms of Epigenetic Regulation 4

1.1.5.1 DNA Methylation 4

1.1.5.2 Histone modifications 5

1.1.5.3 Non-Coding RNA 6

1.2 DNA METHYLATION 7

1.2.1 Biological Importance of DNA Methylation 7

1.2.2 Inheritance, Methylome Establishment and Evolution 7

1.2.3 Alternative DNA Methylation Forms 8

1.3 DNA METHYLATION CHANGES: TWO METHYLATION PARADIGMS[A] 8

1.4 ENVIRONMENTAL INFLUENCE ON PHENOTYPE DIVERSITY: A ROLE FOR SMALL EPIGENETIC CHANGES?[A] 10

1.5 WHAT IS A BIOLOGICALLY MEANINGFUL CHANGE IN METHYLATION LEVEL?[A] 14

1.6 METHYLATION: SINGLE CPG OR CLUSTERS?[A] 16

1.7 THE GLUCOCORTICOID RECEPTOR GENE: TRANSCRIPTION AND TRANSLATION 17

1.8 THESIS' RESEARCH OBJECTIVES 18

1.9 THESIS OUTLINE 20

Chapter 2 Where does transcription start? 5'-RACE adapted to Next-Generation Sequencing **21**

2.1. ABSTRACT 22
2.2. INTRODUCTION 23
2.3. MATERIAL AND METHODS 25
2.3.1. Cell Culture and RNA Extraction 25
2.3.2. RACE-PCR 27
2.3.3. High-Throughput Sequencing (HTS) 28
2.3.4. Bio-Informatics and Statistical Analyses 28
2.3.5. Translational Efficiency of Transcriptional Micro-Variants 29
2.3.6. RNA Structure Prediction 32
2.3.7. Minigene Design & Plasmid Construction 32
2.3.8. Western Blot 32
2.4. RESULTS 32
2.4.1. 5'-RACE Library Sequencing 32
2.4.2. The Mono-Exonic ADRB2R Gene shows Biological Microvariability around One Transcriptional Locus 35
2.4.3. The NR3C1 TSSs are Highly Variable 36
2.4.4. TSS Expression Profiles are Cell Line Specific 42
2.4.5. DAUDI Cells have Highly Variable Biological Replicates 44
2.4.6. Environmental Conditions Influence the TSS Pattern 47
2.4.7. Translational Efficiency Assay 49
2.4.8. NR3C1-Transcript Secondary RNA Structure is Influenced by the 5'UTR Length 51
2.4.9. 5′UTRs Influence the Relative GR Protein Isoform Distribution 51
2.4.10. Evolutionary Analysis of NR3C1 53
2.5. DISCUSSION 54
2.6. ACKNOWLEDGEMENTS 60
2.7. AUTHORS' CONTRIBUTIONS 60
Chapter 3 Genome-wide 5mC and 5hmC DNA profiling of maternally deprived rats 61
3.1. ABSTRACT 62
3.2. INTRODUCTION 63
3.3. MATERIAL AND METHODS 65
3.3.1. Animals & Maternal Deprivation Procedure 65
3.3.2. DNA Extraction and (Hydroxy)Methylated DNA Immunoprecipitation (MeDIP) 65

3.3.3. MeDIP Library Preparation and Sequencing 66
3.3.4. Bio-Informatics and Statistical Analyses 67
3.4. RESULTS 70
3.4.1. Experimental Paradigm 70
3.4.2. MeDIP-Sequencing 70
3.4.3. Group Differences in Distribution 80
3.4.4. Differential Methylation Identification 80
3.4.5. Gene Ontology and Pathway Analysis 94
3.5. DISCUSSION 98
3.6. ACKNOWLEDGMENTS 101
3.7. AUTHORS' CONTRIBUTIONS 101
Chapter 4 Epigenetic consequences of early life H1N1 infection 102
4.1. ABSTRACT 103
4.2. INTRODUCTION 104
4.3. MATERIAL AND METHODS 104
4.3.1. Animal Experiments 104
4.3.2. Cells and Viruses 105
4.3.3. Tissue Culture Infective Dose (TCID50) 105
4.3.4. Perinatal Infections and Challenges 105
4.3.5. Adult Re-exposure 105
4.3.6. Cytokine Analysis 107
4.3.7. DNA Extraction and Methyl-Seq 107
4.3.8. Data Analysis 108
4.3.9. Statistical Analysis 109
4.3.10. PCR Validation 109
4.3.11. Bisulphite Modified Pyrosequencing 111
4.4. RESULTS 112
4.4.1. Early Life H1N1 Infection 112
4.4.2. DNA Methylation of Splenocytes 112
4.4.3. Genome-wide Distribution of DMRs 113
4.4.4. Networks of Differentially Methylated Genes Point to Innate Immune Mechanisms 115
4.4.5. Validation of Differentially Methylated CpGs 116

4.4.6. Programming is not H1N1 Specific 117
4.4.7. H1N1 Programmed Loci are Programmed After polyI:C 118
4.5. DISCUSSION 119
4.6. ACKNOWLEDGEMENTS 121
4.7. AUTHORS' CONTRIBUTION 121
Chapter 5 General Discussion 122
5.1. GENERAL OVERVIEW 123
5.2. TOWARDS A MECHANISM LINKING SUBTLE METHYLATION CHANGES TO PHENOTYPES?[A] 123
5.3. EXPANDING THE MECHANISM FROM THE NR3C1 TO THE COMPLETE TRANSCRIPTOME AND PROTEOME?[B] 126
5.4. RE-DEFINING A 'GENE'[A] 128
5.5. CONCLUDING REMARKS AND FUTURE PERSPECTIVES 130
References 133
Presentations and Meeting participations 155
Publications 157
Erklärung 158

List of Figures

Figure 1: Schematic representation for the principal epigenetic regulation mechanisms. 4

Figure 2: Methylation changes have been associated with adult pathology and environmental factors in association studies 13

Figure 3: A schematic representation of the *NR3C1* structure and the internal ATG translation initiation codons resulting in different translational isoforms 18

Figure 4: UCSC browser views of *ADRB2R* gene and *NR3C1* gene..... 24

Figure 5: A schematic representation of the *ADRB2R* and *NR3C1* gene structure and the 5'-RACE-Sequencing workflow..... 26

Figure 6: Over multiplexing strategy used with Ion Torrent 314 and 316 chips 28

Figure 7: The plasmid 1C GRpcDNA3.1, into which all inserts were cloned 30

Figure 8: The mono-exonic *ADRB2R* shows transcriptional microvariability around one locus 34

Figure 9: The proximal *NR3C1* CpG island TSSs used in three DAUDI technical replicates are plotted, showing common and unique TSSs 35

Figure 10: Microvariable TSS distribution throughout *NR3C1* for A549 biological replicates and Detailed TSS usage pattern for the four cell lines and three treatment condition for exon 1B and 1F. 37

Figure 11: Detailed TSS usage pattern for the four cell lines and three treatment condition for exon 1A and 1I… 38

Figure 12: Detailed TSS usage pattern for the four cell lines and three treatment condition for exon 1D and 1J. 39

Figure 13: Detailed TSS usage pattern for the four cell lines and three treatment condition for exon 1E and 1H 40

Figure 14: Detailed TSS usage pattern for the four cell lines and three treatment condition for exon 1C. 41

Figure 15: Microvariable TSS distribution throughout the *NR3C1* for MCF-7 and T-cell biological replicates are plotted showing common and unique TSSs. 42

Figure 16: Differential expression of TSSs between multiple cell lines. 44

Figure 17: The proximal *NR3C1* CpG island TSSs used in two DAUDI biological replicates are plotted showing common and unique TSSs and their frequencies 45

Figure 18: Microvariable TSS distribution throughout the *NR3C1* under different treatment conditions 46

Figure 19: Differential expression of TSSs between untreated, dexamethasone, Interferon-ɣ and 5-AZA-2'-deoxcytidine exposed DAUDI cells..... 48

Figure 20: Translational efficiency and mRNA stability for exon 1A, 1B and 1C 50

Figure 21: The microvariable TSSs influence the relative abundance of N-terminal protein issoforms and the evolutionary conservation of this region 52

Figure 22: The microvariable TSSs influence the relative abundance of the N-terminal protein isoform. microvariable constructs,1B and 1C 53

Figure 23: Evolutionary conservation analysis of the *NR3C1*'s 5'UTR using the UCSC browser 54

Figure 24: Setting a cut-off value (%) to define a genuine TSS 56

Figure 25: Schematic representation of gene transcription and translation and how it is influenced by methylation 59

Figure 26: A schematic representation of the Maternal Deprivation (MD) model, its workflow and the phenotypic outcomes. 65

Figure 27: Global methylome and hydroxymethylome comparison. 69

Figure 28: MEDIPS parameter settings, balancing resolution and noise against the multiple testing burden, assessing the number of differentially methylated regions (DMR) in function of the optimal window width per read coverage category. 72

Figure 29: MEDIPS parameter settings, balancing resolution and noise against the multiple testing burden, assessing the number of windows to test in function of the optimal window width per read coverage category 73

Figure 30: MEDIPS parameter settings, balancing resolution and noise against the multiple testing burden, assessing the number of windows to test in function of the minimal read coverage per window width category. 74

Figure 31: MEDIPS parameter settings, balancing resolution and noise against the multiple testing burden, assessing the number of differentially methylated regions in function of the minimal read coverage per window width category. 75

Figure 32: Genome-wide domainograph depicting the absolute number of reads per segment along the chromosomes 77

Figure 33: Overall comparison of the methylome and hydroxymethylome within groups. 79

Figure 34: Overall comparison of the methylome and hydroxymethylome profiles between groups 81

Figure 35: The detection and verification of differences in DNA methylation levels between the treatment groups HD and MD 82

Figure 36: The detection and verification of differences in DNA methylation levels between the treatment groups C and HD, and C and MD. 86

Figure 37: The detection and verification of differences in DNA hydroxymethylation levels between the treatment groups HD and MD..... 87

Figure 38: The detection and verification of differences in DNA hydroxymethylation levels between the treatment groups C and HD, and C and MD 93

Figure 39: Early life viral infection mouse model: experimental paradigm and adult immune response to H1N1 re-exposure..... 106

Figure 40: Genome-wide domainograph depicting the absolute number of reads per segment along the chromosomes 113

Figure 41: Manhattan plot visualising the distribution of the RU values genome-wide and functional networks created by Ingenuity, based on the top 400 differentially methylated genes.. 114

Figure 42: Links between genes associated with influenza A and the 400 differentially methylated candidate genes 115

Figure 43: Normalised fold change in gene amplification values in comparison to the RU values of the 11 candidate genes 116

Figure 44: Methylation levels of the candidate genes measured by pyrosequencing. Differences in CpG methylation levels between control and H1N1 programmed mice 117

Figure 45: Methylation levels of the candidate genes measured by pyrosequencing. Differences in CpG methylation levels between control and PolyI:C programmed mice..... 119

Figure 46: A schematic representation of the NR3C1 5' UTR structure, showing the alternative first exons, the structure of the GR mRNA with the internal ATG translation initation codons, and their influence on the different transcriptional isoforms and the frequency of the different protein isoforms..... 125

Figure 47: Proposed mechanism for the creation of phenotype diversity by environmental factors..... 130

List of Tables

Table 1: The 'histone code' of common core histone modifications found in common psychobiological paradigms and phenotypes 5

Table 2: PCR primers..... 27

Table 3: Plasmid variants for translational efficiency analysis..... 30

Table 4: Sequencing summary 71

Table 5: The differentially methylated regions (DMR) for the 3 different group comparisons. For each DMR the exact position, the region width and the methalytion change is given. The change describes the change in 5mC of the first group compared to the second group. 83

Table 6: The differentially hydroxymethylated regions (hDMR) for the 3 different group comparisons. For each hDMR the exact position, the region width and the methalytion change is given. The change describes the change in 5hmC of the first group compared to the second group 88

Table 7: DAVID functional annotation clustering analysis of the differentially (hydroxy)methylated genes between C and HD, C and MD, and HD and MD. 94

Table 8: Ingenuity Pathway Analysis of the differentially (hydroxy)methylated genes between C and HD, C and MD, and HD and MD. 95

Table 9: PCR primers sequences, $MgCl_2$ concentrations and annealing temperatures..... 110

Table 10: Sequencing primers and biotinylated primers, $MgCl_2$ concentration and annealing temperatures 111

Table 11: Summary statistics of Methyl-Seq sequencing runs. 112

Table 12: Cytokine response to polyI:C stimulation in vitro after early life exposure to either H1N1 or polyI:C. 118

Abbreviations

5caC	5-carboxylcytosine
5fC	5-formylcytosine
5mC	5-methylcytosine
5hmC	5-hydroxymethylcytosine
5'-RACE	RNA ligase-Mediated Rapid Amplification of 5′ cDNA Ends
ΔG	Free Energy released on RNA folding
ADRB2R	Mono-exonic Beta-2-Adrenoceptor
AR	Allergic Rhinitis
ARI	Acute viral Respiratory tract Infections
AZA	5-AZA-2'-Deoxycytidine
bp	Base pair
C	Control
CAGE	Cap analysis gene expression
CDS	Complete coding sequence
CGI	CpG Island
CIP	Calf Intestinal Phosphatase
CMV	CytoMegaloVirus promoter
CNS	Central nervous system
CpG	Palindromic CG Dinucleotides
Dex	Dexamethasone
DF	Degrees of Freedom
DHM	Differential HydroxylMethylation
DhMR	Differentially Hydroxymethylated Region
DM	Differential Methylation
DMR	Differentially Methylated Region
DNMT	DNA Methyltransferase
DOHaD	Developmental Origins of Health and Disease
ELA	Early Life Adversity
ELS	Early Life Stress
ENA	European Nucleotide Archive

EWAS	Epigenome-Wide Association Studies
Fwd	Forward
gDNA	Genomic DNA
GIS	Gene Identification Signatures
GSC	Gene Signature Cloning
GWAS	Genome-Wide Association Studies
GR	Glucocorticoid Receptor
HAT	Histone AcetylTransferase
HD	Handling
HDAC	Histone DeACetylase
HPA	Hypothalamus-Pituitary-Adrenal
HTS	High-Throughput Sequencing
IFN-ɣ	Interferon-ɣ
i.n.	Intranasal
i.p.	Intraperitoneal
ISP	Ion Sphere Particles
ISPF	In Silico Phylogenetic Footprints
K	Lysine
Kd	mRNA Decay Constant
LG	Licking/Grooming
lncRNA	Long non-coding RNA
MD	Maternal Deprivation
MeDIP	Methylated DNA ImmunoPrecipitation
MGMT	MethylGuanine-DNA-MethylTransferase
miRNA	MicroRNA
MID	Multiplex Identifier
mQTL	Methylation Quantitative Trait Loci
ncRNA	non-coding RNA
NF-κB	Nuclear factor κB
NGS	Next-Generation Sequencing
nt	Nucleotide
OR	Odds Ratio

PCR	Polymerase chain reaction
piRNA	PiwiRNA
PND	Post-Natal Day
PPD	Post-Partum Depression
PTM	Post-translational modification
PTSD	Post-Traumatic Stress Disorder
R	Arginine
Rev	Reverse
RIN	RNA integrity number
RISC	RNA-Induced Silencing Complex
RITS	RNA-Induced Transcriptional Silencing
RT-qPCR	Quantitative Reverse Transcriptase Polymerase Chain Reaction
RU	Ratio of U-values
SD	Standard Deviation
SE	Standard Error
SEM	Standard Errors of Means
siRNA	Short interfering RNA
SNP	Single Nucleotide Polymorphism
TAP	Tobacco Acid Pyrophosphatase
TCID50	Half maximal Tissue Culture Infectious Doses per ml
TET	Ten-Eleven Translocation
TMM	Trimmed Mean of the M-values
TPCK	L-1-Tosylamido-2-Phenylethyl ChloromethylKetone
TSS	Transcription Start Site
U-value	Unmethylated value
UTR	UnTranslated Region
WGA	Whole Genome Amplification

General Abstract

DNA methylation, through 5-methyl- and 5-hydroxymethylcytosine (5mC and 5hmC) is considered to be one of the principal interfaces between the genome and our environment and it helps explain phenotypic variations in human populations. Initial reports of large differences in methylation level in genomic regulatory regions, coupled with clear gene expression data in both imprinted genes and malignant diseases provided easily dissected molecular mechanisms for switching genes on or off. However, a more subtle process is becoming evident, where small (<10%) changes to intermediate methylation levels were associated with complex disease phenotypes. This has resulted in two clear methylation paradigms. The latter “subtle change” paradigm is rapidly becoming the epigenetic hallmark of complex disease phenotypes, although we were currently hampered by a lack of data addressing the true biological significance and meaning of these small differences.

The initial expectation of rapidly identifying mechanisms linking environmental exposure to a disease phenotype led to numerous observational/association studies being performed. Although this expectation remains unmet, there is now a growing body of literature on specific genes, suggesting wide ranging transcriptional and translational consequences of such subtle methylation changes.

Data from the glucocorticoid receptor (*NR3C1*) has shown that a complex interplay between DNA methylation, extensive 5’UTR splicing and microvariability gives rise to the overall level and relative distribution of total and N-terminal protein isoforms generated. Additionally, the presence of multiple AUG translation initiation codons throughout the complete, processed, mRNA enables translation variability, hereby enhancing the translational isoforms and the resulting protein isoform diversity; providing a clear link between small changes in DNA methylation and significant changes in protein isoforms and cellular locations. Methylation changes in the *NR3C1* CpG island, alters the *NR3C1* transcription and eventually protein isoforms in the tissues, resulting in subtle but visible physiological variability. Implying external environmental stimuli act through subtle methylation changes, with transcriptional microvariability as the underlying mechanism, to fine-tune the total *NR3C1* protein levels.

The ubiquitous distribution of genes with similar structure as *NR3C1*, combined with an increasing number of studies linking subtle methylation changes in specific genes with wide ranging transcriptional and translational consequences, suggested a more genome-wide spread of subtle DNA methylation changes and transcription variability. The subtle methylation paradigm and the biological relevance of such changes were supported by two epigenetic animal models, which linked small methylation changes to either psychopathological or immunological effects. The first model, rats subjected to maternal deprivation, showed long term behavioural

and stress response changes. A second model, exposing mice to early life infection with H1N1, illustrated long-term immunological effects. Both models displayed subtle changes within the methylome. Suggesting/Indicating that early life adversity and early life viral infection 'programmed' the CNS and innate immune response respectively, via subtle DNA methylation changes genome-wide.

The research presented in this thesis investigated the ever-growing roles of DNA methylation; the physiological and functional relevance of subtle small DNA methylation changes genome-wide, in particular for the CNS (MD model) and the immune system (early life viral infection model) ; and the evidence available, particularly from the glucocorticoid of the cascade of events initiated by such subtle methylation changes, as well as addressing the underlying question as to what represents a genuine biologically significant difference in methylation.

Keywords: DNA methylation, hydroxymethylation, EWAS, Association studies, Biomarker, Transcriptional microvariability, Gene-environment interactions, glucocorticoid receptor (*NR3C1*), beta-2-adrenoceptor (*ADRB2R*), transcription start site variability, 5' UTR, 5'-RACE, postnatal viral infection, Developmental origins of health and disease (DOHaD), environmental programming, maternal deprivation model, piRNA, influenza virus A (H1N1)

Chapter 1

General Introduction

This chapter has been published in part as (sections 1.3, 1.4, 1.5 and 1.6):

DNA methylation: conducting the orchestra from exposure to phenotype?

Fleur A. D. Leenen[1,2], Claude P. Muller[1,2] and Jonathan D. Turner[1]

[1] Department of Infection and Immunity, Luxembourg Institute of Health, Esch-Sur-Alzette, L-4354, Grand-Duchy of Luxembourg

[2] Department of Immunology, Research Institute of Psychobiology, University of Trier, Trier, D-54290, Germany

Clin Epigenetics. 2016; 8(1): 92.

doi: 10.1186/s13148-016-0256-8

PMID:27602172; PMCID: PMC5012062

1.1 Epigenetics: An Overview

1.1.1 What is Epigenetics?

The genome, an arrangement of billions of nucleotides (nt), contains coding information controlling the transcriptomic, translational and the proteomic landscape, shapes the phenotypes of each individual. For a long time, the coding information was mainly attributed to the arrangement of the nt in the genome sequence[1–4]. However, over the past decade it has increasingly become accepted that the genome contains additional layers of information aside from the underlying sequence, which is referred to as the epigenome[2,5].

Epigenetics literally means ‘ beyond or above genetics [2,6] and was first described by Conrad Waddington in 1941, whom defined it as ‘the branch of biology which studies the causal interactions between genes and their products which bring the phenotype into being’[7–10]. Over the past decades, epigenetics has been assigned multiple different meanings[8,10]. Today, the original definition coincides with developmental biology[10] and epigenetics is defined as “the study of external environment factors causing heritable changes in gene expression and protein production, without affecting the underlying DNA sequence, and as such defining the resulting behavioural or physiological phenotype”[1,11,12]. Epigenetic regulation exerts control on both transcriptional and translational levels and is involved in processes such as X-chromosome inactivation, genomic imprinting, cell- and tissue-specific gene expression, and silencing of repetitive elements [1,3]. It plays a crucial role in normal developmental and chronical aging processes, as well as orchestrating a continuous adaptation to our environment throughout life[1,10,12–14].

1.1.2 Epigenetic Markers: Inheritance and Reprogramming

Inheritance of epigenetic marks has been described in several different organisms, such as plants, flies, rodents and humans[15]. Both oocytes and spermatocytes transmit epigenetic marks to the subsequent generation. Being specialised cells however, the epigenome requires a resetting to a more basic pluripotent state, in order to assure the full development into an adult organism containing many different cell types[16]. This resetting process, called ’reprogramming’, occurs early in mammalian embryogenesis and erases most of the inherited epigenetic markers[16–19]. Although most of epigenome is erased specific parts, such as imprinted genes, maintain their epigenetic marks. These epigenetic tags are passed on unchanged across generations[18,19]. Later on during embryonic development, prior to blastocyst implantation, individual-specific epigenetic marks are established de novo[161,16,20]. These determine the developmental course of cells until they attain a specialised state. Consequently, epigenetic patterns are cell-

specific[12,20,21]. This re-established epigenome is not a completely fixed entity, but retains some plasticity, enabling it to adapt and adjust in response to its changing environment/environmental challenges[19].

1.1.3 Epigenetic stimuli: the Importance of Type and Time of Exposure

Epigenetics mechanisms underlie the continuous adaptation to our environment throughout life, functioning as an interface between the genome and the environment[1,3,22]. There exists a multitude of environmental stimuli, e.g. of nutritional, chemical, social or physical nature, that can affect the exposed individual's epigenome and possibly the epigenome of its offspring[1,3]. As shown by Lillycrop et al (2005)[23], linking the protein concentration of the maternal diet with different methylation levels in *PPARα* and *NR3C1*, the nature of environmental stimuli affects the epigenetic outcome. Equally important is the time of exposure[1,12,14]. There are three periods of heightened epigenetic plasticity and sensitivity, in utero, the immediate postnatal period or early life, and adolescence [1,12,14,22]. The sensitivity to epigenetic modifications during the in utero period has been underlined by epidemiological observations and animals studies[1,14]. Small or transient environmental changes, such as maternal diet, maternal state of mind, stress or exposure to metals or chemical compounds, impact the offspring's' epigenome inducing phenotypical differences or programming measurable in adulthood[1,14,24]. Early life experiences have been associated with life-long health trajectories and behavioural phenotypes in both animal models as well as humans[12,22,25]. Adolescence, despite being an important neurodevelopmental stage, received only limited attention[1,14,26,27]. Although mainly studied in animal models[1,27], the recent TRAIL studies[28] linked stressful life events or traumatic youth experiences with increased *NR3C1* and *SLC6A4* methylation levels in humans[26,29].

1.1.4 The importance of Epigenetics: Genetics versus Epigenetics?

Milestones such as the first completely sequenced human genome, and the development of new powerful array- and sequencing-based techniques, have strengthened the focus on genetic studies. The genome sequence was anticipated to provide a blueprint for normal development. Assuming that variability in disease heritability was mainly caused by genetic variants, it would enable a better understanding of disease predisposition and development[1,12] Genome-wide association studies (GWAS) have indeed produced strong associations between certain traits, diseases and genetic variants. Yet, these only account for a small fraction of the causality and observable phenotypic diversity [1,12,30,31]. There is now a large body of evidence that suggests the environment-genome interaction, i.e. the epigenome, needs to be considered to understand both pathophysiology, as well as association between traits and diseases[1,3,12,14].

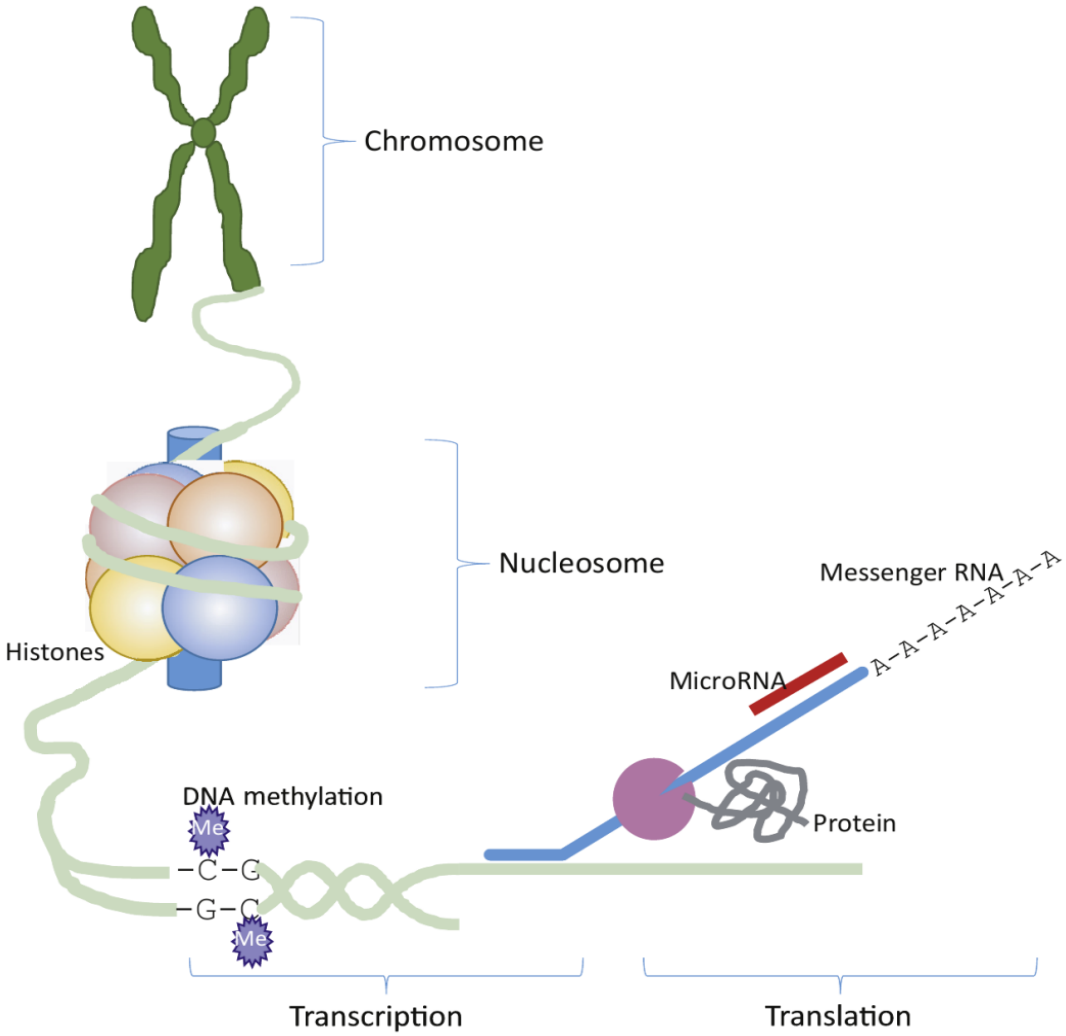

Figure 1: Schematic representation for the principal epigenetic regulation mechanisms. Genomic DNA, packed in chromosomes, is wrapped with 146-147 basepairs around nucleosomes. Histone modifications are considered to be the first level of epigenetic regulation. The second level is DNA methylation of the cytosines of palindromic CpG dinucleotides and the third level of regulation concerns microRNA, affecting the translation process. Image from [1].

1.1.5 The Principal Mechanisms of Epigenetic Regulation

Epigenetic regulation encompasses covalent DNA modifications, DNA packaging and chromatin states, and post-transcriptional regulation. These are mainly performed through the following three functional mechanism: DNA methylation, histone modifications and microRNAs[1,3,5,12,32] (Fig. 1).

1.1.5.1 DNA Methylation

DNA methylation (5-methylcytosine, 5-mC), one of the best studied epigenetic modifications, operates directly on the genomic DNA, where a methyl group is added to the 5' cytosine in a palindromic CG dinucleotide.

It is known to influence gene expression via chromatin re-modelling and gene transcription regulation[13,33–38]. DNA methylation is described more in detail in section 1.2.

1.1.5.2 Histone modifications

Chromatin, a nucleoprotein structure responsible for the DNA organisation within the nucleus, consists of a core unit called the nucleosome (histone octamer) around which 146 DNA base pairs are wrapped. Through its structural conformation, it tightly regulates the access to the underlying DNA sequence, by either effectively blocking or opening for processes such as transcription, damage repair, or recombination [1,4].

Histone modifications are post-translational modifications (PTM) that dynamically remodel chromatin. This vast array of modifications, including covalent phosphorylation, methylation, acetylation, and ubiquitinylation, is often referred to as the 'histone code' (Table 1)[1,5,39–42]. Depending on the chemical nature of the PTM, its position and the nature of the modified histone in general, PTMs affect transcriptional activity either by directly altering the overall chromatin structure, or indirectly by affecting the recruitment of secondary effectors[1,39].

Table 1: The 'histone code' of common core histone modifications found in common psychobiological paradigms and phenotypes[1]

Histone H3						**Histone H4**	**Histone H2B**
Residue	K4	K9	K27	K36	K79	K20	K5
Monomethyl			Activation		Activation	Repression	Activation
Dimethyl	Activation		Repression	Activation	Activation		
Trimethyl	Activation	Repression	Repression		Activation + Repression	Repression	Repression
Acetylation	Activation	Activation	Activation				Activation

Histone acetylation is a transient modification (Table 1), which has a direct impact on the chromatin structure[1,5]. Acetyl groups are either actively added or removed to the lysine (K) residue side chains by histone acetyltransferases (HATs) or histone deacetylases (HDACs) respectively. Acetylation neutralises the lysine's positive charge, weakening the bond between the histones and the negatively charged DNA, rendering the underlying DNA sequence more accessible [1,39].

Histone phosphorylation is a highly dynamic process, mediated by kinases and phosphates, mainly targeting serine, threonine and tyrosine primarily situated in the N-terminal histone tails. By adding a phosphate group, the kinases add a negative charge to the histone, irrevocably altering the chromatin structure[39].

Histone ubiquitination is a dynamic modification, which covalently binds a 76-amino acid ubiquitin polypeptide to lysine by the sequential catalytic actions of an ubiquitin-activating enzyme (E1), an ubiquitin

conjugation enzyme (E2), and an ubiquitin ligase (E3). Due to the considerable size of the ubiquitin peptide, approximately two-thirds of the histone size, ubiquitination or de- ubiquitination is thought to profoundly remodel the chromatin structure and affect the transcriptional activity[1,39,40].

Histone methylation is the process of depositing or removing methylation marks on the side chains of lysine (K) or arginine (R) residues (Table 1), mediated by histone methyltransferases and histone demethylases respectively[1,39]. The lysine residues can be either mono-, di-, or tri-methylated, whereas the arginine residues can be mono-, symmetrically, or asymmetrically dimethylated[39]. Contrary to acetylation, phosphorylation, or ubiquitination, methylation does not alter the histone's charge or remodels the chromatin structure in a direct way. But rather affect the transcriptional activity by regulating the recruitment of the transcriptional machinery, e.g. by increasing the side chains' hydrophobicity and the creation of new binding surfaces for reader proteins[1,39].

1.1.5.3 Non-Coding RNA

There are four non-coding RNAs (ncRNAs) associated with epigenetic mechanism: microRNAs (miRNA), short interfering RNAs (siRNAs), piwi RNAs (piRNAs) and long non-coding RNAs (lncRNAs). All involved in processes such as heterochromatin formation, histone modification, DNA methylation targeting and gene silencing.

miRNA are small non-coding RNA molecules (~22 nucleotides), which modulate the gene expression at post-transcriptional level, by either binding mRNA or targeting specific genes[1,32,43–46]. For the former, miRNA together with the Argonaute protein family forms a RNA-induced silencing complex (RISC), which binds mRNA by imperfect base pairing. The RISC binding blocks the mRNA translation, and starts mRNA deadenylation, degradation, and cleavage[1,45]. miRNA plays a role in the translational control of proteins involved in the HPA axis and pathologies linked to stress[1]. Additionally, studies[47,48] have suggested the implication of miRNA in chromatin remodelling and DNA methylation[32].

siRNA, a molecule of approximately 21 nucleotides, function in a similar way as miRNA to mediate gene expression at post-transcriptional level[48]. Additionally, siRNA was also suggested to be involved DNA methylation and chromatin remodelling. When bound to an RNA-induced transcriptional silencing (RITS) complex, it promotes chromatin condensation and H3K9 methylation[49].

piRNA are small non-coding RNAs (24nt and 32nt) that interact /form a complex with Piwi protein family[50–55]. Their primary roles are the suppression of transposon activity and gene expression regulation, both at transcriptional and post-transcriptional level, in both germline and somatic cells[50–54,56]. Transposon regulation,

either on genomic or epigenetic level, is thought to be important for CNS variability and mosaicism, as well as neural development and plasticity. Dysregulation could compromise cellular homeostasis and possibly resolve into pathology onset [50–52,54,56].

lncRNA, ncRNA with a length of 200 nucleotides or more, form an existential portion of the genome part being transcribed[46,57]. They mainly have regulatory roles and are implicated at almost every level of gene expression, such as network formation of ribonucleoprotein complexes and chromatin regulators, modulating mRNA stability and translation by targeting epigenetic modifications in the nucleus, etc[57,58]. lncRNAs function as chromatin modulators, i.e. they associate with chromatin-modifying proteins, guiding their catalytic activity to specific genomic locations and hence influencing the transcription activity[58].

1.2 DNA Methylation

One of the best studied epigenetic modifications is DNA methylation, a natural covalent DNA modification that operates directly on the DNA sequence. The mechanism consists of the addition of a methyl group on the 5' cytosine of a palindromic CpG pair, i.e. a dinucleotide where the C is immediately followed by a guanine[1,13,33–38,59] (Fig. 1). Approximately 60% to 80% of all CpGs are thought to be methylated. CpG dinucleotides are distributed infrequently throughout the genome. Approximately 98% of the mammalian genome is CpG-deficient, sporting sparsely dispersed single CpGs. These mainly occur in repetitive DNA elements and centromeric regions, and tend to be methylated. The remaining ~2% of the genome features an overrepresentation of CpGs as clusters called 'CpG islands' (CGI's). These CGI's are mainly associated with gene promoter and regulatory regions and thought to be protected from methylation[13,33,35,59,60].

1.2.1 Biological Importance of DNA Methylation

In mammals, CpG methylation plays a central role in the mammalian development and is involved in the regulation of gene expression through processes as chromatin re-modelling, gene transcription regulation, maintenance of X-chromosome inactivation, gene imprinting and tissue-specific gene expression [1,13,33–38,61]. As one of the main epigenetic mechanism and the best studied one, 5mC is considered to be one of the principal interfaces between external environment and the genome. Methylome heterogeneity is partially inherited, but mainly caused by environmental programming, explaining the observed phenotypic variety in mammals[1,12,60,62].

1.2.2 Inheritance, Methylome Establishment and Evolution

As described in paragraph 1.1.2, DNA methylation patterns are established during embryogenesis and early life. The DNA methylation patterns inherited from both gametocytes are almost completely erased during

embryogenesis and individual foetal de novo methylation patterns are re-established prior to the blastocyst implantation[1,20]. This only partial inheritance of the parental methylome is extremely well demonstrated in the case of monozygotic twins, whom share an almost identical methylome, yet exhibit different methylation patterns compared to their parents. The epigenetic discordances observed in monozygotic twins mainly developed only after birth, and tend to increase with age[1,60,63–65].

In general, the methylome is relatively stable over time, with global methylation levels slowly decreasing with age. Demethylation was thought of as a passive process. Recently however, the presence of an active demethylation process gained acceptance. Currently, the methylome is viewed as a stable, but plastic entity that may be reshaped during several well defined periods in time[66–68].

1.2.3 Alternative DNA Methylation Forms

Historically, demethylation was thought to be a passive process that was a direct consequence of the failure to maintain methylation levels during cell division. However, it is now accepted that DNA methylation plasticity is coupled with active demethylation[66–68]. Although its role is not fully elucidated, 5-hydroxymethylcytosine (5hmC) is thought to be the first step of the active DNA demethylation process. 5hmC is generated by oxidation of 5mC by one of the ten-eleven translocation (TET) proteins 1-3, and can sequentially be further oxidised to either 5- formylcytosine (5fC) or 5-carboxylcytosine (5caC). The latter two can be disposed of via processes including base excision or not [36,66,69]. However, although 5hmC has been detected in nearly all tissues, it is particularly enriched in the central nervous system (CNS), where it is thought to also be involved in active transcription of neuronal genes, and brain development [37,66,70].

1.3 DNA Methylation Changes: Two Methylation Paradigms[a]

There are two concurrent paradigms for DNA methylation: the first paradigm is a clear mechanism for switching genes on/off through complete methylation or demethylation of genomic regulatory regions. DNA methylation has long been considered a marker of permanent gene silencing (imprinting) or reactivation[4]. In malignant diseases this simple on/off switch is often observed activating or silencing oncogenes and tumour suppressor genes respectively[71] e.g. O^6-methylguanine-DNA-methyltransferase (*MGMT*) methylation levels vary from 0 to >60%. Although it is not the focus of this review, and has been extensively reviewed and meta-reviewed elsewhere, the principal diagnostic epigenetic cancer biomarkers available such as *VIM, SEPT9, SHOX2, GST1, APC, RASSF1A* share this clear pattern of no or little methylation, and clear (>60%) hypermethylation, with almost nothing in-between[72]. However, this simple paradigm has been challenged, and a second paradigm is emerging. In this second paradigm, intermediary DNA methylation levels are fine-tuned,

a. This section was published in Lleenheer, A. D., Wilmes, G. M. & Turner, J. D. DNA methylation: conducting the orchestra from exposure to phenotype? *Clin. Epigenetics* 8, (2016)

often influenced by the external environment, and are becoming the epigenetic hallmark of many complex non-malignant disorders. In this case, the association of DNA methylation with an observed phenotype occurs through small differences in the methylation level of <10% and often only 1-5%, at single CpGs or over very limited genomic regions[34,73]. Such limited differences in DNA methylation are known to be set during periods of epigenetic sensitivity[1]. Additionally, they have been shown to play a role in creating a large diversity in phenotypes linked to the onset of many complex non-malignant diseases, such as type 2 diabetes, major depression, schizophrenia, hypertension, cardiovascular diseases[15,73]. Epigenetic phenotypes are not necessarily restricted to an exposed individual. Some epigenetic marks are transgenerational, hereby transmitting the phenotypic trait and possible the linked disease to the offspring[4,15,74,75].

This split into two paradigms has been accompanied by the expansion of the roles of 5mC and 5hmC. Both are now considered important factors assuring the quantitative, spatial and temporal regulation of gene expression as well as normal development and differentiation[4,34,59]. By targeting promoter CpGs and CGIs, DNA methylation was mainly thought to interfere with the transcription initiation and consequently gene silencing or reactivating[4,13,76]. Genome-wide analysis techniques showed DNA methylation influences many other mechanisms, such as alternative splicing, alteration of enhancer, insulator and regulatory element function, hence altering gene expression[73,76–78]. For both tissue-specific regulation and non-malignant disorders, changes in gene expression are frequently caused by small changes in methylation levels, often at single CpG dinucleotides or over a limited genomic region. Such small differences have a big impact on the phenotype diversity that is linked to the onset of non-malignant diseases[77,79]. Plasticity in methylation levels allows environmental adaptations, transient changes and long-term alterations of the cell's transcriptomic profile, hereby contributing to the diversity of characteristics, both biochemical and physiological, and hence the phenotypic variations observed in human populations[4,33,34]. These mechanisms have been associated with the onset and maintenance of pathogenesis[33,80,81], and methylation has increasingly been associated with the aetiology and onset of multiple, non-malignant, complex disorders[6,77,80–82].

DNA methylation can be summarised as either discrete hyper- and hypo-methylation coupled with clear gene silencing, and easily dissected molecular mechanisms, or a more subtle complex process where small (<10%) methylation changes are associated with disease phenotypes and many transcriptional processes. This leads us to the fundamental question of the biological significance of such small changes, and how they give rise to the final disease phenotype. There is currently doubt over the true biological relevance of such small changes, if they are genuinely meaningful, what mechanisms link such limited changes in methylation to the phenotype, and how this affects our view of what a gene is. In this review we summarise the pathophysiological and clinical

associations that have been made to small, subtle methylation changes; the ever-growing roles of DNA methylation; and the evidence available, particularly from the glucocorticoid receptor of the cascade of events initiated by such subtle methylation changes, and conclude that such small changes may reflect genuine biological differences.

1.4 Environmental Influence on Phenotype Diversity: A Role for Small Epigenetic Changes?[a]

Environmental influence on DNA methylation, gene expression, phenotype and disease onset have been extensively studied. In the framework of the Developmental Origins of Health and Disease (DOHaD) paradigm in utero or early life conditions program lifelong health trajectories. This paradigm focusses on organisms' biological plasticity to adjust their phenotype to their environment over the short and long term in which epigenetic processes such as DNA methylation are thought to be involved. Mismatches between the pre-/post-natally anticipated and the actual mature environment predisposes organisms to disease (Fig. 2)[30,44,83].

Obesity, hypertension, cardiovascular diseases, diabetes: The prevalence of obesity, hypertension and the accompanying cardiovascular disorders, and diabetes have been associated with early life environmental factors, such as diet, parental diet, and maternal mood during pregnancy (Fig. 2)[20,32]. In the 'small litter' neonatal overfeeding model appetite was dysregulated via hypermethylation of the POMC promoter at the NF-kB and Sp1 binding sites necessary for inducing POMC expression by leptin and insulin[84]. Consequently *POMC* expression will be reduced despite insulin or leptin presence)[20,32]. Parental diet strongly influenced their offspring's methylation profile and phenotype (Fig. 2)[32,74]. Gestational high fat diets increased the offspring's probability of developing obesity, metabolic syndrome, insulin resistance and diabetes in both humans and animal models (Fig. 2)[32,85,86]. Conversely a low-protein maternal diet peri-conceptually or during gestation was associated with lower birth weight, schizophrenia, an increased risk of the offspring developing cardiovascular diseases, hypertension, dyslipidaemia and obesity (Fig. 2)[32,74,87–89]. A well-known natural experiment for transgenerational nutri-epigenomics was the 'Dutch Hunger Winter'. Dutch individuals exposed in utero to malnutrition and their direct descendants had higher rates of obesity (BMI raise of 7.4% in women[90]), hypertension (OR 1.44[91]), an increased risk for cardiovascular disorders (coronary heart disease: OR 3.0) and impaired glucose homeostasis later on in life (glucose tolerance index: prenatally: −21%; late gestation: −4%; mid gestation: −24%; early gestation: −37%)[92,93]. This was accompanied by hypomethylation in *IGF2* (-5.2% to -5.6%) and *INSIGF* (-1.6%), and hypermethylation of *IL10* (2.4%), *ABCA1* (1.7%), *GNASAS* (1.1%) and *LEP* (1.2%) (Fig. 2)[88,89]. Late gestational exposure appeared to be a less sensitive period, as it only affected the

a. This section was published in Leenen, F. A. D., Muller, C. P. & Turner, J. D. DNA methylation: conducting the orchestra from exposure to phenotype? *Clin. Epigenetics* **8**, 87 (2016).

methylation profile of *GNASAS* (-1.1%) from the limited number of target genes investigated[88]. An equally important factor affecting the offspring's methylation profile and phenotype was maternal mental state during pregnancy (Fig. 2). Gestational depression during pregnancy associated with a lower birth weight (OR 3.6, 95% CI: 1.1–11.4), obesity, as well as cardiovascular disorders and diabetes in later life (Fig. 2)[14,94]. This was accompanied by higher *MEG3* methylation levels (2.4%) and decreased methylation of IGF2 (-1.6 %) compared to children with normal birth weight[94]. Offspring with a higher birth weight than normal showed a hypermethylation of *PLAG1* and *PEG10* (5.9% and 3.4% respectively), genes that have previously been linked to the regulation of placental and foetal growth and development, growth in general, and diabetes[94]. Maternal depression during the second trimester of pregnancy on the other hand was linked with hypomethylation of the *SLC6A4* promoter region for both mother and child (Fig. 2)[94,95]. Overall it seems that individuals subjected to poor diets in utero or early life, or born out of mothers suffering from severe depression during gestation develop phenotypes with a higher prevalence of obesity, hypertension and the accompanying cardiovascular disorders and diabetes. It remains unclear, however, whether the methylation changes are part of the mechanism increasing disorder prevalence or rather an additional consequence.

Psychopathology and behaviour: The risk of developing psychopathologies, cognitive, behavioural, anxiety and mood disorders or suicidal tendencies later in life have been related to stressful/traumatic experiences during early development and early life (Fig. 2). These early life periods profoundly affect development of the central nervous system, the limbic structures or hypothalamus-pituitary-adrenal (HPA) axis regulation. Although, the underlying mechanisms are unknown, the detection of methylome and gene expression changes between phenotypes highlights the importance of DNA methylation [34,66,96–99]. *BDNF*, a gene involved in neurodevelopment, neuroplasticity, the onset of psychiatric disorders and suicidal behaviour, has been associated to early life adversity (ELA) (Fig. 2). Rat and mouse models for ELA and depression showed that the epigenetic processes controlling *BDNF* transcription were stress sensitive. The *BDNF* promoter region was hypermethylated (10% to 15% per CpG on average), with the ensuing lower expression levels [97,100]. Similar *BDNF* methylation patterns were observed in post-mortem adult brains from suicide completers[97,98]. HPA-axis and stress response dysregulation have been among the most consistent biological findings in major depression and psychopathology [101]. *NR3C1,* coding the central HPA axis regulating glucocorticoid receptor, was frequently investigated as part of the mechanism linking ELA and the predisposition towards psychopathology or suicide risk (Fig. 2)[97,102,103]. In rat models, ELA caused a hypomethylation of the hippocampal *NR3C1* promoter (2% to 4%), which significantly altered the gene expression and HPA-axis responsivity[13,25,102,104]. Post-mortem brain analyses and clinical studies observed similar trends[25,97,102,103]. Suicide completers with a history of ELA had increased hippocampal *NR3C1* promoter methylation and decreased *NR3C1* expression (Fig. 2)[25,97,102,105].

Similarly, higher hippocampal and leukocyte *NR3C1* methylation levels were observed for healthy individuals previously exposed to ELA. As such, the evidence is now strong that *NR3C1* methylation is part of the panoply of changes linking ELA events to later life psychopathology, although there is no definite evidence as to whether it is a direct mechanism or an additional independent event[103]. HPA-axis changes were not limited to *NR3C1*, ELA and early life stress (ELS) also induced a sustainable hypomethylation of *AVP* (< 10% per CpG position) and *CRH* (< 15% per CpG position[34,106,107], as well as psychopathology associated genes such as *SLC6A4*[108,109]. The methylation status of the *SLC6A4* promoter was shown to be affected by abuse as well as genotype[108]. Although DNA methylation appears to explain the link between ELA and psychopathology through HPA-axis regulation, robust proof of principle remains, however, to be provided as connecting methylome profile alterations and gene expression robustly failed[108]. The link between epigenetic alterations and neuropsychiatric disorders remains unproven.

Asthma and allergic pathologies: Genetic makeup has been seen in many studies to be one of the strongest risk factors for eventually developing allergic symptoms[110,111] consistent with epidemiological evidence of an increased allergic rhinitis (AR) concordance in twin studies[112]. Although many candidate genes have been suggested, genome-wide association studies (GWAS) have not, so far, identified “overlapping and consistent genetic components” [113,114], and epigenetic mechanisms have been proposed to play an equally important role. For example, the promoter methylation level of *NPSR1* showed small but significant differences for persons suffering from severe adult or allergic asthma in children (Fig. 2). *NPSR1*, normally highly methylated (>75%), was hypomethylated by -3.29% and -1.40% for severe adult asthma and allergic asthma in children respectively (Fig. 2)[115]. DNA methylation levels have also been associated with factors such as the current smoking behaviour, parental smoking during infancy and the month in which the sample was taken[115], which are thought to be implicated in the onset of both asthma and allergic diseases.

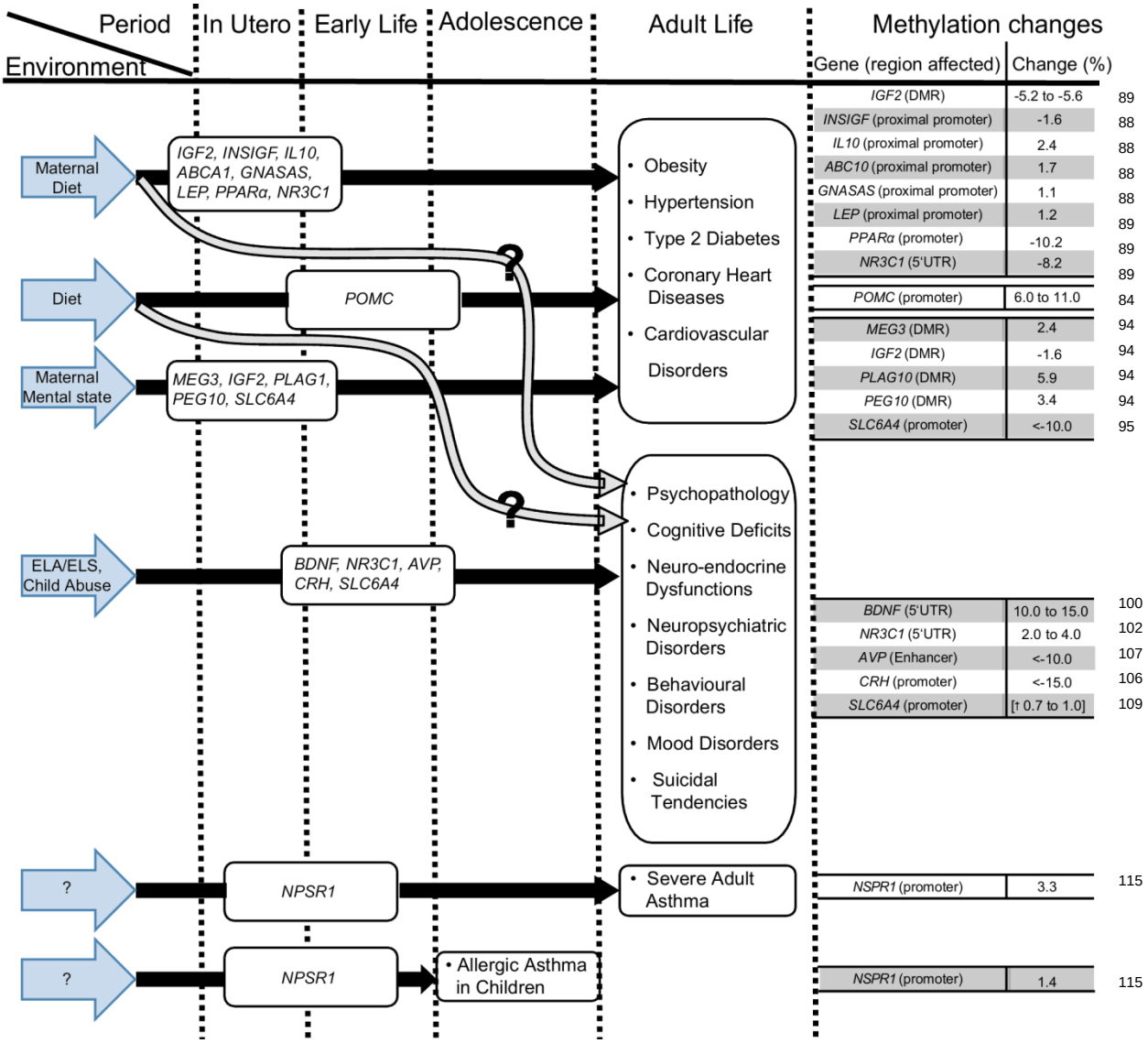

Figure 2: Methylation changes have been associated with adult pathology and environmental factors in association studies. Environmental factors during certain periods of life have been linked to genes or disorders. The changes in methylation are listed next to the phenotype in adult life. ELA Early life adversity; ELS early life stress; † Data are Infinium β-values.[24]

Associations and hypotheses, not mechanisms: The increased number of association studies has given us a better insight of the environmental impact on phenotype development (Fig. 2). Yet, as the majority of these observational studies did not address the underlying mechanisms, we are left with associations and hypotheses. In order to enhance our understanding, future research should address the underlying process and try to provide robust evidence for the exact cascade of events linking environment and phenotype differences. A good example of such a clear link is the viable yellow Agouti (A^{vy}) mouse model, where the offspring's coat colour shifts between yellow and brown due to incomplete erasure of the maternal epiallele during embryogenesis. The Agouti gene has a methylation sensitive intracisternal-A particle retrotransposon inserted at the 5'end that functions as a transcription start site. Large changes in methylation of the A locus from ~70% to ~25% result in a yellow rather than the natural brown coat. The offspring phenotype and methylation level appeared to be

heavily influenced by those of the mother. Oocyte transfer to surrogate mothers of a different epigenetic background, however, was necessary to demonstrate that the offspring epigenotype depended on the incomplete erasure of the maternal methylation during embryogenesis, rather than the uterine environment[116]. For the studies mentioned throughout section 1.4, such detailed mechanistic studies are unfortunately absent.

Currently EWAS data such as those highlighted above are a perfect storm of visibly low methylation levels, of which the biological meaning is uncertain and a large variety of confounding factors influencing their methylation state[117]. There is a void, with limited information or guidelines on how to design and conduct meaningful EWAS. Adopting a set of guidelines or rules for best practices, in a similar manner to GWAS, would benefit EWAS interpretation and increase their relevance.

1.5 What is a Biologically Meaningful Change in Methylation Level?[a]

As we[14] and others[118] have previously noted, there is doubt over the true biological relevance of small changes in absolute methylation levels, and it has been suggested that authors may have increased confidence in the biological significance of methylation differences >10%, and conversely, must treat differences of <5% with extreme caution[119].

Reducing sample variability: Different cell types have specific epigenetic profiles[120], and measuring aggregate levels over a large populations is a major source of variability. Since methylation is essentially binary i.e. in any given cell a specific CpG is either methylated, unmethylated, or potentially hemi-methylated (asymmetric methylation of two alleles), the methylation levels measured simply reflect the proportion of methylated cells in the original sample[14,121–123]. Consequently, minor changes in methylation may actually represent small changes in the cellular composition of the original sample rather than a genuine difference due to the disease or paradigm studied. As an aside, the most widely used sample, blood, is unfortunately one of the most variable, although there is now a well-established procedure that adequately corrects for this variability[124,125].

The impact of the data format: Teasing out the biologically relevant changes in methylation levels is further complicated by the current trend towards reporting fold changes rather than absolute methylation values. The appropriate data to report is naturally specific for the analysis method employed. For example MeDIP-Seq and Infinium arrays (Illumina) give M and β-values that may correlate to the percentage methylation, they are relative values, and they may be considerably different from the direct measurement (e.g. by pyrosequencing) of the absolute methylation levels. Although there is no direct comparison available it has been suggested that “a β-value of 0.8 might correspond to a level of 30% methylation”[118], however, as highlighted above, when

a. This section was published in Leenen, F. A. D., Muller, C. P. & Turner, J. D. DNA methylation: conducting the orchestra from exposure to phenotype? *Clin. Epigenetics* 8, (2016)

methylation levels are low, as in the case of *NR3C1*, a relatively small change in the absolute methylation level will be represented as a wildly exaggerated fold change or percentage increase. In the current situation, where small differences in methylation or low methylation values are being reported, there are additional technical concerns with data analysis and reporting. Illumina β-values are predominantly reported as they can be considered an approximation to the percentage methylation present in the original sample. However, this is only valid for values in the "middle methylation range"[126], with severe heteroscedasticity for low and high methylation values. This has lead authors to suggest that statistical analyses are performed with M-values, but to report β-values[126].

Confounding variables: Interpretation of small methylation changes is further complicated by the numerous sources of epigenetic variability that are currently poorly defined. There is significant evidence that many genetic, demographical, clinical and environmental factors are strong cofounding variables[118]. However, these underappreciated confounding variables all contribute to the overall measured phenotype. This was highlighted by the low intra-individual, but high inter-individual, difference in methylation levels we observed throughout the human brain[127]. Population-wide, 5-mC levels are both reduced and redistributed with age[128], and are generally higher genome-wide in males than females[129,130]. Locus-specific differential hyper- or hypo-methylation has, however, been reported for both men and women[131–134]. Equally, the underlying genomic sequence heavily influences DNA methylation levels. Although there are numerous other examples [135–139], the best estimate is that approximately 2% of the investigated CpGs that cover up to 9.5% of genes represent methylation quantitative trait loci (mQTLs), and may operate over distances up to 5 kb[140]. Our *NR3C1* data demonstrated that methylation of the *NR3C1* promoter 1H was associated with a complete haplotype (haplotype 2), rather than a specific SNP, operating over approximately 3kbp. The effect of the underlying genome sequence is also highlighted by pervasive asymmetric methylation in diploid genomes (i.e. difference between the two alleles), particularly outside imprinted regions[137,141–143]. This asymmetry is known to be regulated by underlying heterozygotic genetic variants. In trans-generational epigenetic inheritance, there is now convincing evidence that it is the genomic sequence, rather than the parental DNA methylation levels that determines 5mC levels during embryogenesis[144]. Furthermore, allele-specific methylation events are found in unrelated individuals with the same haplotype/genotype as well as in multiple inter-individual tissue[137]. Although the evidence for these confounding factors is growing, there are still no population-epigenetics principles available to guide study design, analysis and interpretation. However, we suggest that moving towards sequencing based techniques (whole genome bisulphite sequencing, reduced representation bisulphite sequencing, MeDIP-Seq, etc.) will allow access to the genomic variants that is not available in array-based techniques.

Purpose of epigenetic studies: The current interest in DNA methylation is primarily to exploit its potential as a biomarker. In both malignant and complex non-malignant diseases work has centred on associating methylation changes with the external environment, particularly to exploit the latency between exposure and disease development. In both the DOHaD and "foetal origins" models, early life events induce epigenetic changes that are maintained lifelong. Similarly, many environmental factors e.g. chemical, biological (e.g. toxins, allergens) or heavy metal exposure alter the epigenome, ultimately increasing the risk of developing cancer[145,146], for example, asbestos exposure leads to *DNMT* overexpression, highly specific methylation patterns and eventually malignant pleural mesothelioma[147,148]. In both cases, there is a considerable period of latency from the exposure to clinically discernible disease ranging from a few years (autism, obesity) to many decades (cardiovascular disease, mesothelioma). During this latent period the epigenetic marks are, however, present. If the interest in DNA methylation is solely as a biomarker, then the question of the origin and biological relevance of these changes is somewhat irrelevant. If the observed changes can be robustly validated and replicated, then their simple representation of a change in the sampled cell population may be adequate for their exploitation as a biomarker[72,149]. However, when changes are observed in purified cell populations, such subtle changes in methylation may give significant insight into underlying pathophysiological mechanisms. If we consider post-partum depression (PPD), pre-symptom onset epigenetic markers have been identified, potentially allowing the identification of susceptible women[150]. Although the epigenetic markers had a >80% predictive accuracy and have significant potential as PPD biomarkers, they also provide significant mechanistic insight into the pathophysiology of PPD. It has long been postulated that PPD is linked to the significant fluctuation in hormonal levels during pregnancy, and indeed, the epigenetic marks have all been linked to 17β oestradiol (E2). Although PPD has a range of previously identified biological and environmental risk factors it is unlikely to have a single underlying cause, and the methylation changes identified may represent a 'final common pathway[151] integrating many potential pathways. However, this highlights that differentially methylated regions are not exclusively biomarkers as they are often reported, but may provide significant insight into the underlying mechanisms.

Overall, we are forced to conclude that there is currently no accurate estimate of what represents a genuine, biologically relevant, change in methylation, and what may be ascribed to any of a multitude of external factors. Although it should be emphasised that all these outside factors contribute to measurable differences in the observed phenotype, and that small changes may represent a genuine biological difference.

1.6 Methylation: Single CpG or Clusters?[a]

It is becoming clear that, despite numerous reports of single CpGs associating with disease phenotypes, methylation levels are regulated in clusters. This has brought into question the functional effect of limited changes to the methylation level of single CpG dinucleotides[14,152]. Using the *NR3C1* as an example, that methylation over a region of ~45 consecutive CpGs within one of the many promoter regions efficiently silenced the associated transcripts[153]. However, methylation of smaller regions of ~125bp (~12CpGs) reduced promoter activity by 75%[102]. There is currently no evidence for the *NR3C1* that single CpG methylation has functional effects on gene expression[14]. Both individual[121] and promoter-wide[154] CpG methylation increases have been associated with clinical post-traumatic stress disorder (PTSD) symptoms. Our *NR3C1* methylation data concords with the latter observation, where a strong distance-dependent correlation throughout the *NR3C1* promoter was observed both in man[127,152] and rat [155], suggesting that for the *NR3C1*, methylation occurs in clusters over ~80bp. Similar results have been observed at the whole epigenome level as well as the population level[156,157]. Importantly, at the population level, methylation clusters appeared to behave in a manner similar to genetic variants with multiple clusters of methylation in "linkage-disequilibrium" covering distances up-to 300kbp[158].

1.7 The Glucocorticoid Receptor Gene: Transcription and Translation

A frequently studied gene in epigenetic studies[42,44,102,159] is glucocorticoid receptor (GR) gene or *NR3C1*, which underlies the stress-response through its regulation of the hypothalamic-pituitary-adrenal (HPA) axis[42,159]. The *NR3C1* gene consists of eight constant exons (exon 2-9) and nine untranslated, alternative first exons (1A-1J), each of them having an own promoter (Fig.3A)[160–162]. The generated pre-mRNA is spliced at the 3'UTR end, resulting in one of the main C-terminal transcriptional isoforms GR-α, GR-β or GR-P (Fig.3B). The first two are the pre-dominant forms, generated by the inclusion of either alternative last exon 9α or 9β, and containing the C-terminal ligand binding. GR-P on the other hand lacks both exon 8 and 9[127,153,163–165]. The *NR3C1*'s complex 5'UTR, allowing alternative transcription initiation, combined with alternative splicing of the 3'UTR adds to the diversity of *NR3C1* transcriptome. The resulting transcript heterogeneity has a significant influence/impact on post-transcriptional processes regulating gene expression, and subsequently helps fine-tuning the local GR levels, creating cell- and tissue-specific distribution patterns[160,162,166–170].

a. This section was published in Leenen, F. A. D., Muller, C. P. & Turner, J. D. DNA methylation: conducting the orchestra from exposure to phenotype? *Clin. Epigenetics* 8, (2016)

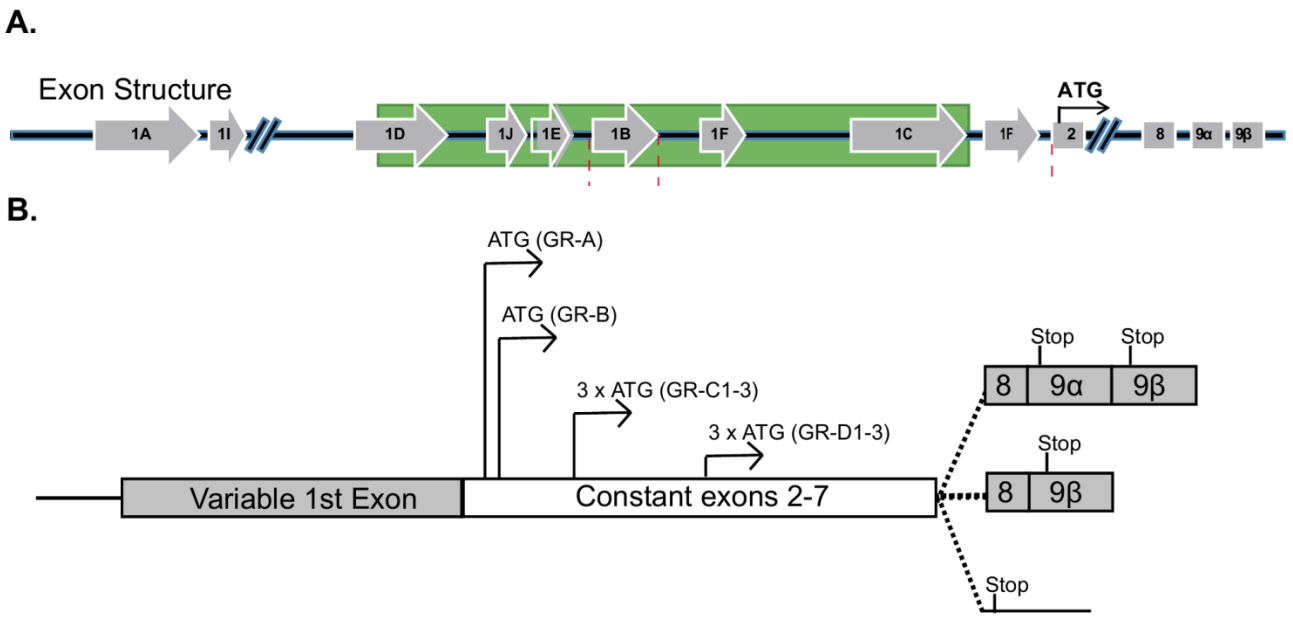

Figure 3: A schematic representation of the *NR3C1* structure and the internal ATG translation initiation codons resulting in different translational isoforms.[24]

Next to alternative transcription initiation and alternative splicing, which operate on a transcriptomic level, the protein isoform diversity is increased by the process of alternative translation initiation. The translation of the GR-α and GR-β transcripts can start at one of the multiple ATG codons within exon 2, hence generating an additional variety by producing N-terminal translational GR isoforms (GR-A, GR-B, GR-C1, GR-C2, GR-C3, GR-D1, GR-D2, GR-D3) (Fig. 3B). All N-terminal GR-α isoforms are functional, as they all contain the identical intact ligand-binding domain for glucocorticoids. Yet, they exhibit different transactivation activities, with each regulating a unique set of genes. Just like the transcript distribution, the translational isoforms show a cell- and tissue-specific pattern[164,165,171].

1.8 Thesis' Research Objectives

Within eukaryotes, gene expression is controlled by a multitude of mechanisms operating on both transcriptional as well as translational level, such as alternative transcription initiation, alternative RNA splicing, mRNA stability, alternative translation initiation etc. The external environment is suggested to affect gene expression and the involved processes through epigenetic mechanism such as DNA methylation, hydroxymethylation, chromatin condensation, ncRNAs and histone modifications. Hereby, re-shaping the transcriptional, translational and proteomic landscape and hence the resulting phenotype. Consequently altering an individual's phenotype, due to specific external environmental factors in time.

DNA methylation has long been thought of as a mechanism switching genes on or off through complete methylation or demethylation. Association studies introduced the idea of subtle changes resulting in

intermediary methylation levels that contributed to the phenotypical diversity. Despite the increase in association studies, the biological relevance of subtle shifts in DNA methylation levels remains elusive. In this thesis, we suggest that DNA methylation occurs in two clear paradigms. The classical paradigm where DNA methylation appears to be either in a hyper-or hypomethylated state with a clear on/off switch of genes. The second paradigm, on which this thesis is focused, is a subtle more complex process where small shifts in methylation level may cause a redistribution of the transcriptional landscape, affect the translational isoform production and re-orchestrate the final proteomic landscape. Therefore, the overall/global objective of this thesis was the examination of the physiological and functional relevance of such subtle small DNA methylation changes. This overall goal has been subdivided into the following smaller sub-objectives:

1. The examination of the role of DNA methylation on mRNA transcripts for model gene *NR3C1*. By adapting the recently introduced high-throughput sequencing (HTS) technique, the transcription initiation process' variability and complexity, and the impact of small DNA methylation could be studied more in greater depth. Previous studies suggested an impact of alternative first exon usage the translational isoform frequency and distribution. Therefore, a series of plasmid constructs, each containing an a first alternative exon variant that was either full exon length or a sequence starting within the exon until its end, were designed to elucidate the link between a highly permissive transcriptional variability, due to shifts in DNA methylation levels, and post-transcriptional regulation of the *NR3C1*.

Throughout the genome, many other genes possess a similar structure to *NR3C1*, hence suggesting the observations made for *NR3C1* could be expanded to other genes as well. To development of the hypothesis that the environment, via small subtle methylation changes, causes a transcriptional distribution that affects the translational isoform distribution and consequently the resulting proteomic landscape, on a genome-wide basis, necessitates suitable models, as outlines in objectives 2 and 3:

2. The second sub-objective consisted of analysing whether maternal deprivation (MD) is a suitable model to analyse the biological relevance of small methylation changes and whether this paradigm holds true for the central nervous system. In this study rats were subjected to MD, a model that mimics one of the most studied psychopathologies, depression, but of which the epigenetic alterations genome-wide are unknown. By combining the methylated DNA enrichment (MeDIP) with HTS, we aimed at detecting differentially methylated loci between groups on a genome-wide basis and link them to the resulting diversity in phenotypes.

3. As severe early life adversity comparable to the MD model is relatively rare in humans, the importance of small DNA methylation shifts was also analysed in a second more clinically and sociologically relevant model, early life infection. To that end, we used a mouse model, exposing animals to early life H1N1 infection and analysing how this modulated their response to viral re-exposure in adulthood. By employing Methyl-Seq, a reduced representation epigenome sequencing technique, we aimed at identifying a network of DNA methylation changes between treatment groups and link them to an altered immunological phenotype.

1.9 Thesis Outline

In **Chapter 1**, a broad notion of the context in which this thesis is embedded is thoroughly described. In **Chapter 2** we present a study which analysed the transcription variability of the *NR3C1*, reported in previous work, in greater depth by adapting the classical RNA ligase-mediated rapid amplification of 5' cDNA ends (5'-RACE) to High-Throughput Sequencing (HTS) identifying the important role it plays in determining the overall protein population. **Chapter 3** reports one of the first studies combining genome-wide methylation and hydroxymethylation profiling after subjecting rats to either MD or handling stress. Presenting a model for the subtle methylation change paradigm in the CNS for future mechanistic studies. **Chapter 4** presents the results of a study analysing the effects of early life infection, with influenza virus H1N1, on the genome-wide methylation profile in mice. The early life infection model, which is a clinically and sociologically more relevant model, assesses whether the subtle methylation paradigm holds true for the immune system. In **Chapter 5**, the major findings are placed into the contemporary context and future perspectives discussed.

Chapter 2

Where does transcription start? 5'-RACE adapted to Next-Generation Sequencing

This chapter has been published as:

Where does transcription start? 5'-RACE adapted to Next-Generation Sequencing

Fleur A. D. Leenen[1,2], Sara Vernocchi[1,2], Oliver E. Hunewald[1], Stephanie Schmitz[1], Anne M. Molitor[1], Claude P. Muller[1,2] and Jonathan D. Turner[1]

[1] Department of Infection and Immunity, Luxembourg Institute of Health, Esch-Sur-Alzette, L-4354, Grand-Duchy of Luxembourg
[2] Department of Immunology, Research Institute of Psychobiology, University of Trier, Trier, D-54290, Germany

Nucleic Acids Res. 2016 Apr 7; 44(6): 2628–2645

doi: 10.1093/nar/gkv1328

PMID: 26615195; PMCID: PMC4824077

2.1. Abstract

The variability and complexity of the transcription initiation process was examined by adapting RNA ligase-mediated rapid amplification of 5' cDNA ends (5'-RACE) to HTS. We oligo-labelled 5'-m^{7}G-capped mRNA from two genes, the simple mono-exonic *Beta-2-Adrenoceptor (ADRB2R)* and the complex multi-exonic *Glucocorticoid Receptor (GR, NR3C1),* and detected a variability in TSS location that has received little attention up to now. Transcription was not initiated at a fixed TSS, but from loci of 4 to 10 adjacent nucleotides. Individual TSSs had frequencies from <0.001% to 38.5% of the total gene-specific 5'm^{7}G-capped transcripts. *ADRB2R* used a single locus consisting of 4 adjacent TSSs. Unstimulated, the *NR3C1* used a total of 358 TSSs distributed throughout 38 loci, that were principally in the 5'UTRs and were spliced using established donor and acceptor sites. Complete demethylation of the epigenetically sensitive *NR3C1* promoter with 5-azacytidine induced 1 new locus and 127 TSSs, 12 of which were unique. We induced *NR3C1* transcription with dexamethasone and Interferon-ɣ, adding 1 new locus and 185 additional TSSs distributed throughout the promoter region. In-vitro the TSS microvariability regulated mRNA translation efficiency and the relative abundance of the different *NR3C1* N-terminal protein isoform levels.

2.2. Introduction

The genome does not only encode mRNA and protein sequences but it contains also the temporal, spatial and quantitative instructions for their expression. This elaborate regulation occurs principally at the transcriptional level, determining both gene expression and transcript diversity. In the simplest case, transcription is initiated from a transcription start site (TSS) after completing the assembly of the competent transcription initiation complex on the associated promoter. Many genes possess a 5' UTR containing multiple alternative first exons, each with its own alternative promoter as a second level of transcriptional complexity. It has been estimated that 58% of the transcribed genes had multiple promoters [172]. The 5' UTR's influence gene expression in a cell- and tissue-specific manner by generating transcriptional variability, i.e. different mRNA variants [167–169,173]. Whilst some alternative 5' UTR first exons may be similar in length and nucleotide (nt) sequence, e.g. the *Pcdh* and *UGT1* gene clusters [170], most alternative 5' UTR first exons differ in length and sequence. These complex 5'UTRs evolved through processes such as gene duplication by recombination, retroposition, intronic deletions, etc. [174–177]. Both alternative splicing and alternative transcription initiation are closely linked and give rise to high complex and diverse transcriptomes and proteomes [164,168,178–181]. Coding 5' UTR first exons generate different mRNA transcript variants and protein isoforms. Although non-coding first exons do not generate protein diversity, they create transcript variability that has significant impact on post-transcriptional gene regulation, including translational efficiency, mRNA processing, stability and export [167,169,170,182,183].

In eukaryotes most promoters are located within CpG-rich regions, whilst conserved, well defined TATA box based promoters are less frequent [5,172]. Ubiquitously expressed genes are primarily associated with CpG islands and variable TSSs, whereas tightly regulated transcripts have TATA box promoters and well-defined TSSs [172]. There is now limited evidence that, irrespective of their location, the site at which transcription is initiated may be variable [172]. This was observed as a series of TSSs over a very small 4-6bp region surrounding the principal TSSs [172].

To further investigate the variability of the transcription start sites, two genes with distinct structures and expression profiles were selected. The *Beta-2 Adrenoceptor* (*ADRB2R;* OMIM 109690), is an intronless single exon gene (Fig. 4A and 5A), with no previously identified transcriptional variability, and a uniform, ubiquitous expression according to the literature [184]. In comparison, the human *glucocorticoid receptor* gene *(GR*; *NR3C1*, OMIM +138040), located within chromosome 5, has a complex 5' structure and a highly variable and tightly regulated, but ubiquitous expression [184]. The *NR3C1* comprises 9 untranslated, alternatively spliced first exons (exon 1A – H) and eight translated exons (exon 2-9), with the translation start site located within exon 2 (Fig.

4B and 5B). All alternative first exons have their own promoter region covering both a CpG island and a distal TATA-like promoter [127,153,160–163,184,185]. They are located either in the distal or the proximal promoter region, 30kb (1A and 1I) and 5kb (1D to 1J) upstream of the translation start site respectively. The latter are contained in a highly conserved 3kb CpG island [127,153,160–163,185]. Regulation of *NR3C1* transcription has been extensively studied. At least 29 transcription factor binding sites have been experimentally confirmed, controlling first exon usage [163]. Additionally, the CpG island promoters were shown to be susceptible to methylation, linking expression levels to the environment, fine-tuning *NR3C1* levels [27,127,153,163].

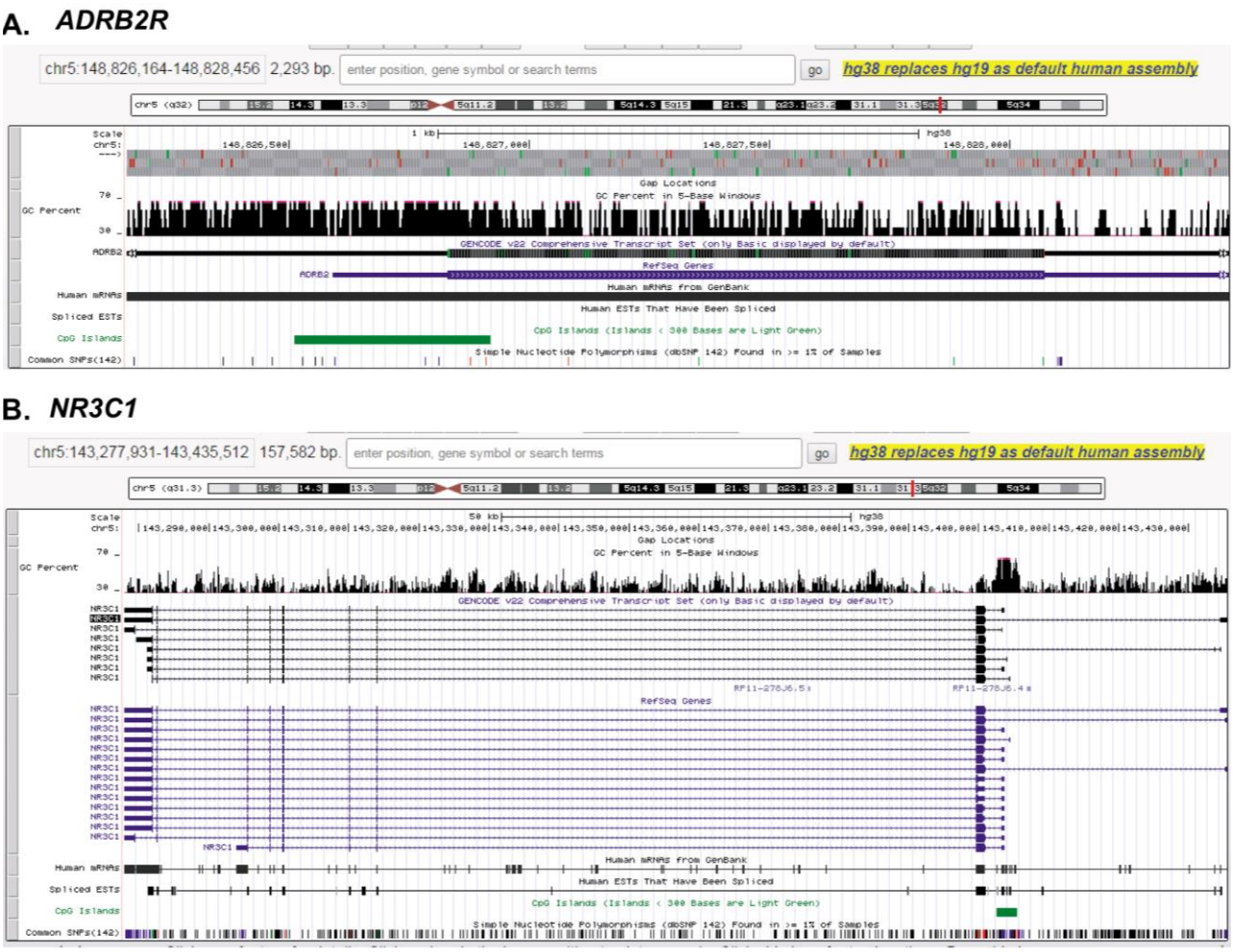

Figure 4: (A) UCSC browser view of *ADRB2R* gene. (B) UCSC browser view of *NR3C1* gene.[166] (Supplementary Data)

By adapting the classical RNA ligase-mediated rapid amplification of 5' cDNA ends (5'-RACE) to Next-Generation Sequencing (NGS) we were able to study the variability and complexity of the *NR3C1* transcription initiation process in greater depth, identifying transcriptional initiation loci that themselves contain many, often adjacent, unique TSSs. The experimental protocol was designed to exclude any other potential interpretation of the HTS results and to minimise any potential ligation bias. The RNA oligo ligation strategy, employing TAP

and CIP treatments, labelled only mature, undegraded mRNAs. By using a common RNA or DNA oligo, potential ligation bias between the sequences was reduced. On top of the experimental precautions, a 0.1% cut-off was introduced to define genuine TSSs, ensuring that errors introduced during the sample preparations are minimal and below our cut-off. The simple mono-exonic gene *ADRB2R* was used as a control. *ADRB2R* showed little transcriptional variability. Our data expands the unique literature TSS to one unique locus consisting of 4 adjacent TSSs. The multi-exonic *NR3C1* gene on the other hand targeted a total of 358 TSSs throughout 38 loci that were cell line and stimuli specific. This microvariability around individual loci was negatively associated with translational efficiency and controlled the relative abundance of *NR3C1* translational isoforms. Although this combination of techniques was initially intend to investigate the particular case of the *GR*, we suggest that our observations can be extended to other genes**.**

2.3. Material and Methods

2.3.1. Cell Culture and RNA Extraction

Three cell lines, DAUDI, MCF-7 and A549, were cultured as previously described [153,186,187]. All culture media were from Lonza (Verviers, Belgium). T cells were isolated from PBMC's by Ficoll-isopaque (GE Healthcare Life Sciences, Amersham, UK) gradient centrifugation and by positive magnetic selection (Miltenyi midiMacs, Miltenyi Biotech GmbH, Cologne, Germany) [162].

Total RNA was extracted from unstimulated DAUDI, MCF-7 and A549 cells, following routine passage using the RNeasy Mini Kit (QIAGEN, Venlo, Netherlands) according to the manufacturer's instructions. 1.5 x 10^7 DAUDI cells were seeded in a $75cm^3$ flask. When they reached 70% confluence, they were stimulated with Interferon ɣ (IFN-ɣ; 6h, 5ng/ml), Dexamethasone (Dex; 6 h, 100nM) or 5-AZA-2'-deoxycytidine (AZA; 72h, 10µM) (Sigma-Aldrich, Diegem, Belgium). Cells were detached from the culture support using trypsin-EDTA (Lonza) and pelleted (5 min, 1 671.6 x g). Subsequently, total RNA was isolated using the RNeasy Mini Kit (QIAGEN). RNA integrity was assessed using the Eukaryote Total RNA Nano assay with a RNA 6000 Nano chip on the Aligent 2100 Bioanalyzer (Aligent Technologies, Diegem, Belgium). The RNA quality assessment was based on the RNA integrity number (RIN). Only samples with a RIN value of >7 were used for further experiments.

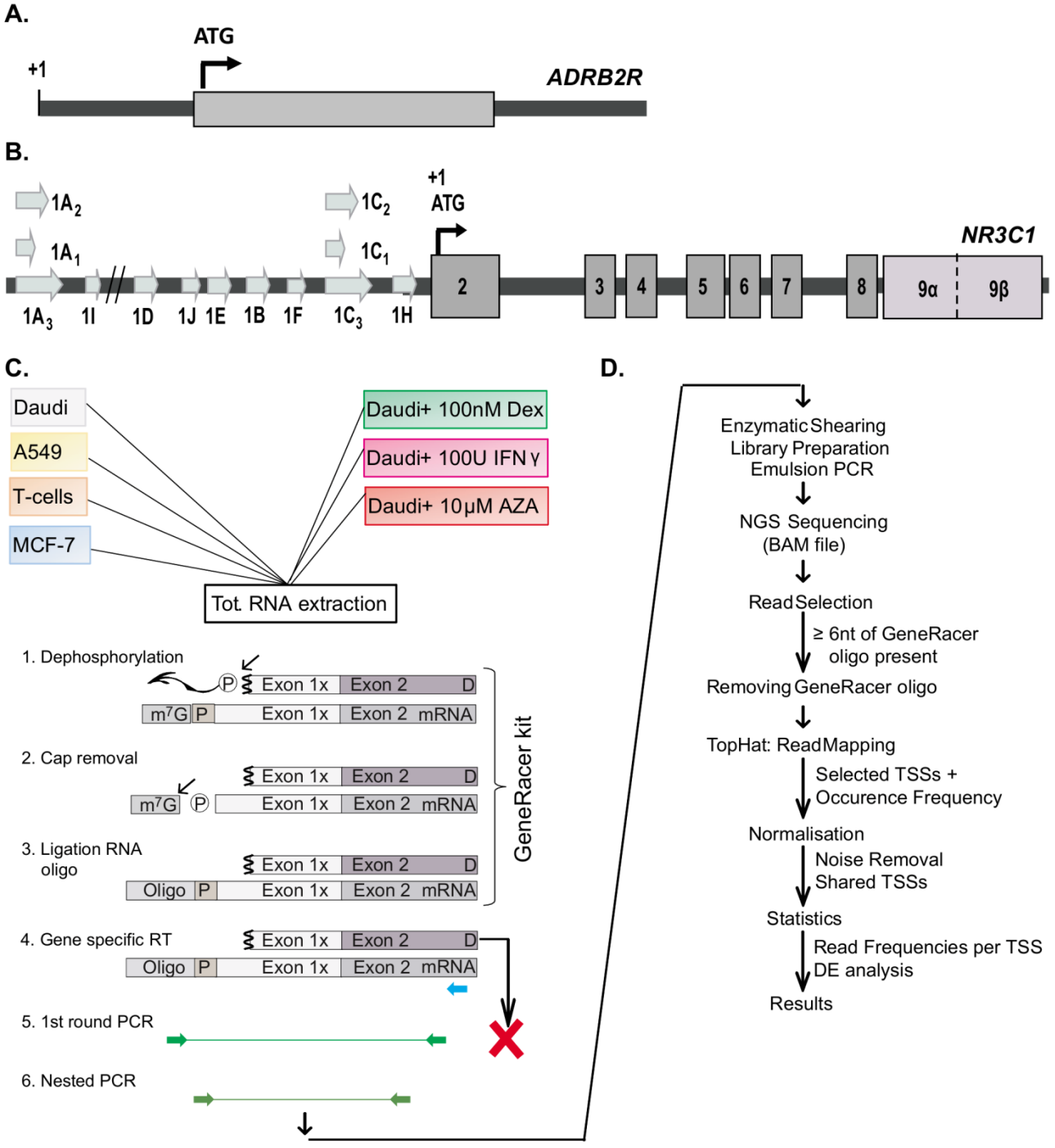

Figure 5: A schematic representation of the ADRB2R and *NR3C1* gene structure and the 5'-RACE-Sequencing workflow. (A) The ADRB2R gene. Nucleotides are numbered with respect to the NCBI reference sequence (NM_000024.5). (B) A schematic representation of the *NR3C1* gene, showing the first exons ( ) in the distal and proximal (CpG island) promoter; the seven common exons ( ); and the two alternative 3' coding exons ( ). Nucleotides are numbered with respect to the ATG translation initiation codon (+1) (C) The experimental workflow from RNA extraction, through the 5'-RACE protocol in order to label the TSSs, the HTS library preparation and the actual sequencing. D, uncapped RNA sequence; M7G mRNA specific 7-methylguanosine cap; P active phosphate; P phosphate diesterbond; oligo generacer specific oligo. (D) The data analysis workflow including read selection, quality control, mapping (TopHat), normalisation, reproducibility and differential expression analysis.[166]

2.3.2. RACE-PCR

To label exclusively the 5' nt of the mRNA, an RNA oligo was ligated in place of the mature mRNA-specific m^7G cap structure, as previously described using the reagents from Invitrogen (Life Technologies, Paisley, UK) [188–190]. Briefly, RNA was treated with calf intestinal phosphatase (CIP) to remove all active 5' mono-phosphates from truncated or otherwise degraded mRNA as well as other RNA's, rendering those sequences unavailable for ligation and leaving only intact capped mRNA unaffected. Subsequently, tobacco acid pyrophosphatase (TAP) was used to remove the 5' cap structure leaving a unique active 5' phosphate on mature mRNA by hydrolysing the pyrophosphate bonds on the m^7G cap triphosphate bridge. An RNA oligo (5′-CGACUGGAGCACGAGGACACUGACAUGGACUGAAGGAGUAGA AA-3′) was ligated to the unique active 5' phosphate using a T4 RNA ligase in a 10µl reaction containing 2ng dephosphorylated, decapped RNA and 0.25µg RNA Oligo, 10X Ligase Buffer, 10mM ATP, 40U/µl RNaseOut and 5U/µl T4 RNA ligase. First-strand cDNA was obtained by reverse transcription of the ligated mRNA using a *NR3C1*-specific primer in exon 2 (5'-CAGTGGATGCTGAACTCTTGG-3', Eurogentec, Seraing, Belgium) or dN6 random hexamer primers (Invitrogen, Life Technologies), for the control gene (*ADRB2R*). Two rounds of PCR amplification were performed with forward primers located within the RNA oligo and reverse primers in exon 2 of the *NR3C1* (Table 2). The control gene was similarly amplified using the same RNA oligo specific forward primers and two *ADRB2R* specific reverse primers (Table 2). The reverse primers were located respectively 29bp downstream and 96bp upstream of the ATG translation start codon. Amplification was performed in 25µl reactions containing 20mM Tris-HCl, 50mM KCl, 2mM $MgCl_2$, 200µM dNTPs, primers (Eurogentec), 1x SYBR green and 1.5U Platinum Taq Polymerase (Life Technologies). Thermal cycling (CFX96, BioRad, Hercules, CA, USA) conditions were 95°C, 2min; 45 cycles of 95°C 20s, Ta 20s, 72°C 90s; and a final elongation step at 72°C for 10min. Nested PCRs were performed using a 1:100 dilution of the 1st round PCR product as a template. Prior to Next Generation Sequencing library preparation, the PCR products were purified using Agent AmPure XP Beads (Analis, Suarlée, Belgium) and quantified with the Quant-iT picogreen dsDNA Assay Kit (Life Technologies) according to the manufacturers' instructions.

Table 2: PCR primers[166] (Supplementary Data)

Primer name	Sequence
GeneRacer 5' primer_fwd	5'-GACTGGAGCACGAGGACACTGA-3'
Exon 2_1st PCR deep seq_rev	5'-GGAACACTGGTCGACCTATTGAGGT-3'
ADRB2R_1st PCR_deep seq_rev	5'-CTTCCATTGGGTGCCAGCAAG-3'
GeneRacer 5' nested_fwd	5'-GGACACTGACATGGACTGAAGGAGTA-3'
Exon 2_2nd PCR deep seq_rev	5'-GGAGTCTGATTGAGAAGCGACAGC-3'
ADRB2R_2nd PCR_deep seq_rev	5'-CTCATTCAGCGGCTGTGGTG-3'

2.3.3. High-Throughput Sequencing (HTS)

The HTS libraries were prepared using the Ion Xpress Plus Fragment Library Kit (Rev. A, Life Technologies) according to the manufacturer's instructions for 100ng gDNA. Briefly, purified PCR products were sheared (Ion Shear Plus Kit, Life Technologies) and Ion Xpress Barcode Adapters (Life Technologies) were ligated to the resulting DNA strands. Using 2% agarose gels (E-gel System, Life Technologies), DNA fragments of 200-350nt length were selected for further amplification. The DNA concentration was estimated with the Agilent High Sensitivity DNA chip on the Agilent Bioanalyzer (Agilent Technologies) and equimolar quantities of each library were pooled (*NR3C1* gene: 4 libraries; *ADRB2R* gene: 12 libraries).

Template preparation was carried out using the Ion OneTouch 200 System Template Kit v2 protocol (Rev. 4, Life Technologies) recommended by the manufacturer. Briefly, a diluted library pool was added to the emulsion mix for DNA clonal amplification on the Ion OneTouch instrument, followed by an enrichment of the template-positive Ion Sphere Particles (ISP) on the Ion OneTouch ES instrument (Life Technologies). To assess the quality and calculate the appropriate library dilution, both unenriched and enriched ISP samples were qualified and quantified performing the Qubit™ dsDNA HS Assay Kits (Life Technologies) using a Qubit 2.0 Fluorimeter (Life Technologies) according to the manufacturer's protocol.

Ion Torrent PGM runs were performed using the Ion PGM 200 Sequencing Kit (PN4474246 Rev. D, Life Technologies) on Ion 314 and 316 Chips (Life Technologies), as simplex and multiplexed runs, with the standard Torrent Suite parameters (Fig. 6).

2.3.4. Bio-Informatics and Statistical Analyses

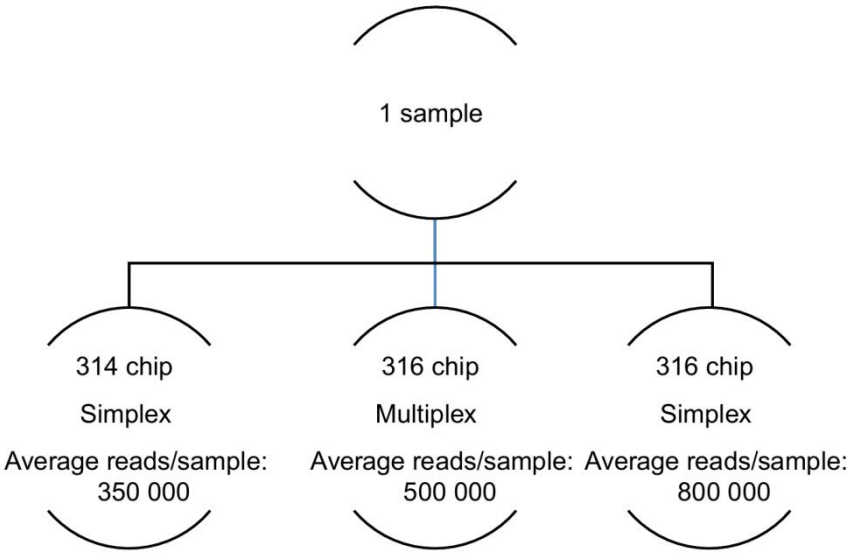

Figure 6: Over multiplexing strategy used with Ion Torrent 314 and 316 chips.[166] (Supplementary Data)

HTS sequencing reads were processed using the default Torrent Suite settings. Sequences containing at least the last 8 nt of the 5' cap oligo after Multiplex Identifier (MID) sorting were retained as oligo-labelled TSSs for further analysis, all non-labelled sequences were discarded. This 5' cap oligo sequence was subsequently trimmed and the reads were mapped against the genomic reference for the *NR3C1* gene (Chr5, hg19, 142 657 496 to 142 850 254) with TopHat software [191,192] (v2.0.3), using default settings. A python script retrieved the TSS for each aligned read. Oligo-labelled TSS reads were also analysed manually with Geneious software (Biomatters, v5.5.6). Throughout this study, *NR3C1* TSSs are annotated with respect to the ATG (+1) translation initiation codon and the *ADRB2R* TSSs with respect to the first nt in the mRNA sequence (NM_000024.5).

TSS count data were normalised using the 'Trimmed Mean of the M-values' (TMM) technique in R (R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/ v3.0.2[193]) using the Bioconductor package NOISeq [194] (v2.6.0). To remove background noise, all TSSs corresponding to a sequence frequency below 0.1% in replicate runs were removed. Data were visualised in Bioconductor packages limma [195] (v3.18.13) and affycoretools (MacDonald, J.W. (2008). affycoretools: Functions useful for those doing repetitive analyses with Affymetrix GeneChips. R package version 1.34.0[196]). Differential expression analysis was performed with Bioconductor package NOISeq [194] (v2.6.0). Differentially expressed TSSs were hierarchically clustered and visualised using CRAN-packages cluster (Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K. (2014) cluster: Cluster Analysis Basics and Extensions. R package version 1.15.2[197]) and pheatmap (Raivo, K. (2013). pheatmap: Pretty Heatmaps. R package version 0.7.7.[198]).

The evolutionary conservation was visualised with the UCSC browser (https://genome-euro.ucsc.edu). Differentially expressed TSS frequencies were plotted against the PHAST phyloP conservation score from the publically available 100 vertebrate genome alignment [199]. In silico Phylogenetic footprints (ISPF) were obtained from a previous report [200].

2.3.5. Translational Efficiency of Transcriptional Micro-Variants

The previously reported full length exons and the CMV promoter were cloned into the synthetic firefly luciferase pGL 4.10 vector (Fig. 7) (Promega, Leiden, Netherlands) [164]. Shorter '5 microvariants of the exons, identified by HTS, were synthesised and also inserted into pGL4.10 (GeneCust, Dudelange, Luxembourg) (Table 3).

Twenty-four hours prior to transfection A549 cells were seeded into 24-well plates ($4x10^4$ cells/well). Cells were transfected with 750ng of 5' UTR constructs, using 0.5µl of PLUS Reagent and 2.0µl of Lipofectamine LTX (Life Technologies) according to the manufacturer's instructions. The *Renilla* luciferase plasmid pGL 4.73 was used as control vector and cells were transfected using a 10:1 ratio of the two plasmids.

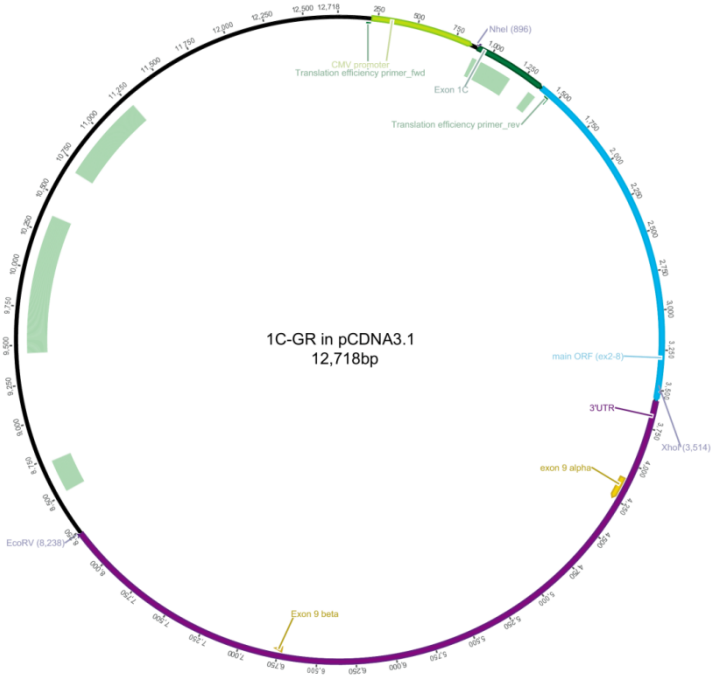

Figure 7: The plasmid 1C GRpcDNA3.1, into which all inserts were cloned.[166] (Supplementary Data)

The firefly and *Renilla* luciferase activity were measured using the Dual-Glo Luciferase assay system (Promega) according to the manufacturer's protocol. The luminescent signal was read with Infinite M200 plate reader (TECAN, Männedorf, Switzerland). The experiments were performed in biological triplicates. Within exon variants, the luminescent signals were subjected to a pairwise multiple comparison using Kruskal-Wallis One Way Analysis of Variance on Ranks with a Tukey post-hoc correction, reporting q-values for a type I error level of 0.05 per comparison.

Table 3: Plasmid variants for translational efficiency analysis[166] (Supplementary Data)

Plasmid Name	Length Insert (nt)	Sequence Insert (5' – 3')
CMV-1A3	94	AGTTGTACCTTAATAACAGGAATTTTCATCTGCCTGGCTCCTTTCCTCAAAGAACAAAGAA GACTTTGCTTCATTAAAGTGTCTGAGAAGGAAG

CMV-1A3	160	AATGAGTGCCTTCTCTGTGCGAGAATGGGGAGGAACAAAATGCAGCTCCTACCCTCCTCG GGCTTTAGTTGTACCTTAATAACAGGAATTTTCATCTGCCTGGCTCCTTTCCTCAAAGAAC AAAGAAGACTTTGCTTCATTAAAGTGTCTGAGAAGGAAG
CMV-1A3	164	ACTTAATGAGTGCCTTCTCTGTGCGAGAATGGGGAGGAACAAAATGCAGCTCCTACCCTC CTCGGGCTTTAGTTGTACCTTAATAACAGGAATTTTCATCTGCCTGGCTCCTTTCCTCAAA GAACAAAGAAGACTTTGCTTCATTAAAGTGTCTGAGAAGGAAG
CMV-1A3	180	ATTAACTTTGATAAGCACTTAATGAGTGCCTTCTCTGTGCGAGAATGGGGAGGAACAAAAT GCAGCTCCTACCCTCCTCGGGCTTTAGTTGTACCTTAATAACAGGAATTTTCATCTGCCTG GCTCCTTTCCTCAAAGAACAAAGAAGACTTTGCTTCATTAAAGTGTCTGAGAAGGAAG
CMV-1A3	981	AGGTTATGTAAGGGTTTGCTTTCACCCCATTCAAAAGGTACCTCTTCCTCTTCTCTTGCTC CCTCTCGCCCTCATTCTTGTGCCTATGCAGACATTTGAGTAGAGGCGAATCACTTTCACTT CTGCTGGGGAAATTGCAACACGCTTCTTTAAATGGCAGAGAGAAGGAGAAAACTTAGATC TTCTGATACCAAATCACTGGACCTTAGAAGGTCAGAAATCTTTCAAGCCCTGCAGGACCG TAAAATGCGCATGTGTCCAACGGAAGCACTGGGGCATGAGTGGGGAAGGAATAGAAACA GAAAGAGGGTAAGAGAAGAAAAAAGGGAAAGTGGTGAAGGCAGGGAGGAAAATTGCTTA GTGTGAATATGCACGCATTCATTTAGTTTTCAAATCCTTGTTGAGCATGATAAAATTCCCAG CATCAGACCTCACATGTTGGTTTCCATTAGGATCTGCCTGGGGGAATATCTGCTGAATCA GTGGCTCTGAGCTGAACTAGGAAATTCACCATAATTAGGAGAGTCACTGTATTTCTCTCCA AAAAAAAAAAAGTTATACCCGAGAGACAGGATCTTCTGATCTGAAATTTTCTTCACTTCTGA AATTCTCTGGTTTGTGCTCATCGTTGGTAGCTATTTGTTCATCAAGAGTTGTGTAGCTGGC TTCTTCTGAAAAAAGGAATCTGCGTCATATCTAAGTCAGATTTCATTCTGGTGCTCTCAGA GCAGTTAGCCCAGGAAAGGGGCCAGCTTCTGTGACGACTGCTGCAGAGGCAGGTGCAGT TTGTGTGCCACAGATATTAACTTTGATAAGCACTTAATGAGTGCCTTCTCTGTGCGAGAAT GGGGAGGAACAAAATGCAGCTCCTACCCTCCTCGGGCTTTAGTTGTACCTTAATAACAGG AATTTTCATCTGCCTGGCTCCTTTCCTCAAAGAACAAAGAAGACTTTGCTTCATTAAAGTGT CTGAGAAGGAAG
CMV-1B	53	AGATGATGCGGTGGTGGGGGACCTGCCGGCACGCGACTCCCCCCGGGCCCAAA
CMV-1B	73	AGCTGAAGACCCGGCCGCCCAGATGATGCGGTGGTGGGGGACCTGCCGGCACGCGACT CCCCCCGGGCCCAAA
CMV-1B	107	AACTTCTCTCCCAGTGCGAGAGCGCGGCGGCGGCAGCTGAAGACCCGGCCGCCCAGAT GATGCGGTGGTGGGGGACCTGCCGGCACGCGACTCCCCCCGGGCCCAAA
CMV-1B	105	CTTCTCTCCCAGTGCGAGAGCGCGGCGGCGGCAGCTGAAGACCCGGCCGCCCAGATGA TGCGGTGGTGGGGGACCTGCCGGCACGCGACTCCCCCCGGGCCCAAA
CMV-1C	71	AAGCTAAGTTGTTTATCTCGGCTGCGGCGGGAACTGCGGACGGTGGCGGGCGAGCGGC TCCTCTGCCAGAG
CMV-1C	73	ACAAGCTAAGTTGTTTATCTCGGCTGCGGCGGGAACTGCGGACGGTGGCGGGCGAGCG GCTCCTCTGCCAGAG
CMV-1C	101	ATATTTCCCTCCTGCTCCTTCTGCGTTCACAAGCTAAGTTGTTTATCTCGGCTGCGGCGG GAACTGCGGACGGTGGCGGGCGAGCGGCTCCTCTGCCAGAG
CMV-1C	479	GGCGCCGCCTCCACCCGCTCCCCGCTCGGTCCCGCTCGCTCGCCCAGGCCGGGCTGCC CTTTCGCGTGTCCGCGCTCTCTTCCCTCCGCCGCCGCCTCCTCCATTTTGCGAGCTCGTG TCTGTGACGGGAGCCCGAGTCACCGCCTGCCCGTCGGGGACGGATTCTGTGGGTGGAA GGAGACGCCGCAGCCGGAGCGGCCGAAGCAGCTGGGACCGGGACGGGGCACGCGCGC CCGGAACCTCGACCCGCGGAGCCCGGCGCGGGGCGGAGGGCTGGCTTGTCAGCTGGG CAATGGGAGACTTTCTTAAATAGGGGCTCTCCCCCCACCCATGGAGAAAGGGGCGGCTG TTTACTTCCTTTTTTTAGAAAAAAAAAATATATTTCCCTCCTGCTCCTTCTGCGTTCACAAG

	CTAAGTTGTTTATCTCGGCTGCGGCGGGAACTGCGGACGGTGGCGGGCGAGCGGCTCC TCTGCCAGAG

2.3.6. RNA Structure Prediction

The free energy released on RNA folding (ΔG) and the resultant secondary structure of the complete *NR3C1* transcripts and of the individual 5′UTRs were calculated using the online RNAfold algorithm (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi)[201]. Default settings were used for all predictions.

2.3.7. Minigene Design & Plasmid Construction

The plasmids were constructed as previously described [164]. Briefly, all constructs contained a first exon variant 1A3, 1B or 1C (Table 3), followed by exon 2 till 8 (NM_000176.2; nt 480-2 673) and the genomic sequence of exon 9α and the corresponding introns (NCBI36/hg18 release March 2006, chr5: 142 637 665 – 142 642 326). A total of 10 different length constructs were prepared. All inserts were synthesized (Genecust, Dudelange, Luxembourg) and subsequently cloned into pcDNA3.1 (-) (Life Technologies, Merelbeke, Belgium).

2.3.8. Western Blot

GR protein isoform quantification was performed as previously described [164]. Briefly, A549 cells were transfected with 500ng DNA using Lipofectamine LTX (Life Technologies) twenty-four hours post-seeding. Total proteins were extracted forty-eight hours post transfection, separated on 4-12% Bis-Tris ZOOM™ gels (Life Technologies) and immunoblotted with primary rabbit-α-GR antibody (P20 clone , epitope within aa 720-769 of the hGRα; Santa Cruz Biotechnologies, Heidelberg, Germany) and Cy5-labelled secondary antibody (GE Healthcare). After washing, mouse anti-β actin (anti-β-actin, Santa Cruz Biotechnologies) probing and secondary goat-anti-mouse Cy3 antibody (GE Healthcare) incubation, the immunoreactive bands on the membrane were read using the Typhoon 9400 imager (GE Healthcare) at excitation wavelength 633nm and 532nm for Cy5 and Cy3 respectively (PMT= 480V; scanning resolution= 50 µm). Band intensities were quantified using ImageJ (NIH, Bethesda, MD, USA) and normalized according to β-actin. Variance was determined with One-Way or Two-Way ANOVAs (Sigmaplot 12.3), pairwise comparisons were performed with the Student-Newman-Keuls test and $p<0,05$ were considered significant.

2.4. Results

2.4.1. 5’-RACE Library Sequencing

Library Preparation and Sequencing Quality To investigate the variability in TSS usage of the *ADRB2R* and *NR3C1* gene, RNA ligase mediated rapid amplification of 5’ cDNA ends (5’-RACE) was adapted to massively

parallel sequencing as outlined in Figure 5C and 5D. Ion Sphere loading densities (ISP) of ~68% were observed, corresponding to > 7.4 x 10^5 reads and > 4.4 x 10^6 reads for *ABRB2R* libraries on 314 and 316 chips respectively. Removing polyclonal reads reduced the numbers of reads to 3.6 x 10^5 and 3.2 x 10^6 reads respectively. *NR3C1* libraries had ISP densities of ~60% with > 3.7 x 10^6 reads per 316 chip. These were reduced to 2.3 x 10^6 after removal of ~23-24% polyclonal reads. Sequence quality, assessed by PHRED values, for both genes was unaffected by the 5'-RACE library preparation. PHRED values were >25 and fell below the acceptable quality threshold of a PHRED score of 20 only after 200nt, as expected from the Ion PGM 200 Sequencing Kit. In all multiplex *ADRB2R* and *NR3C1* sequencing data sets, reads were equally distributed over the different MIDs. To identify m^7G-capped TSSs, only 5' oligo-labelled sequences were analysed. Oligo selection and trimming produced data sets of 170000 to 800000 reads and 193727 to 378859 reads for the GR and *ADRB2R* gene that were retained as labelled TSSs. Because of the shearing step in the library preparation, on average only 40% of these reads (89 000 to 350 000 reads) contained a labelled 5' TSS. However, 97% (186608 to 366431) of these 5' TSS reads were successfully mapped against the GR or the *ADRB2R* gene region. Overall, the aligned reads corresponded to respectively 22% to 65% and 69% to 95% of the initial reads per sample. All raw and aligned sequencing data are available on the European Nucleotide Archive (ENA) of the EMBL-EBI under accession number PRJEB9064.

GC Content Does not Influence Sequencing. The *ADRB2R* and the *NR3C1* differ significantly in their 5' G+C content. This resulted in a somewhat lower loading efficiency (range 52% to 68%) for the *NR3C1* with the higher G+C content than the *ADRB2R* (range 59% to 74%) with the lower G+C content. The sequence quality of both genes was similar, with only slightly higher PHRED values for *ADRB2R* (>28) than for *NR3C1* (>25). Thus the difference in CG content did not seem to affect the sequencing quality.

Sequencing Artefacts do not Perturb TSS Identification. Analysis of the HTS reads revealed no substitutions in either the TSSs of the *ADRB2R* gene (0.187 x 10^6 sequences/21.9 x 10^6 total nt sequenced) or the *NR3C1* gene (0.258 x 10^6 sequences/50.2 x 10^6 total nt sequenced). As expected for Ion Torrent sequencing, however, the rate of insertions and deletions (indels) was high. We observed a 6% indel rate (0.05 insertions per total nt sequenced and 0.005 deletions per total nt sequenced) for *ADRB2R* and 2.20% for *NR3C1* (0.007 insertions per total nt sequenced and 0.004 deletions per total nt sequenced). Importantly, none of these indels were observed in the TSS region, i.e. in the 3' end of the RNA oligo or in the nt immediately downstream of the TSS. The TSSs were also checked for homopolymers. Only one of the 21 most important differentially expressed TSSs identified in the *NR3C1* was part of a two nt homopolymer. The *ADRB2R* gene had no homopolymer in its first exon region. Thus neither substitution, nor indels or homopolymers compromised TSS identification.

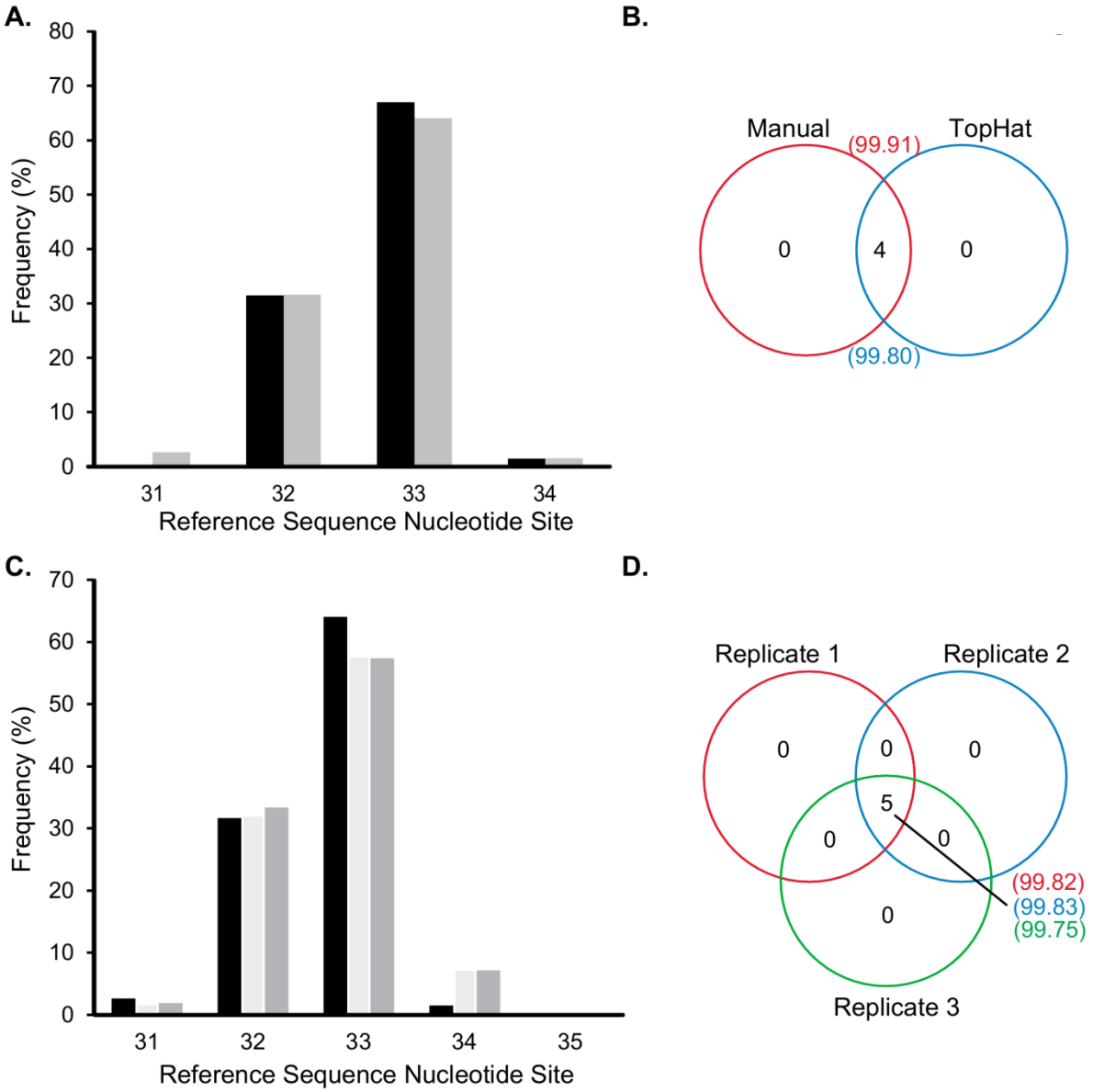

Figure 8: The mono-exonic ADRB2R shows transcriptional microvariability around one locus. (A) Frequency of reads aligned to individual TSSs using TopHat ( ) and a manual approach ( ). (B) The number of TSSs shared between data sets aligned using TopHat and the manual approach. Numbers in brackets are the frequency (%). (D) Bar plot representing the frequency (%) of reads aligned per TSS for three biological replicates. Replicate 1 ( ); Replicate 2 ( ); Replicate 3 ( ) (E) The number of TSSs shared between the biological replicates and the frequency of reads (%) per HTS run corresponding to these TSSs.[166]

The Alignment Method does not affect the TSS Identification. Since the TopHat alignment algorithm is relatively strict, sequences with sequencing errors and indels may have been excluded from the aligned data sets. Therefore, the analysis was repeated using Geneious with less restrictive parameters and the alignment of each sequence was manually verified. The Geneious and TopHat alignments, both mapped similar read numbers (309 739 sequences, 99.76% versus 308 341 sequences, 99.31%) against the *ABRB2R* reference sequence and both approaches identified the same TSSs with virtually identical frequencies (Fig. 8A-B). Both methods selected 7 TSSs: two thirds (64% to 67%) of the reads started at nt position 33, irrespective of the alignment method, One third (range 31.40% to 31.64%) of the reads started at nt position 32. Read numbers

starting at nt 31 and 34 accounted for less than 3%. Percentages of reads with TSSs at nt 30, 35 and 39 were <0.10%, and may represent the error rate in TSS identification. Therefore, a cut-off of <0.10% was applied for all subsequent runs to validate TSSs. As there was virtually no difference in mapping results between the 2 methods, TopHat was used in all subsequent analyses.

Sequencing Depth Requirements. In addition, to determine the read depth necessary to identify valid TSSs, technical triplicates of the *NR3C1* in DAUDI cells were performed on 314 and 316 chips, either in simplex or multiplex format, with similar loading and mapping values (Fig. 6). Despite generating different read numbers for the different HTS conditions, the TSSs pattern between the technical replicates was conserved (Fig. 6and 9). On average 97% of the sequences per replicate run target one of the 103 shared TSSs. Fewer than 4 TSSs were common to only 2 of the replicates, and none had unique TSSs (Fig. 9). These data suggest that for the 3kbp variable *NR3C1* TSS region, a total number of oligo-labelled reads equal to 50x the length of the variable region, is adequate to detect microvariable TSSs.

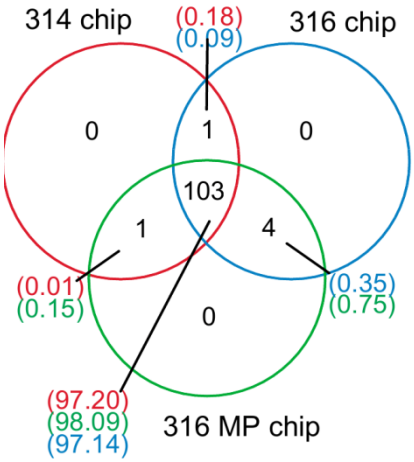

Figure 9: The proximal *NR3C1* CpG island TSSs used in three DAUDI technical replicates are plotted, showing common and unique TSSs. The number of TSSs shared between DAUDI technical replicates (numbers in parentheses are the % of total labelled 5'TSSs.[166] (Supplementary Data)

2.4.2. The Mono-Exonic *ADRB2R* Gene shows Biological Microvariability around One Transcriptional Locus

Analysis of a series of biological replicates revealed that 99.79% (±0.4) of the mRNA sequences used one of 5 consecutive nt (31 to 35) within the published sequence (Fig. 8C). The relative TSS frequencies were

essentially identical between replicates (Fig. 8D). TSS 35 usage was minimal (range 0.02% to 0.14%) and was therefore excluded as a valid TSS. The TSSs 30 and 39, mentioned in the previous section, were excluded as valid TSSs, after the application of the 0.10% cut-off. We conclude that the *ADRB2R* microvariability is limited to 4 nt around a single TSS locus.

2.4.3. The *NR3C1* TSSs are Highly Variable

The *NR3C1* 5' UTR of different cell lines and after different cell treatments were sequenced using the same protocol as for *ADRB2R*. Sequencing libraries were made from templates covering a region from the oligo-labelled TSSs to the start of the common exon 2. The oligo-labelled TSS reads were successfully aligned by TopHat to the *NR3C1*. Each of the 9 *NR3C1* first exons showed a remarkable TSS variability, but all microvariants used the previously published 3' splice donor sites (Fig. 5B). For example, in the biological replicates of A549 cells, we identified 123 to 128 TSSs, of which 96 were shared between replicates. The 96 shared TSSs accounted for ~77% of the reads per run (Fig. 10A -B). About ~17% of the reads used 27 to 32 unique TSSs, accounting each for 0.11% to 4.39% of the total reads per TSS. The remaining ~6% of the reads correspond to TSSs below the 0.10% cut-off and represent the intrinsic identification error rate. The TSSs were distributed in multiple loci per exon throughout the CpG island (Fig. 10A). Each cluster, or locus, consisted of a series of adjacent TSSs, that we term microvariability. For example, exon 1F was previously reported to be 62 nt long, starting at -3208 and ending -3146 [162]. We observed a series of shorter 1F exons, with two clusters of transcriptional loci around TSSs -3205 ±4bp and -3170 ±4bp all sharing the -3536 splice donor site (Fig. 10C). Exon 1B, with a length of 104 nt (-3640 to -3536), had 24 shorter forms all sharing the common splice donor site at -3146 (Fig. 10D) [202] and 11 TSSs immediately upstream of the TSS reported in the literature[161,162]. Similar trends including multiple microvariable transcriptional loci per exon were also observed for the regions corresponding to the other CpG island exons (1C, 1D, 1E, 1H and 1J) and the distal exons (1A and 1I) (Fig. 11-14). Some of the TSSs observed here, were located within a region immediately upstream of the ATG translation initiation codon in exon 2, and interestingly even downstream of the ATG, but still within exon 2. In general, our alternative TSSs were shorter than previously reported [161,162], although, for instance locus B4 and B5 were upstream of the previously identified TSS (Fig. 10D). Both the TSSs used and their variability, were reproducibly cell line dependent. This pattern of reproducible microvariability around multiple transcriptional loci between biological replicates was also observed for MCF-7 and T cells (Fig. 15-16).

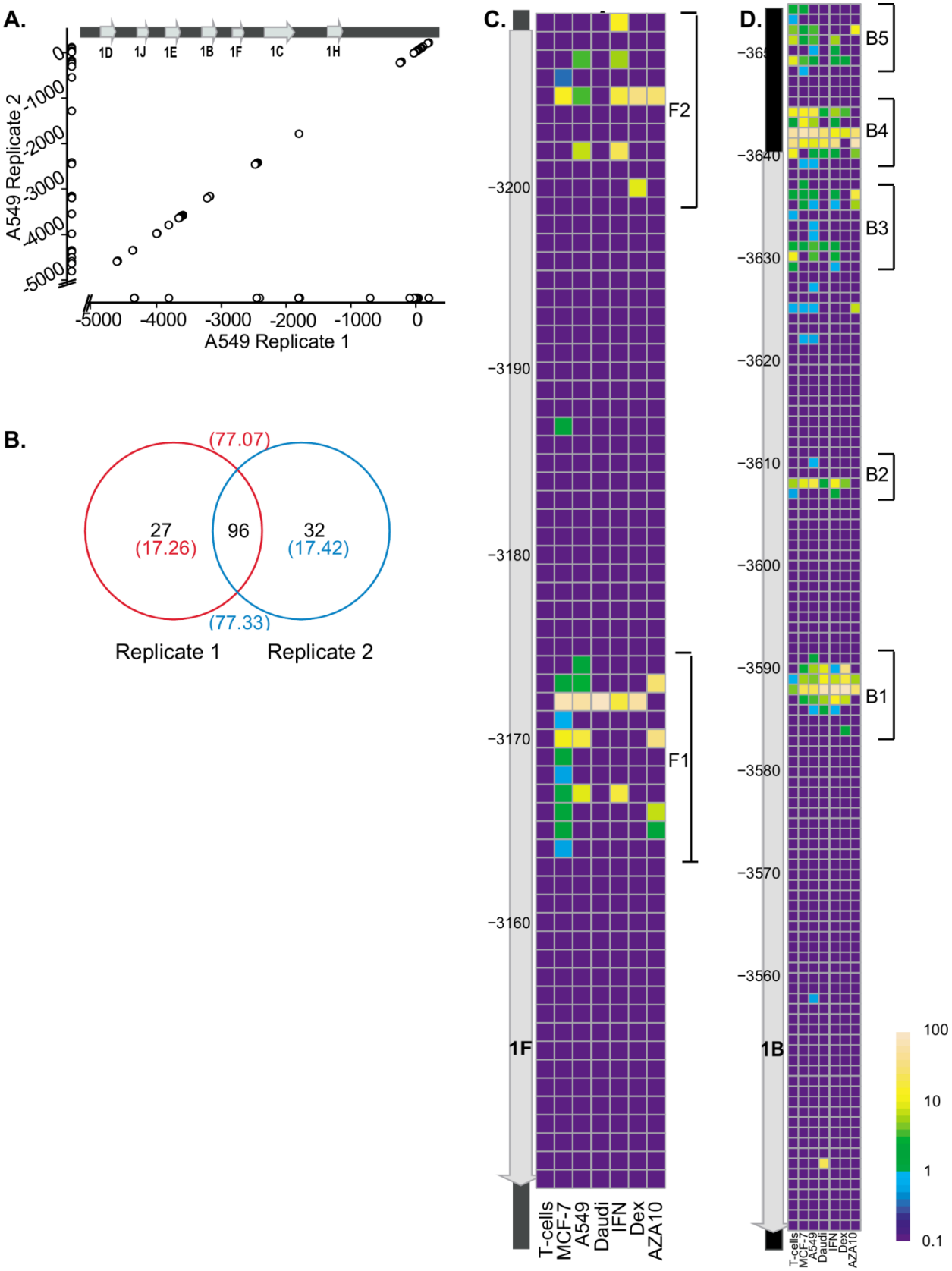

Figure 10: Microvariable TSS distribution throughout *NR3C1* (A) The proximal *NR3C1* CpG island TSSs used in two A549 biological replicates are plotted showing common and unique TSSs. The *NR3C1* TSSs on both axes are annotated with respect to the ATG (±1) translation initiation codon in exon 2. Data points adjacent to the x- and y-axes are unique to the respective replicates. (B) The number of TSSs shared between A549 biological replicates (numbers in parentheses are the % of total oligo-labelled 5' TSSs). (C) Detailed TSS usage pattern for the four cell lines and three treatment condition for exon 1F. (D) Detailed TSS usage pattern for the four cell lines and three treatment condition for exon 1B. Published exon 1F and 1B locations (Turner and Muller 2005[162]) are shown as a grey arrow, on the left of the heatmap. The TSS usage is expressed as the log value of the percentage of TSS expression for specific exon, by colour [0.01% (blue) to 100% (yellow)] of the heatmap.

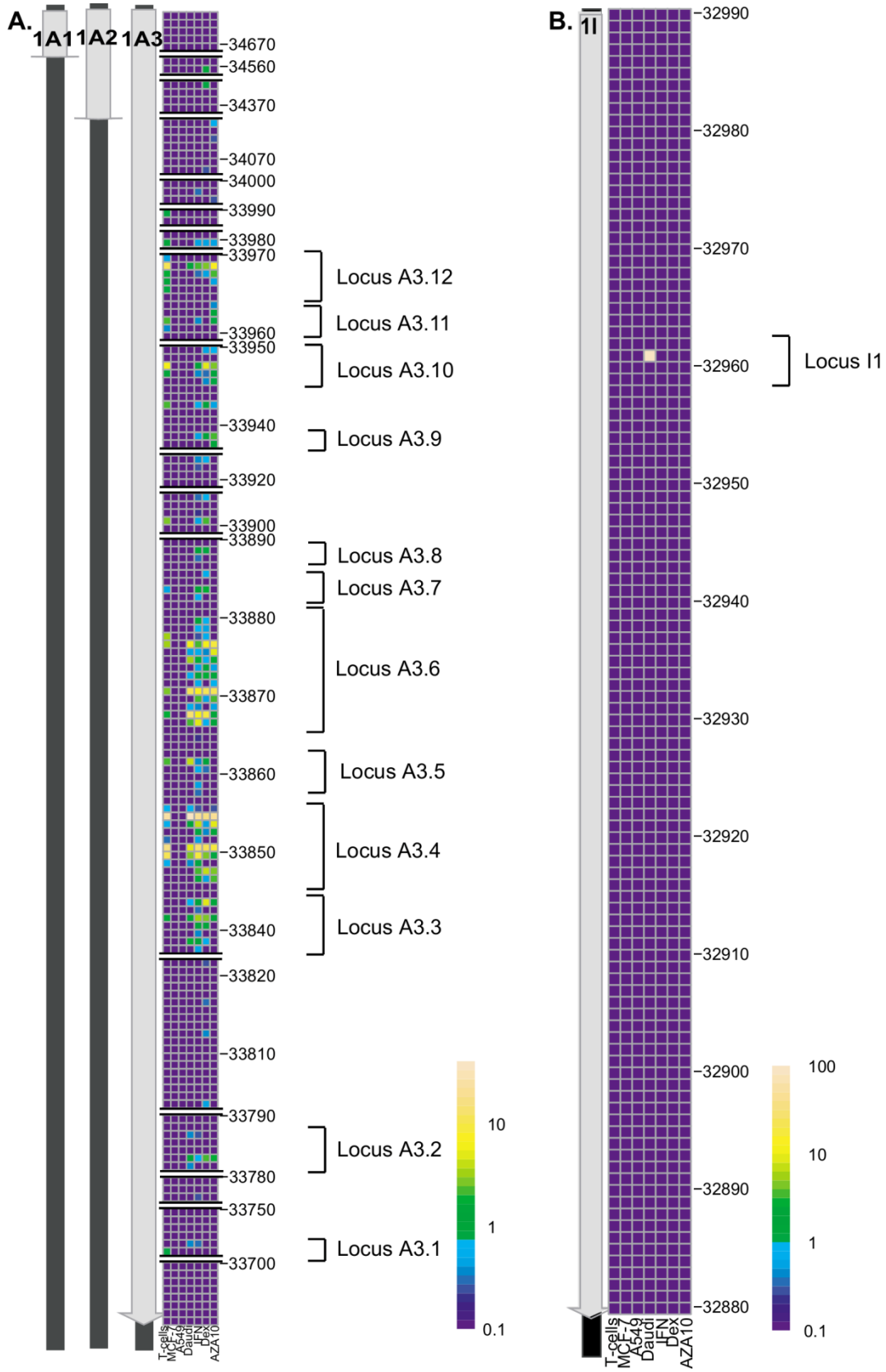

Figure 11: Detailed TSS usage pattern for the four cell lines and three treatment condition for exon 1A and 1I. Published exon locations (Breslin et al, 2001[160]; Turner and Muller 2005[162]; Presul et al, 2007[161]) are shown as a grey arrow, on the left. TSS usage is expressed by colour 0.1% (blue) to 100% (yellow)] logarithmically as the percentage of exon specific transcripts. (Supplementary Data)[166]

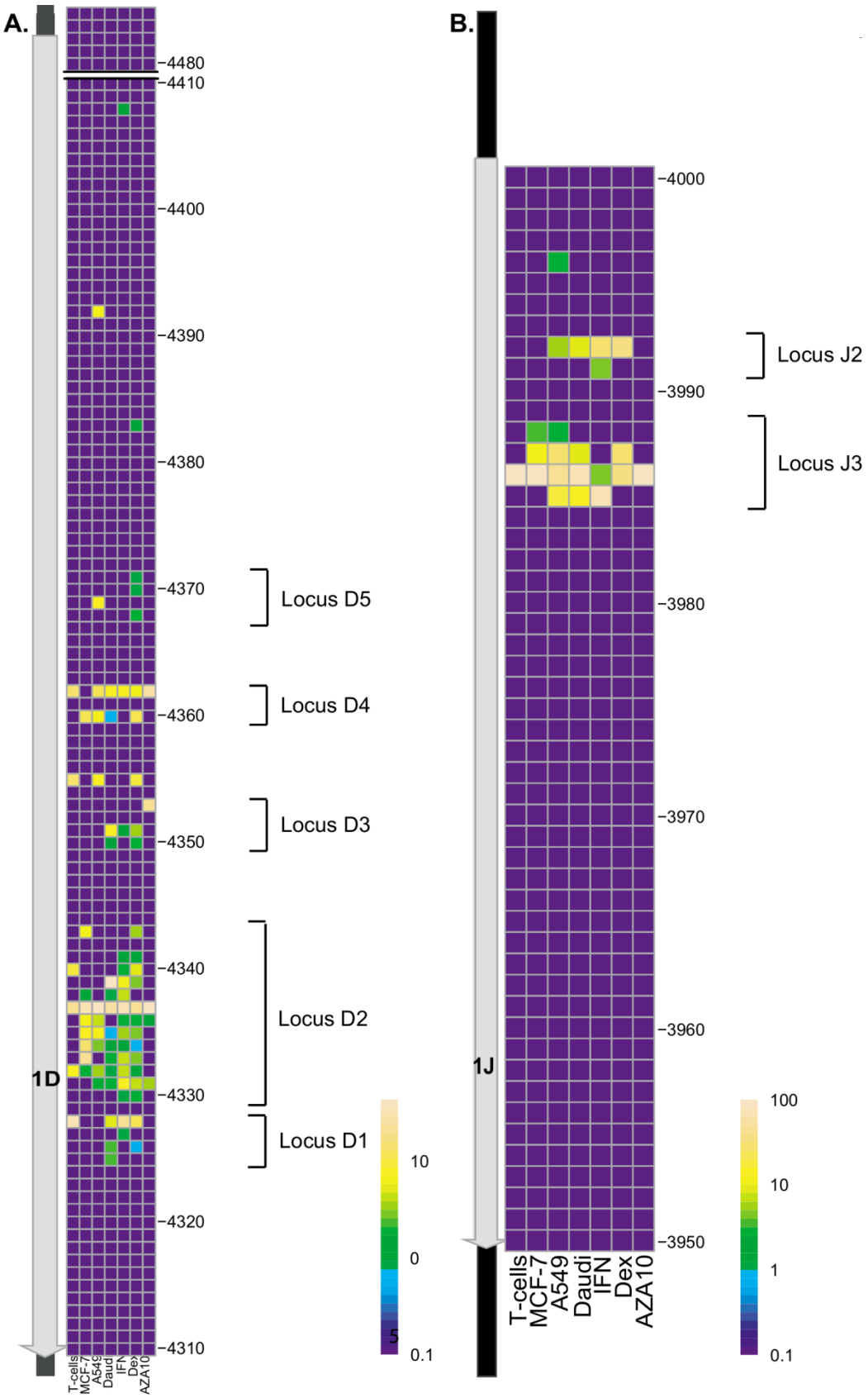

Figure 12: Detailed TSS usage pattern for the four cell lines and three treatment condition for exon 1D and 1J. Published exon locations (Turner and Muller 2005[162]) are shown as a grey arrow, on the left. TSS usage is expressed by colour 0.1% (blue) to 100% (yellow)] logarithmically as the percentage of exon specific transcripts.[166] (Supplementary Data)

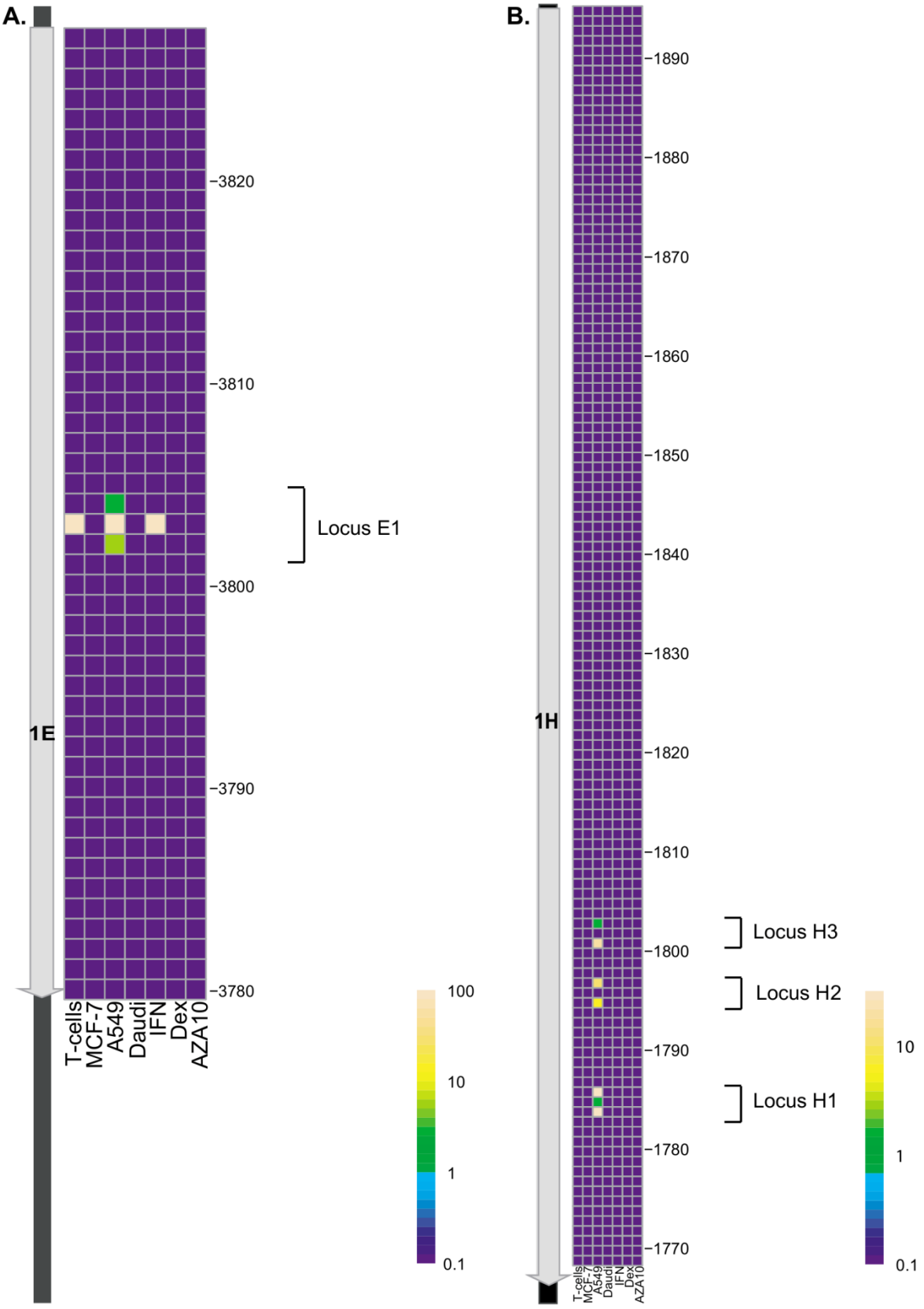

Figure 13: Detailed TSS usage pattern for the four cell lines and three treatment condition for exon 1E and 1H. Published exon locations (Turner and Muller 2005[162]) are shown as a grey arrow, on the left. TSS usage is expressed by colour 0.1% (blue) to 100% (yellow)] logarithmically as the percentage of exon specific transcripts.[166] (Supplementary Data)

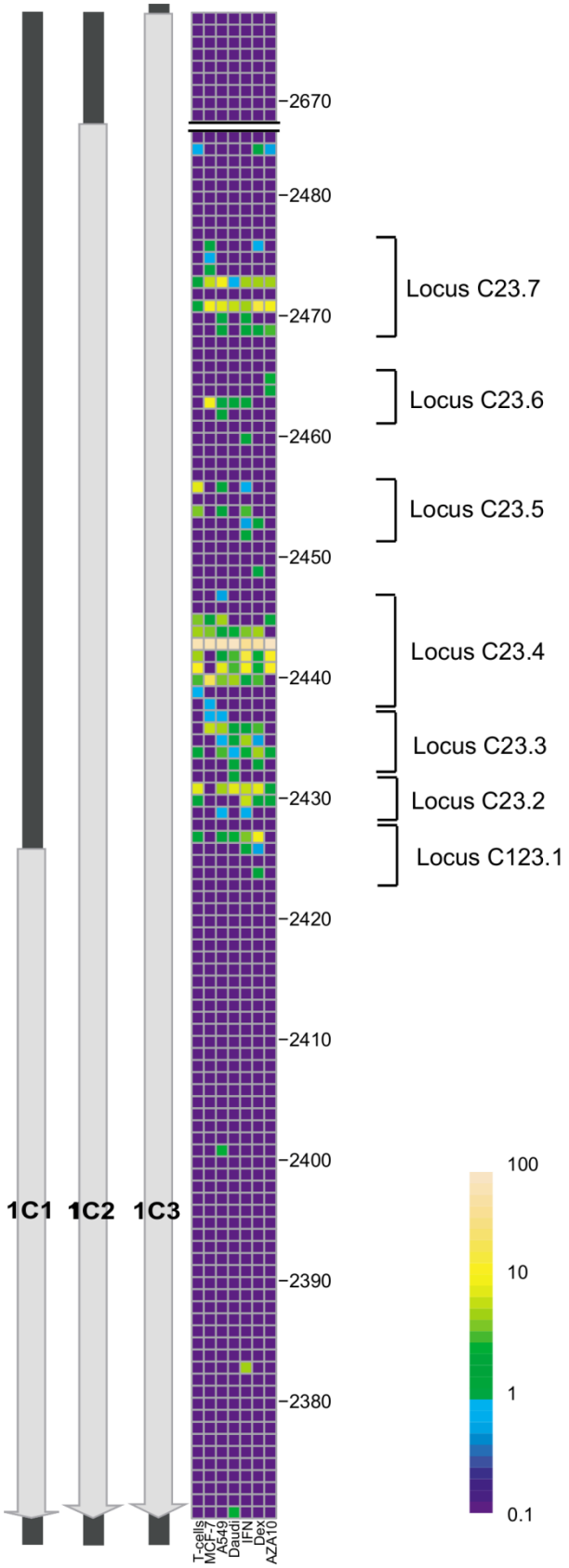

Figure 14: Detailed TSS usage pattern for the four cell lines and three treatment condition for exon 1C. Published exon locations (Turner and Muller 2005[162]) are shown as a grey arrow, on the left. TSS usage is expressed by colour 0.1% (blue) to 100% (yellow)] logarithmically as the percentage of exon specific transcripts.[166] (Supplementary Data)

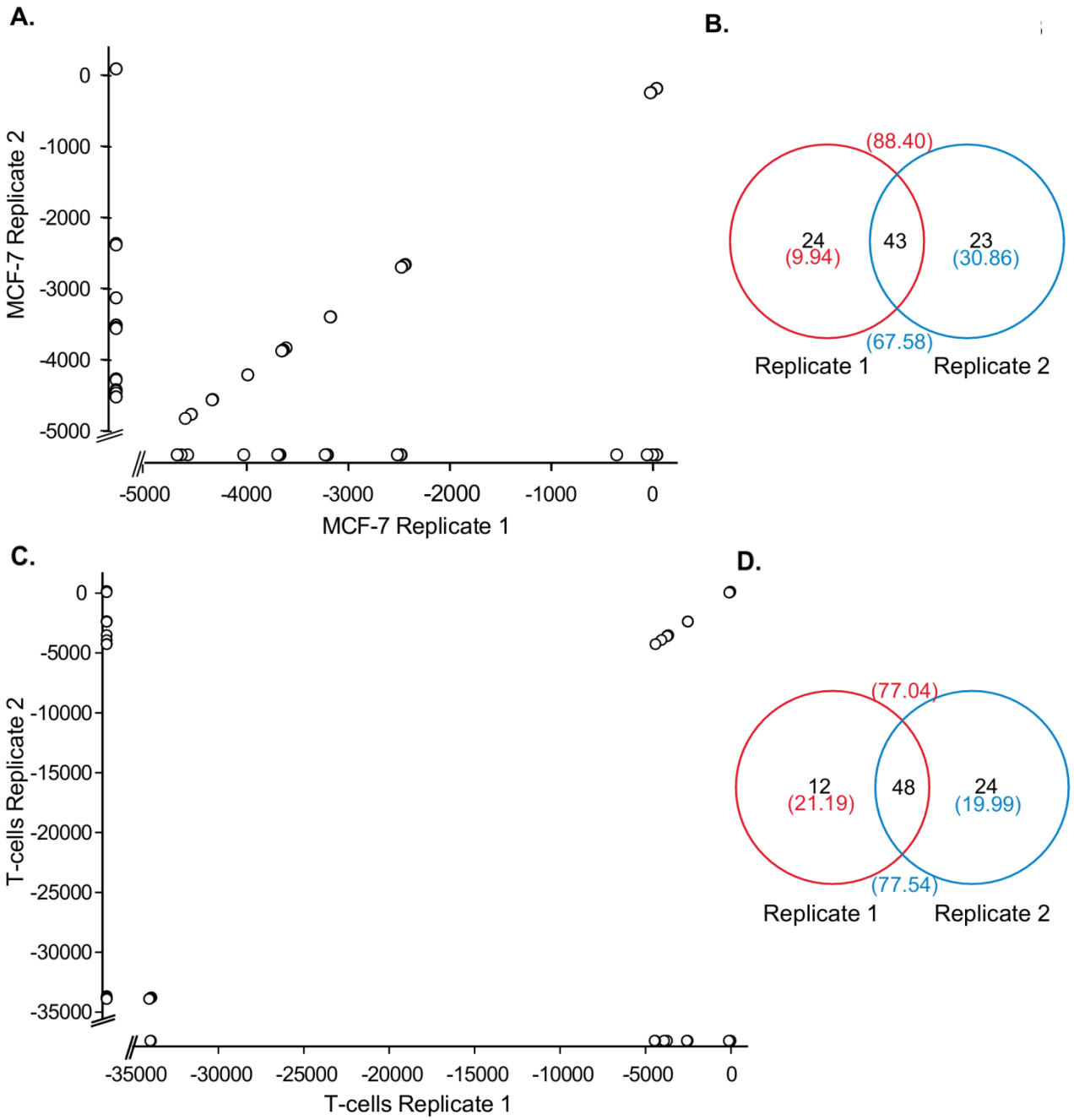

Figure 15: Microvariable TSS distribution throughout the *NR3C1*. (A) The proximal *NR3C1* CpG island TSSs used in two MCF-7 biological replicates are plotted showing common and unique TSSs. The *NR3C1* TSSs on both axes are annotated with respect to the ATG (±1) translation initiation codon in exon 2. (B) The number of TSSs shared between MCF-7 biological replicates (numbers in parentheses are the % of total labelled 5' TSSs). (C) The proximal *NR3C1* CpG island TSSs used in two T-cells biological replicates are plotted showing common and unique TSSs. The *NR3C1* TSSs on both axes are annotated with respect to the ATG (±1) translation initiation codon in exon 2. (D) The number of TSSs shared between T-cells biological replicates (numbers in parentheses are the % of total labelled 5' TSSs).[166]

2.4.4. TSS Expression Profiles are Cell Line Specific

In total we observed 262 discrete TSSs throughout the proximal and distal promoter for the 4 cell lines aggregated (A549, MCF-7, T-cells and DAUDI) (Fig. 16B and D). 66 of these were common to all cell lines and other were shared with at least one other cell line. Each cell line had a small population of unique TSSs (range 6 to 43 TSSs). Pairwise analysis of the 4 cell lines, revealed between 134 and 196 TSSs that were differentially

expressed per comparison (Fig. 16A). Some of these differential TSSs were unique to one cell line, while others were used in several cell lines at similar or different levels. For example, 125 TSSs were shared by both MCF-7 and A549 cells (Fig 16A and D), while 9 were uniquely expressed in MCF-7 and were not seen in any other cell lines. 71 TSSs were solely expressed in A549 and not in MCF-7. 28 of these were also found in other cell lines, leaving 43 TSSs unique to A549 cells. Hierarchical clustering identified the 30 TSSs that most strongly discriminated between the cell lines (Fig. 16B and C). Many of these could be related to transcriptional loci observed for exon 1B and 1F (Fig. 10C-D). When the read frequencies were taken into account, a smaller group of 8 TSSs emerged, which clearly discriminated between the different cell lines (A549, MCF-7, T-cells and DAUDI) (Fig. 16B). The above 8 TSSs were also the most conspicuous within the dendrogram (Fig. 16C). The TSSs 17 and -18 were T cells specific, TSSs -5, -13, -3642 (locus B4) and -33855 were DAUDI cell specific and TSSs -3172 (locus F1) and -2443 (locus C23.4) differentiated best between MCF-7 and the A549 cell line respectively. Except TSSs 17 and -33855, these characteristic TSSs were located within the proximal promoter. TSSs -5, -13, -18 and 17 were located within exon 2. The first three were situated upstream and the latter downstream of the ATG translation initiation codon in exon 2. The TSSs -2443, -3172 and -3642 correspond to exons 1C, 1F and 1B in the proximal promoter region. TSS -33855 corresponds to exon 1A in the distal promoter region. The remaining differentially expressed TSSs also differed between cell lines, yet, in a much less obvious way.

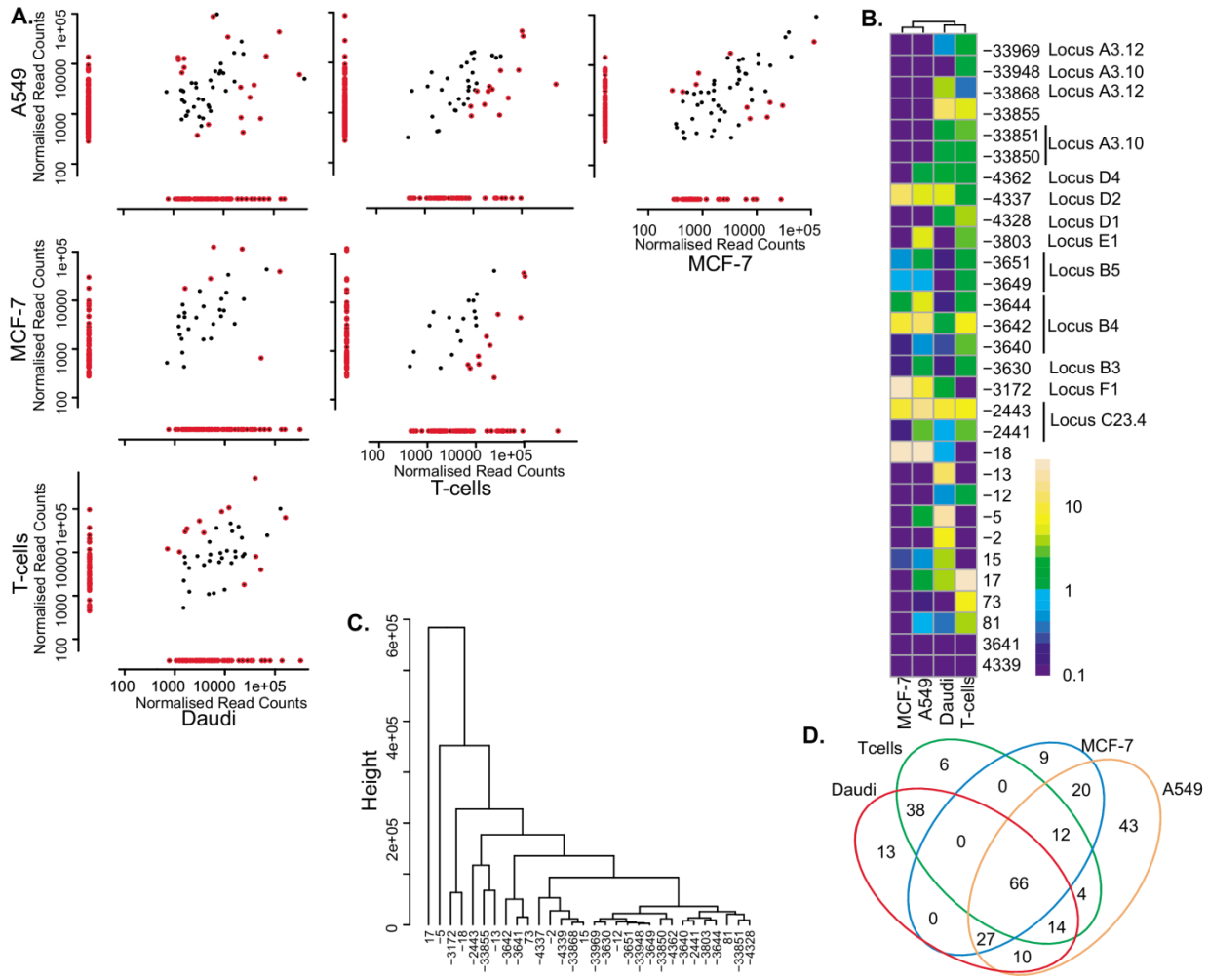

Figure 16: Differential expression of TSSs between multiple cell lines. (A) Pairwise comparisons of normalised read counts for all cell lines. The coloured dots (•) represent the significantly differentially expressed TSSs between 2 cell lines, the black dots (•) represent TSSs that are not differentially expressed between 2 cell lines. Significance was considered for an adjusted p-value of 0.05 after performing the NOISeq proportion test (35) (B) Expression of the 30 most discriminatory TSSs between the different cell lines, as a percentage of the total oligo-labelled TSSs. (C) Hierarchical clustering of the 30 most discriminatory differentially expressed TSSs. (D) The number of unique and common differentially expressed TSSs for the four cell lines. [166]

2.4.5. DAUDI Cells have Highly Variable Biological Replicates

In contrast to other cell lines, the TSS selection patterns in DAUDI cells were less consistent, even in multiple biological replicates (Fig. 17). The DAUDI cell line, a Human Burkitt's lymphoma cell line that reliably expresses the distant 1A exon, selected TSSs in both the distal and proximal promoter regions.

To assess the importance of read depth and determine potential sources of biological variability from technical variability, technical triplicates of the *NR3C1* in DAUDI cells were sequenced on respectively 314, 316 and multiplexed 316 chips with different MIDs. Resulting in respectively 61 807, 2 338 506 and 639 608 reads per run. After applying the 0.10% cut-off, 103 TSS, which corresponded to 97.14% to 98.09% of all reads, were shared between all replicates (Fig. 9). As expected, differential expression analysis did not identify any discriminating or run-specific TSS expression profile. Hereby, indicating that the error induced by differences in

MIDs, adaptors, PCR cycles and Ion Torrent chips was only minimal and confirmed the biological origin of the variability in the DAUDI cells.

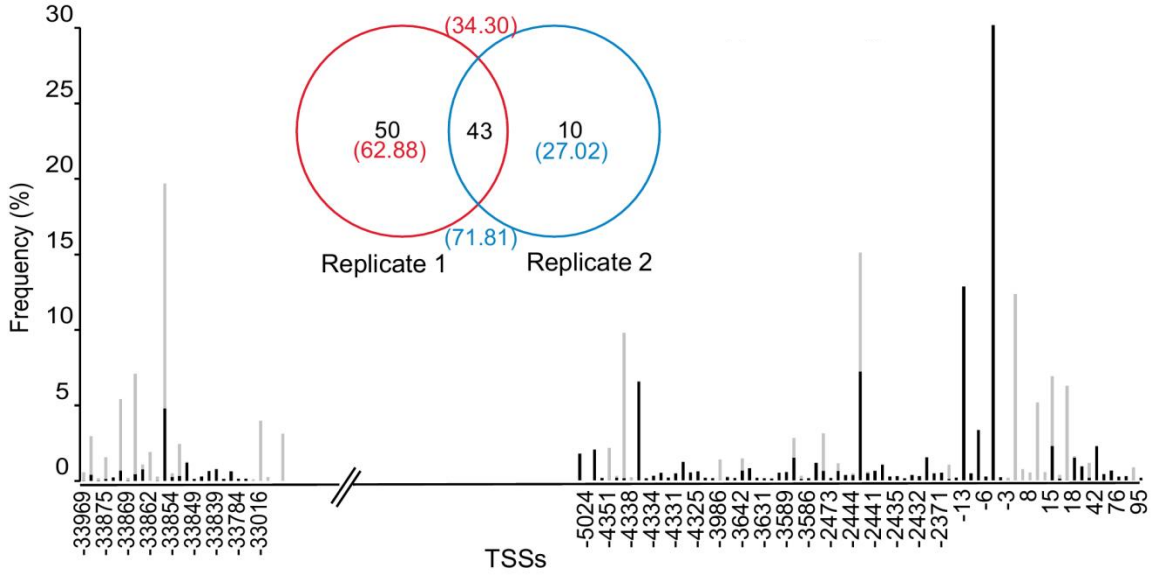

Figure 17: The proximal *NR3C1* CpG island TSSs used in two DAUDI biological replicates are plotted showing common and unique TSSs and their frequencies. The *NR3C1* TSSs on the horizontal axis are annotated with respect to the ATG (±1) translation initiation codon in exon 2. The number of TSSs shared between DAUDI biological replicates (numbers in parentheses are the % of total labelled 5' TSSs).[166] (Supplementary Data)

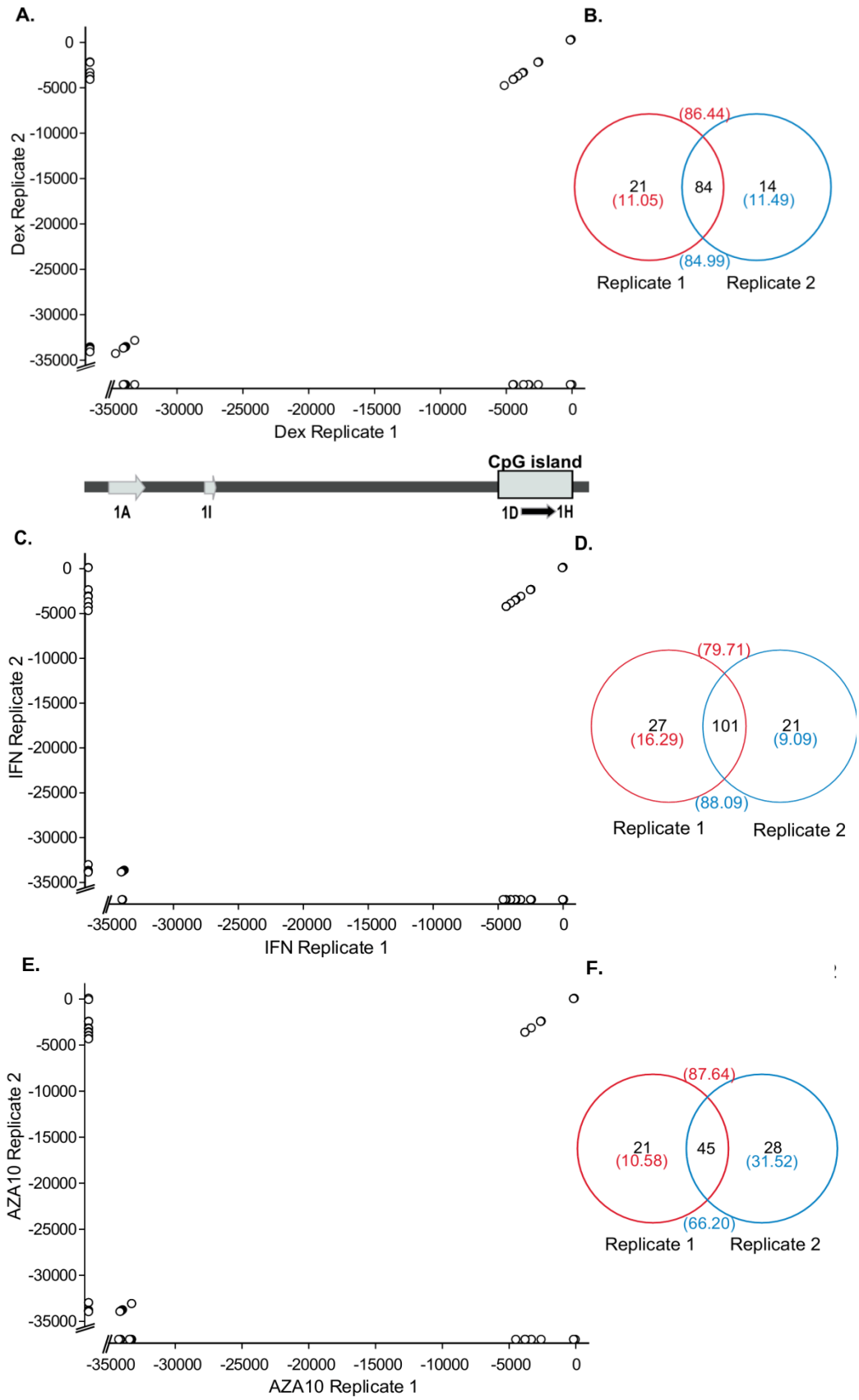

Figure 18: Microvariable TSS distribution throughout the *NR3C1* under different treatment conditions (A) Reproducibility of TSS identification in DAUDI cells after dexamethasone stimulation. The region covered corresponds to both the distal and proximal promoter regions of *NR3C1*. (B) The number of TSSs shared between the two dexamethasone stimulation replicates. (C) The proximal *NR3C1* CpG island TSSs used in two IFN-γ treated DAUDI biological replicates are plotted showing common and unique TSSs. The *NR3C1* TSSs on both axes are annotated with respect to the ATG (±1) translation initiation codon in exon 2. (D) The number of TSSs shared between the IFN-γ treated DAUDI biological replicates. (E) The proximal *NR3C1* CpG island TSSs used in two AZA10 treated DAUDI biological replicates are plotted showing common and unique TSSs. The *NR3C1* TSSs on both axes are annotated with respect to the ATG (±1) translation initiation codon in exon 2. (F) The number of TSSs shared between the AZA10 treated DAUDI biological replicates. (numbers in parentheses are the % of total oligo-labelled 5' TSSs).[166] (Panel C-F: Supplementary Data)

2.4.6. Environmental Conditions Influence the TSS Pattern

To examine the effect of transcriptional stimuli on TSS microvariability, the transcriptionally variable DAUDI cells were exposed to IFN-ɣ, Dex or AZA. These treatments markedly reduced the TSS variability observed in unstimulated DAUDI cells. TSS selection patterns became reproducible between biological replicates, but differed between treatments. When exposed to Dex, DAUDI cell replicates targeted 98 to 105 TSSs, of which 84 were shared between replicates. The 84 shared TSSs accounted for ~85% of the reads per run (Fig. 18A-B). As for the experiments above with stable unstimulated MCF-7, A549 and T cells, each run also selected a small number of unique TSSs (range 14 to 21), but these correspond to no more than 11.49% of the reads and less than 3% of the total reads per TSS. The remaining ~4% of the reads were below the 0.1% cut-off and represent the error rate intrinsic to our 5' labelling technique. Although the majority of the TSSs were located within the proximal promoter region, a smaller number of TSSs were found in the distal *NR3C1* promoter region (Fig. 18A). Similar results were observed for DAUDI cells exposed to IFN-ɣ or AZA (Fig. 18C-F).

In total we observed 234 discrete TSSs throughout the proximal and distal promoter region for the 4 treatments aggregated (Fig. 196B and D). 75 of these were common to all treatments and other were shared with at least one other treatment. Each treatment had a small population of unique TSSs (range 9 to 12 TSSs). Pairwise analysis of the different treatments, revealed between 127 and 179 TSSs that were differentially expressed per comparison (Fig. 19A). Similar to the comparison between cell lines, some of these differential TSSs were unique to one treatment, while others were used in several treatments at similar or different levels. For example, 140 TSSs were shared by both Dex and IFN-ɣ stimulations (Fig. 19A and D), while 28 and 39 TSSs were solely expressed after Dex or IFN-ɣ treatment respectively. 17 of the 28 TSSs induced by Dex were also found in other treatments, leaving 11 TSSs unique to Dex. Similarly, 27 of the 39 TSSs induced by IFN-ɣ were also associated with other treatments, leaving 11 TSSs unique to IFN-ɣ. Hierarchical clustering identified 28 TSSs that most strongly discriminated between treatments (Fig. 19B and C). Many could be related to transcriptional loci observed in exon 1B and 1F (Fig. 10C-D). When the read frequencies were taken into account, a small group of 6 TSSs emerged, which clearly discriminated between the different treatments (Fig. 19B). The above 6 TSSs were also the most conspicuous within the dendrogram (Fig. 19C). The TSSs -5 and -13 were specific to untreated DAUDI cells, TSSs -1 and -2 were DAUDI cells + AZA specific and TSSs -2443 (locus C23.4) and -33855 differentiated best between IFN-ɣ and Dex treatments respectively. TSSs -1, -2, -5, and -13 were situated in exon 2, upstream of its ATG translation initiation codon. The TSS -2443 (locus C23.4) corresponds to exon 1C in the proximal promoter region and TSS -33855 corresponds to exon 1A3 in the distal promoter region. The remaining differentially expressed TSSs also differentiated between different stimulations,

yet, in a much less obvious way. As in our comparison of resting cell lines, we observed transcriptional microvariability around the same loci for the DAUDI cells exposed to either IFN-, AZA or Dex. Although, transcription microvariability was observed to occur in the same loci as unstimulated cells, the differential expression pattern, i.e. the pattern of TSSs within each locus, was treatment specific.

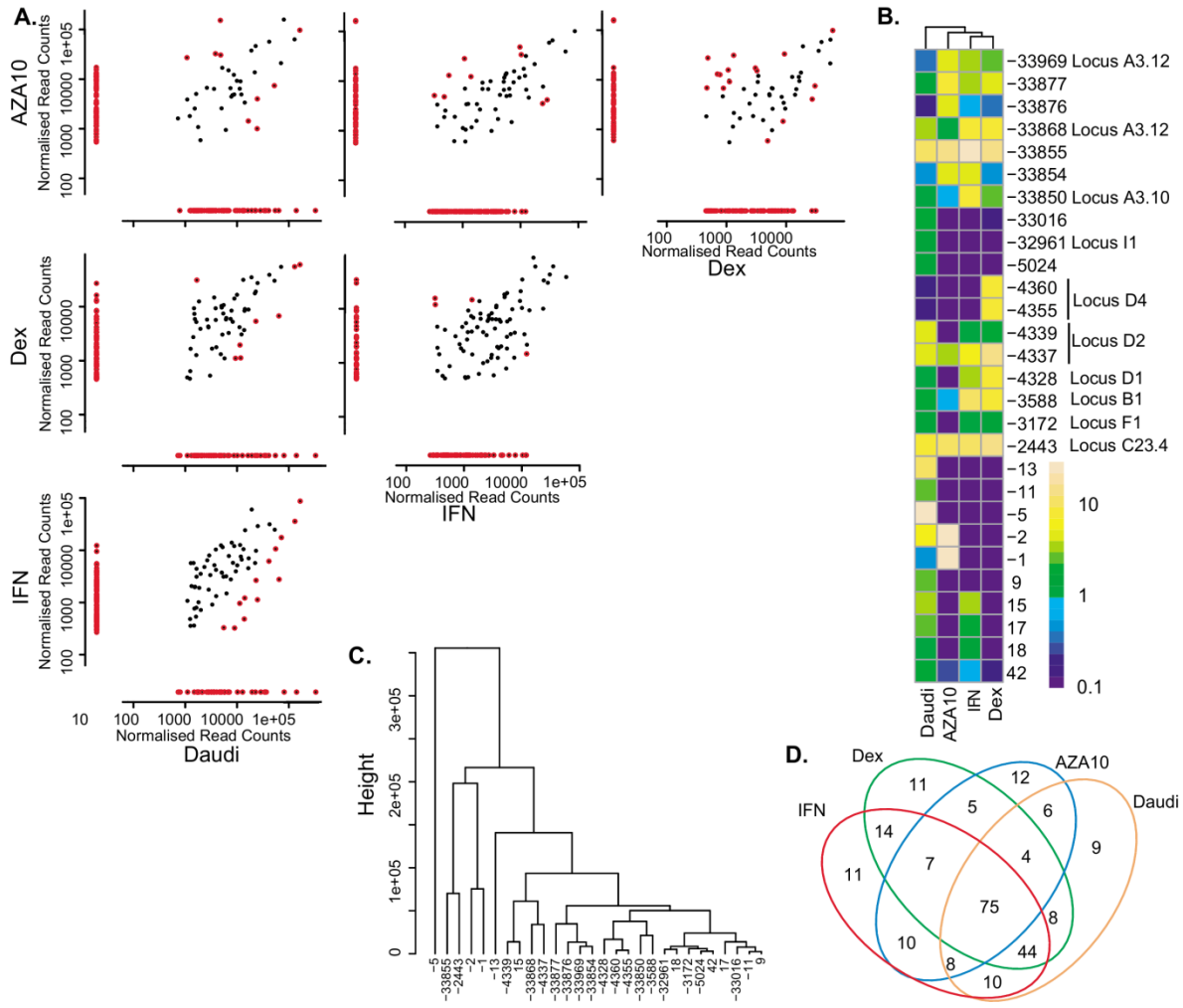

Figure 19: Differential expression of TSSs between untreated, dexamethasone, Interferon-ɣ and 5-AZA-2'-deoxcytidine exposed DAUDI cells. (A) Pairwise comparisons of normalised read counts for all cell lines. Only valid TSSs with a frequency above 0.1% of the total reads were included. The coloured dots (•) represent the significantly differentially expressed TSSs between cell treatment conditions, the black dots (•) represent TSSs that are not non-differentially expressed between treatments. Significance was considered for an adjusted p-value of 0.05 after performing the NOISeq proportion test (35) (B) Heatmap of the 30 most discriminatory TSSs between the different treatment conditions, expressed on a logarithmic scale. (C) Hierarchical clustering of the 30 most discriminatory differentially expressed TSSs. (D) The number of unique and common differentially expressed TSSs between the different treatment conditions untreated, dexamethasone, Interferon-ɣ and 5-AZA-2'-deoxcytidine.[166]

2.4.7. Translational Efficiency Assay

To investigate the functional consequence of *NR3C1* microvariability, we constructed a series of 14 plasmids. Each plasmid contained a first exon variant from of 1A3, 1B or 1C, immediately upstream of a luciferase coding sequence. The first exon microvariants were either (i) full length 1st exon sequences as reported in the literature [161,162], (ii) a sequence starting from a TSS observed within a few nt of the literature start site until the end (e.g. 1B 105 bp and 107 bp), (iii) a sequence from TSSs mid first exon until the end (e.g. 1B 73 bp or 1C 71 bp and 73 bp) or (iv) a sequence starting from TSSs close to the 3' end of the first exon to the end of the exon (e.g. 1A3 94 bp and 1B 53bp). Luciferase assays were performed to evaluate the translational efficiency of the different 5'UTRs (Fig. 20D). The translation efficiency per exon decreased with the increase of construct's length. Translation efficiency of the 1B microvariability variants (highlighted in figure 20D) was much lower than the 1A3 and 1C variants. In all cases there was a considerable effect of the observed transcriptional microvariability on the efficiency of luciferase production (Fig. 20A-C). The highest luciferase signal was observed for the 94 bp long exon 1A3 with a 75 fold increase (q =4.65) compared to the full length 1st exon sequences as reported in the literature [161,162]. Exon 1C (73bp) and exon 1B (107bp) showed the highest luciferase signal of any of the exon 1C and 1B constructs, with a 5-fold (q = 4.32) and a 2-fold increase (q = 2.40) respectively. Aside from the 1B constructs, that did not show any significant effect in luciferase activity, our observations were compatible with our prior report of shorter sequences having a higher activity [164]. Accordingly, only small differences in activity were observed between transcripts of similar length, i.e. with neighbouring TSSs, e.g. 1C 71 bp versus 73 bp (q = 1.44) and 1A3 160 bp versus 164 bp (q = 1.807). The luciferase signal relative to the full length of the 1st exon was in general higher for the exons 1A3 transcriptional variants. The signals measured for exon 1C variants were only slightly higher compared to those of exon 1B variants. Although one could assess the previously published full length exons 1A3 and 1C, respectively 981nt and 479nt, as outliers (Fig. 20D), removing them maintained the negative correlation and resulted in a steeper slope. Overall, translation efficiency increased as 5' UTR length decreased, irrespective of the alternative first exon.

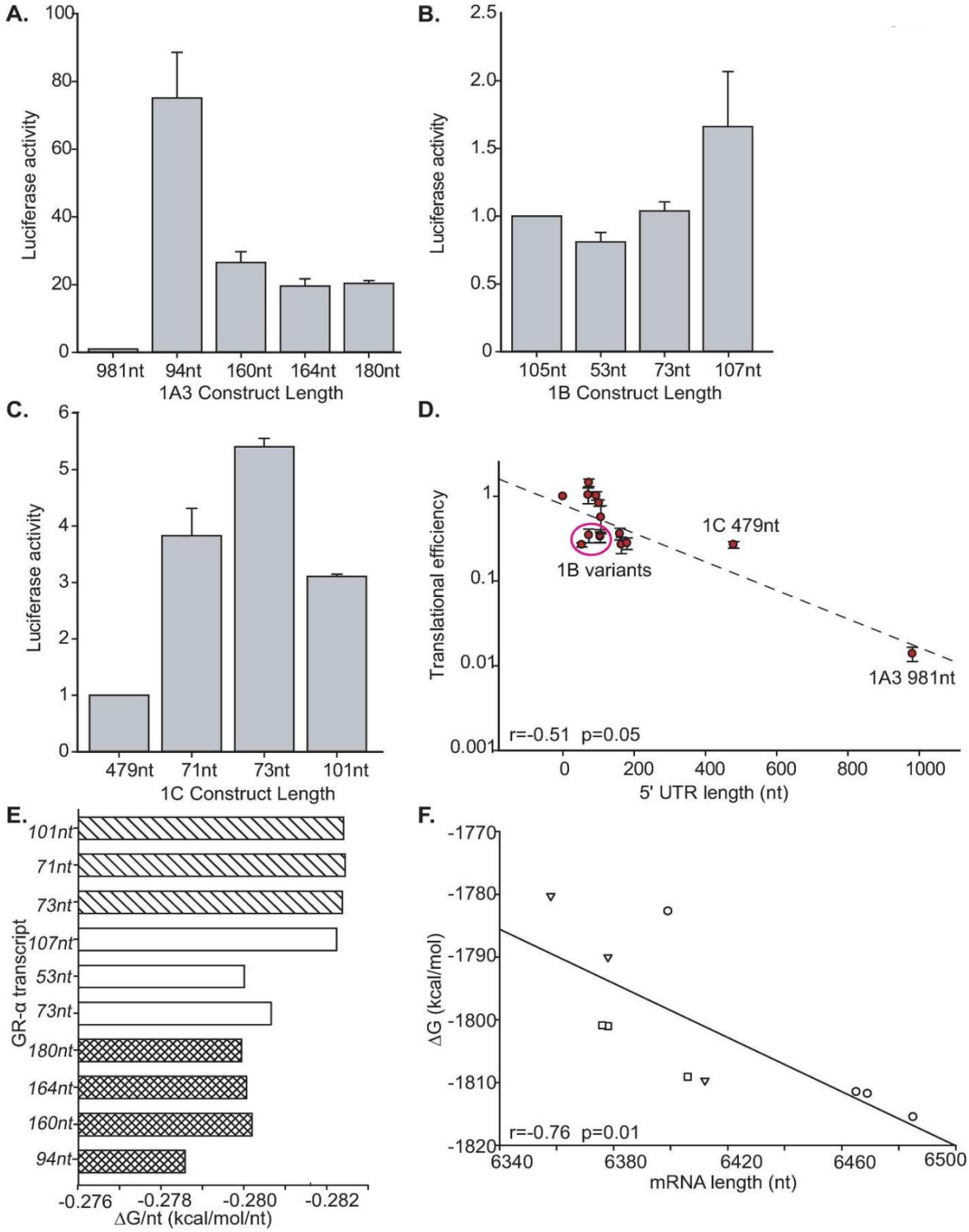

Figure 20: Translational efficiency and mRNA stability for exon 1A, 1B and 1C. (A) The luciferase activity of three replicates per 1A3 construct for different construct lengths. The first construct being the original. The luciferase state of the different constructs was normalised to the luciferase activity of the original full length exons (11) (B) The luciferase activity of three replicates per 1B construct for different construct lengths. The first construct being the original. The luciferase state of the different constructs was normalised to the luciferase activity of the original full length exons (11) (C) The luciferase activity of three replicates per 1C construct for different construct lengths. The first construct being the original. The luciferase state of the different constructs was normalised to the luciferase activity of the original full length exons (11). (D) Translational efficiency values from luciferase are plotted in function of their 5'UTR length expressed in base pairs (bp), presenting a linear decrease in efficiency as the length increases. Exon 1B variant with a lower translation efficiency/nt are circled, and the marked 1C and 1A3 datapoints are the full length literature sequences (20, 40). (E) The free energy of folding (ΔG/nt) for 1A3 microvariable sequences calculated per nucleotide (1A3: ; 1B: ; 1C: constructs. (F) The mRNA length in function of the free folding energy of the mRNA (ΔG) (1A3: ; 1B: ; 1C:). Data in panels A-D are mean ± SD. [166]

2.4.8. *NR3C1*-Transcript Secondary RNA Structure is Influenced by the 5'UTR Length

Secondary structures of full length mature mRNA sequences of *NR3C1*-α transcripts were modelled using the program RNAfold [201]. The *NR3C1*-α transcripts for exon 1A3, 1B and 1C all induce a similar folding. Overall there was a strong negative correlation between the total ΔG and the mRNA length irrespective of its location within the CpG island (r=-0.76) (Fig. 20E). However, as would be expected based on the similarity of their sequences, microvariable sequences clearly cluster by exon. Within these populations the trend was even more pronounced (r=-0.99, r=-0.99, r=-0.99). The free energy per nucleotide (ΔG/nt), being a better indicator of the secondary mRNA structure stability, was anticipated to be similar between the 3 exon *NR3C1* transcripts. The ΔG/nt values range from -0.2824 to -0.2786 for all 3 exon transcripts (Fig. 20F). The ΔG/nt of *NR3C1* 1A3 transcripts are slightly higher than those of the *NR3C1* 1B and 1C transcripts, indicating that the 1A3 transcripts are probably slightly less stable than the 107nt long 1B exon construct or the 1C constructs. For *NR3C1* 1C transcripts the ΔG/nt does not change with mRNA length, for the other exon microvariants, it is always the smallest mRNA variant that has the highest ΔG/nt. However, all of these differences are minimal, therefore we would assume that the differences in stability are minimal.

2.4.9. 5′UTRs Influence the Relative GR Protein Isoform Distribution

To investigate the role of alternative 5′UTRs in translational start site selection, ten microvariable constructs covering the two constitutive [162] and the upstream distal first exon were made. The TSSs covered locus 142 814 191, 142 814 257, 142 814 261 and 142 814 277 for 1A3, 142 783 994, 142 784 014 and 142 784 048 for 1B and 142 782 847, 142 782 849 and 142 782 877 for 1C. Forty-eight hours post transfection all microvariable constructs showed significant differences in the distribution of N-terminal isoforms on Western blots (Fig. 21-22). The most abundant N-isoform for all constructs was GR-A (Fig.21A and 22). For 1A3 and 1C constructs, the GR-A isoform represented 50% of the total *NR3C1* (range 41% to 65% and 42% to 64% respectively) and GR-D represented approximately 19% (range 12% to 36% and 8% to 39% respectively). For the 1B microvariants the GR-A levels were lower, but GR-D levels higher. Taking all the microvariable constructs together, there was a statistical significant transcript variant*protein isoform interaction (DF=39, F=3.6, p-value<0.001, Two-Way ANOVA). This transcript variant*protein isoform interaction was maintained when exon microvariants were treated independently. Detailed examination of the 1A3 variant (DF=12, F=9.064, p-value<0.001, Two-Way ANOVA, Fig. 21A) showed that in the subsequent Student-Newman-Keuls' pairwise comparison (Fig. 21B) there was a significant difference in the level of GR-A for the different microvariable constructs, as well as GR-D and GR-B. The expression of each protein isoform was heavily influenced by the transcript length (Fig. 21B). GR-A levels from the 1A3 94nt variant were

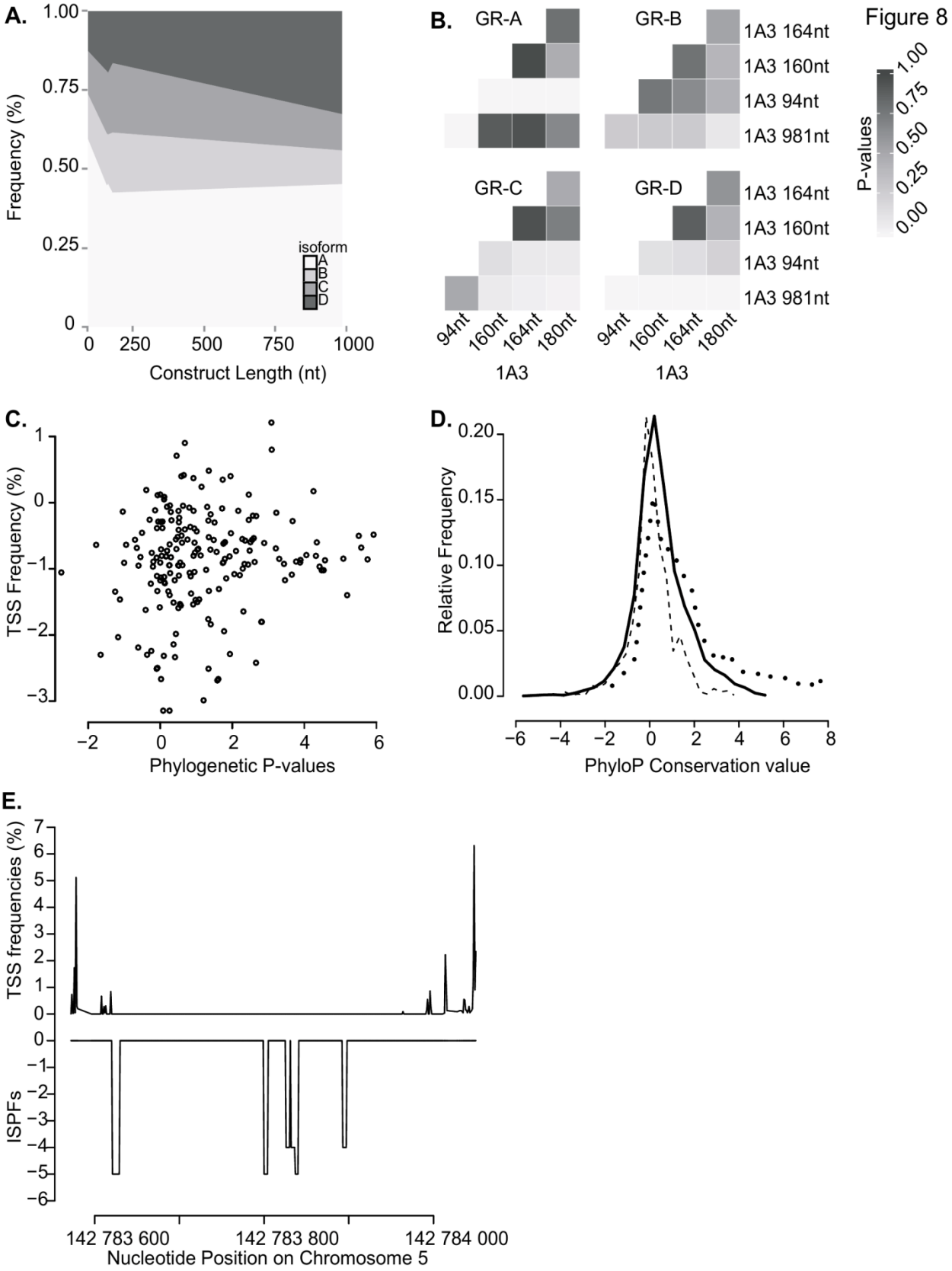

Figure 21: The microvariable TSSs influence the relative abundance of N-terminal protein issoforms and the evolutionary conservation of this region. (A) Areaplot for 1A3 microvariable constructs, indicating the shift of isoform levels according to the construct length. Two-Way ANOVA (DF=12, F=9.064, p<0.01). (B) Pairwise comparison of relative protein levels. Student-Newman Keuls P-values are expressed by grey scale [0.01 (white) to 1 (black)]. (C) The phylogenetic p-value of a specific nucleotide position is plotted in function of the TSS frequency measured at that specific location. (D) The relative frequency distribution for respectively the distal promoter ( ), proximal promoter ( ) and exon 2 region ( ) in function of the PhyloP conservation value. Negative values indicate higher evolution rates, positive values indicate higher conservation rates and zero equal a neutral position concerning evolution rates. (E) For exon 1F and the region before, the in silico phylogenetic footprinting sites (38) the TSS frequencies were plotted against one another.[166]

significantly higher from the other 4 mRNA microvariants (Fig. 21B). The published 1A3 variant (1A3 mini) has significant higher GR-D expression levels compared to all shorter microvariants, additionally its GR-C expression level was significantly lower compared to the 1A3 180nt variant (Fig. 21B). Hence, the considerable differences in translational efficiency amongst the alternative *NR3C1*-transcripts suggests that the 5'UTR variability, and thus TSS microvariability, is involved in the modulation of transcript isoform levels and consequently in protein isoform levels.

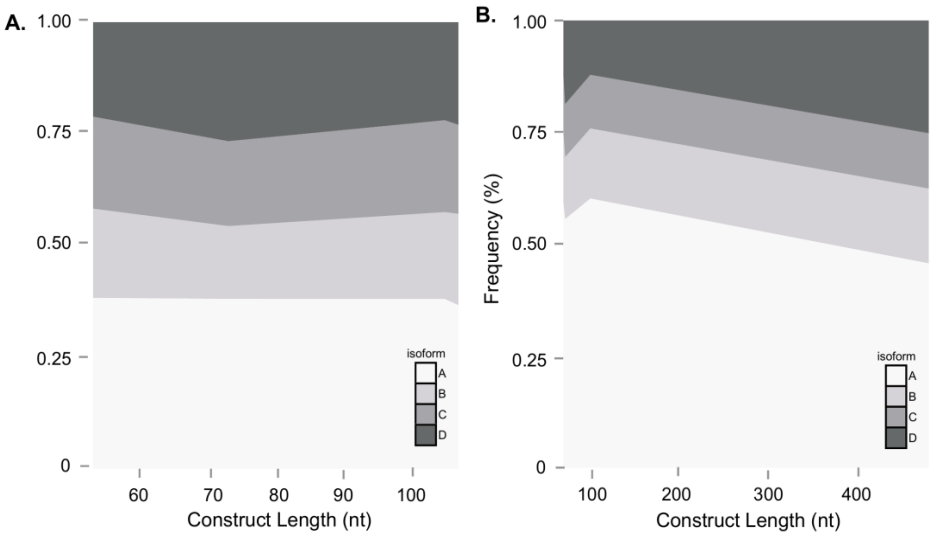

Figure 22: The microvariable TSSs influence the relative abundance of the N-terminal protein isoform. (A) Areaplot of 1B microvariable constructs, indicating the shift of isoform levels according to the construct length. (C) Areaplot of 1C microvariable constructs, indicating the shift of isoform levels according to the construct length.[166] (Supplementary Data)

2.4.10. Evolutionary Analysis of *NR3C1*

The relative conservation of both the variable upstream promoter regions and the conserved *NR3C1* encoding regions were visually investigated from the UCSC 100 vertebrate phyloP conservation data (Fig. 21 and 23). Conservation values within exon 2 are higher and more constant than those in the distal and proximal promoter region, where the conservation pattern was more variable (Fig. 21 and 22). Within both the distal and proximal promoter region the microvariable TSSs did not coincide with regions of increased or decreased evolutionary conservation, but were randomly distributed. This observation was confirmed by plotting TSS usage against phyloP phylogenetic values (Fig. 21C), where no trend was detected. The bulk of the TSSs, irrespectively of their frequency, were centred within the phylogenetic p-value range from -2 till 2, suggesting that irrespective of their frequency, most TSSs were relatively neutral concerning evolution rates, the conservation of these sites is not high, nor is their evolution rate. Throughout the distal and proximal promoter

the phyloP conservation level was distributed around zero (Fig. 21D). Although the coding exon 2 has previously been reported to be highly conserved [203], this was not so. The transcription factor binding sites identified by in silico phylogenetic footprinting [200] did not coincide with nucleotide position of the differentially expressed TSSs (Fig. 21E). Overall it seems that the observed human TSS microvariability does not occur in regions that are evolutionary conserved, suggesting that the pattern of microvariable TSS usage may be species specific.

2.5. Discussion

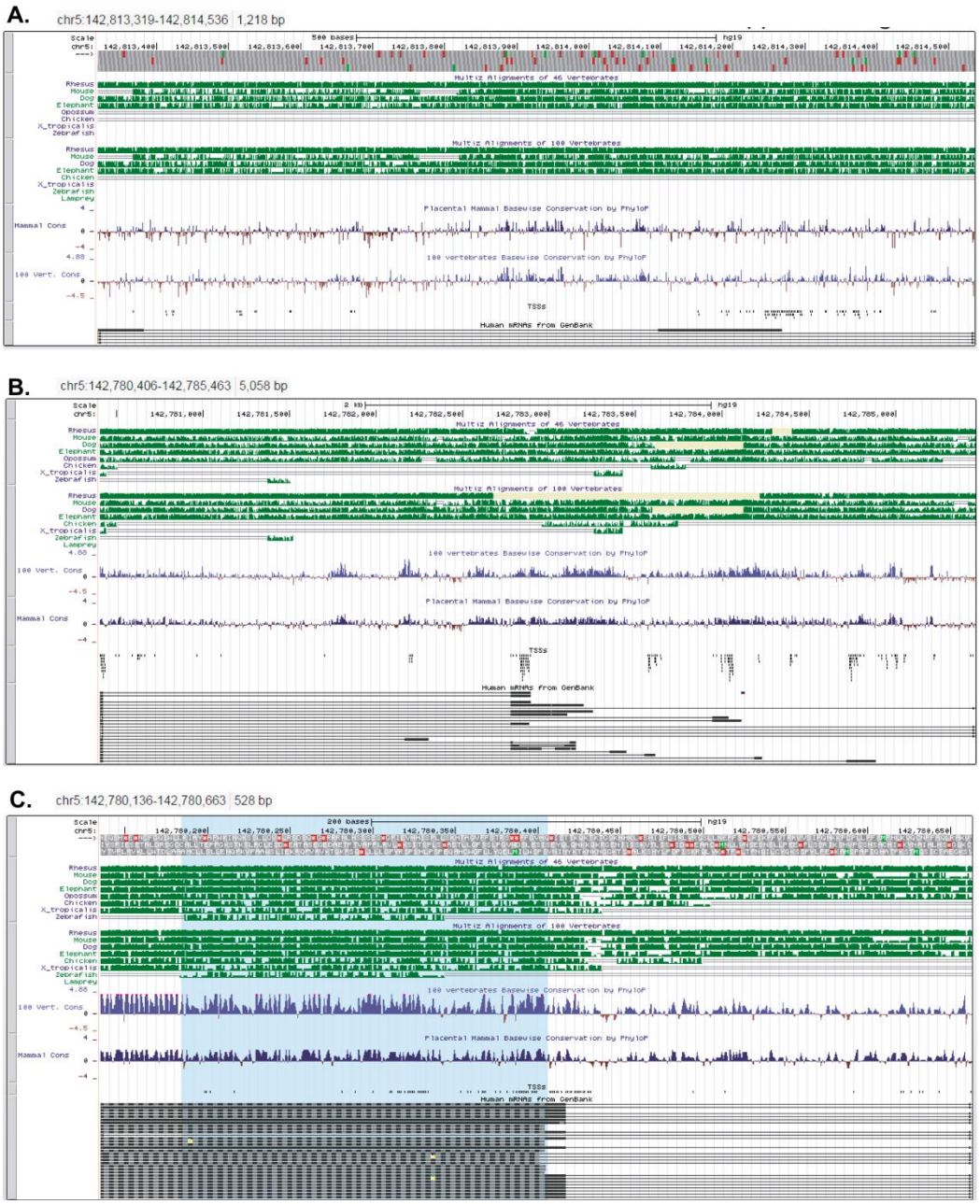

Figure 23: Evolutionary conservation analysis of the *NR3C1*'s 5'UTR using the UCSC browser. (A) The distal promoter region covering exon 1A. (B) The proximal promoter region containing the CpG island. (C) The end of the proximal promoter region and start of the common exon 2 (blue).[166] (Supplementary Data)

We detected a variability in TSS location that has so far received little attention and that appears to be due to a highly permissive transcriptional machinery. The combination of 5' mRNA cap labelling and HTS enabled us to identify clusters of TSSs (loci) consisting of 4-10 adjacent microvariable TSSs. The simple mono-exonic *ADR2BR* utilised a single locus consisting of 4 adjacent TSSs. The multi-promoter *NR3C1* gene targeted 358 TSSs distributed throughout 38 contiguous loci. We observed TSSs with frequencies from 6.74 x 10^{-4} % to 38.5% of the total 5' m^7G transcripts. This expands previous reports from ourselves and others of multiple alternative first exons that temporally, spatially, and quantitatively regulate mRNA levels and isoforms [153,162,164,182,183,204–206], suggesting an almost unlimited transcriptional variability. This transcriptional microvariability had a significant effect on the relative abundance of the final *NR3C1* N-terminal translational isoforms.

The experimental protocol was carefully designed to exclude other potential interpretation of these HTS results. Albeit mRNA extraction and 5' labelling were performed in the presence of RNase inhibitors, it would be conceivable that the 5' m^7G cap and adjacent nt may have been removed by residual RNase activity leading to an apparent ragged TSS pattern. This potential artefact was essentially excluded by the RNA oligo ligation strategy. First, CIP dephosphorylation removed all active 5' mono-phosphates from truncated or otherwise degraded mRNA as well as all other RNAs, leaving only intact capped mRNA unaffected. Subsequently, TAP pyrophosphatase treatment hydrolyses the pyrophosphate bonds in the m^7G cap triphosphate bridge leaving only mature, undegraded mRNAs with a 5' mono-phosphate that is available for oligo ligation [188–190]. Thus the combination of CIP and TAP exclusively labels 5' m^7G capped sequences [207]. The enzymatic mechanisms of both enzymes are well known from their high resolution crystal structures. The latter studies clearly demonstrated RNA binding and formally exclude exonuclease or phosphodiesterase activity [208]. Indeed, 5'-RACE is a well-established technique and incorrect identification of the 5' m^7G capped TSSs has not been previously reported [188–190,207,209,210]. To further reduce the possibility of artefacts due to high sequencing depth, we introduced a 0.1% cut-off to define a genuine TSS. This corresponds to the point of inflection on the frequency/cut-off curve, the point where the phases of the linear regression intersect, above which the numbers of identified TSSs did not significantly decrease (Fig. 24).

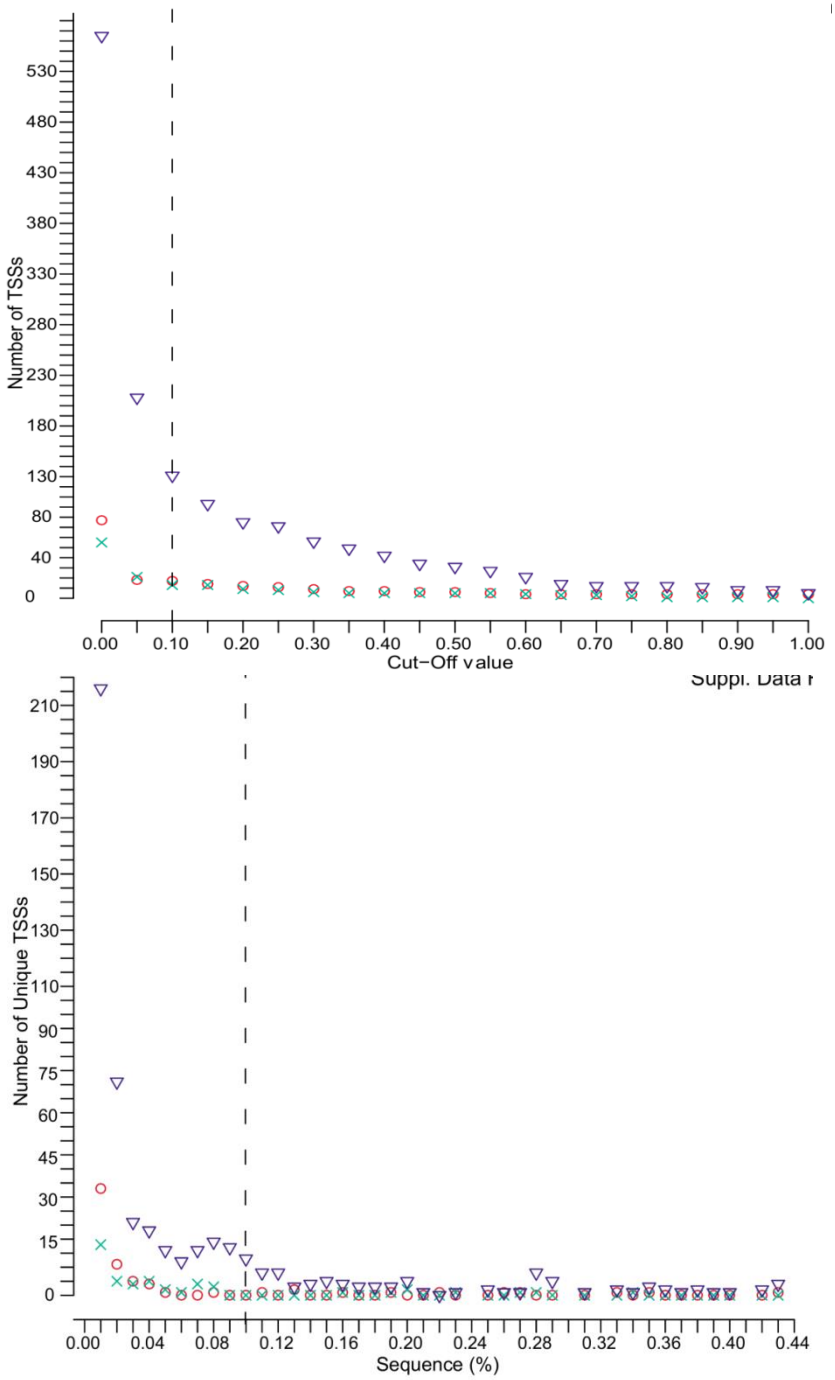

Figure 24: (A) The total number of identified TSSs from the *NR3C1* proximal CpG island are plotted as a function of the cut-off value (%) used to define a genuine TSS. Data are from A549 biological triplicates. (B) The number of unique replicate specific TSSs from the *NR3C1* proximal CpG island are plotted as a function of the cut-off value (%) used to define a genuine TSS. Data are from A549 biological triplicates. (Run 1, ○; Run 2, x ; Run 3: ▽)[166] (Supplementary Data)

Additionally, our experimental protocol was also designed to minimise any potential ligation bias. Both RNA and DNA ligases exhibit intrinsic sequence preferences, particularly with respect to the 3' end of the ligated 3'-5' pair [211–217]. By using a common RNA or DNA oligo with a constant 3' end, we effectively minimised the ligation bias within- and between- samples. Yet sample-specific MID usage, ligating a variable 3' end to the common 5'

end of the template, would still cause ligase induced bias to affect multiplex sequencing experiments [211,213,215]. Hence, sequencing one sample using different MIDs should not result in varying TSS usage profiles because of the common RNA-oligo (Fig. 5C-D), but rather in limited differences in overall relative read numbers because of biased efficiencies during library preparation [211,212]. Both *ADRB2R* and *NR3C1* data sets were screened for ligation bias. When samples were run repeatedly with different MID sequences, no discriminating or run-specific TSS expression profiles were detected. Therefore, we conclude that errors introduced during the sample preparations are minimal and below our 0.1% cut-off. Furthermore, comparisons of sequencing replicates allowed us to distinguish between the low technical variability and high biological variability observed in unstimulated DAUDI cells. Thus our judicious experimental set-up and control experiments demonstrated that the microvariability in TSS usage is a genuine biological phenomenon. Ragged 5'ends of mRNA transcripts so far have been reported in one keynote report in the case of the genes *Postn*, *Myh3* and *Fth1* (1) and recently the importance of alternative 5' end mRNA for transcriptome and proteome diversity has gained attention [168,178]. As such, TSS locations may be considered exact, however, the number of sequences at each position should be considered semi-quantitative, as the minimal bias introduced during the RACE PCR amplification cannot be quantified. HTS-RACE provides an accurate gene specific view of TSS location, and as such is complimentary to genome wide techniques such as CAGE [218], "gene identification signatures (GIS)" and "gene signature cloning (GSC)" [219]. These genome-wide techniques cover the complete transcriptional landscape at a cost of several orders of magnitude more sequencing data, whilst HTS-RACE provides greater insight, due to the greater sequencing depth it permits for single, often weakly expressed genes, as well as allowing longer sequence tags than MmeI digestion [218].

When our observations were compared to the previously published gene and mRNA structures, the simple mono-exonic gene *ADRB2R* had a small increase in complexity, from 1 TSS [184] to a locus of 4 adjacent TSSs. The multi-exonic *NR3C1,* however, displayed a ~30 fold increase in TSSs, going from 9 previously published TSSs [127,153,161–163,184,185] to 358 in 38 loci in the absence of any specific transcriptional stimuli. 66.7% of the newly identified *NR3C1* TSSs were located within the proximal promoter, 16.1% corresponded to the distal promoter, with the remaining 17.2% situated within exon 2. Dex and IFN-ɣ, both ligands of transcription factors activating the *NR3C1* promoter further induced transcription from 1 new locus and 185 additional TSSs distributed throughout the promoter region, giving a ~40 fold overall increase in the number of *NR3C1* TSSs. These TSSs are mostly located within loci utilised also in other cell lines. Thus transcription does not seem to be initiated at a well-defined, fixed, TSS, but their selection seems to move due to a more or less permissive transcription machinery. The effect of the transcription factor ligands Dex and IFN-ɣ may suggest that the transcription factor complex determines the start site of the transcription. Steric effects of binding of transcription factor complexes

to the promoter DNA may perhaps be the simplest explanation for TSS microvariability. Although the microvariability may appear to be to some extent stochastic, we demonstrate that, in the case of *NR3C1*, it also regulates translation. There are several potential scenarios. When microvariable TSSs are upstream of the principal ATG translation initiation codon, the complete coding sequence (CDS) remains available in the mRNA and can be translated, without effect on the protein sequences (Fig. 25). Splicing to the subsequent exon was never affected and consistently performed to the common splice acceptor site. However, we observed for the *NR3C1* that some TSSs were located downstream of the principal translation initiation codon in exon 2. These abridged mRNAs may produce N-terminally truncated protein isoforms from methionine-encoding ATG codons further downstream. Such truncated *GR* isoforms starting at alternative downstream translation initiation codons indeed exist, and are well known as *GR*-A, -B, -C and –D protein isoforms [171] Fig. 25). These internal translation initiation codons are available in the full length mRNA. In line with our previous study [164], where the alternative first exons altered mRNA folding stability, half-life, translation efficiency and protein isoform production in a length-dependent, but sequence-independent manner. We observed identical negative correlations for ∆G - mRNA length, ∆G – translational efficiency and mRNA length – translational efficiency to those in our previous report [164], suggesting that microvariability plays a similar role in the regulation of both transcript and protein levels. We were able to extend this, demonstrating that differences in TSS location of only a few nucleotides within a locus dramatically altered the fine balance between the different N-terminal *GR* isoforms. There are evermore reports of multiple active or alternative initiation codons within a mature mRNA, covering both leaky ribosome scanning and internal ribosome entry in plants [220,221] and mammals [222,223] as well as classical viral IRES. Given that internal ATG codons and methionines are ubiquitous[224,225], we suggest that our observation may apply ubiquitously throughout the transcriptome and throughout evolution. The evolution speed of genes with complex 5'UTRs is negatively correlated with their expression level and is also dependent on functional specialisation of the genes. With a high intron density being one of the characteristics associated with a slower evolution rate [177,226]. We examined the evolutionary conservation of the *NR3C1* CpG island and distal promoter region. As would be expected the non-coding regions were less highly conserved than the coding regions in exon 2. Since the microvariable TSSs did not coincide with either evolutionary conserved transcription factor binding sites or more generally regions of high conservation we suggest that whilst TSS microvariability and its functional consequences are most likely identical between species, the actual TSSs selected will be species specific. These minor changes in TSS dramatically altering the protein isoforms produced and their function may underlay the vastly inflated proteome, significantly increasing the variability from the limited genome to the proteome [164,168,178–181]. The microvariable transcription initiation is an additional mechanism in the spectrum of alternative transcription initiation mechanisms.

The demethylation agent AZA also had a profound effect on TSS usage. Complete de-methylation with AZA

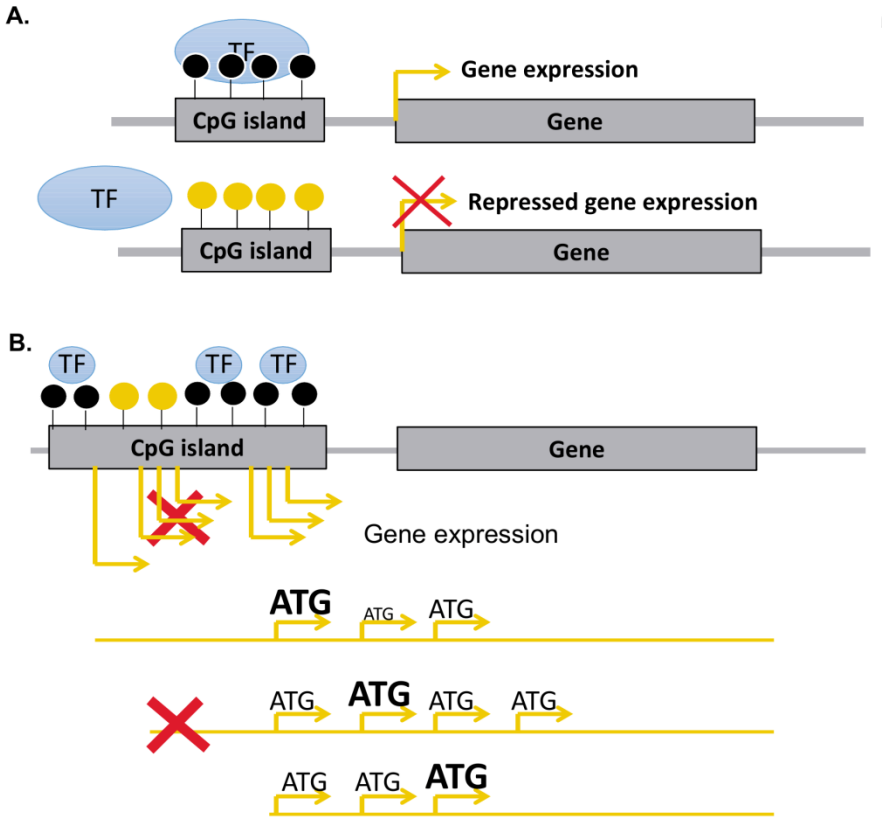

Figure 25: Schematic representation of gene transcription and translation and how it is influenced by methylation. (A) Transcription process. Methylated CpG island stops the transcription factor (TF) from binding and blocks consequently the gene expression. (B) Translation process. Depending on where the transcription initiated the translation start codon within the mRNA differs. (Methylated: ; Non-Methylated:)[166] (Supplementary Data)

induced 12 specific TSSs and another 115 that were also observed for either Dex- or IFN-ɣ- treated DAUDI cells or one of the other cell lines. One new locus was identified. In a similar manner to Dex and IFN- ɣ treatment, demethylation will alter the balance between the different *NR3C1* translational isoforms. Additionally, the vast increase in the overall number of TSSs raises doubts over the functional consequences of single CpG dinucleotide methylation for the regulation of *NR3C1* expression [185,227–229]. We would, however, anticipate that increased methylation levels over a larger cluster of CpG dinucleotides will influence total *NR3C1* levels by silencing one or more loci, concordant with our observations of methylation clusters in several models [153,155]. This raises the interesting hypothesis that the epigenetically controlled response to GC, as previously observed [27,104,127,153,163,230] is due to DNA methylation influencing TSS usage and altering the balance between translational isoforms. There are examples of internal methionine encoding ATGs also serving as secondary translation initiation sites in many vertebrate, invertebrate and plant species [220–223], making them amendable to TSS microvariability induced differential translation initiation.

In conclusion, we suggest that our observations of permissive microvariable transcription may also be expanded to other genes since many possess a structure similar to the *NR3C1* [167,168,170,177,178,182,183,204,231,232]. Our observations further suggest that TSS microvariability is not simply the result of a permissive transcription machinery, but rather a mechanism to fine-tune total protein levels via multiple mRNAs species that differ in stability and in some cases the relative distribution of protein isoforms. We showed that TSS usage can be influenced by transcription factor ligands. Similarly, DNA methylation seems to influence TSS selection, adding another mechanism by which covalent modifications of DNA can regulate gene expression to match physiological requirements.

2.6. Acknowledgements

We are grateful to Hartmut Schächinger for his initiatives within the Trier-Leiden International Research Training Group (IRTG GRK 1389/1), Graduate School of Psychobiology, and the Research Focus Point ("Schwerpunkt") within the University of Trier. We thank Sandra Gohrbandt and Regina Sinner (Institute of Immunology, Luxembourg Institute of Health) of the PGM Ion Torrent platform for their counsel and hands-on help.

2.7. Authors' Contributions

J.D.T. designed and coordinated the study. F.A.D.L., S.V., S.S. and A. M. M carried out the data acquisition. F.A.D.L and O.E.H. performed the data analysis. F.A.D.L. drafted the manuscript, which was critically revised by J.D.T and C.P.M. The study was performed under the supervision of J.D.T and C.P.M.

Chapter 3

Genome-wide 5mC and 5hmC DNA profiling of maternally deprived rats

Manuscript in preparation:

Genome-wide 5mC and 5hmC DNA profiling of maternally deprived rats

Fleur A. D. Leenen[1,2], Sophie B. Mériaux[1], Oliver E. Hunewald[1], Fanny Bonnemberger[1], Torbjørn Breivik[3,4], Claude P. Muller[1,2], Robert Murison[5] and Jonathan D. Turner[1*]

[1] Department of Infection and Immunity, Luxembourg Institute of Health, Esch-sur-Alzette, L-4354, Grand-Duchy of Luxembourg
[2] Department of Immunology, Research Institute of Psychobiology, University of Trier, D-54290 Trier, Germany
[3] Department of Periodontology, Faculty of Dentistry, University of Oslo, Oslo, Norway
[4] Norwegian Defence Research Establishment, Division for Protection, Kjeller, Norway
[5] Department of Biology and Medical Psychology, Faculty of Psychology, University of Bergen, Norway

3.1. Abstract

Early life experiences cause life-long alterations in an individuals' gene expression, which can affect their health later on in life. Exposing rat pups to maternal deprivation (MD) mimics traumatic early life experiences, and induces behavioural, immunological, as well as gene expression changes lifelong. Expanding on our previous report of increased *Nr3c1* mRNA levels, complete 5mC and 5hmC epigenomes were obtained from the hippocampus of rats exposed to MD, handling (HD) or left undisturbed (C) in early life. The 5mC and 5hmC profiles displayed a high concordance between treatments and cytosine modifications. Overall MD or HD induced methylation changes were small, however 50 and 71 clearly defined regions featuring small but distinctive changes in 5mC and 5hmC levels respectively, of which the majority resided in intergenic regions. For stress-related genes including *BDNF*, *AVP*, *CRH*, *NR4A1* and *Igf2*, differential methylation was observed >200kbp from the corresponding genomic loci. However, detailed examination of the region on Chr 18 surrounding *Nr3c1* identified 5 D(h)MRs located significantly further away, approximately 5.7Mbp downstream of the *Nr3c1* locus. Overall, our data show that MD and HD induced epigenetic modifications in regulatory regions more than >200kbp from the closest annotated gene. However, differentially methylated regions appeared to be significantly closer to loci associated with piRNA transcripts. These piRNAs are short regulatory RNAs that have been previously associated with cellular homeostasis, phenotype change, and CNS variability and mosaicism important in the neuronal regulatory system. A gene ontology analysis of the functionally gene annotated D(h)MRs linked MD and HD induced methylation changes to the oxidative phosphorylation, purine metabolism, mitochondrial dysfunction pathways and Huntington's disease. Our data shows that MD and HD induced clear differential methylation that was associated with regulatory RNAs rather than gene loci, and phenotype differences in our maternal handling model were most probably due to post-transcriptional processes rather than direct regulation of specific target genes.

3.2. Introduction

Epigenetic modifications play a crucial role in orchestrating temporal, cell and tissue specific gene expression patterns during normal development and environmental adaptation [12,13,33,34,36,37,59,66,70,233]. One of the best studied epigenetic modifications is DNA methylation (5-methylcytosine, 5mC), which consists of the addition of a methyl group to the cytosine in a palindromic CG dinucleotide, and is known to influences expression patterns through chromatin re-modelling, gene transcription regulation, maintenance of X-chromosome inactivation, gene imprinting and tissue-specific gene expression [13,33–38]. Historically, 5mC levels were considered to be relatively stable throughout time, and in differentiated cells, demethylation was a passive process that was a direct consequence of the failure to maintain methylation levels during cell division. However the dynamic nature of DNA methylation coupled with active demethylation are now accepted [66–68]. Although the exact function of 5-hydroxymethylcytosine (5hmC) has not been elucidated fully, it is gaining acceptance that it marks the first step in DNA demethylation [36,70]. the idea of 5hmC acting as an intermediate in an active enzymatic demethylation process gains more and more acceptance [36,37,66,69,70,234]. 5hmC is generated by oxidation of 5mC by one of the ten-eleven translocation (TET) proteins 1-3. Which sequentially can be oxidised to either 5- formylcytosine (5fC) or 5-carboxylcytosine (5caC). The latter two can be disposed of via processes, either including a base excision or not [36,66,69]. However, although 5hmC has been detected in nearly all tissues, it is particularly enriched in the central nervous system (CNS), where is thought to also be involved in active transcription of neuronal genes, and to have an impact on brain development [37,66,70].

Differences in both 5mC and 5hmC generate phenotypic diversity and have been implicated in the onset of a broad spectrum of mental and physical disorders [12,62,70,233,235,236]. For instance, early life experiences are known to cause life-long alterations in an individuals' gene expression. This changes behavioural responses, the neural system and health later on in life, which coincides/is consistent with the developmental origin of health and disease (DOHAD) hypothesis [22,62,94,123,155,235,237]. Although we are constantly influenced by the environment, there are several periods/windows of increased epigenetic sensitivity and susceptibility to environmental cues. Known sensitive periods are the embryonic and foetal development, postnatal period or early life, and adolescence [12,13,22,32,238]. An important factor during early life, is the quality of parental care, known to affect mental health, brain function, and behaviour into adulthood, for both rats and humans [25,62,123]. Predominant models for assessing early life social life's impact, are the licking/grooming (LG) model and the maternal deprivation (MD) model in rats [25,62,104,123,155]. The LG model uses natural differences in the quality of maternal care, i.e. the intensity in licking and grooming, and demonstrated that these differences affect the neural system, including functions such as stress response, which persists into adulthood [25,104,123]. Offspring that experienced

poor maternal care, showed higher *Nr3c1* methylation levels and an increased hypothalamic-pituitary -adrenal (HPA) response/reactivity towards environmental stress compared to offspring that received high maternal care [62,104,233]. The more-widely used MD model, where postnatal psychosocial stress is induced by daily separation of pups and dams for an extended time, installs long lasting behavioural effects, caused changes in the HPA axis and hence the stress reactivity [155,239,240]. In mice for instance, MD induced DNA methylation changed in promoter regions of genes *AVP*, *NR4A1* and *Nr3c1*. Three genes implicated in the HPA stress response circuit, hence altering the stress reactivity into adulthood [240]. Breivik et al (2015) showed that maternal deprivation during early life affected the *Nr3c1* expression in rat hippocampi and impacts behavioural and immune responses later on in life [155].

Initial hypotheses assumed that the effect of early life experiences would mainly affect brain-specific genes [62,233]. The glucocorticoid receptor (*NR3C1*, *Nr3c1,* GR), a key player in the HPA axis negative feedback loop, has been intensively studied in the context of early life experiences, epigenetic mechanisms and phenotypic diversity resulting from it [25,104,123,127,155,241], often focussing on the gene's promoter region. As extensively discussed in Leenen et al (2016) [242], *NR3C1*'s 5'UTR displayed an almost unlimited transcriptional variability. 5' UTR methylation pattern changes impacted transcription start site usage, and consequently reshaped the translational and proteomic landscape [242]. Throughout the brain, *NR3C1* promoter methylation and expression patterns were shown to be ubiquitous and consistent [105,127,243]. Both patterns were individual specific, with the expression correlating to the overall CpG methylation pattern [127]. Despite the main focus on the promoter region, DMRs were shown to be distributed non-randomly across the whole length of *NR3C1*, including intragenic regions or regions distantly located from transcription start sites [25,123]. Research assessing differential methylation in *NR3C1* in e.g. T-cells and cord blood suggested a system-wide response [62,233,235,244]. A more wide-spread epigenetic programming via a network of genes is also supported by the discovery/presence/existence of additional genes such as *BDNF* and *AVP*, affected by maternal care [25,62,123,233]. Therefore, in this study the *Nr3c1* methylation analysis carried out in Breivik et al (2015) [155], was expanded using a genome-wide approach, enabling a deeper insight in the complex genome-wide response and how subtle methylation changes within such gene networks result in different phenotypes.

3.3. Material and Methods

3.3.1. Animals & Maternal Deprivation Procedure

Animal housing and the maternal deprivation procedure were previously reported [155,245]. Briefly, from each litter, one or two rat pups were randomly and equally assigned to either the handling (HD), maternal deprivation (MD) or control (C) group (Fig. 26). Between postnatal day (PND) 2 and 14, HD and MD pups were removed from their mothers, for 15 minutes or 3 hours respectively at a fixed time every day. Control pups were left undisturbed. Terminal anaesthesia was performed at fourteen weeks of age with Hypnorm-Dormicum (fentanyl/fluanizone, midazolam; 0.2 ml/100g body weight) (Fig. 26).

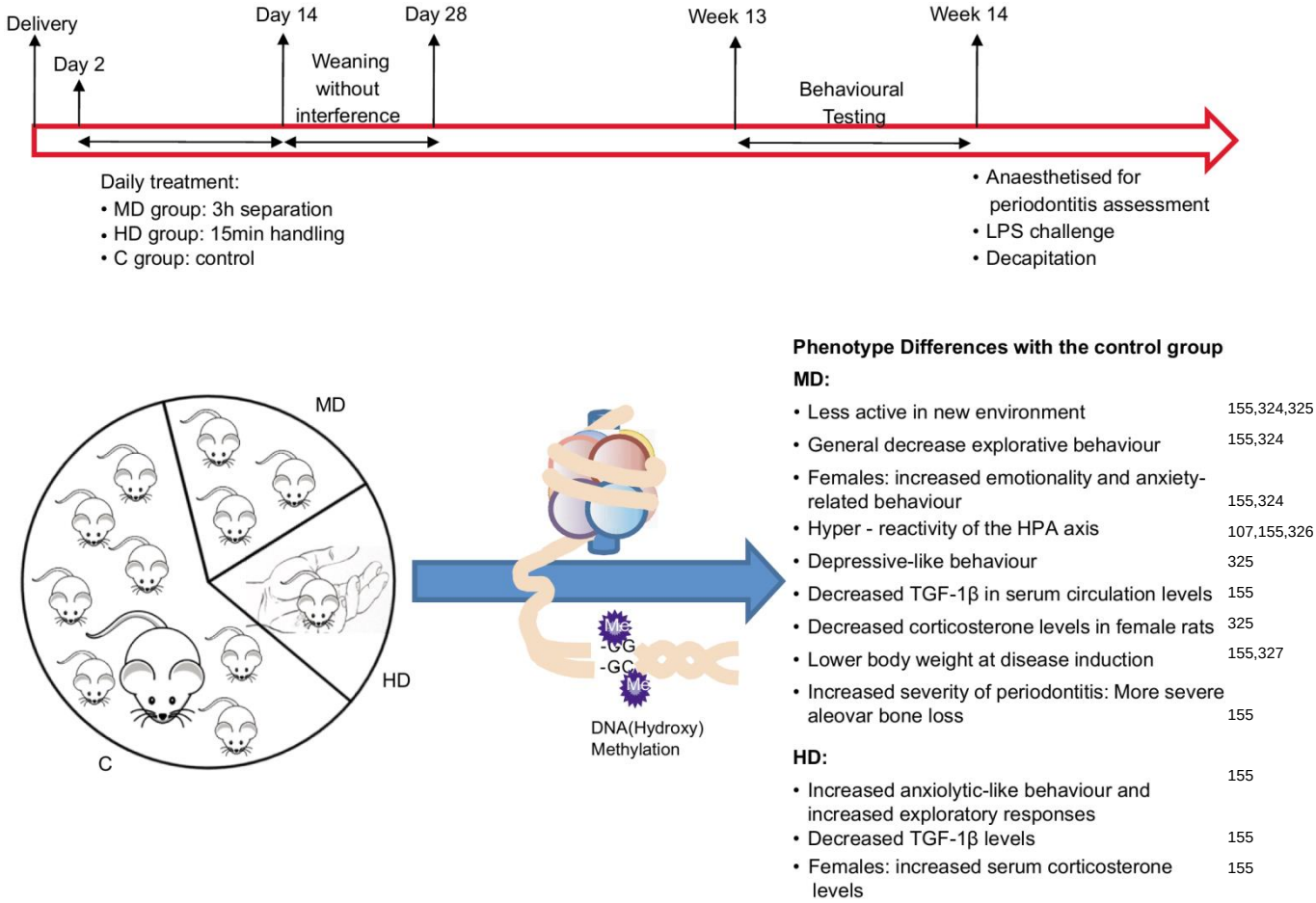

Figure 26: A schematic representation of the Maternal Deprivation (MD) model, its workflow and the phenotypic outcomes. C: Control; HD: Handling; MD: Maternal Deprivation

3.3.2. DNA Extraction and (Hydroxy)Methylated DNA Immunoprecipitation (MeDIP)

Immediately after euthanasia, bilateral hippocampi were removed and genomic DNA (gDNA) extracted using the AllPrep® DNA/RNA Mini kit (Qiagen, Venlo, Netherlands) according to the manufacturer's instructions. DNA concentrations were measured using a NanoDrop Spectrophotometer (ND1000, Isogen Life Science, De

Meern, The Netherlands). The 5-(h)mC changes were expected to be small. Detection of such small modifications benefits from low-variance data, which can be obtained by pooling multiple samples per group. According to the equation of the variance of the estimator of the true distribution mean $\theta = 1/n_p * \left((\sigma_\varepsilon^2)/r_s + \left(\sigma_\xi^2\right)/r_a\right)$, with θ the true distribution mean, n_p the total number of pools, r_s the number of samples per pool, r_a the number of sequenced samples per pool, σ_ε^2 the biological variance and σ_ξ^2 the technical variation, 2 or more samples pooled cause a decrease in variance which leads in turn to an increase in power to identify differentially methylated regions [246–248]. As such, extracted genomic DNA was pooled from all animals within the same treatment group.

MeDIP-Seq was performed as previously described [61]. Briefly, duplicate samples of 2.5µg of gDNA, diluted in 100µl TE buffer, were sonicated to an average length of 200-450bp (Bioruptor® UCD-200, Diagenode, Liège, Belgium). Fragment length and concentration were measured on a high sensitivity bioanalyser chip (Agilent Technologies, Diegem, Belgium). All samples were denatured (95°C for 10min, immediately on ice for 5min). Dynabeads (50µl; M-280 Sheep-anti-Mouse IgG, Invitrogen) were coated with either anti-5-methylcytosine (1µg/µl; clone 33D3, Eurogentec, Seraing, Belgium), or anti-5-hydroxymethylcytosine (1µg/µl, clone 4D9, Eurogentec) antibodies. Denatured DNA was incubated with antibody coated beads in a final volume of 100µl 2x IP buffer (0.05% Triton in PBS, 50µg/ml Yeast tRNA, 1xPBS) for 16h25 at 4°C. Beads were washed twice with 200µl; 0.025% Triton in PBS, followed by two stringent washes using 200µl 0.05% Triton in 1x PBS. DNA was eluted from the beads, by incubating each sample with 7µl proteinase K (10mg/ml, QIAGEN, Venlo, Netherlands) in 200µl digestion buffer (50mM Tris, 10mM EDTA, 0.5%SDS, pH8) for 3hours at 50°C, while being shaken (1400rpm). Duplicate samples were pooled, purified by PEG8000 precipitation (Agencourt®AMPure® XP, Beckman Coulter, Belgium) and the recovered methylated DNA was stored at -20°C until further analysis. Single stranded immunoprecipitated DNA was quantified with a Qubit® ssDNA Assay Kits (Life Technologies) using a Qubit 2.0 Fluorometer (Life Technologies) according to the manufacturer's instructions.

3.3.3. MeDIP Library Preparation and Sequencing

NGS libraries were prepared using the Accel-NGS® 1S Plus DNA library Kit for Illumina® Platforms (Swift Biosciences, MI, USA) according to the manufacturer's instructions. Briefly, 20 µl (4 µg) fragmented and immunoprecipitated DNA was denatured by incubation at 95°C for 2 minutes, and immediately quenched on ice for 2 minutes. After denaturation, DNA was end repaired, 3'end tailed and a truncated adapter ligated to the 3' end. Subsequently, the DNA was elongated, facilitating the ligation of a second truncated 5' adapter. Index primers (I-IL1SP-12A, Swift Biosciences) were added by PCR using the supplied reagents. Cycling conditions

were 98°C for 30s, followed by 8 cycles of 98°C for 10s, 60°C for 30s and 68°C for 60s. The resulting libraries were stored at 4°C until sequencing. Library length distribution was assessed (Bioanalyzer, Agilent Technologies), and quantified (KAPA Library Quantification Kit for Illumina® platforms, KAPABIOSYSTEMS, Massachusetts, USA) prior to 2x125bp paired-end sequencing on an Illumina HiSeq 2500 using the HiSeq SBS V4 chemistry, multiplexed within a single lane by Genewiz (New Jersey, USA). Due to this balanced block sequencing, possible confounding factors due to batch or lane effects are eliminated [249].

3.3.4. Bio-Informatics and Statistical Analyses

The NGS sequencing reads were processed through the CASAVA (v1.8.2, Illumina) pipeline using default settings for standard Basecalling, primary data quality control, demultiplexing and fastq generation. Data were subsequently filtered for a minimum length of 50 nucleotides, and quality trimmed using cutadapt (v1.8.3) [250,251] under default settings. The read quality of the resulting/cleaned fastq files was verified with FastQC (v0.11.4) [251,252]. Reads were aligned against the rat genome (RGSC 6.0/rn6) with Bowtie 2 (v2.1.0) [251,253] and SAMtools (v1.2) [251,254] using default settings.

The mapped read distribution genome-wide was assessed using SeqMonk (v0.34.1, Babraham Institute), by segmenting the genome into 500nt wide windows with 250nt overlap. The read counts/segment were quantified and corrected for the total counts. The CpG and genomic context of the mapped reads were assessed with a custom Python script (Python v2.7, Stichting Mathematisch Centrum Amsterdam, The Netherlands) as previously described by Kirschner et al (2016) [255]. Briefly, based on their CpG density the mapped reads were assigned to either CGI, shore, shelf or open sea. CGI's were characterised as regions of a minimum of 200nt, a GC fraction > 0.5 and an CpG observed-to expected ratio > 0.6. This information and the CGI positions were obtained from the UCSC rn6 CpG Island database. Shores were regions up to 2kb from CGI's, shelves were regions between 2 and 4kb of the CGI's and open sea coincides with regions that are 4kb or further away from CGI's. The genomic context was initially defined as enhancer, promoter, intra- and intergenic regions. A second stratification subdivided the promoter region into 5'UTR, 200bp upstream of TSS and 200-1500nt upstream of the TSS and, the intra-genic region into 1st exon, gene body and 3' UTR. An Ensembl gene and transcript ID were assigned to each fragment. Repetitive elements were analysed with the RGSC 6.0/rn6 UCSC Repeat Masker database.

5(h)mC profiles and differential (hydroxy)methylation (D(h)M) analyses were assessed with the Bioconductor package MEDIPS [256] (v1.20.1) in R (v 3.2.3.) [193], with a window size of 500nt, the minimum read coverage defined as 4% of the window size, extend all reads to a total length of 300nt, replacing stacked reads by one representative (uniq=1), and the paired reads parameter set to true. The MEDIPS profiling and differential

coverage was performed, assuming a negative binomial distribution, using the differential coverage calculation method edgeR at an adjusted p-value of 0.1 (Benjamini-Hochberg). MEDIPS bins the genome according to a fixed window size. To detect the best parameter settings for an optimal trade-off balance between resolution, and noise level and multiple testing burden, several window widths (100 to 900nt) and minimum coverage (1% to 10%) values were tested. The optimal results were observed for a window size of 500nt and a minimum read coverage of 4%, which were used for all subsequent D(h)M analyses. The 5mC and 5hmC profiles were visualised with the UCSC browser (https://genome-euro.ucsc.edu/). Each D(h)MR was assigned an Ensembl Gene ID, Entrez Gene ID or transcript ID when possible (Rnor_6.0, INSDC Assembly GCA_000001895.4). The CpG and genomic context of the D(h)MRs were assessed with a custom bash script that compared the D(h)MRs lists with the previously generated tables containing the CpG and genomic characteristics of the mapped reads.

A functional analysis of all DMR containing genes was performed using Ingenuity Pathway Analysis (IPA, QIAGEN Redwood City, USA, www.qiagen.com/ingenuity) and DAVID (the Database for Annotation, Visualisation and Integrated Discovery) [257,258].

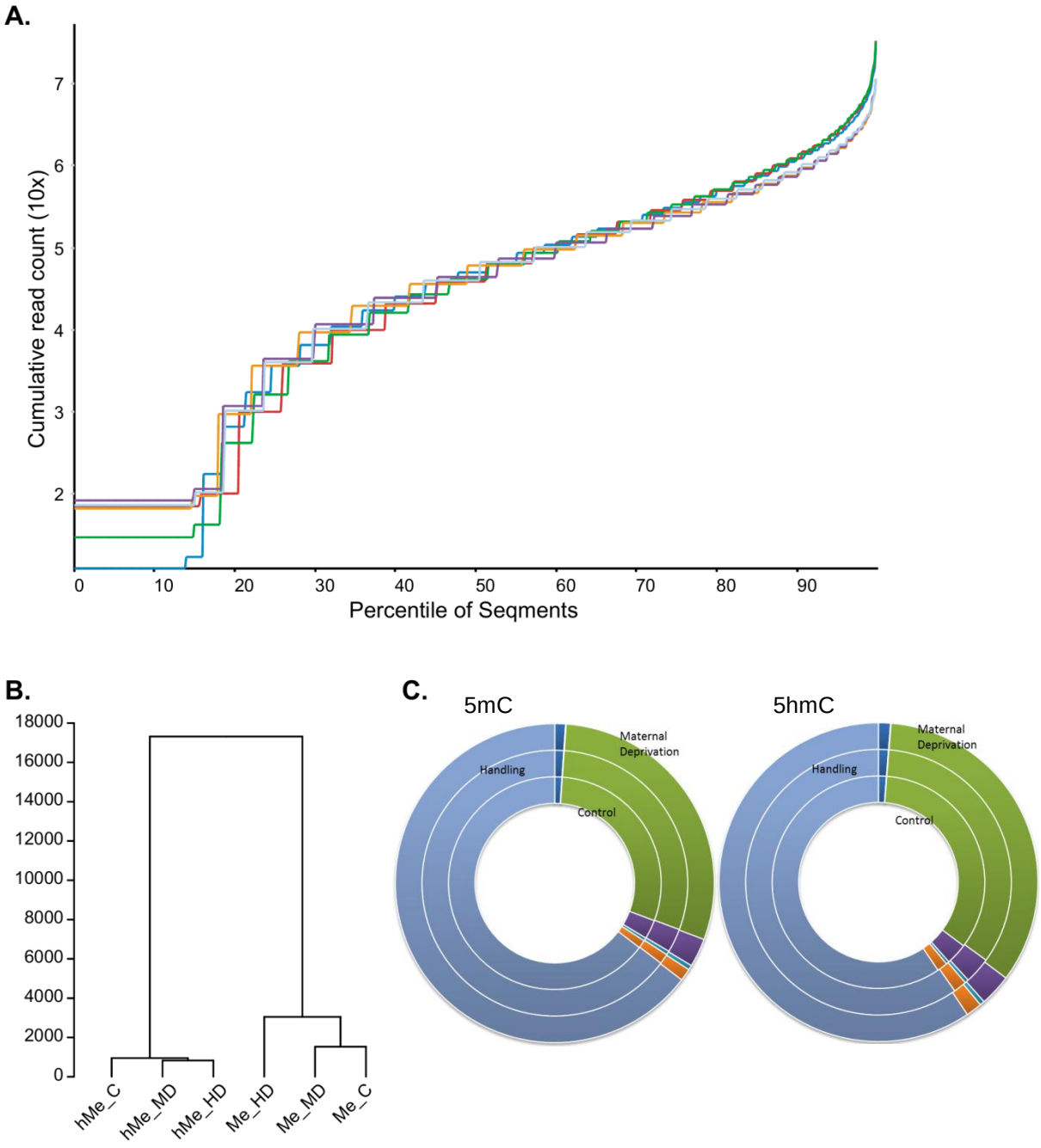

Figure 27: Global methylome and hydroxymethylome comparison. (A) A segment trend plot depicting the log transformed accumulated read numbers pas a function of all segments ranging from the lowest to the highest covered percentile of all 3 groups, for both hydroxy- and methylation. (B) Hierarchical clustering of the treatment groups. C) Infinium gene contexts annotation of the enriched (hydroxy)methylated DNA, sequenced, and aligned to the rat genome. (…), TSS1500; ( ) TSS200; (…) gene body; ( ) UTR; ( ) 1st exon; ( ) 3'UTR; ( ) inter-genic region. Multiple annotations per fragment were possible.

3.4. Results

3.4.1. Experimental Paradigm

The animals used in these experiments have previously been reported [155]. MD exposure from PND 2 to 14 induced more severe periodontitis, a lower weight gain, a significant decrease in T-regulatory cytokines, an increased emotionality and anxiety-related behaviour towards novelty. HD exposed rats, on the other hand, only exhibited altered emotional and anxiety-related behaviours. These previous data also identified methylation differences in *Nr3c1*. HD exposed rats had increased methylation levels in for specific *Nr3c1* regions, whereas the anticipated changes in methylation levels for MD exposed rats remained undetected.

3.4.2. MeDIP-Sequencing

Library Preparation and Sequencing: Genomic DNA was extracted from hippocampi of 21 rats, covering the C, HD and MD groups. Six to eight independent samples were pooled for each experimental group. After preparing sequencing libraries on these pooled DNAs, MeDIP libraries were multiplexed on an Illumina HiSeq 2500. Sequencing yielded 63 309 702 to 130 364 176 raw reads (Table 4). Removal of low-quality and too-short reads reduced read numbers by 1% - 8.5%, resulting in 62 577 420 to 118 998 876 cleaned reads per sample. Initial read analysis showed PHRED values >35 over the entire read and a nucleotide coverage skewed towards GC as expected for all libraries. Cleaned reads were aligned to the rat reference genome (rn6). The overall mapping value was 29.90, between 59 577 460 to 114 343 729 (~95%) successfully mapped reads (Table 4). This included 84.64% (52 367 792 to 93 239 635) unique mappings; 2.36% (1 502 520 to 2 732 723) singletons, and 0.27% (107 841 to 414 095) of MAPQ>5 pairs located on different chromosomes (Table 4). Cumulating the read counts over the different genome percentiles, resulted in 6 similar cumulative paths (Fig. 27A). The paths were largely identical between the data sets, suggesting that no major technical or biological bias had been introduced. All raw and aligned sequencing data are available on the European Nucleotide Archive (ENA) of the EMBL-EBI under the accession number XX.

Table 4: Sequencing summary

Epigenetic Modification	Group	Raw read number	Cleaned read number	Mean PHRED score	Mapped Reads	Mean MAPQ	Unique mappings	Singletons	mate mapped to a different chr (MAPQ>=5)
5mC	C	98 581 624	91 026 104 (92.34%)	36	87 495 523 (96.12%)	28.31	71 274 856 (81.46%)	2 089 147 (2.30%)	318 126 (0.35%)
	HD	130 364 176	118 998 876 (91.28%)	36	114 343 729 (96.09%)	28.36	93 239 635 (81.54)	2 732 723 (2.30%)	414 095 (0.35%)
	MD	75 319 068	70 164 228 (93.16%)	36	67 457 583 (96.14%)	28.17	54 850 791 (81.31%)	1 620 999 (2.31%)	236 744 (0.34%)
5hmC	C	65 372 880	64 725 026 (99.01%)	36	61 674 105 (95.29%)	31.55	54 193 115 (87.87%)	1 566 311 (2.42%)	132 284 (0.20%)
	HD	67 390 058	66 715 602 (99.00%)	36	63 502 119 (95.18%)	31.47	55 718 931 (87.74%)	1 637 625 (2.45%)	128 787 (0.19%)
	MD	63 309 702	62 577 420 (98.84%)	36	59 577 460 (95.21%)	31.55	52 367 792 (87.90%)	1 502 520 (2.40%)	107 841 (0.17%)

Identification of Methylated regions: window size and minimal coverage requirement: To balance resolution and noise against the multiple testing burden when assessing the (hydroxy)methylome using MEDIPS, a series of optimal window width and read coverage values were tested (Fig. 28-31). The D(h)MR number, as well as the windows to test, decreased for an increasing window width, stabilising for window widths ≥ 500nt (Fig. 28-29). For a fixed window width, the D(h)MR number remained constant across the different read coverage values and the number of windows to test declined for increasing read coverage values (Fig. 30-31). The latter stabilised for read coverage values ≥ 4%. Subsequent methylation profiling and D(h)MR analyses were carried with a 500nt window and minimal coverage of 4%, which decreased the multiple testing burden, while maintaining enough sensitivity detecting small changes in 5(h)mC levels.

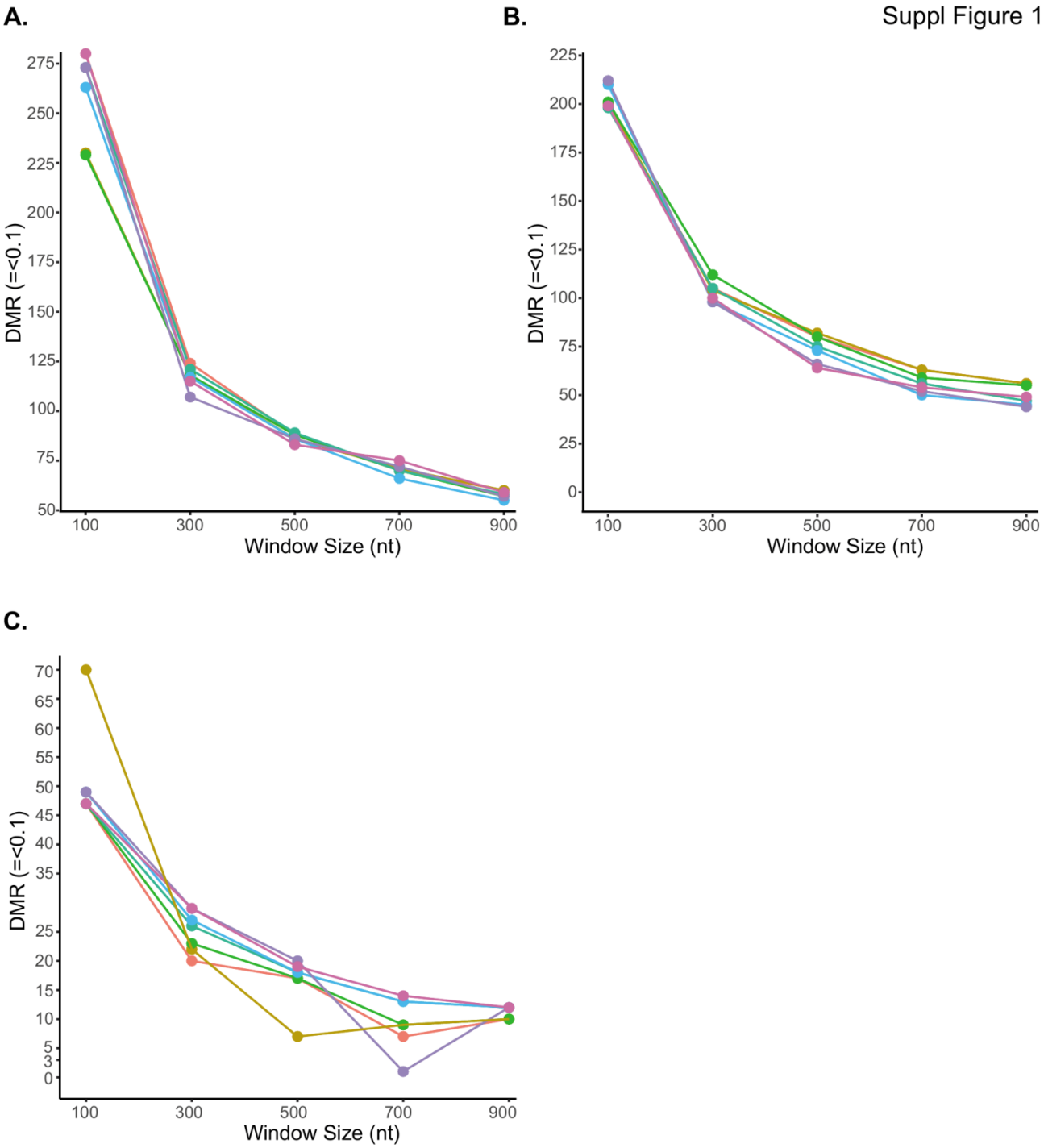

Figure 28: MEDIPS parameter settings, balancing resolution and noise against the multiple testing burden, assessing the number of differentially methylated regions (DMR) in function of the optimal window width per read coverage category. (A) Handling-Maternal Deprivation comparison (B) Control-Handling comparison (C) Control –Maternal Deprivation comparison.10 reads (.); 1% (.); 2% (.); 4% (.); 6% (.); 8% (.); 10% (.).(Manuscript Supplementary Data).

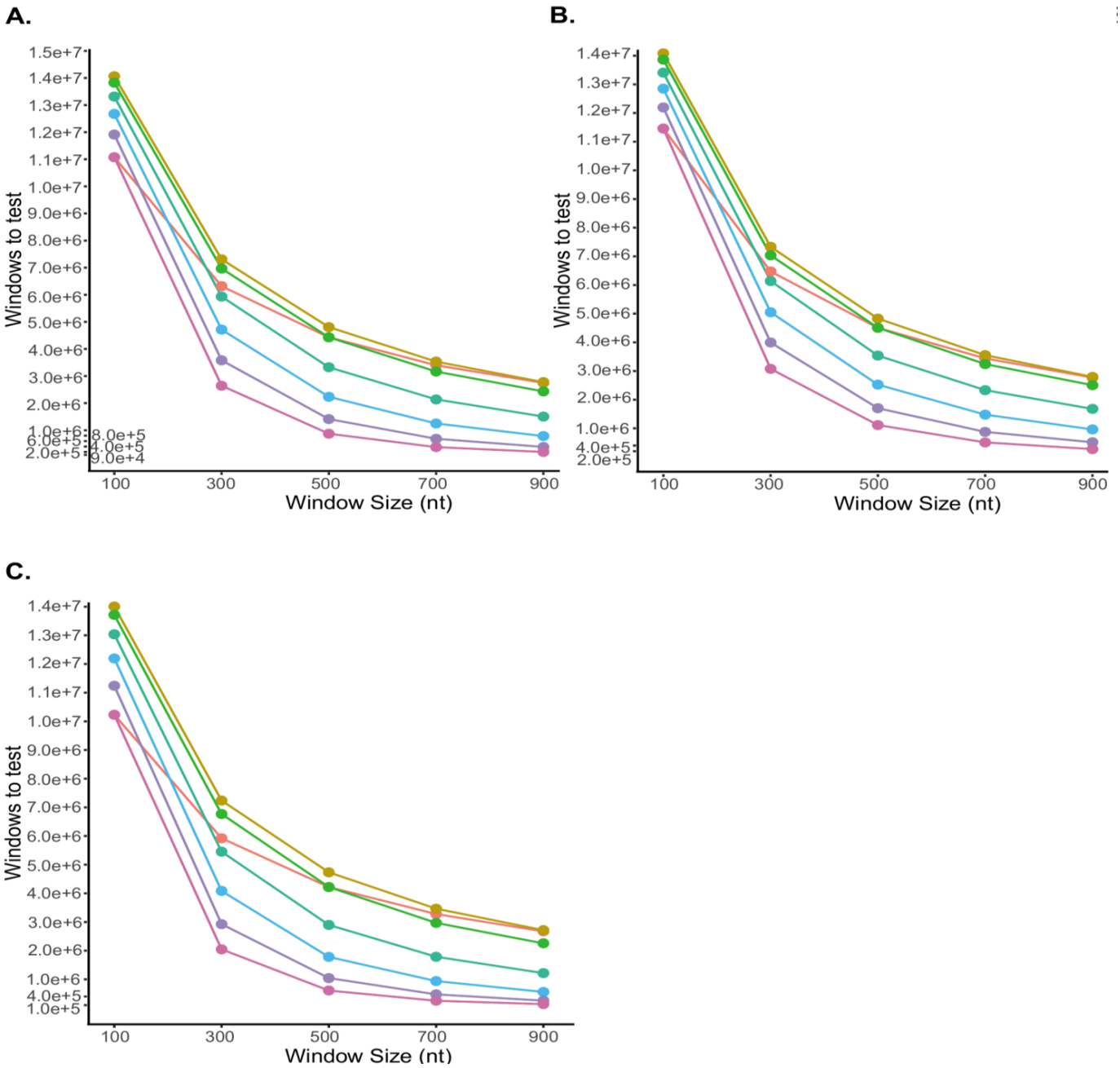

Figure 29: MEDIPS parameter settings, balancing resolution and noise against the multiple testing burden, assessing the number of windows to test in function of the optimal window width per read coverage category. (A) Handling-Maternal Deprivation comparison (B) Control-Handling comparison (C) Control –Maternal Deprivation comparison. 10 reads (…); 1% (…); % (…); % ( ); % (…); % ( ); % (…) (Manuscript Supplementary Data).

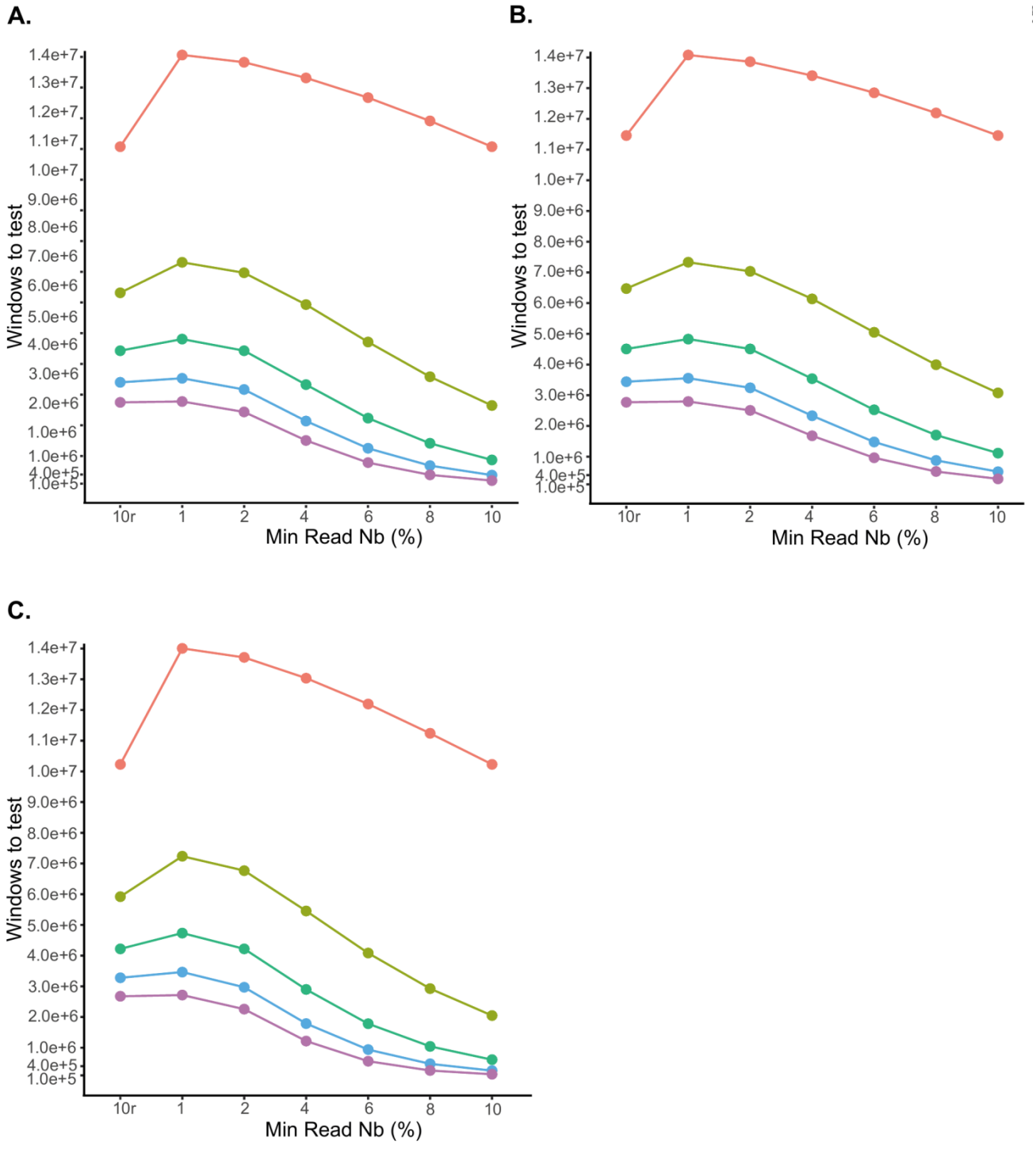

Figure 30: MEDIPS parameter settings, balancing resolution and noise against the multiple testing burden, assessing the number of windows to test in function of the minimal read coverage per window width category. (A) Handling-Maternal Deprivatiopn comparison (B) Control-Handling comparison (C) Control –Maternal Deprivation comparison. 10 reads ( ); 1% ( ); 2% ( ); 4% ( ); 6% ( ); 8% ( ); 10% ( ). (Manuscript Supplementary Data).

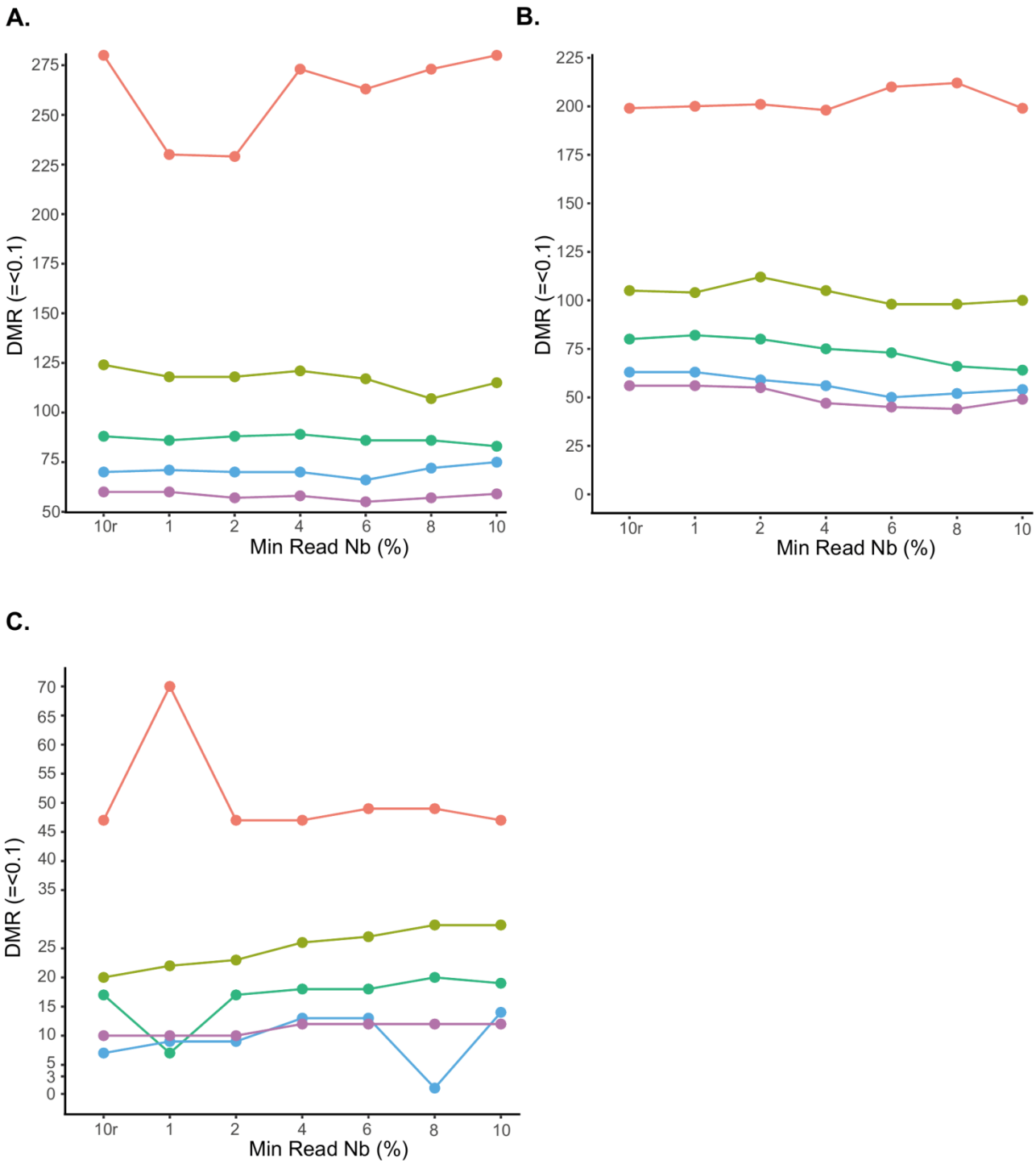

Figure 31: MEDIPS parameter settings, balancing resolution and noise against the multiple testing burden, assessing the number of differentially methylated regions in function of the minimal read coverage per window width category. (A) Handling-Maternal Deprivatiopn comparison (B) Control-Handling comparison (C) Control –Maternal Deprivation comparison. 10 reads (.); 1% (.); 2% (.); 4% (.); 6% (.); 8% (.); 10% (.).(Manuscript Supplementary Data)

3.4.2 Genome-wide Distribution of (Hydroxy)Methylation

Distribution patterns: 5mC and 5hmC profiles were analysed with a domainograph, which depicts the read density along the different chromosomes (Fig. 32). The read density across the genome was overall identical for the three experimental conditions, indicating that the same regions tended to be methylated or hydroxymethylated across treatments. Genome-wide visualisation of (hydroxy)methylation profiles in UCSC confirmed this observation. Across treatments and antibodies, the same genomic regions tend to be methylated or hydroxymethylated, which is to be expected seen their biological connection. Although the six groups displayed highly similar read density profiles, they were hierarchically clustered into two main fractions according to the antibody immunoprecipitation performed (Fig. 27B). Within the 5mC cluster, C and MD displayed more similarity than the HD group. For the 5hmC branch, HD and MD were more closely associated compared to C. Hence, suggesting that despite their similarity, differences detected in 5hmC profiles across treatments were much smaller than for the 5mC profiles (Fig. 27B).

Distribution throughout the genome: Similarly to the read dispersion described in Su et al 2016, [259] the reads cover regions genome-wide, i.e. regions within all chromosomes. As expected, 5mC or 5hmC immunoprecipitated regions were identified from their higher sequencings depths. Analysing the read distribution across different CG rich regions showed that the majority of the reads, for both 5mC and 5hmC, were located within the intergenic regions (65.58%; 34 894 856–66 856 958 fragments), followed by the gene body (33.03%; 18 457 776–30 524 836 fragments) (Fig. 27C). Only a smaller fraction of the reads aligned to 5'UTRs (3.19%; 1 807 078-2 990 388 fragments), 3'UTRs (1.47%; 776 421-1 287 506 fragments), 1st exon regions (0.49%; 269 792- 482 005 fragments) and regions located 1500 to 200nt (TSS 1500; 1.18%; 670 818-1 108 231fragments) and up to 200nt (TSS 200; 0.07%; 36 029-70 783 fragments) upstream of TSSs respectively. The lowest coverage was detected for the TSS 200 regions (Fig. 27C).

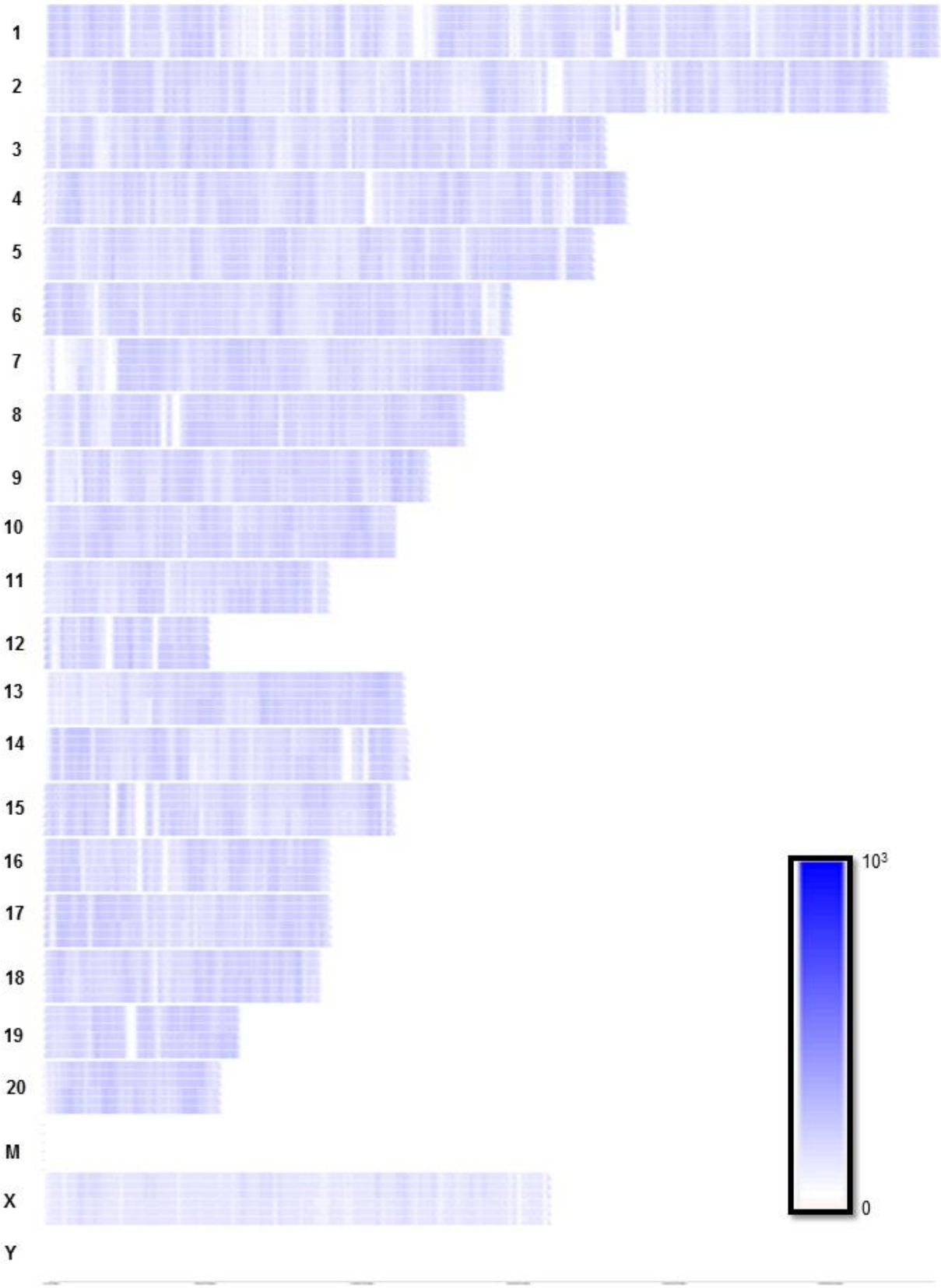

Figure 32: Genome-wide domainograph depicting the absolute number of reads per segment along the chromosomes, from zero (white) to multiple segments (dark blue), for the three different treatment groups (C, HD and MD), for both 5mC and 5hmC.

Cross-Reactivity: To study the methylation and hydroxymethylation pattern genome-wide and the impact of early life experiences on such patterns, DNA was immunoprecipitated with either anti-5-methylcytosine or anti-5-hydroxymethylcytosine antibody, two antibodies targeting very similar molecular structures. Previous data suggests that both antibodies exhibit a high target sensitivity and specificity[55,236,260,261]. In our initial analysis, we identified approximately 5 million enriched genomic regions for both 5mC and 5hmC in all three experimental conditions (Fig. 33). As expected, many of the 5mC and 5hmC enriched windows overlapped, 97.3%, 97.9% and 97.0% of 5mC and 5hmC immunoprecipitated regions were shared within the three groups respectively (Fig. 33A-C). Although approximately 97.4% of the regions coincided for methylation and hydroxymethylation, numerous windows were unique for either 5mC or 5hmC. The C group revealed 70 637 (1.3%) unique 5mC and 70 587 (1.3%) unique 5hmC enriched windows (Fig. 33A), 66 970 (1.3%) and 42 904 (0.8%) were unique for 5mC and 5hmC respectively under HD conditions (Fig. 33B), and MD displayed 58 553 (1.1%) and 100 492 (1.9%) unique 5mC and 5hmC enriched regions respectively (Fig. 33C). This was visualised by the UCSC genome-wide visualisation. Estimates of the 5mC and 5hmC levels of representative regions from the rat genome are shown in figures 33D and 33E, which display enrichment for 5mC, 5hmC or both. As such, our observations suggested/confirmed that antibody cross-reactivity was minimal or non-existent, as shown in previous studies[55,236,260,261].

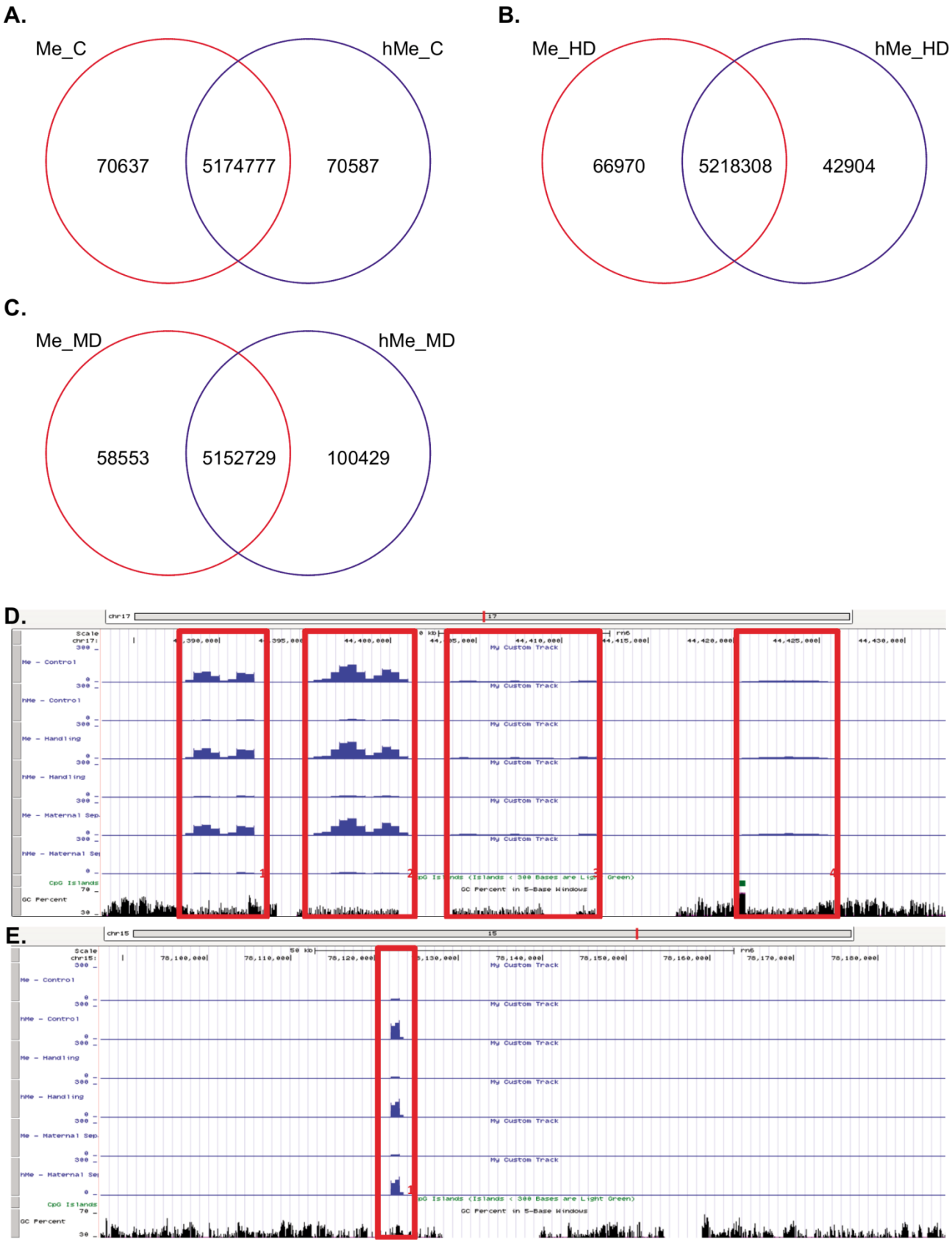

Figure 33: Overall comparison of the methylome and hydroxymethylome within groups. (A) The windows shared between the methylome and hydroxymethylome for the controls (C) (B) The windows shared between the methylome and hydroxymethylome for the handling group (HD). (C) The windows shared between the methylome and hydroxymethylome for the maternal deprivation group (MD). (D) A close-up UCSC browser view of the 5mC and 5hmC profile of the C, HD and MD group in chromosome 17, pinpointing a region exhibiting DNA methylation but no hydroxymethylation (E) A close-up UCSC browser view of the 5mC and 5hmC profile of the C, HD and MD group in chromosome 15, pinpointing a region exhibiting DNA hydroxymethylation but no methylation.

3.4.3. Group Differences in Distribution

5mC: Of the 5.7 x 10^6 windows the genome was divided in, 404 346 (7.0%) windows did not show any methylation presence in any of the three groups (Fig. 34A). 5 336 528 (93.0%) windows appeared to be methylated in at least one of the three conditions, of which 5 134 031 (89.4%) were shared by all three, 74 410 (1.3%) by the C and HD group, 44 725 (0.8%) by the HD and MD group and 18 249 (0.3%) by the C and MD group (Fig. 34A). 18 724 (0.3%), 32 112 (0.6%) and 14 277 (0.2%) methylated windows were unique for the C, HD and MD group respectively (Fig. 34A). Figure 34C shows representative regions from the rat genome within chromosome 1, where the 5mC peaks slightly differ in form and 5mC height.

5hmC: 5hmC data was similar to 5mC. Approximately 7.2% of the windows analysed (413 248) did not exhibit any sign of hydroxymethylation in any of the three groups (Fig. 34B). Of the 5 327 624 (92.8%) windows displaying 5hMC, the majority (5 159 826; 89.9%) was shared by all treatments, 37 405 (0.7%) windows were shared by the C and HD group, 43 734 (0.8%) by the HD and MD group and 31 376 (0.5%) by the C and MD group (Fig. 34B). A small number of windows was unique for each treatment group, 16 755 (0.3%) for C, 20 245 (0.4%) for HD and 18 283 (0.3%) for MD group (Fig. 34B). Figure 34D shows representative regions from the rat genome within chromosome 13, which display small differences in 5hmC level between the 3 treatment groups.

3.4.4. Differential Methylation Identification

The 5 (h)mC profiling of the three groups clearly showed that despite significant similarity, there were small differences in 5(h)mC distribution. To test for differential coverage, 5(h)mC levels were compared between two conditions using MEDIPS. The detected D(h)MRs were verified and validated by quality control plots visualising their characteristics, such as QQ-, volcano and Manhattan plots, under the null distribution, which assumed there were no differential 5(h)mC levels.

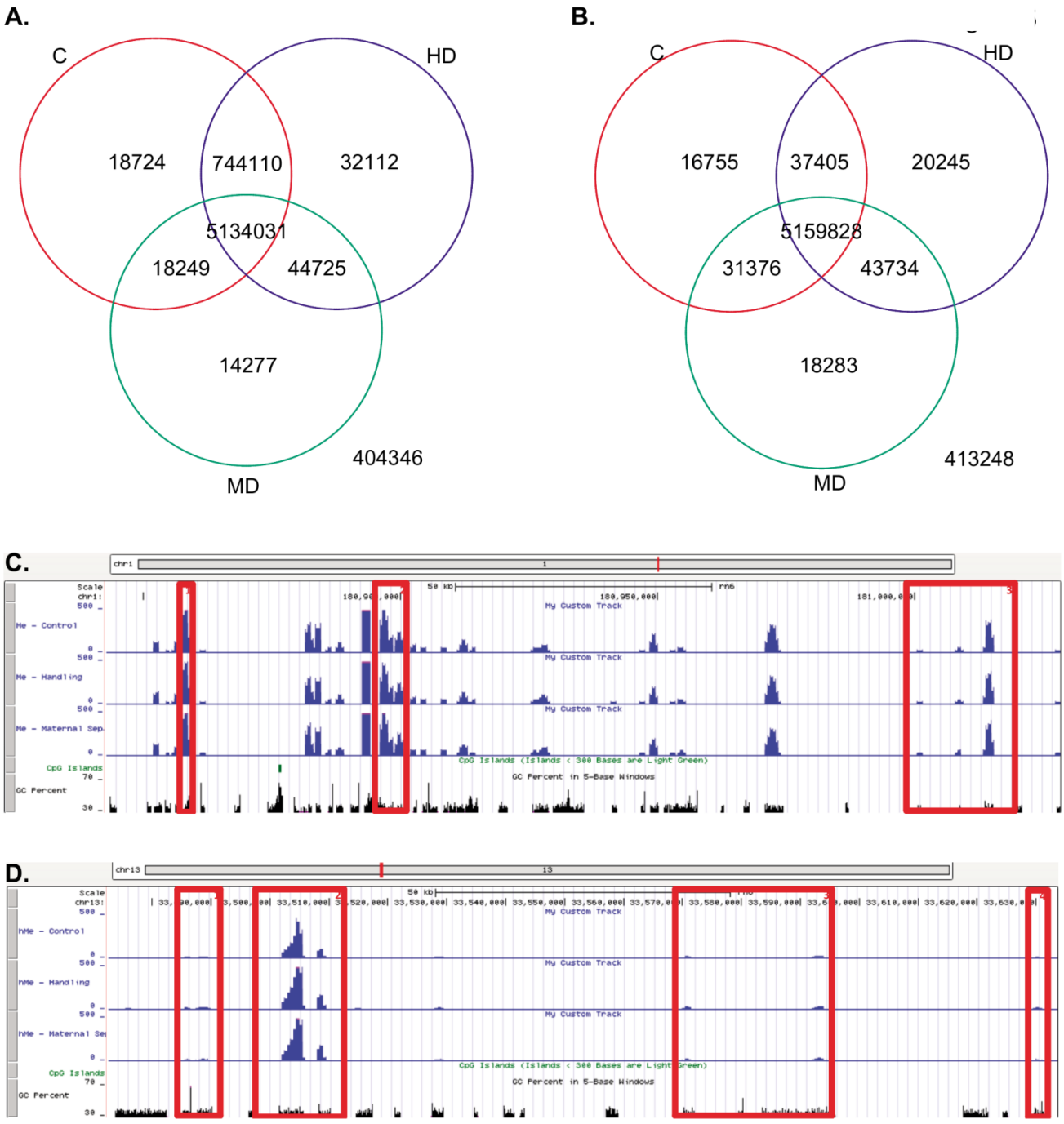

Figure 34: Overall comparison of the methylome and hydroxymethylome profiles between groups. (A) The methylated windows shared between the three groups (B) The hydroxymethylated windows shared between the three groups (C) A close-up UCSC browser view of the 5mC profiles of the C, HD and MD group in chromosome 1 displaying subtle differences in methylation level. (E) A close-up UCSC browser view of the 5hmC profiles of the C, HD and MD group in chromosome 1 displaying subtle differences in hydroxymethylation level. C. Control; HD: Handling; MD: Maternal Deprivation.

5mC: QQ-plots displayed distributions deviant from the null distribution, with an enrichment of small p-values as compared to the expected p-values (Fig. 35A and 36A-B). Indicating that although the majority of windows' 5mC levels did not differ significantly between groups, a small number of windows did. The volcano and Manhattan plots strengthened these observations (Fig. 35B-C and 36C-F). The former displayed approximately symmetrical shapes [range (-3.5,3.5)] (Fig. 35B and 36C-D). The fold change or 5mC difference between HD and MD is not significant for the majority of genomic windows (Fig. 35B). Only, a small subset, coloured in red,

displayed significant small fold changes between both groups. Some larger fold changes were observed, yet were not significant. Hence, implying significant differences between HD and MD concerned mainly genomic windows with small 5mC levels shifts. Similar trends/patterns were also observed for C-HD and C-MD comparisons (Fig. 36C-D). Both comparisons also displayed a few larger significant 5mC level shifts. The Manhattan plots enabled a genome-wide overview of differential windows (Fig. 35C and 36E-F). Most genomic windows presented large p-values for the HD-MD comparison (Fig. 35C). Yet chromosome 1, 3, 7, 13 and M displayed peaks in some regions, indicating the presence of neighbouring windows with small p-values. The C-HD and C-MD Manhattan plot showed similar trends, with peaks observed in chromosomes 1, 3, 5, 7, 13, 20 and M, and chromosomes 1, 3, 5, 7, 13, 20 and M respectively (Fig. 36E-F).

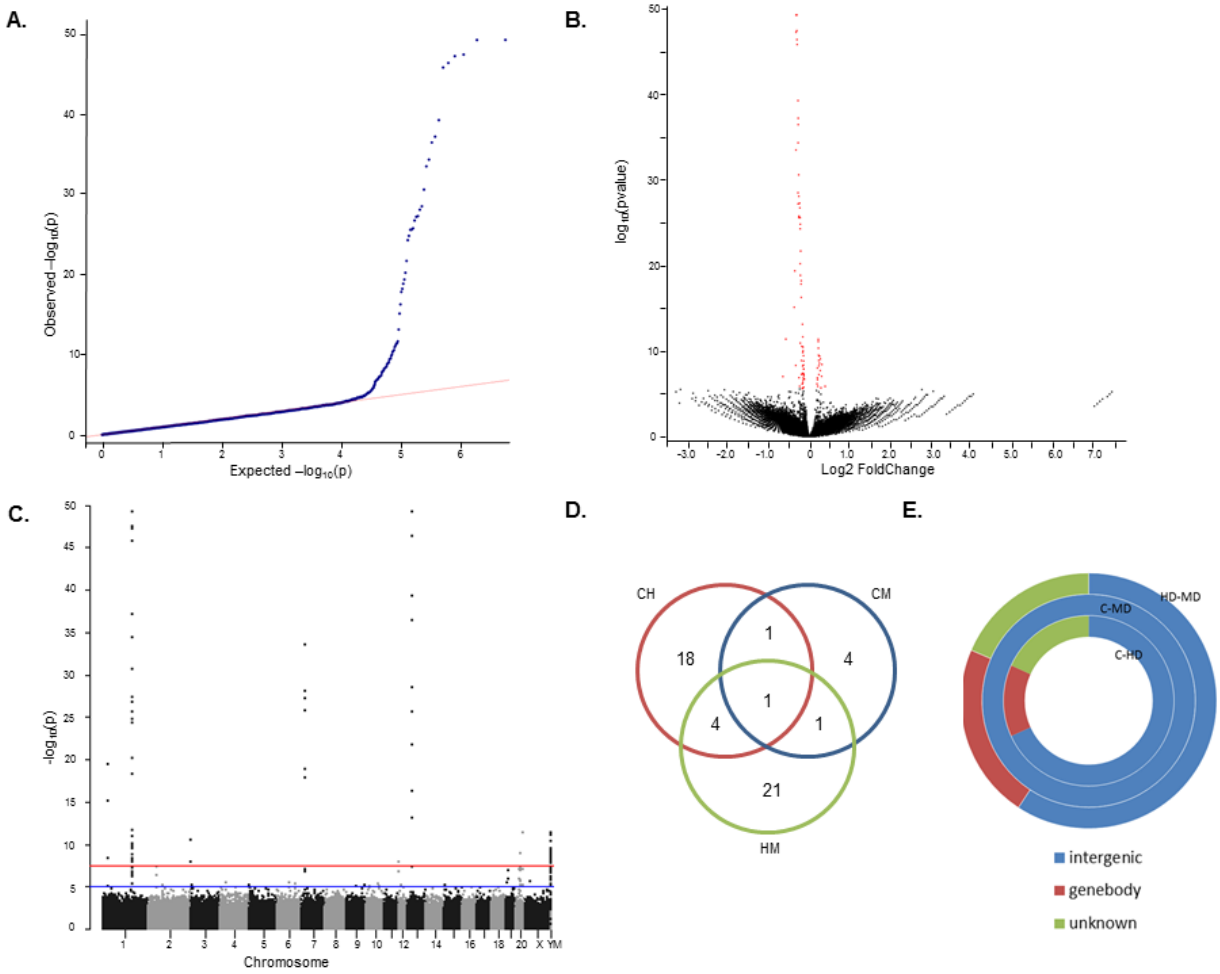

Figure 35: The detection and verification of differences in DNA methylation levels between the treatment groups. (A) A QQ-plot, depicting the observed log values in function of the expected log values for the HM comparison (B) A volcano plot depicting the statistical significance in function of the magnitude of the differences in DNA methylation between HD and MD. (C) A manhattan plot demonstrating the distribution of the DMRs for HM across the rat genome. (D) The DMRs shared between treatment group comparisons. (E) The genomic context of the DMRs. C. Control; HD: Handling; MD: Maternal Deprivation; CH: control-handling comparison; CM: control-maternal deprivation comparison; HM: handling-maternal deprivation comparison. (…) Intergenic; (…) Gene body; (…) unknown.

Table 5: The differentially methylated regions (DMR) for the 3 different group comparisons. For each DMR the exact position, the region width and the methylation change is given. The change describes the change in 5mC of the first group compared to the second group. C: control group, HD: handling group, MD: maternal deprivation group. *clone-based; red text: piRNA (Manuscript Supplementary Data).

Chromosome	Start	Stop	Width	Gene Annotation (Ensembl or Gene Entrez ID)	Change
				Methylation: C - HD	
chr5	6 374 001	6 375 000	1 000	ENSRNOG00000058589*,ENSRNOG00000058268*	loss
chr5	127 472 001	127 472 500	500	ENSRNOG00000015519, ENSRNOG00000012778, ENSRNOG00000012724,ENSRNOG00000012443, ENSRNOG00000056199, ENSRNOG00000011644	
chr20	19 484 001	19 484 500	500	ENSRNOG00000000275, ENSRNOG00000001387, AABR07044752,1	
chrM	2 001	4 000	2 000	ENSRNOG00000031780.1*, ENSRNOG00000030478.3*,	
chrM	5 501	16 500	11 000	ENSRNOG00000029171.3*, ENSRNOG00000043866*, ENSRNOG00000032112*, ENSRNOG00000030644, ENSRNOG00000029301*, ENSRNOG00000033545*, ENSRNOG00000032274*,ENSRNOG00000031033, ENSRNOG00000029677*, ENSRNOG00000033932*, ENSRNOG00000032609*, ENSRNOG00000031333*, ENSRNOG00000029954*, ENSRNOG00000034234, ENSRNOG00000032882*, ENSRNOG00000031685*, ENSRNOG00000030371, ENSRNOG00000029070*, ENSRNOG00000033299, ENSRNOG00000031979, ENSRNOG00000030700, ENSRNOG00000029389*, ENSRNOG00000033615, ENSRNOG00000032320*, ENSRNOG00000031053, ENSRNOG00000029707, ENSRNOG00000033957*, ENSRNOG00000032578*, ENSRNOG00000031667*, ENSRNOG00000029971, ENSRNOG00000029042, ENSRNOG00000032997*, ENSRNOG00000031766, ENSRNOG00000030339*, ENSRNOG00000029145*	
chrX	32 445 501	32 446 000	500	ENSRNOG00000038686, ENSRNOG00000004120	
chrX	33 815 001	33 815 500	500	ENSRNOG00000038654, ENSRNOG00000001387, ENSRNOG00000061508	
chr1	180 644 001	180 647 000	3 000		Gain
chr1	180 806 001	180 807 500	1 500		
chr1	180 808 001	180 810 000	2 000	only mRNA's: DQ620752	
chr1	180 893 001	180 894 000	1 000		
chr1	180 896 001	180 896 500	500		
chr1	273 631 501	273 632 000	500	ENSRNOG00000012084, ENSRNOG00000031381, ENSRNOG00000056862	
chr3	41 001	42 500	1 500	ENSRNOG00000058456*	
chr7	22 324 001	22 324 500	500		

chr7	22 328 001	22 329 500	1 500	only mRNA'S: DQ610624, DQ627099, DQ738923, DQ625003, DQ606303, DQ614554, DQ759772	
chr7	101 696 001	101 696 500	500	only mRNA's: DQ765923, DQ615970, DQ614391, DQ753126, DQ622619, DQ753126	
chr13	33 502 001	33 505 500	3 500	-	
chr13	33 508 001	33 508 500	500		
chr20	27 461 501	27 464 000	2 500		
chr20	27 471 501	27 473 000	1 500	100910945, ENSRNOG00000000277, ENSRNOG00000042916	
chr20	27 474 501	27 475 500	1 000		
Methylation: C – MD					
chr1	180 644 001	180 647 000	3 000		loss
chr1	180 808 001	180 808 500	500	only mRNA's: DQ620752	
chr1	180 893 001	180 893 500	500		
chr7	22 328 001	22 329 500	1 500	only mRNA's: DQ610624, DQ627099, DQ738923, DQ625003, DQ606303, DQ614554, DQ759772	
chr13	33 503 501	33 505 500	2 000	-	
chr13	33 508 001	33 509 000	1 000		
chr14	76 850 501	76 851 000	500	-	gain
Methylation: HD – MD					
chr1	30 985 001	30 986 500	1 500	ENSRNOG00000011622, ENSRNOG00000054753*, ENSRNOG00000042309, ENSRNOG00000012324	
chr1	180 644 001	180 647 000	3 000		loss
chr1	180 703 001	180 706 500	3 500		
chr1	180 805 501	180 810 000	4 500		
chr1	180 808 001	180 810 000	2 000	only mRNA's: DQ620752	
chr1	180 858 001	180 858 500	500		
chr1	180 881 501	180 882 000	500		
chr1	180 893 001	180 894 000	1 000		

chr1	180 896 001	180 897 500	1 500		
chr2	49 364 001	49 365 000	1 000	ENSRNOG00000046657	
chr3	41 001	42 500	1 500	ENSRNOG00000058456*	
chr7	22 323 001	22 324 500	1 500	only mRNA's: DQ610624, DQ627099, DQ738923, DQ625003,	
chr7	22 327 001	22 329 500	2 500	DQ614554, DQ759772	
chr12	3 350 501	3 351 500	1 000	ENSRNOG00000053067*, ENSRNOG00000054841*, ENSRNOG00000061869*	
chr13	33 502 001	33 506 000	4 000	-	
chr13	33 508 001	33 509 000	1 000		
chr19	150 44 501	15 045 500	1 000	367515, ENSRNOG00000046163, ENSRNOG00000057497*, ENSRNOG00000052015*,291863,ENSRNOG00000015519, ENSRNOG00000015438, ENSRNOG00000015438,501233, ENSRNOG00000015438	
chr20	27 462 501	27 463 500	1 000		
chr20	27 471 501	27 472 500	1 000	100910945, ENSRNOG00000000277, ENSRNOG00000042916	
chr20	27 474 501	27 475 500	1 000		
chr20	46 083 501	46 084 500	1 000	ENSRNOG00000000312, ENSRNOG00000059699*, ENSRNOG00000037688*, ENSRNOG00000000308	
chr20	19 483 501	19 484 500	1 000	ENSRNOG00000000275, ENSRNOG00000001387, ENSRNOG00000055901*	
chrM	2 501	5 500	3 000	ENSRNOG00000031780.1*, ENSRNOG00000030478.3*,	gain
chrM	7 501	9 000	1 500	ENSRNOG00000029171.3*,ENSRNOG00000043866*,	
chrM	10 001	11 000	1 000	ENSRNOG00000032112*, ENSRNOG00000030644,	
chrM	11 501	12 000	500	ENSRNOG00000029301*, ENSRNOG00000033545*,	
chrM	13 001	16 000	3 000	ENSRNOG00000032274*, ENSRNOG00000031033, ENSRNOG00000029677*, ENSRNOG00000033932*, ENSRNOG00000032609*, ENSRNOG00000031333*, ENSRNOG00000029954*, ENSRNOG00000034234, ENSRNOG00000032882*, ENSRNOG00000031685*, ENSRNOG00000030371, ENSRNOG00000029070*, ENSRNOG00000033299, ENSRNOG00000031979, ENSRNOG00000030700, ENSRNOG00000029389*, ENSRNOG00000033615, ENSRNOG00000032320*, ENSRNOG00000031053, ENSRNOG00000029707, ENSRNOG00000033957*, ENSRNOG00000032578*, ENSRNOG00000031667*, ENSRNOG00000029971, ENSRNOG00000029042, ENSRNOG00000032997*, ENSRNOG00000031766, ENSRNOG00000030339*, ENSRNOG00000029145*	

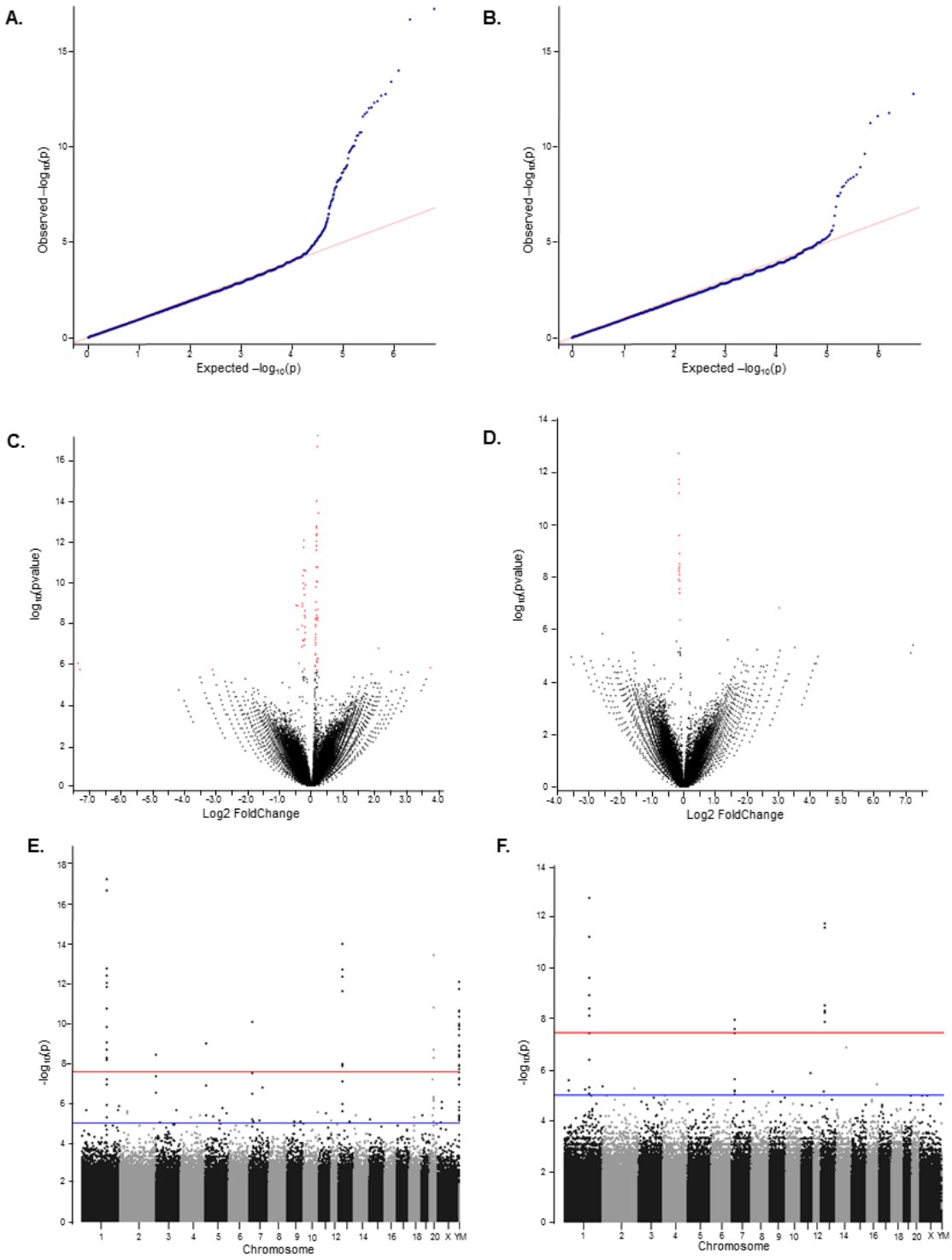

Figure 36: The detection and verification of differences in DNA methylation levels between the treatment groups. A QQ-plot, depicting the observed log values in function of the expected log values for the (A) CH and (B) CM comparison. A volcano plot depicting the statistical significance in function of the magnitude of the differences in DNA methylation between (C) and HD, (D) C and MD. A manhattan plot demonstrating the distribution of the DMRs across the rat genome for (E) CH and (F) CM respectively. C. Control; HD: Handling; MD: Maternal Deprivation; CH: control-handling comparison; CM: control-maternal deprivation comparison; HM: handling-maternal deprivation comparison (Manuscript Supplementary Data).

MEDIPS identified a total of 177 windows where the methylation level changed significantly across the different treatments (Fig. 35D; Table 5). The majority of those were detected for HD-MD, with 86 windows compared to the 73 and 18 windows for C-HD and C-MD respectively. Adjacent differentially methylated

windows were merged into a larger DMR whenever possible. After merging a total of 50 DMRs were identified (Fig. 35D). Comparing HD to MD identified 27 (merged) DMRs, 21 unique DMRs, 1 shared with all group comparisons, 4 shared with C-HD and 1 with C-MD (Fig. 35D). The majority of the DMRs was located in intergenic regions (67.86%; 38 DMR), a small number coincided with the gene body (16.07%; 9 DMRs) or was unknown (16.07%, 9 DMR) (Fig. 35E). Except for the C-MD comparison, where all detected DMRs coincided with the intergenic region (7 DMRs). When comparing HD to MD, 21 DMRs exhibited upregulated 5mC levels (Table 5). C-HD detected 24 DMRs, of which 15 were upregulated in C (Fig. 35D; Table 5). Only 7 DMRs were detected for C-MD, with 6 out of 7 downregulated for C (Fig. 35D; Table 5).

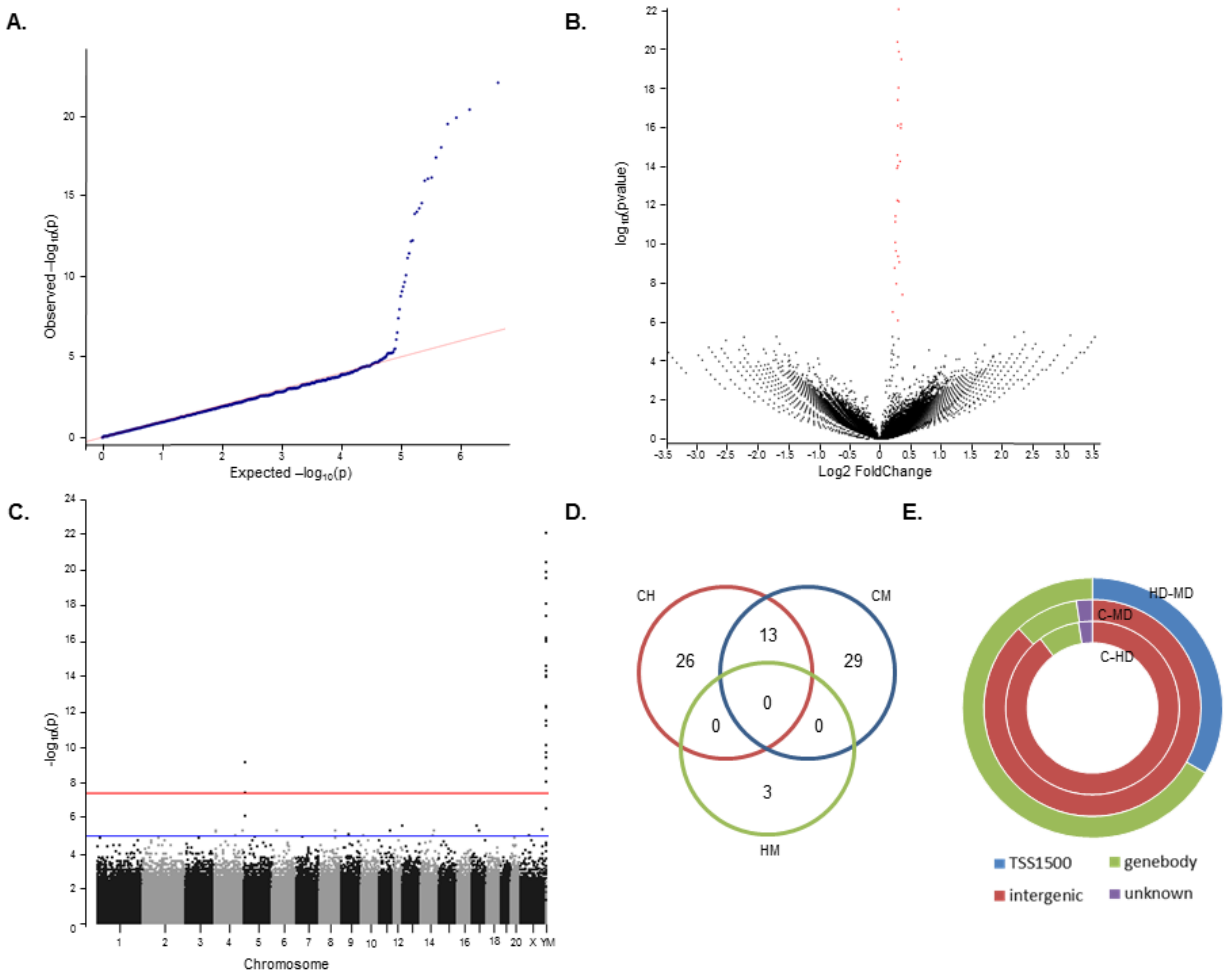

Figure 37: The detection and verification of differences in DNA hydroxymethylation levels between the treatment groups. (A) A QQ-plot, depicting the observed log values in function of the expected log values for the HM comparison (B) A volcano plot depicting the statistical significance in function of the magnitude of the differences in DNA hydroxymethylation between HD and MD. (C) A manhattan plot demonstrating the distribution of the DhMRs for HM across the rat genome. (D) The DhMRs shared between treatment group comparisons. C. Control; HD: Handling; MD: Maternal Deprivation; CH: control-handling comparison; CM: control-maternal deprivation comparison; HM: handling-maternal deprivation comparison. (E) The genomic context of the DMRs. C. Control; HD: Handling; MD: Maternal Deprivation; CH: control-handling comparison; CM: control-maternal deprivation comparison; HM: handling-maternal deprivation comparison. ( ) TSS1500; ( ) Gene body; ( ) Intergenic; ( ) unknown.

Table 6: The differentially hydroxymethylated regions (hDMR) for the 3 different group comparisons. For each hDMR the exact position, the region width and the methylation change is given. The change describes the change in 5hmC of the first group compared to the second group. C: control group, HD: handling group, MD: maternal deprivation group. *clone-based; red text: piRNA (Manuscript Supplementary Data).

Chromosome	Start	Stop	Width	Gene Annotation (Ensembl or Gene Entrez ID)	Change
Hydroxymethylation: C - HD					
chr1	180 644 001	180 647 000	3 000	only mRNA's: DQ620752, AY539949, DQ732421	loss
chr1	180 703 001	180 706 500	3 500		
chr1	180 719 001	180 720 000	1 000		
chr1	180 805 001	180 805 500	500		
chr1	180 806 001	180 809 500	3 500		
chr1	180 858 001	180 859 000	1 000		
chr1	180 884 001	180 884 500	500		
chr1	180 893 001	180 894 000	1 000		
chr1	180 896 501	180 898 500	2 000		
chr1	180 899 501	180 900 500	1 000		
chr1	181 014 501	181 015 000	500		
chr1	181 186 001	181 186 500	500		
chr1	181 198 501	181 200 000	1 500		
chr1	181 286 501	181 287 500	1 000		
chr3	133 567 001	133 567 500	500	ENSRNOG00000047802*	
chr3	133 993 001	133 993 500	500	ENSRNOG00000060762*	
chr3	133 995 501	133 997 000	1 500		
chr3	133 999 001	134 000 000	1 000		
chr3	134 021 001	134 021 500	500		
chr7	22 323 001	22 324 500	1 500	only mRNA's: DQ610624, DQ627099, DQ738923, DQ625003, DQ606303, DQ614554, DQ759772	
chr7	22 326 501	22 329 000	2 500		
chr12	3 350 001	3 353 000	3 000		
chr12	3 355 001	3 356 000	1 000	ENSRNOG00000053067*, ENSRNOG00000054841*, ENSRNOG00000061869*	
chr12	3 356 501	3 357 000	500		
chr13	33 485 501	33 486 000	500		
chr13	33 502 001	33 506 000	4 000		
chr13	33 508 001	33 509 500	1 500	-	
chr13	33 529 001	33 529 500	500		
chr13	33 571 001	33 571 500	500		

chr13	33 593 001	33 594 000	1 000		
chr17	44 391 001	44 391 500	500	ENSRNOG00000056731	
chr18	38 414 501	38 417 500	3 000	ENSRNOG00000039260	
chr18	38 497 501	38 498 500	1 000		
chr20	40 882 501	40 883 000	500	ENSRNOG00000000815	
chrM	2 001	2 500	500	ENSRNOG00000031780.1*, ENSRNOG00000030478.3*, ENSRNOG00000029171.3*,ENSRNOG00000043866*, ENSRNOG00000032112*, ENSRNOG00000030644, ENSRNOG00000029301*, ENSRNOG00000033545*, ENSRNOG00000032274*, ENSRNOG00000031033, ENSRNOG00000029677*, ENSRNOG00000033932*, ENSRNOG00000032609*, ENSRNOG00000031333*, ENSRNOG00000029954*, ENSRNOG00000034234, ENSRNOG00000032882*, ENSRNOG00000031685*, ENSRNOG00000030371, ENSRNOG00000029070*, ENSRNOG00000033299, ENSRNOG00000031979, ENSRNOG00000030700, ENSRNOG00000029389*, ENSRNOG00000033615, ENSRNOG00000032320*, ENSRNOG00000031053, ENSRNOG00000029707, ENSRNOG00000033957*, ENSRNOG00000032578*, ENSRNOG00000031667*, ENSRNOG00000029971, ENSRNOG00000029042, ENSRNOG00000032997*, ENSRNOG00000031766, ENSRNOG00000030339*, ENSRNOG00000029145*	
chr1	81 982 501	81 983 000	500	ENSRNOG00000020310, ENSRNOG00000055761, ENSRNOG00000055650, ENSRNOG00000046112	gain
chr2	28 980 501	28 981 000	500	ENSRNOG00000055452*, ENSRNOG00000015334, ENSRNOG00000014999	
chr14	11 809 501	11 810 000	500	-	
chr17	57 707 501	57 708 000	500	ENSRNOG00000048661, ENSRNOG00000045919, ENSRNOG00000056440*, ENSRNOG00000046403	
Hydroxymethylation: C – MD					
chr1	11 966 501	11 967 000	500	ENSRNOG00000055421*, ENSRNOG00000059852, ENSRNOG00000051831*, 310926, ENSRNOG00000052007*, ENSRNOG00000061685*, ENSRNOG00000047746*, ENSRNOG00000055836*, ENSRNOG00000060657*, ENSRNOG00000059898*, ENSRNOG00000055789*	loss
chr1	128 703 001	128 703 500	500	ENSRNOG00000023274, ENSRNOG00000013877, ENSRNOG00000014030	
chr1	180 590 001	180 590 500	500	only mRNA's: DQ620752, AY539949, DQ732421	

chr1	180 644 001	180 647 000	3 000	
chr1	180 702 501	180 706 500	4 000	
chr1	180 705 501	180 706 500	1 000	
chr1	180 719 001	180 720 000	1 000	
chr1	180 806 001	180 809 500	3 500	
chr1	180 852 501	180 853 000	500	
chr1	180 858 001	180 859 000	1 000	
chr1	180 893 001	180 894 000	1 000	
chr1	180 896 501	180 897 500	1 000	
chr1	180 898 001	180 898 500	500	
chr1	181 077 501	181 078 500	1 000	
chr1	181 186 001	181 187 000	1 000	
chr1	181 198 001	181 199 500	1 500	
chr1	181 215 501	181 216 000	500	
chr1	181 286 501	181 287 500	1 000	
chr2	160 692 001	160 692 500	500	only mRNA's: DQ765923, DQ750390, DQ617315, DQ735565, DQ610613, DQ6106101, DQ735019, AF055714, DQ731535, DQ627773, DQ763101, DQ624671
chr3	133 995 501	133 997 000	1 500	only mRNA's: DQ621077, DQ604474, DQ763579
chr3	133 999 501	134 000 000	500	
chr7	22 323 001	22 324 500	1 500	only mRNA's: DQ610624, DQ627099, DQ738923, DQ625003, DQ606303, DQ614554, DQ759772
chr7	22 326 501	22 329 000	2 500	
chr12	3 350 001	3 352 500	2 500	ENSRNOG00000053067*, ENSRNOG00000054841*, ENSRNOG00000061869*
chr12	3 354 001	3 355 500	1 500	
chr13	33 475 501	33 476 000	500	-
chr13	33 485 501	33 486 000	500	
chr13	33 502 001	33 506 000	4 000	
chr13	33 508 001	33 509 500	1 500	
chr13	33 528 501	33 529 000	500	
chr13	33 593 501	33 594 000	500	
chr17	44 396 501	44 397 500	1 000	ENSRNOG00000056731
chr18	38 414 001	38 415 500	1 500	
chr18	38 416 001	38 417 500	1 500	ENSRNOG00000039260
chr18	38 497 501	38 498 500	1 000	

chr4	6 877 501	6 878 500	1 000	ENSRNOG00000009085, ENSRNOG00000050578,	
chr4	6 879 001	6 880 000	1 000	ENSRNOG00000059699*, ENSRNOG00000009226, ENSRNOG00000025735, ENSRNOG00000009282	
chr4	93 366 501	93 367 000	500	ENSRNOG00000005630*	
chr5	6 374 001	6 374 500	500	ENSRNOG00000058268*	
chr20	19 484 001	19 485 000	1 000	ENSRNOG00000000275, ENSRNOG00000001387, ENSRNOG00000055901*	
chrM	3 001	5 500	2 500	ENSRNOG00000031780.1*, ENSRNOG00000030478.3*, ENSRNOG00000029171.3*, ENSRNOG00000043866*, ENSRNOG00000032112*,ENSRNOG00000030644, ENSRNOG00000029301*, ENSRNOG00000033545*, ENSRNOG00000032274*, ENSRNOG00000031033, ENSRNOG00000029677*, ENSRNOG00000033932*, ENSRNOG00000032609*, ENSRNOG00000031333*, ENSRNOG00000029954*, ENSRNOG00000034234, ENSRNOG00000032882*, ENSRNOG00000031685*,	g a i n
chrM	7 501	15 500	8 000	ENSRNOG00000030371, ENSRNOG00000029070*, ENSRNOG00000033299, ENSRNOG00000031979, ENSRNOG00000030700, ENSRNOG00000029389*, ENSRNOG00000033615, ENSRNOG00000032320*, ENSRNOG00000031053, ENSRNOG00000029707, ENSRNOG00000033957*, ENSRNOG00000032578*, ENSRNOG00000031667*, ENSRNOG00000029971, ENSRNOG00000029042, ENSRNOG00000032997*, ENSRNOG00000031766, ENSRNOG00000030339*, ENSRNOG00000029145*	
Hydroxymethylation: HD – MD					
chr5	6 373 501	6 375 000	1 500	ENSRNOG00000058589*, ENSRNOG00000058268*	
chrM	2 501	5 500	3 000	ENSRNOG00000031780.1*, ENSRNOG00000030478.3*, ENSRNOG00000029171.3*, ENSRNOG00000043866*, ENSRNOG00000032112*,ENSRNOG00000030644, ENSRNOG00000029301*, ENSRNOG00000033545*, ENSRNOG00000032274*, ENSRNOG00000031033, ENSRNOG00000029677*, ENSRNOG00000033932*, ENSRNOG00000032609*, ENSRNOG00000031333*, ENSRNOG00000029954*, ENSRNOG00000034234, ENSRNOG00000032882*, ENSRNOG00000031685*,	g a i n
chrM	7 001	15 500	8 500	ENSRNOG00000030371, ENSRNOG00000029070*, ENSRNOG00000033299, ENSRNOG00000031979, ENSRNOG00000030700, ENSRNOG00000029389*, ENSRNOG00000033615, ENSRNOG00000032320*, ENSRNOG00000031053, ENSRNOG00000029707, ENSRNOG00000033957*, ENSRNOG00000032578, ENSRNOG00000031667*, ENSRNOG00000029971, ENSRNOG00000029042, ENSRNOG00000032997*,	

	ENSRNOG00000031766, ENSRNOG00000030339*, ENSRNOG00000029145*

5hmC: Similarly to the 5mC analyses, QQ-plots displayed distributions deviant from the null distribution, with an enrichment of small p-values as compared to the expected p-values (Fig. 37A; Fig. 38A-B). Indicating that although the majority of windows' 5hmC levels did not differ significantly between groups, a small number of windows did. The volcano and Manhattan plots strengthened these observations (Fig. 37B-C; Fig. 38C-F). The former displayed approximately symmetrical shapes [range (-4.5,4.5)] (Fig. 37B; Fig. 38C-D). The fold change or 5hmC difference between HD and MD is not significant for the majority of genomic windows (Fig. 37B). Only, a small subset, coloured in red, displayed significant small fold changes between both groups (Fig. 37B). Hence, implying significant differences between HD and MD concerned mainly genomic windows with small 5hmC levels shifts. Similar trends/patterns were also observed for C-HD and C-MD comparisons (Fig. 38C-D). Both comparisons also displayed a few larger significant 5hmC level shifts. The Manhattan plots enabled a genome-wide overview of differential windows (Fig. 37C; Fig. 38E-F). Most genomic windows presented large p-values for the HD-MD comparison (Fig. 37C). Yet chromosome 5 and M displayed peaks in some regions, indicating the presence of neighbouring windows with small p-values. The C-HD and C-MD Manhattan plot showed similar trends, with peaks observed in chromosomes 1, 3, 7, 12, 13 and 18, and chromosome 1, 13, 20 and M respectively (Fig. 38E-F).

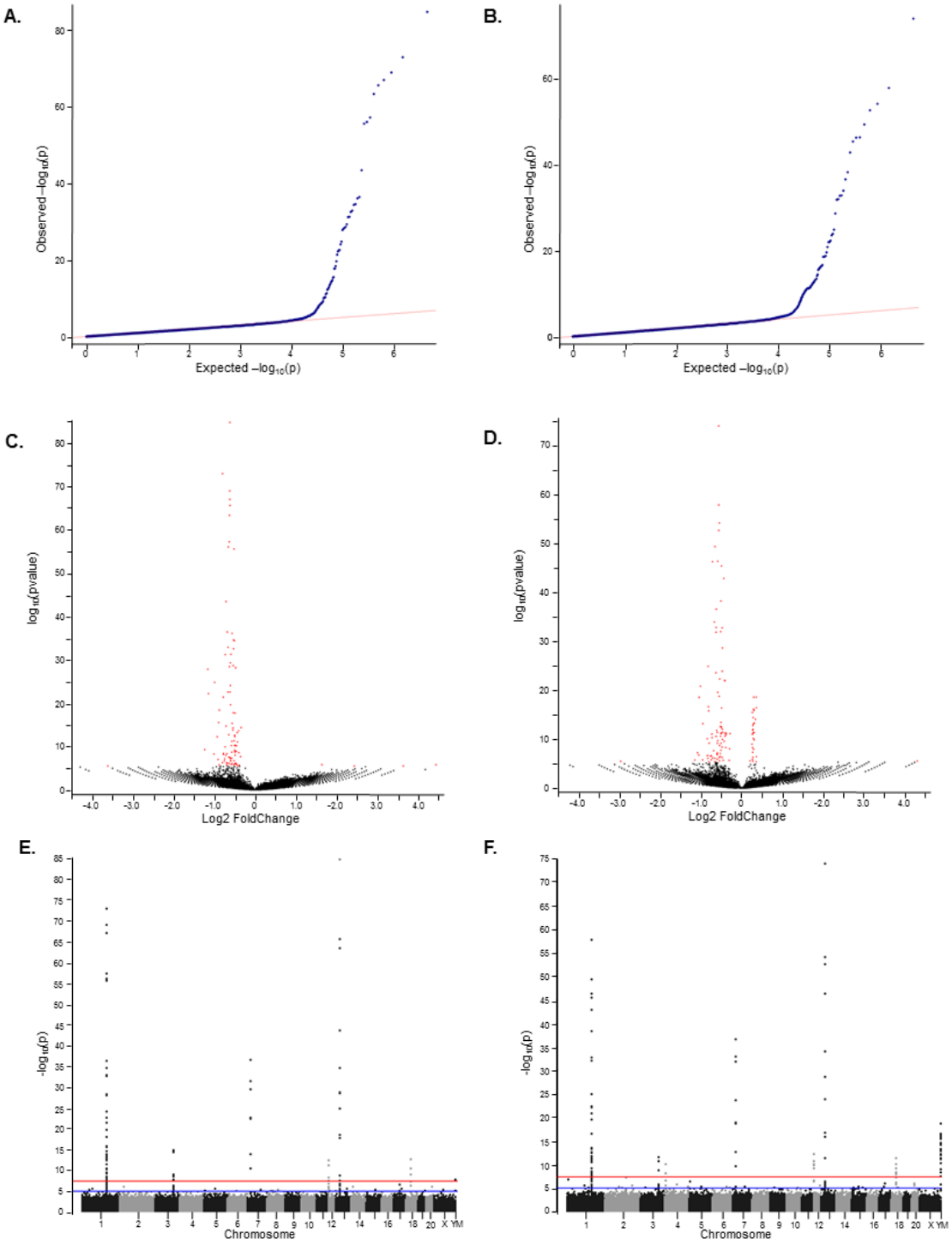

Figure 38: The detection and verification of differences in DNA hydroxymethylation levels between the treatment groups. A QQ-plot, depicting the observed log values in function of the expected log values for the (A) CH and (B) CM comparison. A volcano plot depicting the statistical significance in function of the magnitude of the differences in DNA hydroxymethylation between (C) and HD, (D) C and MD. A Manhattan plot demonstrating the distribution of the DhMRs across the rat genome for (E) CH and (F) CM respectively. C. Control; HD: Handling; MD: Maternal Deprivation; CH: control-handling comparison; CM: control-maternal deprivation comparison; HM: handling-maternal deprivation comparison (Manuscript Supplementary Data).

MEDIPS identified a total of 241 windows where the 5hmC level changed significantly across the different treatments (Fig. 37D; Table 6). Contrary 5mC, the majority of those were detected for C-MD, with 118 windows compared to the 97 windows for C-HD and only 26 widows for HD-MD. Adjacent differentially methylated

windows were merged into a larger DhMR whenever possible. After merging a total 71 DhMRs were identified (Fig. 37D). The majority of the DhMRs was located in intergenic regions (85.71%; 72 DhMR), a small number coincided with the gene body (10.71%; 9 DhMRs) (Fig. 37E). Except for the HD-MD comparison, where detected DhMRs coincided with either gene body (66,67%; 2 DhMRs) or TSS 1500 (33.33%; 1 DhMR). Comparing C to HD identified 39 DhMRs, 26 DhMRs unique and 13 shared with C-MD. When comparing C to HD, 35 DhMRs were downregulated (Table 8). C-MD detected 42 DMRs, of which 35 of the DhMRs were downregulated in C (Fig. 37D; Table 8). Only 3 DhMRs were detected for HD-MD, all upregulated for HD (Fig. 37D; Table 8).

Table 7: DAVID functional annotation clustering analysis of the differentially (hydroxy)methylated genes between C and HD, C and MD, and HD and MD. Control; HD: Handling; MD: Maternal Deprivation; CH: control-handling comparison; CM: control-maternal deprivation comparison; HM: handling-maternal deprivation comparison

Epigenetic Modification	Group	Annotation Cluster	Enrichment score	Count	P-value	Benjamini
5mC	HM	Oxidative phosphorylation	6.95	12	9.55E-17	6.66E-16
		Huntington's disease	2.82	6	4.46E-5	8.87E-5
		Huntington's disease	1.65	6	4.44E-5	8.87E-5
	CM	No clusters selected				
	CH	Generation of precursor metabolites and energy	7.66	13	4.81E-17	4.81E-16
		Oxidative phosphorylation	4.74.	12	8.97E-15	8.72E-13
		Huntington's disease	2.48	6	1.39E-4	4.62E-4
5hmC	HM	Oxidative phosphorylation	8.43	13	2.89E-20	1.45E-19
		Huntington's disease	2.93	6	2.81E-5	4.68E-5
		Electron transport	2.43	7	1.73E-11	9.98E-11
	CM	Generation of precursor metabolites and energy	7.62	13	2.25E-19	3.64E-17
		Huntington's disease	2.77	6	4.44E-5	1.18E-4
		Electron transport	2.25	7	1.48E-10	1.11E-9
	CH	No clusters selected				

3.4.5. Gene Ontology and Pathway Analysis

Annotation D(h)MR: As 5mC and 5hmC regions are often situated outsight of the gene body, the detected D(h)MRs were associated with genes or mRNA transcripts within 200kb range around the D(h)MR position (Table 5-6). The 50 DMRs and 71 DhMRs could be associated with annotated genes, although some were only clone-based. Genes such as *Nr3c1*, *BDNF*, *AVP*, *CRH*, *NR4A1* and *Igf2,* previously reported to be affected by MD models or early life stress [24,25,62,123,155,233,239,240], could not be directly associated with the detected D(h)MR. Almost all mitochondrial genes were present, indicating that the mitochondrial DNA is environmental sensitive and easily subjected to 5(h)mC shifts. Certain D(h)MRs that could not be associated with annotated genes,

could be linked the mRNA transcripts. All of these transcripts belonged to the piRNA family. The remaining D(h)MRs could not be associated with annotated genes or RNA transcripts.

Pathway Analysis: After functional annotation of the D(h)MRS, a gene ontology and network analysis were performed with 'Ingenuity Pathway Analysis®' and DAVID [257,258]. According to DAVID, the functionally annotated genes selected 4 main clusters: oxidative phosphorylation, Huntington's disease, generation of precursor metabolites and energy and electron transport (Table 7). The IPA analysis revealed significant 5mC and 5hmC changes in pathways related to oxidative phosphorylation, purine metabolism and mitochondrial dysfunction (Table 8).

Table 8: Ingenuity Pathway Analysis of the differentially (hydroxy)methylated genes between C and HD, C and MD, and HD and MD. Control; HD: Handling; MD: Maternal Deprivation; CH: control-handling comparison; CM: control-maternal deprivation comparison; HM: handling-maternal deprivation comparison (Manuscript Supplementary Data).

Epigenetic Modification	Group	Top Canonical Pathways (p-value)	Top Diseases and Disorders (p-value)	cellular and molecular functions (p-value)	Physiological Systems functions (p-value)	Top Pathways (p-value)	Top Networks (score)
5mC	CH	Oxidative Phosphorylation (1.03E-19)	-	Molecular Transport (2.74E-02 - 1.84E-02)	-	Oxidative Phosphorylation (3.65E-17)	Cellular Movement, Reproductive System Development and Function, Cellular Development (3)
		Mitochondrial Dysfunction (2.03E-19)	-	Lipid Metabolism (3.64E-02 - 2.74E-02)	-	Purine Metabolism (5.07E-02)	Developmental Disorder, Hereditary Disorder, Metabolic Disease (3)
		Mitochondrial L-carnitine Shuttle Pathway (2.23E-2)	-	Small Molecule Biochemistry (3.64E-02 - 2.74E-02)	-	Fatty Acid Metabolism (1.29E-01)	Developmental Disorder, Hereditary Disorder, Metabolic Disease (2)
		Glutamate Receptor Signalling (7.03E-2)	-	Energy Production (3.64E-02 - 3.64E-02)	-	-	-
		CTLA4 Signalling in Cytotoxic T	-	-	-	-	-

		Lymphocytes (1.11E-1)					
	HM	Oxidative Phosphorylation (2.87E-18)	-	Molecular Transport (3.40E-02 - 1.60E-02)	-	Oxidative Phosphorylation (1.00E-15)	Carbohydrate Metabolism, Lipid Metabolism, Molecular Transport (2)
		Mitochondrial Dysfunction (3.09E-18)	-	Lipid Metabolism (3.40E-02 - 3.40E-02)	-	Purine Metabolism (2.43E-04)	Cardiovascular Disease, Haematological Disease, Metabolic Disease (2)
		Ubiquinol-10 Biosynthesis (Eukaryotic) (1.60E-02)	-	Small Molecule Biochemistry (3.40E-02 - 3.40E-02)	-	Stilbene, Coumarine and Lignin Biosynthesis (2.50E-02)	Developmental Disorder, Hereditary Disorder, Metabolic Disease (2)
		Pyrimidine Deoxyribonucleotides De Novo Biosynthesis I (2.50E-02)	-	-	-	Aminophosphonate Metabolism (3.06E-02)	-
		Pyrimidine Ribonucleotides Interconversion (3.40E-02)	-	-	-	Ascorbate and Aldarate Metabolism (3.40E-02)	-
5hmC	CH	Trans, trans-farnesyl Diphosphate Biosynthesis (2.59E-03)	Endocrine System Disorders (3.62E-03 - 3.62E-03)	Cell-To-Cell Signalling and Interaction (2.59E-03 - 2.59E-03)	Reproductive System Development and Function (5.18E-04 - 5.18E-04)	Biosynthesis of Steroids (1.59E-02)	Nervous System Development and Function, Tissue Morphology, Cell Death and Survival (3)
		Mevalonate Pathway I (6.71E-03)	Metabolic Disease (3.62E-03 - 3.62E-03)	Molecular Transport (2.59E-03 - 2.59E-03)	Haematological System Development and Function (3.10E-03 - 3.10E-03)	Aminosugars Metabolism (9.56E-02)	Cell-To-Cell Signalling and Interaction, Cellular Assembly and Organization, Cellular Development (3)
		Superpathway of Geranylgeranyldiphosphate Biosynthesis I (via	-	Small Molecule Biochemistry (2.59E-	Haematopoiesis (3.10E-03 - 3.10E-03)	-	-

	Mevalonate) (8.77E-03)		03 - 2.59E-03)			
	Superpathway of Cholesterol Biosynthesis (1.44E-02)	-	Cellular Development (9.79E-03 - 3.10E-03)	Humoral Immune Response (3.10E-03 - 3.10E-03)	-	-
		-	Cell Signalling (1.34E-02 - 1.34E-02)	Nervous System Development and Function (9.79E-03 - 9.79E-03)	-	-
CM	Oxidative Phosphorylation (2.71E-20)	Cardiovascular Disease (1.80E-02 - 1.80E-02)	Post-Translational Modification (3.10E-02 - 3.62E-03)	-	Oxidative Phosphorylation (1.14E-17)	Cardiovascular System Development and Function, Organ Morphology, Organismal Development (2)
	Mitochondrial Dysfunction (4.48E-20)	Organismal Injury and Abnormalities (1.80E-02 - 1.80E-02)	Molecular Transport (1.68E-02 - 1.68E-02)	-	Purine Metabolism (4.30E-02)	Developmental Disorder, Hereditary Disorder, Metabolic Disease (2)
	mTOR Signalling (2.09E-02)	Skeletal and Muscular Disorders (1.80E-02 - 1.80E-02)	Cell Death and Survival (1.80E-02 - 1.80E-02)	-	Molecular Mechanisms of Cancer (PI3K DATASET) (3.31E-01)	-
	Breast Cancer Regulation by Stathmin1 (2.24E-02)	-	Cell Signalling (4.84E-02 - 3.10E-02)	-	-	-
	Sonic Hedgehog Signalling (3.10E-02)	-	Protein Synthesis (3.10E-02 - 3.10E-02)	-	-	-
HM	Mitochondrial Dysfunction (2.11E-24)	-	Molecular Transport (1.04E-02 - 1.04E-02)	-	Oxidative Phosphorylation (9.58E-21)	Developmental Disorder, Hereditary Disorder,

						Metabolic Disease (2)
	Oxidative Phosphorylation (6.22E-24)	-	-	-	Purine Metabolism (1.73E-02)	-

3.5. Discussion

The MD model is well known to cause long lasting behavioural effects, changes in the HPA axis reactivity and the stress response, all of which have been previously associated with DNA methylation changes within specific genes [155,239,240]. This is one of the first studies combining both 5mC and 5hmC epigenomes, allowing a direct comparison of the overall methylation and hydroxymethylation profiles and the differential regions between groups, as well as within groups. Adopting a genome-wide approach for methylome and hydroxymethylome investigation, this study enabled a deeper insight in genes/regions susceptible to early life adversity, as well as an evaluation of the effect of subtle methylation changes within such gene networks resulting in different phenotypes. These data suggest the existence of a relation between 5mC and 5hmC, and examines their impact reshaping/remodelling the methylation landscape.

The multiple sequencing-based methylome profiling techniques can be classified as either bisulphite conversion-based or enrichment-based techniques [262–265]. For this study, we opted for the enrichment-based technique MeDIP-Seq, which provides a direct way of comparing 5mC and 5hmC patterns genome-wide. Contrary to bisulphite conversion-based methods, MeDIP-Seq differentiates between the different cytosine modifications such as 5mC, 5hmC, 5caC and 5fC, given the appropriate antibody [262–267]. Although, it only offers a qualitative fragment-based resolution, MeDIP-Seq allows to captures the vast majority of the methylome, including areas outside of annotated genomic regions or in repetitive elements [262,265,266], providing a balance between resolution, coverage, specificity and costs.

Traumatic early life experiences, as mimicked by the MD model, are thought to follow the subtle methylation paradigm, rather than a blunt on/off mechanism [24]. Here, small changes in 5mC levels, fine-tuned gene transcription profiles, consequently altering the transcriptional, translational and proteomic landscape [24,242]. To increase the power of our analysis, the samples were pooled per group in order to obtain low-variance data. The genome-wide profiles indeed displayed great similarity across treatments and cytosine modifications, with only a limited number of clearly defined regions within which the 5(h)mC levels slightly differed, confirming our initial hypothesis that methylation changes would be subtle. For 5mC, most DMRs were detected between HD and MD, whereas very few were revealed for the C-MD comparison. For 5hmC, on the other hand, only a limited

number of DhMRs were detected for the HD-MD comparison, showing almost an opposite pattern in differential regulation between treatments. It is now generally accepted that the environmental manipulation nature of 5mC is coupled with an active demethylation process via 5hmC [36,37,66–70,234]. The overlap between DMRs and DhMRs was limited, yet the presence of multiple DhMRs implied that the 5mC pattern is being actively changed.

We observed that the majority of the D(h)MRs were not part of the gene body itself, but rather located in regions affecting gene transcription activity or intergenic regions/area's. This agrees with the observations made by ENCODE Project Consortium that intergenic regions often exert an important regulatory role in gene expression [268], Approximately, 56% to 70% of D(h)MRs contained multiple genes or RNA transcripts within the 200kbp annotation window, associating the D(h)MRs with more than one gene or RNA transcript. In 13% to 15% of the cases, the D(h)MRs could not be related to any known genomic entity at all, despite browsing a larger genomic area around the gene body, as previously suggested by McGowan et al (2011) [123]. None of the genes previously studied in MD or after early life stress such as *BDNF*, *AVP*, *CRH*, *NR4A1* and *Igf2* [24,25,62,123,155,233,239,240] were geographically linked to our DMRs.

As we have previously reported that MD did not affect *Nr3c1* methylation but impacted transcript levels, we performed a detailed examination of chromosome 18 surrounding *Nr3c1*. Expanding the investigation to approximately 200Mbp, we were unable to identify regions of differential methylation, in accordance with our prior results [155], and in contrast to the reports of McGowan et al [123]. A change in expression, despite no measurable change in the *Nr3c1* 5(h)mC, combined with the presence of D(h)MRs situated up- or downstream of other stress-related genes implies that MD impacts regulatory regions further away from the regulated genes. This pattern appears to be repeated genome-wide, although we observed the D(h)MRs to cluster in a similar manner to McGowan et al [123]. Gene ontology and network analyses of associated genes implied that differences in 5mC and 5hmC across treatments affected gene networks involved in oxidative phosphorylation, Huntington's disease, the generation of precursor metabolites and energy, electron transport, mitochondrial dysfunction and purine metabolism. Many of those were related to mitochondrial genes. Although methylation of mtDNA is somewhat controversial, it has received surprisingly little attention, however, the increasing evidence together with the recent discovery of DNA methylation machinery in mitochondria[269–271], suggests that it as a real phenomenon. There is increasing evidence that mtDNA methylation regulates mitochondrial functions and is involved in many physiological and pathophysiological processes, including neurodegenerative diseases such as Alzheimer and Parkinson disease, psychiatric disorders, dementia and cardiovascular diseases [269–271]. The detection of D(h)MRs covering multiple mitochondrial genes, added to the evidence of mtDNA's involvement in diseases.

Various D(h)MRs did not correspond to annotated genes, but associated with piRNA transcripts. Small non-coding RNAs (24nt and 32nt) known to form a complex by binding Piwi proteins, and to play a role in the regulation of cellular activities [50–55]. Although mainly studied in germlines, they are also present in somatic tissues such as heart, kidney, central nervous system and brain [51–54] Piwi/piRNA complexes are mainly known for their role in transposon silencing/suppression and gene expression regulation, both during and post-transcription, the preservation of the genomic integrity and hence have an impact on the cellular phenotypes [50–54,56]. Transposon regulation, either on genomic or epigenetic level, is thought to be important for CNS variability and mosaicism, and neural development and plasticity. Dysregulation of transposons could compromise cellular homeostasis and possibly resolve into pathology onset [50–52,54,56]. As the majority of piRNAs derive from transposons (genomic clusters), often located within intergenic regions [51,53–55], the differentially 5mC or 5hmC intergenic regions observed, could coincide with shifts in transposon regulation. Suggesting that maternal deprivation influenced piRNA levels, the piwi/piRNA complex formation and consequently its role in cellular homeostasis, change phenotype, and hence influence the CNS variability and mosaicism important in the neuronal regulatory system. Previous studies demonstrated environmental effects on transposon regulation, and subsequently gene expression, through epigenetic programming [51,52,54,56]. Overall, our data implies epigenetic regulation of piRNAs in rat hippocampi due to the early life environment as a significant adaptation mechanism.

Well known effects associated with the MD model, such as changes in HPA responsiveness, alteration of the adult stress reactivity, and conditioning behavioural and immune responses later on in life [104,155,245], have been observed in studies employing Long-Evans, Sprague-Dawley, and Wistar rat strains [245,272–275]. Whereas studies using Lewis or Fischer rat strains did not always induce the anticipated HPA axis effects [155,245,276–281], implying that the sensitivity to maternal deprivation is not only influenced by broad genetic factors and gender, but also by strain, i.e. very specific (inbred) genetic factors. Lewis and Fisher rats seem to be more resistant to the MD model's effects. Our data suggests that piRNA's and their expression act as mediators for the differences observed in HPA axis responsiveness per rat strain.

This study focussed on profiling epigenetic hallmarks. Combining 5mC and 5hmC profiling, enables a better understanding of their distribution across the same loci and the possible interaction between them. Yet, in order to get a better understanding of the effect/impact of these hallmarks, the epigenetic data analysis must in future studies be paired with expression data analysis. Adding transcriptomic data in the future would also confirm the supposed role of the piRNA we observed. Multiple D(h)MRs were associated with piRNA transcripts, suggesting the presence of piRNA in the rat hippocampus and its impact on transcription and translation regulation.

Expression data would be required to validate this hypothesis. Currently our results have not been validated by an additional technique such as microarray, yet by pooling samples during the library preparation, the analysis power increases and the data variance is significantly reduced, increasing the results' trustworthiness[246–248].

Overall, employing the genome-wide approach MeDIP-Seq, we demonstrated that maternal deprivation, or the handling of rat pups affected both their 5mC and 5hmC profiles. The (hydroxy)methylation changes induced by the early life environment were small, supporting the subtle methylation paradigm. D(h)MRs were not directly associated with genes previously involved in CNS modulation or the stress response. Both methylation and hydroxymethylation was primarily associated with non-coding RNA's such as piRNA in intergenic regions, as observed by the ENCODE Project, or within mtDNA. We suggest that differential 5(h)mC may alter the resulting adult phenotype through post-transcriptional regulatory mechanisms.

3.6. Acknowledgments

The authors would like to thank Hartmut Schachinger for his work within the Research Institute of Psychobiology and the Research Focus Point (Schwerpunkt) within the University of Trier. The authors are currently funded by the Fonds National de Recherche, Luxembourg [C12/BM/3985792], and work performed by us and reviewed here through the Fonds National de Recherche [PHD-80-053; TR-PHD-BFR07/127; TR-PHD-BFR-07/043] and the Deutsche Forschungsgemeinschaft (GRK 1389/1).

3.7. Authors' Contributions

J.D.T. designed and coordinated the study. T.B. performed the animal experiments and prepared the DNA samples. F.A.D.L., S.B.M. and F.B. carried out the data acquisition. F.A.D.L performed the data analysis and O.E.H. provided the code assessing the reads' genomic context. The manuscript was drafted by F.A.D.L. and J.D.T and critically reviewed by J.D.T and R.M. The study was performed under the supervision of J.D.T and C.P.M.

Chapter 4

Epigenetic consequences of early life H1N1 infection

This chapter has been submitted to *Epigenomics* and is under revision as:

Epigenetic consequences of early life H1N1 infection

Sophie A. Kirschner[1,2*], Liz Bohnenberger[1*], **Fleur A. D. Leenen[1,2]**, Oliver E. Hunewald[1], Linda D. Theisen[1], Sophie B. Mériaux[1], Claude P. Muller[1,2] and Jonathan D. Turner[1*]

[1] Department of Infection and Immunity, Luxembourg Institute of Health, Esch-Sur-Alzette, L-4354, Grand-Duchy of Luxembourg

[2] Department of Immunology, Research Institute of Psychobiology, University of Trier, Trier, D-54290, Germany

* These authors contributed equally to this work.

4.1. Abstract

Viral respiratory tract infections are highly prevalent during early life and have a long-lasting, profound, impact on both neurodevelopment and the subsequent risk for developing allergy and asthma. Little is known about the long term effects on either the innate or adaptive immune system. BALB/c mice exposed to Influenza A (H1N1, A/Puerto Rico/8/1934) at PND14 had significant increased serum IL-6, MIP-β and RANTES ($p<0.05$) when re-exposed to H1N1 in adulthood, confirming long-term immune programing. Similar effects were observed after re-stimulation with polyI:C and homotypic polyI:C programming and re-stimulation, suggesting preferential programming of the innate immune system. Reduced-representation epigenome sequencing identified a network of methylation changes common to both early life polyI:C and H1N1 programming, leaving a mechanistic 'trace' that remained visible throughout life.

4.2. Introduction

The developmental origins of health and disease (DOHaD) model describes lifelong conditioning of the response to any number of external environmental stimuli by an individual's early life experience[282]. One of the major perinatal events with long term consequences is infection. Although perinatal infections were previously thought to be primarily bacterial[283], acute viral respiratory tract infections (ARI) are now known to be the primary cause of hospitalisation of children under age 1[284]. This is further supported by epidemiological evidence suggesting ~20% of all young children present with influenza each season[285]. Additionally, recent evidence suggests that ARI such as influenza may also have significant long term consequences on both disease susceptibility[286] and neurodevelopment[287,288].

The immune system of both the human neonate and young infants is somewhat peculiar as it is "tolerant of novel, harmless environmental antigens"[289]. Although human infants are born with most immune cell types present, and in normal concentrations, they are dependent upon innate immune mechanisms as they have not developed an adaptive memory, leaving them particularly susceptible to intracellular pathogens[290]. There is growing evidence for the education or 'trained memory' of the innate immune system, particularly through epigenetic mechanisms[291], as well as rapid chromatin re-modelling. For example, effects on methylation of cytokine promoters such as IL-6 promoter during active influenza virus infection or immediately upon dsRNA treatment[292,293].

As DNA methylation remains plastic from conception for approximately 1000 days until 2 years old[294], this opens a window of opportunity for the long-term education of the innate immune system. Once established, differentially methylated regions (DMRs) should be conserved throughout life and create inter-individual epigenetic variation leading to an altered gene expression profile and a different adult phenotype such as an impaired susceptibility to viral infections throughout life.

The very high prevalence of early life viral ARI makes it essential to examine the long term consequences on the immune system. As such, this study was designed to demonstrate that the innate immune system can be educated by an early life early-life acute respiratory tract infection (Influenza A, H1N1) and that this is associated with genome-wide epigenetic changes within splenic lymphocytes.

4.3. Material and Methods

4.3.1. Animal Experiments

BALB/c mice (Harlan Laboratories, Boxmeer, Netherlands) were maintained at 40 ± 5% relative humidity, 22 ± 2°C under timed 12h light/dark cycles in same sex cages. Food and water were available *ad libitum*. New-

born litters were generated by housing breeding couples together for one week, and females were subsequently housed independently until parturition. All animal experiments were performed in compliance with the European Communities Council Directive 86/609/EEC and all national ethical guidelines and regulations.

4.3.2. Cells and Viruses

Influenza virus was cultured as previously reported[295]. Briefly, influenza A virus A/Puerto Rico/8/1934 Influenza H1N1 was grown on MDCK cells in serum-free EMEM supplemented with 2mg/ml BSA (Lonza) and 2µg/ml L-1-tosylamido-2-phenylethyl chloromethylketone (TPCK) treated trypsin (all cell culture reagents Lonza Verviers Belgium). The half maximal mouse lethal dose (MLD50) of A/Puerto Rico/8/1934 Influenza (PR8 pH1N1) was determined on seven week old female pathogen-free BALB/c mice (≥ 4 animals per group) by monitoring body weight and rectal body temperature daily (TH-5 Thermalert Monitoring Thermometer, Phymep, France). Animals were sacrificed if body weight loss exceeded 25%.

4.3.3. Tissue Culture Infective Dose (TCID50)

Half maximal tissue culture infectious doses per ml (TCID50) were used to determine virus titres. Quadruplicates of virus culture supernatant were serially diluted and incubated for 3 days on MDCK cells at 37°C and 5% CO2. Cytopathogenic effects were scored and the TCID50 was determined using ID50 (v5.0, http://www.ncbi.nlm.nih.gov/CBBresearch/Spouge/html_ncbi/html/index/software.html#1). TCID50 was subsequently used in the titration of MLD50

4.3.4. Perinatal Infections and Challenges

On post-natal day 14, pups were randomly assigned to H1N1/Controls (intranasal, i.n.) or polyI:C/Controls (intraperitoneal i.p.) groups as in Fig. 39. The discovery group received 50µl PR8 H1N1 at 1MLD50 or an equivolume of sterile PBS (i.n.). The innate immune validation experiment received polyI:C (0.75mg/kg b.w. at 0.1µg/µL, i.p.) or an equivolume of sterile saline solution (7.5µl/g b.w., 0.9% NaCl, i.p.). Mothers and pups remained undisturbed until weaning at PND21. Animals were subsequently single-sex group housed.

4.3.5. Adult Re-exposure

At PND42, programmed and control mice were re-exposed to either H1N1 or polyI:C. For influenza infections, animals were intranasally infected with 50µl containing 1MLD50 H1N1 and 7 days post-infection, blood was collected from the retro-orbital vein of terminally anesthetized animals, clotted for 1h at 37°C and centrifuged for 30min at 1000rcf. Serum was aliquoted and stored at -80°C until analysis. For polyI:C challenge, animals received an i.p injection of polyI:C (0.75mg/kg b.w. at 0.1µg/µL). 2h post injection, animals were euthanised. Spleens were isolated from all animals and splenocytes isolated by Ficoll-Paque PLUS (VWR,

Leuven, Belgium) gradient centrifugation for 20min at 72g and stored in RNA*later* (Qiagen) at -80C until DNA extractions.

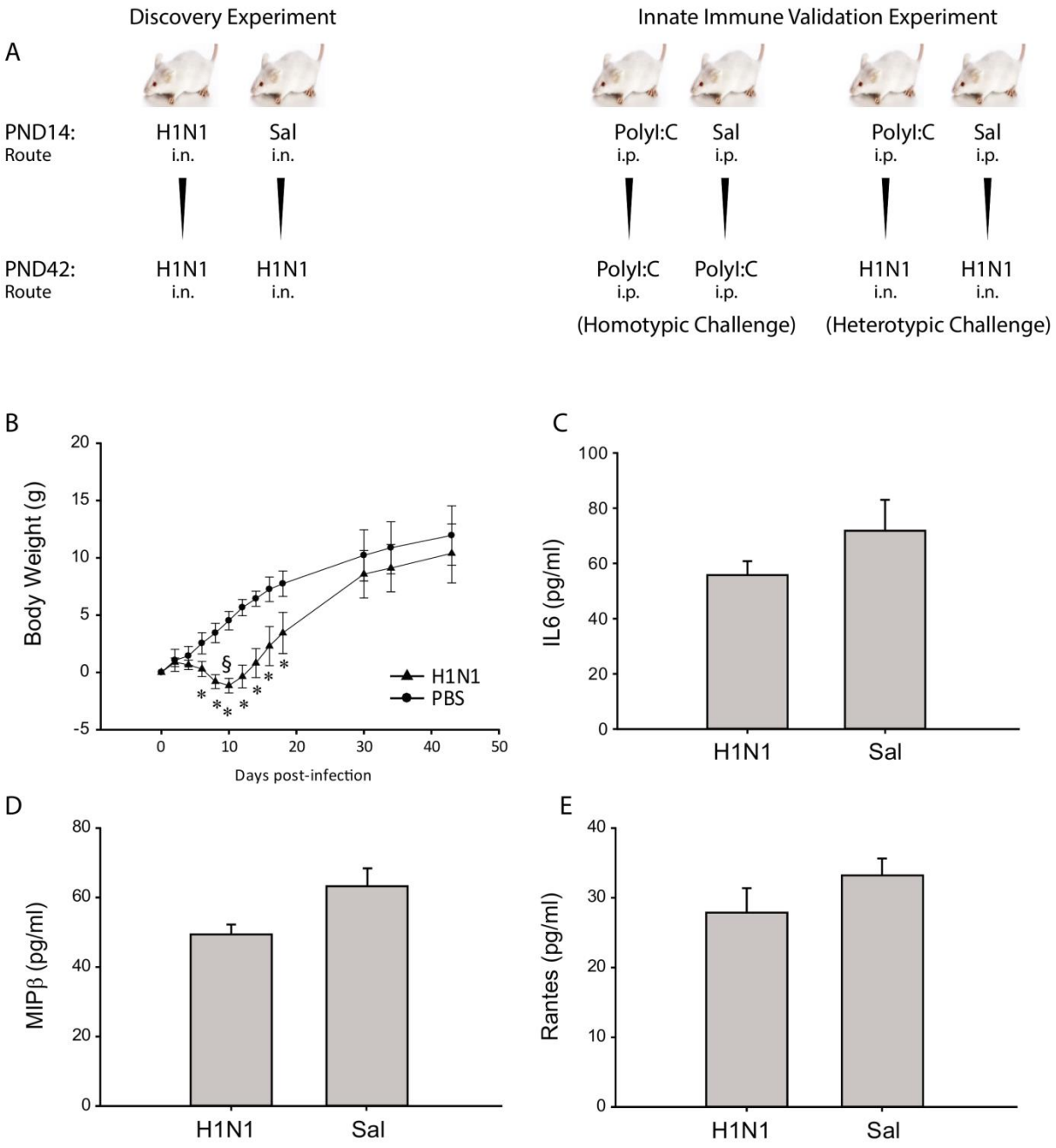

Figure 39: Early life viral infection mouse model: experimental paradigm and adult immune response to H1N1 re-exposure. (A) Paradigm for the discovery and innate immune validation experiments. PND, post natal day; i.n, intranasal; i.p. intraperitoneal. (B) Body weight evolution of mice, from the discovery experiment, exposed to H1N1 ( ) or saline ( ) at PND14. Statistical significance was measured by using 2-Way repeated measures ANOVA (time effect - $p < 0.001$; time*group effect – $p<0.001$) and Holm-Sidak pairwise comparison with overall significance level of 0.05 (*). (C) Serum IL-6 cytokine level measured at PND49, 7 days post-influenza infection (* $p<0.05$) (D) Serum MIP-β levels measured at PND49, 7 days post-influenza infection (* $p<0.05$) (E) Serum RANTES levels measured at PND49, 7 days post-influenza infection (* $p<0.05$).

4.3.6. Cytokine Analysis

Cytokine analysis. IFN-ɣ, IL-1α, IL-1β, IL2, IL-6, IL-10, IL-12p70, MCP1 (CCL2) and MIPβ and RANTES were measured on 50µL 1:4 diluted serum samples using the BD Cytometric Bead Array (CBA) Mouse Flex Sets (BD Biosciences, Erembodegem, Belgium). Data were collected on a BD FACSCanto II flow cytometer and analysed with FCAP Array Software according to the manufacturer's instructions (BD Biosciences).

4.3.7. DNA Extraction and Methyl-Seq

Genomic DNA (gDNA) was extracted from the splenocytes of saline and H1N1 infected mice at PND49 by PureLink™ Genomic DNA Mini Kit (Invitrogen). DNA samples from all animals were quantified (Nanodrop ND-100, Thermo Fisher Scientific) and group pools were made containing an equimolar proportion of each DNA sample at a final concentration of 0.1µg/µl. Pooled DNA (2µg per pool) was digested using 5 units of HpaII or MspI (NEB) for 6 hours and an additional 5U of enzyme was added to the mix after 3h at 37°C in CutSmart® Buffer (NEB) with agitation at 300rpm. After agarose gel size selection (100-400bp, MinElute Gel extraction kit, Qiagen) digested DNA was eluted in 20µl of Low TE.

As DNA methylation changes were expected to be small, their analysis benefits from low-variance data. Variance can be reduced by pooling multiple samples per group. This is described by the equation of the variance of the estimator of the true distribution mean, $\theta = 1/n_p * \left((\sigma_\varepsilon^2)/r_s + \left(\sigma_\xi^2\right)/r_a\right)$ with θ the true distribution mean, n_p the total number of pools, r_s the number of samples per pool, r_a the number of sequenced samples per pool, σ_ε^2 the biological variance and σ_ξ^2 the technical variation. Using this equation, pooling 2 or more samples decreases variance, in turn increasing the statistical power to identify differentially methylated regions[246–248]. As such, extracted genomic DNA was pooled from all animals within the same treatment conditions.

Sequencing was performed and the ratio of U-values analysed as previously reported[255,296,297] with minor modifications. Briefly, digested DNA fragments were end-repaired and sequencing libraries prepared using the Ion Plus Fragment Library Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. After AMPure bead purification (ratio 1:1.8, Beckman Coulter, Inc) fragments with both adaptors (25µl) were amplified for 8 cycles in a 130µl PCR reaction containing 100µl of Platinum PCR SuperMix High Fidelity and 5µl of Library Amplification Primer Mix. Thermocycling was performed for 5min at 95°C followed by 8 cycles of denaturation (15s, 95°C), annealing (15s, 58°C) and extending (1min, 70°C). After size selection (100-300bp; 1% agarose gel; MinElute Gel Extraction Kit, Qiagen) and quantification (Bioanalyser, Agilent Technologies), sequencing template preparation was performed with 17pM of DNA library on the Ion One Touch 2 using the Ion PGM

Template OT2 200 kit (Thermo Fisher Scientific) for fragments of ≈200bp using the manufacturer's fixed protocol. The percentage of template positive ISPs was evaluated by Ion Sphere Quality Control Kit and the Qubit® 2.0 Fluorometer (Thermo Fisher Scientific). The Ion Torrent PGM runs were performed with the Ion PGM 200 Sequencing kit (Thermo Fisher Scientific) on Ion 318v2 BC chips on the Ion PGM™ System. High throughput sequencing (HTS) reads were processed as previously reported by Kirschner et al[255]. Methyl-Seq DMRs were validated by qPCR.

4.3.8. Data Analysis

The standard parameters of Torrent™ Software (v4.0.2) were used for adaptor trimming and output data files was exported without quality filtering. Quality control was performed by counting the number of fragments starting with the cut restriction site '5'-CGG' and ending with the reverse complement restriction site 'CCG-3'. The fragments were mapped to the mouse genome (GRCm38/mm10) of the UCSC Genome Browser via Bowtie2 (Version 2.8.8; John Hopkins University). The fragments were then analysed for their location in a CpG island or a repetitive element and for their annotation to a known transcript or gene ID by comparing the fragments to available UCSC databases. The output of the fragment analyses of the 4 datasets (SH HpaII, HH HpaII, SH MspI, HH MspI) are summarized in 4 'Annotated Tables' as recently described[255]. The fragment positions and their respective coverages were interpreted using the previously reported strategy[297]: The R script merges the HpaII and MspI 'Annotated Tables' by their genomic start position. The script retains only the fragments where the MspI digestion had a minimum coverage (amount of times the same fragments were sequenced/present in the library) of 4. Methylation was estimated from the relative abundance of MspI and HpaII reads. Initially, HpaII reads were normalized per sample to MspI digestion by multiplying HpaII coverage with the ratio: Total MspI coverage/Total HpaII coverage. Then, the relative abundance was calculated from the arctangent evaluating the position of each point in a scatter plot showing the relationship between the number of MspI and normalized HpaII reads at each locus. The greater the angle between the line connecting the point to the origin and the X axis, the higher is the HpaII count compared to the MspI count and the less the sample will be methylated. In addition, the length of this line positively correlates with the confidence of the measurement. The transformation of these arc tangents into a 100 radian scale gives us a value named U-value (Unmethylation-value) for each locus in each of the 4 datasets merged in one table. A U-value of 100 represents the lowest average methylation percentage found in one samples whereas the highest methylation status is set 0. The ratio of the U-value (RU) of a locus in the HH dataset and the same locus in the SH dataset then illustrates the hyper- or hypomethylated status of this CpG locus. The closer RU is to 0, the higher is the hypermethylation and the

higher RU is above 1, the greater is the hypomethylation. RU values around 1 indicate no considerable changes in methylation.

4.3.9. Statistical Analysis

Group differences in cytokine levels were determined by one-way ANOVA at significance level 0.05. Post-hoc Bonferroni correction was applied to correct for multiple testing. If the normality distribution and equal variance assumptions were unfulfilled, a non-parametric Kruskal-Wallis test was used. Differences in methylation at individual CpG dinucleotides were determined with a two-tailed T-test or a Mann-Whitney Rank Sum test based on the fulfilment of distribution and variance assumptions at significance level 0.05. Post-hoc Bonferroni correction was applied to correct for multiple testing.

4.3.10. PCR Validation

Unpooled biological replicates from both H1N1-treated and control groups (1µg each) were digested with 5 units of HpaII (NEB) for 6 hours at 37°C in CutSmart® Buffer (NEB) with agitation at 300rpm. Digestions were supplemented with a further 5U of enzyme after 3h and diluted to 3ng/µl. PCR products were designed to span the 5' restriction sit, and reactions were performed using 2,5U Platinum® Taq DNA Polymerase, 20mM Tris-HCl (pH8,4), 50mM KCl, 200µM of dNTPs (Invitrogen™), 1X GelStar® Nucleic Acid Gel Stain (Lonza), and 1µl sample in a total reaction volume of 25µl. Thermal cycling (CFX96™ Real-Time System, BioRad, Hercules, CA, USA) conditions were 95 C, 2min; 44 cycles of denaturation 95 C (20s); annealing (20s); elongation at 72 C (20s). Primer sequences, MgCl2 concentrations and annealing temperatures are given in Table 9. *Sfi1* was chosen as a methylated reference gene and PCR efficiency was tested using a series of 7 10-fold dilutions revealing that all PCRs were within 1 cycle from the theoretical value. Amplification of the candidate restriction was normalized by the published $2^{-\Delta\Delta CT}$ strategy[298].

As negative control for methylation, whole genome amplification (WGA) was performed using the REPLI-g Mini Kit (Qiagen) following the manufacturer's instructions. As a positive control, 500ng of gDNA was treated with 10U/µl of CpG Methyltransferase (M.SssI; NEB) by adding 5µl of NEBbuffer2 (NEB) and 1X S-adenosylmethionine (SAM; NEB) diluted 1/20. The reaction was incubated for 4h at 37°C, then inactivated for 20min at 65°C. The methylated DNA was purified using AMPure beads (ration 1:1.8) and eluted in Low TE buffer. Both controls (1µg each) were digested with the methylation sensitive HpaII.

Standard errors of the means (SEM) were evaluated between the Ct values of Saline and H1N1 biological replicates, the errors of the ΔCt and ΔΔCt values were calculated via

$\sqrt{SEM(candidate\ gene)^2+SEM(reference\ gene)^2}$ and the error range of the $2^{-\Delta\Delta Ct}$ values was deduced by: $2^{-(\Delta\Delta Ct\text{-}error)} - 2^{-(\Delta\Delta Ct+error)}$.

Table 9: PCR primers sequences, $MgCl_2$ concentrations and annealing temperatures. (Supplementary Data)

			Relative RT-qPCR				
			Sequence (RT-qPCR)	**Product size (bp)**	**Tm (°C)**	**Mg (mM)**	**Oligo (µM)**
Hypermethylated candidates	Cacnb2	Fwd	5'-GGACCTTCCCTGGAGCCA-3'	191	64	1.5	1
		Rev	5'-CCTCTTCGCCCCAGGATC-3'				
	Cct7	Fwd	5'-ATAGAGTGGCGGAAGTGGTC-3'	162	62	1	1
		Rev	5'-CATCATCTTGGAAGCGGCTT-3'				
	Kazalt1	Fwd	5'-GACCCCAGCGCTAACTTCTA-3'	209	62	1.5	2
		Rev	5'-ACAGTGAGGTTAGCGTCCAG-3'				
	Grid1	Fwd	5'-CTGATCCCCGAGGTATTGCT-3'	150	62	2	1
		Rev	5'-AGGATCTCTCTGTGACCCCT -3'				
	Cd276	Fwd	5'-CCCTTTCAGAGCTGGCATTC-3'	170	62	2	0.5
		Rev	5'-CGGGGTGGGTGACTGATTAT-3'				
	Cdx2	Fwd	5'-TACTGCGGAGGACTGACAAA-3'	162	62	2	1
		Rev	5'-ACCATGTACGTGAGCTACCT-3'				
	Mybph	Fwd	Not performed				
		Rev					
	Cxxc5	Fwd	Not performed				
		Rev					
Hypomethylated candidates	DsysI3	Fwd	5'-GATCAGGTGGAGCGAATGGT-3'	226	59	1	0.5
		Rev	5'-CAGCTGGCACCACAAAAGAA-3'				
	Lrp2	Fwd	5'-CCCTCCTTCCTCACTTTGCT-3'	206	64	2	0.5
		Rev	5'-GTTTGGGGACTGGAATGCAG-3'				
	Mfhas1	Fwd	5'-TCTAAGCTCCTTGGACACCC-3'	152	62	2	0.5
		Rev	5'-GTTCTTTGTTGGTCCTGGCC-3'				
	Hacd2	Fwd	5'-CGTCTTTGCCCTCATCCAAG-3'	202	Weak amplification		
		Rev	5'-CCATTCCCCTTGGTCGCT-3'				
	Ddx56	Fwd	5'-CTACGGAGCTGATGAGTCCC-3'	244	59	2	0.5
		Rev	5'-AGTGAGTATGAGGCAGGACG-3'				
	Luc7l	Fwd	Not performed				
		Rev					
	Chsy3	Fwd	Not performed				
		Rev					
Ref. gene	Sfi	Fwd	5'-GGGCTCTGCTGTATGGGTAG-3'	230	60	2	0.5
		Rev	5'-TAGTAGGACGGGGTGGGTAG-3'				

4.3.11. Bisulphite Modified Pyrosequencing

Unpooled biological replicates (400ng) were sodium bisulphite converted using the EpiTect® Bisulphite kit (Qiagen) following the manufacturer's instructions and diluted to a final concentration of 3ng/µl. Primers were designed using 'Methprimer'[299] around the putative differentially methylated CpGs. RT-qPCR was performed as above. Sequencing primers and biotinylated primers were designed via 'PSQ Assay Design' (PyroMarkTM, Biotage), and together with $MgCl_2$ concentration and annealing temperatures are in Table 10. Pyrosequencing was performed on a Pyromark ID (Biotage), and analysed using Pyro-Q-CpG (Biotage).

Table 10: Sequencing primers and biotinylated primers, $MgCl_2$ concentration and annealing temperatures. **–B** Biotin (Supplementary Data)

			Bisulphite modified pyrosequencing					
			Sequence (PCR)	**Product size (bp)**	**Tm (°C)**	**Mg (mM)**	**Oligo (µM)**	**Sequence (Pyrosequencing primers)**
Hypermethylated candidates	Cacnb2	Fwd	5'-GGGTTTAGAAGTTTTAAAGAATTGG-3'	271	60	2	0.5	5'-TAGAAGTTTTAAAGAATTGG-3'
		Rev	5'-CAACAACTCCATCTAACTCTCCTAC-3'**- B**					
	Cct7		No Methprimers available					
	Kazalt1		No Methprimers available					
	Grid1		No Methprimers available					
	Cd276		No Methprimers available					
	Cdx2	Fwd	5'-AGGGTTGAAAGGTGTATATTAAAGT-3'	227	60	2	0.5	5'-TTATATTAATATTATAAGGG-3'
		Rev	5'-AAAAAAACCTCAAAAACCCAAAC-3'**-B**					
	Mybph	Fwd	**B-** 5' ATTTGGGTGAGATAGTATTTTTTTT 3'	292	56	3	0.1	5' CTACCCAACTCATAAAAAAT 3'
		Rev	5' TCTCCCCTACCCAACTCATAAA 3'					
	Cxxc5	Fwd	5' GGTATTAATGTTTTTTTTATATGGTTT 3'	199	56	2	0.5	5' TAATTGGGTTTTTAGAGGTG 3'
		Rev	5' TAACTCCCAATATACCTAACACC 3' **-B**					
Hypomethylated candidates	DsysI3	Fwd	5'-AATAGTAAATTTTTTTAGTTTTTTTT-3'	Weak amplification				
		Rev	5'-ATCTCAAATCAACCATTCCTACCT-3'					
	Lrp2	Fwd	5'-TTTTTTATTTTAAATAATTTTTTTT-3'	Weak amplification				
		Rev	5'-ATACTTTTCTTAAAAACTTTCCCTTTC-3'					
	Mfhas1		No Methprimers available					
	Hacd2	Fwd	5'-GGAAGTTGGGAGGAAATTATTTTTA-3'	186	58	2	1	5'-ATGGAGGTTTTAGTGGGAGTTGTT-3'
		Rev	5'-CACCTCAAACCCAACTAAACTCTAC-3'**-B**					
	Ddx56	Fwd	**B-** 5'-GTTATTTAAAATTATAATAGGATGG-3'	187	54	2	0.5	5'-TTCAAATTACTTAACTCTCA-3'
		Rev	5'-CAAAAAAATAAAAAATAAAATACCC-3'					
	Luc7l	Fwd	5' AAAATAGGAGGTAGGATTTTAGTTTTATAA 3'	146	55	2	1	5' TGATTTTGGGGATTATAGGT -3'
		Rev	5' AAATCAACTAAACTTAATTTTAACC 3' **-B**					
	Chsy3	Fwd	5' AGTGGAGTTAGAGGAGGGGATAT 3' **-B**	299	59	1	0.5	5' CTAAATCCACCATTCTCCTC 3'
		Rev	5' ACTACCACCCTATCACAAAC 3'					

4.4. Results

Neonates that survived the infection started to lose body lost weight from D3 to D10 post infection, subsequently gaining weight but remained significantly lower than in saline treated control until D43 (2-Way repeated measures ANOVA; time effect – $p>0.001$; time*group effect – $p<0.001$, Holm-Sidak pairwise comparison with overall significance level of 0.05; Fig. 39)

4.4.1. Early Life H1N1 Infection

In the initial discovery experiment BALB/c mice exposed to mouse adapted H1N1 Influenza A (A/Puerto Rico/8/1934) at PND14 suffered a mortality rate of 54% ± 21.4%.

At PND42 discovery experiment mice were re-exposing mice to H1N1 Influenza A and the cytokine response measured at 7 days post-infection (PND49). Significantly decreased levels of IL-6, MIP-β and RANTES ($p<0.05$) were observed in mice which received an early-life H1N1 challenge compared to saline exposed controls (Fig.39). Levels of IFN-ɣ, IL-1β, IL-4 and IL-12p70 remained unchanged compared to mice with naïve early-life.

Table 11: Summary statistics of Methyl-Seq sequencing runs. Sequences are available from the European Nucleotide Archive, accession number PRJEB17708 and samples ERS1434981-4. (Supplementary Data)

		Total Reads	cGG&CCG	Internal CCGG	% uncut CCGG	% cut CCGG
Saline	HpaII	5,913,973	3,283,603	722,388	12.21	55.52
	MspI	5,982,643	4,302,181	252,675	4.22	71.91
H1N1	HpaII	5,530,459	3,203,041	650,354	11.76	57.91
	MspI	5,073,180	3,694,192	195,922	3.86	72.84

4.4.2. DNA Methylation of Splenocytes

Methyl-Seq, covering 2.3x106 CpGs, was performed to identify the long-lasting epigenetic imprint of viral infection at PND14. At PND49, splenocytes from 7 saline and 6 H1N1 discovery experiment mice were collected, DNA was pooled in equimolar concentrations and digested by the isoschizomeric restriction enzymes MspI and HpaII. Library sequencing is summarised in table 11, and is in line with our previous Methyl-Seq libraries[255]. The aligned sequencing files have been deposited in the European Nucleotide Archive under the accession number PRJEB17708, and sample accession numbers ERS1434981-4.

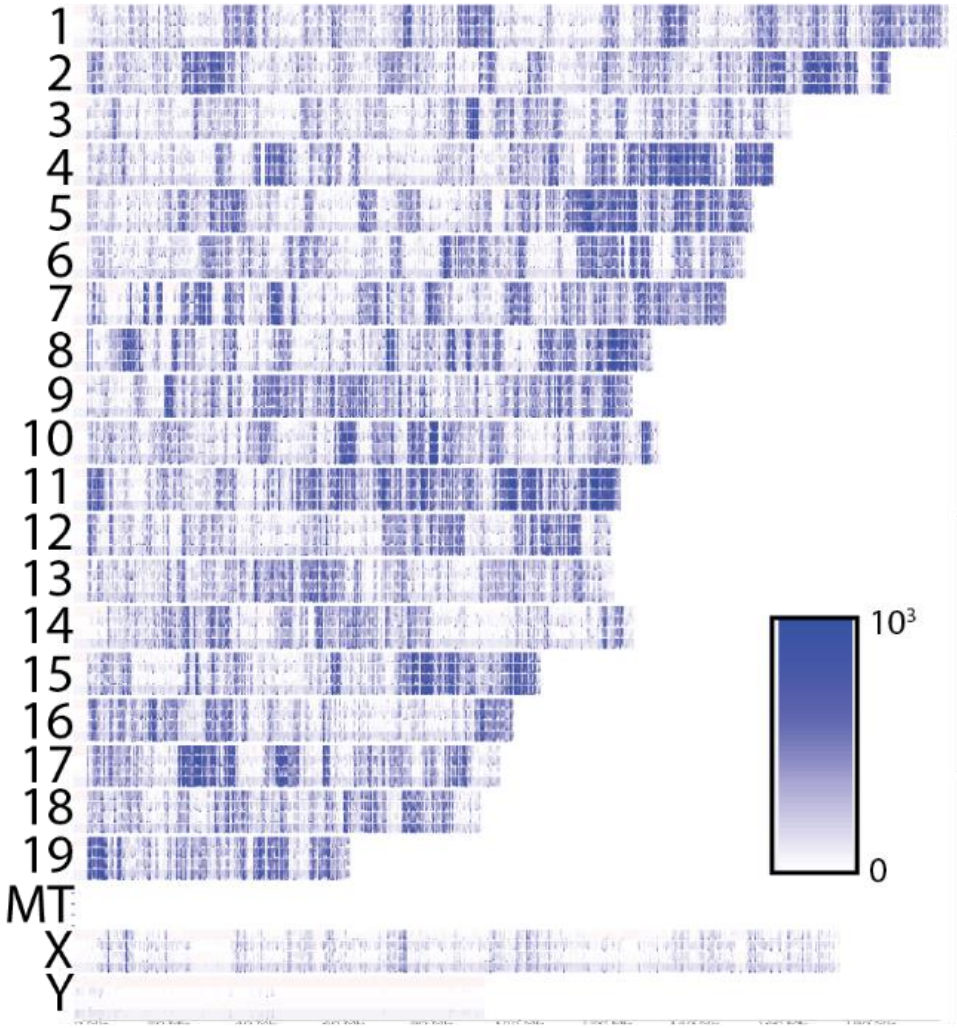

Figure 40: Genome-wide domainograph depicting the absolute number of reads per segment along the chromosomes, from zero (white) to multiple segments (10^3; dark blue).

4.4.3. Genome-wide Distribution of DMRs

The 2.3x106 CpG dinucleotides interrogated by MspI/HpaII digestionl[255] were uniformly distributed throughout the complete genome (Fig. 40). After excluding low coverage fragments[296], 1660 and 5011 hyper and hypomethylated loci were identified (Fig. 40). As sequencing was performed on pooled samples, differential methylation was plotted genome-wide as the ratio of the U values (Fig. 41A, [296]), and DMRs were distributed genome-wide, rather than concentrating on particular loci. Early life H1N1 exposure induced hypermethylation ranging up-to to 7 fold (RU=0.14 *CaCnb2*) change in methylation. Similarly, hypomethylated loci upon H1N1 exposure had fold changes up-to 9 (RU=8.8, *Wrap73*).

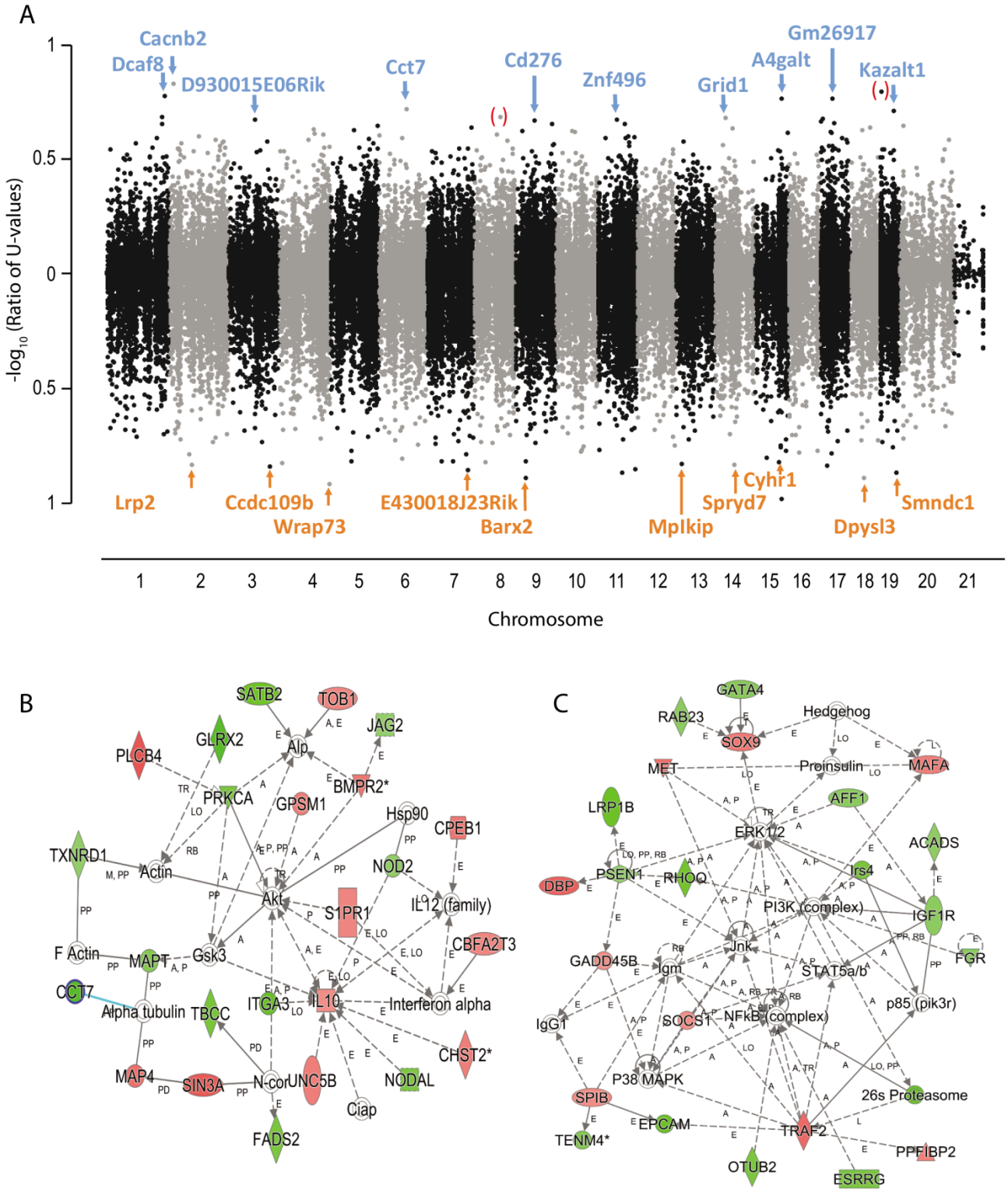

Figure 41: (A) (A) Manhattan plot visualising the distribution of the RU values across chromosomes. The top hyper- and hypomethylated candidates are represented in blue and orange respectively. (•) represent loci that could not be annotated to a gene. (B and C) Two top ranked Ingenuity Pathway Analysis functional networks of differentially methylated genes centred around *Akt* and. *NFκB.* Hypomethylated genes of each network are represented in green and hypermethylated in red.

4.4.4. Networks of Differentially Methylated Genes Point to Innate Immune Mechanisms

The top 400 differential methylated genes were organized in functional networks by Ingenuity Pathway Analysis® (Fig 41B-C). The statistically significant diseases and functions were biased towards related gene expression/transcription and activation functions (P-values from 5.6x 10^{-5} to 4.1x10^{-7}; 36-53 genes, 9-13% of the dataset, per function). The first network (Ingenuity score 33, 24 focus molecules; Fig. 41B) is biased towards the most strongly differentially methylated loci, containing 69% of the top 40 differentially methylated genes. This network, organised around *Akt*, suggests that *INFα*, *IL12* as well as *IL10* will be differently regulated as a result of the H1N1 exposure. IL12p70 was increased in both homotypic (H1N1/H1N1) and heterotypic (H1N1/polyI:C) scenarios (Fig. 39). The second network (Ingenuity score 32, 24 focus molecules; Fig. 41C) with 66% candidates formed a web around *NFκB*, *Jnk*, *STAT5a/b*, *PI3K*, *ERK1/2* and *IgM*. To find possible associations with influenza A infection, the top 400 differential methylated genes were compared to known H1N1 infection related genes. The 49 linked genes (12.25% of the dataset) formed a network containing *NFκB*, *IL6* and *10*, *TLR3*, *IFN-ɣ* and *β1* (Fig. 42).

Figure 42: The top 400 differentially methylated genes organised in functional networks using Ingenuity Pathway Analysis®. (A) Providing links between genes associated with influenza A (white) and the 400 differentially methylated candidate genes (green/red). (B) Ingenuity network created from the subset of linking genes (panel A) in the 2 gene groups. Hypomethylated genes of each network are represented in green and hypermethylated in red. (Supplementary Data)

4.4.5. Validation of Differentially Methylated CpGs

Based on their functional relevance combined with a high difference in methylation, 11 candidate positions were chosen for validation by RT-qPCR targeting the putative differentially digested restriction sites: *Cacnb2*, *Cct7*, *Kazalt1*, *Grid1*, *Cd276* and *Cdx2* for hypermethylation and *Dpysl3*, *Lrp2*, *Mfhas1*, *Hacd2* and *Ddx56* for hypomethylation. Six were validated by relative quantification (Fig. 43). Four candidate positions were additionally chosen for validation by bisulphite modified pyrosequencing targeting the putative differentially methylated CpG. Three out of the four genes tested (*Cdx2*, *Hacd2*, *Ddx56*) showed the correct direction of change in mean methylation level, even if absolute levels were low and did not reach statistical significance (<1%, 0.3<p<0.1, Fig. 44).

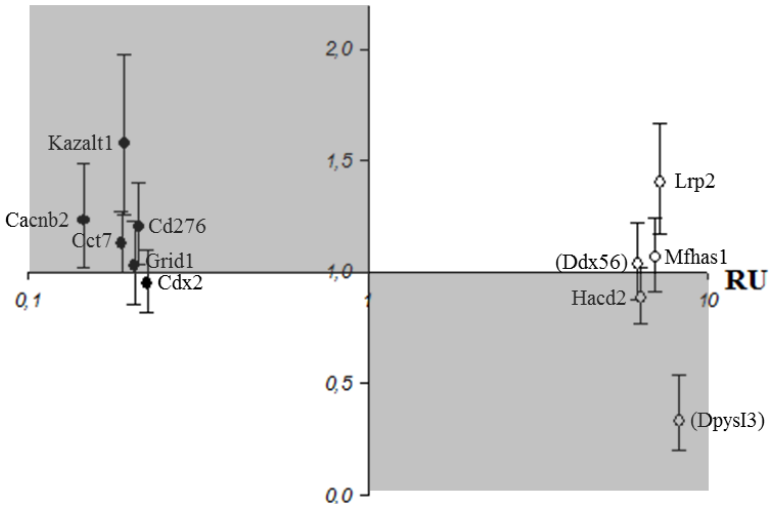

Figure 43: Normalised fold change in gene amplification $2^{-\Delta\Delta Ct}$ values) in comparison to the RU values of the 11 candidate genes. Grey quadrants indicate validation of the hyper- or hypomethylated candidates (RU>1 and $2^{-\Delta\Delta Ct}$<1 are hypomethylated candidates (º); RU<1 and $2^{-\Delta\Delta Ct}$>1 are hypermethylated candidates (•)). (Supplementary Data)

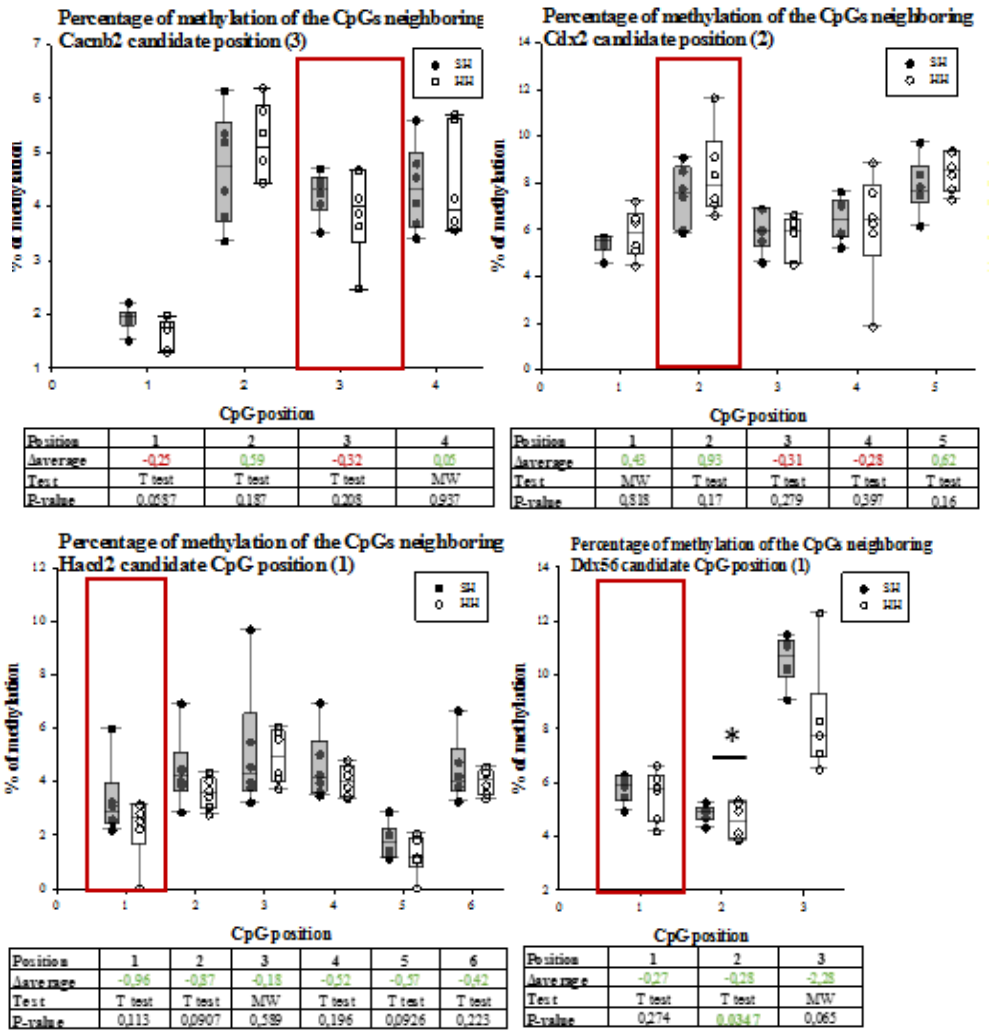

Figure 44: Methylation levels of the candidate genes measured by pyrosequencing. Differences in CpG methylation levels between control and H1N1 programmed mice for 4 Methyl-Seq candidate genes *Cacnb2* (A), *Cdx2* (B), *Hacd2* (C) and *Ddx56* (D). T-test or Mann-Whitney(MW) if assumptions were not fulfilled (*≤0.5; **≤0.01). (Supplementary Data)

4.4.6. Programming is not H1N1 Specific

Innate immune system programming was confirmed in the innate immune validation experiment (Fig. 39) by adult exposure at PND42 to polyI:C of both H1N1 and polyI:C programmed mice. Heterotypic polyI:C re-stimulation of H1N1 programmed mice significantly increased IL1α, IL2, MIPα, and IL12p70 levels (Table 12, $p<0.05$). Similarly, homotypic polyI:C exposure and re-stimulation significantly increased levels of MCP1, IL12p70 and RANTES (Table 12, $p<0.05$). Significantly, IL12p70 was increased in both homotypic and heterotypic scenarios, and is a key element of the predicted network of epigenetically modified genes (Fig. 41B).

Table 12: Cytokine response to polyI:C stimulation *in vitro* after early life exposure to either H1N1 or polyI:C. [a] – One-way anova; [b] – One-way anova on ranks

Early life	Mean (SD)	Mean (SD)	P-value[a]	
H1N1	Control (saline)	Exposed		
	IL6	14.2 (4.8)	16.7 (2.9)	0.122
	IL2	9.2 (1.6)	11.4 (1.8)	0.0159
	MCP1	889.1 (333.4)	1006.3 (574.7)	0.630b
	IL10	40.9 (12.8)	49.4 (19.4)	0.253
	MIPα	15.4 (4.4)	20.4 (4.6)	0.0242
	TNFα	26.5 (10.5)	31. 9 (10.2)	0.170
	MIPβ	200.2 (60.0)	259.0 (74.1)	0.0567
	IL12p70	24.8 (7.9)	38.9 (13.7)	0.0269
	RANTES	1489.5 (602.6)	1215.2 (615.6)	0.184
	IL1α	9.9 (1.0)	14.2 (1.7)	0.000273
	IL1β	20.5 (6.2)	33.896 (15.260)	0.0667
PolyI:C	Control (Saline)	Exposed		
	IL6	14.2 (4.8)	16.7 (2.9)	0.937b
	IL2	13.3 (1.7)	13.6 (1.2)	0.363
	MCP1	1284.8 (363.3)	2133.0 (357.2)	0.0011
	IL10	60.0 (12.2)	73.7 (7.7)	0.065b
	MIPα	22.7 (9. 5)	20.4 (0.7)	0.699b
	TNFα	39.8 (5.4)	48.4 (9.4)	0.156
	MIPβ	320.7 (111.2)	428.9 (117. 9)	0.206
	IL12p70	51.9 (20.4)	56.6 (20.4)	0.050
	RANTES	1088.5 (181.3)	1666.7 (294.8)	0.00109
	IL1α	16.5 (1.1)	15.7 (0.90	0.59
	IL1β	37.8 (3.3)	45.2 (18.2)	0.39

4.4.7. H1N1 Programmed Loci are Programmed After polyI:C

Pyrosequencing was performed on four target genes *Cacnb2*, *Cdx2*, *Hacd2* and *Ddx56* after early life polyI:C exposure. Methylation levels at 16 of the 18 CpG dinucleotides examined were significantly reduced at PND42 (Fig. 45, $p<0.05$), and 13 CpG dinucleotides remained statistically significant after Bonferroni post-hoc correction ($p<0.0031$, Fig. 45).

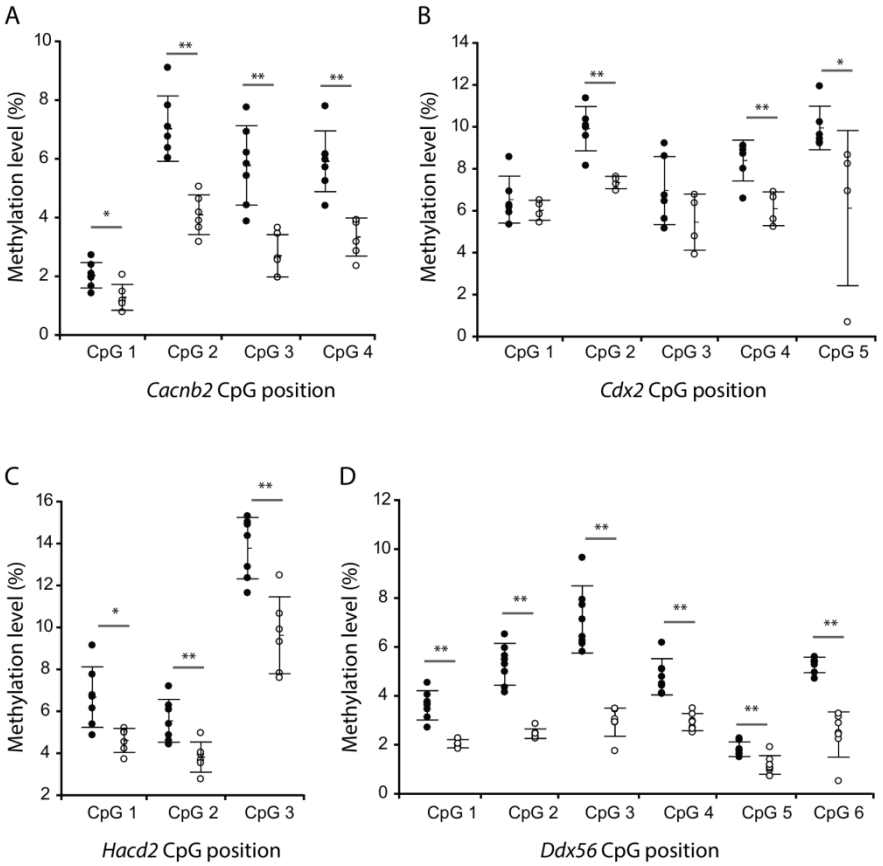

Figure 45: Methylation levels of the candidate genes measured by pyrosequencing. Differences in CpG methylation levels between control and PolyI:C programmed mice for 4 Methyl-Seq candidate genes *Cacnb2* (A), *Cdx2* (B), *Hacd2* (C) and *Ddx56* (D). T-test or Mann-Whitney(MW) if assumptions were not fulfilled (*≤0.5; **≤0.01).

4.5. Discussion

In this study we demonstrated that early life H1N1 infection modulated cytokine secretion after H1N1 re-exposure in adulthood. This appears to be a learned or programmed innate immune response as a similar effect was observed after heterotypic adult re-stimulation (H1N1/polyI:C) and homotypic polyI:C programming and re-stimulation (polyI:C/polyI:C). Reduced representation epigenome sequencing identified a network of methylation changes common to both early life polyI:C and H1N1 programming through which early life immune challenge left a mechanistic 'trace' that remained visible throughout life.

Since the seminal observation of Ellis et al[300] that early-life exposure to the dsRNA mimetic polyI:C significantly attenuated the febrile response to a subsequent (adult) polyI:C challenge coupled to an exaggerated corticosterone response[300], efforts have concentrated on the neurodevelopmental and neuroimmune consequences of polyI:C exposure rather than long term immunological effects. In the equivalent bacterial paradigm, LPS exposure leads to long-term programming of both the febrile response and the HPA axis. Such exposure increases the HPA axis response, correspondingly attenuating the cytokine response (e.g.

TNF- α, IL-6 and IL-1β)[300] and reducing the fever response to subsequent adult re-exposure[301]. Adrenalectomy and corticosterone supplementation or pharmacological glucocorticoid receptor blockage (e.g. RU486) abolished this attenuation of cytokine secretion suggesting heightened HPA axis responses mediate this effect.

In our model, early life influenza infection similarly modulated the cytokine response to homotypic re-exposure. Since influenza infection is primarily cleared by the innate immune system (reviewed in Schmolke et al (2010)[302]) even though the adaptive immune response is essential for developing immunological memory[303,304], higher IL-6 levels are known to be protective against primary H1N1 infection, enhancing both innate immunity and viral clearance[305]. Similarly, RANTES (CCL5) is upregulated in airway epithelial cells upon H1N1 infection[306], recruiting protective eosinophils to the site of infection, decreasing viral loads [307]. Since influenza A is known to signal though TLR3 a dsRNA-specific sentinel located on the endosomal membrane[308], we hypothesised that this was primarily an innate immune process. Confirming this hypothesis, exposure to the TLR3 specific stimulant polyI:C in adulthood induced clear differences in cytokine levels after early-life H1N1 exposure, although this was surprisingly in opposite directions to the homotypic polyI:C challenge. This can be explained by the more complex recognition of H1N1 infection by the innate immune system compared to polyI:C. Throughout the replication cycle of H1N1 in cells, at least three different receptor families can recognize the viral infection: TLRs (3, 7 and 8), RIG-1 and the NOD-like receptor family member NOD, LRR- and pyrin domain-containing 3 (NLRP3)[309] compared to the unique TLR3 for polyI:C. Although the effects of homotypic polyI:C and H1N1 programming on the cytokine production did not converge, epigenetic differences were clearly observed after both exposures. As the complete murine adaptive immunity response only develops at PND30[310], our study hereby supports the hypothesis of an epigenetic regulation of the innate memory called trained immunity[291,311].

Early life programming appears to lower the cytokine response to subsequent infections. Superficially, this may appear somewhat antithetical, however, on many occasions long term exposure to high cytokine levels has been shown to be deleterious. These changes are, however, accompanied by a tangled network of methylation changes surrounding innate immune regulatory genes, particularly *NF-κB*. Although our data are preliminary, we suggest that since both chemokines induced by H1N1 interact with CCR5 (RANTES/CCL5 and MIP1β/CCL4) and signal downstream to NFκB, an equilibrium has been found where despite lower circulating cytokine levels, epigenetic regulation of genes interacting with NFκB may result in the same physiological effect, while avoiding repeated exposure to high cytokine levels. Based on these data, we hypothesise that programmed differences in eosinophil recruiting chemokines such as RANTES and MIP1β secretion may partly explain the epidemiological link between early life ARI and asthma/allergy. We suggest that increased levels of

RANTES and MIPβ levels will further recruit eosinophils into the airway increasing the risk of asthma, therefore increasing the risk of asthma or exacerbating symptoms after disease onset.

As from prior data the methylation changes were anticipated to be small[166] and the number of available experimental samples was limited, they were pooled to reduce the intragroup variance and increase DMR detection power[246–248]. This study also leaves open the question of the mechanisms involved in the epigenetic modifications during viral infection. One possible mechanism related to influenza A virus infection and host responses was proposed by Fang et al. (2012)[312], attributing a role to miR29 in blocking DNMTs provoking demethylation in the *COX2* promoter and increasing the subsequent *IFN-γ* production[312]. Furthermore, the consequences of differential methylation need to be shown on both the mRNA and protein level in order to identify the functional consequences of early-life influenza infection.

In summary, these preliminary data show that a single early life exposure to viral infection programs the innate immune response to both homo- and hetero-typic challenges in adulthood that is observed as altered secretion of key mediating cytokines upon homotypic re-infection later in life, coupled with clear epigenetic differences in splenic lymphocytes.

4.6. Acknowledgements

We would like to thank Stéphanie Willieme for her technical assistance, and we are grateful to Hartmut Schächinger for his initiatives within the Graduate School of Psychobiology, and the Research Focus Point ("Schwerpunkt") within the University of Trier. This work was financially supported by AFR PhD grant N°1176135 to SK (Fonds National de Recherche, Luxembourg), the Luxembourg Institute of Health (LIH) and the Ministry of Higher Education and Research of Luxembourg. FL, JDT and CPM were funded through C12/BM/3985792 "EpiPath" (Fonds National de Recherche, Luxembourg). JDT is a management board member of the EU-funded COST action CM1409.

4.7. Authors' Contribution

J.D.T. designed and coordinated the study. S.A.K. and L.D.T. performed the animal experiments. S.A.K., L.B. and S.B.M carried out the data acquisition. The data analysis was performed by O.E.H., F.A.D.L, L.B. and S.A.K. S.A.K., J.D.T. and F.A.D.L. drafted and critically reviewed the manuscript. The study was performed under the supervision of J.D.T and C.P.M.

Chapter 5

General Discussion

This chapter has been published in part as (sections 5.2, 5.3 and 5.4):

DNA methylation: conducting the orchestra from exposure to phenotype?

Fleur A. D. Leenen[1,2], Claude P. Muller[1,2] and Jonathan D. Turner[1]

[1] Department of Infection and Immunity, Luxembourg Institute of Health, Esch-Sur-Alzette, L-4354, Grand-Duchy of Luxembourg

[2] Department of Immunology, Research Institute of Psychobiology, University of Trier, Trier, D-54290, Germany

Clin Epigenetics. 2016; 8(1): 92.

doi: 10.1186/s13148-016-0256-8

PMID:27602172; PMCID: PMC5012062

5.1. General Overview

In this thesis, we demonstrated that small DNA methylation changes have a genuine biological relevance for the model gene we studied (*NR3C1*). The glucocorticoid receptor displayed a higher transcription initiation variability than previously assumed and this transcription initiation pattern was affected by small shifts in methylation levels. Furthermore, this transcription microvariability was shown to affect the translational regulation and hence the relative abundance of protein isoforms. This suggested that external environmental stimuli acted through subtle methylation changes and transcription variability as a mechanism to fine-tune the total *NR3C1* protein levels. To expand the findings for *NR3C1* and investigate the relevance of small DNA methylation changes on a genome-wide basis, we introduced two epigenetic models. Rats subjected to the maternal separation paradigm, displayed long term behavioural and stress response changes. Complete methylome analyses identified many subtle differences, suggesting that the early life adversity (MD) 'programmed' the CNS via subtle changes in methylation levels. A second model, exposing mice to early life infection with H1N1, illustrated long-term immunological effects and displayed a network of subtle methylation changes, including innate immune regulatory genes. Indicating that early life viral infection 'programmed' the innate immune response via subtle DNA methylation changes genome-wide. Both models link small methylation level changes to either psychopathological or immunological effects, hence suggesting that the paradigm of small DNA methylation changes fits.

5.2. Towards a Mechanism Linking Subtle Methylation Changes to Phenotypes?[a]

The work presented in Chapter 2 builds upon many years research within the host group on the GR. The glucocorticoid receptor (*NR3C1*, GR), has well characterised transcriptional and translational variability. The association of receptor levels and variants with disease[162,163,185] has made it a particularly useful model to explore both the functional relevance and the effects of small methylation changes[153,164,166,187,200], and the association between methylation and pathology at the single gene level[105,127,152,313]. The *NR3C1* 5' structure, containing multiple alternative non-coding first exons (1A to 1J) with a multitude of transcription factor binding sides (Fig. 46A), was initially reported by to be responsible for the quantitative, spatial and temporal expression of the *NR3C1*[162,163,185]. Data presented in Chapter 2 however, demonstrated that *NR3C1*'s transcription was exceptionally permissive rather than being initiated at fixed positions (Fig. 46). We observed a total of 358 statistically significant transcription start sites (TSS) located in 38 contiguous loci in the absence of any particular stimuli, with a further 185 stimuli specific[166]. For instance, demethylation with 5-AZA-2'-deoxycytidine (AZA) had a profound influence on the TSSs used, with 127 stimuli-specific TSSs induced by demethylation. This

a. This section was published in Leenen, F. A. D., Muller, C. P. & Turner, J. D. DNA methylation: conducting the orchestra from exposure to phenotype? *Clin. Epigenetics* 8, (2016)

permissivity, covering a large 3kbp region, is called transcriptional microvariability (Fig. 46)[166]. Although such microvariability appears to be stochastic, in the case of the *NR3C1*, I showed that it had ait has a significant effect on translation. Small differences in TSS location (<10nt) within any given locus redirected ribosomes to initiate translation from internal (downstream) ATG codons, altering the balance of the translational GR isoforms produced (Fig. 46)l[32]. A shift in TSS location resulted in an altered mRNA secondary structure and half-life, and influences the overall translational efficiency in a "length-dependent, but sequence-independent manner" [164,166]. These data are also significant because they demonstrate how the *NR3C1* microvariability vastly inflates the associated proteome. The GR is classically cytosolic, however we have previously demonstrated that the membrane bound form of the receptor (mGR[88]) is derived from the classical *NR3C1* gene[187],. and further refined its molecular origin to the epigenetically regulated alternative first exon, 1D[164]. In Chapter 2 we demonstrated that the *NR3C1* microvariability influenced not only the final protein form, but also the final cellular distribution of the GR proteins (Fig. 46)[32]. Our data lead me to conclude that physiological differences in glucocorticoid secretion and response were the result of DNA methylation altering TSS/first exon usage, with the consequentially proteomic difference. Importantly, my data suggested that, at least for the *NR3C1*, neither single nor clusters of CpGs that are methylated switch off transcription of any particular splice variant, rather, they orchestrate the final proteomic landscape, and potentially alter the splicing internally or at the 3' end. Part of the publication Leenen, F. A. D., Muller, C. P. & Turner, J. D. DNA methylation: conducting the orchestra from exposure to phenotype? Clin. Epigenetics 8, (2016)

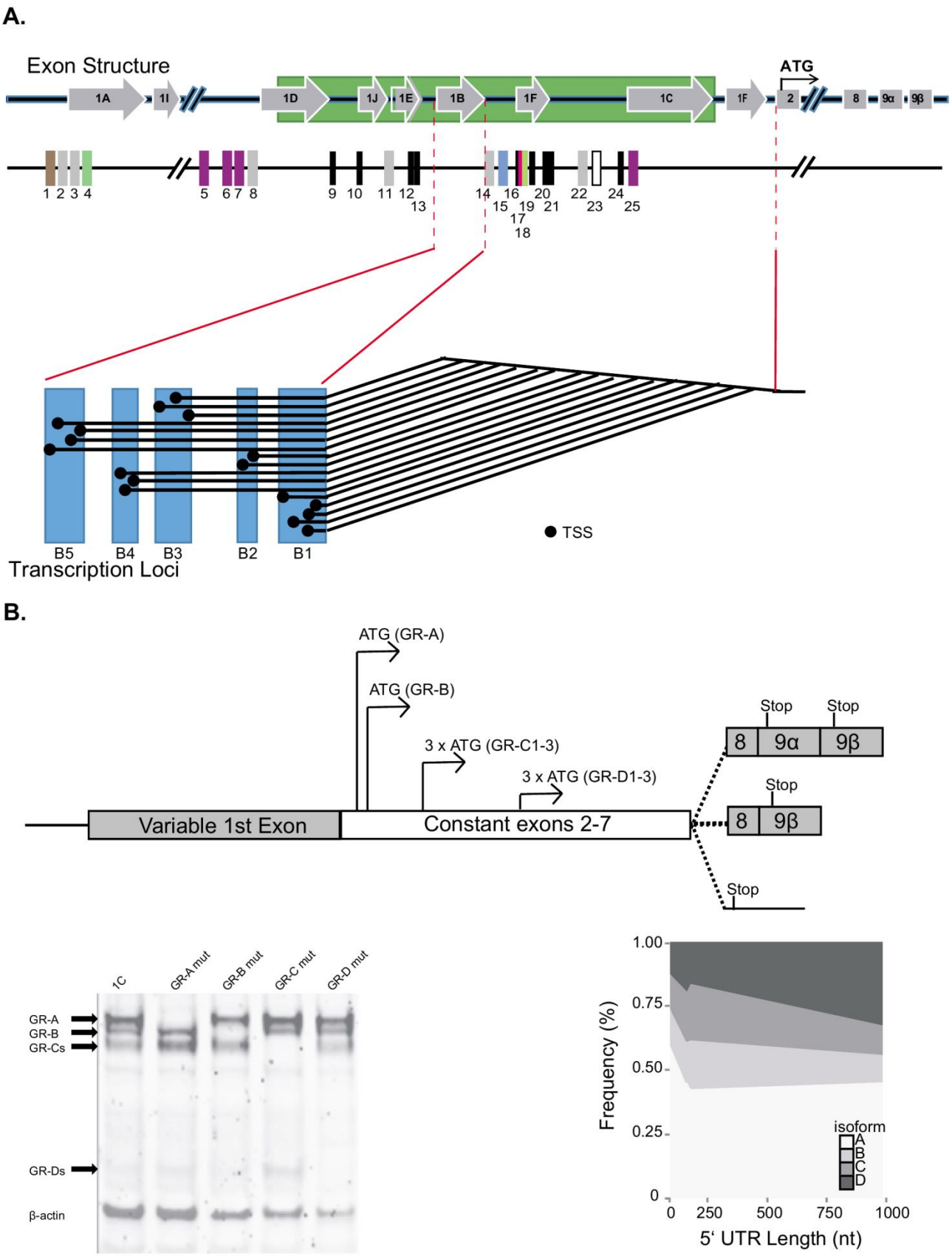

Figure 46: Panel A) A schematic representation of the *NR3C1* 5’ UTR structure, showing the alternative first exons (1A-1J, CpG island:), transcription factor binding sites (1-25), transcriptional loci (B1-B5), and microvariable transcription start sites (•). Transcription factor binding sites: ( ) IRF-1 and IRF-2 (position 1); ( ) glucocorticoid response elements (GRE, positions 2, 3, 8, 11, 14 and 22); ( ) c-Myb, c-Ets1/2 and PU1 (position 4); ( ) Ying Yand 1 (positions 5, 6, 7 and 25); ( ) Sp1 binding sites (positions 9, 10, 12, 13, 16, 19, 20, 21 and 24); ( ) Ap-1 (position 15); ( ) NGFI-A binding site (position 17); ( ) glucocorticoid response factor-1 (GRF-1, position 18); ( ) Ap-2 (position 23). Panel B) Structure of the GR mRNA with the internal ATG translation initation codons, a western blot demonstrating the different transcriptional isoforms (from [166] with permission), and the frequency of the different protein isoforms with increasing 5’UTR length (adapted from [164] with permission).[24]

5.3. Expanding the Mechanism from the *NR3C1* to the Complete Transcriptome and Proteome?[b]

It has been recognised for many years that both the complexity and phenotypic diversity increase as the relative size of non-coding genomic regions and the regulatory elements and variability within them increase throughout evolution[314,315]. Features like alternative first exons, transcriptional microvariability, and alternative in-frame downstream ATG initiation codons are found genome-wide. ~60% of all genes are thought to possess a highly variable 5' structure with many alternative first exons[172], and this transcription variability has been reported in many cases to be responsible for spatio-temporal gene expression patterns[167]. Similarly, multiple alternative in-frame ATG translation initiation codons within mature mRNA are found ubiquitously through evolution, occurring in many plant, invertebrate and vertebrate species[220–222].

In light of data on the origin and evolution of new TSSs and exons in different species[168,177,178,181] transcriptional microvariability is not unexpected, and now several reports have observed transcription starting over small multiple small loci genome-wide[172] and in model organisms[316]. Multiple alternative in-frame ATG translation initiation codons have been observed in a wide range of genes. Although no systematic review of their occurrence has been performed, they are thought to be ubiquitous, and cover both leaky ribosome scanning and internal ribosome entry[225]. These observations and the ubiquity of the features, made researchers suggest that the 5'UTR, together with intergenic regions and the 75% of the human and mouse genomes that are transcribed are the key to understanding the vastly inflated proteome[166,314,317]. It therefore seems logical that the mechanisms outlined for *NR3C1* above should be expandable to the complete transcriptome and proteome. Irrespective of whether 5' variability comes from the mRNA structure, the TSS location, transcriptional microvariability, or alternative mRNA splicing, this variability will give rise "to high complex and diverse transcriptomes and proteomes"[166] (Fig. 47).

In order to expand the subtle methylation paradigm genome-wide, reliable epigenetic animal models were required to confirm the relevance of subtle methylation changes and the applicability of the transcription microvariability as a proposed mechanism. The DOHaD paradigm describes the conditioning of individual's lifelong health trajectories by their in utero or early life experiences[30,44,83]. In Chapter 3 and 4, the MD model in rats and the perinatal infection model in mice respectively, were introduced to assess whether the subtle methylation change paradigm holds true genome-wide.

The MD model, mimicking depression, one of the best studied psychopathologies, was shown to induce long lasting behavioural effects, affect the stress reactivity, and to induce DNA methylation changes in promoter

b. Paragraph 1 and 2 of this sectionwere published in Leenen, F. A. D., Muller, C. P. & Turner, J. D. DNA methylation: conducting the orchestra from exposure to phenotype? *Clin. Epigenetics* 8, (2016)

regions of specific genes[155,239,240]. Although the main focus has been on the promoter region, differences in methylation occurred across the whole length of *NR3C1*, including intragenic regions or regions distantly located from transcription start sites[25,62,123,233,235,244], supporting a more wide-spread epigenetic programming due to maternal care. The data presented in Chapter 3 enabled a deeper insight in genes/regions susceptible to early life adversity genome-wide, as well as an evaluation of the effect of subtle methylation changes, by employing a genome-wide enrichment-based approach MeDIP-Seq. The 5mC and 5hmC profiles revealed a randomly distributed read density across all chromosomes. Similar regions tended to be (hydroxy)methylated across maternal care conditions, with only a limited number of regions exhibiting slightly different 5(h)mC levels. They were found in promoter, intragenic and intergenic regions, not all necessarily annotated. These findings were a first hint at the biological relevance of subtle DNA methylation changes for the CNS. None of the annotated D(h)MRs was linked to DMRs or genes previously reported by early life adversity studies[24,25,62,123,155,233,239,240]. Although no real changes could be detected in *NR3C1*'s 5(h)mC profile, the transcript levels did alter[155]. Additionally, the detected D(h)MRs, although not coinciding with previous revealed D(h)MRs associated with MD or early life stress, occurred both up- and down-stream of different stress-related genes. Suggesting that MD affected regulatory genes although more distantly linked to regulatory genes of the CNS and the stress response. Many of the D(h)Mrs were situated in mtDNA. Although mtDNA methylation only recently gained attention, it has been suggested to play a role in the regulation of mitochondrial functions, but also to be implicated in physiological and pathophysiological processes, including neurodegenerative diseases such as Alzheimer and Parkinson disease, dementia and psychiatric disorders[269–271]. The presence of multiple D(h)MRs in the MD model added to this suggestion. Various D(h)MRs detected in the MD model, associated with piRNA transcripts. Together with Piwi proteins they form a complex which is mainly known for its role in transposon regulation and gene expression regulation[50–56]. With the majority of piRNAs deriving from transposons, often intergenically situated,[51,53–55] and transposon regulation assumed to be implicated in CNS variability and mosaicism, and neural development and plasticity[50–52,54,56], the intergenic with piRNA transcript associated D(h)MRs could coincide with regulatory regions for transposon regulation. Implying that the epigenetic regulation of piRNAs as a consequence of the early life environment, acts as an adaptation mechanism. Overall, the data presented in Chapter 3 profiled epigenetic hallmarks by combining 5mC and 5hmC profiling. The by the early life environment induced changes in (hydroxy)methylation levels were small, hence supporting the subtle methylation paradigm. The detected D(H)MRs were not directly linked genes previously involved in CNS modulation or the stress response. Yet seem to be either situated in regions that were either distantly linked to regulatory genes of the CNS and the stress response, mtDNA or piRNA, which are both thought to exert regulatory functions directly or indirectly. Hence, the findings presented in Chapter 3 suggest that the subtle

(hydroxy)methylation changes caused by maternal deprivation may be involved in the regulation/modulation of the CNS, either direct or indirectly.

With early life infection being one of the major perinatal events with long term consequences[283–285], the mouse model subjected to perinatal viral infection with H1N1 presented a clinically and sociologically relevant model to study the biological relevance of subtle methylation changes genome-wide. The data presented in Chapter 4 demonstrated that early life H1N1 infection modulated the cytokine secretion upon homotypic re-exposure in later life. As the adaptive immunity response was not yet developed during early life viral infection, the model suggests that the innate immune response was modulated. The similar reaction upon re-exposure to either homo-typic or hetero-typic stimulation (TLR3 ligand polyI:C or H1N1) supported this. I hypothesised that a plausible mechanism underlying the development of this phenotype would be subtle changes in DNA methylation. Employing a restriction enzyme based approach (Methyl-Seq), revealed that DMR occurred in CpGs across the whole genome, presenting a uniformly genome-wide DMR distribution. They were found in promoter, intragenic regions, but also in not-annotated intergenic regions. The exhibited changes in methylation levels between control and H1N1 groups tended to be small, with a 7- and 8-fold change, for hyper- and hypomethylation respectively, as largest alterations. The changes could be associated with gene networks surrounding innate immune regulatory genes. A first one organised around *Akt* involving also genes such as *INFα*, *IL12* and *IL10*. A second network build included genes such as *NFκB*, *Jn*k, *STAT5a/b*, *PI3K*, *ERK1/2* and *IgM*. Overall, the H1N1 infected mouse model demonstrated that viral infections during early life program the innate adult response later on, with DNA methylation as underlying mechanism. As such, the H1N1 model shown that the subtle methylation change paradigm holds true for the immune system, by underlining the biological relevance of small methylation changes genome-wide in modelling the immunological response and thus the resulting phenotype.

The two models, maternal deprivation and early life infection both successfully induced subtle changes in DNA methylation, concurring with the central hypothesis of this thesis that during epigenetically sensitive periods, external environmental stimuli induce subtle changes in DNA methylation that will associate with the eventual phenotype. As such, both of these models can now be used in subsequent studies to further dissect the molecular mechanisms linking the environment to the phenotype. The techniques developed in Chapter 2 to assess RNA 5' microvariability will be a key element of such studies.

5.4. Re-defining a 'Gene'[a]

The significant increase in transcriptional and translational complexity observed for the *NR3C1* concords with the recent movement towards re-defining a "gene". While the definition of "gene" has changed considerably

a. This section was published in Leenen, F. A. D., Muller, C. P. & Turner, J. D. DNA methylation: conducting the orchestra from exposure to phenotype? *Clin. Epigenetics* 8, (2016)

over the last century, the current definition used worldwide for genome annotation is “a DNA segment that contributes to phenotype/function. In the absence of demonstrated function a gene may be characterized by sequence, transcription or homology”[318]. This definition has come under scrutiny over the last decade[319,320]. Large-scale sequencing projects such as ENCODE/GENCODE have identified several phenomena that are changing our perception of what a gene is, including universal alternative splicing, pervasive and intergenic transcription, and dispersed patterns of transcription regulation[321–323]. Gerstein et al metaphorically described the classical definition of a gene as “subroutines in the genomic operating system”[319]. This analogy was further broken down into the genome being a complete human “operating system” and with gene being a clear, well-defined “subroutines” where a genomic region is assembled as in a homologous manner to computer code, with transcription and translation considered the homologues of calling and running a subroutine. In this analogy gene elements (5’, 3’UTR, intron, exon etc...) were considered as the syntax. GENCODE and subsequent data have called this neat definition into question. The vastly inflated transcriptome and proteome suggest that the process is rather “higgledy-piggledy” or stochastic, with the gene “subroutine” very poorly defined with many starting points. Post GENCODE the definition of a gene was simplified taking into account this variability as “a gene is a genomic sequence (DNA or RNA) directly encoding functional product molecules, either RNA or protein”[319]. The two definitions can be compared to strict Boolean or fuzzy logic. This definition is amenable to the integration of data, such as ours, from the epigenetic regulation of the *NR3C1*, as it would appear that a combination of genetic and epigenetic variants underpin and orchestrate the “higgledy- piggledy” or fuzzy processes into a concerted, specific response to the external environment.

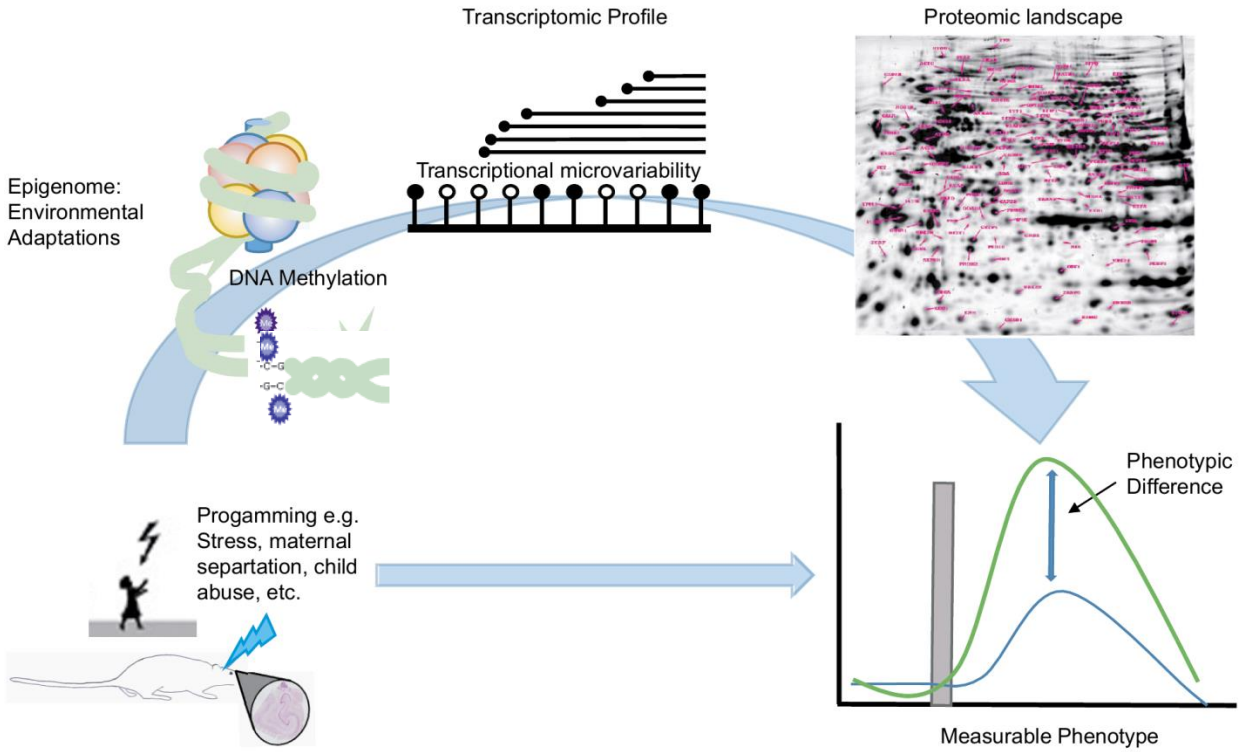

Figure 47: Proposed mechanism for the creation of phenotype diversity by environmental factors. Epigenetic marks, such as DNA methylation, are proposed to influence transcriptional variability and hence the proteomic landscape, resulting in a phenotype diversity.[24]

5.5. Concluding Remarks and Future Perspectives

In this thesis we have demonstrated *NR3C1*'s transcription was extremely permissive and did not start at well-defined, fixed TSS. A permissivity that significantly affected the translation. Small shifts in TSS usage, due to DNA methylation, altered the translational GR isoform balance, and affected the mRNA 's secondary structure stability, its half-life, and the translational efficiency. Which in turn influenced the final protein form and *NR3C1*'s final cellular distribution. Consequently, our data showed that subtle changes in DNA methylation altered the TSS usage and therefore translational variability and the resulting proteomic landscape. The ubiquitous distribution of genes with similar structure as *NR3C1*, suggested that transcription microvariability could be mechanism to fine-tune total protein levels that is more wide-spread throughout the genome.

Moreover, the data illustrated that maternal separation or handling of rat pups modulated their methylation and hydroxymethylation profiles, via subtle changes of (hydroxy)methylation levels. Hereby supporting the subtle methylation paradigm and being one of the first studies to our knowledge combining 5mC and 5hmC profile analysis. The detected differential regions were mainly located in intergenic regions, as observed by the

ENCODE Project Consortium, and often associated with more distant regulatory regions of the genome, non-coding piRNA's or mtDNA. Suggesting that they may exert regulatory functions directly or indirectly.

Furthermore, we have shown the subtle methylation paradigm also holds true for the immune system. Using the early life infection model, we demonstrated that a single early life exposure to viral infection programmed the innate immune response to both homo- and hetero-typic challenges of mice later on in life. The secretion of key mediating cytokines upon homotypic re-infection later in life was altered and could be coupled with clear epigenetic differences in splenic lymphocytes. Therefore supporting the subtle methylation paradigm and the biological relevance of such small DNA methylation changes.

Overall, it has become clear that DNA methylation occurs either as discrete hyper- and hypo-methylation coupled with a clear on/off switch of genes as often observed in oncogenes, and easily dissected molecular mechanisms, or in a second paradigm as a more subtle complex process where small (<10%) methylation changes have been associated with divers phenotypes and epigenetic programming events. In this thesis we provided two suitable animal models, the MD model and early life viral infection model, to further investigate the biological relevance and function of small DNA methylation differences. Our data demonstrated that the alterations in methylation are subtle. Yet in order to decipher the underlying mechanism and the functional consequences of these, future studies employing these models would need to focus on the effect of the methylation changes on both translational and proteomic level.

Despite the observational association studies' aim to increase our understanding of the environmental impact on phenotype development, the underlying mechanisms linking subtle methylation changes to an eventual phenotype remained unaddressed (Fig. 47). Consequently, hampering our interpretation of the associations with subtle changes in methylation due to a lack of data addressing the true biological relevance and function of such small differences. We are now starting to gain insight into the function and relevance of such small changes in methylation from genes, such as *NR3C1*. These data suggest that the 5′ UTR is the key to controlling gene expression. Small changes in methylation throughout this region impact mechanisms such as alternative splicing and transcriptional microvariability, altering enhancer and insulator use, and the function of regulatory elements. Methylation of single CGs affect the TSS usage within a gene promoter region, i.e. silence a specific location, whereas methylation of multiple closely related CG's will rather silence a transcription loci, i.e. a whole site of adjacent TSSs. Recent studies demonstrate that small changes in methylation levels seem to be regulated in clusters rather than single CpGs. But whether they act as single CpGs or in clusters, these small changes do not function as an on/off switch, rather redistributing the transcriptional landscape, affect translational isoform production, and orchestrating the final proteomic landscape. Hence, this thesis suggests

that the observed *NR3C1* transcription microvariability can be expanded as a genome-wide mechanism through which subtle DNA methylation changes can act on the resulting phenotype.

Technologies such as high-throughput sequencing (HTS) have enabled researchers to study these subtle methylation changes in greater detail genome-wide. In order to confirm the subtle methylation paradigm in the MD and early life viral infection models and to study the transcription initiation in *NR3C1* in greater depth, we combined existing techniques such as 5-RACE and MeDIP with HTS. Techniques that can readably employed in future epigenetic studies analysing genome-wide DNA methylation patterns or the transcriptional variability. Both techniques also will also be valuable for the study of alternative covalent DNA modifications, such as 5mC, 5fC or 5CaC. Enabling an epigenome-wide overview, as well as a deeper insight in the DNA methylation mechanisms underlying changes. Additionally, our techniques in combination with the emerging single cell HTS, will allow a far more in depth analysis of this phenomenon and its importance overall as well as on the single-cell level.

References

1. Turner, J. D., Kirschner, S. A., Molitor, A. M., Evdokimov, K. & Muller, C. P. in *International Encyclopedia of Social and Behavioural Sciences* (ed. Wright, J. D.) 839–847 (Elvsevier, 2015).
2. Eckhardt, F., Beck, S., Gut, I. G. & Berlin, K. Future potential of the Human Epigenome Project. *Expert Rev. Mol. Diagn.* **4,** 609–618 (2004).
3. Jirtle, R. L. & Skinner, M. K. Environmental epigenomics and disease susceptibility. *Nat. Rev. Genet.* **8,** 253–262 (2007).
4. Mazzio, E. A. & Soliman, K. F. A. Basic concepts of epigenetics: impact of environmental signals on gene expression. *Epigenetics* **7,** 119–30 (2012).
5. Siggens, L. & Ekwall, K. Epigenetics, chromatin and genome organization: recent advances from the ENCODE project. *J. Intern. Med.* **276,** 201–14 (2014).
6. Laker, R. C. & Ryall, J. G. DNA methylation in skeletal muscle stem cell specification , proliferation and differentiation. **2016,** (2016).
7. Waddington, C. H. *The Strategy of The Genes*. (George Allen & Unwin Ltd., 1957).
8. Bird, A. Perceptions of epigenetics. *Nature* **447,** 396–8 (2007).
9. Dupont, C., Armant, D. R. & Brenner, C. A. Epigenetics: Definition, Mechanisms and Clinical Perspective. *Semin Reprod Med* **27,** 351–357 (2009).
10. Feinberg, A. P. Phenotypic plasticity and the epigenetics of human disease. *Nature* **447,** 433–440 (2007).
11. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* **33 Suppl,** 245–54 (2003).
12. Nagy, C. & Turecki, G. Sensitive periods in epigenetics: bringing us closer to complex behavioral phenotypes. *Epigenomics* **4,** 445–457 (2012).
13. García-Carpizo, V., Ruiz-Llorente, L., Fraga, M. & Aranda, A. The growing role of gene methylation on endocrine function. *J. Mol. Endocrinol.* **47,** (2011).
14. Vinkers, C. H. *et al.* Traumatic stress and human DNA methylation : a critical review. *Epigenomics* **7,** 593–608 (2015).
15. Guerrero-Bosagna, C., Weeks, S. & Skinner, M. K. Identification of genomic features in environmentally induced epigenetic transgenerational inherited sperm epimutations. *PLoS One* **9,** (2014).
16. Gluckman, P. D., Hanson, M. A., Cooper, C. & Thornburg, K. L. Effect of In Utero and Early Life Conditions on Adult Health and Disease. *N. Engl. J. Med.* **359,** 61–73 (2008).

17. Gluckman, P. D., Hanson, M. A., Buklijas, T., Low, F. M. & Beedle, A. S. Epigenetic mechanisms that underpin metabolic and cardiovascular diseases. *Nat.Rev.Endocrinol.* **5,** 401–408 (2009).

18. Branco, M. R., Oda, M. & Reik, W. Safeguarding parental identity: Dnmt1 maintains imprints during epigenetic reprogramming in early embryogenesis. *Genes Dev.* **22,** 1567–1571 (2008).

19. Kanherkar, R. R., Bhatia-Dey, N. & Csoka, A. B. Epigenetics across the human lifespan. *Front. cell Dev. Biol.* **2,** 49 (2014).

20. Lillycrop, K. a & Burdge, G. C. The effect of nutrition during early life on the epigenetic regulation of transcription and implications for human diseases. *J. Nutrigenet. Nutrigenomics* **4,** 248–60 (2011).

21. Jablonka, E. & Lamb, M. J. The inheritance of acquired epigenetic variations. *Int. J. Epidemiol.* **44,** 1094–1103 (2015).

22. Roth, T. L. & Sweatt, J. D. Epigenetic mechanisms and evironmental shaping of the brain sensitive periods of development. *J. Child Psychol. Psychiatry* **52,** 398–408 (2011).

23. Lillycrop, K. A., Phillips, E. S., Jackson, A. A., Hanson, M. A. & Burdge, G. C. Dietary protein restriction of pregnant rats induces and folic acid supplementation prevents epigenetic modification of hepatic gene expression in the offspring. *J. Nutr.* **135,** 1382–1386 (2005).

24. Leenen, F. A. D., Muller, C. P. & Turner, J. D. DNA methylation: conducting the orchestra from exposure to phenotype? *Clin. Epigenetics* **8,** (2016).

25. Suderman, M. *et al.* Conserved epigenetic sensitivity to early life experience in the rat and human hippocampus. *Proc. Natl. Acad. Sci.* **109,** 17266–17272 (2012).

26. van der Knaap, L. J. *et al.* Glucocorticoid receptor gene (NR3C1) methylation following stressful events between birth and adolescence. The TRAILS study. *Transl. Psychiatry* **4,** e381 (2014).

27. Witzmann, S. R., Turner, J. D., Mériaux, S. B., Meijer, O. C. & Muller, C. P. Epigenetic regulation of the glucocorticoid receptor promoter 1(7) in adult rats. *Epigenetics* **7,** 1290–301 (2012).

28. Ormel, J. *et al.* The TRacking Adolescents' Individual Lives Survey (TRAILS): Design, current status, and selected findings. *J. Am. Acad. Child Adolesc. Psychiatry* **51,** 1020–1036 (2012).

29. van der Knaap, L. J. *et al.* Adverse Life Events and Allele-Specific Methylation of the Serotonin Transporter Gene (SLC6A4) in Adolescents: The TRAILS Study. *Psychosom. Med.* **77,** 246–55 (2015).

30. Gluckman, P. D., Hanson, M. A. & Buklijas, T. A conceptual framework for the developmental origins of health and disease. *J. Dev. Orig. Heal. Dis. J. Dev. Orig. Heal. Dis.* **1117325213,** 6–18 (2010).

31. Rakyan, V. K., Down, T. a, Balding, D. J. & Beck, S. Epigenome-wide association studies for common human diseases. *Nat. Rev. Genet.* **12,** 529–41 (2011).

32. Lillycrop, K. a. Effect of maternal diet on the epigenome: implications for human metabolic disease. *Proc.*

Nutr. Soc. **70,** 64–72 (2011).

33. Illingworth, R. *et al.* A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol.* **6,** 0037–0051 (2008).
34. Neidhart, M. *DNA Methylation and Complex Human Diseases*. (Academic Press, 2016).
35. Serre, D., Lee, B. H. & Ting, A. H. MBD-isolated genome sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. *Nucleic Acids Res.* **38,** 391–399 (2009).
36. Nestor, C. E. *et al.* Rapid reprogramming of epigenetic and transcriptional profiles in mammalian culture systems. *Genome Biol.* **16,** 11 (2015).
37. Wen, L. & Tang, F. Genomic distribution and possible functions of DNA hydroxymethylation in the brain. *Genomics* **104,** 341–346 (2014).
38. Li, S. *et al.* Hippocampal increase of 5-hmC in the glucocorticoid receptor gene following acute stress. *Behav. Brain Res.* **286,** 423 (2015).
39. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21,** 381–395 (2011).
40. Boutou, E. *et al.* in *Advances in DNA Repair* (ed. Chen, C. C.) 81–94 (InTech, 2015). doi:10.5772/61582
41. Lillycrop, K. A., Hoile, S. P., Grenfell, L. & Burdge, G. C. DNA methylation, ageing and the influence of early life nutrition. *Proc Nutr Soc* **73,** 413–421 (2014).
42. Mathews, H. L. & Janusek, L. W. Epigenetics and psychoneuroimmunology: Mechanisms and models. *Brain. Behav. Immun.* **25,** 25–39 (2011).
43. Chuang, J. C. & Jones, P. A. Epigenetics and microRNAs. *Pediatr. Res.* **61,** 24–29 (2007).
44. Godfrey, K. M., Lillycrop, K. A., Burdge, G. C., Gluckman, P. D. & Hanson, M. A. Epigenetic mechanisms and the mismatch concept of the developmental origins of health and disease. *Pediatr. Res.* **61,** 31–36 (2007).
45. Kim, V. N., Han, J. & Siomi, M. C. Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol* **10,** 126–139 (2009).
46. Weichenhan, D. & Plass, C. The evolving epigenome. *Hum. Mol. Genet.* **22,** R1-6 (2013).
47. Kim, D. H., Saetrom, P., Snøve, O. & Rossi, J. J. MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 16230–5 (2008).
48. Bayne, E. H. & Allshire, R. C. RNA-directed transcriptional gene silencing in mammals. *Trends Genet.* **21,** 370–373 (2005).
49. Carthew, R. W. & Sontheimer, E. J. Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136,** 642–

655 (2009).

50. Qureshi, I. A. & Mehler, M. F. Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease. *Nat. Rev. Neurosci.* **13,** 528–541 (2012).

51. Zuo, L., Wang, Z., Tan, Y., Chen, X. & Luo, X. piRNAs and Their Functions in the Brain. *Int J Hum Genet* **16,** 53–60 (2016).

52. Rajan, K. S. & Ramasamy, S. Retrotransposons and piRNA: The missing link in central nervous system. *Neurochem. Int.* **77,** 94–102 (2014).

53. Dharap, A., Nakka, V. P. & Vemuganti, R. Altered expression of PIWI RNA in the rat brain after transient focal ischemia. *Stroke* **42,** 1105–1109 (2011).

54. Lee, E. J. *et al.* Identification of piRNAs in the central nervous system Identification of piRNAs in the central nervous system. 1090–1099 (2011). doi:10.1261/rna.2565011.3

55. Meister, G. Argonaute proteins: functional insights and emerging roles. *Nat. Rev. Genet.* **14,** 447–459 (2013).

56. Rajasethupathy, P. *et al.* A role for neuronal piRNAs in the epigenetic control of memory- related synaptic plasticity. *Cell* **149,** 693–707 (2012).

57. Tsoi, L. C. *et al.* Analysis of long non-coding RNAs highlights tissue-specific expression patterns and epigenetic profiles in normal and psoriatic skin. *Genome Biol.* **16,** 24 (2015).

58. Mercer, T. R. & Mattick, J. S. Structure and function of long noncoding RNAs in epigenetic regulation. *Nat. Struct. & Mol. Biol.* **20,** 300–307 (2013).

59. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13,** 484–92 (2012).

60. Fraga, M. F. *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 10604–9 (2005).

61. Elliott, G. *et al.* Intermediate DNA methylation is a conserved signature of genome regulation. *Nat. Commun.* **6,** 6363 (2015).

62. Szyf, M. The early life social environmnent and DNA methylation - DNA methylation mediating the long-term impact of social environments early in life. *Epigenetics* **6,** 1–8 (2011).

63. Kaminsky, Z. A. *et al.* DNA methylation profiles in monozygotic and dizygotic twins. *Nat. Genet.* **41,** 240–245 (2009).

64. Bell, J. T. & Spector, T. D. A twin approach to unraveling epigenetics. *Trends Genet.* **27,** 116–125 (2011).

65. Castillo-Fernandez, J. E., Spector, T. D. & Bell, J. T. Epigenetics of discordant monozygotic twins: implications for disease. *Genome Med.* **6,** 60 (2014).

66. Li, S. *et al.* Genome-wide alterations in hippocampal 5-hydroxymethylcytosine links plasticity genes to acute stress. *Neurobiol. Dis.* **86,** 99–108 (2016).

67. Feng, J. *et al.* Dnmt1 and Dnmt3a are required for the maintenance of DNA methylation and synaptic function in adult forebrain neurons. *Nat. Neurosci.* **13,** 423–430 (2010).

68. Sweatt, J. D. Dynamic DNA methylation controls glutamate receptor trafficking and synaptic scaling. *J. Neurochem.* **137,** 312–330 (2016).

69. Wu, S. C. & Zhang, Y. Active DNA demethylation: many roads lead to Rome. *Nat. Rev. Mol. Cell Biol.* **11,** 607–20 (2010).

70. Wang, T. *et al.* Genome-wide DNA hydroxymethylation changes are associated with neurodevelopmental genes in the developing human cerebellum. *Hum. Mol. Genet.* **21,** 5500–5510 (2012).

71. De Smet, C., Lurquin, C., Lethé, B., Martelange, V. & Boon, T. DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter. *Mol. Cell. Biol.* **19,** 7327–35 (1999).

72. Mikeska, T. & Craig, J. M. DNA methylation biomarkers: Cancer and beyond. *Genes (Basel).* **5,** 821–864 (2014).

73. Levenson, V. V. DNA methylation as a universal biomarker. *Expert Rev. Mol. Diagn.* **10,** 481–488 (2010).

74. Roseboom, T. J., Painter, R. C., Van Abeelen, A. F. M., Veenendaal, M. V. E. & De Rooij, S. R. Hungry in the womb: What are the consequences? Lessons from the Dutch famine. *Maturitas* **70,** 141–145 (2011).

75. Guerrero-Bosagna, C., Settles, M., Lucker, B. & Skinner, M. K. Epigenetic transgenerational actions of vinclozolin on promoter regions of the sperm epigenome. *PLoS One* **5,** 1–17 (2010).

76. Yagi, S. *et al.* DNA methylation profile of tissue-dependent and differentially methylated regions (T-DMRs) in mouse promoter regions demonstrating tissue-specific gene expression. *Genome Res.* **18,** 1969–1978 (2008).

77. Liyanage, V. R. *et al.* DNA modifications: function and applications in normal and disease States. *Biol.* **3,** 670–723 (2014).

78. Carrió, E. & Suelves, M. DNA methylation dynamics in muscle development and disease. *Front. Aging Neurosci.* **7,** 1–12 (2015).

79. Levenson, V. V. & Melnikov, A. a. DNA Methylation as Clinically Useful Biomarkers—Light at the End of the Tunnel. *Pharmaceuticals* **5,** 94–113 (2012).

80. Zhang Y, Z. C. Role of DNA methylation in cardiovascular diseases. *Clin Exp Hypertens.* **30,** 1–7 (2016).

81. Moarii, M., Boeva, V., Vert, J.-P. & Reyal, F. Changes in correlation between promoter methylation and gene expression in cancer. *BMC Genomics* **16,** 873 (2015).

82. Ulahannan, N. & Greally, J. M. Genome-wide assays that identify and quantify modified cytosines in human disease studies. *Epigenetics Chromatin* **8,** 5 (2015).

83. Gluckman, P. D., Hanson, M. A. & Mitchell, M. D. Developmental origins of health and disease: reducing the burden of chronic disease in the next generation. *Genome Med.* **2,** 14 (2010).

84. Plagemann, A. *et al.* Hypothalamic proopiomelanocortin promoter methylation becomes altered by early overfeeding: an epigenetic model of obesity and the metabolic syndrome. *J. Physiol.* **587,** 4963–4976 (2009).

85. Dunn, G. A. & Bale, T. L. Maternal high-fat diet effects on third-generation female body size via the paternal lineage. *Endocrinology* **152,** 2228–2236 (2011).

86. Dunn, G. A. & Bale, T. L. Maternal high-fat diet promotes body length increases and insulin insensitivity in second-generation mice. *Endocrinology* **150,** 4999–5009 (2009).

87. Bygren, L. O. Intergenerational Health Responses to Adverse and Enriched Environments. *Annu Rev Public Heal.* **34,** 49–60 (2013).

88. Tobi, E. W. *et al.* DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. *Hum. Mol. Genet.* **18,** 4046–4053 (2009).

89. Heijmans, B. T. *et al.* Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc. Natl. Acad. Sci.* **105,** 17046–17049 (2008).

90. Ravelli, A. C. & Osmond, C. Obesity at the age of 50 years in men and women exposed to famine prenatally. *Am J Clin Nutr* **70,** 811–816 (1999).

91. Stein, A. D., Zybert, P. A., Van Der Pal-De Bruin, K. & Lumey, L. H. Exposure to famine during gestation, size at birth, and blood pressure at age 59 y: Evidence from the dutch famine. *Eur. J. Epidemiol.* **21,** 759–765 (2006).

92. De Rooij, S. R. *et al.* Impaired insulin secretion after prenatal exposure to the Dutch famine. *Diabetes Care* **29,** 1897–1901 (2006).

93. De Rooij, S. R. *et al.* Glucose tolerance at age 58 and the decline of glucose tolerance in comparison with age 50 in people prenatally exposed to the Dutch famine. *Diabetologia* **49,** 637–643 (2006).

94. Liu, Y. *et al.* Depression in pregnancy, infant birth weight and DNA methylation of imprint regulatory elements. *Epigenetics* **7,** 735–746 (2012).

95. Devlin, A. M., Brain, U., Austin, J. & Oberlander, T. F. Prenatal exposure to maternal depressed mood and the MTHFR C677T variant affect SLC6A4 methylation in infants at birth. *PLoS One* **5,** 2–9 (2010).

96. Kundakovic, M. *et al.* DNA methylation of BDNF as a biomarker of early-life adversity. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 6807–13 (2015).

97. Labonté, B. & Turecki, G. in *The Neurological Basis of Suicide* (CRC Press/Taylor & Francis, 2012). at <http://www.ncbi.nlm.nih.gov/books/NBK107193/>

98. Keller, S. *et al.* Increased BDNF promoter methylation in the Wernicke area of suicide subjects. *Arch. Gen. Psychiatry* **67,** 258–267 (2010).

99. Stenz, L. *et al.* BDNF promoter I methylation correlates between post-mortem human peripheral and brain tissues. *Neurosci. Res.* **91,** 1–7 (2015).

100. Roth, T. L., Lubin, F. D., Funk, A. J. & Sweatt, J. D. Lasting Epigenetic influence of early-life adversity on the BDNF gene. *Biol. Psychiatry* **65,** 760–769 (2009).

101. Pariante, C. M. & Lightman, S. L. The HPA axis in major depression: classical theories and new developments. *Trends Neurosci.* **31,** 464–468 (2008).

102. McGowan, P. O. *et al.* Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nat. Neurosci.* **12,** 342–348 (2009).

103. Tyrka, A. R., Price, L. H., Marsit, C., Walters, O. C. & Carpenter, L. L. Childhood adversity and epigenetic modulation of the leukocyte glucocorticoid receptor: preliminary findings in healthy adults. *PLoS One* **7,** e30148 (2012).

104. Weaver, I. C. G. *et al.* Epigenetic programming by maternal behavior. *Nat. Neurosci.* **7,** 847–54 (2004).

105. Alt, S. R. *et al.* Differential expression of glucocorticoid receptor transcripts in major depressive disorder is not epigenetically programmed. *Psychoneuroendocrinology* **35,** 544–56 (2010).

106. Elliott, E., Ezra-Nevo, G., Regev, L., Neufeld-Cohen, A. & Chen, A. Resilience to social stress coincides with functional DNA methylation of the Crf gene in adult mice. *Nat. Neurosci.* **13,** 1351–1353 (2010).

107. Murgatroyd, C. *et al.* Dynamic DNA methylation programs persistent adverse effects of early-life stress. *Nat Neurosci* **12,** 1559–1566 (2009).

108. Vijayendran, M., Beach, S., Plume, J. M., Brody, G. & Philibert, R. Effects of genotype and child abuse on DNA methylation and gene expression at the serotonin transporter. *Child Neurodev. Psychiatry* **3,** 55 (2012).

109. Beach, S. R. H., Brody, G. H., Todorov, A. A., Gunter, T. D. & Philibert, R. A. Methylation at SLC6A4 is linked to family history of child abuse: an examination of the Iowa Adoptee sample. *Am J Med Genet B Neuropsychiatr Genet* **153B,** 710–713 (2010).

110. Kim, D. K., Choi, S. H., Yu, J., Yoo, Y. & Koh, Y. Y. Bronchial responsiveness to methacholine and adenosine 5'- monophosphate in atopic and non-atopic preschool children with recurrent wheezing. *Clin.*

Exp. Allergy **37,** 15–21 (2007).

111. Blumenthal, M. N. The role of genetics in the development of asthma and atopy. *Curr. Opin. Allergy Clin. Immunol.* **5,** 141–5 (2005).

112. Kim, J. S. *et al.* Dissociation between the prevalence of atopy and allergic disease in rural China among children and adults. *J. Allergy Clin. Immunol.* **122,** 929–935 (2008).

113. Li, J. Y. *et al.* Association between DNA hypomethylation at IL13 gene and allergic rhinitis in house dust mite-sensitized subjects. *Clin. Exp. Allergy* **46,** 298–307 (2016).

114. Ho, S.-M. Environmental Epigenetic of Asthma – An update. **126,** 453–465 (2010).

115. Reinius, L. E. *et al.* DNA Methylation in the Neuropeptide S Receptor 1 (NPSR1) Promoter in Relation to Asthma and Environmental Factors. *PLoS One* **8,** (2013).

116. Morgan, H. D., Sutherland, H. G., Martin, D. I. & Whitelaw, E. Epigenetic inheritance at the agouti locus in the mouse. *Nat. Genet.* **23,** 314–318 (1999).

117. Birney, E., Smith, G. D. & Greally, J. M. Epigenome-wide Association Studies and the Interpretation of Disease -Omics. *PLoS Genet.* **12,** 1–9 (2016).

118. Hogg, K., Price, E. M. & Robinson, W. P. Improved reporting of DNA methylation data derived from studies of the human placenta. *Epigenetics* **9,** 333–337 (2014).

119. Michels, K. B. *et al.* Recommendations for the design and analysis of epigenome-wide association studies. *Nat. Methods* **10,** 949–55 (2013).

120. Zhang, B. B. *et al.* Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm. *Genome Res.* **23,** 1522–1540 (2013).

121. Labonté, B., Azoulay, N., Yerko, V., Turecki, G. & Brunet, A. Epigenetic modulation of glucocorticoid receptors in posttraumatic stress disorder. *Transl. Psychiatry* **4,** e368 (2014).

122. Labonte, B. *et al.* Differential glucocorticoid receptor exon 1(B), 1(C), and 1(H) expression and methylation in suicide completers with a history of childhood abuse. *Biol. Psychiatry* **72,** 41–48 (2012).

123. McGowan, P. O. *et al.* Broad epigenetic signature of maternal care in the brain of adult rats. *PLoS One* **6,** e14739 (2011).

124. Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13,** 86 (2012).

125. Houseman, E. A., Molitor, J. & Marsit, C. J. Reference-free cell mixture adjustments in analysis of DNA methylation data. *Bioinformatics* **30,** 1431–9 (2014).

126. Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11,** 587 (2010).

127. Cao-Lei, L. *et al.* Glucocorticoid receptor gene expression and promoter CpG modifications throughout the human brain. *J. Psychiatr. Res.* **47,** 1597–607 (2013).

128. Liu, L. *et al.* Insufficient DNA methylation affects healthy aging and promotes age-related health problems. *Clin. Epigenetics* **2,** 349–360 (2011).

129. Fuke, C. *et al.* Age related changes in 5-methylcytosine content in human peripheral leukocytes and placentas: An HPLC-based study. *Ann. Hum. Genet.* **68,** 196–204 (2004).

130. Shimabukuro, M. *et al.* Global hypomethylation of peripheral leukocyte DNA in male patients with schizophrenia: A potential link between epigenetics and schizophrenia. *J. Psychiatr. Res.* **41,** 1042–1046 (2007).

131. Sandovici, I. *et al.* Interindividual variability and parent of origin DNA methylation differences at specific human Alu elements. *Hum. Mol. Genet.* **14,** 2135–2143 (2005).

132. Sarter, B. *et al.* Sex differential in methylation patterns of selected genes in Singapore Chinese. *Hum. Genet.* **117,** 402–403 (2005).

133. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, Burton J, Cox TV, Davies R, Down TA, Haefliger C, Horton R, Howe K, Jackson DK, Kunde J, Koenig C, Liddle J, Niblett D, Otto T, Pettett R, Seemann S, Thompson C, West T, Rogers J, Olek A, Ber, B. S. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet.* **38,** 1378–85 (2006).

134. El-Maarri, O. *et al.* Gender specific differences in levels of DNA methylation at selected loci from human total blood: A tendency toward higher methylation levels in males. *Hum. Genet.* **122,** 505–514 (2007).

135. Bjornson, T. H. *et al.* Intra-individual Change Over Time in DNA. *Jama* **299,** 2877–2883 (2013).

136. Bock, C., Walter, J., Paulsen, M. & Lengauer, T. Inter-individual variation of DNA methylation and its implications for large-scale epigenome mapping. *Nucleic Acids Res.* **36,** (2008).

137. Gertz, J. *et al.* Analysis of dna methylation in a three-generation family reveals widespread genetic influence on epigenetic regulation. *PLoS Genet.* **7,** (2011).

138. Heijmans, B. T., Kremer, D., Tobi, E. W., Boomsma, D. I. & Slagboom, P. E. Heritable rather than age-related environmental and stochastic factors dominate variation in DNA methylation of the human IGF2/H19 locus. *Hum. Mol. Genet.* **16,** 547–554 (2007).

139. Xie, H. *et al.* Genome-wide quantitative assessment of variation in DNA methylation patterns. *Nucleic Acids Res.* **39,** 4099–4108 (2011).

140. Wagner, J. R. *et al.* The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. *Genome Biol.* **15,** R37 (2014).

141. Choufani, S. *et al.* A novel approach identifies new differentially methylated regions (DMRs) associated

with imprinted genes A novel approach identifies new differentially methylated regions (DMRs) associated with imprinted genes. 465–476 (2011). doi:10.1101/gr.111922.110

142. Kerkel, K. *et al.* Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat. Genet.* **40,** 904–908 (2008).

143. Schalkwyk, L. C. *et al.* Allelic Skewing of DNA Methylation Is Widespread across the Genome. *Am. J. Hum. Genet.* **86,** 196–212 (2010).

144. Tang, A. *et al.* Analysis of a four generation family reveals the widespread sequence-dependent maintenance of allelic DNA methylation in somatic and germ cells. *Sci. Rep.* **6,** 19260 (2016).

145. Choi, J. D. & Lee, J.-S. Interplay between Epigenetics and Genetics in Cancer. *Genomics Inf.* **11,** 164–173 (2013).

146. Huang, S. Genetic and non-genetic instability in tumor progression: Link between the fitness landscape and the epigenetic landscape of cancer cells. *Cancer Metastasis Rev.* **32,** 423–448 (2013).

147. Christensen, B. C. *et al.* Epigenetic profiles distinguish pleural mesothelioma from normal pleura and predict lung asbestos burden and clinical outcome. *Cancer Res.* **69,** 227–234 (2009).

148. Kohno, H. *et al.* Abberant promoter methylation of WIF-1 and SFRP1, 2, 4 genes in mesothelioma. *Oncol. Rep.* **24,** 423–431 (2010).

149. Robles, A. I. & Harris, C. C. Integration of multiple 'OMIC' biomarkers: A precision medicine strategy for lung cancer. *Lung Cancer* (2016). doi:10.1016/j.lungcan.2016.06.003

150. Guintivano, J., Arad, M., Gould, T. D., Payne, J. L. & Kaminsky, Z. A. Antenatal prediction of postpartum depression with blood DNA methylation biomarkers. *Mol. Psychiatry* **19,** 560–567 (2014).

151. Kimmel, M., Kaminsky, Z. & Payne, J. L. Biomarker or pathophysiology? The role of DNA methylation in postpartum depression. *Epigenomics* **5,** 473–475 (2013).

152. Li-Tempel, T. *et al.* The cardiovascular and hypothalamus-pituitary-adrenal axis response to stress is controlled by glucocorticoid receptor sequence variants and promoter methylation. *Clin. Epigenetics* **8,** 12 (2016).

153. Cao-Lei, L. *et al.* Transcriptional control of the human glucocorticoid receptor: identification and analysis of alternative promoter regions. *Hum. Genet.* **129,** 533–43 (2011).

154. Yehuda, R. *et al.* Lower Methylation of Glucocorticoid Receptor Gene Promoter 1F in Peripheral Blood of Veterans with Posttraumatic Stress Disorder. *Biol. Psychiatry* **77,** 356–364 (2015).

155. Breivik, T., Gundersen, Y., Murison, R., Turner, J. D. & Muller, C. P. Maternal Deprivation of Lewis Rat Pups Increases the Severity of Experi- mental Periodontitis in Adulthood. *Open Dent. J.* **9,** 65–78 (2015).

156. Ball, M. P. *et al.* Targeted and genome-scale strategies reveal gene-body methylation signatures in

human cells. *Nat. Biotechnol.* **27,** 361–8 (2009).

157. Bell, J. T. *et al.* DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol.* **12,** R10 (2011).

158. Liu, Y. *et al.* GeMes, clusters of DNA methylation under genetic control, can inform genetic and epigenetic analysis of disease. *Am. J. Hum. Genet.* **94,** 485–495 (2014).

159. Roth, T. L. Epigenetic mechanisms in the development of behavior: advances, challenges, and future promises of a new field. *Dev Psychopathol.* **25,** 1279–1291 (2013).

160. Breslin, M. B., Geng, C. D. & Vedeckis, W. V. Multiple promoters exist in the human GR gene, one of which is activated by glucocorticoids. *Mol. Endocrinol.* **15,** 1381–95 (2001).

161. Presul, E., Schmidt, S., Kofler, R. & Helmberg, A. Identification, tissue expression, and glucocorticoid responsiveness of alternative first exons of the human glucocorticoid receptor. *J. Mol. Endocrinol.* **38,** 79–90 (2007).

162. Turner, J. D. & Muller, C. P. Structure of the glucocorticoid receptor (NR3C1) gene 5' untranslated region: identification, and tissue distribution of multiple new human exon 1. *J. Mol. Endocrinol.* **35,** 283–92 (2005).

163. Turner, J. D. *et al.* Transcriptional control of the glucocorticoid receptor: CpG islands, epigenetics and more. *Biochem. Pharmacol.* **80,** 1860–1868 (2010).

164. Turner, J. D., Vernocchi, S., Schmitz, S. & Muller, C. P. Role of the 5'-untranslated regions in post-transcriptional regulation of the human Glucocorticoid Receptor. *Biochim. Biophys. acta - Gene Regul. Mech.* **1839,** 1051–1061 (2014).

165. Lu, N. Z. & Cidlowski, J. A. Glucocorticoid receptor isoforms generate transcription specificity. *Trends Cell Biol.* **16,** 301–307 (2006).

166. Leenen, F. A. D. *et al.* Where does transcription start? 5'-RACE adapted to next-generation sequencing. *Nucleic Acids Res.* **44,** 2628–2645 (2015).

167. Pal, S. *et al.* Alternative transcription exceeds alternative splicing in generating the transcriptome diversity of cerebellar development. *Genome Res.* **21,** 1260–72 (2011).

168. Shabalina, S. a., Spiridonov, A. N., Spiridonov, N. a. & Koonin, E. V. Connections between alternative transcription and alternative splicing in mammals. *Genome Biol. Evol.* **2,** 791–799 (2010).

169. Hughes, T. A. Regulation of gene expression by alternative untranslated regions. *Trends Genet.* **22,** 119–22 (2006).

170. Zhang, T., Haws, P. & Wu, Q. Multiple Variable First Exons : A Mechanism for Cell- and Tissue-Specific Gene Regulation. *Genome Res.* **14,** 79–89 (2004).

171. Lu, N. Z. & Cidlowski, J. A. Translational Regulatory Mechanisms Generate N-Terminal Glucocorticoid Receptor Isoforms with Unique Transcriptional Target Genes. *Mol. Cell* **18,** 331–342 (2005).

172. Carninci, P. *et al.* Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* **38,** 626–35 (2006).

173. Ayoubi, T. A. Y. & Van De Ven, W. J. M. Regulation of gene expression by alternative promoters. *FASEB J.* **10,** 453–60 (1996).

174. Brosius, J. & Gould, S. J. On 'genomenclature': a comprehensive (and respectful) taxonomy for pseudogenes and other 'junk DNA'. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 10706–10710 (1992).

175. Brosius, J. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. *Gene* **238,** 115–134 (1999).

176. Brosius, J. Many G-protein-coupled receptors are encoded by retrogenes. *Trends Genet.* **15,** 304–305 (1999).

177. Merkin, J. J. J., Chen, P., Alexis, M. S. S., Hautaniemi, S. K. K. & Burge, C. B. B. Origins and Impacts of New Mammalian Exons. *Cell Rep.* **10,** 1992–2005 (2015).

178. Shabalina, S. a., Ogurtsov, A. Y., Spiridonov, N. a. & Koonin, E. V. Evolution at protein ends: Major contribution of alternative transcription initiation and termination to the transcriptome and proteome diversity in mammals. *Nucleic Acids Res.* **42,** 7132–7144 (2014).

179. Schor, I. E., Gómez Acuña, L. I. & Kornblihtt, A. R. in *Bulletin du cancer* **95,** 33–41 (2013).

180. Kornblihtt, A. R. *et al.* Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.* **14,** 153–65 (2013).

181. Irimia, M., Rukov, J. L., Penny, D. & Roy, S. W. Functional and evolutionary analysis of alternatively spliced genes is consistent with an early eukaryotic origin of alternative splicing. *BMC Evol. Biol.* **7,** 188 (2007).

182. Pankratova, E. V. Alternative promoters in expression of genetic information. *Mol. Biol.* **42,** 371–380 (2008).

183. Resch, A. M., Ogurtsov, A. Y., Rogozin, I. B., Shabalina, S. A. & Koonin, E. V. Evolution of alternative and constitutive regions of mammalian 5'UTRs. *BMC Genomics* **10,** 162 (2009).

184. Su, A. I. *et al.* A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 6062–7 (2004).

185. Turner, J. D., Schote, A. B., Macedo, J. A., Pelascini, L. P. L. & Muller, C. P. Tissue specific glucocorticoid receptor expression, a role for alternative first exon usage? *Biochem. Pharmacol.* **72,** 1529–37 (2006).

186. Theisen, L. L. & Muller, C. P. EPs® 7630 (Umckaloabo®), an extract from Pelargonium sidoides roots, exerts anti-influenza virus activity in vitro and in vivo. *Antiviral Res.* **94,** 147–56 (2012).

187. Vernocchi, S. *et al.* Membrane glucocorticoid receptor activation induces proteomic changes aligning with classical glucocorticoid effects. *Mol. Cell. proteomics* **12,** 1764–1779 (2013).

188. Maruyama, K. & Sugano, S. Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. *Gene* **138,** 171–174 (1994).

189. Volloch, V., Schweitzer, B. & Rits, S. Ligation-mediated amplification of RNA from murine erythroid cells reveals a novel class of β globin mRNA with an extended 5'-untranslated region. *Nucleic Acids Res.* **22,** 2507–2511 (1994).

190. Schaefer, B. C. Revolutions in rapid Amplification of cDNA Ends: New Strategies for Polymerase Chain Reaction Cloning of Full-Length cDNA Ends. *Anal. Biochem.* **227,** 255–273 (1995).

191. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25,** 1105–11 (2009).

192. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7,** 562–78 (2012).

193. R Core Team. R: A language and environment for statistical computing. (2014). at <https://www.r-project.org/>

194. Tarzona, S., Garcia-Alcade, F., Dopazo, J., Ferrer, A. & Conesa, A. Differential expression in RNA-seq : A matter of depth. *Genome Biol.* **21,** 2213–2223 (2011).

195. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43,** 1–26 (2015).

196. MacDonald, J. W. affycoretools: Functions useful for those doing repetitive analyses with Affymetrix GeneChips. (2008).

197. Maechler, M., Rousseuw, P., Struyf, A., Hubert, M. & Hornik, K. cluster: Cluster Analysis Basics and Extensions. (2014).

198. Raivo, K. pheatmap: Pretty Heatmaps. (2013). at <http://cran.r-project.org/package=pheatmap>

199. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15,** 1034–50 (2005).

200. Turner, J. D., Pelascini, L. P. L., Macedo, J. a & Muller, C. P. Highly individual methylation patterns of alternative glucocorticoid receptor promoters suggest individualized epigenetic regulatory mechanisms. *Nucleic Acids Res.* **36,** 7207–18 (2008).

201. Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R. & Hofacker, I. L. The Vienna RNA websuite.

Nucleic Acids Res. **36,** 70–74 (2008).

202. Breslin, M. B. & Vedeckis, W. V. The human glucocorticoid receptor promoter upstream sequences contain binding sites for the ubiquitous transcription factor, Yin Yang 1. *J. Steroid Biochem. Mol. Biol.* **67,** 369–81 (1998).

203. Stolte, E. H., Verburg van Kemenade, B. M. L., Savelkoul, H. F. J. & Flik, G. Evolution of glucocorticoid receptors with different glucocorticoid sensitivity. *J. Endocrinol.* **190,** 17–28 (2006).

204. Araujo, P. R. *et al.* Before It Gets Started: Regulating Translation at the 5' UTR. *Comp. Funct. Genomics* **2012,** 475731 (2012).

205. Turner, J. D., Schote, A. B., Keipes, M. & Muller, C. P. A new transcript splice variant of the human glucocorticoid receptor: Identification and tissue distribution of hGRΔ313-338, an alternative exon 2 transactivation domain isoform. *Ann. N. Y. Acad. Sci.* **1095,** 334–341 (2007).

206. Roni, V., Carpio, R. & Wissinger, B. Mapping of transcription start sites of human retina expressed genes. *BMC Genomics* **8,** 42 (2007).

207. Suzuki, Y., Yoshitomo-nakagawa, K., Maruyama, K. & Suyama, A. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. *Gene* **200,** 149–156 (1997).

208. Coleman, J. E. Structure and mechanism of alkaline phosphatase. *Annu. Rev. Biophys. Biomol. Struct.* **21,** 441–83 (1992).

209. Rittié, L. & Perbal, B. Enzymes used in molecular biology: a useful guide. *J. Cell Commun. Signal.* **2,** 25–45 (2008).

210. Castora, F. J. in *Recombinant DNA Principles and Methodologies* (eds. Greene, J. J. & Rao, V. B.) 768 (CRC Press, 1998). at <http://www.google.lu/books?hl=en&lr=&id=YPGw2q935MIC&pgis=1>

211. Alon, S. *et al.* Barcoding bias in high-throughput multiplex sequencing of miRNA. *Genome Res.* **21,** 1506–11 (2011).

212. Song, Y., Liu, K. J. & Wang, T.-H. Elimination of ligation dependent artifacts in T4 RNA ligase to achieve high efficiency and low bias microRNA capture. *PLoS One* **9,** e94619 (2014).

213. Berry, D., Ben Mahfoudh, K., Wagner, M. & Loy, A. Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Appl. Environ. Microbiol.* **77,** 7846–9 (2011).

214. Jayaprakash, A. D., Jabado, O., Brown, B. D. & Sachidanandam, R. Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Res.* **39,** e141 (2011).

215. Van Nieuwerburgh, F. *et al.* Quantitative bias in Illumina TruSeq and a novel post amplification barcoding strategy for multiplexed DNA and small RNA deep sequencing. *PLoS One* **6,** e26969 (2011).

216. Hafner, M., Renwick, N., Brown, M. & Mihailović, A. RNA-ligase-dependent biases in miRNA

representation in deep-sequenced small RNA cDNA libraries. *Rna* 1697–1712 (2011). doi:10.1261/rna.2799511.Croce

217. Zhuang, F., Fuchs, R. T. & Robb, G. B. Small RNA expression profiling by high-throughput sequencing: implications of enzymatic manipulation. *J. Nucleic Acids* **2012,** 360358 (2012).

218. Kodzius, R. *et al.* CAGE: cap analysis of gene expression. *Nat. Methods* **3,** 211–22 (2006).

219. Carninci, P. *et al.* The transcriptional landscape of the mammalian genome. *Science* **309,** 1559–63 (2005).

220. Wamboldt, Y. *et al.* Participation of Leaky Ribosome Scanning in Protein Dual Targeting by Alternative Translation Initiation in Higher Plants. *Plant Cell Online* **21,** 157–167 (2009).

221. Tournillon, A.-S. *et al.* The alternative translated MDMX ^{p60} isoform regulates MDM2 activity. *Cell Cycle* **14,** 449–458 (2015).

222. Bohrer, A.-S. *et al.* Alternative translational initiation of ATP sulfurylase underlying dual localization of sulfate assimilation pathways in plastids and cytosol in Arabidopsis thaliana. *Front. Plant Sci.* **5,** 1–10 (2015).

223. Zhuo, R.-G. *et al.* The isoforms generated by alternative translation initiation adopt similar conformation in the selectivity filter in TREK-2. *J. Physiol. Biochem.* (2015). doi:10.1007/s13105-015-0422-z

224. Cagney, G., Amiri, S., Premawaradena, T., Lindo, M. & Emili, A. In silico proteome analysis to facilitate proteomics experiments using mass spectrometry. *Proteome Sci.* **1,** 5 (2003).

225. Ghesquiere, B. *et al.* Redox Proteomics of Protein-bound Methionine Oxidation. *Mol. Cell. Proteomics* **10,** M110.006866-M110.006866 (2011).

226. Wolf, Y. I., Novichkov, P. S., Karev, G. P., Koonin, E. V & Lipman, D. J. The universal distribution of evolutionary rates of genes and distinct characteristics of eukaryotic genes of different apparent ages. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 7273–7280 (2009).

227. Hopkins-Donaldson, S. *et al.* Silencing of death receptor and caspase-8 expression in small cell lung carcinoma cell lines and tumors by DNA methylation. *Cell Death Differ.* **10,** 356–64 (2003).

228. Zhang, Z., Chen, C. & Manev, H. DNA methylation as an epigenetic regulator of neural 5-lipoxygenase expression: evidence in human NT2 and NT2-N cells. *J. Neurochem.* **88,** 1424–1430 (2004).

229. Appleton, K. *et al.* Phase I and pharmacodynamic trial of the DNA methyltransferase inhibitor decitabine and carboplatin in solid tumors. *J. Clin. Oncol.* **25,** 4603–4609 (2007).

230. Moser, D. *et al.* The glucocorticoid receptor gene exon 1-F promoter is not methylated at the NGFI-A binding site in human hippocampus. *World J. Biol. Psychiatry* **8,** 262–8 (2007).

231. Landry, J.-R., Mager, D. L. & Wilhelm, B. T. Complex controls: the role of alternative promoters in

mammalian genomes. *Trends Genet.* **19,** 640–8 (2003).

232. Shabalina, S. a. *et al.* Distinct patterns of expression and evolution of intronless and intron-containing mammalian genes. *Mol. Biol. Evol.* **27,** 1745–1749 (2010).

233. Szyf, M. DNA methylation, the early-life social environment and behavioral disorders. *J. Neurodev. Disord.* **3,** 238–49 (2011).

234. Thomson, J. P. *et al.* Dynamic changes in 5-hydroxymethylation signatures underpin early and late events in drug exposed liver. *Nucleic Acids Res.* **41,** 5639–5654 (2013).

235. Tyrka, A. R. *et al.* Methylation of exons 1D, 1F, and 1H of the glucocorticoid receptor gene promoter and exposure to adversity in preschool-aged children. *Dev. Psychopathol.* **27,** 577–585 (2015).

236. Guo, J. U., Su, Y., Zhong, C., Ming, G. L. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145,** 423–434 (2011).

237. Langley-Evans, S. C. & McMullen, S. Developmental origins of adult disease. *Med. Princ. Pract.* **19,** 87–98 (2010).

238. Essex, M. J. *et al.* Epigenetic Vestiges of Early Developmental Adversity: Childhood Stress Exposure and DNA Methylation in Adolescence. *Child Dev.* **84,** 58–75 (2013).

239. Paternain, L. *et al.* Methyl donor supplementation in rats reverses the deleterious effect of maternal separation on depression-like behaviour. *Behav. Brain Res.* **299,** 51–58 (2016).

240. Kember, R. L. *et al.* Maternal separation is associated with strain-specific responses to stress and epigenetic alterations to Nr3c1, Avp, and Nr4a1 in mouse. *Brain Behav.* **2,** 455–467 (2012).

241. Palma-Gudiel, H., Córdova-Palomera, A., Leza, J. C. & Fañanás, L. Glucocorticoid receptor gene (NR3C1) methylation processes as mediators of early adversity in stress-related disorders causality: A critical review. *Neurosci. Biobehav. Rev.* **55,** 520–535 (2015).

242. Leenen, F. A. D. *et al.* Where does transcription start? 5′-RACE adapted to next-generation sequencing. *Nucleic Acids Res.* **44,** 2628–2645 (2016).

243. Klok, M. D. *et al.* Decreased expression of mineralocorticoid receptor mRNA and its splice variants in postmortem brain regions of patients with major depressive disorder. *J. Psychiatr.* **45,** 871–878 (2011).

244. Oberlander, T. F. *et al.* Prenatal exposure to maternal depression, neonatal methylation of human glucocorticoid receptor gene (NR3C1) and infant cortisol stress responses. *Epigenetics* **3,** 97–106 (2008).

245. Lariviere, W. R. Inflammation-Susceptible Lewis Rats Show Less Sensitivity Than Resistant Fischer Rats in the Formalin Inflammatory Pain Test and With Repeated Thermal Testing. *J. Neurophysiol.* **95,** 2889–2897 (2006).

246. Halder, R. *et al.* DNA methylation changes in plasticity genes accompany the formation and maintenance of memory. *Nat Neurosci* **19,** 102–110 (2016).

247. Peng, X. *et al.* Statistical implications of pooling RNA samples for microarray experiments. *BMC Bioinformatics* **4,** 26 (2003).

248. Kendziorski, C., Irizarry, R. A., Chen, K.-S. S., Haag, J. D. & Gould, M. N. On the utility of pooling biological samples in microarray experiments. *Proc Natl Acad Sci U S A* **102,** 4252–4257 (2005).

249. Auer, P. L. & Doerge, R. W. Statistical design and analysis of RNA sequencing data. *Genetics* **185,** 405–16 (2010).

250. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10–12 (2011).

251. Tange, O. GNU Parallel - The Command-Line Power Tool. *;login USENIX Mag.* **36,** 42–47 (2011).

252. Andrews, S. FastQC: a quality control tool for high throughput sequence data. (2010). at <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>

253. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–360 (2012).

254. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

255. Kirschner, S. A. *et al.* Focussing reduced representation CpG sequencing through judicious restriction enzyme choice. *Genomics* **107,** 109–119 (2016).

256. Lienhard, M., Grimm, C., Morkel, M., Herwig, R. & Chavez, L. MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. *Bioinformatics* **30,** 284–6 (2014).

257. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37,** 1–13 (2009).

258. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4,** 44–57 (2009).

259. Su, Y. *et al.* Genome-wide DNA methylation profile of developing deciduous tooth germ in miniature pigs. *BMC Genomics* **17,** 134 (2016).

260. Habib, M., Fares, F., Bourgeois, C. A., Bella, C. & Bernardino, J. DNA Global Hypomethylation in EBV-Transformed Interphase Nuclei. *Exp. Cell Res.* **249,** 46–53 (1999).

261. Li, W. & Liu, M. Distribution of 5-Hydroxymethylcytosine in Different Human Tissues. *J. Nucleic Acids* **2011,** 1–5 (2011).

262. Taiwo, O. *et al.* Methylome analysis using MeDIP-seq with low DNA concentrations. *Nat. Protoc.* **7,** 617–

36 (2012).

263. Li, D., Zhang, B., Xing, X. & Wang, T. Combining MeDIP-seq and MRE-seq to investigate genome-wide CpG methylation. *Methods* (2014). doi:10.1016/j.ymeth.2014.10.032

264. Zhao, M.-T., Whyte, J. J., Hopkins, G. M., Kirk, M. D. & Prather, R. S. Methylated DNA immunoprecipitation and high-throughput sequencing (MeDIP-seq) using low amounts of genomic DNA. *Cell. Reprogram.* **16,** 175–84 (2014).

265. Walker, D. L. *et al.* DNA methylation profiling: comparison of genome-wide sequencing methods and the Infinium Human Methylation 450 Bead Chip. *Epigenomics* 1–16 (2015). doi:10.2217/EPI.15.64

266. Staunstrup, N. H. *et al.* Genome-wide DNA methylation profiling with MeDIP-seq using archived dried blood spots. *Clin. Epigenetics* **8,** 81 (2016).

267. Shen, L. *et al.* Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* **153,** 692–706 (2013).

268. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **7414,** 57–74 (2012).

269. van der Wijst, M. G. P. & Rots, M. G. Mitochondrial epigenetics: An overlooked layer of regulation? *Trends Genet.* **31,** 353–356 (2015).

270. Byun, H.-M. *et al.* Effects of Air Pollution and Blood Mitochondrial DNA Methylation on Markers of Heart Rate Variability. *J. Am. Heart Assoc.* **5,** e003218 (2016).

271. Iacobazzi, V., Castegna, A., Infantino, V. & Andria, G. Mitochondrial DNA methylation as a next-generation biomarker and diagnostic tool. *Mol. Genet. Metab.* **110,** 25–34 (2013).

272. Hunt, R. L., Ladd, C. O. & Plotsky, P. M. in *Encyclopedia of Stress. Volume 2* (ed. Fink, G.) 699–707 (Academic Press, 2000).

273. King, J. A. & Edwards, E. Early Stress and Genetic Influences on Hypothalamic – Pituitary – Adrenal Axis Functioning in Adulthood. **85,** 79–85 (1999).

274. Lehmann, J., Stöhr, T. & Feldon, J. Long-term effects of prenatal stress experience and postnatal maternal separation on emotionality and attentional processes. *Behav. Brain Res.* **107,** 133–144 (2000).

275. Plotsky, P. M. & Meaney, M. J. Early, postnatal experience alters hypothalamic corticotropin-releasing factor (CRF) mRNA, median eminence CRF content and stress-induced release in adult rats. *Mol. Brain Res.* **18,** 195–200 (1993).

276. Amkraut, A. A., Solomon, G. F. & Kraemer, H. C. Stress, early experience and adjuvant-induced arthritis in the rat. *Psychosom. Med.* **33,** 203–14 (1971).

277. Ellenbroek, B. a & Cools, a R. The long-term effects of maternal deprivation depend on the genetic

background. *Neuropsychopharmacology* **23,** 99–106 (2000).

278. Breivik, T., Opstad, P. K., Gjermo, P. & Thrane, P. S. Effects of hypothalamic-pituitary-adrenal axis reactivity on periodontal tissue destruction in rats. *Eur. J. Oral Sci.* **108,** 115–22 (2000).

279. Breivik, T., Thrane, P. S., Gjermo, P. & Opstad, P. K. Glucocorticoid receptor antagonist RU 486 treatment reduces periodontitis in Fischer 344 rats. *J. Periodontal Res.* **35,** 285–90 (2000).

280. Breivik, T., Gundersen, Y., Opstad, P. K. & Fonnum, F. Chemical sympathectomy inhibits periodontal disease in Fischer 344 rats. *J. Periodontal Res.* **40,** 325–330 (2005).

281. Breivik, T., Gundersen, Y., Osmundsen, H., Opstad, P. K. & Fonnum, F. Chronic treatment with the glutamate receptor antagonist MK-801 alters periodontal disease susceptibility. *J. Periodontal Res.* **40,** 28–35 (2005).

282. Gluckman, P. D., Hanson, M. A. & Buklijas, T. A conceptual framework for the developmental origins of health and disease. *J Dev Orig Heal. Dis* **1,** 6–18 (2010).

283. Osrin, D., Vergnano, S. & Costello, A. Serious bacterial infections in newborn infants in developing countries. *Curr Opin Infect Dis* **17,** 217–224 (2004).

284. Hall, C. B. *et al.* The burden of respiratory syncytial virus infection in young children. *N Engl J Med* **360,** 588–598 (2009).

285. Zhang, T. *et al.* Influenza-associated outpatient visits among children less than 5 years of age in eastern China, 2011-2014. *BMC Infect Dis* **16,** 267 (2016).

286. Oren, E. *et al.* Cough during infancy and subsequent childhood asthma. *Clin Exp Allergy* **45,** 1439–1446 (2015).

287. Jurgens, H. A., Amancherla, K. & Johnson, R. W. Influenza infection induces neuroinflammation, alters hippocampal neuron morphology, and impairs cognition in adult mice. *J Neurosci* **32,** 3958–3968 (2012).

288. Meyer, U. Prenatal poly(i:C) exposure and other developmental immune activation models in rodent systems. *Biol Psychiatry* **75,** 307–315 (2014).

289. Carraro, S., Scheltema, N., Bont, L. & Baraldi, E. Early-life origins of chronic respiratory diseases: understanding and promoting healthy ageing. *Eur Respir J* **44,** 1682–1696 (2014).

290. Marodi, L. Neonatal innate immunity to infectious agents. *Infect Immun* **74,** 1999–2006 (2006).

291. Netea, M. G., Latz, E., Mills, K. H. & O'Neill, L. A. Innate immune memory: a paradigm shift in understanding host defense. *Nat Immunol* **16,** 675–679 (2015).

292. Tang, B. *et al.* Interleukin-6 expression was regulated by epigenetic mechanisms in response to influenza virus infection or dsRNA treatment. *Mol Immunol* **48,** 1001–1008 (2011).

293. Zhang, Q. *et al.* Tet2 is required to resolve inflammation by recruiting Hdac2 to specifically repress IL-6.

Nature **525,** 389–393 (2015).

294. Koletzko, B. *et al.* The power of programming and the earlynutrition project: Opportunities for health promotion by nutrition during the first thousand days of life and beyond. *Ann. Nutr. Metab.* **64,** 187–196 (2014).

295. Theisen, L. L. *et al.* Tannins from Hamamelis virginiana bark extract: characterization and improvement of the antiviral efficacy against influenza A virus and human papillomavirus. *PLoS One* **9,** e88062 (2014).

296. Brunner, A. L. *et al.* Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver. *Genome Res* **19,** 1044–1056 (2009).

297. Suzuki, M. *et al.* Optimized design and data analysis of tag-based cytosine methylation assays. *Genome Biol.* **11,** R36 (2010).

298. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25,** 402–8 (2001).

299. Li, L.-C. & Dahiya, R. MethPrimer: designing primers for methylation PCRs. *Bioinformatics* **18,** 1427–1431 (2002).

300. Ellis, S., Mouihate, A. & Pittman, Q. J. Early life immune challenge alters innate immune responses to lipopolysaccharide: implications for host defense as adults. *Faseb J* **19,** 1519–1521 (2005).

301. Saia, R. S., Bertozi, G., Cunha, F. Q. & Carnio, E. C. Estradiol and thermoregulation in adult endotoxemic rats exposed to lipopolysaccharide in neonatal life. *Acta Physiol* **203,** 429–439 (2011).

302. Schmolke, M. & Garcia-Sastre, A. Evasion of innate and adaptive immune responses by influenza A virus. *Cell Microbiol* **12,** 873–880 (2010).

303. Epstein, S. L., Lo, C. Y., Misplon, J. A. & Bennink, J. R. Mechanism of protective immunity against influenza virus infection in mice without antibodies. *J Immunol* **160,** 322–327 (1998).

304. Graham, M. B. & Braciale, T. J. Resistance to and recovery from lethal influenza virus infection in B lymphocyte-deficient mice. *J Exp Med* **186,** 2063–2068 (1997).

305. Dienz, O. *et al.* Essential role of IL-6 in protection against H1N1 influenza virus by promoting neutrophil survival in the lung. *Mucosal Immunol* **5,** 258–266 (2012).

306. Matsukura, S. *et al.* Expression of RANTES by normal airway epithelial cells after influenza virus A infection. *Am J Respir Cell Mol Biol* **18,** 255–264 (1998).

307. Adamko, D. J., Yost, B. L., Gleich, G. J., Fryer, A. D. & Jacoby, D. B. Ovalbumin sensitization changes the inflammatory response to subsequent parainfluenza infection. Eosinophils mediate airway hyperresponsiveness, m(2) muscarinic receptor dysfunction, and antiviral effects. *J Exp Med* **190,** 1465–1478 (1999).

308. Le Goffic, R. *et al.* Cutting Edge: Influenza A virus activates TLR3-dependent inflammatory and RIG-I-dependent antiviral responses in human lung epithelial cells. *J Immunol* **178,** 3368–3372 (2007).

309. Iwasaki, A. & Pillai, P. S. Innate immunity to influenza virus infection. *Nat Rev Immunol* **14,** 315–328 (2014).

310. Landreth, K. S. Critical windows in development of the rodent immune system. *Hum Exp Toxicol* **21,** 493–498 (2002).

311. Saeed, S. *et al.* Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity. *Science (80-.).* **345,** 1251086 (2014).

312. Fang, J. *et al.* Epigenetic Changes Mediated by miR29 Activate Cyclooxygenase-2 and Interferon-λ1 Production during Viral Infection. *J. Virol.* **86,** 1010–1020 (2012).

313. Elwenspoek, M. M. C., Hengesch, X., Schächinger, H., Muller, C. P. & Turner, J. D. Increased methylation of the GR 1F promoter in post-institutionalized adults. *Psychoneuroendocrinology* **61,** 64 (2015).

314. Barrett, L. W., Fletcher, S. & Wilton, S. D. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell. Mol. life Sci.* **69,** 3613–3634 (2012).

315. Levine, M., Tjian, R. & Tijan, R. Transcription regulation and animal diversity. *Nature* **424,** 147–151 (2003).

316. Matteau, D. & Rodrigue, S. in *DNA-Protein Interactions: Principles and Protocols* (eds. Leblanc, B. P. & Rodrigue, S.) **1334,** 143–159 (Springer New York, 2015).

317. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489,** 101–8 (2012).

318. Wain, H. M. *et al.* Guidelines for Human Gene Nomenclature. *Genomics* **79,** 464–470 (2002).

319. Gerstein, M. B. *et al.* What is a gene , post-ENCODE? History and updated definition. *Genome Res.* **17,** 669–681 (2007).

320. Laubichler, M. D., Stadler, P. F., Prohaska, S. J. & Nowick, K. The relativity of biological function. *Theory Biosci.* **134,** 143–147 (2015).

321. Shafik, A., Schumann, U., Evers, M., Sibbritt, T. & Preiss, T. The emerging epitranscriptomics of long noncoding RNAs. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1859,** 59–70 (2016).

322. de Andres-Pablo, A., Morillon, A. & Wery, M. LncRNAs, lost in translation or licence to regulate? *Curr. Genet.* 1–5 (2016). doi:10.1007/s00294-016-0615-1

323. Berretta, J. & Morillon, A. Pervasive transcription constitutes a new level of eukaryotic genome regulation. *EMBO Rep.* **10,** 973–82 (2009).

324. McEwen, B. S., Eiland, L., Hunter, R. G. & Miller, M. M. Stress and Anxiety: Structural plasticity and

epigenetic regulation as a consequence of stress. *Neuropharmacology* **62,** 3–12 (2012).

325. De La Fuente, M. *et al.* Early maternal deprivation in rats: A proposed animal model for the study of developmental neuroimmunoendocrine interactions. *Ann. N. Y. Acad. Sci.* **1153,** 176–183 (2009).

326. Huot, R. L., Gonzalez, M. E., Ladd, C. O., Thrivikraman, K. V. & Plotsky, P. M. Foster litters prevent hypothalamic-pituitary-adrenal axis sensitization mediated by neonatal maternal separation. *Psychoneuroendocrinology* **29,** 279–289 (2004).

327. Barreau, F., Ferrier, L., Fioramonti, J. & Bueno, L. Neonatal maternal deprivation triggers long term alterations in colonic epithelial barrier and mucosal immunity in rats. *Gut* **53,** 501–506 (2004).

Presentations and Meeting participations

2012:

- Participation at “Antigen Processing and Presentation in Health and Disease” meeting, Luxembourg, LU, 2012.11.15-16
- Participation at Sar-Lor-Lux meeting of Virology, Nancy, FR, 2012.11.28

2013:

- Participation at doctoral symposium ‘Genetics meets Psychology’, Trier, GE, 2013.10.31

2014:

- Participation at the UniGR-Workshop Systems Biology, Epigenetics & Systems Analysis, Saarbrücken, GE, 2014.01.22
- Participation at 4th Clinical Epigenetics Society International Meeting, Dusseldorf, GE, 2014.03.6-7 and poster presentation entitled: “Epigenetics on the Ion Torrent: MEDIP-Seq and LUMA-Seq”
- Participation at COST meeting "NGS Data after the Gold Rush", Norwich, UK, 2014.05.05-08 and oral presentation entitled: ”Transcription microvariability detection by NGS”
- Participation at Doctoral Colloquium Trier, Trier, GE, 2014.07.14
- Participation at Sar-Lor-Lux meeting of Virology, Nancy, FR, 2014.09.11 and oral presentation entitled: “Transcription microvariability detection by NGS”
- Participation at Life Science PhD days, Luxembourg, LU, 2014.09.15 and oral presentation entitled: “Transcriptional Microvariability within the Glucocorticoid Receptor”
- Participation at TRIERER STRESSFORSCHUNG QUO VADIS, Trier, GE, 2014.12.08

2015:

- Participation at TRIERER STRESSFORSCHUNG QUO VADIS - part 2, Trier, GE, 2015.03.17
- Participation at Keystone Symposia: DNA methylation and Epigenetics, Keystone, Colorado, USA, 2015.03.29-04.03 and poster presentation: “Early Life- and Adolescent- Stress and the Methylome”
- Participation at Treffen des Forschungsschwerpunkts "Psychobiologie des Stresses", Trier, GE, 2015.06.08
- Participation at Life Science PhD days, Luxembourg, LU, 2015.09.21 and poster presentation entitled: “Early Life and Adolescent Stress and the Methylome”

- Participation at 7th Next Generation Sequencing Congress, 3rd Single Cell Analysis Congress and Genome Editing Congress, London, UK, 2015.11.12-13 and poster presentation entitled: “Transcriptional Microvariability within the Glucocorticoid Receptor”

2016:

- Participation at 6th Clinical Epigenetics Society International Meeting, Dusseldorf, GE, 2016.03.03-04
- Participation at Symposium “Digitalisierung der Medizin Bayern – Fokus klinische Patientdaten”, Munich. GE, 2016.06.30

Publications

Fleur A. D. Leenen, Sara Vernocchi, Oliver E. Hunewald, Stephanie Schmitz, Anne M. Molitor, Claude P. Muller, Jonathan D. Turner (2016) “Where does transcription start? 5′-RACE adapted to next-generation sequencing”

Nucl. Acids Res.44(6): 2628-2645

Fleur A. D. Leenen, CP Muller, JD Turner (2016) “DNA methylation: conducting the orchestra from exposure to phenotype?”

Clin Epigenetics. 2016; 8(1): 92

Sophie A. Kirschner, Liz Bohnenberger, **Fleur A. D. Leenen**, Oliver E. Hunewald, Linda D. Theisen, Sophie B. Mériaux, Claude P. Muller and Jonathan D. Turner (2017) “Epigenetic consequences of early life H1N1 infection”

(Submitted to Epigenomics)

Fleur A. D. Leenen, Sophie B. Mériaux, Oliver E. Hunewald, Torbjørn Breivik, Claude P. Muller, Robert Murison, Jonathan D. Turner (2017) “Genome-wide 5mC and 5hmC DNA profiling of maternally deprived rats”

(Manuscript in Preparation)

Erklärung

Hiermit versichere ich, dass ich die vorliegende Dissertationsschrift selbständig verfasst un keine anderen als die angegebenen Hilfsquellen verwendet habe. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Luxembourg, February 2017

_____

Fleur A. D. Leenen