From Department of Medicine Solna
Translational Immunology Unit
Karolinska Institutet, Stockholm, Sweden

Epigenetic Mechanisms of Asthma and Allergy

Nathalie Acevedo

Karolinska Institutet

Stockholm 2015

Published by Karolinska Institutet
Printed by E-print

ISBN 978-91-7549-830-0

Epigenetic Mechanisms of Asthma and Allergy

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Nathalie Acevedo Caballero

Principal Supervisor:
Professor Annika Scheynius
Karolinska Institutet
Department of Medicine Solna
Division of Translational Immunology

Co-supervisor(s):
Professor Juha Kere
Karolinska Institutet
Department of Biosciences and Nutrition
Division of Molecular Genetics

Associate Professor Cilla Söderhäll
Karolinska Institutet
Department of Biosciences and Nutrition
Division of Molecular Genetics

Opponent:
Professor John Holloway
University of Southampton, UK
Faculty of Medicine

Examination Board:
Associate Professor Maja Jagodic
Karolinska Institutet
Department of Clinical Neuroscience

Professor Claes Wadelius
Uppsala University
Department of Immunology,
Genetics and Pathology

Associate Professor Lennart Nilsson
Linköping University
Department of Clinical and
Experimental Medicine

A Luis Caraballo, mentor y amigo cuyo ejemplo y enseñanzas
fueron esenciales para iniciar mis estudios de doctorado

"Whether men will be able to survive the changes of environment that their own skill has brought about is an open question. If the answer is the affirmative, it will be known some day; if not, not" - B. Russell (1955)

ABSTRACT

Allergic diseases such as atopic eczema (AE) and asthma are chronic inflammatory disorders in which the regulation of the immune system is altered, in concert with disease-specific defects in target organs. Their inception requires gene-environment interactions but the molecular mechanisms contributing to the pathogenesis remain unclear. During the last decades there was a dramatic increase in the prevalence of allergic diseases worldwide and several environmental conditions have been implicated (*e.g.* pollution, changes in lifestyle, reduced biodiversity). These exposures are supposed to affect leukocyte function, possibly through epigenetic mechanisms, which are the ultimate connection between the cell and the environment. The overall aim of this thesis was to elucidate the DNA methylation signatures as an epigenetic marker associated to cell lineage **(I)** and increasing age **(IV)** in primary blood leukocytes, and to compare DNA methylation **(II, III)** and miRNA expression **(II)** between healthy controls (HC) and patients with AE and asthma in connection with changes in mRNA expression.

In **study I**, we found a particular DNA methylation landscape for each population of blood leukocytes that discriminates their lineage and that lymphoid cells are more methylated than myeloid cells. At the DNA methylation level, $CD19^+$ B cells are the most different among all leukocytes. This study revealed that differences in cell composition may bias DNA methylation analyses conducted in DNA extracted from whole blood and provided a catalog of lineage-specific CpG sites. In **study II**, we compared DNA methylation and miRNA levels between AE patients and HC in four T cell populations isolated from peripheral blood. We discovered that circulating skin-homing memory $CD4^+CLA^+$ T cells isolated from AE patients have significant DNA methylation differences in 49 CpG sites and differential expression of 16 miRNAs compared to HC. Integrative data analyses with the differentially expressed mRNAs revealed that epigenetic signatures in AE patients are connected to the IL-10/IL-13 pathway. In **study III**, we analyzed the relation between genotype and epigenotype in the *GSDMB/ORMDL3* asthma-susceptibility locus. We found significant differences in DNA methylation between children with controlled persistent asthma and HC. Methylation levels in three of these differentially methylated CpG sites correlate with the mRNA levels of *ORMDL3*. Causal inference tests revealed that both, single nucleotide polymorphisms and non-polymorphic CpG sites have independent effects on *ORMDL3* mRNA expression. We also discovered that $CD8^+$ T cells contain a differentially methylated region in the 5'UTR of *ORMDL3* that is less methylated compared to other leukocyte subtypes, suggesting that methylation differences in *ORMDL3* might promote lymphocyte driven inflammation. In **study IV**, we performed a pilot genome-wide DNA methylation analysis in whole blood, aiming to elucidate which genes had longitudinal changes in DNA methylation between 3 and 60 months after birth in healthy children. After filtering by lineage-CpG sites, we found 330 age-methylated CpG sites and 464 age-demethylated CpG sites that reflect the immunological window and are located on genes encoding for cytokines and chromatin remodelers.

In summary, we discovered DNA methylation signatures in blood leukocytes associated to terminal differentiation and development, and demonstrated that compared to HC, patients suffering of AE and asthma have differences in DNA methylation and miRNA levels that might have functional impact on mRNA expression. The findings of this thesis provide new insights on genomic loci containing epigenetic differences between allergic patients and HC, that may be studied in the future as candidates for the development of preventive strategies, disease biomarkers or novel pharmacological therapies.

LIST OF SCIENTIFIC PAPERS

I. Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlén SE, Greco D, Söderhäll C, Scheynius A, Kere J. **Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility**. PLoS One. 2012;7(7):e41361.

II. Acevedo N, Bruhn S, Andersson A, Katayama S, Krjutshkov K, Wickberg G, Lundeberg L, Söderhäll C, Kere J, Greco D, Scheynius A. **Differential DNA methylation and miRNA expression in skin-homing $CD4^{+}CLA^{+}$ T cells of atopic eczema patients**. Manuscript.

III. Acevedo N, Reinius LE, Greco D, Gref A, Orsmark-Pietras C, Persson H, Pershagen G, Hedlin G, Melén E, Scheynius A, Kere J, Söderhäll C. **Risk of childhood asthma is associated with CpG site polymorphisms, regional DNA methylation and mRNA levels at the *GSDMB/ORMDL3* locus.** Human Molecular Genetics. 2015; 24(3):875-890.

IV. Acevedo N, Reinius LE, Vitezic M, Fortino V, Söderhäll C, Honkanen H, Veijola R, Simell O, Toppari J, Ilonen J, Knip M, Scheynius A, Hyöty H, Greco D, Kere J. **Age-associated DNA methylation changes in immune genes, histone modifiers and chromatin remodeling factors within 5 years after birth in human blood leukocytes.** Under revision in Clinical Epigenetics.

Related publications not included in this thesis

Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, Reinius L, Acevedo N, Taub M, Ronninger M, Shchetynsky K, Scheynius A, Kere J, Alfredsson L, Klareskog L, Ekström TJ, Feinberg AP. **Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis.** Nat Biotechnol. 2013;31(2):142-7.

Reinius LE, Gref A, Sääf A, Acevedo N, Joerink M, Kupczyk M, D'Amato M, Bergström A, Melén E, Scheynius A, Dahlén SE; BIOAIR Study Group, Pershagen G, Söderhäll C, Kere J. **DNA methylation in the Neuropeptide S Receptor 1 (*NPSR1*) promoter in relation to asthma and environmental factors.** PLoS One. 2013;8(1):e53877.

Acevedo N, Sääf A, Söderhäll C, Melén E, Mandelin J, Pietras CO, Ezer S, Karisola P, Vendelin J, Gennäs GB, Yli-Kauhaluoma J, Alenius H, von Mutius E, Doekes G, Braun-Fahrländer C, Riedler J, van Hage M, D'Amato M, Scheynius A, Pershagen G, Kere J, Pulkkinen V. **Interaction between retinoid acid receptor-related orphan receptor alpha (*RORA*) and neuropeptide S receptor 1 (*NPSR1*) in asthma.** PLoS One. 2013;8(4):e60111.

Bornelöv S, Sääf A, Melén E, Bergström A, Torabi Moghadam B, Pulkkinen V, Acevedo N, Orsmark Pietras C, Ege M, Braun-Fahrländer C, Riedler J, Doekes G, Kabesch M, van Hage M, Kere J, Scheynius A, Söderhäll C, Pershagen G, Komorowski J. **Rule-based models of the interplay between genetic and environmental factors in childhood allergy.** PLoS One. 2013;8(11):e80080.

Fuertes E, Söderhäll C, Acevedo N, Beckere A, Brauer M, Chan-Yeung M, Dijk N, Heinrich J, de Jongste J, Koppelman G, Postma D, Kere J, Kozyrskyj A, Pershagen G, Sandford A, Standl M, Tiesler C, Waldenberger M, Westman M, Carlsten C, Melén E. **Associations between the 17q21 region and allergic rhinitis in five birth cohorts**. J Allergy Clin Immunol. Epub pii: S0091-6749(14)01188-9.

CONTENTS

LIST OF ABBREVIATIONS

A	adenine
AE	atopic eczema
ALADDIN	Assessment of Lifestyle and Allergic Disease During INfancy
BAMSE	Swedish abbreviation for Children, Allergy, Milieu, Stockholm, Epidemiology
BAL	bronchoalveolar lavage
BH	Benjamini-Hochberg (statistical correction)
BHR	bronchial hyperresponsiveness
bp	base pairs
C	cytosine
CA	controlled persistent asthma
CLA	cutaneous lymphocyte antigen
CGI	CpG island
DHS	DNAse hypersensitivity sites
DIPP	DIabetes Prediction and Prevention study
DNA	deoxyribonucleic acid
DNMTs	DNA methyltransferases
G	guanine
GO	gene ontology
GWAS	genome wide association study
HC	healthy controls
ICS	inhaled corticosteroids
IL	interleukin
LD	linkage disequilibrium
LINE	long interspersed nucleotide elements
MBPs	methyl binding proteins
methQTL	methylation quantitative trait loci
miRNA	microRNA
mRNA	messenger RNA
NK	natural killer
nt	nucleotides
ORMDL3	ORMDL sphingolipid biosynthesis regulator 3
PBMC	peripheral blood mononuclear cells
PcG	polycomb group
qPCR	quantitative PCR
RNA	ribonucleic acid
SAM	S-adenosyl methionine
SINE	short interspersed nucleotide elements
SLE	systemic lupus erythematosus
SNP	single nucleotide polymorphism
T	thymidine
TF	transcription factor

TSS	transcription start site
UTR	untranslated region

1 Introduction

The deoxyribonucleic acid (DNA) contains the genetic information encoded by the sequential order of four nucleotides: adenine (A), guanine (G), thymidine (T) and cytosine (C). The human haploid genome has 3200 million nucleotides, but only 2 to 3% of it contains information resulting into proteins and the rest encode for non-coding ribonucleic acids (RNAs) and a huge amount of regulatory elements that have just started to become discovered[1]. The DNA sequence is 99.9% identical between humans, but mutations affecting only a single gene, are able to cause molecular defects leading to monogenic diseases like cystic fibrosis, hemophilia or sickle cell anemia. In complex traits such as atopic eczema (AE) and asthma, single nucleotide substitutions and other genetic variants may predispose to the phenotype and influence the response to treatments[2,3]. However, no genetic variant has been found necessary and sufficient to cause these complex diseases and typically, the heritable component does not follow Mendel's law of inheritance. A body of research conducted during the last decades strongly indicates that the molecular mechanisms leading to complex inflammatory diseases involve many more elements than variations in the DNA sequence[4,5]. Then, the information encoded by the nucleotides is part of a more general code, coordinated by a complex network of chemical modifications that make this "book of life" functional. Parts of that network are the epigenetic modifications, which include DNA methylation, histone modifications and non-coding RNAs[6]. These modifications recruit molecular complexes that activate or silence genes and explain how the cells contained in the human body, having (almost) the same DNA sequence, differentiate in around 400 different cell types with diverse phenotypes and expressed genes[7] (Figure 1).

The term "epigenetics" was coined in 1942 by Conrad Hal Waddington[8] and defined as "the branch of biology which studies the interactions of genes and their environment that bring the phenotype into being". With marvelous anticipation he predicted that once a progenitor cell take a differentiation pathway, among several possibilities, it progress downhill in the epigenetic landscape until a differentiated stage[9]. It is known today that terminally differentiated cells can become progenitors by resetting the chromatin structure and the DNA methylation landscape, like the case in induced-pluripotent stem cells[10], but the concept is still the same: epigenetic modifications acquired during differentiation remain as a cellular memory, which is heritable, and determine which genes are expressed, when they are expressed and where. The chemical modifications in the DNA bases or the chromatin, change gene expression and cell phenotype without changing the nucleotide sequence, and are

heritable by mitosis and/or meiosis. In recent years many new mechanisms like non-coding RNAs (including miRNAs) are being considered as epigenetic modifications because they are involved in maintaining the cell memory and are heritable[11]. The epigenetic code make cells to remember their phenotype by keeping the activation of genes that must be active and repressing those that are not needed. At the time I write this thesis, there is a body of knowledge on how epigenetic modifications work under physiological conditions in a number of organisms. However, with the exception of cancer, very little is know about the role of epigenetic modifications in the pathogenesis of human complex diseases.

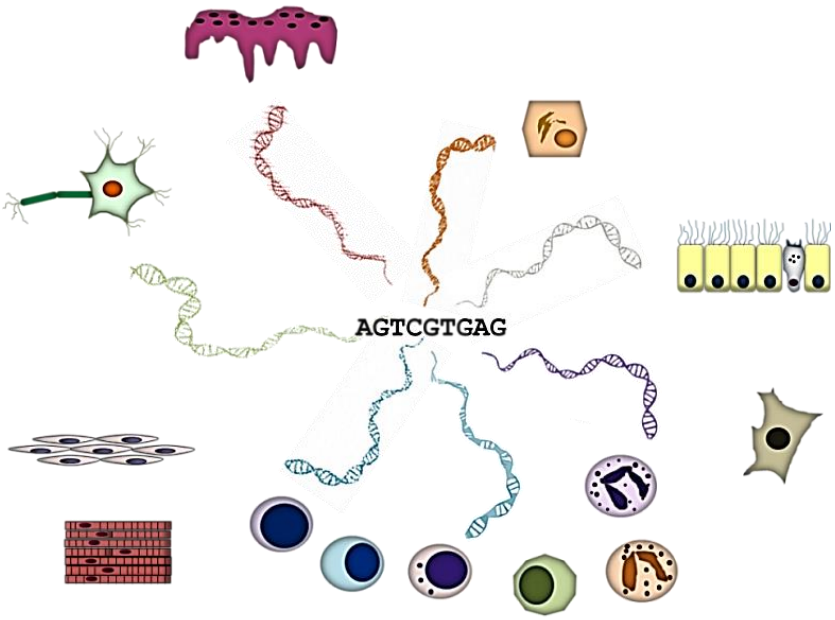

Figure 1
The human genome with almost the same DNA sequence in each cell results in millions of epigenomes with particular epigenetic landscapes according to cell lineage. Even within the same cell, the epigenome is remodeled by time, developmental stages and stimuli from the environment.

There are two ways of influencing the pathogenesis of complex diseases: genetic mechanisms and the effects of the environment. In this scenario, the epigenetic modifications are an interphase between the two general mechanisms. Environmental factors from diet, pollutants, the microbiome, or pharmacological compounds, can induce epigenetic modifications that result in changes in the phenotype, sometimes of clinical relevance[12]. By their plasticity, and being reversible and heritable, the

epigenetic modifications might hold the clue to understand key molecular events leading to complex inflammatory diseases. In the case of allergic diseases such as AE or allergic asthma, there are reasons to believe that epigenetic modifications may be involved in their pathogenesis: *first*, the discordance in the presence of the allergic phenotype between monozygotic twins[2]; *second*, differences in the age of disease-onset and the spontaneous relapse over time in certain patients[13]; and *third*, the changes in disease prevalence that followed industrialization and other global trends[14]. Given the main role of inflammation in AE and asthma, the peripheral blood leukocytes are relevant target cells to study the epigenetic mechanisms contributing to disease predisposition or perpetuation. This PhD thesis aimed to identify DNA methylation signatures that occur under physiological conditions in peripheral blood leukocytes during terminal differentiation and childhood, as well as to identify epigenetic signatures associated to AE and asthma. This would improve our understanding of complex diseases, will help to detect objective markers of disease and facilitate the development of novel therapies.

1.1 DNA methylation

This epigenetic modification consists of the covalent binding of a methyl group (CH_3) to DNA bases. The most studied form is the methylation of carbon 5 in the pyrimidine ring of cytosines (5-methylcytosine), when this nucleotide is followed by a guanine forming a CpG site (Figure 2a). The 5-methylcytosine was discovered in 1948 by Rollin Hotchkiss by using paper chromatography[15] , and it was not until 1975 that the connection between methylation and gene expression was first proposed[16,17]. The methyl group does not affect the pairing of the 5-methylcytosine to the complementary strand, but it is positioned in the major groove of the DNA where it may affect gene expression by recruiting methyl binding proteins, and other complexes like the polycomb group (PcG), histone deacetylases (HDACs), histone methyltransferases, and chromatin remodelers. Altogether, they promote silent chromatin states and affect the accessibility of the underlying locus to the transcriptional machinery[18]. The interconnection between DNA methylation and other epigenetic layers like histone modifications, results in strong functional effects on mRNA expression. DNA methylation plays an essential role in the genome stability, chromatin condensation, X-chromosome dosage compensation (X-inactivation), coordinated expression of imprinted genes and the regulation of cell-specific gene expression during the different phases of the cell cycle[19].

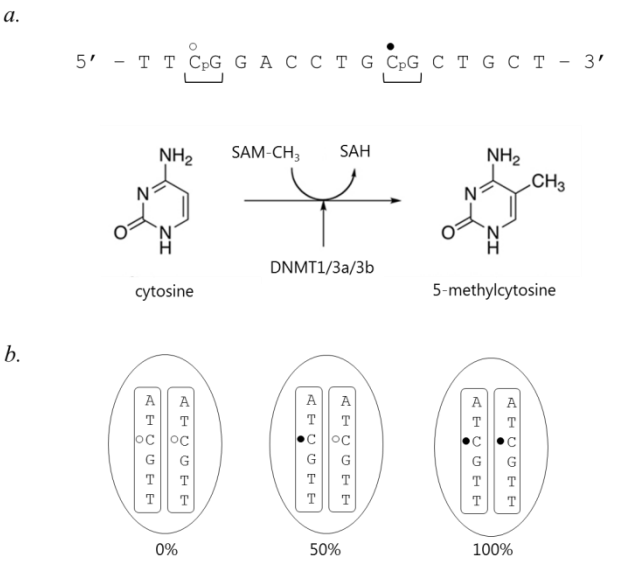

Figure 2

a. The nucleotide sequence reads 5' to 3' and there are two CpG sites formed by a C followed by G; unmethylated cytosine is indicated by a white circle; methylated cytosine is indicated by a filled black circle. The methyl group (CH_3) is taken by the DNA methyltransferases (DNMTs) from the S-adenosyl methionine (SAM) and added at the 5-carbon of cytosine. SAH: S-adenosyl homocysteine.

b. The methylation status of any given CpG site at the single cell level. Each vertical sequence represents the paternal and maternal chromosome in 5' to 3' orientation. A CpG site is unmethylated (white circle) or methylated (filled black circle) if the status of the cytosine is equal in the maternal and paternal chromosomes. There are situations when a CpG site is methylated in just one chromosome but not in the other as in imprinting. The percentage of methylation is indicated assuming symmetrical methylation on the complementary strand (not shown).

Since a single cell has two chromosomes from each parent, a given CpG site can be "unmethylated" if the methyl group is absent in the two parental strands or methylated if the methyl group is present in the two parental strands; under some conditions like imprinting, a CpG site can be methylated in just one parental chromosome (Figure 2b). Hemimethylation occurs if one strand is methylated but the complementary is not, and probably reflects post-replication lag in the maintenance of DNA methylation or *de novo* events. DNA methylation is one of the most studied epigenetic modifications by the strategic advantage that it can be robustly measured by a number of current methodologies[20], and by its stability under clinical collection conditions and storage. The DNA methylation measurements can be expressed as a quantitative value (*e.g.* percentage, Illumina β-value, M-value), that reflects the average methylation level of a specific CpG site in all the cells contained in any given sample.

It is estimated that there are 28 million CpG sites in the human genome and ~80% of them are methylated. More than one-third of the methylated CpG sites are located in retrotransposons and regions with highly repetitive DNA like Alu sequences, long interspersed nucleotide elements (LINE), short interspersed nucleotide elements (SINE) and centromeres. DNA methylation in these repetitive sequences is essential in maintaining genome stability[21,22]. Besides, there are short stretches of DNA sequence with a high content of CpG sites, called CpG islands (CGI) and defined as intervals of 200 to 500 base pairs (bp) with a CG percentage above 50% and an observed/expected CG ratio above 60%[23]. The CGI indicate the promoters of more than 50% of all human genes. Most CGI are unmethylated, even when the gene is transcriptionally inactive[24]. Aberrant methylation of promoter-associated CGI is typically observed in tumors. The CGI located inside the genes (intragenic) or in "apparently" non coding regions (intergenic) are usually methylated. There is evidence that they are biologically relevant and can be methylated in a tissue-specific manner[25]. The 2 kb regions immediately flanking the upstream and downstream of a CGI are called CGI shores (Figure 3a). They have lower content of CpG sites compared to a CGI. However, it has been found that CGI shores contain the highest variation in DNA methylation between cell types and also between healthy and cancer cells[26]. The regions of 2-4 kB immediately adjacent to the CGI shore are called CGI shelves. The rest is annotated as non-CGI related.

Methylated DNA in the proximity of transcriptional start sites (TSS) and in the surrounding of the first exon has been historically associated with suppression of gene expression[27]. However, DNA methylation levels and mRNA levels do not always follow an inverse relationship. Indeed, the location of the CpG sites within the gene structure have been found to play a crucial role in determining if a DNA methylation mark will lead to up-regulation or down-regulation of gene expression. The CpG sites mapped to coding genes can occur in promoters, untranslated regions (UTR), introns or exons (Figure 3b). In general, CpG sites located in gene bodies are strongly associated to increased gene expression in dividing cells[28,29] while if located in a promoter, DNA methylation can lead to either up-regulation or down-regulation of gene expression. DNA methylation in the gene body is very intriguing because this feature is conserved in evolution and seems to regulate the mRNA level of genes in the mid-expression range. Many tissue-specific differentially methylated regions have been found in the gene bodies, including intragenic CGIs[30]. Two functions have been found to be related with gene body methylation: regulation of alternative splicing and the prevention of transcriptional initiation in the gene body[31].

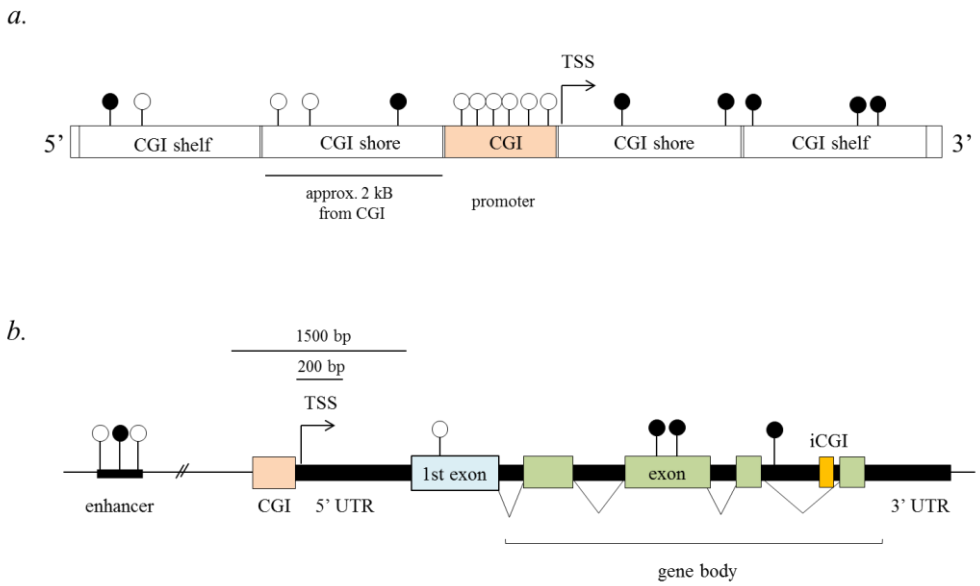

Figure 3
a. Location of CpG island (CGI), CGI shores and CGI shelves in relation to the transcriptional start site (TSS). CGI are usually CG-rich regions and their CpG sites are unmethylated (white lollipops). CpG island shores and shelves have less CpG sites compared to CGI but they are more frequently related to specific DNA methylation patterns depending on tissue or disease. CGI shores have been found to have strong effects on mRNA expression. Methylated CpG sites are indicated by black lollipops.
b. Location of CpG sites according to gene structure. Gene body is defined as the region immediately after the first exon. Methylation in intragenic CpG islands (iCGI) correlates with silencing of the associated gene[32].

1.1.1 The DNA methylation machinery

From early developmental stages through childhood and senescence the DNA methylation marks are written and erased, copied and edited. Some regions are very dynamic and can be modified by age or environmental exposures[33]. The DNA methylation machinery is responsible for keeping the fidelity of the code after each cell division or to modify it if needed. In eukaryotic cells the methyl groups are added after replication by DNA methyltransferases (DNMTs), a family of enzymes that catalyzes the transfer of a methyl group from S-adenyl methionine (SAM) to cytosines. DNMT1 is responsible for copying the methylation pattern from the parental strand to the daughter strand during mitosis, and is considered the maintenance methyltransferase (Figure 4). This enzyme is located in the replication fork and has preference by the hemimethylated DNA[34]. The DNA methyltransferases DNMT3A and DNMT3B provide *de novo* methylation during embryogenesis or in

response to environmental stimuli[35]. DNMT3A is expressed ubiquitously, while DNMT3B is poorly expressed by the majority of differentiated tissues with the exception of the thyroid, testes, and bone marrow[36]. Studies in knockout mice support that *Dnmt3b* is required during early development, whereas *Dnmt3a* is required for normal cellular differentiation[37]. The DNMT3L is homologous to DNMT3A and DNMT3B, but lacks the catalytic domain of DNMTs. Nevertheless, this enzyme stimulates the methyltransferase activity of DNMT3A and B[38] and it is required to establish maternal and paternal genomic imprints. DNMT3L is mainly expressed during early development and is restricted to germ cells and thymus in adulthood[39]. Targeted *de novo* DNA methylation may occur mediated by transcription factors (TFs) that either recruit DNMT3A or bind to the DNA sequence and protect the underlying locus from *de novo* methylation. CGI appear to be protected from methylation by TF binding, and when the TF binding sites are mutated the CGI are unable to retain their unmethylated state. The changes in the expression of TFs that occur during cell differentiation or aging may expose CpG sites to be targeted by the DNMTs machinery[40].

The removal of methyl groups (demethylation) also plays an essential role for cell biology and can occur by passive or active mechanisms. Passive demethylation results of decreased DNMT1 activity during cell replication, which renders the daughter strands unmethylated. Therefore, a CpG site initially methylated is replicated as unmethylated during successive divisions (Figure 4). Active demethylation requires several steps of enzymatic reactions that convert the 5-methylcytosine to cytosine. One step involve the addition of a hydroxyl group by ten-eleven translocator proteins (TET) that convert 5-methylcytosine to 5-hydroxi-methylcytosine (5hmC)[41]. The 5hmC can be further oxidized by TET proteins to 5-formyl-cytosine and 5-carboxyl-cytosine. Alternatively, 5-methylcytosine can be deaminated by the activation induced deaminase (AID/APOBEC) to form 5-hydroxymethyluracil. Thereafter the thymine DNA glycosylase (TDG) and base excision repair (BER) glycosylases replace these intermediates by cytosine and complete the DNA demethylation[42].

Right after fertilization the DNA methylation marks are globally deleted in the zygote (with some exceptions), which result in a hypomethylated genome with totipotent capacity to generate any cell type in the preimplantation embryo[43]. The demethylation of the paternal genome is an active process that removes DNA methylation marks within hours after fertilization while the maternal genome is passively demethylated during cleavages up to the blastocyst stage[44,45]. In the gastrulating embryo the DNA methylation marks are replenished by *de novo*

methylation. Thereafter, DNA methylation participates in defining the epigenetic landscape of each cell lineage (as *n* methylomes, Figure 4), guiding the differentiation processes, reinforcing cell commitments and preventing regression to undifferentiated states. A second phase of re-programming occurs in the primordial germ cells during embryogenesis where DNA methylation patterns are erased in coding genes and some repetitive elements[46]. After birth, some loci exhibit changes in DNA methylation in differentiated cells that affect single CpG sites or regions[33].

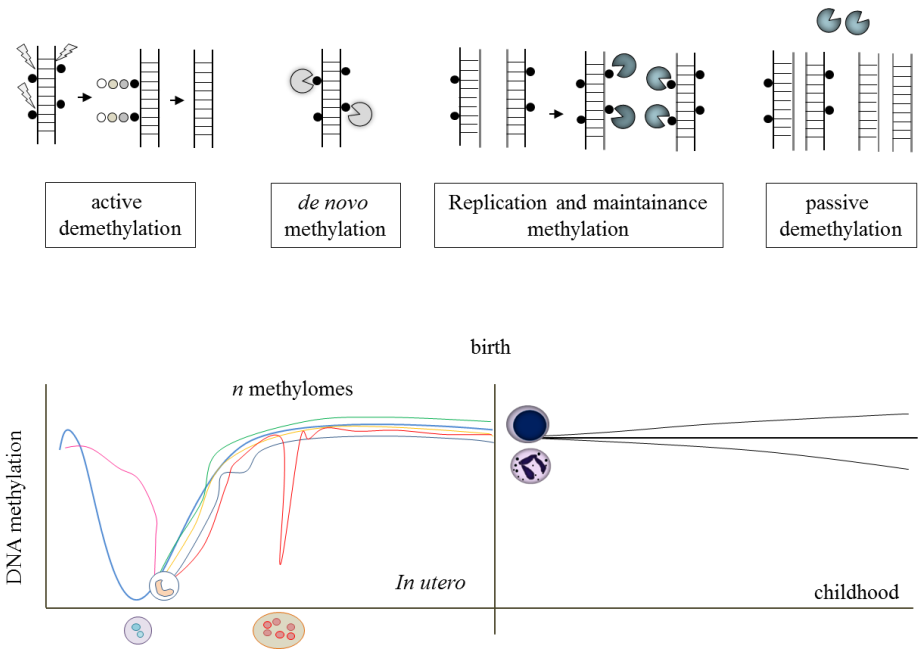

Figure 4
DNA methylation levels are in constant change during development, in response to physiological processes and the environment. Filled black circles indicate methyl groups; gray circles indicate chemical intermediates of 5-methylcytosine; lightning indicate TET proteins; pie symbol indicate DNMTs; the daughter strand after DNA replication is presented in gray.

The DNA methylation marks are read by at least three families of proteins: the methyl-CpG-binding domain (MBD) proteins[47], the ubiquitin-like, containing PHD and RING finger domain (UHRF) proteins and zinc finger proteins. MBD proteins include MeCP2 along with MBD1, MBD2, MBD3 and MBD4. This family mediates transcriptional repression by recruiting chromatin remodeling complexes and histone deacetylases (HDAC)[48]. MBD4 is a repairing enzyme that preferentially recognizes

deaminated 5-methycytosines. The UHRF protein family (particularly UHRF1) actually recognizes hemimethylated DNA during DNA replication and recruits DNMT1 to catalyze the methyl-transference reaction[49,50]. The zinc finger proteins bind to methylated DNA by a zinc-finger domain and include Kaiso, ZBTB4, and ZBTB38[51]. They can bind single methylated CpG sites or two consecutively methylated CpG sites and repress transcription in a DNA methylation-dependent manner. Overall, methyl binding proteins provide the strongest link between DNA methylation and histone modifications, the combinations involved can modify gene expression and maintain the methylation code[52].

1.1.2 DNA methylation and chromatin remodeling

The DNA methylation patterns contribute to regulate gene expression and thereby cell-specific functions. However, since the DNA is wrapped in high order structures like nucleosomes and chromatin, a gene can only be transcribed if the chromatin structure changes to allow transcription[53]. The nucleosome is the functional unit of the chromatin, and consists of 147 bp wrapped in an octamer of globular proteins called histones (one H3 H4 tetramer and two H2A-H2B dimers). Those are positively charged and have a great affinity for the DNA which is negatively charged. The aminoacids of the histone tails are targets of post-translational modifications (*e.g.*, acetylation, methylation, ubiquitination, phosphorylation, sumoylation) that enable or prevent the access of regulatory proteins and the transcriptional machinery to the DNA[54]. Approximately 60 aminoacid residues can be modified in core histones but only a small number of modifications have been associated to changes in transcriptional activity[55]. In most cases histone acetylation and phosphorylation lead to gene activation whereas histone methylation, desacetylation, and ubiquitination lead to gene inactivation. Histone modifications are added and removed by a newtwok of enzymes that work in coordination with chromatin remodelers[56]**.** Histone acetylation is mediated by lysine acetyltransferases (KAT) which transfer an acetyl group from acetyl-CoA to an ε-amino groups in lysines. Removal of acetyl groups is catalized by histone deacetylases (HDAC). Moreover, histones can be mono-, di- or trimethylated at lysine residues by lysine methyltransferases (KMTs), which transfer a methyl group from SAM to ε-amino groups in lysines. The removal of methyl groups from histones is catalyzed by lysine demethylases (KDM). The lysine-specific demethylase 1 (LSD1 or KDM1A) removes mono- and di- methylation and the jumonji demethylase family is able to remove trymethylation[57,58]. The methylation of lysines can lead to activation or repression depending on the residue

that is modified, for instance H3K4me3 is correlated to active gene expression while H3K9me2/3 is a repressive mark. Histone modifications can also recrutit DNMT3A which *de novo* methylate the CpG sites wrapped in that histone[59].

The changes in the structure of chromatin are mediated by ATP-driven molecular machines called chromatin remodelers (CR)[60]. They translocate the nucleosomes or evict them from the DNA, thereby affecting the accessibility of the transcriptional and the repairing machineries to the template DNA. The catalityc subunits associate to non-catalityc ones in complexes. CR are divided in four main families based on the sequence of their ATPase subunit: switching defective/sucrose nonfermenting (SWI/SNF), imitation switch (ISWI), CHD (chromo domain helicase DNA-binding) and inositol requiring 80 (INO80). The combined effects of CR, histone modifiying enzymes, chaperones and DNMTs are transient and very dynamic. The cell phenotype and/or cell-specific gene-expression are the outcomes. By the time that a sample of differentiated cells is assessed by current methodologies many of these molecular complexes (and their potential variations under disease traits) can have vanished. In this scenary, DNA methylation marks at single nucleotide resolution or in differentially methylated regions, may represent a vestige of other molecular events that underwent in a locus. This concept may be important to bear in mind when interpreting the effect sizes of comparing DNA methylation levels.

1.1.3 The polycomb group (PcG)

The polycomb group is a family of transcription regulatory proteins that play a crucial role in developmental processes and cell differentiation by creating and maintaining repressive chromatin environments. PcG proteins associate in molecular complexes that allow them to regulate the transcription of a great number of genes from the early embryogenesis, through infancy and adulthood. The polycomb was first described in *Drosophila* as repressors of the homeotic genes (Hox) in the bithorax complex[61]. The PcG mediate gene silencing by remodeling the chromatin and are classified as epigenetic regulatory proteins. PcG proteins form two major polycomb repressive complexes: PRC1 and PRC2 that are compositionally and functionally conserved in flies and humans[62,63]. In humans, PRC1 is formed by CBXs (polycomb homolog), PHC1,2, and 3 (polyhomeotic homologs), Ring1a and Ring1b (as dRING homologs) and, BMI1 and six minor others (as posterior sex combs homologs). The human PRC2 includes EZH1/2 (as enhancer of zeste), SUZ12 (as supresor of zeste 12), embryonic ectoderm development (EED) and RbBP4[64] (Figure 5). A known function of PcG proteins is the catalysis of H3K27. First PRC2 binds to

chromatin and the catalytic subunit EZH2 trimethylates H3K27, then H3K27me3 is recognized by the CBX components of PRC1, the ligases RING1/2 monoubiquitinates H2A on K119 resulting in chromatin compation[65].

At present, it is still unclear how the PRC2 is directed to its target promoters in mammals, but some studies in mouse stem cells have shown the participation of proteins like JARID2[66] and polycomb like proteins (Pcl)[67]. Furthermore, subunits of the PcG directly interact with the DNMTs and play a crucial role in regulating genome-wide and locus-specific methylation patterns. For instance EZH2 serves as a scaffold and direct link between methylated histones and the DNA methylation machinery at repressed promoters. The binding of DNMTs to repressed genes depends on the presence of EZH2[68]. Most of the knowledge currently available on PcG proteins is related to their functions in regulating stem cell activity, developmental processes and cancer[69]. There are many other family members of the PcG and associated molecules which biological functions and interactions are less known. A summary of the human polycomb complexes and some of their known interactors are presented in Figure 5.

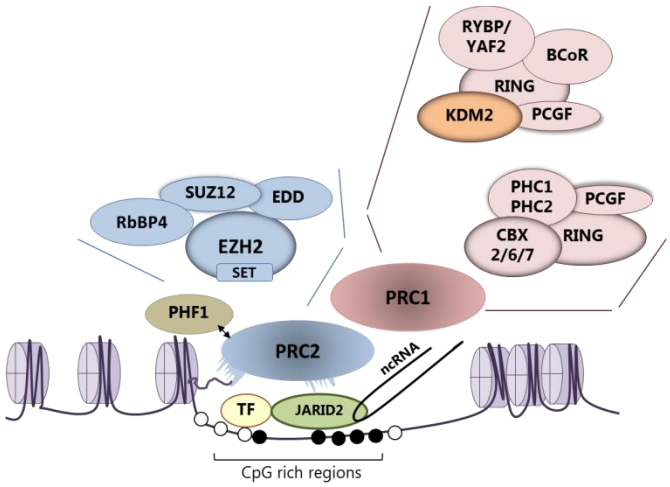

Figure 5
The components of the human polycomb group (PcG). DNA is represented by a purple line turning around histone cores (lilac). CpG sites are denoted by white circles (unmethylated) and filled black circles (methylated). Transcription factors (TF), Jumonji AT rich interactive domain 2 (JARID2) and non-coding RNAs bind to CpG rich regions and recruit the PRC2 (blue) and the canonical and non-canonical versions of the PRC1 (pink). PHF1: PHD finger protein 1; KDM2: Lysine (K)-specific demethylase 2.

1.1.4 5-methylcytosine as an alternative allele

An additional level of complexity is added by single nucleotide polymorphisms (SNPs) that coincide with CpG sites and influence *in cis* the levels of DNA methylation in other CpG sites located in their surroundings. DNA methylation has been found as a heritable continuous quantitative trait in several tissues including brain and peripheral blood leukocytes[70]. Some of the genetically regulated DNA methylation is directly connected with variation in gene expression[71]. The majority of CpG sites are non-polymorphic, but 5-methylcytosine is prone to mutagenic events that cause C to T transitions. When a SNP coincides with a CpG site (for example a C is changed to T), the CpG site is removed and there is no possibility of methylation for the carriers of the T allele (TpG) (Figure 6a). Alternatively, if the polymorphic variant is a C or a G, they can create a CpG site. As mentioned above, a methylated cytosine can bind nuclear proteins that do not bind to unmethylated cytosines, and therefore can modify the binding of TF or recruit other nuclear molecules[72,73] (Figure 6b). This setting configures the 5-methylcytosine as the third allele.

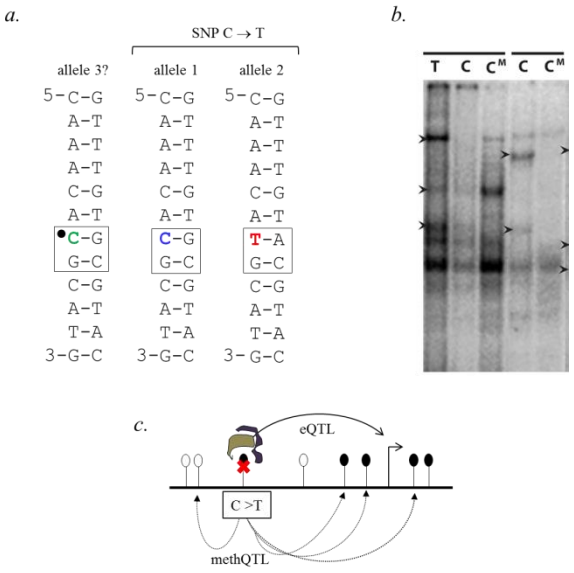

Figure 6

a. A CpG-site SNP. The cytosine has the possibility to be methylated (filled black circle) as the 3rd allele.
b. Electrophoretic mobility shift assay comparing nuclear protein binding between probes with allele T, allele C or the methylated allele C (C^M), modified from Reinius *et al.*[72]
c. Schematic model of a CpG-site SNP affecting the DNA methylation levels of other non-polymorphic CpG sites in its surroundings (methQTL) by the differential recruitment of nuclear proteins by the polymorphic allele; white lollipops indicate unmethylated CpG sites; black filled lollipops indicate methylated CpG sites; eQTL: expression quantitative trait locus.

Alternatively, SNPs can influence regional DNA methylation levels in other CpG sites located in their surrounding[74] (Figure 6c). These CpG site-SNPs are quantitative trait locus for methylation (methQTLs). Studies have shown that the *cis* methQTL effect occurs in a range of 45 bp from the site in question and is most likely in CpG sites located outside the CGI[75], although effect-windows of up to 1 Mb have been described[70]. The association between alleles and regional DNA methylation levels highlight the fact that the underlying DNA sequence (even if it is not directly modified in terms of the nucleotides), can definitively influence the epigenetic modifications placed on it (in particular allele-specific DNA methylation[76]). It has been estimated that about 20% of the differences in DNA methylation levels between individuals has a genetic association[77].

1.1.5 DNA methylation and leukocyte differentiation

The process of cellular differentiation is the hallmark of epigenetic regulation and indeed, epigenetic changes mediate and maintain the fate that multipotent hematopoietic progenitors will follow in a differentiation pathway towards developing into a lymphocyte or a neutrophil. Studies on the epigenetic marks (including CpG sites) that may categorize cell populations according to their lineage have been under extensive research for more than two decades. A comprehensive methylome map of hematopoietic progenitors revealed that DNA methylation correlated with gene expression and is directly implicated in lineage choice[78]. Indeed, alterations in the DNA methylation patterns of hematopoietic progenitors have been found to cause many hematological malignancies[79,80]. Studies in mice and human cells have shown that DNA methylation increases with the commitment to the lymphoid lineage and decreases with the myeloid differentiation program[78]. The analysis of candidate genes has identified some CpG sites and genes that can be used to predict lineage commitment in human leukocytes[81-83]. However, little is known about the genomewide DNA methylation landscape of terminally differentiated leukocytes that circulate in peripheral blood.

In a healthy adult about 10^{12} new blood cells are produced every day by the bone marrow. The presence of colony stimulating factors and cytokine signals in the microenvironment, induce epigenetic changes that determine the paths that the hematopoietic progenitors will follow until a differentiated state and confer their cell-identify and cell-specific functions. Lymphoid cells originate from the common

lymphoid progenitor and include T cells, B cells, natural killer cells (NK)[84] and other populations of innate lymphoid cells[85]. Myeloid cells originate from the common myeloid progenitor (granulocyte-macrophage lineage) and include granulocytes (neutrophils, eosinophils and basophils), monocyte/macrophages and mast cells. The dendritic cells (DC) originate from both, lymphoid and myeloid progenitors, and are found in very small proportions in blood while they are in transit to the tissues as plasmacytoid DCs ($CD303^+$), or myeloid DCs ($CD1c^+$ and $CD141^+$). Since lymphoid and myeloid cells have profound differences in their DNA methylation levels, the estimates of DNA methylation obtained from whole blood can be affected by cell heterogeneity, especially if the targeted CpG sites coincide to those implicated in cell linage. By the time I started my thesis, this limitation was not widely recognized. Several studies were already published studying methylation in DNA in samples obtained from whole blood[86], and many research groups around the world were in the transition to evaluate the DNA samples left from the genome-wide association studies (GWAS) era with the arrays for DNA methylation-wide association studies. Blood contains a complex mixture of leukocytes that account for 0.1 to 0.2% of all components and are divided in lymphocytes and myelocytes; the rest are erythrocytes (93-96%) and platelets (4-7%). Peripheral blood leukocytes are distributed in different proportions that vary depending on the individual, age, and the presence of infections or inflammatory conditions (Table 1). The immuno-phenotype of the leukocytes can be identified by flow cytometry with the use of cell surface markers that are characteristic of each cell type.

T lymphocytes are distinguished from B cells and NK cells by the presence of the T cell receptor (TCR) on their cell surface. They are divided in T helper cells, which express the CD4 glycoprotein ($CD4^+$ cells) and T cytotoxic cells expressing the CD8 glycoprotein ($CD8^+$), and are present in peripheral blood in a rough 2:1 ratio[87]. Both subsets of T cells can be divided in conventional naïve if they express the $CD45RA^+$ marker, or antigen-experienced memory T cells if they express $CD45RO^+$. The latter are antigen specific cells that originate from naïve T cells upon antigen stimulation and can live for long periods of time as central memory T cells or effector memory T cells. Other T cell populations in minor proportions in blood include regulatory T cells (inducible and natural), natural killer T (NKT) cells, scarce gamma delta T cells and $CD4^+CD8^+$ T cells. It is worth to keep in mind that even within the same population (*e.g.* $CD4^+$ T cells), the circulating T cells correspond to a mixture of diverse developmental stages, activation status and imprinted cytokine patterns (*e.g.* Th1, Th2, Th17, Th9, induced T regs).

Table 1. Frequencies of leukocytes in human peripheral blood (adults)

	frequency ($x10^6$ cells/mL)	% of circulating leukocytes	Conditions
lymphocytes	1.1-3.5	14-47	
T cells	0.54-1.79	7-24	↑ viral infections
$CD4^+$ T	0.30-1.50	4-20	
naïve $CD4^+$ T	0.08-0.76	1-10	
memory $CD4^+$ T	0.25-0.81	3-11	
regulatory $CD4^+CD25^+$	0.007-0.05	0.1-0.7	
$CD8^+$ T	0.14-0.82	2-11	
naïve $CD8^+$ T	0.03-0.21	0.4-2.6	
memory $CD8^+$ T	-	1-7	
B cells	0.07-0.53	1-7	↑infectious mononucleosis (EBV)
naïve B cells	0.05-0.37	0.7-4.9	
memory B cells	-	0.2-1.7	
NK cells	0.08-0.43	1-6	
myelocytes			
monocytes	0.20-0.90	2-12	↑SLE, ↑RA
granulocytes	2.13-6.35	35-80	↑bacterial infections
neutrophils	2.09-5.97	30-80	↑bacterial infections ↑COPD
eosinophils	0.03-0.30	0-7	↑allergy, asthma, AE ↑helminthic infection
basophils	0.01-0.08	0-2	

COPD: chronic obstructive pulmonary disease; DC: dendritic cells; EBV: Epstein-Barr virus; NK: natural killer; RA: rheumatoid arthritis; SLE: systemic lupus erythematosus. Myeloid DC represent 0.3 to 0.9% of all leukocytes; plasmacytoid DC represent 0.2 to 0.6% of all leukocytes. The values presented in this table were obtained from stemcelltechnologies™.

The $CD19^+$ lymphocytes found in peripheral blood correspond to sub-populations of B cells that have differentiated from the early B cell precursor into mature B lymphocytes and emigrated from the bone marrow to the secondary lymphoid organs and tissues. Peripheral blood B cells can be grouped in immature/transitional (5.4%), naïve (64%), memory B-lymphocytes (31%), and plasmablasts (2.1%). Based on the differential expression of immunoglobulin heavy chains, the B cells can be further divided in IgM^+, IgD^+, IgG^+, IgA^+ and very scarce IgE^+ cells.

The natural killer cells (NK) play critical roles in innate immunity, shape adaptive immune responses and even participate as immune-regulatory cells. There are at least five subtypes of NK cells in human peripheral blood based on the expression of CD56 and CD16. The $CD56^{dim}CD16^{bright}$ subset represents 90% of circulating NK cells; they contain the inhibitory killer cell receptors (KIR), have strong cytotoxic activity and high concentration of perforins and granzymes[88]. A maximum of 10% of all circulating NK cells are $CD56^{bright}$ in combination with $CD16^-$ or $CD16^{dim}$.

Regarding myeloid cells, human monocytes circulate few days in blood before they migrate to tissues and develop into different types of macrophages. Based on the expression of the cell surface marker CD14, they are divided into classical ($CD14^{++}CD16^{-}$), intermediate ($CD14^{++}CD16^{+}$) and non-classical ($CD14^{+}CD16^{++}$) monocytes[89]. It is worth to mention that monocytes have a non-segmented nucleus and by their density they isolate with the peripheral blood mononuclear cell fraction (PBMC).

The peripheral blood granulocytes include the neutrophils, eosinophils and basophils. The neutrophils are the most abundant cell population in blood and the daily production can reach up to $2x10^{11}$ cells. Once released from the bone marrow as polymorphonuclear (segmented) cells, they can survive up to 4 to 5 days. They have low transcriptional activity but store several pro-inflammatory molecules and enzymes in their granules. Neutrophils are the first line of defense against bacterial infections by acting as phagocytes and creating extracellular traps (NETs)[90]. Eosinophils are innate immune cells with lobulated nucleus that have a life span of 2 to 5 days. Their cytoplasmic vesicles are loaded with cytotoxic enzymes that are crucial in the defense to parasites; however, eosinophilia is a common feature of allergic diseases induced by the production of IL-5[91]. Basophils are the least common of the granulocytes and account for less than 1 percent of all leukocytes. They express the FcεRI on the surface and their activation leads to granule exocytosis. Basophils release heparin, histamine and elastase and are potent producers of IL-4 and IL-13. They share many features with mast cells (present in tissues) but the latter constitute a different cell lineage[91].

1.1.6 DNA methylation and human diseases

Mutations in genes encoding DNMTs and other enzymes of the human methylation machinery usually result in large-scale methylation defects, and if the conceptus is viable, into clinical syndromes like immunodeficiency centromeric instability and facial abnormalities (DNMT3B, OMIM#242860); Rett syndrome (MeCP2, OMIM#312750); and hematological malignancies (DNMT3A, OMIM# 602769). There are other conditions in which the DNA methylation landscape is affected by epimutations that lead to imprinting disorders such as the Beckwith-Wiedemann syndrome, Silver-Russell syndrome, Prader-Willi and Angelman syndrome among others[92]. DNA methylation defects have been also associated to human diseases with repeat-instability involving folate-sensitive fragile sites (i.e. Fragile-X syndrome) and methylatable CGG repeats[93]. One of the best demonstrated effects of altered DNA methylation in human diseases is the development of malignancies[94], with genome-

wide hypomethylation and locus-specific hypermethylation of tumor-suppressor genes as hallmarks. Hypomethylation usually occurs in repetitive regions and retrotransposons resulting in genomic instability, aberrant activation of oncogenes and loss of imprinting. Gene-specific methylation usually occurs in the promoter regions of tumor suppressor genes[95]. For the interested reader, the database Pubmeth (http://www.pubmeth.org/) contains curated information on the genes with altered DNA methylation in diverse types of cancer. Research on the field of schizophrenia, bipolar disorder, autism spectrum disorders, anxiety and depression suggested that subtle DNA methylation changes are involved in the pathogenesis of complex traits[96].

Regarding complex inflammatory diseases, differences in DNA methylation has been found in systemic lupus erythematosus (SLE); for instance, the genomes of the T cells in SLE patients have global hypomethylation and the levels of DNMT1 are reduced[97]. Differences in DNA methylation have been described for Type I diabetes[98]. However, the differences in methylation that have been detected between patients affected by complex inflammatory diseases and controls are typically very small. Some studies applied cut-offs of a mean difference in DNA methylation levels of 5% or 10% between affected and non-affected groups at single nucleotide resolution, but the functional outcomes of these small DNA methylation changes are still largely unknown.

1.2 microRNAs

In the last years new empirical evidence in plants, fungi and metazoans revealed that non-coding RNAs (ncRNAs) play a role as epigenetic regulators[11,99,100]. Depending on their length in nucleotides (nt) these molecules can be grouped in small non-coding RNAs (<30 nt) and long non-coding RNAs (>200 nt). Both groups have shown to have a role in regulating chromatin structure, heterochromatin formation, histone modifications, DNA methylation targeting and gene silencing. The group of small ncRNAs contains the miRNAs, short interfering RNAs and the piwi-interacting RNAs. There is still controversy whether those molecules can be regarded as epigenetic mechanisms, but indeed they modify gene expression without affecting the DNA sequence and are heritable[11,101]. In the case of miRNAs, experiments in plants have shown that miR-165 and miR-166 are required for the methylation at the *PHABULOSA* (*PHB*) gene in *Arabidopsis*. They interact with the newly processed *PHB* mRNA and alter the chromatin in the *PHB* DNA template[102]. Other studies have shown that miRNAs control the expression of epigenetic regulators by directly

targeting DNA methyltransferases, histone deacetylases and PcG genes[103]. A curated list of examples how miRNAs regulate epigenetic modifications can be found in the EpiMiR database[104]. The most studied ncRNAs are the miRNAs and in the following lines I will refer to their biology.

miRNAs are small non-coding RNA molecules (~22 nucleotides, single-stranded) that regulate the translation of protein coding genes. This mechanism of post-transcriptional regulation implicates complementary base-pairing between the miRNA and the target RNA molecule. In animals the interaction between the miRNA and the mRNA target require as few as 8 nucleotides (the seed region) and usually the complementary site is located on the 3'UTR of the target mRNA. Any given miRNA can have hundreds of mRNA targets, and a mRNA transcript can be targeted by several miRNAs. It has been estimated that each conserved vertebrate miRNA can have ~400 targets[105,106]. The nomenclature of miRNAs is simply sequential, lettered suffixes denote closely related mature sequences (*e.g.* miR-18a and miR-18b). When two miRNA sequences (22 nt) originate from the opposite arms of the same precursor but it is not possible to determine which is the predominant one, the suffix 5p (from the 5' arm) and 3p (from the 3'arm) are used.

1.2.1 Generation of miRNAs

miRNA genes are encoded in the nuclear DNA and represent ~4% of the human genes. They are encoded in different parts of the genome either in dedicated miRNA genes, usually found in intergenic regions, or by sequences embedded in the antisense strand of a host mRNA coding gene. About 40% of all miRNAs are produced in polycistronic units with multiple loops. The miRNA genes as any other gene are subject to epigenetic regulation by DNA methylation and histone modifications. The miRNA genes are transcribed by the RNA polymerase II into a long RNA with hairpin structure, a 5' 7-methylguanosine cap and a 3' polyadenylated tail (primary miRNA). The pri-miRNA is processed in the nucleus by the microprocessor complex (Drosha and DGCR8). The result is a stem-loop structure of approximately 70 nt called the pre-miRNA that is exported to the cytoplasm by the karyopherin exportin 5 and the Ran-GTP complex (Figure 7). Then the pre-miRNA is processed by the enzyme Dicer that generates the miRNA as a double stranded RNA of 21-23 nucleotides in length. The two strands are separated by helicases, after unwind the active or "guide" strand is loaded into the RNA-induced silencing complex (RISC) and repress the target mRNA by blocking translation or causing transcript degradation. The inactive strand is removed and degraded[107].

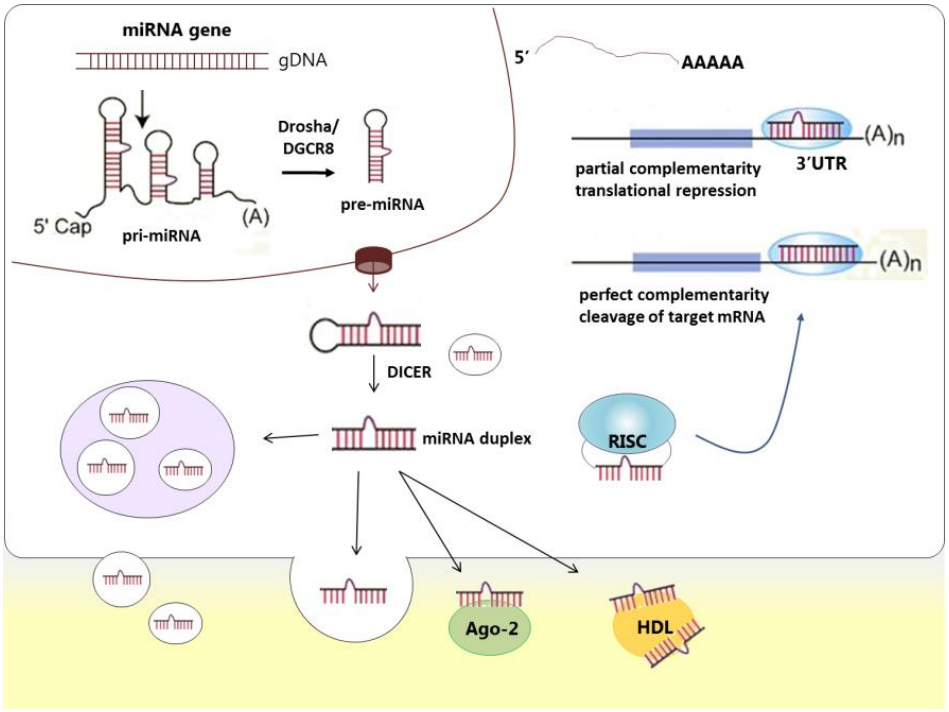

Figure 7
The microRNA gene (usually as a polycistronic unit) encodes the primary miRNA transcript (pri-miRNA) after transcription by RNA polymerase II. This transcript may contain several stem-loop structures that are cleaved by the RNA-polymerase type III Drosha to form the pre-miRNA. This step requires the presence of the cofactor DGCR8. After being exported from the nucleus the pre-miRNA is processed by DICER to form a double strand RNA duplex with the mature miRNA and its complementary strand. The RISC complex then recruits the strand with the lowest thermodynamic stability at its 5' end. Indeed both strands (5p and 3p) can have functional effects and be more represented in some tissues. The mature miRNAs can be also released to the extracellular space in microvesicles or in molecular complexes with the protein Argonaute-2 (Ago) or high-density lipoprotein (HDL). DGCR8: DiGeorge syndrome critical region gene 8.

1.2.2 Biological effects of miRNAs

It has been calculated that miRNAs control up to one third of all human genes[108]. The initial observations on the biological role of miRNAs were done in the nematode *Caenorhabditis elegans*[109], and during the last years several studies have demonstrated their critical role in development, cellular differentiation, and the maintenance of tissue-specific expression profiles in humans[110]. The mechanisms of post-transcriptional regulation by miRNAs are predominantly based on inhibition of

translation. Therefore miRNAs are generally regarded as negative regulators of gene expression by: a) co-translational protein degradation; b) inhibition of translation initiation, and c) premature termination of translation (ribosome drop-off). Depending on the degree of complementarity between the miRNA and the target, animal miRNAs can also induce cleavage (under perfect duplex), degradation of mRNAs or can silence their targets by sequestering the mRNA in discrete cytoplasmic foci that exclude the translational machinery [111]. In recent studies it has been observed that miRNAs can also interact with the 5'UTR of the target genes and with the coding regions (CDS). Depending on the characteristics of these interactions, the miRNAs can increase the translation of a target mRNA by recruiting protein complexes to the AU-rich elements on the transcript[112]. This is important when evaluating the effects of differentially expressed miRNAs/mRNAs by cross-pairing. It is worth to mention that given the nature of the complementarity between the miRNA and its targets, the biological effect of a miRNA represents the synergy of moderate effects on multiple mRNAs. Some miRNAs has been found targeting genes that control chromatin structure like PcG genes (*EZH2* and *Bmi-1*), histone deacetylases (*HDAC1* and *HDAC4*) and DNMTs (*DNMT3A* and *DNMT3B*)[113]. Provided that miRNAs not only target mRNA transcripts, but also other non-coding RNAs and proteins, the possibilities for miRNA-mediated effects goes beyond post-transcriptional regulation of mRNA transcripts. Moreover, the miRNA effects may act on the same cell that produces them, but if released to the extracellular compartment (Figure 7), they can be internalized and regulate biological processes in distant cells.

1.2.3 miRNAs as biomarkers and therapeutic targets

Alterations in the miRNA levels or its regulators have been implicated in almost all types of human cancer. Moreover, altered expression of miRNAs has been detected in serum and plasma from patients with non-oncological pathologies like active tuberculosis, acute myocardial infarction, hepatitis and sepsis[114]. The Human miRNA disease database (HMDD) provides a comprehensive tool to explore the relation of human diseases and miRNAs. Profiling miRNA expression patterns can classify hundreds of tumor samples to their tissue of origin[115], even in cases of unknown primary with an accuracy of 90%[116]. After these initial observations hundreds of studies have described the use of miRNAs as potential diagnostic tools, especially in cancers, because they can be detected in biofluids like plasma, serum, saliva and

urine. The study of miRNAs is of great interest because their potential as non-invasive biomarkers to help the diagnosis and the prognosis of inflammatory diseases, to monitor the response to treatments or to be used in patient stratification and individualized medicine. The comparative profiling of miRNAs between patients and controls provide the possibility to identify critical miRNAs involved with the disease and to develop molecules that can modify their production. Approaches to block the production of a specific miRNA include miRNA mimics or anti-miRs by using locked nucleic acids (LNA), antagomirs, and miR sponges. Two antagonists have reached clinical trials (i.e. miravirsen, MRX34), but most miRNA therapeutics are still in preclinical development[117,118].

1.3 Allergic diseases

The immune cells are usually beneficial when they participate in protection from infections and tumors but can generate several clinical problems when they recognize autoantigens or innocuous environmental molecules that are not pathogenic. An allergic disease is a clinical condition in which the immune system sustains an altered response to molecules that otherwise are tolerated by the majority of the population. This inflammatory response induced and maintained by defective immuno-regulatory networks, lead to the expansion of inflammatory cell populations and to tissue changes that affect organ function and result in clinical symptoms. The most common allergic diseases include atopic eczema, allergic asthma, allergic rhinoconjuntivitis and food allergy. They usually co-segregate with atopy, an intermediate trait defined as the genetic predisposition to produce elevated levels of circulating total IgE and/or allergen-specific IgE[119,120]. In the last decades there was a dramatic increase in the prevalence of allergic diseases worldwide suggesting that something has changed in our modern environments that lead to the dysregulation of the immune responses[121]. During the last years, epigenetic modifications started to be investigated as an interphase between the environment and the susceptibility to allergic diseases[122]; an updated summary of these studies have been recently published by Harb *et al*[123] and Lockett *et al*[124]. However, a PubMed search with the term "allergy epigenetics" (February 2015) yielded 213 publications, whereof 97 were review articles. Thus, original research in the field is still very scarce, and almost half of all publications represent reviews that typically point out our ignorance and need for research.

1.3.1 Atopic eczema

Atopic eczema (AE), also called atopic dermatitis, is a common inflammatory skin disease characterized by intense pruritus and xerosis with disease onset in 90% of cases before the age of 5 years[125]. The majority of children with AE have a spontaneous remission by puberty but at least one third of the patients continue to have eczema or relapsing AE in adulthood[126]. By still unknown reasons, the prevalence of AE has increased during the last four decades in industrialized countries, affecting today 15 to 30% of children and 2 to 10% of adults[14,127]. The exact molecular mechanisms that cause AE are largely unknown but a combination of skin-barrier defects with alterations in immune regulation is critical for disease development. The majority of AE patients present elevated levels of circulating total IgE and allergic sensitization to food and environmental allergens. Genetic factors are the most prominent determinants for the development of AE, with a concordance rate in monozygotic twins ranging from 72 to 86%[128]; however, the susceptibility loci identified so far account for 14.4% of AE heritability[129]. Several genes have been associated to AE including loss-of-function mutations in the filaggrin gene (*FLG*) on chromosome 1q21.3[130] and other components of the skin barrier function such as the serine peptidase inhibitor, kazal type 5 (*SPINK5*), and epidermal collagen[131]. Still, not all individuals with mutations in skin-related genes develop AE thereby additional alterations in the immune system are required to disease inception. Indeed, genetic association with the *RAD50/IL13* locus and HLA alleles have been detected and replicated in European populations[132], and the most significant susceptibility regions as detected by GWAS contain genes that underline the importance of the immune dysregulation[133].

The histologic features of acute eczematous lesions are epidermal intercellular edema and a prominent perivascular infiltrate of lymphocytes, monocytes, dendritic cells, mast cells, and few eosinophils in the dermis. In subacute and chronic stages the plaques are lichenified and excoriated, the epidermis is thickened and its upper layer is hypertrophied[125]. The current knowledge on the pathogenesis of AE indicates that a defective skin-barrier (deficiency of ceramides and antimicrobial peptides such as cathelicidins, plus defects in structural components in the desmosomes) lead to trans-epidermal water loss and increased penetration of allergens, microbes and irritants. These exposures activates the keratinocytes to produce increased amounts of thymic stromal lymphopoietin (TSLP) and the dendritic cells (DC) to produce alarmins (IL-25, IL-33), chemokines and growth factors that lead to a Th2 polarization of the T cell responses in the acute lesions with predominance of IL-4, IL-5, and IL-13 production[134]. In the chronic lesions there is a predominance of Th1 cells and

increased production of interferon gamma, IL-12, IL-5, and granulocyte macrophage colony-stimulating factor (GM-CSF). An alternative hypothesis sustain that immunological defects lead to IgE mediated allergic sensitization and upon exposure to the triggers, the inflammation alters the epithelial barrier, however, early onset atopic dermatitis usually emerges in the absence of detectable IgE sensitization[125].

Common triggers of flares include heat, sweating, anxiety and infections. Among all these exposures, colonization of skin by bacteria and fungi has shown to have important roles. Indeed, more than 90% of AE patients have skin colonization by *Staphylococcus aureus*, and even in the absence of signs of infection, the patients improve with antibiotic treatment[135]. *Malassezia ssp.*, is a commensal fungus that colonizes our skin after birth, but is also associated with several skin diseases among those AE[136]. Around 50% of AE patients have detectable specific-IgE sensitization to *Malassezia* allergens and the levels of IgE correlate with disease severity[137,138]. Gene expression profiling of skin samples from AE patients with positive patch-test reactions to *Malassezia* revealed the reciprocal differential expression of pro-inflammatory genes and lipid-metabolism genes that resembles the observations in lesional AE skin, supporting the triggering role of *Malassezia* in the disease[139].

The diagnosis of AE is based on clinical criteria and the SCORAD (Scoring atopic dermatitis) system is used in research studies to assess and monitor the disease as objectively as possible[140]. The skin lesions have a typical age-dependent distribution with facial, scalp, and extensor involvement in infants and young children, and predominant flexural involvement in older children and adults. Pruritus is a hallmark of the disease. Acute lesions are characterized by pruritic papules with erythema, excoriations, and serous exudate, whereas chronic AE is characterized by areas of lichenification and fibrotic nodules, often accompanied by acute lesions. The disease severity varies between patients and from time to time in the same person. Most AE patients have mild to moderate disease; but there is a subgroup of patients that develop severe symptoms. In those the flare-ups can last several weeks or more, and cover many areas of skin. Even worse, some of the patients with severe AE are refractory to current available treatments. A link with autoimmune mechanisms has been suggested in severe AE by the detection of serum specific IgE antibodies to proteins from keratinocytes and endothelial cells[141]. Some of these autoallergens have molecular mimicry with microbial compounds such as *Malassezia* allergens, as described for the IgE-mediated and T cell-mediated autoimmunity against manganese superoxide dismutase[142] and thioredoxin[143,144].

The first-line treatment of AE includes skin-hydration, topical corticosteroids, and topical calcineurin inhibitors (tacrolimus)[145]. Antihistamines are used to treat sleep

disturbances and antibiotics and antifungals to treat skin infections. Furthermore, allergen-specific immunotherapy has proved to be of use in difficult-to-manage AE patients with allergic sensitization[146]. Treatments targeting the immune dysregulation include the human monoclonal antibody against the alpha subunit of the IL-4 receptor (dupilumab) that blocks the functions of IL-4 and IL-13[147]. Up to date, no treatment is curative, therefore there is a great need to identify pathogenic events that can be modified by primary intervention or can provide clues on how to protect the skin-barrier from inflammation and extrinsic triggers.

1.3.1.1 CLA^+ T cells critical players in AE

Circulating T cells can infiltrate human skin as a mechanism of immunosurveillance. They are identified by the expression of the cutaneous lymphocyte-associated antigen (CLA), an inducible cell surface glycoprotein that has been extensively characterized as skin-homing receptor[148]. In healthy subjects, CLA is expressed by approximately 5 to 15% of peripheral blood T cells[149] of the memory subset $CD45RO^+$ but not in $CD45RA^+$ naïve T cells. Moreover, CLA is detected in almost all T cells infiltrating skin in benign and malignant conditions[150]. The ligand of CLA is E-selectin (CD62E) which is expressed on the endothelium of postcapillary venules at sites of acute inflammation as well as in chronically inflamed skin. The interaction between CLA and E-selectin allows the transmigration of CLA^+ T cells from the peripheral circulation to the dermis[151]. Previous studies have shown that AE patients have a higher percentage of circulating CLA^+ T cells compared with HC[152,153]. Also, that the T-cell receptor (TCR) repertoire in CLA^+ cells from AE patients recognize superantigens of *Staphylococcus aureus*, the *Malassezia sympodialis* allergen Mala s 13 (with cross-reactivity to human thioredoxin) and mite allergens[152,154-156].

The $CLA^+CD45RO^+$ T cells in the peripheral blood of AE patients induce IgE production in B cells and enhance eosinophil survival[157]. They also display features of *in vivo* activation like increased levels of CD25, CD40L and HLA-DR and spontaneous release of IL-13[158,159]. Circulating $CD4^+CLA^+$ T cells in patients with acute AE produce significantly higher levels of IL-4, IL-13 and tumor necrosis alpha compared to patients in the chronic phase of the disease or healthy controls[160]. CLA^+ T cells express the chemokine receptors CCR4 and CCR10, and are attracted to skin by CCL17, CCL22 and CCL27[161]. These chemokines are produced by keratinocytes, vascular endothelial cells, Langerhans DC and inflammatory dendritic epidermal cells[162]. After skin homing, the CLA^+ T cells form dermal infiltrates, secrete interferon gamma and make keratinocytes susceptible to apoptosis[163].

Cutaneous inflammation in AE lead to the production of chemokines that allow continuous migration of skin-reactive CLA^+ T cells from the peripheral blood to the inflamed lesions (Figure 8). Once in the skin, the CLA^+ T cells are activated by the encounter with their cognate allergens, auto-allergens[164] and superantigens. A portion of the CLA^+ T cells returns to the blood via the thoracic duct[165] and display features of recent activation and spontaneous production of cytokines. The fact that *S. aureus* superantigens up-regulate the expression of the CLA antigen on T cells[166] reflects that this population is directly influenced by environmental cues of the skin microbiome. CLA^+ T cells infiltrate skin under physiological conditions, but whether they harbor intrinsic molecular alterations in patients with AE remains to be evaluated.

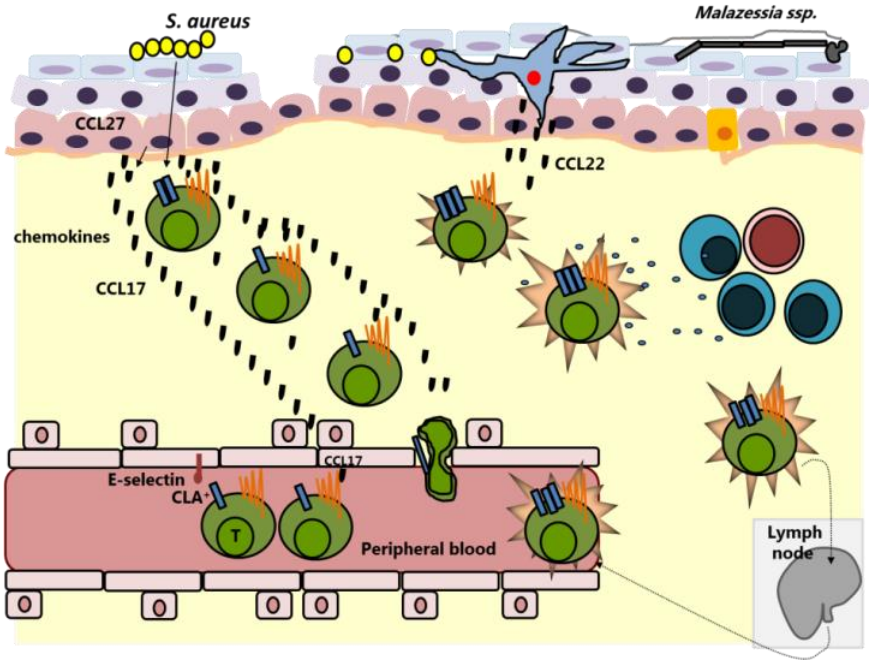

Figure 8
CLA^+ T cells (in green) recirculate between peripheral blood and skin.

1.3.2 Asthma

Asthma is a chronic inflammatory disorder in the airways where inflammation leads to airway hyper-responsiveness and recurrent episodes of wheezing, breathlessness, chest tightness and coughing. However, asthma is a syndrome and an umbrella term for many sub-phenotypes thought to originate from different molecular alterations[167]. It is a polygenic and multifactorial disease resulting of complex gene-gene and gene-environment interactions, but the exact initial mechanisms that trigger the pathological processes remain unclear. The genetic associations with genes implicated with immune pathways and the airway response to environmental exposures suggest that leukocytes and bronchial epithelial cells play a major role in the origin of this disease. Several initiatives have aimed to classify asthma based on the presence of allergic sensitization or atopy (e.g. allergic asthma vs. non-allergic asthma) or the response to treatments (e.g. controlled asthma vs. steroid-resistant asthma). Besides, great efforts in asthma research are oriented to discover biomarkers that can objectively help to better classify the different sub-phenotypes[168] or help to identify groups of patients with particular susceptibilities.

Asthma has become epidemic reaching 20 to 30% prevalence in UK, Australia, New Zealand or Brazil and according to the world health organization (WHO) it is estimated to affect 235 million people worldwide. Exacerbating factors include the exposure to tobacco smoke, allergens (*e.g.* from house dust mite, cat, dog, grass, pollen, molds, food, etc.), respiratory infections including viruses, exercise and hyperventilation, weather changes, sulfur dioxide, air pollution, drugs and additives. Early-onset asthma is more likely to be atopic and related to genetic risk factors while adult-onset asthma is more related to smoking and triggered by environmental exposures[169,170]. The symptoms can fluctuate over time, with periods of remission and relapse[171]. In some patients the disease can shift to another phenotype like allergic rhinitis or even disappear as "spontaneous cure". There are gender differences in the natural history of asthma; males are more susceptible to the disease during early childhood while females develop more symptoms after puberty and are prone to have late-onset asthma, some times in connection to obesity and hormonal changes.

In addition, there are parent-of-origin effects with the antecedent of maternal asthma having stronger effects than paternal asthma on the susceptibility to early onset phenotypes[170,172]. Studies in mice suggest that epigenetic mechanisms may explain the "maternal effect" in asthma, since the splenic $CD11c^+$ DCs of neonates born of allergen-sensitized mothers had substantial changes in DNA methylation compared to the offspring of non-sensitized mice, and even when they have not been in contact

with allergens, the "pro-allergenic DC" can induce allergic inflammation after adoptive transfer[173].

Nowadays the American Thoraxic Society (ATS) and the Global initiative for asthma (GINA) classify asthmatic patients depending on their symptoms and the ability of current medications to manage the disease. However, these classifications do not consider underlying mechanisms of pathogenesis. Most patients can be controlled by current available treatments, but 5 to 10% has the severe-uncontrolled form of the disease, non-sensitive to steroids[174,175]. The pharmacological treatment for the control of asthma include inhaled corticosteroids (ICS), leukotriene modifiers, long-acting inhaled β-agonists, and for some patients, systemic steroids, theophylline, cromones or the use of monoclonal anti-IgE antibodies. Medications for relieving the symptoms include rapid-acting inhaled β_2-agonists, systemic steroids and anticholinergics. Unfortunately, none of those medications are curative, and allergen-specific immunotherapy is the only that can modify the course of the disease in some patients with allergic sensitization[176]. Therefore, the discoveries of new therapeutical targets that can modify the natural history of asthma or prevent its inception are highly needed, in combination with diagnostic tools (such as biological markers) that allow a better definition of phenotypes and individualized therapy.

For the phenotype of early-onset asthma (childhood asthma), the current hypothesis on the pathogenesis indicates that patients harbor alterations in immune genes and defects in the epithelial/mesenchymal trophic unit in the lung[177]. In this scenario, intrinsic defects in the bronchial epithelium lead to a dysregulation in cytokine production upon exposure to environmental factors like viruses or airborne allergens, and this thereby allow extensive infiltration (non-IgE mediated) of immune inflammatory cells on the airways[178]. This occurs in the context of type 2 inflammation and bronchial eosinophilia (which is observed in the majority but not in all patients). The chronic stimuli of cytokines and other inflammatory factors produced by the epithelium and the cell infiltrate, lead to the differentiation of fibroblast to myofibroblast, bronchial smooth cell hyperplasia and the dysregulation of other cells in the bronchia, overall leading to bronchial remodeling and the perpetuation of the pathological mechanisms[179]. The damage of the epithelium also promotes IgE sensitization in individuals with atopic predisposition. The discovery of the molecular mechanisms underlying these changes will be helpful for managing this still puzzling disorder.

1.3.2.1 The asthma susceptibility gene *ORMDL3*

In 2007 a GWAS revealed a previously unknown region on chromosome 17q21 associated with childhood asthma[180]. This finding has been replicated by more than 20 independent studies conducted in populations with different genetic backgrounds. Some of the alleles conferring risk for asthma in the 17q21 locus have been associated with the mRNA levels of the underlying genes *ORMDL3* (ORMDL sphingolipid biosynthesis regulator 3) and *GSDMB* (gasdermin B). Alleles in the 17q21 locus may influence the outcome of environmental exposures, for instance, Smit *et al.*, described that the effect of bronchiolitis on the predisposition to childhood asthma is higher in carriers of the risk genotypes in the 17q21 locus[181]. The linkage disequilibrium (LD) within this region is very high and the causative gene explaining the association with asthma remains unclear. Epigenetic mechanisms were proposed to participate by the observations that asthma-associated SNPs affect genotype dependent DNA-protein interactions, nucleosome positioning and insulator binding[182]. Also, by studies in human lymphoblastoid cell lines showing that methylation levels in *GSDMB* differ depending on the gene haplotypes[183]. Among the five genes narrowed to the asthma-susceptibility region in the 17q21 locus, *ORMDL3* has been considered the most promising functional candidate. The gene spans ~6.5 kb and is transcribed from the minus strand in two known transcript variants. The human ORMDL3 belongs to a family of ubiquitously expressed transmembrane proteins anchored to the endoplasmic reticulum (ER)[184], and the translated human isoforms contain 153 and 137 amino acids, respectively.

The connection of *ORMDL3* with inflammatory mechanisms was suggested by the genetic associations with other chronic inflammatory conditions like inflammatory bowel disease[185], ankylosing spondylitis[186] and rheumatoid arthritis[187]. Moreover, *in vitro* experiments demonstrated the functional connection of ORMDL3 in inflammatory pathways. Breslow *et al.*, described that ORM proteins are critical mediators of sphingolipid homeostasis and raised the hypothesis that sphingolipid dysregulation contribute to the development of childhood asthma[188]. ORMDL3 binds and suppress the serine palmitoyl transferase, a limiting enzyme in *de novo* synthesis of sphingolipids, which induce bronchial hyper-reactivity in murine models of asthma. An alternative mechanism on the function of ORMDL3 was presented by Cantero-Recasens *et al.*, which revealed that ORMDL3 regulates calcium (Ca^{2+}) homeostasis in the ER[189] and facilitates the unfolded protein response[185]. The experiments by Carreras-Sureda *et al.*, confirmed that ORMDL3 modify calcium signaling in T-cells[190]. These findings are particularly interesting because T cell activation rely on depletion of Ca^{2+} from the ER and the generation of cytoplasmic Ca^{2+} signals that allow the translocation of transcription factors[191].

1.4 Epigenetics and environmental exposures

Exogenous environmental exposures have been found to change the epigenome at DNA methylation, miRNA and histone levels. These exposures include pollutants (*e.g.*, metals, polycyclic aromatic hydrocarbons, dioxin-like chemicals, polychlorinated biphenyls, phthalates and pesticides), tobacco smoke and pharmacological drugs, whereof many have been found to predispose to cancer and complex inflammatory diseases[192,193]. The first experimental indications that the environment induces changes in DNA methylation were obtained in the agouti mouse, in which the feeding of pregnant mothers with a high methyl diet was able to change the color of the fur from yellow to wild-type agouti in the pups[194]. Since the establishment of DNA methylation marks requires the availability of components like methionine, vitamin D, vitamins B6, B12, and folate, nutrition is one of the critical environmental exposures shaping the epigenome[195,196]. The hypothesis on the developmental origins of health and disease is based on the observations that early environmental exposures can induce epigenetic modifications that will influence traits later in life, including metabolic responses[197]. Furthermore, air pollution is among the best characterized factors associated to DNA methylation changes in humans (including particulate matter $PM_{2.5}$ and PM_{10}), as well as metals including arsenic[198]. DNA methylation changes induced by these contaminants have been detected in LINE-1 and Alu sequences as well as in coding genes, particularly related to the synthesis of nitric oxide and the response to oxidative stress[198]. Drugs like hydralazine or procainamide induce SLE in predisposed individuals by mechanisms that involve inhibition of DNMT1 and decreased DNMT1 levels, respectively[199]. In addition, the exposure of pregnant rats to the endocrine disruptor vinclozolin during the period of sex determination induced spermatogenic defects and other abnormalities that were even detectable in the F3 and F4 generations[200].

In the last years, many other factors have been associated with differences in DNA methylation including stressful events[201,202] and season of conception[203]. The exposure to allergens has been associated to changes in DNA methylation and miRNA expression in mice models after sensitization and challenge with house dust mite extracts[204]. Indeed, mice exposed to a combination of diesel exhaust particles and allergens of *Aspergillus fumigatus* shown the highest levels of total IgE and pronounced methylation differences in *infg* and *il4*[205]. As I will discuss later in the Results and Discussion section, the simultaneous effects of multiple environmental exposures on the epigenome are the clues and the challenge when conducting epigenetic studies in humans. Among these, the relationship between the commensal microbes that reside on the epithelial surfaces and epigenetic changes in human cells

is one of the most enigmatic. The gut microbiota can produce metabolites that induce epigenetic changes in the colonic epithelium[206]. These observations open the possibility that microorganisms on the skin and in the bronchial airways may induce epigenetic modifications in cells located in those tissues, but this remains to be demonstrated.

2 Aims

The overall aim of this thesis was to investigate the contribution of DNA methylation and miRNA expression on leukocyte biology and to identify differences between allergic patients and healthy controls that could reveal new insights on pathophysiology, disease biomarkers and therapies.

The specific aims were:

Study I to test the hypothesis that peripheral blood leukocyte populations have different DNA methylation signatures associated to cell lineage.

Study II to compare DNA methylation and miRNA levels in peripheral blood lymphocytes of atopic eczema (AE) patients and healthy controls.

Study III to explore the biological interplay between genotypes, DNA methylation and mRNA expression in the asthma susceptibility genes *GSDMB/ORMDL3* between asthmatic children and healthy controls.

Study IV to analyze longitudinal trends in DNA methylation levels that occur in peripheral blood leukocytes between 3 and 60 months after birth in healthy children.

3 Materials and Methods

This section summarizes the samples and methodologies applied to achieve the specific aims of this thesis. Detailed information can be found in the papers I to IV.

Study populations and blood samples

MALF Study	Case-control dataset of adult AE patients and HC from the Stockholm area. Six male HC (age 38±13.6 years) provided a 450 ml blood donation to sort seven leukocyte populations (Paper I and III). Ten AE patients and 10 HC, age and sex-matched, were recruited for a cross-sectional comparison of DNA methylation and miRNA levels in peripheral blood T cells in connection with mRNA expression levels (Paper II).
BAMSE	Prospective birth-cohort study in Stockholm initiated in 1994 with available DNA samples extracted from whole blood at 8 years of age for genetic association studies (n=1953) and bisulfite treated DNA samples for methylation analyses by EpiTYPER® (n=291), and pyrosequencing (n=383), (Paper III).
Swedish Search	Multicenter, nationwide, cross-sectional study on children (age 12.6±3.9 years) with problematic uncontrolled severe asthma, age-matched peers with controlled persistent asthma, and healthy controls. A subset of 46 children with DNA and RNA samples from whole blood, as well as cell counts at time of sampling was studied for DNA methylation and mRNA levels in the 17q21 locus (Paper III).
DIPP	Population-based prospective follow-up of Finnish children to evaluate risk factors for Type I diabetes. DNA samples were extracted from whole blood in ten healthy girls within seven time points after birth (3, 6, 12, 24, 36, 48, and 60 months) for measuring longitudinal DNA methylation levels (Paper IV).

Experimental procedures

Ficoll separation	Starting from a 450 ml blood donation, the blood was diluted 1x in sterile PBS. PBMCs and granulocyte fractions were separated by density gradient centrifugation using Ficoll-Paque Plus™ (Paper I and II).
Cell sorting	Different cell populations from PBMCs and granulocytes were labelled with monoclonal antibodies conjugated to microbeads for positive and negative selection by magnetic associated cell sorting using commercially available kits from Miltenyi Biotec (Paper I and II).

Flow cytometry	Phenotype of harvested cells after sorting was verified by staining with fluorescently labelled antibody panels and expression of cell surface markers evaluated in a FACS calibur (Paper I and II).
DNA extraction	DNA extractions were performed by using commercially available kits: QIAmp DNA micro kit (Paper I, II) and FlexiGene kit (Paper IV). Samples included in paper III were extracted at the Karolisnka Institutet biobank using standard methods.
RNA extraction	To preserve miRNAs the RNA isolation was performed by the acid guanidinium thiocyanate-phenol-chloroform method (Paper II). For the mRNA measurements presented in Paper III, RNA was previously isolated by using the RiboPure™ Blood extraction kit[207].
Bisulfite treatment	Treatment of genomic DNA with bisulfite that convert the unmethylated cytosines to uracil but leave the methylated cytosines unaffected. For all the samples presented in this thesis we used the EZ-96 DNA methylation kit from Zymo Research Corporation (Paper I-IV).
HumanMethylation450 BeadChip	Microarray platform from Illumina covering 485 755 CpG sites distributed in 99% of RefSeq genes. It detects the methylation level at a single base resolution. The 50 bp infinium probe query a C/T polymorphism created by bisulfte conversion of unmethylated cytosines (Paper I-IV).
Agilent 8 x 60K array	Microarray platform from Agilent technologies with 4774 probes covering 2006 human miRNAs (as miRbase v19). The protocol requires 100 ng of total RNA as input material and uses Cyanine3-pCp labelling chemistry and hybridization on microarray slides (Paper II).
STRT sequencing method	Genome profilling of transcriptional initiation at the 5' end of transcripts with poly A+ tail. The protocol is a modification of the single-cell tagged reverse transcription (STRT) method starting from 10 ng of total RNA as input (Paper II).
Quantitative PCR (qPCR)	To investigate gene expression levels of miRNA (Paper II) and mRNAs (Paper II and III); miRNA analysis were done by using the miRCURY LNA™ Universal RT microRNA PCR kit (Exiqon). For mRNA targets we used validated Taqman probes with beta 2 microglobulin (*B2M*) (Paper II) and cyclophilin A (*PPIA*) as endogenous controls (Paper III).
Sequenom EpiTYPER®	Quantitative analysis of DNA methylation levels of targeted regions in *GSDMB* and *ORMDL3*. Primers were designed in Epidesigner and *in silico* evaluated using MassArray (Paper III).

Pyrosequencing	DNA sequencing method based on the principle of "sequencing by synthesis" and the detection of pyrophosphate release on nucleotide incorporation. Since it uses a pre-defined sequence, this method was used to measure DNA methylation levels in CpG sites that coincide with polymorphisms (CpG site SNPs, Paper III).

Bioinformatics

Quantile normalization	Non-parametric method to normalize a matrix of probe level intensities based on the concept of quantile-quantile plot extended to n dimensions. This compensate for systematic technical differences between microarrays by forcing the intensities of all probes in one standard distribution (Paper I and IV).
SWAN-normalization	Subset-quantile within array normalization for Illumina Infinium HumanMethylation450 Bead Chip that allow the Infinium type I and type II probes within a single array to be normalized together and substantially reduces the differences in beta value (Paper II).
Batch effect correction	Method to combine data from different microarrays and adjust the data by the non-biological experimental variation between arrays batches and other covariates. The tool ComBat was used (Paper II).

Statistics

Moderate t-statistics	Ratio of the $\log_2$ expression/methylation level to its standard error. The standard errors are moderated across genes borrowing information from the ensemble of genes and aid with the inference about individual genes (Paper II and IV).
Bayes contrasts	Empirical Bayes shrinkage of the standard errors towards a common value (Paper I, II, and IV).
Causal inference tests	Series of conditional association analysis that implement a formal statistical hypothesis test, resulting in a p-value. The data include genotype, possible causal mediator (e.g. methylation) and one outcome of interest (e.g. mRNA expression), see Paper III.
Multiple testing corrections	Re-calculate probabilities obtained from a statistical test which was repeated multiple times. The correction by Benjamini-Hochberg (also known as q-value) and the Bonferroni correction were used (Paper II, III, and IV).

4 Results and Discussion

This section summarizes the main findings of this PhD thesis. Detailed information can be found in the papers I to IV.

4.1 Discovering the genomewide DNA methylation landscapes of sorted human blood leukocytes (Paper I)

The primary goal of this study was to identify differentially methylated CpG sites according to cell-lineage in sorted leukocytes from human peripheral blood. A total of 60 DNA samples were isolated, including unfractionated whole blood, PBMCs, granulocytes and seven sorted cell populations from six adult healthy male blood donors. Genomewide DNA methylation levels were measured by the HumanMethylation450 BeadChip from Illumina. This study showed, for the first time, that the hierarchical clustering of blood leukocytes based on their global DNA methylation levels clearly recapitulates their hematopoietic origin, supporting that each population of terminally differentiated leukocytes have a particular DNA methylation landscape that can discriminate their lineage. Indeed, the DNA methylation differences were more prominent among distinct cell types from the same individual than within the same cell-type from different donors. The DNA methylation marks acquired during hematopoiesis are typical for each leukocyte lineage, and as expected, the closer the relation between cells, the lower is the number of differentially methylated CpG sites (Figure 9).

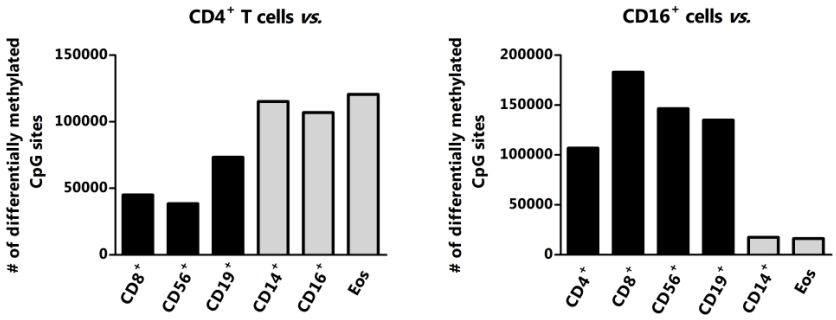

Figure 9
Total numbers of differentially methylated CpG sites between $CD4^+$ T cells *vs.* cytotoxic T cells ($CD8^+$), NK cell ($CD56^+$), B cells ($CD19^+$), monocytes ($CD14^+$), neutrophils ($CD16^+$), and eosinophils (Eos) (left panel), and between neutrophils ($CD16^+$) and the other sorted populations (right panel). Black bars indicate populations of lymphoid origin and grey bars indicate populations of myeloid origin.

It is worth to highlight that $CD19^+$ B cells have a very particular DNA methylation landscape that separate them from other lymphoid cells on the second principal component. The discovery of the differentially methylated CpG sites among these sorted blood populations is of great utility to understand which pathways and genes may be of functional relevance in terminally differentiated leukocytes under physiological conditions. These loci deserve to be further evaluated by *in vitro* experiments to explore their functional effects on leukocyte functions and may be also of interest in transdifferentiation experiments.

Moreover, the results of this study indicated that most of the differentially methylated CpG sites between the sorted populations and whole blood (unmethylated and methylated) were located in intragenic regions (gene bodies) and were not associated to CGI. These observations are in agreement with those by Irizarry *et al* [26] and Deaton *et al.*[32], and strongly support that the mechanisms implicating DNA methylation in cell differentiation involve other genomic regions than promoter-associated CGI.

In agreement with previous observations in hematopoietic progenitors[78] this study revealed remarkable differences in the methylation status of the lineage-related CpG sites depending if the populations have a lymphoid or myeloid origin. For instance, approximately 85% of the CpG sites with significant methylation differences between $CD16^+$ neutrophils and whole blood were unmethylated, while in $CD4^+$ T cells and $CD8^+$ T cells the percentages of unmethylated CpG sites with differences compared to whole blood were 9.4% and 6.8%, respectively. Conversely, the majority of CpG sites with significant differences between $CD4^+$ T cells and whole blood were methylated (Figure 10a).

Our observations on a demethylated genome in human neutrophils compared to whole blood, were confirmed by Rönnerblad *et al.*, who described that differentially methylated CpG sites generally show decreased methylation during granulopoiesis[208]. Recent studies in sorted populations from healthy individuals indicate that hypomethylated regions in neutrophils have effects on gene expression[209]. The hypomethylated genome of neutrophils may reflect a lower requirement of the DNA methylation machinery in connection with a high cell turnover. A more stochastic drift may be expected in neutrophils compared to the DNA methylation landscape of lymphocytes.

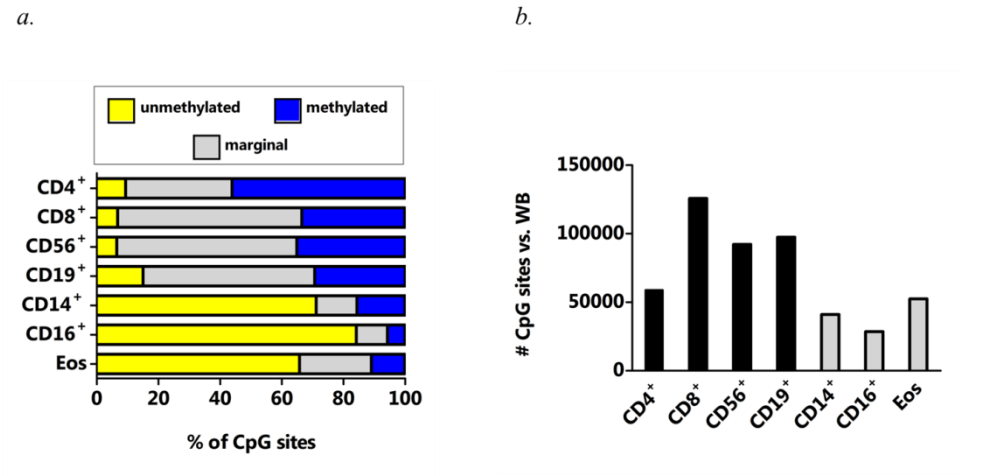

Figure 10

a. The distribution of the differentially methylated CpG sites between sorted populations and whole blood based on their methylations status (unmethylated, marginal or methylated).

b. Total numbers of differentially methylated CpG sites based on the comparison of M-values between whole blood (WB) and T helper cells ($CD4^+$), cytotoxic T cells ($CD8^+$), NK cell ($CD56^+$), B cells ($CD19^+$), monocytes ($CD14^+$), neutrophils ($CD16^+$), and eosinophils (Eos).

The discovery of differentially methylated CpG sites according to cell lineage in sorted populations of terminally diferentiated blood leukocytes also motivated the development of algorithms to correct for differences in cell composition when whole blood is used as the source of DNA[210,211]. Especially because for some cell types like $CD8^+$ T cells or $CD19^+$ B cells the number of differentially methylated CpG sites compared to whole blood coincide with ~100 000 CpG sites in the assay (Figure 10b). Our results created awareness in the research community[212] on the bias that cell heterogeneity can cause in DNA methylation studies and the need to sort more defined cell types[213]. This is particularly important as many research centers around the world have readily available samples to profile DNA methylation levels between patients and controls, but most of them had isolated the DNA from whole blood. The effect of cell heterogeneity may be even more prominent when studying complex inflammatory diseases that change the blood differential cell counts like SLE[214]. The Houseman algorithm uses the methylation profiles of lineage specific CpG sites to calibrate a regression model that deconvolute the observed methylation levels based on the proportion of cell types. However, this approach requires advanced knowledge of the dominant cell types in the sample and their methylation profiles to seed the algorithm[210]. For some tissues like placenta, adipose or tumor tissues the relevant underlying cell types may be unknwon, therefore the field is developing towards

reference-free-cell mixture adjustements[215]. Other approaches like FaST-LMM-EWASher, apply advanced statistical modelling to correct for cell-type composition without the need of pure cell profiles; however, it removes large-scale methylation changes from the analysis assuming they are caused by cell heterogeneity and therefore may not be suitable for studying cancer or methylation differences between tissues[216]. In summary this study provided important insights on the DNA methylation landscapes that are characteristic of each leukocyte population in human blood, and provided a catalog of loci containing differentially methylated CpG sites that has been of great value for the development of approaches that correct for differential cell composition in genomewide DNA methylation studies. Future research should be warranted to elucidate the effects of these differentially methylated CpG sites on gene expression and leukocyte functions.

4.2 Genomewide analysis of DNA methylation and miRNA levels in sorted T cells of AE patients (Paper II)

Studies have reported that differences in DNA methylation[217-219] and miRNA expression[220-222] might play a role in the pathophysiology of AE. However, no study has yet conducted a comparative multi-omic profiling of genomewide DNA methylation, miRNA expression and mRNA levels in AE patients and HC. Based on the results of Paper I, it became evident that such study requires the use of sorted cell populations, at least in a discovery phase. We here tested the hypothesis that epigenetic changes associated to AE can be detected in peripheral blood lymphocytes. Four populations of T cells ($CD4^+$, $CD4^+CD45RA^+$ naïve, $CD4^+CLA^+$ and $CD8^+$) were magnetically sorted from whole blood of 10 AE patients and 10 age-matched HC. DNA methylation levels were measured using the HumanMethylation450 BeadChip. The global expression of approximately 2000 miRNAs was measured by a commercial microarray platform from Agilent Technologies. For the first time, this study revealed that circulating skin-homing memory $CD4^+CLA^+$ T cells isolated from AE patients had significant DNA methylation differences in 49 CpG sites and differential expression of 16 miRNAs compared to HC.

Among the loci harboring the differentially methylated CpG sites between AE patients and HC we found the genes encoding for interleukin 13 (*IL13*) and the alpha chain of the IL-10 receptor (*IL10RA*). However, for the majority of the CpG sites the DNA methylation differences between AE patients and HC were small (about 1.5 fold). Moreover, none of the genes containing differentially methylated CpG sites showed significant differences in mRNA expression; with the exception of *IL13* that

was found upregulated in the $CD4^+CLA^+$ T cells of AE patients by qPCR but not by STRT. The decreased DNA methylation levels in the promoter of *IL13* detected in this study may be related with the IL-13 dominated Th2 pattern of the CLA^+ T cells that has been previously reported in AE patients[157,159,160]. Another remarkable finding was the identification of differentially methylated CpG sites in several genes encoding for enzymes involved in carbohydrate metabolism and with functional enrichment in the Golgi membrane. A question that remains to be explored is which is the functional outcome (if any) of these small DNA methylation differences and their relationships with the disease. The fact that some of these genes are know to be co-expressed (like *IL10RA* and *SH2B3*) and that many belong to similar biological pathways suggest that these DNA methylation differences might not be spurious observations. As I mentioned in section 1.3.1.1, the $CD4^+CLA^+$ T lymphocytes sorted from blood are memory cells with the possibility to recirculate between the skin, the lymph nodes and the peripheral blood, therefore, it is feasible that the DNA methylation differences detected in this study, are just the reflection of the activation status in the circulating cells of AE patients. Under this hypothesis, the differentially methylated CpG sites in the $CD4^+CLA^+$ cells may be remaining markers, a vestige, of a series of molecular events that may have already happened within the cells by the time they received their priming in the lymph nodes or when contacted by their cognate allergens/antigens in the skin.

This notion is suggested by the fact that the overlap between the differentially methylated CpG sites detected in $CD4^+CLA^+$ T cells and the other 3 lymphocyte populations (especially $CD4^+CD45RO^+$ naïve T cells) was very small. Indeed, some DNA methylation differences between AE patients and HC may occur earlier in the lymphocyte development, as suggested by the detection of three loci that are differentially methylated between AE patients and HC but only in $CD4^+CD45^+$ naïve T cells; as well as by the detection of one CpG site located in the gene paraoxonase 1 (*PON1*) that was significantly more methylated in all four T cell populations of AE patients compared to HC.

The comparative analysis of miRNA expression revealed that only the $CD4^+CLA^+$ T cells contained differentially expressed miRNAs between AE patients and HC. Ten were up-regulated and six were downregulated in the patients. We confirmed significant differential expression between AE patients and HC for miR-21-3p, miR-130b-3p, miR-150-5p, and miR-1275 by qPCR (although only eigth out of sixteen differentially expressed miRNAs were tested). The miRNAs were predicted to target some of the differentially methylated genes and some of the differentially expressed genes including *SOD2* encoding for superoxide dismutase 2 (mitocondrial), *PRDX3*

encoding for peroxiredoxin 3 and *HMGB2* encoding for high mobility group box 2. Gene ontology (GO) analysis with the combined list of differentially methylated genes, miRNA targets and differentially expressed genes revealed a significant enrichment in the biological processes of "immune response", "response to external biotic stimulus" and "response to molecules of bacterial origin". One possible interpretation for these observations is that differential methylation, miRNA and mRNA expression in the $CD4^+CLA^+$ T cells of AE patients are the outcome of environmental cues that are associated to the disease, namely, the exposure of AE patients to skin-associated micro-organisms in the context of skin-barrier defects. The results of this study revealed that epigenetic marks associated to AE can be detected in circulating T cells. Studies on lesional and non-lesional skin biopsies obtained from the AE patients and HC are ongoing to visualize the $CD4^+CLA^+$ T cell infiltrate and to evaluate if the differentially expressed miRNAs detected in circulating cells can be also detected by *in situ* hybridization. Moreover, the differentially expressed miRNAs are being measured in plasma samples from this dataset to investigate if they can be detected in circulation. The integrative analyses presented in Paper II are just an initial attempt to evaluate the relationship between the three data layers but additional pipelines are needed for more comprehensive data integration.

In conclusion, this study revealed for the first time that $CD4^+CLA^+$ T cells of AE patients contain cell-specific DNA methylation differences in genes related to Th2 cytokines, intracellular signaling and carbohydrate metabolism. Moreover, 16 miRNAs are differentially expressed in $CD4^+CLA^+$ T cells of AE patients compared to HC, and predicted to target differentially expressed genes at the mRNA level. The miRNA/mRNA regulatory loops are now under study.

4.3 DNA methylation levels in the *GSDMB/ORMDL3* locus between asthmatic patients and healthy controls (Paper III)

The *ORMDL3* gene is one of the most replicated susceptibility locus for childhood asthma[223], but its DNA methylation landscape in primary human leukocytes was unknown. The aim of this study was to compare the DNA methylation levels of CpG sites in *ORMDL3*, and the nearby gene *GSDMB*, between asthmatic children and healthy controls. The results revealed that some of the most replicated asthma-associated SNPs (i.e. rs7216389, rs4065275 and rs12603332) coincide with CpG sites and given the strong linkage disequilibrium in this locus, the carrier status of the polymorphic alleles configures different methylation patterns in these CpG-site SNPs. The children carrying the genotype combination confering increased risk for

asthma (rs7216389-TT/rs4065275-GG/rs12603332-CC) had 5-6% less methylation in two non-polymophic CpG sites located in the 5'UTR of *ORMDL3*, suggesting that asthma-associated CpG-site SNPs do not only affect the methylation status in the site they coincide with, but are methQTL for other CpG sites in this region. Although it was not possible to point out a causative variant for the genetic association with asthma or the association with DNA methylation levels, this phase of the study indicated that the 5'UTR of *ORMDL3* may contain the genomic interval with potential differences in methylation related to the asthma phenotype. It is worth to highlight that DNA methylation levels were measured in DNA samples extracted from whole blood when the children were 8 years of age, but the majority of participants in the patient group had received a diagnosis of asthma by age 4 years. The BAMSE birth-cohort was not conceived for this type of epigenetic studies and sorted cell populations or cell counts were not available for this dataset.

Therefore, to obtain a broader view of the DNA methylation landscape of *ORMDL3*, we studied DNA methylation in combination with genotypes and gene expression in 46 children (15 HC and 31 asthmatics) from the Swedish Search Study. We found significant differences in the DNA methylation levels of five CpG sites between asthmatic children and healthy controls, that remained significant after adjusting by cell counts at the time of blood sampling. Three of these CpG sites also correlated with the mRNA levels of *ORMDL3*. Interestingly, the DNA methylation differences were specific between the phenotype of controlled-persistent asthma and HC, but not with the phenotype of severe asthma. The DNA methylation differences albeit statistically significant after adjustement by cell counts and genotype, were in the range of 3 to 6% between patients and controls. Given the small sample size and the fact that these measurements were done in whole blood it is feasible that stratification during the regression analysis may have resulted in type I error. Nevertheless, we found in a parallel sample, analyzed for mRNA expression, that asthmatic patients with the phenotype of controlled persistent asthma but not severe asthma have a significant up-regulation in the mRNA levels of *ORMDL3*. These findings in the Swedish Search Study strongly indicates that methylation in *ORMDL3* could be implicated in the pathophysiology of this phenotype, and are in line with previous studies suggesting that *ORMDL3* is a susceptibility gene that can be predictive of persistent wheezing phenotypes[224]. Further studies performed in this dataset by our research group, also support that patients with severe uncontrolled asthma may be affected by other susceptibility loci[207]. Furthermore, causal inference tests revealed that both, SNPs and the DNA methylation in non-polymorphic CpG sites, have independent effects on *ORMDL3* mRNA expression, but the effects of the SNPs are stronger than CpG sites. Another prominent finding in this study was the discovery

of correlation between distant CpG sites that follow a pattern of "linkage" that recapitulates the inverse pattern of gene expression at this locus[182]. The finding of correlation between CpG sites has been also observed by Liu *et al* at the genomewide level[225].

Since the samples used for the comparison between asthmatic children and HC were obtained from whole blood, the data obtained in Paper I was used to compare the DNA methylation patterns observed in BAMSE and Swedish Search with that in sorted cells. Overall, the analysis of sorted populations confirmed that *GSDMB* is mainly methylated and that *ORMDL3* contains a CGI that is homogeneously unmethylated in all cell types. Moreover, we discovered that $CD8^+$ T cells contain a differentially methylated region in the 5'UTR of *ORMDL3* that is less methylated than in other leukocyte subtypes. Indeed, the two non-polymorphic CpG sites that are associated with the carrier status of the risk alleles for asthma are embedded in this $CD8^+$ T cell differentially methylated region. These findings suggest that DNA methylation differences in *ORMDL3* might have an impact on T-cells and promote lymphocyte driven inflammation. The asthma susceptibility alleles may promote increased mRNA expression and DNA methylation changes in the $CD8^+$ T cells of the carriers, making those individuals more susceptible to a dysregulated inflammatory response by $CD8^+$ T cells upon viral infections. This finding provides insights that could explain the previously observed association between *ORMDL3* with asthma only in the context of personal antecedents of rhinovirus wheezing illness[226], as well as with early-onset phenotypes[227].

4.4 On the dynamics of DNA methylation in human blood leukocytes between 3 and 60 months after birth (Paper IV)

Prospective longitudinal analyses on the dynamics of DNA methylation in early childhood are essential to identify genes that might be epigenetically modified during this period of life, and if disturbed, might contribute to the susceptibility to complex inflammatory diseases, including allergic phenotypes. Paper IV was a pilot genomewide DNA methylation analysis in whole blood, aiming to elucidate which genes have longuitudinal changes in DNA methylation between 3 and 60 months after birth in healthy children. A total of 60 serial DNA samples distributed within seven time points (3, 6, 12, 24, 36, 48 and 60 months after birth) were included in the analysis. Based on the results of paper I, and to avoid as much as possible the confounding effects of differences in cell composition, all the CpG sites showing significant differences due to age (n=853) were filtered against a list of CpG sites with DNA methylation differences in sorted leukocytes, that serve as cell type

classifiers. The present study revealed a catalog of 794 age-modified CpG sites that robustly reflect the changes in DNA methylation levels that occur in human blood leukocytes within 3 to 60 months after birth. Of these, 330 CpG sites (41.6%) are age-methylated and 464 CpG sites (58.4%) are age-demethylated.

Notably, we found that the genomic location of age-modified CpG sites differ depending whether the CpG sites become age-methylated or age-demethylated. Age-methylated CpG sites are more likely found in gene bodies and within +5 to +50 kb downstream of the transcriptional start site (TSS) compared to age-demethylated CpG sites, which are more frequently within -5kb and +5 kb from the TSS and are associated to promoters and DNAse hypersensitivity sites (DHS). The results of the gene ontology analysis made evident that genes containing age-methyated CpG sites are significantly over-represented in developmental processes and morphogenesis of anatomical structures. This is expected because after lineage choice the peripheral blood leukocytes may not require these developmental genes. Still, it was interesting that among all the systems, age-methylated CpG sites are significantly over-represented in genes related to neuronal functions and the formation of neuronal structures. Genes harboring age-demethylated CpG sites in blood leukocytes are over-represented in immune response genes reflecting the immunological window. Indeed, the DNA methylation levels of some genes that have been previously associated with chronic inflammatory diseases, like the interferon regulatory factor 5 (*IRF5*), the nucleotide-binding oligomerization domain-containing protein 2 (*NOD2*) and the prostaglandin E receptor 4 (*PTGER4*) are modified during childhood. The expression data from the FANTOM5 project[228] revealed that many of the genes containing age-modified CpG sites are expressed in leukocyte populations. These findings open new perpectives on the possibility to use environmental exposures to modulate the dynamics of DNA methylation on immune genes and to design future strategies for primary intervention in susceptible children.

Moreover, this study revealed for the first time the longitudinal changes in DNA methylation that occur in genes encoding chromatin remodelers like *SMARCD3* (SWI/SNF related, matrix associated, actin-dependent regulator of chromatin, subfamily D, member 3), in components related to the PcG proteins (including *CBX7*, *PHF1*, *RNF2*, *KDM2A*, *KDM2B* and *JARID2*) and histone deacetylases (*HDAC4*). We also identified a number of transcription factors that have longitudinal changes in DNA methylation during childhood and may be of relevance for the maturation of the immune response [229]. The results of this study suggest that age-related changes in DNA methylation during this period of life may not be only due to a stochastic drift but rather correspond to a program with potential relevance for

leukocyte biology. Moreover, the 794 age-modified CpG sites identified in this study may be very useful in developing tools to correct for age effects when performing DNA methylation studies in children.

4.5 Methodological considerations

The studies included in this doctoral thesis contain several methodological limitations that are discussed in each of the papers (I-IV). In the following lines I will comment their potential impact when interpreting the results and the scope of this work.

First, the number of samples included in the studies was small. Nevertheless, a sample size may be appropriate or not depending on the research question and the effect sizes. A total sample size of sixty samples from six healthy male individuals was enough for studying cell-lineage specific DNA methylation levels (Paper I). In this case, DNA methylation differences associated to leukocyte lineage differed more between cell types than between individuals. The clustering of cell types based on DNA methylation was later validated in 20 independent individuals in an ongoing study (data not shown). In Paper II and Paper III the aim was to compare DNA methylation levels between groups of patients and controls. In this case, it should be noted that the sample size was underpowered to draw convincing results, especially because the observed differences between groups are typically small. Due to the small number of participants, it is not possible to rule out type II error by lack of statistical power. Several logistic and methodological reasons allowed us only to study these numbers of individuals; nevertheless, the methodological design included a careful selection of participants based on the strategy of the extreme phenotypes, as an attempt to compensate the small sample size with including more homogeneous groups of patients.

In paper IV, a total of sixty samples were analyzed and this might have limited the power to detect age-associated changes in methylation. This is to our knowledge, however, the largest number of repeated DNA samples from the same individuals analyzed ever for longitudinal changes in genomewide DNA methylation during childhood. Thus, the availability of measurements over many time points compensates by the small number of children. In combination with a very astringent statistical model, the small sample size could explain why the list of age-modified CpG sites was shorter in this study compared to previous reports[230]. We replicated previously known age-modified loci in this dataset, like the differences in the AT rich interactive domain 3A (*ARID3A*) and discovered many new loci that require further verification. Given the small sample size and to avoid the confounding effects

that gender differences may introduce in the DNA methylation measurements, it was decided to include only males in Paper I and II and only girls in Paper IV. Second, it may appear counterintuitive based on the results of Paper I, but DNA extracted from whole blood was used in Paper III to compare DNA methylation between asthmatic children and controls, and in Paper IV to evaluate longitudinal changes in DNA methylation. The selection of those samples was based on availability in the framework of birth cohorts that were not originally intended for epigenetic studies but that provide the unique opportunity to use DNA and extensive clinical records. I think this is a common limitation for many ongoing studies worldwide that are using datasets from the GWAS era to study epigenetics. Adjustments by cell counts (Paper III) or verification of the differentially methylated CpG sites in the catalog of cell-lineage CpG sites (Paper III and Paper IV) were used as strategies to avoid confounding. It cannot be ruled out that some of the observations resulted of type I error by compositional differences in blood. This is particularly the case for the comparisons performed in BAMSE (Paper III) and in DIPP (Paper IV).

Another source of variation that must be discussed belongs to the methodologies that were implemented to access DNA methylation, miRNA expression and other biological readouts. DNA methylation analysis can be performed by numerous methods including methyl binding proteins, restriction enzymes or bisulfite sequencing. For genomewide screenings, the Human Methylation450 BeadChip from Illumina was the most robust and comprehensive microarray technology by the time that these analyses were done. However, even though it contains CpG sites distributed in 99% of the genes described in RefSeq, it is biased towards a higher number of CpG sites in the proximities of promoter-associated CGI. This may affect the possibilities to detect relevant DNA methylation differences in intragenic regions, especially when comparing methylation between cases and controls. Provided that there is a high degree of correlation between CpG sites spanning within the same regions, most probably the sites included in the array have allowed fishing regional signals. During the last years, a number of methods were evolving for analyzing 450K data in terms of pre-processing the raw data, normalization and analysis on differential methylation[231]. The different approaches implemented in Papers I-IV reflect some of these changes.

A substantial amount of work was invested in the technical validation of 450K measurements by other methods like EpiTYPER® and pyrosequencing, and many of the observations were consistent. In our hands, EpiTYPER® proved to have a very small standard deviation between duplicates in most of the CpG sites. However, it should be noted that the margin of technical variation for these methods (especially

pyrosequencing) may encompass the observed DNA methylation differences between cases and controls. The bisulfite treatment of DNA samples could also introduced sources of bias. In paper II, miRNA levels were measured by a microarray platform from Agilent Technologies. This was found to have the most consistent performance in miRNA detection over qPCR, sequencing, and other microarray platforms[232]. These results were consistent and some of the differentially expressed miRNA indeed validated by qPCR. The samples used for miRNA analysis were extracted using the acid guanidinium thiocyanate-phenol-chloroform method (without columns) and this procedure resulted to be very effective to obtain consistent amounts of miRNAs with broader representation of miRNAs species. Moreover, the extractions were done from normalized cell numbers, with a minimum of 2.5 million cells as input; therefore it is very unlikely that the miRNA analysis have been affected by low input cell numbers. Furthermore, the mRNA analyses were done using a modified STRT method which by sequencing the first 37 bp of the 5'-end of polyA-tailed RNAs in 10 ng total RNA provides an indication of transcription start sites and expression levels[233]. In contrast to microarray technologies, this sequencing protocol has a broader dynamic range, higher sensitivity and is less susceptible to cross-hybridizing molecules. Some of the mRNA expression differences between AE-patients and HC by STRT were validated by qPCR (Paper II).

5 Concluding remarks and future perspectives

Beyond the epigenetic mechanisms discussed in this thesis (i.e. DNA methylation and miRNAs), it becomes evident that we are just at the very beginning. The evidence accumulated until now supports that the complexity of the genome involves many other layers of regulation, that being studied under the current reductionist approaches can only provide a limited view of the cell biology. The technological advances in the coming years will be oriented to develop methodologies that allow the simultaneous analysis of diverse layers of epigenetic regulation with deeper resolution, and the integration of their biological readouts taking into account emergent processes. The genome research will be focused on more global mechanisms that can explain the unsolved questions by the conventional view that we have of the genes and their close regulators. This is especially important when the profiling of patients with complex inflammatory diseases and healthy controls repeatedly revealed small epigenetic differences between groups, within diverse loci, but without obvious biological readouts. Hopefully new analytical models will help to disentangle their combined effects under the principles of systemic biology. In addition, it can be anticipated that many more epigenetic mechanisms are to be discovered, as well as the functions of other chemical modifications on cytosines (like 5-hydroxymethilcytosine) and the role of the "transcribed-non-translated" genome.

The discoveries generated during the past four years in the framework of this thesis opened new theoretical insights on the mechanisms involved in leukocyte physiology and the pathogenesis of allergic diseases. One of the main conclusions is that the epigenetic modifications may not be the determining events leading to altered cell phenotypes (or "causative"), but rather components in a series of events that configure cell physiology and its altered status in the presence of pathology. The environment, assumed as the global network of signals that are sensed by cells, induce changes in the epigenetic landscape and is probably the key behind the modest differences that we detect in our samples. Future translational studies aiming to elucidate mechanisms in multifactorial diseases like asthma or AE should focus on disentangle the genomic changes that follow the interactions between the immune cells and their environment, to unravel the exposome. If we consider the epigenetic changes as a metabolic outcome of environmental effects, there is a unique opportunity to use them as objective measurements of more global processes.

Moreover, the implementation of longitudinal studies with extensive sampling before and after disease inception will be crucial to understand at the molecular level when

do human complex inflammatory diseases start, because once the symptoms are established it is very difficult to differentiate the epigenetic mechanisms that occur at early stages on the pathogenesis from those resulting of ongoing inflammation or treatments. In this sense, the evaluation of *in utero* exposures and even the sampling of the parents during pre-conception and pregnancy must be considered. Ongoing studies in the Assessment of Lifestyle and Allergic Disease During INfancy (ALADDIN) birth-cohort, including samples during pregnancy and delivery are targeting these questions[234]. But, how far in the developmental history of an individual can we trace an epigenetic mark? Lamarck and others[235,236] proposed the transgenerational inheritance of acquired characters. Only the future will reveal if some of the epigenetic marks involved in the pathogenesis of allergic inflammation can remain after the two steps of genome resetting and be transmitted by meiosis from parents to siblings.

Overall, our results support a transformation in the current view of the genome, traditionally observed as a static entity only modified by sequence variations, and studied with analytical tools that were developed based on the mechanisms leading to monogenic diseases. If we imagine the environment and the cell as a continuum with the epigenetic modifications as an interphase, a challenging question to be sorted in the future is: how to distinguish the basal noise from those epigenetic changes that result in changes of the cell phenotype? Or from epigenetic changes that contribute to human complex diseases?

The concepts discussed in this thesis, only known by experts in cell biology until few years ago, have permeated several disciplines of the biomedical sciences and created an impact in the strategies to study complex diseases. Provided that we and others have detected epigenetic modifications associated to asthma and AE, there are possibilities to test their use as biomarkers of sub-clinical phenotypes or be targeted by novel therapies. The discovery of significant DNA methylation differences in the subgroup of children with controlled persistent asthma but not in those with severe asthma (Paper III), suggested that molecules like ORMDL3 may be implicated in the pathogenesis of sub-phenotypes and therefore a critical candidate for personalized medicine. We also found several genes belonging to common biological pathways that harbor DNA methylation differences in patients with AE. The following step will be to identify critical, non-redundant bottlenecks within these pathways to design novel specific therapies or disease preventing strategies. We also found that DNA methylation in several immune genes and chromatin remodelers is dynamic within childhood (Paper IV). More studies are needed to elucidate if the disease-

associated epigenetic changes can be modulated by environmental exposures and if, hopefully, are amenable to be modified by primary interventions.

Last but not least, I would like to summarize some practicalities to consider when designing future epigenetic studies:

a) The ideal strategy to perform epigenetic studies should be to obtain purified cells from the organs affected by the disease under study. For instance, to investigate epigenetic mechanisms in AE, the profiling on sorted blood leukocytes can be correlated with the study of epigenetic modifications in skin biopsies of lesional and non lesional tissues. Studies investigating epigenetic modifications associated with asthma should assess the cellular elements in the BAL as well as bronchial biopsies. For instance, there is evidence that some epigenetic differences occurring in asthmatic patients can only be detected in cells from the lungs but not in PBMCs[237].

b) Even if the affected tissues are available, it should be kept in mind that a biopsy usually contains different cell types and by studying such tissue there is always the possibility that the signals get diluted between affected and non-affected cells. Therefore it is essential to apply appropriate methodologies to control for tissue heterogeneity. This is especially important when the disease-associated epigenetic differences are more likely to be restricted to particular cell types as it was found in Paper II.

c) The use of whole blood as the source of DNA should be carefully considered and must take into account the characteristics of the loci under study. Similarly as for tissue-derived samples, it is necessary to use appropriate algorithms to correct by cell composition when using DNA samples extracted from whole blood[210,211,213,216].

d) Some of the limitations imposed by the study of bulk of cells can be solved in the future by profiling of epigenetic signatures at the single cell level[238].

e) It is necessary to plan and collect the samples by using reagents and appropriate conservation and storage procedures that allow preserving the different epigenetic modifications in clinical settings. This is especially relevant for transitory or less stable marks as histone modifications or modifications in DNA bases apart of 5-methylcytosine.

f) It is very important to study primary cells, meaning that samples should be obtained directly from the individuals under study. The procedures to culture or to immortalize the cells alter their epigenetic profiles to a point in which they no longer

resemble the primary cell type and therefore are not a reliable source of DNA for epigenetic studies[239]

g) Since the epigenetic modifications are extremely susceptible of being modified by environmental factors, for instance, season of sample collection[72,203] all epigenetic studies should be conducted with appropriate pairing by age, gender, ethnicity and sociodemographic factors between cases and controls. It should be considered that optimal population based-studies on epigenetic variation may require new analytical strategies that differ to those implemented in genetic epidemiology[240].

Nowadays, several consortia like the roadmap epigenomics project (http://www.roadmapepigenomics.org/) and the international human epigenome consortium (http://www.ihec-epigenomes.org/) are mapping human epigenomes for normal and diseased cell types, thus many new discoveries regarding the role of DNA methylation in leukocyte function are to come within the next years.

6 Populärvetenskaplig sammanfattning*

Allergiska sjukdomar såsom atopiskt eksem och astma är kroniska, inflammatoriska sjukdomar där balansen mellan de mekanismer som reglerar immunsystemet påverkats samtidigt som det finns defekter i målorganen såsom huden vid atopiskt eksem eller epitelet och den glatta muskulaturen i luftrören vid astma. Ärftlighet i form av variationer i DNA-sekvensen i kombination med exponering för miljöfaktorer kan utlösa eller bidra till sjukdomen. Allergen från t.ex. björkpollen och jordnötter, är exempel på utlösande miljöfaktorer och virussjukdomar i lungorna kan vara bidragande. Interaktion mellan ärftlighet och miljö kan ge upphov till förändringar hos celler och molekyler som vi ännu inte förstår. Dessa förändringar kan skada cellens funktion och bidra till utveckling av allergiska symptom. Under de senaste fyrtio åren har förekomsten av allergiska sjukdomar ökat dramatiskt. Detta kan tyda på förändringar i livsstil och våra moderna miljöer som därmed påverkar balansen i vårt immunförsvar. Förbättrad hygien har medfört en mindre exponering för mikroorganismer och parasiter samt en minskning av den biologiska mångfalden. Några forskare menar att dessa miljöförändringar påverkar cellernas funktion i immunförsvaret möjligen via epigenetiska mekanismer. Exempel på epigenetiska förändringar är modifiering av DNA och kromatin. Dessa modifieringar kan påverka uttrycket av gener utan att förändra den underliggande DNA-sekvensen och överförs från cell till cell när dessa delas.

Målsättningen med denna avhandling har varit att studera eventuella skillnader i epigenetiska faktorer mellan allergiska patienter och friska personer samt DNA-metyleringsmönster i de olika vita blodkropparna för att öka kunskapen om sjukdomsmekanismer och därmed finna nya strategier för prevention och behandling.

I studie I fann vi att varje enskild studerad population av vita blodceller har karaktäristiska mönster av DNA-metylering som kan relateras till cellhärkomst. Lymfoida celler är mer metylerade än de myeloida. Studien visar att när man studerar DNA metylering i blodprover utan att isolera specifika celltyper kan resultaten påverkas mer av de cellspecifika skillnaderna i metyleringsmönster snarare än den sjukdom man studerar.

I studie II fann vi att en variant av T-celler, så kallade $CD4^{+}CLA^{+}$, skiljer sig epigenetiskt mellan patienter med atopiskt eksem och friska kontrollpersoner. Faktorerna inkluderade DNA-metylering och uttrycket av 16 mikroRNA. De gener

som var berörda av dessa skillnader är nu identifierade och kan i framtiden tjäna som nya markörer för att diagnostisera atopiskt eksem eller agera mål för nya behandlingar.

I studie III analyserade vi effekten av ärftliga variationer i DNA-sekvensen samt förändringar i DNA-metyleringen i kromosomregionen där bl a generna *GSDMB/ORMDL3* är lokaliserade, en region som kopplats till ökad risk för astma. Vi fann skillnader i DNA-metylering hos barn med lindrig, kontrollerad, ihållande astma jämfört med friska barn. De metyleringsnivåer i *ORMDL3* som visade skillnad hos de astmatiska barnen jämfördes med genuttrycket för generna i regionen. Genen som kodar för *ORMDL3* befanns mindre metylerad i cytotoxiska $CD8^+$ T-celler som skyddar oss mot virusinfektioner. Tillsammans indikerar dessa fynd att skillnader i DNA-metylering i genen som kodar för *ORMDL3* skulle kunna bidra till de mekanismer som gör att barn i tidig ålder lättare utvecklar astma efter virusinfektioner.

Allergiska sjukdomar kan både uppstå och försvinna under barndomen alternativt kvarstå upp i vuxen ålder. I studie IV studerade vi DNA-metyleringsnivåerna i vita blodkroppar över tid för att finna genregioner som uppvisar förändringar i tidig ålder, mellan 3 och 60 månader efter ett friskt barns födelse. Vi fann 330 DNA regioner som visar ökad metylering under denna tidsperiod och 464 regioner där metyleringen minskar. De gener som finns i dessa regioner påverkar bl.a. mognaden av immunsystemet och reglerar genomet.

Sammanfattningsvis visar dessa studier att DNA-metyleringsmönster skiljer sig mellan olika typer av vita blodkroppar och att det finns regioner i DNA som erhåller förändrade metyleringsnivåer under barndomen hos friska barn. Vi har också visat epigenetiska skillnader mellan friska personer och allergiska patienter med atopiskt eksem och barn med astma. Dessa fynd ger oss ledtrådar för att kunna utveckla nya och bättre biomarkörer och behandlingar för dessa sjukdomar.

*Översatt från spanska av Kesrtin Lagercrantz. Redigerad av Lovisa E. Reinius, Annika Scheynius och Cilla Söderhäll.

7 Resumen de divulgación científica

Las enfermedades alérgicas como el eczema atópico (EA) y el asma son enfermedades inflamatorias crónicas en las cuales los pacientes sufren de defectos en los mecanismos que regulan el sistema inmune, y al mismo tiempo, tienen defectos en células relacionadas con los organos afectados por la enfermedad, por ejemplo la piel en el caso del EA o células del epitelio o el musculo liso de los bronquios en el caso del asma. En la aparición de estas enfermedades participan, por una parte, la susceptibilidad genética que esta codificada por variaciones en la secuencia del ADN y por otra la exposición a factores ambientales que desencadenan la enfermedad, por ejemplo, la exposición a altas concentraciones de alergenos, infecciones virales, el contacto con el humo de cigarrillo, etc. De las interacciones entre los genes y el ambiente resultan alteraciones moleculares que dañan el funcionamiento celular y finalmente llevan a los síntomas alérgicos, los cuales todavía no están bien comprendidos. Durante las últimas cuatro décadas ocurrió un aumento en la prevalencia de enfermedades alérgicas en el mundo. Muchos factores se han asociado con este fenomeno, por ejemplo la contaminación ambiental, los cambios en el estilo de vida, el mejoramiento de la higiene que ha llevado a una menor exposición a bacterias y parásitos y la reducción de la biodiversidad. Algunas hipotesis sugieren que esos cambios ambientales afectan las funciones de las células del sistema inmune como los linfocitos, al parecer mediante mecanismos epigeneticos. Estos son modificaciones químicas en el ADN y la cromatina que pueden cambiar la expressión de los genes sin cambiar la secuencia de ADN y que se transmiten de célula a célula cuando estas se dividen. Hay evidencias para pensar que las modificaciones epigeneticas ayudan a entender las alteraciones genomicas que llevan a las enfermedades alérgicas e inflamatorias. El objetivo general de esta tesis doctoral fue elucidar las diferencias en la metilación del ADN que se relacionan con el linaje celular y la edad en linfocitos de sangre periferica; además comparar los niveles de metilación del ADN y de expression de microRNAs entre personas sanas y pacientes alérgicos.

En el estudio I se descubrió que cada población de leucocitos circulantes tiene patrones caracteristicos en la metilación del ADN relacionados con su linaje celular. Además, que las celulas de origen linfoide están más metiladas de las células de origen mieloide. Este estudio tambien reveló que cuando se hacen estudios de metilación usando muestras de sangre completa los resultados pueden afectarse por diferencias en la proporcion de las diferentes células.

En el estudio II se descubrió que una población de linfocitos T llamada $CD4^{+}CLA^{+}$ contiene diferencias significativas entre pacientes con eczema atópico y controles sanos, tanto en la metilación del ADN como en la expression de 16 microRNAs. Se identificaron los genes afectados por esas diferencias y algunos de ellos pudieran ser investigados en el futuro como nuevos marcadores de la enfermedad o candidatos para ser bloqueados por nuevos tratamientos.

En el estudio III se analizó el efecto de las variaciones en la secuencia de ADN que predisponen al asma y los cambios en la metilación del ADN en una región que confiere susceptibilidad al asma denominada *GSDMB/ORMDL3*. Se encontraron diferencias significativas en la metilación entre niños con asma controlada persistente y los controles sanos. Los niveles de metilación en los sitios CpG con diferencias en los niños asmaticos se correlacionaron con los niveles de expresión del gen *ORMDL3*. En este estudio tambien descubrimos que el gen *ORMDL3* esta menos metilado en los linfocitos citotoxicos $CD8^{+}$ generalmente asociados con la protección contra las infecciones virales. Estos hallazgos sugieren que las diferencias de metilación en el gen *ORMDL3* afectan principalmente este tipo celular, siendo uno uno de los mecanismos que predisponen al asma de inicio muy temprano e inducida por infecciones virales.

Una caracteristica interesante de las enfermedades alérgicas es que pueden aparecer y luego desaparecer durante la infancia o seguir un curso crónico. En el estudio IV se realizó un análisis preliminar de los niveles de metilación global en leucocitos circulantes con el fin de identificar los genes que tienen cambios en la metilación entre los 3 y los 60 meses después del nacimiento en niños sanos. Se encontraron 330 sitios CpG que aumentan su metilación durante ese periodo y 464 sitios CpG que la tienen disminuida. Los genes que contienen esos cambios estan relacionados con la maduración del sistema inmune y con muchas moleculas que remodelan el genoma.

En resumen, los estudios que componen esta tesis doctoral llevaron al descubrimiento de marcas en la metilación del ADN particulares en cada tipo de leucocitos en sangre periferica y además a la identificación de un catalogo de genes que cambian su metilación durante la infancia en las personas sanas. Además, revelaron diferencias entre personas sanas y pacientes alérgicos. Esos hallazgos proporcionan nuevas pistas para el desarrollo de ensayos que sirvan como biomarcadores de esas enfermedades asi como para el desarrollo de nuevos tratamientos.

8 Acknowledgements

The experiences of this doctoral thesis exceed all my expectations. I must said that the last five years have been fantastic as were the great people that I had the honor to work with. I am very grateful to Karolinska Institutet, the Swedish Foundation for Strategic Research and to all members of Epigene that made this possible, especially to:

Annika Scheynius, my main supervisor, for giving me the opportunity to join your laboratory and to work with you. I am a very fortunate student for having such dedicated, serious and responsible supervisor. Thank you very much for caring about my education, for teaching and guide me with patience, but overall for showing me that one should pursue one's goals in life doing good science, publishing, revisiting the hypothesis and keep questioning. I will never forget our scientific discussions and the great times that we spent planning, doing and interpreting the results of our studies. My eternal gratitude for supporting all my initiatives during these years.

Juha Kere, my co-supervisor, for inspiring me to look beyond the horizon and for always coming to our meetings with smart ideas and a smile. I admire your charm towards science and your goodness to everyone. Thanks for your trust, for hosting me at the JKE lab and for giving me the opportunity to get in contact with so many molecular techniques, high throughput methodologies and frontline research. The privilege to work with you and your team has left an everlasting imprint in my scientific development.

Cilla Söderhäll, my co-supervisor, thank you very much for your commitment and support with all my academic responsibilities, for your dedication to carefully read and correct the manuscripts, and the help with several methodological aspects of the projects and the sample sets; but overall, thanks for teaching me that good science can only be achieved with structured work and perseverance.

Lovisa E. Reinius thanks for your support during these years, especially for introducing me to R and the world of microarrays and scripts. I really enjoyed working with you. **Catharina Johansson**, you are fantastic, thanks for always being there to help and for your support with the recruitment of the participants in the MALF study. **Dario Greco**, for doing the bioinformatic calculations, but overall for the insightful discussions and your advices, keep doing great work as scientist. **Mauro D'Amato**, for your time and advices, and teaching me the toolbox from quantitative PCR to cell lines and luciferase assays. **Ville Pulkkinen**, thanks for great times studying the genetics of *RORA*; it was a fantastic experience to work and write

together. **Anna Andersson**, for your support at the lab and endless patience. **Sören Bruhn**, you have been a key player for the CLA and the ALADDIN project, thank you very much for your critical views on the projects and your hard work at the lab. **Erik Melén** and **Anna Gref**, for your support with the statistical analysis. **Göran Pershagen** and **Sven-Erik Dahlén,** for insightful discussions on the epigene seminars. **Helena Persson** for your advices with the miRNA isolations and analytical help with the FANTOM5 data. **Shintaro Katayama**, for your contributions with the analysis of the STRT data, we still have a number of manuscripts to write together. **Kaarel Krjutskov,** I really admire your dedication to fix these sequencing protocols, thanks for helping with the RNA profiling of the CLA cells. **Ingegerd Fransson**, for your help with the DNA preparations in the DIPP study and library preparations for STRT profiling. **Morana Vitezic**, thank you very much for all the support with the analysis of the FANTOM5 data in the DIPP project. I would also like to thank all the co-authors of the papers included in this thesis and the collaborators that provide the clinical samples with special regards to **Gunilla Hedlin**, **Heikki Hyöty** and **Mikael Knip**.

Many of the findings written in this book are the result of the skillful and dedicated work of the personnel at the Mutation Analysis Core Facility (MAF) with especial regards to **Kristina Duvefelt, Lotta Rönn, Jessica Lindvall and Gunnar Falk;** as well as at the personnel at the Bioinformatics and Expression Core Facility (BEA), with special regards to **Susann Fält, David Brodin, Marika Rönnholm, Patrick Muller, and Fredrik Fagerström-Billai.**

I also thank present and past member of the JKE lab and colleagues at the Department of Biosciences and Nutrition, for the support during the last years and making my life very sweet and full of joy. Also, to present and past colleagues at L2:04 and the infrastructure group. Finally, I want to give special regards to my friends **Luis Caraballo, Nancy et Hervé Zeller, Karen Sunahara, Marisa Baptista, Kurt Arkestål, Lenis Alvarez, Patricia Torregrosa, Kerstin Lagercrantz, Liliana Landazabal, Vladimir Pabón-Martinez, Ghazaleh Asadi, Isabel Tapia**, **Andrea Bieder**, **Botond Rajna, Germán Camargo**, **Saioa Goñi, Marcela Gonzalez** and **Hector Espinoza**. También quiero agradecer a todos los miembros de mi familia por su apoyo durante estos años, en especial honrar a mi madre **Rosario Caballero,** quien me inculcó el amor por el estudio y a mi abuelita **Ana Catalina Tovio** por cuidarme con esmero y apoyar todos mis planes. Eres una de las personas que más admiro y fuiste mi inspiración para empezar una nueva vida lejos de mi país en busca de una meta. Te adoro con toda mi vida. A mis tías y tíos, primos y primas, los quiero mucho!

9 References

1 Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921, doi:10.1038/35057062 (2001).

2 Ober, C. & Yao, T. C. The genetics of asthma and allergic disease: a 21st century perspective. *Immunol Rev* **242**, 10-30, doi:10.1111/j.1600-065X.2011.01029.x (2011).

3 Holloway, J. W., Yang, I. A. & Holgate, S. T. Genetics of allergic disease. *J Allergy Clin Immunol* **125**, S81-94, doi:10.1016/j.jaci.2009.10.071 (2010).

4 Feinberg, A. P. Phenotypic plasticity and the epigenetics of human disease. *Nature* **447**, 433-440, doi:10.1038/nature05919 (2007).

5 Petronis, A. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature* **465**, 721-727, doi:10.1038/nature09230 (2010).

6 Portela, A. & Esteller, M. Epigenetic modifications and human disease. *Nat Biotechnol* **28**, 1057-1068, doi:10.1038/nbt.1685 (2010).

7 Forrest, A. R. *et al.* A promoter-level mammalian expression atlas. *Nature* **507**, 462-470, doi:10.1038/nature13182 (2014).

8 Waddington, C. H. The epigenotype. *Endeavor* **1**, 18-20 (1942).

9 Waddington, C. H. The strategy of the genes. A discussion of some aspects of theoretical biology. With an appendix by H. Kacser. *London: George Allen & Unwin, Ltd.*, (1957).

10 Papp, B. & Plath, K. Reprogramming to pluripotency: stepwise resetting of the epigenetic landscape. *Cell Res* **21**, 486-501, doi:10.1038/cr.2011.28 (2011).

11 Rassoulzadegan, M. *et al.* RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature* **441**, 469-474, doi:10.1038/nature04674 (2006).

12 Jirtle, R. L. & Skinner, M. K. Environmental epigenomics and disease susceptibility. *Nat Rev Genet* **8**, 253-262, doi:10.1038/nrg2045 (2007).

13 Panettieri, R. A., Jr., Covar, R., Grant, E., Hillyer, E. V. & Bacharier, L. Natural history of asthma: persistence versus progression-does the beginning predict the end? *J Allergy Clin Immunol* **121**, 607-613, doi:10.1016/j.jaci.2008.01.006 (2008).

14 Asher, M. I. *et al.* Worldwide time trends in the prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and eczema in childhood: ISAAC Phases One and Three repeat multicountry cross-sectional surveys. *Lancet* **368**, 733-743, doi:10.1016/S0140-6736(06)69283-0 (2006).

15 Hotchkiss, R. D. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *J Biol Chem* **175**, 315-332 (1948).

16 Holliday, R. & Pugh, J. E. DNA modification mechanisms and gene activity during development. *Science* **187**, 226-232 (1975).

17 Riggs, A. D. X inactivation, differentiation, and DNA methylation. *Cytogenet Cell Genet* **14**, 9-25 (1975).

18 Jones, P. L. *et al.* Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* **19**, 187-191, doi:10.1038/561 (1998).

19 Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* **13**, 484-492, doi:10.1038/nrg3230 (2012).

20 Bock, C. Analysing and interpreting DNA methylation data. *Nat Rev Genet* **13**, 705-719, doi:10.1038/nrg3273 (2012).

21 Walsh, C. P., Chaillet, J. R. & Bestor, T. H. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* **20**, 116-117, doi:10.1038/2413 (1998).

22 Schulz, W. A., Steinhoff, C. & Florl, A. R. Methylation of endogenous human retroelements in health and disease. *Curr Top Microbiol Immunol* **310**, 211-250 (2006).

23 Gardiner-Garden, M. & Frommer, M. CpG islands in vertebrate genomes. *J Mol Biol* **196**, 261-282 (1987).

24 Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* **39**, 457-466, doi:10.1038/ng1990 (2007).

25 Illingworth, R. *et al.* A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol* **6**, e22, doi:10.1371/journal.pbio.0060022 (2008).

26 Irizarry, R. A. *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* **41**, 178-186, doi:10.1038/ng.298 (2009).

27 Orphanides, G. & Reinberg, D. A unified theory of gene expression. *Cell* **108**, 439-451 (2002).

28 Hellman, A. & Chess, A. Gene body-specific methylation on the active X chromosome. *Science* **315**, 1141-1143, doi:10.1126/science.1136352 (2007).

29 Ball, M. P. *et al.* Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol* **27**, 361-368, doi:10.1038/nbt.1533 (2009).

30 Maunakea, A. K. *et al.* Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466**, 253-257, doi:10.1038/nature09165 (2010).

31 Shukla, S. *et al.* CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**, 74-79, doi:10.1038/nature10442 (2011).

32 Deaton, A. M. *et al.* Cell type-specific DNA methylation at intragenic CpG islands in the immune system. *Genome Res* **21**, 1074-1086, doi:10.1101/gr.118703.110 (2011).

33 Rakyan, V. K., Down, T. A., Balding, D. J. & Beck, S. Epigenome-wide association studies for common human diseases. *Nat Rev Genet* **12**, 529-541, doi:10.1038/nrg3000 (2011).

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

35 Yanagisawa, Y., Ito, E., Yuasa, Y. & Maruyama, K. The human DNA methyltransferases DNMT3A and DNMT3B have two types of promoters with different CpG contents. *Biochim Biophys Acta* **1577**, 457-465 (2002).

36 Xie, S. *et al.* Cloning, expression and chromosome locations of the human DNMT3 gene family. *Gene* **236**, 87-95 (1999).

37 Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-257 (1999).

38 Jia, D., Jurkowska, R. Z., Zhang, X., Jeltsch, A. & Cheng, X. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature* **449**, 248-251, doi:10.1038/nature06146 (2007).

39 Aapola, U. *et al.* Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. *Genomics* **65**, 293-298, doi:10.1006/geno.2000.6168 (2000).

40 Lienert, F. *et al.* Identification of genetic elements that autonomously determine DNA methylation states. *Nat Genet* **43**, 1091-1097, doi:10.1038/ng.946 (2011).

41 Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930-935, doi:10.1126/science.1170116 (2009).

42 Cortellino, S. *et al.* Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146**, 67-79, doi:10.1016/j.cell.2011.06.020 (2011).

43 Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat Rev Genet* **14**, 204-220, doi:10.1038/nrg3354 (2013).

44 Mayer, W., Niveleau, A., Walter, J., Fundele, R. & Haaf, T. Demethylation of the zygotic paternal genome. *Nature* **403**, 501-502, doi:10.1038/35000654 (2000).

45 Ratnam, S. *et al.* Dynamics of Dnmt1 methyltransferase expression and intracellular localization during oogenesis and preimplantation development. *Dev Biol* **245**, 304-314, doi:10.1006/dbio.2002.0628 (2002).

46 Lees-Murdock, D. J. & Walsh, C. P. DNA methylation reprogramming in the germ line. *Epigenetics* **3**, 5-13 (2008).

47 Hendrich, B. & Bird, A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol* **18**, 6538-6547 (1998).

48 Ballestar, E. & Wolffe, A. P. Methyl-CpG-binding proteins. Targeting specific gene repression. *Eur J Biochem* **268**, 1-6 (2001).

49 Bostick, M. *et al.* UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* **317**, 1760-1764, doi:10.1126/science.1147939 (2007).

50 Bianchi, C. & Zangi, R. How to distinguish methyl-cytosine from cytosine with high fidelity. *J Mol Biol* **424**, 215-224, doi:10.1016/j.jmb.2012.09.024 (2012).

51 Filion, G. J. *et al.* A family of human zinc finger proteins that bind methylated DNA and repress transcription. *Mol Cell Biol* **26**, 169-181, doi:10.1128/MCB.26.1.169-181.2006 (2006).

52 Rothbart, S. B. *et al.* Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. *Nat Struct Mol Biol* **19**, 1155-1160, doi:10.1038/nsmb.2391 (2012).

53 Cockerill, P. N. Structure and function of active chromatin and DNase I hypersensitive sites. *Febs J* **278**, 2182-2210, doi:10.1111/j.1742-4658.2011.08128.x (2011).

54 Tessarz, P. & Kouzarides, T. Histone core modifications regulating nucleosome structure and dynamics. *Nat Rev Mol Cell Biol* **15**, 703-708, doi:10.1038/nrm3890 (2014).

55 Bell, O., Tiwari, V. K., Thoma, N. H. & Schubeler, D. Determinants and dynamics of genome accessibility. *Nat Rev Genet* **12**, 554-564, doi:10.1038/nrg3017 (2011).

56 Ho, L. & Crabtree, G. R. Chromatin remodelling during development. *Nature* **463**, 474-484, doi:10.1038/nature08911 (2010).

57 Berry, W. L. & Janknecht, R. KDM4/JMJD2 histone demethylases: epigenetic regulators in cancer cells. *Cancer Res* **73**, 2936-2942, doi:10.1158/0008-5472.CAN-12-4300 (2013).

58 Black, J. C., Van Rechem, C. & Whetstine, J. R. Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Mol Cell* **48**, 491-507, doi:10.1016/j.molcel.2012.11.006 (2012).

59 Cheng, X. Structural and functional coordination of DNA and histone methylation. *Cold Spring Harb Perspect Biol* **6**, doi:10.1101/cshperspect.a018747 (2014).

60 Chen, T. & Dent, S. Y. Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nat Rev Genet* **15**, 93-106, doi:10.1038/nrg3607 (2014).

61 Lewis, E. B. A gene complex controlling segmentation in Drosophila. *Nature* **276**, 565-570 (1978).

62 Shao, Z. *et al.* Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* **98**, 37-46, doi:10.1016/S0092-8674(00)80604-2 (1999).

63 Levine, S. S. *et al.* The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans. *Mol Cell Biol* **22**, 6070-6078 (2002).

64 Di Croce, L. & Helin, K. Transcriptional regulation by Polycomb group proteins. *Nat Struct Mol Biol* **20**, 1147-1155, doi:10.1038/nsmb.2669 (2013).

65 Francis, N. J., Kingston, R. E. & Woodcock, C. L. Chromatin compaction by a polycomb group protein complex. *Science* **306**, 1574-1577, doi:10.1126/science.1100576 (2004).

66 Pasini, D. *et al.* JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* **464**, 306-310, doi:10.1038/nature08788 (2010).

67 Hunkapiller, J. *et al.* Polycomb-like 3 promotes polycomb repressive complex 2 binding to CpG islands and embryonic stem cell self-renewal. *PLoS Genet* **8**, e1002576, doi:10.1371/journal.pgen.1002576 (2012).

68 Vire, E. *et al.* The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439**, 871-874, doi:10.1038/nature04431 (2006).

69 Aloia, L., Di Stefano, B. & Di Croce, L. Polycomb complexes in stem cells and embryonic development. *Development* **140**, 2525-2534, doi:10.1242/dev.091553 (2013).

70 Sun, Y. V. The Influences of Genetic and Environmental Factors on Methylome-wide Association Studies for Human Diseases. *Curr Genet Med Rep* **2**, 261-270, doi:10.1007/s40142-014-0058-2 (2014).

71 Zhang, D. *et al.* Genetic control of individual differences in gene-specific methylation in human brain. *Am J Hum Genet* **86**, 411-419, doi:10.1016/j.ajhg.2010.02.005 (2010).

72 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**, e53877, doi:10.1371/journal.pone.0053877 (2013).

73 Giacca, M. *et al.* A protein target site in an early replicated human DNA sequence: a highly conserved binding motif. *Biochem Biophys Res Commun* **165**, 956-965 (1989).

74 Zhi, D. *et al.* SNPs located at CpG sites modulate genome-epigenome interaction. *Epigenetics* **8**, 802-806, doi:10.4161/epi.25501 (2013).

75 Gibbs, J. R. *et al.* Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet* **6**, e1000952, doi:10.1371/journal.pgen.1000952 (2010).

76 Shoemaker, R., Deng, J., Wang, W. & Zhang, K. Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. *Genome Res* **20**, 883-889, doi:10.1101/gr.104695.109 (2010).

77 McRae, A. F. *et al.* Contribution of genetic variation to transgenerational inheritance of DNA methylation. *Genome Biol* **15**, R73, doi:10.1186/gb-2014-15-5-r73 (2014).

78 Ji, H. *et al.* Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature* **467**, 338-342, doi:10.1038/nature09367 (2010).

79 Milani, L. *et al.* DNA methylation for subtype classification and prediction of treatment outcome in patients with childhood acute lymphoblastic leukemia. *Blood* **115**, 1214-1225, doi:10.1182/blood-2009-04-214668 (2010).

80 Lubbert, M., Mertelsmann, R. & Herrmann, F. Cytosine methylation changes during normal hematopoiesis and in acute myeloid leukemia. *Leukemia* **11 Suppl 1**, S12-18 (1997).

81 Janson, P. C. *et al.* Profiling of CD4+ T cells with epigenetic immune lineage analysis. *J Immunol* **186**, 92-102, doi:10.4049/jimmunol.1000960 (2011).

82 Wilson, C. B., Rowell, E. & Sekimata, M. Epigenetic control of T-helper-cell differentiation. *Nat Rev Immunol* **9**, 91-105, doi:10.1038/nri2487 (2009).

83 Decker, T. *et al.* Stepwise activation of enhancer and promoter regions of the B cell commitment gene Pax5 in early lymphopoiesis. *Immunity* **30**, 508-520, doi:10.1016/j.immuni.2009.01.012 (2009).

84 Yang, Q., Jeremiah Bell, J. & Bhandoola, A. T-cell lineage determination. *Immunol Rev* **238**, 12-22, doi:10.1111/j.1600-065X.2010.00956.x (2010).

85 Artis, D. & Spits, H. The biology of innate lymphoid cells. *Nature* **517**, 293-301, doi:10.1038/nature14189 (2015).

86 Terry, M. B., Delgado-Cruzata, L., Vin-Raviv, N., Wu, H. C. & Santella, R. M. DNA methylation in white blood cells: association with risk factors in epidemiologic studies. *Epigenetics* **6**, 828-837 (2011).

87 Zhu, J., Yamane, H. & Paul, W. E. Differentiation of effector CD4 T cell populations (*). *Annu Rev Immunol* **28**, 445-489, doi:10.1146/annurev-immunol-030409-101212 (2010).

88 Poli, A. *et al.* CD56bright natural killer (NK) cells: an important NK cell subset. *Immunology* **126**, 458-465, doi:10.1111/j.1365-2567.2008.03027.x (2009).

89 Ziegler-Heitbrock, L. *et al.* Nomenclature of monocytes and dendritic cells in blood. *Blood* **116**, e74-80, doi:10.1182/blood-2010-02-258558 (2010).

90 Kolaczkowska, E. & Kubes, P. Neutrophil recruitment and function in health and inflammation. *Nat Rev Immunol* **13**, 159-175, doi:10.1038/nri3399 (2013).

91 Stone, K. D., Prussin, C. & Metcalfe, D. D. IgE, mast cells, basophils, and eosinophils. *J Allergy Clin Immunol* **125**, S73-80, doi:10.1016/j.jaci.2009.11.017 (2010).

92 Girardot, M., Feil, R. & Lleres, D. Epigenetic deregulation of genomic imprinting in humans: causal mechanisms and clinical implications. *Epigenomics* **5**, 715-728, doi:10.2217/epi.13.66 (2013).

93 Robertson, K. D. DNA methylation and human disease. *Nat Rev Genet* **6**, 597-610, doi:10.1038/nrg1655 (2005).

94 Watanabe, Y. & Maekawa, M. Methylation of DNA in cancer. *Adv Clin Chem* **52**, 145-167 (2010).

95 Feinberg, A. P. & Tycko, B. The history of cancer epigenetics. *Nat Rev Cancer* **4**, 143-153, doi:10.1038/nrc1279 (2004).

96 Grayson, D. R. & Guidotti, A. The dynamics of DNA methylation in schizophrenia and related psychiatric disorders. *Neuropsychopharmacology* **38**, 138-166, doi:10.1038/npp.2012.125 (2013).

97 Lei, W. *et al.* Abnormal DNA methylation in CD4+ T cells from patients with systemic lupus erythematosus, systemic sclerosis, and dermatomyositis. *Scand J Rheumatol* **38**, 369-374, doi:10.1080/03009740902758875 (2009).

98 Stefan, M., Zhang, W., Concepcion, E., Yi, Z. & Tomer, Y. DNA methylation profiles in type 1 diabetes twins point to strong epigenetic effects on etiology. *J Autoimmun* **50**, 33-37, doi:10.1016/j.jaut.2013.10.001 (2014).

99 Wollmann, H. & Weigel, D. Small RNAs in flower development. *Eur J Cell Biol* **89**, 250-257, doi:10.1016/j.ejcb.2009.11.004 (2010).

100 Grishok, A., Tabara, H. & Mello, C. C. Genetic requirements for inheritance of RNAi in C. elegans. *Science* **287**, 2494-2497 (2000).

101 Grandjean, V. *et al.* The miR-124-Sox9 paramutation: RNA-mediated epigenetic control of embryonic and adult growth. *Development* **136**, 3647-3655, doi:10.1242/dev.041061 (2009).

102 Bao, N., Lye, K. W. & Barton, M. K. MicroRNA binding sites in Arabidopsis class III HD-ZIP mRNAs are required for methylation of the template chromosome. *Dev Cell* **7**, 653-662, doi:10.1016/j.devcel.2004.10.003 (2004).

103 Wiklund, E. D., Kjems, J. & Clark, S. J. Epigenetic architecture and miRNA: reciprocal regulators. *Epigenomics* **2**, 823-840, doi:10.2217/epi.10.51 (2010).

104 Dai, E. *et al.* EpimiR: a database of curated mutual regulation between miRNAs and epigenetic modifications. *Database (Oxford)* **2014**, bau023, doi:10.1093/database/bau023 (2014).

105 Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215-233, doi:10.1016/j.cell.2009.01.002 (2009).

106 Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of mammalian microRNA targets. *Cell* **115**, 787-798 (2003).

107 Berezikov, E. Evolution of microRNA diversity and regulation in animals. *Nat Rev Genet* **12**, 846-860, doi:10.1038/nrg3079 (2011).

108 Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15-20, doi:10.1016/j.cell.2004.12.035 (2005).

109 Lee, R. C., Feinbaum, R. L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843-854 (1993).

110 Ambros, V. The functions of animal microRNAs. *Nature* **431**, 350-355, doi:10.1038/nature02871 (2004).

111 Eulalio, A., Huntzinger, E. & Izaurralde, E. Getting to the root of miRNA-mediated gene silencing. *Cell* **132**, 9-14, doi:10.1016/j.cell.2007.12.024 (2008).

112 Vasudevan, S., Tong, Y. & Steitz, J. A. Switching from repression to activation: microRNAs can up-regulate translation. *Science* **318**, 1931-1934, doi:10.1126/science.1149460 (2007).

113 Sato, F., Tsuchiya, S., Meltzer, S. J. & Shimizu, K. MicroRNAs and epigenetics. *Febs J* **278**, 1598-1609, doi:10.1111/j.1742-4658.2011.08089.x (2011).

114 De Guire, V. *et al.* Circulating miRNAs as sensitive and specific biomarkers for the diagnosis and monitoring of human diseases: promises and challenges. *Clin Biochem* **46**, 846-860, doi:10.1016/j.clinbiochem.2013.03.015 (2013).

115 Lu, J. *et al.* MicroRNA expression profiles classify human cancers. *Nature* **435**, 834-838, doi:10.1038/nature03702 (2005).

116 Rosenfeld, N. *et al.* MicroRNAs accurately identify cancer tissue origin. *Nat Biotechnol* **26**, 462-469, doi:10.1038/nbt1392 (2008).

117 van der Ree, M. H. *et al.* Long-term safety and efficacy of microRNA-targeted therapy in chronic hepatitis C patients. *Antiviral Res* **111**, 53-59, doi:10.1016/j.antiviral.2014.08.015 (2014).

118 Ling, H., Fabbri, M. & Calin, G. A. MicroRNAs and other non-coding RNAs as targets for anticancer drug development. *Nat Rev Drug Discov* **12**, 847-865, doi:10.1038/nrd4140 (2013).

119 Johansson, S. G. *et al.* A revised nomenclature for allergy. An EAACI position statement from the EAACI nomenclature task force. *Allergy* **56**, 813-824 (2001).

120 Weidinger, S., Baurecht, H., Naumann, A. & Novak, N. Genome-wide association studies on IgE regulation: are genetics of IgE also genetics of atopic disease? *Curr Opin Allergy Clin Immunol* **10**, 408-417, doi:10.1097/ACI.0b013e32833d7d2d (2010).

121 Haahtela, T. *et al.* The biodiversity hypothesis and allergic disease: world allergy organization position statement. *World Allergy Organ J* **6**, 3, doi:10.1186/1939-4551-6-3 (2013).

122 Kuriakose, J. S. & Miller, R. L. Environmental epigenetics and allergic diseases: recent advances. *Clin Exp Allergy* **40**, 1602-1610, doi:10.1111/j.1365-2222.2010.03599.x (2010).

123 Harb, H. & Renz, H. Update on epigenetics in allergic disease. *J Allergy Clin Immunol* **135**, 15-24, doi:10.1016/j.jaci.2014.11.009 (2015).

124 Lockett, G. A. *et al.* Epigenomics and allergic disease. *Epigenomics* **5**, 685-699, doi:10.2217/epi.13.68 (2013).

125 Bieber, T. Atopic dermatitis. *N Engl J Med* **358**, 1483-1494, doi:10.1056/NEJMra074081 (2008).

126 Lammintausta, K., Kalimo, K., Raitala, R. & Forsten, Y. Prognosis of atopic dermatitis. A prospective study in early adulthood. *Int J Dermatol* **30**, 563-568 (1991).

127 Flohr, C. & Mann, J. New insights into the epidemiology of childhood atopic dermatitis. *Allergy* **69**, 3-16, doi:10.1111/all.12270 (2014).

128 Thomsen, S. F. *et al.* Importance of genetic factors in the etiology of atopic dermatitis: a twin study. *Allergy Asthma Proc* **28**, 535-539, doi:10.2500/aap2007.28.3041 (2007).

129 Ellinghaus, D. *et al.* High-density genotyping study identifies four new susceptibility loci for atopic dermatitis. *Nat Genet* **45**, 808-812, doi:10.1038/ng.2642 (2013).

130 O'Regan, G. M., Sandilands, A., McLean, W. H. & Irvine, A. D. Filaggrin in atopic dermatitis. *J Allergy Clin Immunol* **122**, 689-693, doi:10.1016/j.jaci.2008.08.002 (2008).

131 Soderhall, C. *et al.* Variants in a novel epidermal collagen gene (COL29A1) are associated with atopic dermatitis. *PLoS Biol* **5**, e242, doi:10.1371/journal.pbio.0050242 (2007).

132 Weidinger, S. *et al.* A genome-wide association study of atopic dermatitis identifies loci with overlapping effects on asthma and psoriasis. *Hum Mol Genet* **22**, 4841-4856, doi:10.1093/hmg/ddt317 (2013).

133 Paternoster, L. *et al.* Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. *Nat Genet* **44**, 187-192, doi:10.1038/ng.1017 (2012).

134 Oyoshi, M. K., He, R., Kumar, L., Yoon, J. & Geha, R. S. Cellular and molecular mechanisms in atopic dermatitis. *Adv Immunol* **102**, 135-226, doi:10.1016/S0065-2776(09)01203-6 (2009).

135 Gong, J. Q. *et al.* Skin colonization by Staphylococcus aureus in patients with eczema and atopic dermatitis and relevant combined topical therapy: a double-blind multicentre randomized controlled trial. *Br J Dermatol* **155**, 680-687, doi:10.1111/j.1365-2133.2006.07410.x (2006).

136 Saunders, C. W., Scheynius, A. & Heitman, J. Malassezia fungi are specialized to live on skin and associated with dandruff, eczema, and other skin diseases. *PLoS Pathog* **8**, e1002701, doi:10.1371/journal.ppat.1002701 (2012).

137 Schmid-Grendelmeier, P., Scheynius, A. & Crameri, R. The role of sensitization to Malassezia sympodialis in atopic eczema. *Chem Immunol Allergy* **91**, 98-109, doi:10.1159/000090246 (2006).

138 Kekki, O. M. *et al.* Sensitization to Malassezia in children with atopic dermatitis combined with food allergy. *Pediatr Allergy Immunol* **24**, 244-249, doi:10.1111/pai.12057 (2013).

139 Saaf, A. M. *et al.* Global expression profiling in atopic eczema reveals reciprocal expression of inflammatory and lipid genes. *PLoS One* **3**, e4017, doi:10.1371/journal.pone.0004017 (2008).

140 Severity scoring of atopic dermatitis: the SCORAD index. Consensus Report of the European Task Force on Atopic Dermatitis. *Dermatology* **186**, 23-31 (1993).

141 Mittermann, I. *et al.* Autoimmunity and atopic dermatitis. *Curr Opin Allergy Clin Immunol* **4**, 367-371 (2004).

142 Schmid-Grendelmeier, P. *et al.* IgE-mediated and T cell-mediated autoimmunity against manganese superoxide dismutase in atopic dermatitis. *J Allergy Clin Immunol* **115**, 1068-1075, doi:10.1016/j.jaci.2005.01.065 (2005).

143 Limacher, A. *et al.* Cross-reactivity and 1.4-A crystal structure of Malassezia sympodialis thioredoxin (Mala s 13), a member of a new pan-allergen family. *J Immunol* **178**, 389-396 (2007).

144 Hradetzky, S. *et al.* Differential cytokine induction by the human skin-associated autoallergen thioredoxin in sensitized patients with atopic dermatitis and healthy control subjects. *J Allergy Clin Immunol*, doi:10.1016/j.jaci.2014.10.038 (2014).

145 Ring, J. *et al.* Guidelines for treatment of atopic eczema (atopic dermatitis) part I. *J Eur Acad Dermatol Venereol* **26**, 1045-1060, doi:10.1111/j.1468-3083.2012.04635.x (2012).

146 Darsow, U. Allergen-specific immunotherapy for atopic eczema: updated. *Curr Opin Allergy Clin Immunol* **12**, 665-669, doi:10.1097/ACI.0b013e3283588cf4 (2012).

147 Beck, L. A. *et al.* Dupilumab treatment in adults with moderate-to-severe atopic dermatitis. *N Engl J Med* **371**, 130-139, doi:10.1056/NEJMoa1314768 (2014).

148 Berg, E. L. *et al.* The cutaneous lymphocyte antigen is a skin lymphocyte homing receptor for the vascular lectin endothelial cell-leukocyte adhesion molecule 1. *J Exp Med* **174**, 1461-1466 (1991).

149 Picker, L. J. *et al.* Differential expression of homing-associated adhesion molecules by T cell subsets in man. *J Immunol* **145**, 3247-3255 (1990).

150 Picker, L. J., Michie, S. A., Rott, L. S. & Butcher, E. C. A unique phenotype of skin-associated lymphocytes in humans. Preferential expression of the HECA-452 epitope by benign and malignant T cells at cutaneous sites. *Am J Pathol* **136**, 1053-1068 (1990).

151 Biedermann, T. *et al.* Targeting CLA/E-selectin interactions prevents CCR4-mediated recruitment of human Th2 memory cells to human skin in vivo. *Eur J Immunol* **32**, 3171-3180, doi:10.1002/1521-4141(200211)32:11<3171::AID-IMMU3171>3.0.CO;2-4 (2002).

152 Torres, M. J. *et al.* Circulating CLA+ lymphocytes from children with atopic dermatitis contain an increased percentage of cells bearing staphylococcal-related T-cell receptor variable segments. *Clin Exp Allergy* **28**, 1264-1272 (1998).

153 Teraki, Y., Hotta, T. & Shiohara, T. Increased circulating skin-homing cutaneous lymphocyte-associated antigen (CLA)+ type 2 cytokine-producing cells, and decreased CLA+ type 1 cytokine-producing cells in atopic dermatitis. *Br J Dermatol* **143**, 373-378 (2000).

154 Balaji, H. *et al.* Malassezia sympodialis thioredoxin-specific T cells are highly cross-reactive to human thioredoxin in atopic dermatitis. *J Allergy Clin Immunol* **128**, 92-99 e94, doi:10.1016/j.jaci.2011.02.043 (2011).

155 Davison, S., Allen, M., Vaughan, R. & Barker, J. Staphylococcal toxin-induced T cell proliferation in atopic eczema correlates with increased use of superantigen-reactive Vbeta-chains in cutaneous lymphocyte-associated antigen (CLA)-positive lymphocytes. *Clin Exp Immunol* **121**, 181-186 (2000).

156 Santamaria Babi, L. F. *et al.* Circulating allergen-reactive T cells from patients with atopic dermatitis and allergic contact dermatitis express the skin-selective homing receptor, the cutaneous lymphocyte-associated antigen. *J Exp Med* **181**, 1935-1940 (1995).

157 Akdis, C. A., Akdis, M., Simon, H. U. & Blaser, K. Regulation of allergic inflammation by skin-homing T cells in allergic eczema. *Int Arch Allergy Immunol* **118**, 140-144, doi:24051 (1999).

158 Akdis, M. *et al.* Skin homing (cutaneous lymphocyte-associated antigen-positive) CD8+ T cells respond to superantigen and contribute to eosinophilia and IgE production in atopic dermatitis. *J Immunol* **163**, 466-475 (1999).

159 Akdis, M., Akdis, C. A., Weigl, L., Disch, R. & Blaser, K. Skin-homing, CLA+ memory T cells are activated in atopic dermatitis and regulate IgE by an IL-13-dominated cytokine pattern: IgG4 counter-regulation by CLA- memory T cells. *J Immunol* **159**, 4611-4619 (1997).

160 Antunez, C. *et al.* Different cytokine production and activation marker profiles in circulating cutaneous-lymphocyte-associated antigen T cells from patients with acute or chronic atopic dermatitis. *Clin Exp Allergy* **34**, 559-566, doi:10.1111/j.1365-2222.2004.1933.x (2004).

161 Homey, B., Steinhoff, M., Ruzicka, T. & Leung, D. Y. Cytokines and chemokines orchestrate atopic skin inflammation. *J Allergy Clin Immunol* **118**, 178-189, doi:10.1016/j.jaci.2006.03.047 (2006).

162 Gunther, C. *et al.* CCL18 is expressed in atopic dermatitis and mediates skin homing of human memory T cells. *J Immunol* **174**, 1723-1728 (2005).

163 Akdis, M., Trautmann, A., Klunker, S., Blaser, K. & Akdis, C. A. Cytokine network and dysregulated apoptosis in atopic dermatitis. *Acta Odontol Scand* **59**, 178-182 (2001).

164 Heratizadeh, A. *et al.* The role of T-cell reactivity towards the autoantigen alpha-NAC in atopic dermatitis. *Br J Dermatol* **164**, 316-324, doi:10.1111/j.1365-2133.2010.10090.x (2011).

165 Clark, R. A. *et al.* The vast majority of CLA+ T cells are resident in normal skin. *J Immunol* **176**, 4431-4439 (2006).

166 Zollner, T. M., Nuber, V., Duijvestijn, A. M., Boehncke, W. H. & Kaufmann, R. Superantigens but not mitogens are capable of inducing upregulation of E-selectin ligands on human T lymphocytes. *Exp Dermatol* **6**, 161-166 (1997).

167 Wenzel, S. E. Asthma phenotypes: the evolution from clinical to molecular approaches. *Nat Med* **18**, 716-725, doi:10.1038/nm.2678 (2012).

168 Lotvall, J. *et al.* Asthma endotypes: a new approach to classification of disease entities within the asthma syndrome. *J Allergy Clin Immunol* **127**, 355-360, doi:10.1016/j.jaci.2010.11.037 (2011).

169 Tan, D. J. *et al.* Age-of-asthma onset as a determinant of different asthma phenotypes in adults: a systematic review and meta-analysis of the literature. *Expert Rev Respir Med* **9**, 109-123, doi:10.1586/17476348.2015.1000311 (2015).

170 London, S. J., James Gauderman, W., Avol, E., Rappaport, E. B. & Peters, J. M. Family history and the risk of early-onset persistent, early-onset transient, and late-onset asthma. *Epidemiology* **12**, 577-583 (2001).

171 van den Toorn, L. M., Overbeek, S. E., Prins, J. B., Hoogsteden, H. C. & de Jongste, J. C. Asthma remission: does it exist? *Curr Opin Pulm Med* **9**, 15-20 (2003).

172 Demenais, F., Chaudru, V. & Martinez, M. Detection of parent-of-origin effects for atopy by model-free and model-based linkage analyses. *Genet Epidemiol* **21 Suppl 1**, S186-191 (2001).

173 Fedulov, A. V. & Kobzik, L. Allergy risk is mediated by dendritic cells with congenital epigenetic changes. *Am J Respir Cell Mol Biol* **44**, 285-292, doi:10.1165/rcmb.2009-0400OC (2011).

174 Barnes, P. J. & Woolcock, A. J. Difficult asthma. *Eur Respir J* **12**, 1209-1218 (1998).

175 Currie, G. P., Douglas, J. G. & Heaney, L. G. Difficult to treat asthma in adults. *Bmj* **338**, b494, doi:10.1136/bmj.b494 (2009).

176 Calderon, M. A. *et al.* Allergen-specific immunotherapy for respiratory allergies: from meta-analysis to registration and beyond. *J Allergy Clin Immunol* **127**, 30-38, doi:10.1016/j.jaci.2010.08.024 (2011).

177 Holgate, S. T. *et al.* Local genetic and environmental factors in asthma disease pathogenesis: chronicity and persistence mechanisms. *Eur Respir J* **29**, 793-803, doi:10.1183/09031936.00087506 (2007).

178 Lambrecht, B. N. & Hammad, H. The airway epithelium in asthma. *Nat Med* **18**, 684-692, doi:10.1038/nm.2737 (2012).

179 Holgate, S. T. The sentinel role of the airway epithelium in asthma pathogenesis. *Immunol Rev* **242**, 205-219, doi:10.1111/j.1600-065X.2011.01030.x (2011).

180 Moffatt, M. F. *et al.* Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* **448**, 470-473, doi:10.1038/nature06014 (2007).

181 Smit, L. A. *et al.* 17q21 variants modify the association between early respiratory infections and asthma. *Eur Respir J* **36**, 57-64, doi:10.1183/09031936.00154509 (2010).

182 Verlaan, D. J. *et al.* Allele-specific chromatin remodeling in the ZPBP2/GSDMB/ORMDL3 locus associated with the risk of asthma and autoimmune disease. *Am J Hum Genet* **85**, 377-393, doi:10.1016/j.ajhg.2009.08.007 (2009).

183 Berlivet, S. *et al.* Interaction between genetic and epigenetic variation defines gene expression patterns at the asthma-associated locus 17q12-q21 in lymphoblastoid cell lines. *Hum Genet* **131**, 1161-1171, doi:10.1007/s00439-012-1142-x (2012).

184 Hjelmqvist, L. *et al.* ORMDL proteins are a conserved new family of endoplasmic reticulum membrane proteins. *Genome Biol* **3**, RESEARCH0027 (2002).

185 McGovern, D. P. *et al.* Genome-wide association identifies multiple ulcerative colitis susceptibility loci. *Nat Genet* **42**, 332-337, doi:10.1038/ng.549 (2010).

186 Laukens, D. *et al.* Evidence for significant overlap between common risk variants for Crohn's disease and ankylosing spondylitis. *PLoS One* **5**, e13795, doi:10.1371/journal.pone.0013795 (2010).

187 Kurreeman, F. A. *et al.* Use of a multiethnic approach to identify rheumatoid- arthritis-susceptibility loci, 1p36 and 17q12. *Am J Hum Genet* **90**, 524-532, doi:10.1016/j.ajhg.2012.01.010 (2012).

188 Breslow, D. K. *et al.* Orm family proteins mediate sphingolipid homeostasis. *Nature* **463**, 1048-1053, doi:10.1038/nature08787 (2010).

189 Cantero-Recasens, G., Fandos, C., Rubio-Moscardo, F., Valverde, M. A. & Vicente, R. The asthma-associated ORMDL3 gene product regulates endoplasmic reticulum-mediated calcium signaling and cellular stress. *Hum Mol Genet* **19**, 111-121, doi:10.1093/hmg/ddp471 (2010).

190 Carreras-Sureda, A. *et al.* ORMDL3 modulates store-operated calcium entry and lymphocyte activation. *Hum Mol Genet* **22**, 519-530, doi:10.1093/hmg/dds450 (2013).

191 Hogan, P. G., Lewis, R. S. & Rao, A. Molecular basis of calcium signaling in lymphocytes: STIM and ORAI. *Annu Rev Immunol* **28**, 491-533, doi:10.1146/annurev.immunol.021908.132550 (2010).

192 Baccarelli, A. & Bollati, V. Epigenetics and environmental chemicals. *Curr Opin Pediatr* **21**, 243-251 (2009).

193 Collotta, M., Bertazzi, P. A. & Bollati, V. Epigenetics and pesticides. *Toxicology* **307**, 35-41, doi:10.1016/j.tox.2013.01.017 (2013).

194 Wolff, G. L., Kodell, R. L., Moore, S. R. & Cooney, C. A. Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice. *Faseb J* **12**, 949-957 (1998).

195 Chmurzynska, A. Fetal programming: link between early nutrition, DNA methylation, and complex diseases. *Nutr Rev* **68**, 87-98, doi:10.1111/j.1753-4887.2009.00265.x (2010).

196 Heijmans, B. T. *et al.* Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A* **105**, 17046-17049, doi:10.1073/pnas.0806560105 (2008).

197 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**, 5R-10R, doi:10.1203/pdr.0b013e318045bedb (2007).

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

199 Richardson, B. Primer: epigenetics of autoimmunity. *Nat Clin Pract Rheumatol* **3**, 521-527, doi:10.1038/ncprheum0573 (2007).

200 Anway, M. D., Cupp, A. S., Uzumcu, M. & Skinner, M. K. Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science* **308**, 1466-1469, doi:10.1126/science.1108190 (2005).

201 McGowan, P. O. *et al.* Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nat Neurosci* **12**, 342-348, doi:10.1038/nn.2270 (2009).

202 Lam, L. L. *et al.* Factors underlying variable DNA methylation in a human community cohort. *Proc Natl Acad Sci U S A* **109 Suppl 2**, 17253-17260, doi:10.1073/pnas.1121249109 (2012).

203 Waterland, R. A. *et al.* Season of conception in rural gambia affects DNA methylation at putative human metastable epialleles. *PLoS Genet* **6**, e1001252, doi:10.1371/journal.pgen.1001252 (2010).

204 Shang, Y. *et al.* Epigenetic alterations by DNA methylation in house dust mite-induced airway hyperresponsiveness. *Am J Respir Cell Mol Biol* **49**, 279-287, doi:10.1165/rcmb.2012-0403OC (2013).

205 Liu, J. *et al.* Combined inhaled diesel exhaust particles and allergen exposure alter methylation of T helper genes and IgE production in vivo. *Toxicol Sci* **102**, 76-81, doi:10.1093/toxsci/kfm290 (2008).

206 Yang, T., Owen, J. L., Lightfoot, Y. L., Kladde, M. P. & Mohamadzadeh, M. Microbiota impact on the epigenetic regulation of colorectal cancer. *Trends Mol Med* **19**, 714-725, doi:10.1016/j.molmed.2013.08.005 (2013).

207 Orsmark-Pietras, C. *et al.* Transcriptome analysis reveals upregulation of bitter taste receptors in severe asthmatics. *Eur Respir J* **42**, 65-78, doi:10.1183/09031936.00077712 (2013).

208 Ronnerblad, M. *et al.* Analysis of the DNA methylome and transcriptome in granulopoiesis reveals timed changes and dynamic enhancer methylation. *Blood* **123**, e79-89, doi:10.1182/blood-2013-02-482893 (2014).

209 Zilbauer, M. *et al.* Genome-wide methylation analyses of primary human leukocyte subsets identifies functionally important cell-type-specific hypomethylated regions. *Blood* **122**, e52-60, doi:10.1182/blood-2013-05-503201 (2013).

210 Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**, 86, doi:10.1186/1471-2105-13-86 (2012).

211 Jaffe, A. E. & Irizarry, R. A. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol* **15**, R31, doi:10.1186/gb-2014-15-2-r31 (2014).

212 Adalsteinsson, B. T. *et al.* Heterogeneity in white blood cells has potential to confound DNA methylation measurements. *PLoS One* **7**, e46705, doi:10.1371/journal.pone.0046705 (2012).

213 Liu, Y. *et al.* Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol* **31**, 142-147, doi:10.1038/nbt.2487 (2013).

214 Absher, D. M. *et al.* Genome-wide DNA methylation analysis of systemic lupus erythematosus reveals persistent hypomethylation of interferon genes and compositional changes to CD4+ T-cell populations. *PLoS Genet* **9**, e1003678, doi:10.1371/journal.pgen.1003678 (2013).

215 Houseman, E. A., Molitor, J. & Marsit, C. J. Reference-free cell mixture adjustments in analysis of DNA methylation data. *Bioinformatics* **30**, 1431-1439, doi:10.1093/bioinformatics/btu029 (2014).

216 Zou, J., Lippert, C., Heckerman, D., Aryee, M. & Listgarten, J. Epigenome-wide association studies without the need for cell-type composition. *Nat Methods* **11**, 309-311, doi:10.1038/nmeth.2815 (2014).

217 Liang, Y. *et al.* Demethylation of the FCER1G promoter leads to FcepsilonRI overexpression on monocytes of patients with atopic dermatitis. *Allergy* **67**, 424-430, doi:10.1111/j.1398-9995.2011.02760.x (2012).

218 Luo, Y., Zhou, B., Zhao, M., Tang, J. & Lu, Q. Promoter demethylation contributes to TSLP overexpression in skin lesions of patients with atopic dermatitis. *Clin Exp Dermatol* **39**, 48-53, doi:10.1111/ced.12206 (2014).

219 Rodriguez, E. *et al.* An integrated epigenetic and transcriptomic analysis reveals distinct tissue-specific patterns of DNA methylation associated with atopic dermatitis. *J Invest Dermatol* **134**, 1873-1883, doi:10.1038/jid.2014.87 (2014).

220 Rebane, A. *et al.* MicroRNA-146a alleviates chronic skin inflammation in atopic dermatitis through suppression of innate immune responses in keratinocytes. *J Allergy Clin Immunol* **134**, 836-847 e811, doi:10.1016/j.jaci.2014.05.022 (2014).

221 Sonkoly, E. *et al.* MiR-155 is overexpressed in patients with atopic dermatitis and modulates T-cell proliferative responses by targeting cytotoxic T lymphocyte-associated antigen 4. *J Allergy Clin Immunol* **126**, 581-589 e581-520, doi:10.1016/j.jaci.2010.05.045 (2010).

222 Sonkoly, E., Stahle, M. & Pivarcsi, A. MicroRNAs: novel regulators in skin inflammation. *Clin Exp Dermatol* **33**, 312-315, doi:10.1111/j.1365-2230.2008.02804.x (2008).

223 Dijk, F. N., de Jongste, J. C., Postma, D. S. & Koppelman, G. H. Genetics of onset of asthma. *Curr Opin Allergy Clin Immunol* **13**, 193-202, doi:10.1097/ACI.0b013e32835eb707 (2013).

224 Granell, R. *et al.* Examination of the relationship between variation at 17q21 and childhood wheeze phenotypes. *J Allergy Clin Immunol* **131**, 685-694, doi:10.1016/j.jaci.2012.09.021 (2013).

225 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, doi:10.1016/j.ajhg.2014.02.011 (2014).

226 Caliskan, M. *et al.* Rhinovirus wheezing illness and genetic risk of childhood-onset asthma. *N Engl J Med* **368**, 1398-1407, doi:10.1056/NEJMoa1211592 (2013).

227 Halapi, E. *et al.* A sequence variant on 17q21 is associated with age at onset and severity of asthma. *Eur J Hum Genet* **18**, 902-908, doi:10.1038/ejhg.2010.38 (2010).

228 A promoter-level mammalian expression atlas. *Nature* **507**, 462-470, doi:10.1038/nature13182 (2014).

229 Lim, P. S., Li, J., Holloway, A. F. & Rao, S. Epigenetic regulation of inducible gene expression in the immune system. *Immunology* **139**, 285-293, doi:10.1111/imm.12100 (2013).

230 Alisch, R. S. *et al.* Age-associated DNA methylation in pediatric populations. *Genome Res* **22**, 623-632, doi:10.1101/gr.125187.111 (2012).

231 Morris, T. J. & Beck, S. Analysis pipelines and packages for Infinium HumanMethylation450 BeadChip (450k) data. *Methods* **72**, 3-8, doi:10.1016/j.ymeth.2014.08.011 (2015).

232 Mestdagh, P. *et al.* Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study. *Nat Methods* **11**, 809-815, doi:10.1038/nmeth.3014 (2014).

233 Islam, S. *et al.* Highly multiplexed and strand-specific single-cell RNA 5' end sequencing. *Nat Protoc* **7**, 813-828, doi:10.1038/nprot.2012.022 (2012).

234 Slaats, G. G. *et al.* DNA methylation levels within the CD14 promoter region are lower in placentas of mothers living on a farm. *Allergy* **67**, 895-903, doi:10.1111/j.1398-9995.2012.02831.x (2012).

235 Waddington, C. H. Selection of the genetic basis for an acquired character. *Nature* **169**, 278 (1952).

236 Daxinger, L. & Whitelaw, E. Understanding transgenerational epigenetic inheritance via the gametes in mammals. *Nat Rev Genet* **13**, 153-162, doi:10.1038/nrg3188 (2012).

237 Stefanowicz, D. *et al.* DNA methylation profiles of airway epithelial cells and PBMCs from healthy, atopic and asthmatic children. *PLoS One* **7**, e44213, doi:10.1371/journal.pone.0044213 (2012).

238 Junker, J. P. & van Oudenaarden, A. Every cell is special: genome-wide studies add a new dimension to single-cell biology. *Cell* **157**, 8-11, doi:10.1016/j.cell.2014.02.010 (2014).

239 Thompson, T. M. *et al.* Comparison of whole-genome DNA methylation patterns in whole blood, saliva, and lymphoblastoid cell lines. *Behav Genet* **43**, 168-176, doi:10.1007/s10519-012-9579-1 (2013).

240 Mill, J. & Heijmans, B. T. From promises to practical strategies in epigenetic epidemiology. *Nat Rev Genet* **14**, 585-594, doi:10.1038/nrg3405 (2013).