TESIS DE DOCTORADO

EPIGENETIC MODULATION OF THE IMMUNE SYSTEM IN EARLY CHILDHOOD: FROM INFECTIOUS DISEASES TO VACCINES

Sara Pischedda

ESCUELA DE DOCTORADO INTERNACIONAL DE LA UNIVERSIDAD DE SANTIAGO DE COMPOSTELA

PROGRAMA DE DOCTORADO EN MEDICINA MOLECULAR

SANTIAGO DE COMPOSTELA

2021

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

DECLARACIÓN DO AUTOR/A DA TESE

D./Dna. **Sara Pischedda**

Título da tese **Epigenetic modulation of the immune system in early childhood: from infectious diseases to vaccines**

Presento a miña tese, seguindo o procedemento axeitado ao Regulamento, e declaro que:

1) A tese abarca os resultados da elaboración do meu traballo.
2) De ser o caso, na tese faise referencia ás colaboracións que tivo este traballo.
3) Confirmo que a tese non incorre en ningún tipo de plaxio doutros autores nin de traballos presentados por min para a obtención doutros títulos.
4) A tese é a versión definitiva presentada para a súa defensa e coincide a versión impresa coa presentada en formato electrónico

E comprométome a presentar o Compromiso Documental de Supervisión no caso de que o orixinal non estea na Escola.

En **Santiago de Compostela, 12 de Maio de 2021**.

Sinatura electronica

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

AUTORIZACIÓN DO DIRECTOR / TITOR DA TESE

Epigenetic modulation of the immune system in early childhood: from infectious diseases to vaccines

D. Antonio Salas.

D. Federico Martinon Torres

INFORMAN:

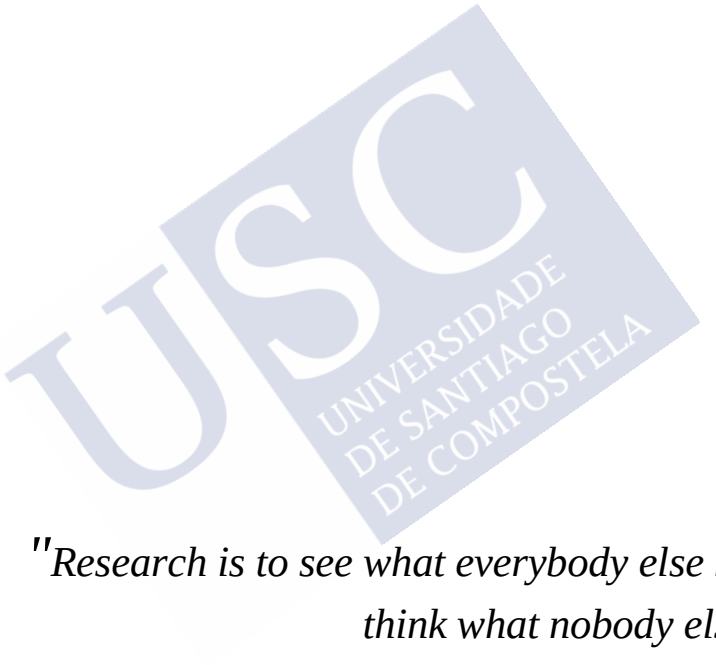
Que a presente tese, correspóndese co traballo realizado por Dna. Sara Pischedda, baixo a miña dirección/titorización, e autorizo a súa presentación, considerando que reúne os requisitos esixidos no Regulamento de Estudos de Doutoramento da USC, e que como director desta non incorre nas causas de abstención establecidas na Lei 40/2015.

De acordo co indicado no Regulamento de Estudos de Doutoramento, declara tamén que a presente tese de doutoramento é idónea para ser defendida en base á modalidade de Monográfica con reproducción de publicaciones, nos que a participación do/a doutorando/a foi decisiva para a súa elaboración e as publicacións se axustan ao Plan de Investigación.

En Santiago de Compostela, 12 de Mayo de 2021

Sinatura electronica

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA



"Research is to see what everybody else has seen, and to think what nobody else has thought."

- Albert Szent-Gyorgyi

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

To my family

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

AKNOWLEDGMENT

Desde cuando empecé a escribir la tesis, tenía claro que la parte de los agradecimientos iba a ser la más difícil de escribir, simplemente porque sabía que iba a ser imposible resumir todo lo que significaron estos años para mí, en pocas líneas. Dejé mi tierra y mi familia a los 18 años y entre todos los lugares donde viví, solo Galicia, y todas las maravillosas personas que cruzaron mi camino en estos 5 años, hicieron que volviera a sentir esa sensación de casa, de HOGAR, que había dejado en Cerdeña. Se dice que si elige un trabajo que te gusta, no tendrás que trabajar ni un día de tu vida, y este pensamiento me acompaño de manera constante en estos últimos años. Encontrar mi camino no fue fácil, me caí y me perdí bastantes veces. Pasé por momentos de desanimo donde quería dejarlo todo, y donde no me sentía capaz o a la altura del puesto que tenía, pero al final logré encontrar lo que me gustaba, lo que me apasionaba y lo que me daba el estímulo de levantarme y empezar el día.

Todo llega para quien sabe esperar, pero creo que si llegué hasta aquí hoy es solo gracias a todas las personas que compartieron conmigo este viaje, que me hicieron amar la ciencia y que me acompañaron y soportaron durante este proceso.

Quiero agradecer en primer lugar a mis directores de tesis Antonio Salas y Federico Martinon; gracias por vuestro continuo apoyo, por vuestras enseñanzas y por vuestro ejemplo que cada día me proporciona el espíritu y la curiosidad que mueven los científicos.

Gracias Toño, porque si no fuera por ti y por esa respuesta tan positiva, a la solicitud de Antonio Torroni, que en el lejano 2015 te preguntaba si querías una estudiante erasmus en tu laboratorio para unas prácticas, ahora no estaría aquí acabando mi doctorado. Fue desde esa primera experiencia en tu grupo que inició mi camino hacia la investigación. Fuiste el primero en creer en mí y siempre fuiste capaz con tu disponibilidad y tu profesionalidad, de animarme a seguir adelante. Gracias por enseñarme la paciencia y la perseverancia necesaria que hacen falta en la investigación.

Gracias Federico, por acogerme en tu grupo. Gracias por todas las nuevas ideas, por los retos y por el ejemplo constante de poder siempre con todo. Gracias por seguir creyendo en mí y en mis capacidades.

Quiero agradecer a ambos la oportunidad que me diste de hacer una estancia en Oxford. Algo desde allí cambió y permitió que viera con más claridad el camino que quería seguir.

Gracias a mis compañeros de viajes, Alberto, Jacobo, Miriam, Majo, Mirian, Patri, Sandra, Ruth, Xabi, Laura y Juli. Cada uno de vosotros hizo que mis días de trabajos fueran más ligeros y llevaderos. Gracias por vuestra experiencia, vuestros consejos y vuestro apoyo. Gracias por vuestro tiempo, vuestra amistad y por todos esos momentos de risas que hacen los días de trabajo mucho más agradables. Aunque nombré solo la parte del grupo del laboratorio, quiero agradecer a cada miembro del grupo GENVIP: sois mi segunda familia española y no podía desear compañeros de trabajo mejores.

Dados mis orígenes y dedicando este trabajo a mi familia, me gustaría para estos agradecimientos, poder utilizar mi lengua materna, el italiano.

Se mi avessero chiesto 10 anni fa di fare una previsione sul mio futuro, mai e poi mai avrei pensato di ritrovarmi qui, alla fine di uno dei percorsi piú intensi che io abbia mai vissuto e che mi ha segnata immensamente, facendomi maturare e finalmente credere in me stessa.

Tutto ció non sarebbe stato possibile senza il sostegno, l'appoggio e il supporto costante dei miei genitori. Grazie mamma, grazie babbo per l'amore incondizionato, grazie per tutti i sacrifici e le rinunce fatti con la speranza di dare ai vostri figli un futuro migliore. Se oggi sono riuscita a raggiungere i miei obbiettivi, lo devo principalmente a voi. Mi avete sempre stimolata e spronata a superare i miei limiti e ad aspirare al meglio. Non deve essere stato facile spingere due figli di soli 18 anni ad andar via di casa per cercare un futuro migliore. Grazie per essere stati a volte cosí duri, solo ora capisco quanto sia stato difficile per voi e quanto sia stato utile e produttivo per noi.

Grazie Marco, per essere un esempio costante! Sei e continuerai ad essere il mio punto di riferimento, il mio porto sicuro. Grazie per avermi insegnato e mostrato quanto sia importante essere felici e sentirsi liberi. Sono cosí orgogliosa di avere un fratello come te e anche se ci sono voluti anni per arrivare ad avere il meraviglioso rapporto che abbiamo, rifarei tutto da capo, ma sempre e solo con te al mio fianco. Tu, mamma e babbo siete la mia unica costante in un mondo fatto di tante variabili.

Grazie Francesca, non so come sarebbero stati questi anni senza la tua presenza. Ci sei stata sempre, nei momenti belli ma sopratutto in quelli piú duri, e continui a farlo riuscendo a rendere le mie giornate migliori. Mi ascolti fino allo sfinimento, trovi sempre le parole giuste per tirarmi su di morale, per strapparmi un sorriso e per farmi sentire meno sola. Con te ho la continua conferma che l'amicizia supera ogni limite, ogni distanza e ogni ostacolo.

Grazie a tutte le amiche e gli amici lontani che non smettono di far parte della mia vita nonostante la distanza. Grazie a tutti i nuovi amici e le nuove conoscenze che ogni giorno incrociano la mia strada.

Y por último, y no por importancia, gracias Jesús, mi compañero de vida. Gracias por seguir a mi lado, gracias por la felicidad que solo tú sabes regalarme con una sonrisa. Si ahora Galicia es mi hogar, es también gracias a ti.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

INDEX

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Figure legend

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Table legend

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

ABBREVIATIONS

5cC:	5-Carboxylcytosine
5fC:	5-Formylcytosine
5hmC:	5-Hydroxymethylcytosine
5mC:	5-Methycytosine
AD:	Alzheimer's Disease
aDMR:	Aging-specific Differentially Methylated Region.
AEC:	Airway epithelial cell
AID:	Autoimmune disorder
ALRI:	Acute Lower Respiratory Infection
ANOVA:	Analysis of Variance
APC:	Antigen-presenting cell
ASM:	Allele-specific Methylation
AUC:	Area under the curve
CCA:	Chimpanzee coryza agent
cDMR:	Cancer-specific Differentially Methylated Region
CGIs:	CpG Island
CH3:	Methyl group
Chr:	Chromosome
CLP:	Common Lymphoid Progenitor
CMP:	Common Myeloid Progenitor
CRC:	Colorectal Cancer
CR:	Conversion Reagent
DC:	Dendritic cell
DMPs:	Differentially Methylated Probes/Positions
DMR:	Differentially Methylated Regions
DNAm:	DNA Methylation
DNMTs:	DNA Methyltransferase

EBV:	Epstein-Barr Virus
EMA:	European Medicines Agency
ENCODE:	Epigenetics of DNA elements
EWAS:	Epigenome-Wide Association Studies
FANTOM:	Functional Annotation of the Mammalian Genome
FDA:	Food and Drug Administration
FDR:	False Discovery Rate
FSC:	Functional Class Scoring
GO:	Gene Ontology
GSEA:	Gene Set Enrichment Analysis
GWAS:	Genome-Wide Association Studies
HBV:	Hepatitis B Virus
HD:	Huntinghton's Disease
HIV:	Human Immunodeficiency Virus
HLA:	Human Leukocyte Antigens
HPV:	Human Papilloma Virus
HSC:	Hematopoietic Stem Cell
HSM:	Haplotype-specific methylation
ICRs:	Imprinting control regions
iDMR:	imprinting-specific differentially methylated region
IFN-γ:	Interferon-γ
Ig:	Immunoglobulin
IL:	Interleukin
ILC:	Innate Lymphoid cell
lncRNA:	long non-coding RNA
MAF:	Minor Allele Frequency
MERS-CoV:	Middle East Respiratory Syndrome - Coronavirus
methOR:	Methylation odds ratio
methQTLs:	Methylation quantitative trait loci
MHC:	Major Histocompatibility Complex
miRNA:	Micro RNA

MVP:	Methylation Variable Position
nAb:	Neutralizing Antibody
ncRNA:	non-coding RNA
NGS:	Next Generation Sequencing
NK:	Natural killer cell
NP:	Nasopharyngeal
NS:	Non-structural protein
OPA:	Opsonophagocytic Assay
ORA:	Over-Representation Analysis
ORF:	Open Reading Frame
PAH:	Pulmonary Arterial Hypertension
PCA:	Principal Component Analysis
PCGs:	Primordial Germ Cell
PCV:	Pneumococcal Conjugate Vaccine
PD:	Parkinson's Disease
Post-F:	Post-fusion
PPi:	Pyrophosphate Molecules
PPRs:	Pattern Recognition Receptors
Pre-F:	Pre-fusion
PTMs:	Post-Translational Modifications
QC:	Quality Control
RA:	Rheumatoid Arthritis
rDMR:	Reprogramming-specific Differentially Methylated Region
ROC:	Receiver Operating Characteristic
RRBS:	Reduced-representation Bisulfite Sequencing
RSV:	Respiratory Syncytial Virus
SAM:	S-adenyl methionine
SARS-CoV:	Severe Acute Respiratory Syndrome Coronavirus 1
SARS-Cov2:	Severe Acute Respiratory Syndrome Coronavirus 2
SBQ:	Subset Quantile Normalization
siRNA:	Short Interfering RNA

SLE:	System Lupus Erythemateosus
SNP:	Single Nucleotide Position
tDMR:	Tissue-specific Differentially Methylated Region
TGF-β:	Transforming Growth Factor-Beta
TET:	Ten-Eleven-Translocation
Th:	Helper T cell
TLR:	Toll-like Receptor
TNF:	Tumor Necrosis Factor
Treg:	Regulatory T cell
TSS:	Transcription Start Site
TT:	Tetanus Toxoid
UTR:	Untranslated Regions
VMR:	Variable Methylated Regions
WGBS:	Whole-Genome Bisulfite Sequencing
Xist:	X-inactive specific transcript

ABSTRACT

Epigenetics referred to all the heritable changes in expression (i.e. function) of genes that are not encoded in the DNA sequence and form a mechanism whereby past environmental events may be encoded and modify future cellular responses. Epigenetic mechanisms have repressive and permissive effects on gene expression, and thus play a key role in cellular differentiation, by activating/repressing different sets of genes. Epigenetic patterns are different between individuals and change over time influenced by intrauterine and postnatal environmental factors such as nutrition, toxins or drugs, and illnesses. Emerging evidence suggests a pivotal role of epigenetic in human pathologies and the regulation of the immune system in health and disease. It has been observed that the most important cell development processes are regulated by epigenetic changes. Many functional differences observed in the immune system of neonates and adults seem to be associated with epigenetic modifications of genes that control inflammation and immune response. In addition, changes in the immune system are thought to increase susceptibility to infection, meaning that the modulation of epigenetic mechanisms that govern the immune cell phenotype and function allows the external environment to influence the immune response outcome.

The immune system shows an unlimited capacity to respond to environmental stimuli. Its innate and adaptative potential allows the protection of the organism against danger, but also the maintenance of tolerance and the control of cellular integrity. The main feature of the immune system is the immunological memory, referred to as the capacity to rapidly respond with increased intensity upon secondary challenge with the same antigen, a mechanism that forms the basis of vaccination. In early life, the developing immune system presents a very low level of maturity that is reflected by an increased susceptibility to infections and a decreased vaccination response observed in infants. In addition to the multiple factors within the innate and adaptive arm of the immune system that have been identified contributing to this immaturity, the epigenetic mechanisms seem to be implicated and contribute to this regulation.

Epigenetics behaves as a dynamic interface between genome and environment. When an infection occurs, the aberrant epigenetic regulation of host defense cells associated is likely to play an important role in disease progression and development. Pieces of evidence support the implication of epigenetic mechanisms in the modulation of the interaction between host and

pathogen, as well as in the variable immune response observed after vaccination. In the last decade, it has been shown that epigenetic processes play a crucial role in host-pathogen interactions, and several studies demonstrate how pathogens have evolved mechanisms to exploit the host epigenetic system to improve their survival and maximize their transmission. It seems that histone modifications and chromatin remodeling, as well as DNA methylation, are the first key target during bacterial and viral infection.

One of the most studied epigenetic mechanisms is DNA methylation that involves the addition of methyl groups to the DNA molecule; it is considered the major epigenetic factor influencing gene activities. In recent years, changes in genomic DNA methylation in response to bacterial and viral infections have been described. However, there are very little evidence and studies on the role of the epigenetic regulation of immune response to infection or vaccine in early childhood. What is more, there is very limited knowledge about how the epigenome changes over time, as well as which effects different antigenic stimuli can have on the epigenome.

The new field of epigenomics, focused on the study of all epigenetic modifications that occur in a cell, is starting to be used as a discovery instrument to better understand the epigenetic contributions to human diseases. The epigenomic identification of qualitative and quantitative measures of epigenetic changes is likely to exert a higher level of biological characterization of potential disease predisposition biomarkers and provide great potential for advancing preclinical knowledge and predict disease outcomes.

The main aim of the present thesis is to provide insights on the role of epigenetics processes, specifically the role of DNA methylation, on the relationship between infection and host disease outcomes, as well as on the investigation of the immune system response to vaccination that leads to a highly variable response in children. What is more, the project aims to elucidate epigenetic changes that occur over time in early childhood.

Specifically, to explain the role of the epigenome on the modulation of immune response after infection, the respiratory sequelae observed after RSV infection are investigated considering a study cohort of children affected by RSV and exhibiting different types of recovery after the infection. To assess the role of epigenetic mechanisms in the modulation of the different immune responses observed after vaccination, instead, a study cohort of healthy infants that received a booster of PCV13 was analyzed. Of these groups of children, age-associated DNA methylation changes were also studied testing two different timepoints.

To achieve all the purposes of the study the Infinium Illumina Methylation BeadChip platform was implemented. This new cutting-edge technology, ideal for epigenome-wide association studies, shows a high-quality genome-wide coverage that allows access to single CpG methylation status located at various regulatory regions throughout the human genome. With this technology, it was possible to interrogate the methylation level of more than 850,000 positions distributed in the whole genome.

The results obtained by these studies revealed an important contribution of changes in DNA methylation in the regulation of several immune processes which drive the differential immune response observed in children after foreign stimuli.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

RESUMEN (GALEGO)

Introdución: A epixenética refírese a todos os cambios herdados na expresión de xenes que non están codificados na secuencia de ADN e forman un mecanismo polo cal os eventos ambientais pasados poden codificarse e modificar as respostas celulares futuras. Os mecanismos epixenéticos teñen efectos represivos / inhibitorios e permisivos na expresión xénica e, polo tanto, xogan un papel fundamental na diferenciación celular, ao activar / reprimir diferentes conxuntos de xenes. Os patróns epixenéticos son diferentes entre os individuos e cambian co paso do tempo, influídos por factores ambientais intrauterinos e posnatais como a nutrición, as toxinas ou as drogas e enfermidades. A metilación do ADN é un dos mecanismos epixenéticos máis estudados e implica a adición de grupos metilo á molécula de ADN; considérase o principal factor epixenético que inflúe nas actividades xénicas e rexistráronse cambios na metilación do ADN xenómico nos últimos anos en resposta a infeccións bacterianas e virais. Non obstante, hai moi poucas evidencias e estudos sobre o papel da modulación epixenética da resposta inmune á infección ou vacinación na primeira infancia. Ademais, hai un coñecemento moi limitado sobre como o epixenoma cambia co paso do tempo ou sobre o efecto que diferentes estímulos antixénicos poden ter no epixenoma.

Dado que os cambios epixenéticos son necesarios para o desenvolvemento e a saúde normais e están asociados a procesos esenciais implicados na fisioloxía celular, non é de estrañar que, cando se alteren, poidan ser responsables dalgúns estados da enfermidade. A interrupción de calquera dos tres sistemas que contribúen a alteracións epixenéticas pode causar unha activación anormal ou silenciamento dos xenes. Estas interrupcións asociáronse especialmente co cancro, pero tamén con síndromes que implican inestabilidades cromosómicas e retraso mental, así como enfermidades neurolóxicas e autoinmunes. O novo campo da epixenómica, que consiste no estudo de todas as modificacións epixenéticas que se producen nunha célula, pode usarse como ferramenta de descubrimento para comprender mellor as contribucións epixenéticas ás enfermidades humanas. É probable que a identificación epixenómica de medidas cualitativas e cuantitativas dos cambios epixenéticos exerza un maior nivel de caracterización biolóxica dos potenciais biomarcadores de predisposición da enfermidade e proporcione un gran potencial para avanzar no coñecemento preclínico e predicir os resultados da enfermidade.

Evidencias recentes suxiren un papel fundamental para a epixenética na regulación do sistema inmune na saúde e nas enfermidades. Os procesos de desenvolvemento celular máis importantes observáronse regulados por cambios epixenéticos. Moitas diferenzas funcionais observadas no sistema inmunitario de recen nados e adultos parecen estar asociadas a modificacións epixenéticas de xenes que controlan a inflamación e a resposta inmune. Ademais, parece que os cambios no sistema inmunitario aumentan a susceptibilidade á infección, o que significa que a modulación dos mecanismos epixenéticos que rexen o fenotipo e a función das células inmunes permiten que o ambiente externo inflúa no resultado da resposta.

O sistema inmune mostra unha capacidade ilimitada para responder aos estímulos ambientais. O seu potencial innato e adaptativo permite a protección do organismo contra o perigo, pero tamén o mantemento da tolerancia e o control da integridade celular. A principal característica do sistema inmunitario é a memoria inmunolóxica, coñecida como a capacidade de responder rapidamente con maior intensidade a unha segunda exposición co mesmo antíxeno, un mecanismo que constitúe a base da vacinación. Nos primeiros anos de vida, o sistema inmune en desenvolvemento ten un nivel de madurez moi baixo, o que se reflicte nunha maior susceptibilidade e vulnerabilidade ás infeccións e unha menor resposta á vacinación. Ademais dos múltiples factores do sistema inmune innato e adaptativo identificados e que contribúen a esta inmadurez, os mecanismos epixenéticos están implicados e contribúen a esta regulación.

De feito, a epixenética compórtase como unha interface dinámica entre o xenoma e o medio ambiente e, polo tanto, no caso dunha infección, a regulación das células de defensa do hóspede está directamente relacionada co desenvolvemento da enfermidade. Hai evidencias que apoian a implicación de mecanismos epixenéticos na modulación da interacción entre o hóspede e o patóxeno, así como na resposta inmune variable observada despois da vacinación. Na última década, demostrouse que os procesos epixenéticos xogan un papel crucial nas interaccións hóspede-patóxeno e varios estudos demostran como os patóxenos desenvolveron mecanismos para explotar o sistema epixenético do hóspede para mellorar a súa supervivencia e maximizar a súa transmisión. Varios estudos da literatura revelaron que moitos virus teñen a capacidade de manipular marcas epixenéticas do hóspede, e isto probablemente contribúa ao establecemento da latencia e algún papel patóxeno. Entre eles, o virus respiratorio sincitial (VRS) demonstrou a capacidade de modular a epixenética do hóspede, levando a unha regulación aberrante da expresión xénica.

VRS é un patóxeno común que infecta practicamente a todos os nenos aos dous anos de idade e é a principal causa mundial de hospitalización de bebés. É a principal causa de infeccións respiratorias inferiores agudas (ALRI) en nenos pequenos e está asociada a morbilidade e mortalidade na infancia. Nunha revisión sistemática de 2015, estimouse que aproximadamente 34 millóns de novos episodios de ALRI en nenos en todo o mundo eran atribuíbles a VRS, un gran número que resultou en 3,2 millóns de ingresos hospitalarios e case 60.000 mortes infantís mundiais cada ano. Desafortunadamente, aínda non hai ningún tratamento eficaz nin vacinas.

Ademáis da carga aguda do VRS, un número crecente de evidencias dos datos da epidemioloxía apoia que a infección por VRS nos primeiros 3 anos de vida pode estar directamente correlacionada con morbilidades respiratorias a longo prazo, como sibilancias e asma recorrentes. Obsérvase que o VRS, como outros virus respiratorios, causa un fenómeno "hit and run", caracterizado polo aumento do risco de desenvolver sibilancias e asma recorrentes na infancia despois da infección, como fenotipo permanente que persiste moito despois da eliminación do virus. O sibilancia é o típico son agudo e sibilante que se produce durante a respiración. O asma, por outra banda, caracterízase por anomalías na función pulmonar que inclúen unha obstrución variable das vías aéreas e unha reactividade bronquial aumentada. Xeralmente, recoñecer o asma adoita ser obvio e a maioría das veces os pacientes asmáticos informan de episodios de sibilancias; con todo, é moi difícil predicir e distinguir que nenos presentarán síntomas só na vida temperá, dos que presentarán - síntomas persistentes e que poden desenvolver sibilancias definitivas ou asma.

O risco de sibilancias e / ou incidencia de asma estivo cada vez máis relacionado cunha combinación de factores xenéticos e ambientais, así como coa gravidade da infección respiratoria. Fixéronse varios esforzos para establecer as causas últimas da relación entre a infección e as secuelas respiratorias posteriores, pero aínda son necesarios outros estudos complementarios para comprender o mecanismo subxacente.

Como se describiu anteriormente, a inmadurez observada no sistema inmunitario no inicio da vida aumenta, non só, a vulnerabilidade dos bebés ás enfermidades infecciosas, senón que tamén reflexa a súa reducida resposta inmune á vacinación.

A pesar dos inmensos avances científicos sobre o papel dos cambios epixenéticos nos procesos infecciosos, non hai tantos coñecemento sobre o impacto das vacinacións no panorama epixenético, así como do papel dos cambios epixenéticos en resposta ás vacinas.

Na última década, varios estudos centráronse na heteroxeneidade observada nas respostas á vacina, un factor crucial, con importantes implicacións para a distribución da protección nos individuos e para o alcance da inmunidade de rabaño alcanzada cando a vacina é amplamente utilizada nunha poboación. As variacións observadas nas respostas á vacina teñen consecuencias tanto para a eficacia protectora como para a duración da protección. Demostrouse que durante a vacinación infantil un número elevado de nenos está a acadar niveis de anticorpos de máis de 100 veces superiores á inmunización que outros. A razón desta enorme variación non se comprende completamente. Os principais factores que inflúen nas respostas da vacina humoral e celular nos humanos inclúen factores intrínsecos (como a idade, o sexo, a xenética e as comorbilidades) e os factores extrínsecos (como a microbiota, as infeccións e os antibióticos). Os factores ambientais, así como os factores de comportamento e nutrición tamén xogan un papel na forma en que os individuos responden ás vacinas.

As infeccións pneumocócicas son un importante problema de saúde pública en todo o mundo que causan morbilidade e mortalidade elevadas en nenos pequenos que sofren pneumonía, meninxite e septicemia, provocando un millón de mortes en nenos menores de cinco anos. Crese que os nenos pequenos, sendo os pacientes máis afectados, son os principais transmisores destas infeccións en toda a poboación. Polo tanto, manter niveis axeitados de anticorpos nesta categoría de poboación é a clave para bloquear a transmisión do organismo e alcanzar unha protección completa.

As vacinas contra os polisacáridos cápsulares do neumococo que se empregan xeralmente para inmunizar adultos non se atoparon nin inmunoxénicas nin protectoras en nenos pequenos debido ás malas respostas dos anticorpos. Polo tanto, na última década, a investigación centrouse no desenvolvemento de vacinas inmunóxenicas adicionais contra o neumococo para proporcionar inmunidade a longo prazo en nenos de menos de 2 anos. O enfoque máis prometedor centrouse no desenvolvemento dunha vacina conxugada proteína-polisacárido para diferentes serotipos neumocócicos que causan a maioría das infeccións nos nenos. As vacinas conxugadas contra o neumococo (PCV) son altamente eficaces na prevención de infeccións neumocócicas e obsérvase que proporcionan inmunidade rápidamente e durante o segundo ano de vida.

Obxectivos: Debido á importancia do sistema inmunitario na primeira liña de defensa contra as infeccións virais e tamén á falta de estudos centrados na modulación da resposta

inmune na primeira infancia, o principal obxectivo do presente estudo é proporcionar información sobre o papel dos procesos epixenéticos, en concreto o papel da metilación do ADN, na relación entre a infección e os resultados da enfermidade do hóspede, así como na modulación da resposta do sistema inmunitario á vacinación.

En concreto, para abordar o obxectivo xeral do proxecto, definíronse catro obxectivos específicos que se detallan a continuación.

• Avaliar os mecanismos de regulación xénica a nivel de metilación que están implicados nos fenotipos diferenciais que se poderían manifestar despois da infección por VRS.

• Atopar unha relación entre os cambios de metilación despois da infección por VRS documentada e o desenvolvemento posterior de sibilancias e / ou asma.

• Avaliar os cambios de metilación que se producen no inicio da vida para identificar vías que mostran cambios significativos ao longo do tempo.

• Elucidar as diferenzas de metilación entre as persoas que responden con diferente magnitude á vacina de reforzo PCV13 para buscar biomarcadores epixenéticos asociados a unha resposta inmune máis robusta.

Material y Métodos: Para acadar todos os propósitos do estudo, implementouse a plataforma Infinium Illumina Methylation BeadChip.

Antes de empezar co microarray de Illumina, as mostras de sangue recolléronse nun tubo EDTA, procesáronse e almacenáronse a 80º despois da recollida. Despois da extracción de ADN, avaliáronse as medidas de pureza e concentración de ADN con Nanodrop xunto co ensaio PicoGreen. O seguinte paso foi a conversión de bisulfito de ADN xenómico de mostras de sangue completo. Esta técnica considérase un enfoque estándar para investigar o estado de metilación dos sitios de CpGs en todo o xenoma. Basicamente consiste na conversión de residuos de citosina non metilada en uracilo mediante desaminación utilizando bisulfito de sodio, deixando a citosina metilada (5-mC) sen afectar. Deste xeito, os uracilos amplifícanse na reacción posterior de PCR como timinas, mentres que os residuos de 5-mC ou 5-hmC seguen sendo citosinas.

O ADN tratado foi entón hibridado co Illumina Infinium Methylation BeadChip e a matriz foi imaxinada usando o sistema Illumina iScan no que se cuantificou o porcentaxe de estado de metilación de cada sitio CpG para todo o grupo de estudo.

Esta nova tecnoloxía de vangarda, ideal para estudos de asociación de todo o epixenoma, mostra unha cobertura de todo o xenoma de alta calidade que permite acceder ao estado de metilación de cada CpG situado en varias rexións reguladoras de todo o xenoma humano. Con esta tecnoloxía foi posible analizar o nivel de metilación de máis de 850.000 posicións distribuídas por todo o xenoma. Os ficheiros de intensidade bruta preprocesáronse e transformáronse en valores β e M usando os paquetes R. As intensidades dos valores β estimáronse a partir da relación de intensidade dos sinais metilados sobre os sinais totais (metilados e non metilados) para cada posición xenómica, o que representa a porcentaxe de metilación das células sanguíneas nunha determinada citosina e para un único individuo.

Despois do control de calidade, os datos brutos normalizáronse e seguiron un proceso de filtrado e corrección do efecto causado pola variabilidade técnica. Despois de estimar a composición celular, a identificación de sondas ou posicións diferencialmente metiladas (DMPs) e de rexións diferencialmente metiladas (DMRs) entre os diferentes grupos do estudo realizouse mediante un modelo de regresión lineal. Finalmente realizouse a análise de enriquecemento de rutas biolóxicas potencialmete implicadas.

Resultados: A primeira parte do estudo céntrase en interrogar os perfís de metilación do ADN de todo o xenoma de 77 bebés infectados por VRS e agrupados en controis (nenos que despois do VRS tiveron unha recuperación normal) e casos (nenos que despois de VRS desenvolveron secuelas respiratorias definidas como sibilancias ou asma) para comprender como a epixenética inflúe nas distintas secuelas respiratorias desenvolvidas despois da infección por VRS. Despois de axustar o modelo lineal para a composición de tipo celular, idade e sexo, obtivéronse 5.097 DMPs estadísticamente significativas entre controis e casos (FDR <0,01), que corresponderon a 3.278 xenes únicos. Entre os DMPs máis significativos identificáronse dous CpG interesantes localizados no lugar de inicio da transcrición (TSS1500) dos xenes EYA3 e o DDX27, e que mostran un patrón de metilación reducida no grupo de nenos con secuelas respiratorias cando se comparan cos que teñen unha normal recuperación despois da infección por VRS. O cambio de metilación observado nestes xenes, ambos implicados en procesos inflamatorios, podería revelar unha posible desregulación xénicos en bebés que manifestan secuelas respiratorias. Para abordar a posible significación biolóxica do impacto dos CpG diferencialmente metilados, levouse a cabo un análise do enriquecemento do conxunto de xenes e de rutas biolóxicas . As categorías con maior representación foron aquelas relacionadas

con varios procesos do ciclo celular, como o punto de control do ciclo celular, o dano no ADN e o punto de control de integridade do ADN.

Para a segunda parte do estudo, 24 nenos que aos 12 meses de idade recibiron a vacina conxugada contra o neumococo (PCV13), como parte dun ensaio clínico previo, seleccionáronse en función da resposta de anticorpos IgG e clasificáronse en altos ou baixos respondedores. As mostras de sangue tomadas aos 12 e 24 meses de idade foron avaliadas mediante análise de microarrays epixenéticos de ADN. O primeiro paso da análise centrouse nos cambios de metilación entre os dous puntos de tempo (12 vs. 24 meses). Despois de corrixir o modelo lineal para a composición de tipo celular, o xénero e a resposta á vacina, realizouse unha análise emparellada que revelou 721 DMPs, asociados a 421 xenes únicos, estatisticamente significativos entre os 12 e os 24 meses de idade. O resultado da análise de rutas biolóxicas revelou un enriquecemento significativo en 151 vías de "Gene Ontology"(http://geneontology.org/), sendo as vías máis significativas aquelas relacionadas coa regulación da activación das células T, a adhesión célula-célula e a regulación da produción de citoquinas.

O seguinte paso da análise realizouse para avaliar se as diferenzas nas respostas dos anticorpos fronte aos antíxenos da vacina se asociaron aos patróns de metilación, comparando os perfís do epixenoma sanguíneo previos á vacinación dos respondedores altos cos dos respondedores baixos, no tempo basal (12 meses). Neste caso non se detectaron DMP estadísticamente significativas. Non obstante, empregando un umbral de significación menos restrictivo (valor *P* non corrixido <0,01), detectaronse 4.067 sitios CpG diferencialmente metilados.

De estes DMPs, seleccionamos as posicións que mostraban as maiores diferencias de metilación entre respostas altas e baixas á vacina e centrámonos nas rexións que presentaban múltiples DMPs en xenes implicados nos procesos do sistema inmunitario. Así, foi posible observar 4 posicións dentro do xene HLA-DPB1 que presentaban unha metilación reducida en persoas con unha alta resposta á vacina y además dúas CpGs asociadas coa transcrición xénica da interleucina 6 (IL6). Estas últimas mostraron unha hipermetilación nos nenos con respostas altas á vacina, con respecto aos nenos que responderon á vacina de forma máis baixa. A hipometilación de HLA-DPB1, en pacientes con unha alta resposta á vacina, podería fortalecer a resposta inmune total como se demostrou previamente na contribución do HLA á inmunidade celular do hóspede en resposta a outras vacinas. Sen embargo, as posicións hipermetiladas

dentro do locus IL6 asociadas con altos respondedores están anotadas no xenoma como posibles lugares de regulación da transcrición xénica "enhancers" polo que poderían estar modulando a expresión de IL6, unha citoquina proinflamatoria que participa nunha amplia variedade de procesos inmunolóxicos e que además tamén xoga un papel importante na mediación da resposta inmune innata e adaptativa. Polo tanto, a expresión diferencial de IL6 podería afectar á respostas inmune á vacina.

Conclusions: O presente estudo ofrece novas ideas sobre o papel da metilación do ADN na regulación do sistema inmunitario na primeira infancia. A resposta á infección natural e á vacinación pode levar a secuelas da enfermidade ou a unha mala resposta á vacina, respectivamente, debido á inmadurez do sistema inmunitario dos nenos e as nenas. Os resultados incluídos nesta estudio proporcionan evidencias sobre como a epixenética pode contribuír ás diferentes secuelas observadas despois das infeccións, así como á resposta diferencial á vacinación. A detección de CpGs específicas localizadas en xenes potencialmente implicados no desenvolvemento de secuelas respiratorias despois da infección, ofrece a posibilidade de identificar biomarcadores epixenéticos específicos que se poden emplear na clínica para predicir con maior precisión diferentes condicións patolóxicas despois da infección por VRS. Do mesmo xeito, o descubrimento de sitios específicos de CpGs asociados coa resposta inmune á vacinación suxire que os patróns epixenéticos poden influír na produción de anticorpos e poderían usarse para avaliar a inmunidade frente á vacinación PCV.

A pesar de que son necesarias investigacións adicionais con cohortes máis grandes para axudar a esclarecer con maior profundidade o papel da epixenómica na inmunidade, este estudio ofrece, a nivel global, fortes evidencias sobre o papel das modificacións epixenéticas na regulación da resposta inmune dos nenos relacionada tanto coa infección natural como coa vacinación.

1
INTRODUCTION

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

1.1 EPIGENETICS

"The branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being". This is the first definition of Epigenetics, whose term was coined by Waddington in 1942 [1]. In line with this, epigenetics is considered in a wide sense as a bridge between genotype and phenotype and Waddington does not refer to epigenetics as alterations in genetic inheritance, rather in the "epigenetic landscape"[2], with the idea that distinct epigenetic patterns govern gene expression and focusing on the role of epigenetics in embryonic development. Literally, epigenetics means "beside or on the top of genetics" and throughout the years, this definition has evolved since it indicated a phenomenon implicated in a wide variety of biological processes. Arthur Riggs in 1996 depicted epigenetics as "the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in the DNA sequence"[3], leaving an open interpretation of how it works and highlighting that epigenetics is not the inheritance of mutational changes. Regardless of the exact definition, today when we speak about Epigenetics, we refer to all heritable changes that occur in gene expression without changing the DNA sequence.

The epigenome, considered as the set of all epigenetic modifications associated with genomic DNA, integrates the information encoded in the genome with all the molecular and chemical signals of cellular, and extracellular origin. The epigenome allows the organism to adapt and evolve expressing a set of features or phenotypes that develop in response to environmental stimuli. Therefore, the epigenome, unlike the genome, is characterized by a dynamic and flexible response to intra- and extra-cellular stimuli, through cell-cell contact, or entirely by the environment to which the organism is exposed. To understand the mechanisms and cellular pathways that lead the effect of an epigenetic factor to create global or specific epigenetic changes, Kanherkar et al., in 2014 hypothesized two mechanisms: the first one, based on a direct effect, proposes that the epigenetic factor directly alters the state of epigenetic enzymes, or directly causes a change in a biochemical process that results in an altered availability of a substrate, or a specific metabolite required for a particular epigenetic tag. The second hypothesis, based on an indirect model, suggests that the exposure to a factor influences cellular signaling pathways bringing to the altered expression of growth factors, receptors, and ion channels, which in turn modify transcription factor activity at gene promoters. With more

chronic exposure, the transcription factors and other gene regulatory proteins could recruit epigenetic enzymes resulting in the addition or removal of epigenetic tags. In this way cells adapt to the persistent gene expression changes by causing permanent modifications, leading to enduring alteration of the affected epigenetic network [4].

Each cell in the organism presents an identical and stable genotype, while the phenotype manifested by an organism is flexible and variable because of the changes of gene expression in response to environmental cues. In the cell, epigenetic mechanisms play a function of control regulating gene expression and silencing. This control varies along with tissues and is crucial in cell differentiation [5]. Additionally, differences in gene expression between cells, which are driven by epigenetic modifications, result in the unique function of specific cell types [6].

Three main different epigenetic mechanisms that control gene expression have been identified: DNA methylation, histone modification, and non-coding RNA (ncRNA)-associated gene silencing.

1.1.1 Post-translational modifications of histones

Histone modifications, also called post-translation modification (PTMs) of histone proteins are the most variable epigenetic mechanism and mainly regulate the accessibility of chromatin for the transcription machinery. Histones are the small proteins that are found together with DNA to form in the nucleus the chromatin structure, which is made of histone units called nucleosomes. Each nucleosome presents an 8-subunit histone core (H3, H4, H2A, H2B) around which 147 base pairs of DNA are wrapped [7, 8].

Two major states can be adopted by the chromatin: heterochromatin that is a compact form resistant to the binding of various proteins, such as the transcriptional machinery; and euchromatin, a relaxed form of chromatin that is accessible to modifications and transcriptional processes. Until 1990 it was thought that histones have only a structural role without having any role in gene regulation [9], then it has become very clear that histones not only play an active role in gene regulation but also in DNA damage repair, in DNA recombination, and DNA replication, representing a crucial factor of epigenetic regulation in cell differentiation and growth [10].

The deposition of a specific modification on the histone tail is thought to specify a ‘code’ that dictates the regulatory features of a gene [10]. Most of the histones modifications categories

are flexible and dynamic and are able to appear and disappear on chromatin within minutes of stimulus arriving at the cell surface. Among the most known there are acetylation, methylation, and phosphorylation [11] (Figure 1).

Generally, histone modifications are associated with active transcription, and this is the case of the acetylation of H3 and H4, commonly referred to as euchromatin modifications; instead, modifications that occur to inactive genes or regions, such as H3 K9me and H3 K27me, are often defined heterochromatin modifications.

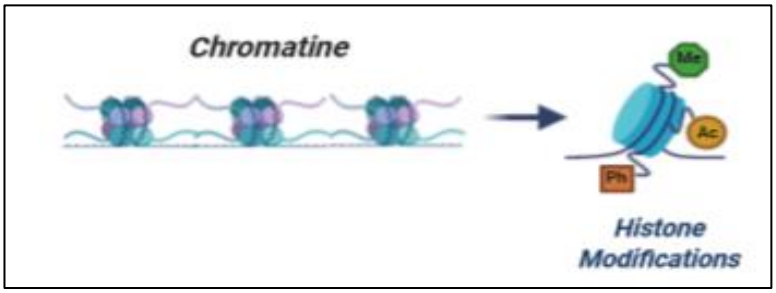


Figure 1: The schematic representation of histone modifications. Self-made figure. Created with BioRender.

Histone modifications are mainly distributed in distinct localized patterns within the upstream region, the core promoter, the 5′ end of the open reading frame (ORF), and the 3′ end of the ORF. The location of the modifications is fundamental for the effect of that change on the transcription [12].

Histone acetylation, common in lysine residues, is the most studied histone PTMs, and is a reversible modification that has an activating or repressing role depending on its acetylation state; its effect is observed not only at the level of chromatin structure but particularly on the transcription regulatory protein binding DNA, leading to a direct modulation and regulation of gene expression.

Similar to DNA methylation, histone methylation involves the addition of a methyl group on three main basic residues that are lysines, arginines, and histidines. It has been proposed that histone methylation confers both stability and reversibility of gene regulation to an organism, suggesting that to maintain normal biological function an appropriate balance between stable and dynamic histone methylation is necessary [13].

Histone phosphorylation takes place, predominantly, but not exclusively, on serines, threonines, and tyrosines. It has been observed that phosphorylated residues are involved in chromatin condensation associated with mitosis and meiosis, as well as in chromatin relaxation linked to transcription activation [14].

1.1.2 RNA-mediated gene silencing

The most recent epigenetic mechanism that is emerging as an important regulator of gene expression and is arousing a great interest as an epigenetic modifier is the RNA, from the small to the long non-coding RNA (lncRNA). Despite most of the human genome is transcribed, only around 2% of these transcripts are translated into proteins. The remaining transcripts, collectively called non-coding RNAs, can be divided into two groups according to their length: short RNAs less than 200 nucleotides in length and the larger transcripts defined lncRNA. Many of these ncRNAs are involved in the regulation of gene expression and have been linked to a wide range of molecular biological processes, that go from nucleus-specific roles in modulating transcription, epigenetic modification, and DNA looping, to involvement in alternative splicing and RNA-RNA network cross-talk [15, 16].

It has been demonstrated that ncRNAs can promote gene silencing by regulating the structure of the chromatin [17-19]. For instance, the non-coding RNA, X-inactive specific transcript (Xist), plays a key role in silencing the inactive X chromosome in females. More recently, short ncRNAs, including short interfering RNA (siRNA) and microRNA (miRNA) molecules, which target homologous genes for silencing, have been identified. The miRNA molecules act silencing and inhibiting gene expression by promoting the degradation of the controlled messenger RNA (mRNA) or restricting translation initiation and activation under certain conditions. These miRNAs, even though represent only 1% of the genome, are found to target 30% of genes [20]. Evidence from studies in plants demonstrated that these siRNAs are capable to silence homologous DNA regions recruiting DNA methyltransferases and histone-modifying enzymes, although it is not yet clear whether this occurs similarly in mammalian systems. Furthermore, a recent controversial finding suggests that these short ncRNAs may also be able to activate some genes [21].

1.1.3 DNA methylation

DNA methylation (DNAm) is one of the most studied, stable, and accessible epigenetic mechanisms. Discovered in mammals in the 40s during the identification of DNA as the genetic material, it is considered the major epigenetic factor influencing gene expression, being involved in cell differentiation and gene regulation. DNA methylation depicts the addition of a

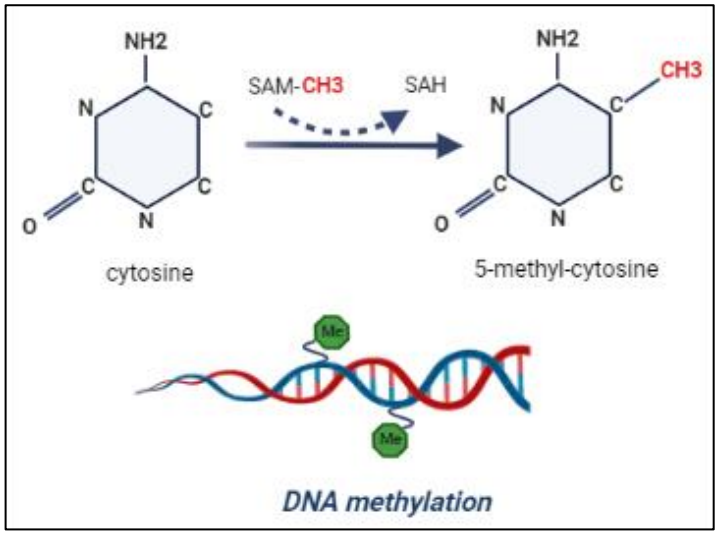


Figure 2: The schematic representation of DNA methylation. Self-made figure. Created with BioRender.

methyl group (CH3) from an S-adenyl methionine (SAM) to the fifth carbon of a cytosine residue to form 5-methyl-cytosine (5mC) (Figure 2).

DNA methylation, like the other epigenetic marks, is heritable through cell division, thus it is an important component of the cellular memory mechanism that maintains cell identities. However, it needs reprogramming to its totipotent state to produce the next generation. The process of demethylation and remethylation during the mammalian life cycle can occur during two periods: during germ cell development or early embryogenesis [22, 23]. The epigenetic reprogramming in the germline involves the deletion of somatic methylation patterns in primordial germ cells (PGCs) and the subsequent establishment of sex-specific germ cell methylation patterns, including methylation marks in imprinting control regions (ICRs). On the contrary, epigenetic reprogramming in early embryogenesis involves the erasure of most methylation marks inherited from the gametes at preimplantation stages and the re-establishment of global DNA methylation patterns upon implantation [24].

The process of DNA methylation is strictly regulated by several molecules among which DNA methyltransferases (DNMTs) are the most important. Three main different DNA methyltransferase, essential for mammalian development, are involved in the DNA methylation process: Dnmt1 whose function is to copy the DNA methylation pattern from the parental DNA strand onto the newly synthesized daughter strand during DNA replication, and Dnmt3a and Dnmt3b that are known as "*de novo* DNMTs" and establish a new methylation pattern to unmodified DNA (Figure 3).

DNA methyltransferases are extensively involved in the development of the embryo and when cells are completely differentiated, the expression of the DNMTs is remarkably reduced. They play an important role in maintaining genomic integrity, which damage or disruption may bring to chromosome instability and tumor progression. It is well established that DNMTs are

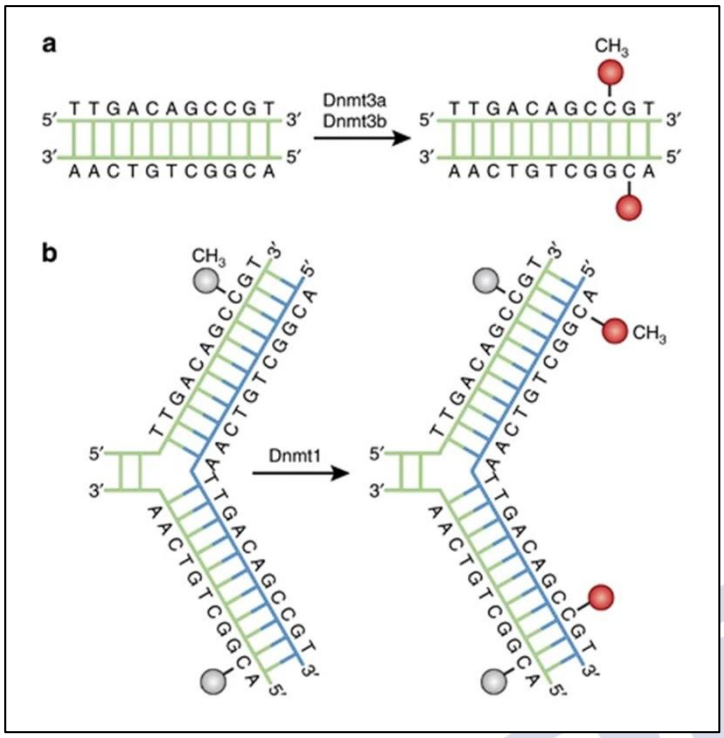


Figure 3: DNA methylation pathways: (a) Dnmta and Dnmt3b establish a new methylation pattern to unmodified DNA, while (b) Dnmt1, during DNA replication, copies the DNA methylation pattern from the parental DNA strand onto the newly synthesized daughter strand. DNA Methylation and Its Basic Function. Moore et al. Publisher: Springer Nature Adaptations/modifications - Springer Nature allows adaptation of figures for style and formatting purposes under this license under the condition that this does not alter the meaning of the content.

required for maintaining chromosome stability, through the transcriptional silencing of imprinted genes, and the silencing of genes on the inactive X chromosome and transposable elements. The target sites of DNMTs are often palindromic sequences, that can exist in three methylation states: unmethylated, where both strands of the DNA are not methylated, hemimethylated, where only one strand of the DNA is methylated, and fully methylated, with both strands presenting methylation.

Dnmt1 is the most abundant methyltransferase in adult cells. Its expression is activated by cell cycle-dependent transcription factors in the S phase and thus it is highly expressed in most mitotic cells. Dnmt1 is specific for cytosines restricted to CpGs sites and prefers hemimethylated CpGs (DNA with only one stand methylated) over unmethylated substrates.

After DNA replication, the newly synthesized strand is not methylated, therefore Dnmt1 binds to these hemimethylated CpG sites and methylates the cytosine on the newly synthesized strand. This maintains established CpG methylation patterns through mitosis.

De *de novo* DNMTs, Dnmt3a, and Dnmt3b were identified for the first time in 1989 [25] and have the same affinity for hemimethylated and no-methylated DNA. They can also modify cytosines that are not localized in a non-CG context, such as CA, CC, and CT, despite the preference for CpG sites. It has been demonstrated that Dnmt3a is maternally provided and predominates in oocytes and early preimplantation embryos. Because of its origin, it plays a crucial role in the establishment of the differential DNA methylation patterns in male and

female gametes [26]. Dnmt3b, instead, is robustly expressed by the blastocyst stage [27]. Both Dnmt3a and Dnmt3b are essential for early development, and it is suggested that their loss or deletion causes embryonic lethality [28]. A third *de novo* DNA methyltransferase, called Dnmt3l, has been identified as an important activating cofactor of Dnmt3a/b, although its absence of enzymatic activity. Indeed, it lacks the characteristic N-terminal catalytic domain but it can increase 15-fold the catalytic activity of Ddnmt3a and Dnmt3b [29].

The maintenance vs. *de novo* function of these enzymes is not absolute; in fact, Dnmt1 can function also as a *de novo* DNA methyltransferase and its overexpression leads to *de novo* methylation of CpG islands [30]. Similarly, Dnmt3a and Dnmt3b can perform the role of a maintenance DNMT.

In the last few years, evidence has emerged on the role of these enzymes as demethylases. Indeed, mammalian Dnmt1, Dnmt3a, and Dnmt3b can convert 5mC to cytosine, in an enzymatic process $Ca2^{+}$ ion-dependent [31], that consists essentially of the enzymatic reversion of the Cytosines-5mC through demethylation. This kind of active DNA demethylation is mediated by TET enzymes, members of the Ten-Eleven Translocation (TET) family, which induce the oxidation of 5mC to 5-hydroxymethylcytosine (5hmc), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC).

1.1.3.1 CpG Island, Shore, Shelf, and OpenSea regions

Although the brain contains the highest levels of DNA methylation of any tissue in the body, the methylated cytosine only accounts for ~1% of nucleic acids in the human genome [32]. This methylation occurs predominantly in repetitive elements and regions in which the density of CpG is low. Conversely, CpG-rich regions called ‘CpG islands’ (CGIs), which are stretches of DNA of around 1,000 base pairs long, have a higher CpG density than the rest of the genome but often are free of methylation [33].

Most CGI reside within the majority of gene promoters, approximately 70% [34], such as the promoters for housekeeping genes that are often embedded in CGI [35]. CGIs, especially those associated with promoters, are highly conserved between mice and humans, and their location and preservation throughout evolution imply functional importance for these regions. Although ~50% of CpG islands contain known transcription start sites, CGIs are often devoid of common promoter elements such as TATA boxes [36]. Since many transcription factors

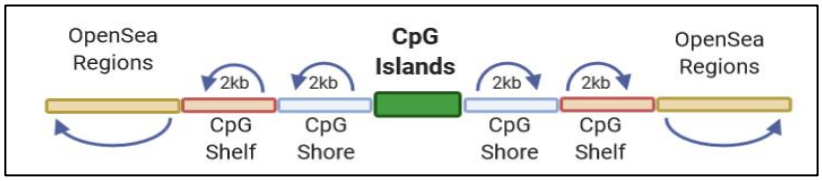


Figure 4: Schematic of CpG annotation in the CpG island context.
Self-made figure. Created with BioRender.

binding sites are GC rich, CpG islands are likely to enhance the accessibility of DNA and promote transcription factor binding. In addition to CGI, other regions exist in the genome where DNA methylation occurs with high frequency. The closest regions to CGI are called CpG island shores, located as far as 2 kb from CpG islands. They have highly conserved patterns of tissue-specific methylation and like CpG islands, their methylation is highly correlated with reduced gene expression [37, 38]. Next, it can be possible to observe the CpG island shelf, found 2 kb regions upstream and downstream of the CpG island shores, and the open sea regions, considered isolated CpG sites in the genome that do not have a specific classification (Figure 4).

1.1.3.2 The role of DNA methylation in gene expression

In the analysis of DNA methylation changes, two important concepts need to be taking into account: hypomethylation that defines a low methylation status, and hypermethylation which depicts a pattern with a high level of methylation. Since the discovery of DNA methylation, the two major changes that were described from the plethora of studies performed in cancer were the hypermethylation of CpG island resulting in stable silencing of gene expression [39] because a block of genes transcription, and the hypomethylation of CpG that was generally associated with chromosomal instability and possibly active, and overexpression of genes [40]. Nevertheless, in the last decades, the understanding of the effect of DNA methylation on gene expression has rapidly changed. It is clear that DNA methylation is strictly correlated with the regulation of gene expression and it is well demonstrated that it is also implicated in the silencing of retroviral elements, in genomic imprinting, and in X chromosome inactivation. However, with the advent of high-throughput technologies as well as New Generation Sequencing (NGS), a more sophisticated aspect of DNA methylation has been revealed, suggesting a possible role for this phenomenon in splicing regulation [41], nucleosomes positioning [42], and the recruitment of transcription factors [43]. Together, these multiple and variable functions suggest that the process of DNA methylation can be considered as a cellular epigenetic memory [44].

According to recent evidence, the dogma explaining the classical correlation between DNA methylation and gene expression is no longer sustainable and whole-genome studies reveal that the effect of DNA methylation is not restricted to the promoter regions, but epigenetic changes occurring in low CpG density regions, in the gene body, and intergenic regions have also great importance and impact on the regulation of gene expression. Although currently, the exact role of methylation in gene expression remains unknown, it is evident that proper DNA methylation is essential for cell differentiation and embryonic development. Moreover, depending on the underlying genetic sequences, DNA methylation may exert different influences on gene activities in different genomic regions.

The real mechanism that explains how methylation in CpG islands regulates gene expression is still being uncovered, with the accepted hypothesis that methylation occurring in CGIs impair the binding of the transcription factor and recruit repressive methyl-binding proteins, leading to the stable silencing of gene expression. However, CpG islands, especially those associated with gene promoters, are rarely methylated, and a growing body of evidence demonstrated that the association between gene expression and DNA methylation can be both positive and negative. Moreover, studies on colorectal cancers (CRC) showed that most of the aberrant methylation in CRC cells doesn't occur in the promoters and CpGs island, but in the CpGs shore found 2kb from the islands [37], indicating a distinct methylation profile in cancer, as well as for CpG shelf and open sea regions whose changes in methylation have been associated with hepatocellular carcinoma [45], and metastatic melanomas [46]. Furthermore, the methylation of CpG shores in cell differentiation is observed to be more dynamic than the methylation of CpG islands, suggesting a decisive role of CpG shore methylation on cell fate determination [47].

Although DNA methylation in the promoter region is the most studied affecting gene expression, other gene regions could be affected by DNA methylation leading to a dysregulation of gene expression: for instance, it has been observed that transcriptional enhancers can physically interact with gene promoters to support tissue-specific differentiation. Enhancers are the distal cis-acting regulatory elements capable to enhance transcription through the recruitment of transcription factors and co-activator proteins. Enhancers regulate the expression of their target genes over distances, going from several hundreds of base pairs to a few megabases, and consequently, if they are absent, the transcription of associated genes is often weakened. In general, enhancers exhibit low CpG density, therefore several studies

support the idea that enhancer methylation, like promoters, has characteristic DNA methylation patterns [48], and hypomethylation of these regions correlated with active gene expression [49, 50]. Importantly, such enhancer methylation was shown to be more closely associated with gene expression changes in cancer cells than promoter methylation itself [51] and can contribute to elucidating the interaction between distal DNA methylation sites and germline polymorphisms as a possible explanation of the biology behind predisposition to cancer [52].

As well as enhancer, also DNA methylation in the gene body has been fully studied because taxonomically widespread, even though the exact function remains enigmatic. The term gene body is used to refer to the region of the gene past the first exon that includes together introns and exons. DNA methylation in gene bodies is surprisingly abundant and evidence suggests that this is associated with a higher level of gene expression in dividing cells [53, 54], even though it can interfere also with transcription elongation [55]. The different behavior observed in the activities of DNA methylation in promoters when compared with gene bodies has been called the DNA methylation paradox [56]. In 2012 Jjingo et al., performed a meta-analysis of genome-wide methylation, to better understand the presence and role of DNA methylation in gene-body [57]. The explanation model they proposed argued that the repression of intragenic transcription associated with gene-body methylation is largely epiphenomenal, which means that this methylation occurred predominantly because of the accessibility of the DNA to methylating enzyme complexes.

1.1.3.3 DNA methylation and Aging

The relationship between DNAm and age has been strongly studied because it has been demonstrated that changes in DNA methylation occur with age and across the entire life [58, 59]. With the advent of the array technology, several studies showing DNA methylation changes at specific CpGs in various age-related diseases were reported [60-62]. Using machine learning technology, it was possible to build different epigenetic clocks based on datasets containing DNA methylation data together with chronological age information with the idea to identify which CpG sites in the human genome predict chronological age with high accuracy. Currently, there exist several DNA methylation clocks able to predict age from the different number of specific CpGs and in different kinds of tissues. The most widely used epigenetic clock has been attributable to Horvath that in 2013 developed an age estimator for the microarray 27K and 450K based on 353 CpGs sites [63], the so-called pan-tissue epigenetic

clock. The clock was designed considering 51 different types of tissues and 8,000 individuals of different ages and shows a high correlation with the chronological clock across the life cycle. What is interesting is that, while many of the CpG sites individually show only a weak correlation with age, this subset of CpG sites selected by Horvath predicts chronological age with great accuracy. Currently, it is well known that DNA methylation-derived epigenetic clocks are better in estimating the real chronological age than transcriptomic and proteomic data, or telomere length [64]. However, it was observed that some variability in this initial clocks' age estimation existed, which was identified to be a measure capturing individual variation in biological age. The deviation and difference observed between this epigenetically measured age and the actual chronological age is defined as a process of deceleration or acceleration. Age acceleration was found to be associated with mortality and other age-related phenotypes or diseases that have been observed in different kinds of cancers, as well as in different disorders, such as Down syndrome [65] and HIV [66].

In predicting age using the Horvath clock, a big issue was observed with the new EPIC microarray platform because, among the 353 CpGs considered for the age prediction [63], 19 CpGs were absent in the EPIC data leading to the doubt if this epigenetic clock could be suitable also for the novel data. In a study conducted in 2018, Mc Ewen and colleagues demonstrated that the lack of the clock-CpGs on the EPIC array did not compromise the utility of the epigenetic age predictors. Furthermore, they observed small differences in the DNAm age estimate between data pre-processing methods, implying that although the methods assessed here differed in mean values, the trends in respect to chronological age were consistent across methods [67].

The majority of the studies assessing age-associated DNA methylation changes, and found in the literature, focused on adults, leading to very limited knowledge on how the epigenome changes over time during infancy and what implications and consequences could have for disease-based pediatric epigenetic studies. In 2011 Martino et al., assessed the temporal changes over time between birth and five years of age in a small number of individuals demonstrating that immune system-related pathways show significant changes over time [68]. More recently, in 2015, another study performed on infant human blood leukocytes, revealed that DNA methylation changes during early childhood affected genes involved in inflammatory diseases and genes encoding histone modifiers and chromatin remodeling factors [69]. Last year, Péréz et al., with a paired-analysis of 11 children at 3 different time-points (at birth, 5 and 10 year)

demonstrated that epigenetic remodeling is dramatically reduced after the first 5 years of life; moreover, they observed that hypermethylation changes were associated with repressive genomic locations and genes with developmental and cell signaling functions, while hypomethylation changes were linked to enhancer regions and genes with immunological and mRNA and protein metabolism functions [70].

1.1.4 Inheritance and Imprinting

Epigenetic inheritance indicates the transmission of information beyond the DNA sequence during cell division and from one generation to the next. It is a crucial process to maintain differential gene expression patterns in development and disease, where epigenetic marks, established during embryonic development are stably inherited during mitosis, allowing cell differentiation and growth. This epigenetic inheritance, better defined as mitotic stability allows the cell, together with genetic information, to inherit also the patterns of gene expression that characterize differentiated cells and that is not encoded in the nucleotide sequence of DNA; besides, an environmental factor that modifies the epigenome of a somatic cell during development is stably replicated and permanently influences the somatic cell differentiation and function throughout life. Therefore, even after an early life exposure, the altered epigenome will continue to regulate gene expression of that cell population. This inherited epigenetic signature could be present in all tissues including the germline, although the measure of this phenomenon is not fully known. Despite the fact it is strongly affirmed that epigenetic inheritance is restricted to mitosis, few studies suggest that in mammals, epigenetic marks established during the life of an organism could be passed through meiosis on to the next generation in the so-called transgenerational inheritance [71, 72]. In this kind of inheritance well documented in plants and nematodes, during meiosis, most of the epigenetic marks are reprogrammed during gametogenesis and early embryogenesis, and this resetting of the epigenome is of great importance. Most of the epigenetic modifications are depleted, to restore totipotency but some structures are conserved, resulting in a kind of memory of the epigenetic state for the following generations. What remains unclear is to know if the transgenerational inheritance is the result of an incomplete erasure, or if it represents an evolutionary advantage to preserve the epigenetic marks that reflect the parents' experience [73].

The strongest example of epigenetic transmission during development and across generations is represented by “imprinted genes”. The phenomenon of imprinting is the best explanation of the regulatory function of DNA methylation in mammals since it is essential for normal mammalian growth and development. Discovered in 1984 following two simultaneous experiments of McGrath and Barton [74, 75], the feature that characterizes imprinted genes is that their expression is limited to only one of the two parental chromosomes. One copy is active and transcribed into proteins, whereas the other undergoes silencing owing to the presence of a stable mark “imprint”, which interferes with regulating binding-proteins avoiding transcription. Albeit the nature of imprinting is currently under investigation, pieces of evidence suggested that imprints originated in primordial germ cells, erased and re-established during oogenesis or spermatogenesis, and maintained in all somatic tissues throughout development [76, 77]. This germline reprogramming could represent a barrier to transgenerational epigenetic inheritance, notwithstanding, imprinted loci can resist the post-zygotic phase, maintaining DNA methylation of ICRs. For imprinted genes, the expression of the two gene copies is determined by the sex of the parent; a group of genes present in the same chromosomal region is inhibited in a coordinated way by methylation of imprinting center, often called differentially methylated regions (DMRs), that often overlap CpG islands. Recent studies suggest that transcription of DMRs in the oocyte may target them for subsequent CpG modification by maintaining an open chromatin structure accessible to *de novo* methylation [78].

The most known example of imprinting is the X chromosome silencing where DNA methylation mediates gene dosage control via inactivation of the second X chromosome in females. In this process, a cascade of events, involving the activation of a cis-acting noncoding RNA, (Xist), which coats the X chromosome and the induction of displacement of transcription factors, and chromatin changes, bring to the CpG methylation at promoter CGIs.

In both imprinting and X-inactivation, long non-coding RNAs [79], like the Xist transcript in X-chromosome inactivation [80], may play a regulatory role in gene expression. It has been proposed that these long non-coding RNAs serve to recruit chromatin-modifying complexes that would in turn silence or activate genes found within the allelically regulated clusters. The conformation of the chromatin conformation is likely to be fundamental for the imprinting of some maternally imprinted genes. Specifically, it has been observed that deficiency of histone H3-lysine 4 (H3K4) demethylase, essential for these processes, results in embryonic stem cell lethality during early differentiation [81, 82]. At the moment, it is known that the recognition of

specific DNA sequence motifs in imprinting centers, the expression of factors that mark maternal chromosomes, and chromatin remodeling, are the mechanisms involved in the methylation of specific DMRs during imprinting.

Understanding epigenetic inheritance and memory is crucial to distinguish the cause or the effect of epigenetic variation in the disease process. To determine whether the variation is present before the development of a disease, it is useful to consider if the epigenetic footprint has been inherited and hence be present in all tissues including the germline, in case that it happens in early development [83] if it is limited to one or few tissues, in case that it appears postnatally or during adult life [84], or if it is environmentally-induced, either by adult life-style related factors developmental reprogramming.

1.1.5 Epigenetics in Cancer and complex diseases

Since epigenetic changes are required for normal development and health and are associated with essential processes involved in cell physiology, it is not surprising that when altered, they can be responsible for some disease states. The disruption of any of the three systems that contribute to epigenetic alterations can cause abnormal activation or silencing of genes. These disruptions have been associated notably with cancer, but also with syndromes involving chromosomal instabilities, and mental retardation, as well as neurological and autoimmune diseases (Table 1).

Table 1: List of some of the principal syndromes and disorders associated with the disruption of the epigenetic system. Disorders are classified into categories and for each group, etiology and consequences of epigenetic changes are provided.

Disease	Etiology	Consequences	References
Cancer			
Primary breast and ovarian carcinomas	Promoter hypermethylation of BRCA1 gene	Loss of BRCA1 transcript	[85]
Papillary renal cell carcinoma	Hypermethylation of RASSF1A	Inactivation of RASSF1A	[86]
Primary hepatocellular carcinoma	Promoter methylation of p53 gene	Reduction of p53 gene expression	[87]
Coloretal Cancer	Hypomethylation of CDH3 (P-cadherin) promoter	Induction of CDH3 expression	[88]
Retinoblastoma	Methylation of RB1 and other tumor suppressor genes	Allelic inactivation of the RB1 gene.	[89]
Chromosomal Instability Disorders			
Prader-Will syndrome	Deregulation of one or more imprinted genes	Loss of the paternal contribution	[90]
Angelman's syndrome		Loss of the maternal contribution	[91]

Disease	Etiology	Consequences	References
Fragile X chromosome	Promoter DNA hypermethylation of *FMR1* gene	*FMR1* gene silencing	[92]
ATR-X syndrome	Hypomethylation of certain repeat and satellite sequences	Aberrant expression of several genes in *ATRX*-deficient patients	[93]
Metabolic Disorders			
Type II diabetes	Hypermethylation of the *PPARGC1A* promoter	Reduction of PPARGC1A gene expression	[94]
Obesity	Altered DNA methylation of HIF3 in adipose tissue and blood cells	High level of BMI	[95]
Neurological Disorders			
Schizophrenia	Hypermethylation of RELN gene	Decreased RELN expression	[96]
Epilepsy	Increased DNA methylation levels in hippocampal tissues	Increased DNA methyltransferase activity, disruption of adenosine homeostasis, and spontaneous recurrent seizures.	[97]
Neurodegenerative Disorders			
Huntington's disease	Genome-wide demethylation	Instability of CTG/CAG trinucleotide repeats	[98]
Amyotrophic lateral sclerosis	DNA methylation of *EAAT2* gene promoter	Silent state of the human *EAAT2* gene	[99]
Alzheimer's disease	Global DNA hypomethylation and specific gene hypermethylation	Transcription activation or repression	[100]
Parkinson's disease	Activation or inhibition of DNA methylation	Regulation of α-synuclein gene expression	[101]
Autoimmune Diseases			
Rheumatoid arthritis	Aberrant methylation status in T cells	Alteration of gene expression and generation of autoreactive T cells	[102]
Systemic lupus erythematosus	Hypomethylation of $CD4^+$	Overexpression of genes involved in disease progression	[103]
Systemic sclerosis			
Sjogren's syndrome	Hypomethylation of CD70	Gene's overexpression	[104]
Autoimmune thyroid diseases	Global DNA hypomethylation	Overexpression of genes important for a correct immune function	[105]
Inflammatory Bowel Diseases	K8 and K16 acetylation	Inflammatory gene expression	[106]
Type I Diabetes	Hypermethylation of FOXP3 in $CD4^+$ cells and histone methylation	Increased risk for the development and progression of vascular complications	[107]

The first studies observing the loss of DNA methylation in tumor samples as compared with the level of DNA methylation in normal tissue samples in colon cancer date back to 1983 [108]. Cancer, among other disorders, is the most described example of how the failure of the proper maintenance of heritable epigenetic marks can result in inappropriate activation or

inhibition of various signaling pathways [109, 110]. These epigenetic alterations are capable to interact at all stages of cancer development, promoting and triggering cancer progression. Although the genetic origin of cancer is widely accepted, recently several studies have suggested that epigenetic alterations may play a key role in the initiating events typical of some forms of cancer [111]. These pieces of evidence have led to insights into the role of epigenetics in the initiation and propagation of cancer. Although the major effect observed in cancer is a global loss of DNA methylation of repetitive DNA sequences and demethylation of coding regions and introns regions of DNA that allow alternative versions of the messenger RNA that is transcribed from a gene, nowadays most studies focus on DNA hypermethylation in promoter regions resulting in gene silencing and leading to inactivation of tumor suppressor genes. Epigenetic changes, unlike genetic mutations, are potentially reversible and can be restored to their normal state by epigenetic therapy that makes such initiatives promising and therapeutically relevant [112].

Discovering the importance of epigenetics in human disease initiation and progression started in oncology, but many other syndromes and diseases are now involved, such as neurodevelopment and neurodegenerative disorders, cardiovascular diseases, and autoimmune disorders. It is well known that epigenetic mechanisms play a significant role in several essential neurodevelopmental mechanisms and are intrinsically related to neurodegenerative and neurodevelopmental diseases [113-115].

In Alzheimer's disease (AD) shreds of evidence demonstrated that changes in DNA methylation levels of several genes related to sensory perceptions, cognition, and neuroplasticity were associated with a reduction in the expression of the corresponding RNA transcripts and proteins involved in the development of the AD [116, 117]. Patients with Parkinson's Disease (PD) are found to present an aberrant level of DNA methylation [118] or dysregulation of specific miRNAs [119] that could contribute to the onset and development of the sporadic or the familial form of the disease. Huntington's Disease (HD) and Multiple Sclerosis (MS) are other two neurodegenerative disorders where the epigenetic machinery that controls the gene expression and the pathologic process of these disorders is altered [120, 121]. Epigenetic aberrant changes can be observed also in multiple target tissues from patients affected by metabolic disorders, such as type 2 diabetes, where epigenetic modifications may act as the cause or the consequence of the disease and obesity, where a pivotal role is recognized for DNA methylation of genes encoding for a key enzyme in metabolism [122].

Autoimmune disorders (AID) are a group of complex conditions characterized by high morbidity and high variability across ages, gender, and human populations. Together with genetic predisposition and environmental factors, epigenetics has been suggested to contribute to AID development and pathogenesis. For instance, the aberrant methylation of specific immune genes is found to contribute to the inflammatory environment in Rheumatoid Arthritis (RA) [123] and the unusual acetylation of specific histones could be associated with higher disease activity in patients with Systemic Lupus Erythematosus (SLE) [124].

1.1.6 Epigenomics and Epigenome-wide Association Studies (EWAS)

The term "epigenetics" has been formally used for decades referring to genome-wide chromatin modification profiles mainly including DNA methylation and histone modifications, which control the accessibility of the chromatin to DNA replication and repair and transcription machinery, modulating genome activities. Recently, the newer term "epigenomics" has started to be adopted by the research community to indicate studies, on the genome-wide level, that focus on the analysis and the study of the epigenome which encloses together DNA methylation, histone modifications, and noncoding RNAs mechanisms [125, 126]. Therefore, while epigenetics focuses on processes that regulate gene expression, epigenomics is concentrated on the analysis of epigenetic changes across the entire genome.

In line with epigenomics and the advent of the high-throughput technologies from microarray to next-generation sequencing, the term "epigenetic epidemiology" was started to be embraced in the literature to describe the need to understand the interaction between the inheritance of epigenetic information and the mechanism of inherited traits including the risk of disease [127]. Therefore, it is crucial to determine the contribution of epigenetic mechanisms to reveal epigenotypes that might be involved in the development or progression of diseases and complications. To this aim, epigenome-wide association studies (EWAS) developed for the first time in 2008, providing a large-scale systematic approach to uncovering epigenetic variants underlying common diseases [128].

EWAS, like GWAS, aim to examine the epigenetic state at many different loci in several individuals and evaluate whether any of these loci are associated with a trait [129, 130]. EWAS typically focus on the association between DNA methylation in a specific tissue and the presence of a disease or other characteristics such as environmental pollutants. EWAS in DNA

methylation address to test whether the DNA methylation at individual or groups of adjacent cytosines in the genome differed systematically in those with the phenotype of interest with the idea to select specific genes within the genome that may play a key role in the disease risk or etiology.

Although EWAS show the potential to capture the integrated effects of both nature and nurture on disease development [131], compared to GWAS it presents several limitations that are associated with the plasticity of epigenetic modifications. In fact, DNA methylation marks are flexible, dynamic, and change both over time and in response to environmental exposures. Therefore, the choice of an appropriate study is fundamental when performing and analyzing EWAS. Several epidemiologic study designs are available for GWAS, but the most frequently used for EWAS is the case-control study, where unrelated participants are selected on the basis of a specific phenotype and are distinguished in case (subjects with a disease or a specific phenotype) and controls (subjects without that feature). These kinds of studies, like also the birth cohort studies, the familial-based studies, and the twin's studies suffer from the potential for reverse causality or causation [132], which means that epigenetic changes may be a cause of the exposure and disease development or a consequence and effect of the disease state. For this reason, it is important to establish a potential causal link between DNA methylation and a phenotype of interest. Other important concerns specific to EWAS that should be considered are the choice of the sample source, because of the high tissue specificity of the epigenetic marks and the confounders from genetics and environmental factors that are hard to control. For instance, it is important to check for sample heterogeneity by applying deconvolution methods. Furthermore, environmental factors or other particular hallmarks that could contribute to phenotypic variation have to be considered, as well as aging or allele-specific DNA methylation. Finally, the verification and validation of the obtained results are crucial to minimize false negative or false positive findings.

1.2 EPIGENETIC REGULATION OF IMMUNE RESPONSE

1.2.1 The immune system

The mammalian immune system has evolved several strategies to protect the host against microbial infections. The immune response to antigenic molecules occurs mainly through two distinct mechanisms of defense: the innate and the adaptative immune systems [133]. The first one is basically composed of the barrier functions of the body surfaces and mucosal epithelia of the respiratory, gastrointestinal, and reproductive tracts. The key components of the innate immune response are the macrophages, the cells that are immediately triggered when a foreign microorganism is recognized and bind the surface receptor of the cell. Their main role is the secretion of cytokines that when released in response to pathogenic constituents initiate the process of inflammation. The adaptative immune system, instead, is partially induced by the innate immune response and it is specific, which means that it is directed to one foreign identity that is specifically recognized. Generally, the adaptive immune response provides lifelong protective immunity in case of reinfection from the same pathogen. The key element of the acquired immune system is the antibody, which oppositely to macrophages are produced only after infection, and are specific for the infecting pathogen [134].

Innate immunity mainly involves macrophages, innate lymphoid cells (ILCs), natural killer cells, dendritic cells, and granulocytes, which are a diverse collection of white blood that includes neutrophils, basophils, and eosinophils. Adaptive immune responses, instead, depend upon lymphocytes, which provide lifelong immunity following exposure to infectious diseases or vaccination. Two different types of lymphocytes can be distinguished: B cells whose main role is the production of antibodies and T cells, subdivided into two subgroups, the cytotoxic T cells, and the helper T cells, which directly kill infected host cells, activating other immune cells, and producing cytokines. All immune cells originated from a common hematopoietic stem cell (HSC), which gives rise to common lymphoid progenitor (CLP), which are the precursors of B cells, T cells, ILCs cells and natural killer cells, and common myeloid progenitor (CMP) cells, which, instead, generate granulocytes and monocytes/macrophages [135].

The immune response that is initiated soon after an infection involves both innate and adaptive mechanisms. The response of the innate immune system is immediately triggered

when a set of pattern recognition receptors (PRRs) perceive the presence of external pathogens or toxins. One of the best characterized PRR is the TLR-4 which recognizes pathogen-associated molecular patterns derived from various microbes. The adaptative immune response, instead, is activated when an immature dendritic cell (DC) takes up a pathogen in infected tissue, becomes activated, and matures into a highly effective antigen-presenting cell (APC) activating pathogen-specific lymphocytes. Macrophages are APCs, as the mature DCs that are the most powerful antigen-presenting cells that activate CD4 and CD8 T cells. CD4 T cells when activated, differentiate in a broad range of subset cells which activate macrophages, stimulate B-cell differentiation, and regulate autoimmunity and inflammation. CD8 T cells, instead, when activated, work in the elimination of the viral infected cells. The communication that occurs between immune cells and their activation, migration, maturation, and differentiation is coordinated by a variety of cytokines including interleukins and chemokines. The immune response at the level of cytokine production is controlled by the major histocompatibility complex (MHC) [136], a group of genes that encode proteins on the host cell surface which play a critical role in the processing and presentation of antigens to T cells. Two different classes of MHC molecules regulate the delivery of short peptides to the cell surface for recognition by CD8+ and CD4+ T-cells respectively: molecules of class I that are intracellular or endogenous and those of class II that are extracellular or exogenous molecules [137] (Figure 5).

The induction of long-lived memory after the infection is the principal characteristic of vaccination, and B cells can contribute to the establishment of memory, allowing the rapid production of protective antibodies immunoglobulin G (IgG) [138]. Although it is thought that memory is a typical hallmark of the adaptive immune system, recently it has become clear that innate cells such as natural killer, monocytes, and macrophages can also “remember” previous experience, enhancing responsiveness when they reencounter pathogens; this program of innate immune memory is called trained immunity [139].

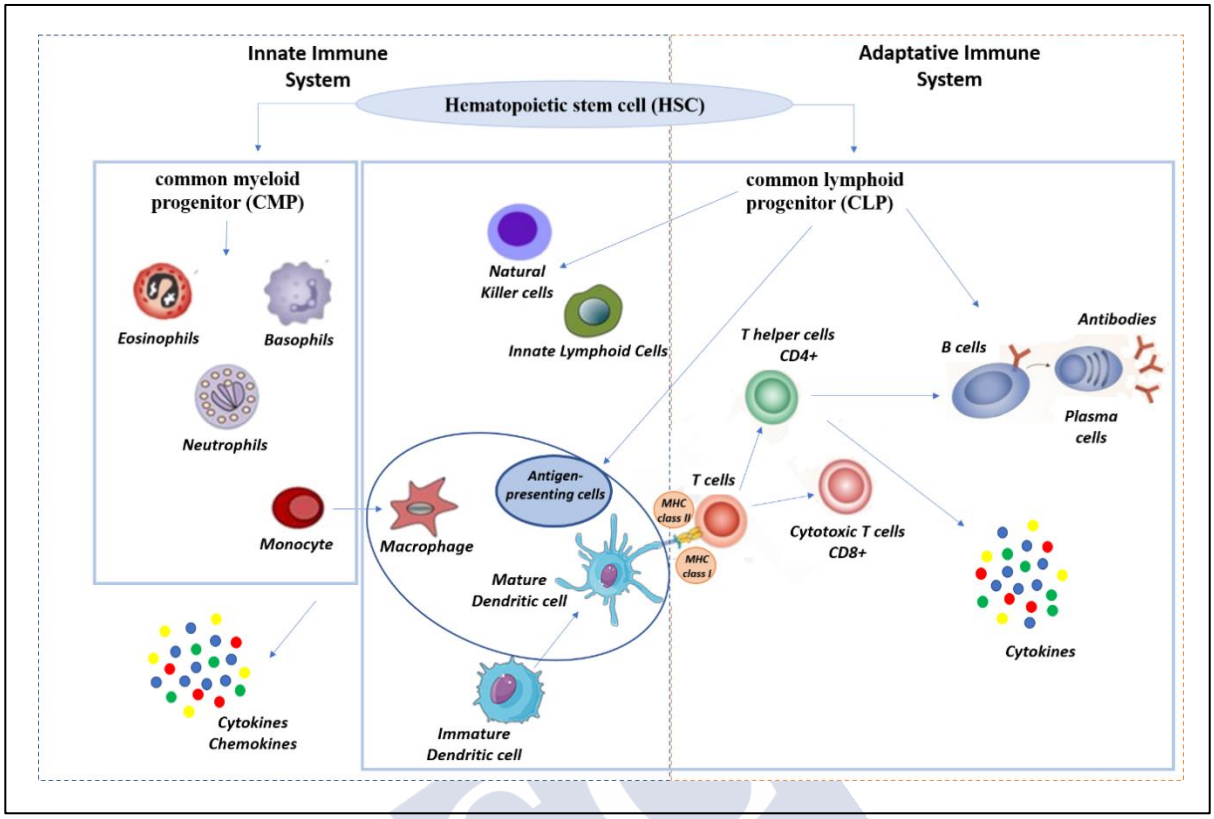


Figure 5: Schematic of immune system. Self-made figure. Created with BioRender.

1.2.2 The role of epigenetic reprogramming in immune cells differentiation

Epigenetic regulation is a fundamental factor for the orchestration of physiological immune responses and maintenance of immunological memory. For several years, a plethora of studies brought the attention to the role of epigenetic events in the development of cancer and other related diseases. However, in the last decade, a great interest has grown in the ability of infections, as well as vaccines, to alter the human epigenome and consequently modulate the activity of the immune system. In the last years, researchers address understanding how natural infection and vaccination can alter the epigenome, modulating the initial immune response and the longitudinal disease risk. A growing body of evidence suggested that promoters and enhancers of different regulators of the immune response are strongly susceptible to modulation by epigenetic marks.

The differentiation of the immune cell from a multipotent hematopoietic stem cell through the lymphoid or myeloid lineages is the first step where transcription factors and epigenetic modifiers collaborate to activate and simultaneously silence genes of a particular lineage, at the expense of the other precursor cell. Each developmental process is regulated by epigenetic

mechanisms, from immune cell differentiation to cell fate decision. For instance, it has been observed that the process of stem cell self-renewal depends strongly on DNA methylation, which is associated with repressed gene expression [140, 141].

Concerning innate immune cells, macrophage polarization, the process by which macrophages adopt different functions, induces two extreme states of activation, called M1, which is responsible to kill microbes, and M2, involved in tissue function preservation under physiological conditions and stress. The activation and the effector function of these two states are governed by epigenetic modifiers that can determine the fate of macrophages, whether in the healthy tissues or at the site of tissue injury [142]. A similar regulation has been observed also for adaptative immune cells, as B cells and T cell development and maturation are under the control of epigenetic mechanisms: in CD4+ helper T cell differentiation, a lineage-specific demethylation pattern is observed in the three major Th subsets, Th1, Th2, and Th17. It has been observed that in Th1 and Th2 cells there are a set of enhancers and other boundary elements whose state of methylation confers different functions, thus a hypomethylation pattern in Th1 CD4+ cells have potential functions in the initiation and maintenance of the transcription of Interferon-γ (IFN-γ), while hypermethylation of the same elements results in the silencing of the interferon in Th2 cells [143]. In the epigenetic regulation of lymphocyte cell subsets, several studies focus on the key role of DNMTs in demethylation or *de novo* methylation events during immune cell differentiation: for instance, a block in DNA methylation maintenance by Dnmt1 decreased HSC self-renewal potential bringing to defective differentiation patterns [144], while DNMT3a deficiency in T cells, blocks their differentiation into terminal effector and memory precursor phenotypes [145]. Epigenetic reprogramming of transcriptional pathways has also been found to be a critical determinant for trained immunity, through the acquisition of H3K27ac at distal enhancers, and the consolidation of histone H3K4me3 at the promoters of stimulated genes in immune cells [146].

1.2.3 Epigenetic manipulation of the immune system during infectious disease

Young children are characterized by immature immune systems, which lead to a higher risk of developing life-threatening diseases compared with adults. Deficits observed in the number of immune cells are more striking in newborn and preterm infants that show weak bactericidal functions, poor responses to inflammatory stimuli, and diminished chemotaxis. In monocytes

and macrophages, they show reduced expression of TLR4 expression with impaired innate signaling pathways, leading to a decrease of cytokine responses when compared with adults. Moreover, they exhibit a reduced T-cell response towards Th2, which is associated with a diminishing in the production of IL-12 and IFNα by the neonatal antigen-presenting cells (APC) [147]. This could have an effect on the immune response bringing newborns to be more susceptible to microbial infections and allergic reactions. What is more, in early life the production of antibodies is reduced, such for instance for some IgG isotype, considered as the most effective immunoglobulin against some polysaccharides, and this causes the reduction of TI-2 antigen response [148].

Despite many years of research, the specific characteristics of the immune system in early childhood that underpins increased susceptibility to infection have still not been fully elucidated [149].

Epigenetic regulation behaves as a dynamic interface between genome and environment and, therefore, in the case of viral or bacterial infection, this regulation in host defense cells is directly related to disease development [150]. Pieces of evidence support the implication of epigenetic mechanisms in the modulation of the interaction between host and pathogen [151]. In the last decade, it has been shown that epigenetic processes play a crucial role in host-pathogen interactions, and several studies demonstrate how pathogens have evolved mechanisms to exploit the host epigenetic system to improve their survival and maximize their transmission. It seems that histone modifications and chromatin remodeling, as well as DNA methylation, are the first key target during bacterial and viral infection [152, 153].

Several studies in the literature revealed that many viruses have the ability to manipulate host epigenetic marks, and this probably contributes to the establishment of latency and some pathogenic role [154, 155]. One of the master viruses in exploiting and influencing the host epigenome by reprogramming cells to cause long-lasting, oncogenic phenotypes is the Epstein-Barr virus (EBV) [156]. After EBV infection, a typical epigenetic switch occurs where the naked viral DNA assembles into nucleosomes; here the CpG methylation progressively increases over time since it is required for the productive phase of the viral lifecycle and its long-term persistence [157, 158]. Likewise, the family of adenovirus is found to provoke changes in histone acetylation, causing the repression of genes involved in the inhibition of the infection [159]. The infection of Human Papillomaviruses (HPVs) induces aberrant expression

of many proteins involved in epigenetic regulation, including DNA methyltransferases, histone-modifying enzymes and, subunits of chromatin remodeling complexes, influencing host cell transcription program [160]. Hepatitis B Virus (HBV) is able to modulate different host epigenetic mechanisms after the infection, for instance through the protein HBx which is capable to dysregulate several cellular pathways changing expression patterns of miRNAs, or affecting histone methyltransferases [161].

RNA viruses, like DNA viruses, are shown to have developed complex processes aimed to antagonize the signaling components that regulate the host's innate immune response. These mechanisms are designed to regulate the host epigenome, and control host immune antiviral defense processes, to improve virus replication and pathogenesis. Coronavirus and Influenza Virus family can antagonize the host gene expression by inducing histone modifications [162]. The NS1 influenza protein can mimic the histone tail of the H3 histone, blocking the expression of the human PAF1 transcription elongation complex (hPAF1C) that has a crucial role in the antiviral response [163]. In SARS-CoV and MERS-CoV infection, it has been shown that the host antiviral IFN response is manipulated to allow the viral replication through the inhibition of the ISG effectors production that shows a delay 24/48h after the infection [164] avoiding the initiation of the pathogen-driven immune response. In 2017, Shafer et al., using a ChIP-PCR approach, observed that cells infected with MERS-CoV exhibited increased levels of H3K27me3 mark thus preventing open chromatin and inducing downregulation of ISG expression [162]. In the recent SARS-CoV2, genome-wide DNA methylation study on the blood mononuclear cells of 9 terminally-ill patients, revealed that severe SARS-CoV2 infected subjects present particular DNAm signatures characterized by hypermethylated IFN-related genes and hypomethylated inflammatory gene [165].

1.2.3.1 Respiratory Syncytial Virus

Identified for the first time in 1956, its association with respiratory illness among infants was not immediately clear. In fact, only after its first appearance in chimpanzee coryza suffering from colds and coryza in 1986 [166], it was confirmed that it was the agent causing respiratory illness in humans. For this reason, when the neutralizing antibody to chimpanzee coryza agent (CCA) was detected in most school-aged children, the virus was renamed respiratory syncytial virus (RSV) to depict its clinical manifestations. Today it is known that RSV is the common pathogen that infects virtually all children by two years of age, and the

leading global cause of hospitalization of infants. RSV is the principal cause of acute lower respiratory infections (ALRI) in young children worldwide, and it is associated with significant morbidity and mortality in childhood [167, 168]. According to the predominant pathophysiologic processes, RSV caused ALRIs can be classified into RSV acute bronchiolitis that mainly affects the smallest branches of the bronchial tree in children < 2 years of age and show a predominant obstructive pattern with resulting hypercapnia, and RSV with a more restrictive pattern associated to hypoxia and apnea.

1.2.3.1.1 Epidemiology

In a systematic review of 2015 [169], it was estimated that approximately 34 million worldwide new ALRI episodes in children were attributable to RSV, a huge number that resulted in 3.2 million hospital admissions and almost 60,000 global childhood deaths each year. The highest burden of childhood RSV ALRI occurs during the first year of life when children are most vulnerable.

RSV, being a seasonal virus, shows extremely variable epidemiology, exhibiting distinct seasonal cycles, with the highest peaks during the cold season in temperate temperature and during the rainy season in tropical regions.

In 2017, the first robust report on global country-specific RSV seasonality was provided with the idea to improve the prevention, diagnosis, and treatment of RSV infection promptly and in most countries, not only where RSV incidence was recorded through own surveillance systems [170].

1.2.3.1.2 RSV Properties and clinical features

RSV is an enveloped RNA virus belonging to the Paramyxoviridae family, that exists in humans as two antigenic subgroups A and B [171]. Each one of these main strains divides into genotypes (at least 5 for strain A with 22 subtypes and 4 for strain B with 6 subtypes), one or two of which will circulate predominantly each year and will be displaced by another subtype the next year. This displacement of strains allows the virus to fight against natural immunity by changing the predominant strain in each infectious season. RSV genome is encapsulated by the nucleocapsid N protein, which, together with the phosphoprotein P and the large polymerase subunit L, constitute the minimum unit for RNA replication. RSV present

two non-structural proteins, NS1 and NS2, whose functions are unknown, although NS1 appears to be a negative regulatory factor for RNA synthesis. Even if the mechanism of action has not been completely understood, RSV encodes three surface envelope proteins that are components of the virion: the attachment protein G, which determines the absorption of the virus, the fusion protein F, and the small hydrophobic protein SH.

The 298-amino acid glycoprotein G is responsible for viral attachment to cells and is the target for neutralizing and protective immune responses. It can differ by up to 47% in the A and B strains, whilst G proteins from the same strain can differ by as much as 20%. Variability found in the G protein gene is concentrated in the extracellular domain, composed of two hypervariable regions separated by a central immutable 13 amino-acid region strictly conserved in all human RSV strains [172]. From alternative initiation codons on the same gene, two different forms of the G protein can be synthesized. One is membrane-bound (mG) and mediates viral attachment while the other is secreted (sG) and is thought to act as an antigen decoy to bind G-specific antibodies and helping the virus to evade neutralization [173].

The fusion RSV F protein is present as inactivated form (F0) and is activated when F0 is cleavaged into two linked subunits, F1 and F2. The homotrimer F1-F2 is assembled in the virus particle in an unstable prefusion conformation (pre-F). Once RSV bind to cellular receptors, the F protein induces the fusion of viral and cell membranes leading to the transfer of viral genetic material [174]. During this process, F undergoes a series of conformational changes, which results in a highly stable conformation called the postfusion conformation (post-F). One of the most interesting roles of F protein is the promotion of fusion of infected and adjacent cell membranes causing the formation of syncytia, which represent the main hallmark of the RSV cytopathic effect and are necessary for cell-to-cell viral transmission. The interaction between the F protein and a small GTPase, called RhoA, facilitates RSV-induced syncytium formation [175].

The precise role of the third transmembrane protein, SH is unknown at the moment. It is clear that SH protein is not required for viral replication or syncytium formation, but it seems to facilitate fusion [176]. Indeed, when recombinant RSV without the SH gene is inoculated intranasally into mice, replication in the lower respiratory tract occurred normally like wild-type, but replication in the upper respiratory tract is found to be restricted 10-fold. For this

interesting role, this site-specific attenuation of viral replication, caused by SH deficiency may have implications for future vaccine development.

1.2.3.1.3 Pathogenesis

Humans are the only source of RSV infection, and transmission is produced through direct or mediated contact from contaminated surfaces between the mucous membranes of the eyes, mouth, or nose of a healthy person and the respiratory secretions of infected subjects. The clinical symptoms generally begin after an incubation period of 4 to 6 days and in children, the first symptoms settled in the upper respiratory tract followed few days with others in the lower tract. Rhinitis, cough, fever, and dyspnea are the peculiar features of the infection [177], although, in some patients, the disease course may progress to the lower airways causing wheezing, jugular, and intercostal retractions, hypoxemia, or respiratory distress that needs hospital admission [178].

A growing body of evidence suggested that the severity of RSV bronchiolitis occurring in infancy could be associated with immunological mechanisms. In fact, RSV infections appear with more frequency in the first years of life when the immune system is immature, and despite the possession of maternally derived specific RSV antibody.

In RSV infection, the first innate mechanism of defense is represented by the pulmonary surfactant which is a layer of phospholipids and surfactant proteins. The recognition of RSV products by the host cell is mediated by two important pattern-recognition receptors, the Toll-like receptor 4 (TLR-4), and its co-receptor CD14, both essential for the innate response activated by the RSV F proteins [179]. The most important cellular components of the innate response in RSV infection are neutrophils, macrophages, eosinophils, and natural killer cells.

Neutrophils are considered the predominant airway leukocytes in RSV bronchiolitis [180] as they were observed to play a crucial role in the pathological changes that occur in RSV bronchiolitis [181]. Neutrophils are capable to promote RSV lung pathophysiology producing and releasing pro-inflammatory cytokines and cytotoxic molecules which contribute to mucus hyperproduction and airway epithelial cell (AEC) death [182]. Because of their role in RSV infection, their depletion has been found to decrease the level of some enzymes involved in disrupting chromatin packaging, in the airways, and reducing the Tumor necrosis factor (TNFα) levels attenuating airway production in the lungs of RSV infected mice [183].

Macrophages along with other respiratory epithelial cells are the first cells to attack RSV in the airways [184]. They respond to infection by secreting cytokines that up-regulate the immune response increasing vascular permeability and causing the activation and recruitment of lymphocytes, neutrophils, NK cells, and possibly eosinophils to the site of infection.

Eosinophils play a significant role in RSV infection due to their mechanism of action in viral bronchiolitis and asthma. They possess antiviral activity [185], and it is believed that these cells contribute to tissue damage, as well as to respiratory dysfunction through degranulation of their secretory proteins and enzymes [186].

The first studies focused on the role of natural killer (NK) cells in RSV infection, showed that they accumulate in the lung in the first few days of RSV infection and are responsible for much of the early production of IFN-γ [187]. The number of NK cells in the blood is significantly lower in infants hospitalized for RSV bronchiolitis, suggesting that NK cells are recruited from the peripheral circulation into other tissues, probably the lungs, as has been demonstrated in 2012 by Li and colleagues which observed a significant increase of NK cells percentage in the lungs 72h after RSV infection, leading to the activation of most NK cells [188]. However, NK cells are found to play a dual role in viral infection: protecting against viruses thanks to their cytotoxicity and inducing injury by inhibiting antibody responses and secreting pro-inflammatory factors [189]. The role of the released IFN-γ is critical in promoting the activation of the adaptive responses and the inhibition of antibody responses in neonates. Indeed, it has been demonstrated that RSV-infected NK cells are more prone to produce IFN-γ than uninfected cells while maintaining an unaltered cytotoxic response, a combination that could contribute to the severity of the disease [190].

The adaptive immune response consists of immunological memory and may be divided into humoral and cell-mediated responses. In RSV infection, the humoral response carried out by T cell, is primarily involved in protective immunity whilst the cell-mediated response controlled by activated B cells promotes viral clearance.

New-born babies present specific RSV neutralizing antibodies from the placental transfer of maternal immunoglobulin. Most of the severe RSV diseases are observed between 2 and 6 months of age, when protection from maternal antibody should be present, suggesting that anti-RSV antibodies play a role in the immunopathogenesis of RSV bronchiolitis. RSV infection causes the production of serum antibodies in even the youngest children, although the antibody

titers produced in infants are low compared to older children and adults. Humans generally develop antibodies to most RSV proteins, even if the F and G proteins stimulate the production of potent neutralizing antibodies (nAb), important in protective immunity. The F protein, because of its high conservation across RSV strains, is considered the primary antigenic target for nAb, in contrast to the G protein that being more variable between RSV subgroups is considered a less attractive vaccine target. Besides, evidence suggested that the pre-F appears to be a more effective nAb than the post-F form, as a subunit vaccine in experimental animals [191], as well as the pre-F displays more potent neutralizing epitopes than the post-F form [192].

In the RSV infection different classes of immunoglobulins are produced: during the primary infection, the IgM antibodies are the first to be found within a few days from the infection and remain detectable for 1–2 weeks. The IgG antibodies, instead, appear in the second week, reaching high peaks in the fourth week and declining after 1 or 2 months. The level of IgA serum antibody response in infants is more variable and may not occur. Serum RSV-neutralising antibodies seem to have a protective effect since it is observed that children with high titers (> 1/100) are less likely to develop bronchiolitis than those with lower titers. According to recent clinical trials of parenteral administration of RSV-neutralising antibodies to high-risk infants, it was demonstrated protection from severe disease.

Another interesting immunoglobulin class is the IgE, whose role in RSV bronchiolitis as well as in the subsequent development of wheeze is still under debate. Welliver et al., found RSV IgE in the secretions of infants during the recovery phase of RSV bronchiolitis. Levels of IgE were also high in infants with acute bronchiolitis and correlated with the degree of hypoxia.

Cell-mediated immunity in RSV disease demonstrated its importance in children with deficient cellular immunity who shed virus for several months compared to immunocompetent individuals who clear RSV within 21 days after the infection. The cell-mediated immune response is classified into CD8+ cytotoxic lymphocytes (CTL) and CD4+ T helper cells. Both these cell types have antiviral and immunopathogenic capabilities. CD4+ T helper cells are subdivided into Th1 and Th2 lymphocytes on the basis of cytokine secretion. Th1 responses are characterized by IFN-γ, IL-2, and TNF-α secretion; they have important roles in immune modulation and their responses are generally associated with a reduced disease during RSV infection [193]. Th2, instead, was shown to contribute to the severity of the RSV disease in

mice, and are characterized by IL-4, IL-5, IL-10, and IL-13 secretion. Recently, other three T cell subsets have been described: Th9, Th17, and Th22. Th9 like Th2 seems to contribute to the pathogenesis of RSV bronchiolitis in humans [194]; while Th17 and Th22 known to play an important role in the pathogenesis of inflammatory disorders of the gastrointestinal tract and other autoimmune diseases are still under investigation for their possible role in the pathogenesis of RSV infections.

1.2.3.1.4 Prevention and treatment of RSV disease

Unfortunately, so far no specific treatment or vaccine prevention is available for RSV infections. Severe cases of RSV generally require supportive therapy, such as fluid therapy oxygen supplementation or ventilatory support.

The only antiviral agent approved by the FDA for the treatment of severe RSV infections is the aerosolized Ribavarin, a synthetic guanosine analog and broad-spectrum antiviral agent [195]. However, because of the growing concern about its efficacy and its high cost, the treatment is limited for immunocompromised patients and selected young children at high risk for serious RSV disease [196].

Even if vaccines are considered one of the most cost-effective health interventions for infectious diseases, at the moment there are no vaccines available for RSV. Palivizumab is the only humanized monoclonal IgG1 antibody targeting specific viral mechanisms controlling infection and it has been developed for prophylactic use. It is active against both A and B subtypes of RSV and can be administered either intramuscularly or intravenously. Palivizumab is available for infants with high-risk infection, and it is administered monthly during the RSV season, nevertheless because of its high-cost, its dosing requirements and some side effect is not approved by the Food and Drug Administration (FDA) for the target population [197].

The global burden of the disease caused by RSV has become increasingly recognized, leading to many clinical trials of vaccine candidates. The search for a suitable vaccine is very challenging and several RSV immunoprophylaxis and vaccine candidates, are currently in development [198, 199]. The identification of a population with a higher risk of developing respiratory diseases that require hospitalization is essential to plan strategies for prevention such as vaccination or monoclonal antibody use. In the last decade, there has been a strong consensus to focus on children younger than six months of age, when the risk of severe RSV-

associated respiratory disease is highest. This need for immediate protection and the poor efficacy of active immunization lead the maternal immunization and passive prophylaxis to become a greater priority in reducing the burden of childhood pneumonia.

1.2.3.1.5 Asthma and Wheezing

Besides the acute burden of RSV, a growing body of evidence from epidemiology data supports that RSV infection in the first 3 years of life can be directly correlated with long-term respiratory morbidities, such as recurrent wheezing and asthma [200, 201]. It is observed that RSV, like other respiratory viruses, causes a "hit and run" phenomenon, characterized by the increased risk of developing recurrent wheezing and asthma in childhood after the infection, as a permanent phenotype that persists long after virus clearance [202]. Wheezing is the typical high-pitched, whistling sound made during breathing. The Tucson Children's Respiratory Study distinguishes different wheezing phenotypes according to the appearance of wheezing during life: never, transient early, late-onset, and persistent wheezing [203]. Asthma, on the other hand, is characterized by abnormalities in lung function that include variable airway obstruction and an increased bronchial reactivity [204]. Generally, recognizing asthma is usually obvious and most of the time asthmatic patients report wheezing episodes; however, it is very challenging to predict and distinguish which children will present only early-life symptoms, whose symptoms persist, and who may develop definitive wheeze or asthma [205].

The risk of wheezing and/or asthma incidence has been increasingly related to a combination of genetic and environmental factors as well as the severity of the respiratory infection [206-208]. Several efforts have been made to establish the ultimate causes of the relation between the infection and the subsequent respiratory sequelae, but further complementary studies are still necessary to understand the underlying mechanism. It has been hypothesized that prenatal interaction between the maternal and the child immune system as well as the high cytokine production can affect the lung structure and function inducing changes in the regulation of immune response to the infection [209]. What is more, the age at which children are infected by RSV seems to be crucial in the disease outcome, as infants infected before 6 months of age show a higher risk to develop bronchiolitis that usually requires hospitalization [210].

It has been observed that wheezing and allergic sensitization during childhood were associated with elevated IgE levels [211]. However, no serum virus-specific anti-RSV IgE was detected in acute or convalescent samples of infants with RSV infection suggesting that these respiratory sequelae after the infection have not been associated with an increased risk of atopy or higher serum IgE levels. According to several researchers in some children with virus-induced wheezing early in life, the viral infection led to a specific inflammatory response that caused the long-term airway obstruction. This non-atopic wheezing phenotype seems to be associated with a type of wheezing that is less severe and less prevalent than the atopic phenotype. Several genetic variants have been proposed as candidates' genes for non-atopic forms of wheezing or asthma in childhood: for instance, IL-8 is believed to play a role in the pathogenesis of bronchiolitis and it is observed that this chemokine has been associated preferentially with virus-induced inflammation being one of the major mediators of the inflammatory response [212]. Other studies focused on the role of the cytokine IL-4 and IL-13 in the generation of Th2 responses, considered the major determinants of atopic asthma: in fact, IL-4 is essential for the maturation of naive T cells into Th2 cells and is responsible for the production of IgE, while IL-13 inhibits the production of pro-inflammatory chemokines and is also involved in MHC class II expression. Polymorphisms on these genes are found to contribute significantly to bronchial hyperresponsiveness and susceptibility to atopic wheezing and asthma [213].

However, it remains under considerable debate whether the association between RSV and subsequent recurrent wheezing and or asthma is causal. A retrospective cohort study with more than 95,000 infants demonstrated that infants born 3 months before the RSV season showed the greatest risk for hospitalization as well as the highest risk to develop asthma between the ages of 4 and 5 years, suggesting causality [214]. While a Danish study with more than 18,000 twins revealed that severe RSV-induced illnesses were associated with recurrent wheezing during early childhood, but this was attributable to genetic predisposition for both more severe RSV-induced illness and asthma [215].

1.2.3.1.6 RSV and epigenetic modulation

The immune system plays a pivotal role in the sequelae observed in patients suffering from respiratory infections. It is well documented that the modulation of epigenetic mechanisms governs the immune cell phenotype and function allowing the external

environment to influence the immune response outcome [216]. As previously described for other DNA and RNA viruses, also RSV infection could modulate the host epigenetics leading to aberrant regulation of gene expression. For instance, Elgislouli and colleagues exploring the methylation status of a particular enhancer region in the perforin 1 gene (PRF1), an essential cytotoxic protein for the control of viral infection, observed an alteration of its methylation pattern after a severe RSV infection. Infants hospitalized with exacerbated RSV-induced bronchiolitis exhibited decreased methylation of the perforin-1 enhancer after 4 years of follow-up, suggesting that immune response changes due to RSV could persist even years after infection [217, 218]. Therefore, changes in the expression of this gene have been related to the pathogenesis of asthma. Besides, Wang et al., [172] showed that cultured bronchial human epithelial cells infected with RSV presented a high level of NODAL (a member of the TGF-Beta superfamily), and this resulted in increased Th2 and Th17 skewing of T cells. In murine models, it was observed overexpression of Kdm5b, an H3K4 demethylase, in dendritic cells after an RSV infection [219]; this demethylase has the potential to repress transcription of type I IFN and other innate cytokines, causing a decrease of pro-inflammatory cytokines and expression of a Th2 phenotype. Methylation of H3K4 in regulatory T cells by the histone methyltransferase SMYD3 is necessary to control inflammation in the lungs after RSV infection [220].

1.2.4 Epigenome and vaccine response

As it has been previously described, the immaturity observed in the immune system early in life, increases, not only, the vulnerability of infants to infectious diseases, but also reflects their reduced immune response to vaccination [221].

Despite significant research into the role of epigenetic changes induced by infection, less is known about the impact of vaccinations on the epigenetic landscape as well as the role of epigenetic changes in response to vaccines.

In the last decade, several studies have focused on the heterogeneity observed in vaccine responses, a crucial factor with important implications for the distribution of protection in individuals and for the extent of herd immunity achieved when a vaccine is widely used in a population. The variations observed in vaccine responses have consequences for both protective efficacy and the duration of protection. It has been demonstrated that during childhood

vaccination a high number of children are achieving more than 100-fold higher antibody levels post-immunization than others [222, 223]. The reason for this huge variation is not fully understood. The main factors influencing humoral and cellular vaccine responses in humans include intrinsic factors (such as age, sex, genetics, and comorbidities), and extrinsic factors (such as microbiota, infections, and antibiotics). The environmental factors, as well as behavioral and nutritional factors also play a role in how individuals respond to vaccines [223].

In recent years, a different group of studies, mostly conducted in adults, has begun to elucidate the genetic determinants of immune responses to vaccines [224, 225]. Studies on twins were the first to provide insights on the variance in antibody levels for mumps, rubella, and hepatitis B vaccines, due to genetic factors [226-228]. Recently, the genetic variations associated with variable vaccine response have been started to be evaluated in infants after childhood immunization. In 2019, O'Connor and colleagues conducted a genome-wide association study to evaluate the persistence of immunity to capsular group C meningococcal (MenC), to Haemophilus influenzae type b, and to tetanus toxoid (TT) vaccines. Their study revealed associations between SNPs within the SIRP locus and total MenC-specific IgG concentrations, as well as SNPs within the Human Leukocyte Antigen (HLA) gene complex and the persistence of TT-specific immunity. The critical role of the HLA system in the control of antibody response to different vaccines has been extensively studied and HLA polymorphisms have evolved to provide diversity in host responses to multiple, varied, and frequently changing pathogen antigens [229-231].

It is extensively known that epigenetic alterations are the results of aging and diseases, and the decline of immune response is one of the biological systems more affected by aging. Age-dependent alterations in the epigenetic status of some genes may result in the decline of immune function and this might contribute to the increased incidence of diseases in adults and old people [232]. The aberrant gene expression in immune system cells caused by epigenetic changes may contribute to loss of immune tolerance, inflammation, and autoimmunity, resulting in immune senescence. In a study of healthy people of 50–74 years, it has been demonstrated that a specific group of CpG sites, when hypomethylated, is able to reduce the humoral immune response to influenza vaccination [233]. Another study comparing the response to influenza vaccine in older and younger populations described a great epigenetic remodeling in vaccine responders in people over 50 years [234]. Most of the studies found in the literature on the role of epigenetic modification in the modulation of vaccine response are focused on adults. Only in

2014, Lu et al., conducted a study on twenty-five infants classified as good and non-/low-responders and observed an association between poor HBV vaccine responses and the hypomethylation of RNF39 (Ring Finger Protein 39), which is a transcription factor in the major histocompatibility complex (MHC) class I region [235].

1.2.4.1 Pneumococcal Conjugate Vaccine

The encapsulated bacterium Streptococcus pneumoniae is the leading cause of the common pneumococcal infections considered as a major public health problem worldwide causing high morbidity and mortality in young children who suffer pneumonia, meningitis, and septicemia, leading up to 1 million deaths in children under the age of five years [236, 237]. Young children, being the most affected patients, are thought to be the main transmitters of these infections throughout the population [238]. Therefore, maintaining suitable levels of antibodies in this population category is the key to block the transmission of the organism and reach complete protection.

The capsular polysaccharide pneumococcal vaccines generally used to immunize adults have been found to be neither immunogenic nor protective in young children due to poor antibody responses. Therefore, in the last decade, research has focused on the development of additional immunogenic pneumococcal vaccines to provide long-term immunity in children < 2 years of age. The most promising approach focused on the development of a protein-polysaccharide conjugate vaccine for different pneumococcal serotypes that cause most of the infections in children. Pneumococcal conjugate vaccines (PCVs) are found to be highly effective in preventing pneumococcal infections and are observed to provide immunity rapidly and during the second year of life.

The conjugated vaccine against pneumococcus employs the same technology used for the successful vaccine against Haemophilus influenzae type b [239]. It basically consists of an immunogenic but inactive protein that is covalently coupled to the polysaccharide layer of the selected strains of pneumococci. The conjugated antigen can induce a more powerful, T-cell-based immune response in infants, which is developed by the time they are 2 months of age.

The first PCV introduced for the first time in USA in 2000 was the 7-valent vaccine (PCV7), which comprised the serotypes 4, 6B, 9V, 14, 18C, 19F, and 23F. Since 2009 newer pneumococcal conjugate vaccines have been introduced: PCV10 whose coverage was extended for 3 additional serotypes (1, 5, and 7F) and PCV13 sharing the same serotypes of PCV10 plus

3, 6A, and 19A. A growing body of evidence in the literature suggests that immunogenicity variation and persistence exist among different pneumococcal serotypes. For this reason, different clinical trials have been performed in order to assess the immunogenicity and the reactogenicity of the pneumococcal conjugate vaccines, as well as to investigate their interchangeability. In 2010 Snape et al., [240] performed a phase III, randomized, double-blind, active-controlled study where 278 were randomly allocated to receive either PCV13 or PCV7 at 2, 4, and 12 months of age with the idea to assess the immunogenicity and tolerability of PCV13. Some years later, a follow-on study evaluates and demonstrates the antibody persistence against most serotypes until the age of at least 3.5 years of age in the same individuals.

In 2016, with the idea to interchange PCVs to improve vaccine immunobiology, Trück and colleagues performed a new clinical trial to investigate the potential of a booster of PCV at 12 months of age in terms of immunogenicity and antibody persistence [241]. This time 178 children were randomized to receive at 12 months of age a booster dose of either PCV10 or PCV13. Blood and serum samples of the participants were collected before at 12 months (basal samples), 1 month later (13 months), and 24 months after the booster of PCV10 or PCV13. The immune response was measured with ELISA, with post-booster serotype-specific OPA and memory B cells responses. The results of the study demonstrated that at 1-month post-booster a robust antibody response was induced by both vaccines. PCV10 was found to be inferior to PCV13 only for two serotypes (5 and 9F), then PCV13 was observed to be more immunogenic in terms of post-booster IgG responses.

With pneumococcal conjugate vaccine for several serotypes, the concentration of type-specific IgG produced in response to vaccination is predictive of subsequent nasopharyngeal (NP) carriage of that serotype, with higher responses predicting a reduced risk of carriage. The anticapsular immunoglobulin G (IgG) is the key mechanism of protection against each of these endpoints, which protects in a serotype-specific fashion. In a control randomized trial data, it has been suggested that the magnitude of the IgG response to pneumococcal conjugate vaccines varies across individuals within a population, and strong responders to the capsular polysaccharide of any serotypes in the vaccine tended to respond strongly to that of other serotypes [242].

2
JUSTIFICATION AND OBJECTIVES OF THE STUDY

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

The immune system is principally involved in the protection of the organism against exogenous factors, but also in the maintenance of tolerance and in the control of cellular integrity. Its main feature is immunological memory, referred to as the capacity to rapidly respond with increased intensity upon secondary challenge with the same antigen, a mechanism that forms the basis of vaccination.

Epigenetic mechanisms dynamically regulate gene expression and transcription; epigenetic patterns are different between individuals, even between identical twins, and change over time influenced by intrauterine and postnatal environmental factors such as nutrition, toxins or drugs, and illnesses. Emerging evidence suggests the pivotal role of epigenetic in human pathologies and the regulation of the immune system in health and disease. It has been observed that the most important cell development processes such as early progenitor commitment, lineage fate decision, and differentiation are regulated by epigenetic changes. Many functional differences observed in the immune system of neonates and adults seem to be associated with epigenetic modifications of genes that control inflammation and immune response. What is more, changes that occur in the immune system are thought to increase susceptibility to infection, meaning that external environmental factors the modulate the epigenetic mechanisms that regulate the immune cell phenotype and function could influence the immune response outcome. Epigenetic regulation behaves as a dynamic interface between genome and environment and, therefore, in the case of infection, the regulation in host defense cells is directly related to disease development. Pieces of evidence support the implication of epigenetic mechanisms in the modulation of the interaction between host and pathogen, as well as in the variable immune response to vaccination.

One of the most studied epigenetic mechanisms that involve the addition of methyl groups to the DNA molecule, is DNA methylation, considered as the major epigenetic factor influencing gene activities. Concerning its role in pathogenesis processes, changes in genomic DNA methylation have been described in recent years in response to bacterial and viral infections. However, it was unknown whether the infection induced any type of alteration in this host methylation-mediated transcriptional regulation mechanism. Moreover, there is very limited knowledge about how the epigenome changes over time, as well as there are very few studies assessing the effect of the epigenome on the immune response to vaccination and whether different antigenic stimuli have a divergent effect on the epigenome. The new field of epigenomics can be used as a discovery instrument to better understand the epigenetic

contributions to human infectious diseases. The epigenomic identification of qualitative and quantitative measures of epigenetic changes is likely to exert a higher level of biological characterization of potential disease predisposition biomarkers and provide great potential for advancing preclinical knowledge as well as predict disease outcomes.

Because of its fundamental role in epigenetic regulation, the analysis of DNA methylation is of great interest to understand biological processes such as DNA replication, gene expression, cell differentiation, or the molecular basis of many diseases, including cancer. As can be observed in the following material and methods section, numerous techniques have been developed for the study of DNA methylation, both at the regional, global, and genomic levels. The Infinium Illumina Methylation BeadChip platform used in this project is a cutting-edge technology with a high-quality genome-wide coverage that allows access to single CpG methylation status located at various regulatory regions throughout the human genome with a relatively low cost, and with a great utility in clinical studies. This technology makes possible the interaction between epigenetics and immunology which could help to understand the molecular basis behind the regulation of different physiological processes and translate this knowledge into clinical applications.

2.1 PRIMARY OBJECTIVE

Due to the importance of the immune system in the first line of defense against viral infections and also because of the lack of studies focused on the modulation of the immune response in early childhood, the main objective of the present study is to provide insights into the role of epigenetics processes, specifically the role of DNA methylation, on the relationship between infection and host disease outcomes, as well as on the modulation of the immune system response to vaccination.

2.2 SECONDARY OBJECTIVES

To address the overall objective of the project, four specific objectives have been defined, and are detailed below.

- Evaluate the mechanisms of gene regulation at the methylation level that are involved in the differential phenotypes that could be manifested after RSV infection.
- Find a relationship between methylation changes after documented RSV infection, and posterior development of wheezing and/or asthma.
- Assess the methylation changes that occur early in life to identify pathways that show significant changes over time.
- Elucidate the methylation differences between high and low PCV13 booster vaccine responders in order to search for epigenetic biomarkers that are associated with a more robust immune response.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

3
MATERIALS AND METHODS

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

3.1 STUDY DESIGN

3.1.1 Respiratory Sequelae after RSV

In the first part of the project, a prospective, longitudinal, observational study of infants and children admitted to the hospital for a respiratory infection due to RSV from 2010 to 2015 was carried out.

A clinical follow-up of these children was performed for more than 3 years after the infection, to monitor any development of wheezing or asthma. Furthermore, information on any wheezing episode was obtained by interviewing the family pediatricians of all enrolled children.

Patients selected according to their clinical course were classified into three different subgroups that matched perfectly for age, gender, and disease severity:

- Recurrent wheezing RSV cases: RSV positive patients who have developed at least 3 episodes of wheezing;
- Asthma RSV cases: RSV positive patients who have developed asthma after RSV infection (diagnosed by a pulmonologist or with a response to treatment).
- Not-wheezing RSV cases (control group): RSV positive patients who have not developed recurrent wheeze/asthma after the infection.

A wheezing episode was defined as a respiratory episode with wheezing for more than 1 day. The interval between two episodes was defined as a period of at least 7 days without respiratory symptoms. Recurrent wheeze was defined as three or more episodes of wheezing during the first year of life.

Patients with any underlying/chronic disease that could interfere with the results were excluded from the study. Two patients' cohorts were selected for this study: the first one was the study cohort (Table 2), and the second one was the validation cohort (Table 3).

Blood for DNA methylation analysis was collected for a total of 77 samples in the study cohort: 45 RSV cases (36 recurrent wheezing and 9 asthma cases after RSV infection) and 32 control samples identified as RSV infected patients that did not develop wheezing and/or asthma; and 43 total samples in the validation cohort: 22 RSV cases (12 recurrent wheezing, 10 asthma cases after RSV infection) and 21 control samples with a normal recovery. The study was approved by the Ethics Committee of Clinical Investigation of Galicia (CEIC 26-July-2017, reg 2017/398).

Table 2: Clinical features of patients of the study cohort and their classification in control, wheezing, and asthma.

	Control (*n* = 32)	**Wheezing (*n* = 36)**	**Asthma (*n* = 9)**
Ethnicity			
Western Europe	28 (87.5%)	33 (91.7%)	8 (88.9%)
Southern Europe	2 (6.2%)	0	1 (11.1%)
Southern America	1 (3.1%)	0	0
Gypsy	0	3 (8.3%)	0
Other	1 (3.1%)	0	0
Sex			
Male	20 (62.5%)	23 (63.9%)	7 (77.8%)
Female	12 (37.5%)	13 (36.1%)	2 (22.2%)
Age years (mean [SD])	0.60 [0.42]	0.49 [0.30]	0.45 [0.30]
Past medical history prior to RSV admission			
Dermatitis	2 (6.2%)	10 (27.8%)	4 (44.4%)
Alimentary allergies	1 (3.1%)	2 (5.6%)	3 (33.3%)
Stational allergies	0	1(2.8%)	2 (22.2%)
Asma	0	0	0
Obesity	0	1 (2.8%)	0
Vitamin D supplement	22 (68.8%)	31 (86.1%)	7 (77.8%)
Premature new-born	4 (12.5%)	3 (8.3%)	1 (11.1%)
Admissions prior to RSV	11 (34.4%)	10 (27.8%)	3 (33.3%)
Annual bronchitis prior the RSV admission	3 (9.4%)	15 (41.7%)	7 (77.8%)
Family history			
Asthma	4 (12.5%)	9 (25.0%)	6 (66.7%)
Respiratory Problems	7 (21.9%)	13 (36.1%)	4 (44.4%)
Vitamin D supplement during pregnancy	3 (9.4%)	3 (8.3%)	1 (11.1%)
Clinical characteristics of the RSV hospitalization			
Respiratory distress			
Mild	10 (31.2%)	3 (8.3%)	1 (11.1%)
Moderate	17 (53.1%)	28 (77.8%)	6 (66.7%)
Severe	5 (15.6%)	5 (13.9%)	2 (22.2%)
Oxygen Need	26 (81.2%)	23 (63.9%)	5 (55.6%)
Respiratory support			
Non invasive	4 (12.5%)	5 (13.9%)	3 (33.3%)
Mechanical	2 (6.2%)	0	0
Diagnosis			
Bronquiolitis	28 (87.5%)	28 (77.8%)	7 (77.8%)
Bronchospasm	0	1 (2.8%)	0
Pneumonia	3 (9.4%)	2 (5.6%)	0
Other	0	5 (13.9%)	2 (22.2%)
Bacterial superinfection suspected	23 (71.9%)	10 (27.8%)	3 (33.3%)
Follow up 3 years			
Admissions during the 3 years follow-up	7 (21.9%)	11 (30.6%)	4 (44.4%)
Annual bronchitis during the 3 years follow-up	7 (21.9%)	34 (94.4%)	8 (88.9%)

Table 3: Clinical features of patients of the validation cohort and their classification in control, wheezing, and asthma.

	Control (*n* = 21)	**Wheezing (*n* = 12)**	**Asthma (*n* = 10)**
Ethnicity			
Western Europe	19 (90.5%)	10 (8.3%)	9 (90%)
Southern America	0	0	0
Gypsy	1 (4.8%)	1 (8.3%)	1(10%)
Northern Africa	1 (4.8%)	1 (8.3%)	0
Sex			
Male	12 (57.1%)	7 (58.3%)	6 (60%)
Female	9 (42.9%)	5 (41.7%)	4 (40%)
Age years (mean [SD])	0.30 [0.41]	0.32 [0.44]	0.19 [0.19]
Past medical history prior to RSV admission			
Dermatitis	9 (42.9%)	5 (41.7%)	2 (20%)
Alimentary allergies	2 (9.5%)	0	1 (10%)
Stational allergies	4 (19%)	2 (16.7%)	2 (20%)
Asma	0	0	0
Obesity	0	0	0
Vitamin D supplement	11 (52.4%)	6 (50%)	4 (40%)
Premature new-born	5 (23.8%)	1 (8.3%)	2 (20%)
Admissions prior to RSV	1 (4.8%)	1 (8.3%)	1 (10%)
Annual bronchitis prior the RSV admission	3 (14.3%)	1 (8.3%)	2 (20%)
Family history			
Asthma	8 (38.1%)	3 (25%)	8 (80%)
Respiratory Problems	3 (14.3%)	0	4 (40%)
Vitamin D supplement during pregnancy	2 (9.5%)	0	0
Clinical characteristics of the RSV hospitalization			
Respiratory distress			
Mild	3 (14.3%)	3 (25%)	1 (10%)
Moderate	10 (47.6%)	5 (41.7%)	4 (40%)
Severe	8 (38.1%)	4 (33.3%)	5 (50%)
Oxygen Need	21 (100%)	12 (100%)	13 (100%)
Respiratory support			
Non invasive	9 (42.9%)	5 (41.7%)	3 (30%)
Mechanical	4 (19%)	3 (25%)	3 (30%)
Diagnosis			
Bronquiolitis	19 (90.5%)	11 (91.7%)	10 (100%)
Bronchospasm	0	1 (8.3%)	0
Pneumonia	1 (4.8%)	0	0
Bacterial superinfection suspected	2 (9.5%)	5 (41.7%)	3 (30%)
Follow up 3 years			
Admissions during the 3 years follow-up	1 (4.8%)	1 (8.3%)	5 (50%)
Annual bronchitis during the 3 years follow-up	15 (71.4%)	12 (100%)	10 (100%)

3.1.2 Vaccine Response to PCV13*

In the second part of the project, 24 infants were selected from a previous clinical trial, where a group of 178 children was randomized to receive a booster dose of either the 10- or the 13-valent pneumococcal conjugate vaccine (PCV10/PCV13) at 12 months of age following the scheduled dose of PCV13 at 2 and 4 months of age [243, 244]. Detailed immune responses were measured at three-time points, before and at 1 and 12 months after the booster. At each study time point, anti-polysaccharide serum Immunoglobulin G (IgG) and opsonophagocytic assay (OPA) titers to all PCV13 serotypes were determined, and antigen-specific memory B cell responses were measured for selected pneumococcal polysaccharides. For this study, 24 individuals among those who had received a PCV13 booster dose at 12 months of age (n=90) (baseline group) were selected based on their aggregated IgG antibody response and classified into high and low responders. For this categorization, the intensity of the polysaccharide-specific IgG response was measured for each PCV13 recipient and vaccine serotype at 1 month following the booster. The overall IgG antibody response was then calculated through the determination of median rank for all serotypes. After excluding participants with missing serum or DNA samples, 38 individuals remained, of whom the 12 high and 12 low responders were selected (Figure 6 and Table 4). Of these 24 children, the DNA methylation profile was examined at two different time points: at 12 months, before the booster of PCV13, and at 24 months, after the vaccine booster.

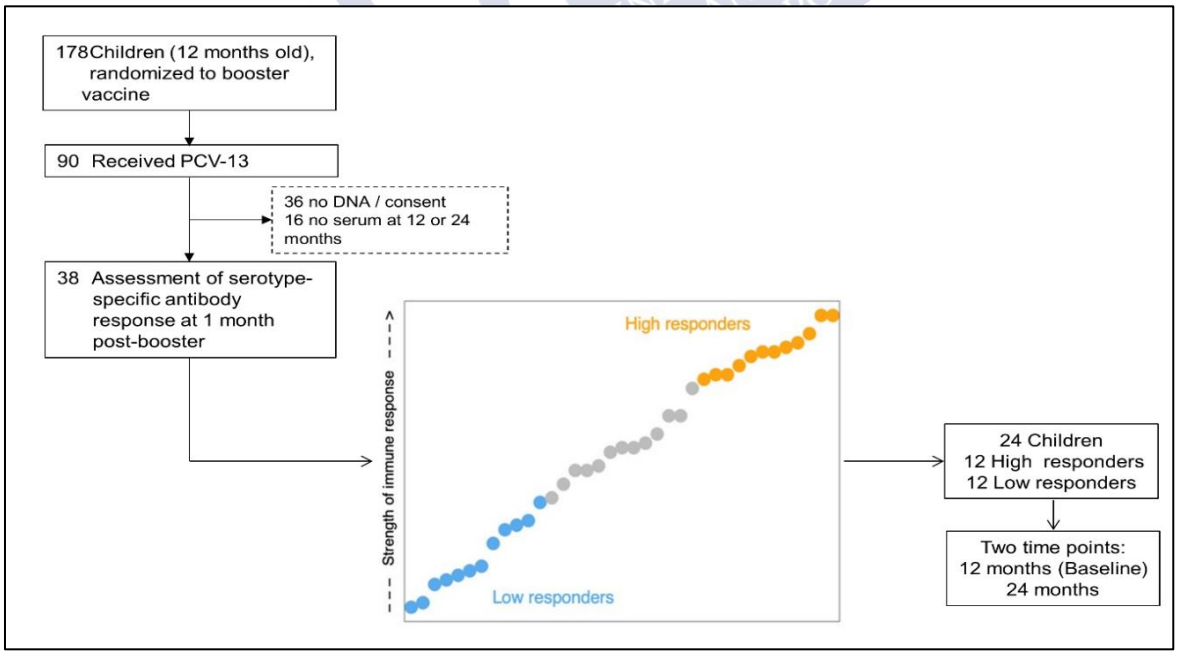


Figure 6: Flow diagram with the study design.*

* Pischedda, S., O'Connor, D., Fairfax, B.P. *et al.* Changes in epigenetic profiles throughout early childhood and their relationship to the response to pneumococcal vaccination. *Clin Epigenet* 13, 29 (2021). https://doi.org/10.1186/s13148-021-01012-w. (see Section 7. ANNEX)

Table 4: List of the participants of the study with clinical information

Response	Sex	Age at visit 1 (months)	Age at visit 2 (months)	Ethnicity
High	f	12.46	23.82	Indian/British
High	f	12.16	23.46	European/White west African
High	f	12.43	23.72	White Caucasian/European
High	f	12.75	24.05	White Caucasian/European
High	f	12.2	23.98	White Caucasian/European
High	m	12.72	24.57	White Caucasian/European
High	m	11.77	23.52	White Caucasian/European
High	m	12.72	24.25	White Caucasian/European
High	m	13.28	24.61	White Caucasian/European
High	m	12.75	24.28	White Caucasian/European
High	m	12.82	24.61	White Caucasian/European
High	m	12.85	24.08	White Caucasian/European
Low	f	12.56	23.89	White Caucasian/European
Low	f	12.49	23.85	White Caucasian/European
Low	m	12.39	23.95	White Caucasian/European
Low	m	12.16	23.66	Asian South East Asian Heritage
Low	m	12.98	24.74	White Caucasian/European
Low	m	12.39	23.46	White Caucasian/European
Low	m	12.00	23.26	White Caucasian/European
Low	m	13.05	24.38	White South African
Low	m	12.46	23.49	White Caucasian/European
Low	m	12.36	24.21	Canadian/British
Low	m	13.18	24.51	White Caucasian/European
Low	m	12.92	24.9	White Caucasian/European

3.2 SAMPLE PROCESSING

3.2.1 DNA isolation from blood samples

Blood samples were collected in EDTA tube, processed, and stored at 80º after collection. Genomic DNA was isolated using the Wizard® Genomic DNA Purification Kit (PROMEGA) QIAamp DNA mini kit (Qiagen) according to the manufacturer's instructions. Briefly, the purification was performed following three important steps consisting of cell lysis, nuclei lysis and protein precipitation, and finally DNA precipitation and rehydration.

3.2.2 DNA purification and concentration measurement

Detecting and quantitating the amount of DNA is extremely important for many molecular biological analysis methods and often it is the cause of the good success of the experiment. In the present study, after DNA extraction, DNA purity and concentration measurements were evaluated with Nanodrop in conjunction with the PicoGreen assay.

NanoDrop spectrophotometer is considered one of the most used methods to estimate nucleic acid concentration through the measurement of sample absorbance at 260 nm. The absorbance ratio 260/280 and 260/230 are used to determine DNA purity and the presence of contaminants in the biological samples during the DNA extraction process. Following the protocol recommendation, the NanoDrop was used at first to evaluate DNA quantification and quality after DNA blood extraction. Before proceeding with bisulfite conversion, the quality and the quantification of DNA samples were again assessed with the PicoGreen assay which provides a highly sensitive way to quantify dsDNA with minimal consumption of the sample. The main advantage of this assay is the ability to overcome the typical drawbacks produced by the general UV spectroscopy, such as the contribution of signal from single-stranded DNA (ssDNA) and other contaminants, such as protein and extraction buffers, and the inability to distinguish between DNA and RNA. The PicoGreen reagents don’t interfere with free nucleotides or other contaminants and provide a more accurate quantitation for the success of downstream experiments. In the present project, the Quan-iT PicoGreen dsDNA Assay (Life 138 Technologies) whom necessary reagents are all provided in the kit, was implemented.

3.2.3 Bisulfite conversion

Sodium bisulfite conversion is the step to which genomic DNA from whole blood or saliva samples was subjected before proceeding with the Illumina microarray. This technique is considered a gold standard approach to investigate the methylation state of CpGs sites within the whole genome. It basically consists of the conversion of unmethylated cytosine residues to uracil through deamination using sodium bisulfite, while leaving the methylated cytosine (5-mC) unaffected. In this way, the uracils are amplified in the subsequent PCR reaction as thymines, whereas 5-mC or 5-hmC residues remain cytosines.

The EZ-96 DNA Methylation kit (Zymo Research Corp) adopted here to perform bisulfite conversion is based on a three-step reaction that takes place between cytosine and sodium bisulfite where cytosine is converted in uracil. The technique performed following the manufacturer's recommendations for Infinium assays briefly consists of:

i. the addition of an M-dilution buffer to each DNA sample in a Conversion plate;
ii. the incubation of the plate at 37º for 15 minutes;
iii. the addition of a conversion reagent (CR) to each sample in the plate after incubation;
iv. the incubation of the plate in the dark at 50º for 12-15 hours;
v. further incubation of the samples at 4º for 10 minutes;
vi. the addition of a binding buffer to each well of a Zymo-Spin™ I-96 Binding Plate on a Collection Plate followed by centrifugation of the plate at ≥ 3,000 x g for 5 minutes;
vii. the addition of a wash buffer and second centrifugation like the previous one;
viii. the addition of an M-Desulphonation Buffer to each well, and incubation at room temperature for 15-20 minutes, and further centrifugation at ≥ 3,000 x g for 5 minutes.
ix. the addition of the M-Wash Buffer and the subsequent centrifugation repeated twice;
x. the positioning of the Zymo-Spin™ I-96 Binding Plate onto an Elution Plate, the addition of an M-Elution Buffer directly to the binding matrix in each well, and centrifugation for 3 minutes at ≥ 3,000 x g to elute the DNA.

After this last process DNA is ready for later use. Complete bisulfite conversion is necessary for a successful and robust DNA methylation study.

3.3 DNA METHYLATION PROFILING

Several technologies have been developed to map DNA methylation on large scale. The main aim of DNA methylation profiling is generally the detection of differentially methylated positions (DMPs) o differentially methylated regions (DMRs) that can be used as candidate biomarkers for future analysis. In 2010 Bock and collaborators with a disease-centered case-control study evaluated a comprehensible comparison among the different kinds of methods existing to study DNA methylation such as the methylation-specific restriction enzyme digestion, the bisulfite DNA sequencing, or the methylation-specific PCR [245]. The method used depends on different factors that go from the availability of the DNA, the total number of targets being analyzed, and the desired specificity. In recent years novel microarrays and sequencing technology have emerged to provide high-throughput access to specific areas of the genome at high resolution and in large samples.

Regardless of the methods used to map DNA methylation, the treatment of DNA with bisulfite is the foundation for most downstream assays geared toward analyzing changes in DNA methylation patterns. Bisulfite modification, as mentioned before, is a “gold standard” approach to which samples are submitted before methylation analysis to allow the recognition of uracils as thymines and 5-mC or 5-hmC as cytosines, thereby differentiating between methylated and unmethylated cytosine. Because the estimated level of DNA methylation depends on bisulfite treatment, it is crucial to ensure complete conversion of non-methylated cytosines before performing other types of analysis.

Sequencing-based methods are the most used to map DNA methylation, and bisulfite sequencing is considered today the standard unsupervised approach to study gene expression regulation. It includes the Whole-Genome Bisulfite Sequencing (WGBS) and the Reduced-Representation Bisulfite Sequencing (RRBS) and involves the library preparation with the adaptor ligation step and PCR amplification. The main limitation of this technology is the large amount of sample input required by the platform, as well as the complexity of the bioinformatic analysis.

Microarray-based DNA methylation profiling is a new field, recently developed, to have access to many genes' methylation status or to the entire genome. Besides, arrays offer a reliable cost-effective way of analyzing multiple targets on a large scale. Illumina has adapted its GoldenGate SNP genotyping assay [246] to recognize bisulfite-converted DNA for the

investigation of DNA methylation status [247, 248]. Illumina platforms allow high sample throughput with single-CpG resolution.

The Illumina BeadArray assay uses oligonucleotides conjugated to bead types to measure specific target sequences, measuring multiple beads per bead type, and it is able to combine the bead-based array platform with scalable automation for sample handling and data processing. With the GoldenGate Illumina assay, it was possible to detect 1,536 different CpG sites showing changes in their methylation status simultaneously using 200 ng of human genomic DNA [247]. The biggest advantage of this technology is the possibility to build probes that can specifically interrogate CpG sites in the genome, opening new avenues to large-scale discovery, validation, and clinical application for DNA methylation biomarkers in human health and disease. Besides, the technology should prove useful for comprehensive methylation analyses in large populations.

3.3.1 Illumina Infinium Methylation BeadChip

Although effective, these microarray platforms required large amounts of sample and rely heavily on bioinformatic analyses, making them difficult to use in large-scale studies where sample amounts may be limited. To overcome these challenges, in 2008 Illumina launched a new methylation BeadChip array, based on the Infinium technology, previously described for SNP genotyping [249]. To interrogate individual CpG sites within a given DNA sample, the Infinium assay design uses beads that display long, target-specific probes. According to the protocol, the previously bisulfite-converted sample is amplified and enzymatically fragmented; an endpoint fragmentation is used to prevent over fragmentation. The DNA samples are then precipitated and resuspended in a hybridization buffer before dispensing them onto the BeadChips. In the hybridization step, samples are loaded into the Hyb Chamber inserts divided by an IntelliHyb® seal (or gasket), that are then incubated in the Illumina Hybridization Oven for 16–24 hours at 48°C. The amplified and fragmented DNA samples anneal to locus-specific 50-mers attached to each bead of the chip during hybridization. To prepare samples for the staining and extension process, the BeadChips are washed to remove unhybridized and non-specifically hybridized DNA samples and assembled into flow-through chambers. Single-base extension of the oligos on the BeadChip, at the 3'terminus of the oligonucleotide using fluorescently labeled ddNTPs, determines the genotyping' of the C/T conversion that results from bisulfite conversion. Finally, after extension, the array is fluorescently stained and scanned

using iScan or HiScan SQ system (Illumina), which uses a laser to excite the fluorophore of the single-base extension product on the beads. The high-resolution images of the light emitted from the fluorophores are recorded by the scanner; the measure of the level of methylation for the interrogated locus is determined by calculating the ratio of the fluorescent signals from the methylated vs. unmethylated sites [250]. In this scanning procedure, decoded data (DMAP) files, provided by Illumina and containing information of the bead-type for each position on the BeadChip are required to generate raw data files (IDAT files).

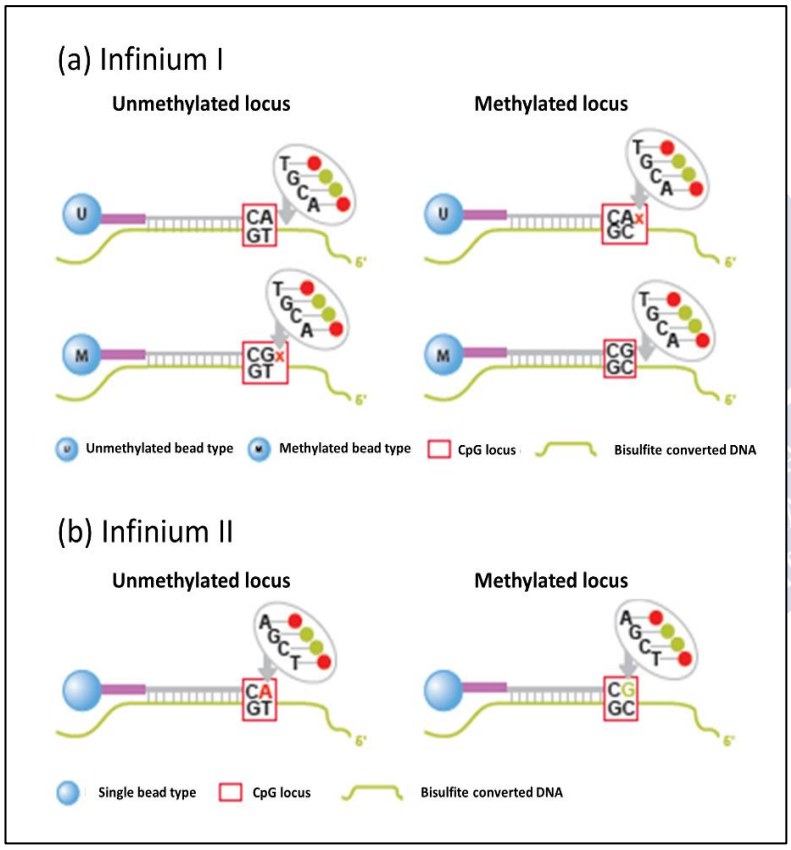


Figure 7: Illumina Infinium Assay. (a) type I,each individual CpG is interrogated using two bead types: methylated (M) which match with methylated CpG site and unmethylated (U) which join the unmethylated CpG site. When the unmethylated CpG target site matches with the U probe, single-base extension and detection occur. On the contrary, if unmethylated CpG has a single-base mismatch to an M probe, the extension was inhibited. The reverse occurs with methylated CpGs.(b) Infinium type II where each target CpG is interrogated using a single bead type that distinguishing between "methylated" and "unmethylated" through different dye colors (green and red). Illumina HumanMethylation BeadChip for Genome-Wide DNA Methylation Profiling: Advantages and Limitations. Kazuhiko Nakabayashi. Publisher: Springer Nature. Adaptations/modifications - Springer Nature allows adaptation of figures for style and formatting purposes under this license under the condition that this does not alter the meaning of the content.

HumanMethylation BeadChip uses two assay designs: Infinium I consisting of two site-specific probes per CpG locus: one "unmethylated" and one "methylated"; both bead types will incorporate in the CpG locus the same labeled nucleotide, determined by the base preceding the interrogated "C", and therefore will be detected in the same color channel (Figure 7a). This assay assumes that the methylation status of adjacent CpG sites tends to be correlated. The Infinium II assay, instead, requires only one bead type per CpG, distinguishing between "methylated" and "unmethylated" through different dye colors (green

and red) (Figure 7b) [251]. In this assay, the 3' terminus of the 50-mer probes presents a cytosine complementary to the guanine of the CpG site investigated. A single-base extension of G or A base occurs, depending on the complementarity to either the C (methylated) or T (unmethylated). Infinium II probes can contain up to 3 CpG sites within a 50-mer probe sequence and the underlying CpG sites may be represented by "degenerate" R-bases (A or G base), allowing multiple combinations of oligos attached to the bead. This factor allows the assessment of the methylation status of a query site independently of the status of neighboring CpG sites thus increasing the number of CpG sites interrogated. The main difference between the two assays is that the Infinium II design, because of its chemistry, can be applied whenever possible, while the Infinium I design is required generally to measure the DNA methylation levels of CpG sites within CpG-rich regions such as CpG islands. Dedeurwaerder and colleagues demonstrated that Infinium II assays show larger variance and are less sensitive for the detection of extreme methylation values than the Infinium I, probably because of the dual-channel read-out; this explains why the Infinium I assay is a better estimator of the true methylation state [252]. The first Illumina Methylation assay based only on the Infinium I technology was the Human Methylation BeadChip 27K able to interrogate 27,578 CpG sites associated with 14,495 protein-coding gene promoters [247]. The results achieved with this first Infinium assay, although similar to those produced on other DNA methylation platforms show great advantages as they require only 4 days to produce methylome profiles of human samples using only 250–500 ng of genomic DNA as a starting material. For these reasons Illumina in the last decade has developed other two Infinium Methylation assay platforms: the HumanMethylation450 BeadChip in 2011 [253] and its successor HumanMethylationEPIC in 2015 [254], both including the two distinct probe types, Infinium I (n=135,501) and Infinium II (n=350,076) [251]. The Infinium Methylation 450K BeadChip, considered one of the most popular and cost-effective tools available for DNA methylation studies, has been designed with the idea to combine the benefits of Infinium chemistry with expanded genome coverage to provide high quality, genome-wide content with target selection. The array allows researchers to interrogate more than 485,000 methylation loci per sample at single-nucleotide resolution with a coverage of 482,421 CpG sites, including 90% of the sites on the 27K array and 3,343 sites corresponding to CNG targets.

The Infinium Methylation 450K BeadChip shows a gene coverage that includes a total of 21,231 UCSC RefGenes, plus additional genes and transcripts not covered by the UCSC

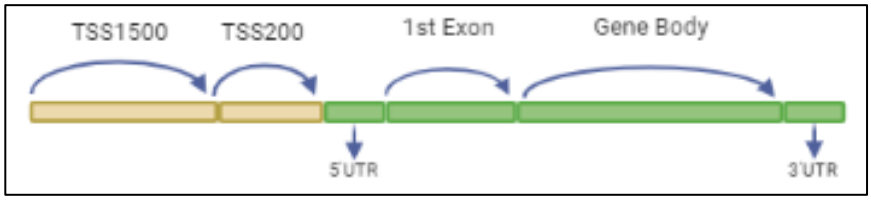


Figure 8: Genomic context where CpGs can be located.
Self-made figure. Created with BioRender.

database. To obtain a better comprehensive evaluation of gene region methylation, probes covering gene regions were subdivided into sub-regions targeting: promoters, region 200 base pairs within the transcription start site (TSS), and region 1500 base pairs within the TSS excluding the TSS 200 region (designated as TSS200 and TSS1500, respectively), the 5′ and 3′ untranslated regions (UTR), the first exon and gene body (Figure 8).

The coverage of CpG island, considered as regions of DNA greater than 500 bp with a GC base composition greater than 50% and a CpG observed/expected ratio of more than 0.6 [255], was of 26,658 CpG islands covered overall with an average of 5.63 sites each. The coverage of "CpG island shore" includes 28,249 "north" or upstream and 25,761 "south," or downstream, found immediately outside of the CpG islands, were targeted with averages of 2.93 and 2.81 sites, respectively. The "CpG island shelves", the 2 kb regions upstream and downstream of the "CpG island shores", were targeted with global averages of 2.07 and 2.03 sites each ("north" and "south," respectively).

Recently Illumina has released the EPIC BeadChip Array, allowing the analysis of more than 860,000 CpG sites including more than 90% of Human Methylation 450K content and 413,743 additional CpG sites. The newly added CpG sites include 35,000 CpG sites located in enhancer regions identified by the FANTOM5 (http://fantom.gsc.riken.jp/5/) and the ENCODE projects (http://www.encodeproject.org/). The 413,743 CpG newly added positions in the EPIC microarray are localized among all 22 autosomal chromosomes in addition to the X and Y sex chromosomes, where chromosome 1 contains the most positions (39,087, 9.4%) and chromosome Y the fewest (179, 0.04%). In 2018 Solomon et al., analyzing umbilical cord blood and whole blood of children at age of 14 observed a high correlation between EPIC and 450K, as well as a high correlation of cell-type proportion estimation [256]. However, they discovered that probes of the Infinium assay II (72% for 450K and 84% for EPIC) presented higher individual site correlations than type I between the two platforms and also report a relatively small subset of CpG sites with consistently large Beta value differences.

3.3.1.1 Beta values and M values

DNA methylation values, recorded for each locus in each sample, can be described as Beta values, or M values. As described by Du et colleagues in 2010, Beta value is measured as:

$$\text{Beta} = \frac{\max(\text{methylated}, 0)}{\max(\text{methylated}, 0) + \max(\text{unmethylated}, 0) + \alpha}$$

where methylated and unmethylated are the intensities measured by methylated and unmethylated probes, respectively [257]. In order to prevent negative values after the background adjustment, any negative values will be reset to 0. The constant offset α (by default, α = 100) to the denominator is added as a recommendation of Illumina to regularize Beta value when both methylated and unmethylated probe intensities are low. The Beta value statistic results in a number between 0 and 1, or 0 and 100%, where ideally 0 represents that no positions are methylated in any of the cells and 1 defines positions that are methylated in all cells in the sample.

The M value, instead, is calculated as the log2 ratio of the intensities of methylated probe versus unmethylated probe:

$$M = \frac{\max(\text{methylated}, 0) + \alpha}{\max(\text{unmethylated}, 0) + \alpha}$$

M values close to 0 depict a similar intensity between methylated and unmethylated probes, indicating that the CpG site is about half-methylated when considering a proper normalization for the intensity data has been by Illumina GenomeStudio or some other external normalization algorithm. Positive M values indicate that there are more methylated than unmethylated molecules, while negative M values display the opposite.

In statistical data analyses carried out in microarray data (such as linear models or ANOVA), it is usually assumed that the variable variances are approximately constant, assuming homoscedasticity. The violation of this assumption, which is described as heteroscedasticity in statistics, imposes serious challenges when applying these analyses to

high-throughput data [258]. To check the homoscedasticity of the data is important to visualize the relations between mean and standard deviation (Figure 9).

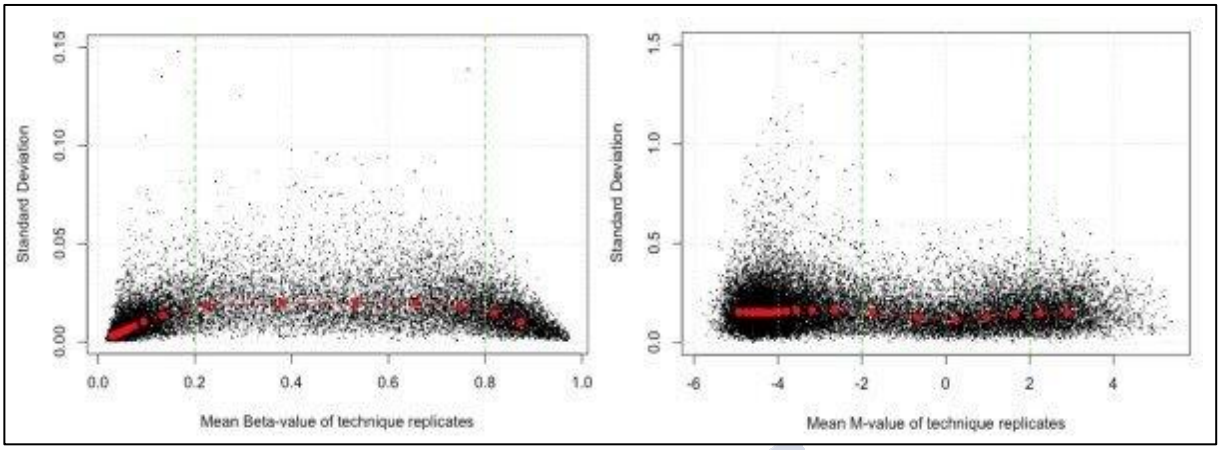


Figure 9: Mean and standard deviations o Beta and M values. The standard deviation of Beta value shows heteroscedasticity in the low and high methylation range, while the standard deviation of M value is maintained approximately constant across the entire methylation range for M values. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. Pan Du et al. Publisher: Springer Nature. Creative Commons. This is an open access article distributed under the terms of the Creative Commons CC BY license.

It is quite visible that the standard deviation of Beta value is greatly compressed in the low (between 0 and 0.2) and high (between 0.8 and 1) ranges. This confirms that significant heteroscedasticity in the low and high methylation range is present for Beta values means. However, when converting Beta value to M value the problem of heteroscedasticity is effectively overcome. Therefore, for the most statistical model used to analyze microarray data the M value statistic is more suitable due to its homoscedastic assumptions. Beta values are preferred for data visualization and are used in text and figures since they are more intuitive for biological interpretation.

3.3.1.2 450K and EPIC Beadarray Platform

The main concept behind the methylation profiling platform is that of common arrays, which are made of oligos bound to silica beads that are randomly assembled into wells. Illumina Infinium BeadChips are sophisticated technologies that use three-micron silica beads attached to hundreds of thousands of copies of a specific oligonucleotide containing long 50-mer locus-specific sequences concatenated to a decoding sequence. The complexity of this assay is that the beads used present a long-target specific probes design to correspond to each CpGs. While the HumanMethylation27, was built using only Infinium I, for the 450K and EPIC Methylation

BeadChip two types of Infinium technology are combined. Physically, each sample is measured on a single "array". For the 450K design, 12 arrays can be observed on a single physical "glass slide" or beadchip, which is arranged in 6 rows and 2 columns (Figure 10a), while the EPIC array has been characterized by 8 arrays per slide (Figure 10b). An IntelliHyb seal separates the sample sections of the slide to allow running multiple samples simultaneously. Because the microwells are in a higher number than the distinct bead types, multiple copies of each bead type are present in the array. This redundancy improves robustness and measurement precision.

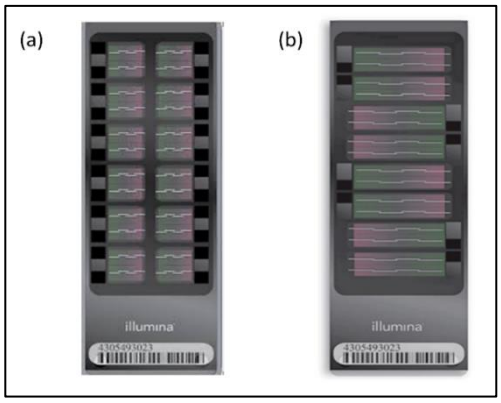


Figure 10: (a) The Infinium HumanMethylation450 BeadChip, and (b) the Infinium HumanMethylation EPIC BeadChip (Illumina). Images from Illumina for public general use https://www.illumina.com/company/news-center/multimedia-images.html

3.3.1.3 Processing of raw data

In DNA methylation studies, choosing the right pre-processing pipeline must be done carefully to maximize the statistical power of the analysis and the robustness of the results. DNA Methylation data pre-processing and normalization is critical when analyzing the genome-wide methylation Illumina 450K and the EPIC array data since there exist several challenges including technical challenges dealing with batch effects and the underlying biochemistry employed by the array methods. Therefore, it is crucial to focus on the entire framework for dealing with methylation data. In the present project, we followed the framework of DNA methylation analysis reviewed and illustrated by Wilhelm-Benartzi in 2013 [259]. All the processes, involving DNA methylation quality control, pre-processing, normalization, bias correction, and downstream analysis were performed in R using different Bioconductor packages. Several methods and different R packages have been proposed for this kind of analysis. In Figure 11 it can be observed a brief diagram with the pipeline performed to process DNA methylation data built considering the *minfi* User's guide for 450K arrays [260] and the work of Fortin et al., [261] for EPIC arrays. Here, for the pre-processing of raw data, the package minfi (v1.30.0), considered one of the most appropriate and suitable R packages for the 450K and EPIC data investigation, was applied.

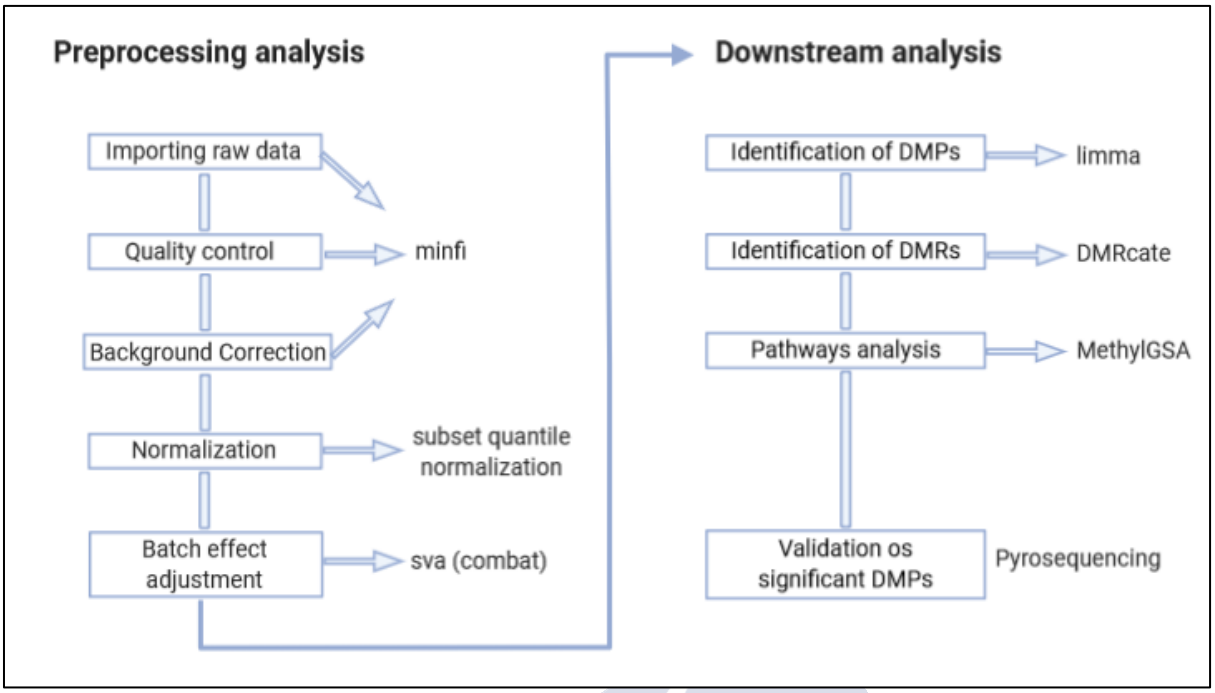


Figure 11: Pipeline used to process DNA methylation data from Illumina Infinium Methylation 450K BeadChip and Infinium Methylation EPIC BeadChip. Self-made figure. Created with BioRender.

The IDAT files representing the different color channels prior to normalization are the input to start preprocessing analysis with *minfi*. These raw intensity data are the most complete raw methylation data derived from arrays as they also include the measurements on the control probes. IDAT files were loaded in R together with a "Sample Sheet" CSV file based on a sample sheet file provided by Illumina that describes the layout of the experiment. In this file, it is possible to find the information for each sample related to its position on the plate and the array, however, this file can be completed by adding other important features, such as the phenotype, the age, and the gender of the samples analyzed that could be useful in downstream analysis.

The following step in this preliminary process is the annotation of each CpG of the methylation array; for this purpose, two types of annotation packages can be used: the "manifest", containing the array design, and the "annotation" package including information about the position of the methylation loci in the genome, about the genomic features they map to and the known SNPs they can overlap. These files are essential to identify and classify the significant DMPs and DMRs. Here, the annotation packages I*lluminaHumanMethylation450kanno.ilmn12.hg19* for the Illumina Infinium Methylation 450K BeadChip [262] and *IlluminaHumanMethylationEPICanno.ilm10b4.hg19* for EPIC BeadChip [263] were adopted, both based on the coordinates from Human Genome assembly hg19 (Table 5).

Table 5: Infinium MethylationEPIC and Methylation 450K Manifest Column Headings for Differentially Methylated Positions

Column Headings	Description
CpG_ID	Unique identifier from the Illumina CG database. (The probe ID).
Chr	Chromosome containing the CpG (Build 37).
Pos	Chromosomal coordinates of the CpG (Build 37).
Strand	The Forward (F) or Reverse (R) designation of the Design Strand.
AddressA_ID	For Infinium I beadtypes, this is the Address ID for the probe specific for the A allele. For Infinium II beadtypes, this is the Address ID for the probe used for both A and B alleles (in this case, AddressB_ID and AlleleB_ProbeSeq columns will be empty).
AddressB_ID	For Infinium I beadtypes, the address ID for the probe specific for the B allele.
ProbeSeqA	The sequence of the probe identified in the AddressA_ID column.
ProbeSeqB	For Infinium I beadtypes, the sequence of the probe identified in the AddressB_ID column.
Infinium_Design_Type	Infinium I (2 probes/locus) or Infinium II (1 probe/locus).
Next_Base	For Infinium I probes, the nucleotide immediately following the CpG. Blank for Infinium II.
Color_Channel	For Infinium I probes, the color channel of the Next_Base signal.
SNP_ID	rsids of SNPs located in the probe. Multiple listings of SNP rsids are allowed.
SNP_DISTANCE	Distance of SNPs from query base of the probe. Multiple listings of the distance values are associated with rsid.
SNP_MinorAlleleFrequency	Minor allele frequency of SNPs. Multiple listings of the minor allele frequencies are associated with rsid.
UCSC_CpG_Islands_Name	Chromosomal coordinates of the CpG Island from UCSC.
Relation_to_UCSC_CpG_Island	The location of the CpG relative to the CpG island.
Forward_Sequence	Plus (+) strand (HapMap) sequence (5'-3') flanking the CG.
SourceSeq	The original, genomic sequence used for probe design before bisulfite conversion
UCSC_RefGene_Name	Target gene names, from the UCSC database. Multiple listings of the same gene name indicate splice variants.
UCSC_RefGene_Accession	The UCSC accession numbers of the target transcripts. Accession numbers are in the same order as the target gene transcripts.
UCSC_RefGene_Group	Gene region feature category describing the CpG position, from UCSC. Features are listed in the same order as the target gene transcripts. TSS200 = 0–200 bases upstream of the transcriptional start site (TSS).
Phantom5_Enhancers	Chromosomal coordinates from the FANTOM consortium of enhancer regions associated with FANTOM5promoters.
DMR	Differentially methylated regions (experimentally determined).
450k_Enhancer	Predicted enhancer elements as annotated in the original 450K design (informatically determined by the ENCODE Consortium) are marked True.
HMM_Island	Hidden Markov Model Islands. Chromosomal map coordinates of computationally predicted CpG islands.
Regulatory_Feature_Name	Chromosomal map coordinates of the regulatory feature (informatically determined by the ENCODE Consortium).
Regulatory_Feature_Group	Description of the regulatory feature referenced in Regulatory_Feature_Name, as provided by the Methylation Consortium.
GencodeBasicV12_NAME	Target gene names, from the basic GENECODE build. Multiple listings of the same gene name indicate splice variants.

Column Headings	Description
GencodeBasicV12_Accession	The basic GENECODE accession numbers of the target transcripts. Accession numbers are in the same order as the target gene transcripts.
GencodeBasicV12_Group	Gene region feature category describing the CpG position, from basic GENECODE. Features are listed in the same order as the target gene transcripts.
GencodeCompV12_NAME	Target gene names, from the complete GENECODE build. Multiple listings of the same gene name indicate splice variants.
GencodeCompV12_Accession	The complete GENECODE accession numbers of the target transcripts. Accession numbers are in the same order as the target gene transcripts.
GencodeCompV12_Group	Gene region feature category describing the CpG position, from complete GENECODE. Features are listed in the same order as the target gene transcripts.
DNase_Hypersensitivity_NAME	Chromosomal coordinates of DNase hypersensitivity regions from ENCODE.
DNase_Hypersensitivity_Evidence_Count	Number of experimental sources supporting the identification of each DNase hypersensitivity region from ENCODE.
OpenChromatin_NAME	Chromosomal coordinates of open chromatin regions from ENCODE.
OpenChromatin_Evidence_Count	Number of experimental sources supporting the identification of each open chromatin region from ENCODE.
TFBS_NAME	Chromosomal coordinates of transcription factor binding sites from ENCODE.
TFBS_Evidence_Count	Number of experimental sources supporting the identification of each transcription factor binding site from ENCODE.
Methyl27_Loci	CpG's carried over from the HumanMethylation27 array (92% carryover) are marked "True".
Methyl450_Loci	CpGs carried over from the HumanMethylation450 array (94% carryover) are marked True.
Random_Loci	CpG loci chosen randomly by consortium members during the design process are marked True.

3.3.1.3.1 Sample's quality control

The Infinium arrays contain several control probes to control the data quality, including sample-independent and -dependent controls. To detect poorly performing samples in Illumina arrays, *minfi* has provided a quality control report based on the intrinsic control probes present on the array (such as staining, hybridization, bisulfite conversion, and negative controls) together with the ability to eliminate probes and samples according to their probe signal intensity. A series of plots can be constructed to visualize the quality control check: the QC plot uses the log median intensity in both the methylated (M) and unmethylated (U) channels and when plotting the two medians against each other, good samples cluster together, while failed samples tend to separate due to their lower median intensities. The density plot, instead, is used to represent the methylation Beta density for all samples, usually colored by sample group. In case of concern about outlier samples, the "bean" plot shows each sample in its own section.

Sex represents one of the strongest factors influencing the genomic distribution of DNA methylation and sample traits across differential effect studies; it has a particular impact and effect on predisposition, development, and prognosis for many diseases. The getSex function of *minfi* allows the estimation of the copy number of sex chromosomes from the total signal intensities (the sum of methylated and unmethylated signals) for the sex chromosomes and predicts sex distinguishing two different clusters of points corresponding to which gender the samples belong, expecting one X chromosome (chrX) and one Y chromosome (chrY) for typical males and two chrX and no chrY for typical females.

3.3.1.3.2 Probe quality control and filtering processes

For probes quality control the array contains several internal control probes that can be used to assess the quality control of different sample preparation steps (bisulfite conversion, hybridization, etc.). The values of these control probes can be plotted specifying the control probe type. Before downstream analysis, it is important to filter out probes with potential errors, and this is an essential step in data pre-processing steps. The first filter involves the removal of probes that have a detection *P*-value below a certain prespecified threshold, which according to Illumina's recommendation is *P*-value <0.05. Samples with > 1% of sites with high detection p-value should be removed. The detection *P*-value is a statistical calculation providing the probability that the signal from a given probe is greater than the average signal from the negative control probes which are sequences randomly permutated that should not hybridize to the DNA template. The detection *P*-value is generated for every CpG in every sample, as an indicative measure of the quality of the signal. In *minfi* it is calculated comparing the total signal (Methylated +Unmethylated) for each probe to the negative control probes whose mean signal defines the background signal level. Very small *P*-values reflect a reliable signal.

The second step in the filtering process consists of the exclusion of probes with a bead count < 3, which means probes that are not represented by a minimum of 3 beads on the array in at least 5% of samples per probe. The third filter involves the removal of probes located in sex chromosomes as they can affect the global results, because of the influence of the sex factor on the downstream results. Also, sites containing SNPs or with a minor allele frequency (MAF) of at least 5%, should be removed. In fact, the presence of SNPs at the CpG loci targeted by Illumina Infinium probes is one of the most important confounding variables in microarray analysis, especially in the case where there is a comparison between the methylation level of individuals from genetically non-homogeneous populations. Probes targeting CpG sites that

overlap known SNPs usually are referred to as polymorphic CpGs. The difficulty in distinguishing between true differential methylation at a CpG site and the presence of a SNP at that same site may create bias in methylation data, bringing to erroneous conclusions from affected epigenetic studies. Daca-Roszak et al., [264] showed that more than 68% of interrogated CpGs carried SNPs with strongly differentiating allele frequencies in inter-population comparisons. The characterization of the theoretical and practical impact of SNPs on the methylation data is important for the interpretation of population data. Here, the approach used to remove all probe locations known to hide human genetic variants before investigating methylation is the dropLociWithSNP() function from *minfi*. From a study of 2014 [265] 190,672 probes that include more than 70K target, CpG SNPs from the Illumina 450K methylation array data were observed, accounting for a loss of 39% of the available CpG sites.

The last filtering consists of the elimination of probes known to have cross-reaction. Cross-reactivity is considered one of the major technical artifacts in DNA methylation analysis; it has been demonstrated that cross-reactivity is the result of probes targeting repetitive sequences or hybridizing to sequences that are highly homologous to the target sequences or are pseudogenes. This could generate false signals, resulting in spurious conclusions and a lack of validation in downstream analysis. The sites affected by cross-reactivity could reflect CpGs with different methylation statuses or non-CpGs that can be detected as fully methylated or unmethylated loci. The cross-reactive probes were originally discovered in 2013 by Chen and colleagues, which, investigating sex-associated DNA methylation on autosomes, observed that the top candidate CpGs were targeted by probes with sequences that also mapped to the sex chromosomes with high identity matches [266]. They reported that 6% of the Illumina 450K microarray probes were cross-reactive, similarly to the results obtained analyzing the Human Methylation 27K two years before [267], where 6-10% of probes were found to be non-specific and mapping to highly homologous genomic sequences. Concerning the Human Methylation EPIC BeadChip, among the 863,904 CpG-targeting probes present on the array, only 4.9% of total CpG-targeting probes were identified as potentially cross-hybridizing according to a study of 2016 [268].

After all these filtering processes, quality control was performed again to visualize the data, before starting with the downstream analysis. The check of these factors is fundamental as they can affect the determination of Beta or M values, lead, in case, to a false interpretation of Illumina microarray results.

3.3.1.3.3 Signal and background correction

The Infinium Human Methylation 450K and EPIC BeadChip are two-color assays and technical noise may be introduced by the variation in background fluorescence signal across arrays and color channel where the intensities measured might be imbalanced. Therefore, a color balance adjustment should be necessary and important if the color effect is inconsistent across samples. There exist a kind of within-array normalization process that allows correcting for background fluorescence and dye bias; it is called the single-sample Noob (ssNoob) method, and it is particularly suitable and useful for analyzing large datasets obtained from different batches and when integrating data from multiple generations of Infinium methylation arrays, that is when combining data from different Infinium platforms. This background subtraction approach allows background noise estimation, removing it for each sample separately, and provide dye-bias normalization using a control probes subset to detect dye bias. The background-corrected intensities are then normalized using one of the numerous normalization processes.

3.3.1.3.4 Normalization processes

When looking at the density profile of M values, a mixture of the distribution of Infinium I and Infinium II could be seen as an intrinsic characteristic of the array. In fact, in the Infinium methylation 450K and EPIC BeadChip there is a combination of two different assay chemistries, which may cause a bias in the analysis if all signals are considered together as a unique source of methylation measurement. The bias induced is the same that can be created when comparing probes from two different technologies with different signal distributions using a single statistical approach. Infinium I and Infinium II probes can be considered as two different subarrays, and their intensities distributions should be normalized.

Several methodologies have been developed to correct the differences between Infinium I and Infinium II. When Dedeurwaerder and colleagues observed that signals of Infinium I were more stable than Infinium II signals, they create a peak-based approach, to adapt the density distribution of Infinium II probes based on those of Infinium I. However, the choice of the correct normalization methods depends, principally, on the source of the samples and particularly on the results researchers are expected to find. For instance, it has been demonstrated that if a known large-scale biological methylation difference between samples is expected, such as cancer vs.

normal samples, or if a comparison between different tissues/cell types is performed, the functional normalization is more suitable for this kind of analysis [269]. In the case in which global differences between samples or tissues are not expected, the quantile normalization process is the best choice. In gene expression analysis the subset quantile normalization (SQN) is used to normalize the gene expression signal with the help of negative control probes that are supposed to produce constant measurements across samples [270]. Following this approach, Touleimat and Tost in 2012 launched a Subset Quantile Normalization for methylation data based on the assumption that Beta values of each CpG from the same biological category should have the same density distribution [271]. Because of the two types of annotations provided by Illumina provides, they implemented two versions of the SQN approach. The first one based on the 'relation to CpG' annotation (S shore, S shelf, N shore, N shelf, and distant), and the second one is related to the association to gene sequence annotation (body, 5´-UTR, 3´-UTR, 1st exon). From their results, the SQN method based on the relation to CpG annotations successfully achieved the complete correction of the shift between the probe signals of Infinium I and Infinium II probe signals and perform an efficient quantile normalization.

3.3.1.3.5 Cell type heterogeneity

Because of the extreme cell-type specificity of DNA methylation [272], variations in cell-type composition between phenotypes can confound analyses and bring to a wrong interpretation of the results. In fact, the majority of tissues accessible for DNA methylation studies, such as saliva, whole blood, placenta, or many others, are characterized by mixtures of different cell types, with different degrees of variation. Therefore, mixed cell methylation data exhibit a profile which is a convolution of the profiles from the constituent cell types. This produces several issues when conducting EWAS with mixed cell tissue samples. First of all, methylation differences between cell types could introduce a large amount of variation which probably is unrelated to the phenotype of interest [273]. Then, the phenotype changes related to the composition of cell-type are represented as methylation changes in the mixture of the cell and could contribute to false associations at these loci. Finally, phenotype associations with less prevalent cell types are unlikely to be detected compared to that of more common cell types. Beta values from mixed cell samples have been designed as a linear combination of the underlying cell-type-specific methylation levels, weighted by the cell-type proportions of the sample [274].

Blood cell subclasses originated through a developmental process known as hematopoiesis; the white blood cells (leukocyte), are critical in the host response to pathogens, and their population composition reflects disease states and can be modulated by signaling cascades that prompt migration of whole classes of cells into or out of tissues. For this reason, counting blood cells is a crucial step necessary to adjust for individual differences in cellular heterogeneity in the blood sample from which genomic DNA was extracted. Several DMRs have been identified as reliable biomarkers of individual human white blood cell types. There exist assays identifying cell-specific DMRs useful for quantifying individual cell types in human tissues and peripheral blood [275]. However, the biggest limitation of these assays lies in detecting the relative proportion of one individual cell type compared with all others. Houseman et al., in 2012 employed the concept of DMRs as markers of immune cell identity to estimate the proportions of immune cells in whole blood [276]. The strength of this method is that DNA methylation signature is associated with each of the principal immune components of whole blood such as B cells, T cells, granulocytes, monocytes, and NK cells.

Following this approach, before checking for differentially methylated positions and differentially methylated regions, the FlowSorted.Blood.450K and FlowSorted.Blood.EPIC packages for the estimation of blood cell type composition were implemented. Both packages contain appropriate data for deconvolution of adult blood samples used in EWAS data. The basic function of these packages estimateCellCounts, for 450K data, and the modified estimateCellCounts2, for EPIC data, work applying the Houseman regression calibration algorithm to the Illumina 450K and EPIC microarray for deconvoluting heterogeneous tissue sources like blood. Briefly, the function makes use of an object created from the DNA methylation study of blood and returns the relative proportions of leukocyte components in each sample. In the Illumina 450K, the type of immune cells returned are T-lymphocytes (CD4+ and CD8+), Natural Killer (NK) (CD56+) cells, B cells (CD19+), monocytes (CD14+), and granulocytes while in the Illumina EPIC BeadChip, the identified cells are T-lymphocytes (CD4+ and CD8+), Natural Killer (CD56+) cells, B cells (CD19+), monocytes (CD14+) and neutrophils.

3.3.1.3.6 Batch effect correction

A study involving a consistent number of samples assumes that samples need to be allocated into different chips or processed at different times. In high-throughput experiments, genomic data are often produced in batches because of logistical or practical restrictions. This inevitable logistics or technical variation can cause the so-called "batch effects" that are reported in the literature to cause significant heterogeneity across batches of data.

Batch effects are the technical artifacts that are not associated with the underlying biology, but rather with unrelated factors, such as laboratory conditions, experiment time, reagent lots, and/or laboratory personnel differences. When these experiment batches are confounded with the variable or the phenotype of interest, differences between biological groups may be artifacts. Batch effects often result in discrepancies and inconsistencies in the statistical distributions across data from different technical processing batches thus having an unfavorable impact on downstream analysis. This impact sometimes can be so profound that, without appropriate correction, it may lead to inaccurate conclusions or a significant reduction in the power for true biological signal detection. Microarray data provided a good example of composition of batch effects, showing that while the overall distribution of samples may be normalized to the same level across batches, individual genes may still be affected by batch-level bias. Infinium Illumina microarray Methylation 450K and EPIC BeadChip platforms accommodate measurement of different samples on a single "chip," organized into many columns and rows, creating a possible batch effect. One way to bypass the problem of batch effect in methylation data is to randomly distribute samples across the BeadChip avoiding the possibility that a single chip could contain only one phenotype of interest. In any case, since batch effects can affect different probes in different ways, they often cannot fully address by routine normalization methods, and special techniques to correct them are needed. Several batch effect correction methods have been developed for gene expression microarray [277, 278], albeit the appropriateness of these methods for the methylation data has not been evaluated.

Here the package *sva* and its function ComBat was the bioinformatic tool used to correct for batch effect caused by the positions of samples on the arrays. ComBat [279] is currently considered one of the most popular batch effect adjustment methods used when the batch covariate comes from known sources. The input M values were corrected and the new measurements, free of batch effects obtained as the output of the function were used in the following downstream analysis.

3.3.1.4 Downstream analysis

3.3.1.4.1 Differentially Methylated Positions/Probes (DMPs)

The main focus of most DNA methylation studies is the identification of differentially methylated probes, considered as all CpGs sites that are found to be significantly different between two groups or two phenotypes of interest, and that can be used as biomarkers or assist in the diagnostics or treatment of various diseases. According to Illumina's recommendation, the threshold used for statistical analysis in DMPs detection is *P*-value<0.05, and the absolute difference in the Delta Beta, considered as the difference of means of the Beta value of a single position between the group of interest, should be > 0.2. However, according to the type of study, these threshold values can be variable.

To identify significative DMPs the *limma* package [280] was used, assuming a linear model where the M values of each probe were used as quantitative dependent variables in all analyses. The application of *limma* package makes possible to include covariate in the definition of the design matrix that represents the model to be fit.

In the first part of the study, the linear model was corrected for gender, age, and cell composition. After running the linear model, with the M value, corrected for batch effects, an empirical Bayes approach was applied to moderate standard errors, and moderated t-statistics and associated *P*-values were obtained for each CpG site. In the second part of the study, a paired analysis was conducted, to detect differences between the two time-points. The model was adjusted considering three different covariates: responder, gender, and cell composition. Differentially methylated probes were filtered by significance threshold, and all significant DMPs with a resulting *P*-adjusted<0.01, and a Delta Beta value>0.10 were considered suitable results of our analysis. *P*-values were adjusted for multiple testing using the Benjamini–Hochberg method [281].

3.3.1.4.2 Differentially Methylated Regions (DMRs)

Differentially methylated regions are referred to as contiguous regions in the genome where sites or a group of sites close together have different methylation patterns between phenotypes. The presence of multiple adjacent CpGs can occur in different contexts, and according to this, DMRs can be subdivided into iDMR (imprinting-specific DMR), tDMR

(tissue-specific DMR), rDMR (reprogramming-specific DMR), cDMR (cancer-specific DMR), aDMR (aging-specific DMR).

Identifying DMRs among multiple tissues contributes to a comprehensive analysis of epigenetic differences among human tissues. DMRs may occur throughout the whole genome, but they show particularly importance when occurring around the promoter regions of genes, within the body of genes, and at the intergenic regulatory region. DMRs detected between cancer and normal samples demonstrate aberrant methylation in cancers [282].

Different methods have been developed to look at the DMRs. Directly, it is possible to look at DMRs according to the annotation of CpGs, searching for positions settled in the same Island regions or the close gene positions. On the other hand, other methods, based on statistical approaches have been developed to search for differentially methylated regions such as (bumphunter, Comb-p, DMRcate, DMRfinder and seqlm.). Here a data-driven approach called DMRcate [283] was used as it was considered one of the most common methods to detect DMRs in Whole Genome Bisulfite Sequencing and Illumina Infinium Array. This method does not consider all annotations except for chromosomal coordinates (spatial annotations). In fact, DMRcate intentionally avoids the distinction between probes that have an explicit gene and/or CpG-island annotation (as provided by Illumina) and those that do not. CpGs sites with a corrected *P*-value (FDR correction) are selected as significant CpGs. Those regions which contain at least 3 CpGs inside were considered DMRs. They were selected according to the statistical significance defined as Stouffer transformed Benjamini-Hochberg (BH) adjusted *P*-value<0.01.

3.3.1.4.3 Enrichment Pathways Analysis

Once the analysis of differentially methylated positions and regions between groups of samples has been completed, there will be a long list of significant CpG that need to be interpreted. The investigation of every single gene requires a large amount of manual literature searching to interpret. A standard approach to solve this issue and to gain an understanding of the epigenetic changes is pathway enrichment analysis, which helps to reduce the large gene list as a smaller list of easily interpretable pathways that may be enriched for differential methylation in the experiment. Different approaches have been developed to study enrichment pathways analysis in gene expression microarrays and RNA-sequencing. Usually,

gene set testing can be performed with two different approaches named over-representation analysis (ORA), which basically consists of the separation of genes differentially expressed from genes not-differentially expressed according to a pre-specified *P*-value threshold, and the functional class scoring (FCS) that instead of using only significant differentially expressed genes, utilizes the entire gene list to compute a gene set level *P*-value via permutation. In this way, FCS allows to detect small changes in gene expression and incorporates the dependence structure among the genes in a gene set, oppositely to ORA that can lead to loss of information when ignoring genes that are found marginally above the *P*-value threshold.

Unlike gene expression data, DNA methylation can occur anywhere on the genome and the number of CpGs profiled and gene length is not of one-to-one correspondence. In fact, EWAS data result in multiple CpG association *P*-values per gene since several differential CpGs can belong to the same gene, as well as one single CpG position can be found in more than one gene. Therefore, the biggest challenge in performing gene set testing in DNA methylation data is how to assign differentially methylated features to genes and how to adjust for the number of CpGs instead of gene length.

To our knowledge, there are very few gene set testing approaches arranged specifically for DNA methylation data. One of the most recent R package used here is called *methylGSA*, and it has been proposed for gene set enrichment analysis (GSEA) of methylation array data [284]. *MethylGSA* uses two different functions to perform gene set testing analysis: methylglm, similar to GOglm [285], and methylRRA. methylglm employs a logistic regression model to adjust for the number of CpGs; it works only for FCS so the entire list of genes is used in the model, as potentially the whole list of CpG could be of interest. methylRRa, instead, adjusts multiple *P*-values for each gene by Robust Rank Aggregation followed by either over-representation analysis, where *P*-values are corrected for multiple testing through false discovery rate (FDR), or functional class scoring in combination with GSEA, where gene list is ranked by z-scores, resulting from the conversion of *P*-values. MethylGSA package allows the specification of the array type used, 450K or EPIC; it supports pathways from "GO", "KEGG", and "Reactome and the most interesting feature, is the possibility to include in the specificity of the functions, the option to consider the different genomic context when selecting the gene list like promoter regions (CpGs within genes belonging to "TSS1500", "TSS200", "1stExon", or "5'UTR"), gene body or all together. Here the threshold adopted to select pathways significantly enriched is FDR *P*-value<0.05.

3.3.1.4.4 Principal Component analysis

Principal Component Analysis (PCA) is a statistical method used to simplify and reduce the complexity of sample spaces with many dimensions while preserving the majority of information and variability. Large datasets are increasingly common and widespread and are often difficult to interpret and analyze. PCA belongs to a family of techniques known as unsupervised learning whose main objective is extracting information from the predictors, for example, to identify subgroups. The PCA method, therefore, allows to "condense" the information provided by multiple variables into only a few components. PCA works projecting mathematically the original data set, which may have involved many variables, into the reduced PCA space, using combinations of the original variables. The accounted variability associated with each dimension decrease when moving from the first component to the last.

In the present study, PCA was considered to compare DNA methylation profiles from individuals of different phenotypes according to our study design.

3.3.1.4.5 ROC curves

The diagnostic efficiency of the candidate DMPs is evaluated as the agreement percentage using receiver operating characteristic (ROC) curve analyses.

The ROC curve is one of the most popular tools used to evaluate the diagnostic power of a selected biomarker. The ROC is basically used with three specific objectives: to determine the cut-off value with the highest sensitivity and specificity, to evaluate the ability of the diagnostic test to differentiate healthy versus sick individuals, and to compare the discriminative capacity of two or more diagnostic tests that express their results as continuous scales. ROC curves are created by plotting the rate of True positive and the rate of False Positive of a biomarker. They can be represented along the two axes, in order to study their relationships. The true positive rate is also known as sensitivity, which is the probability of detection. The false-positive rate instead is known as specificity, which is the probability of false alarms and is calculated as (1 - specificity). Since the abundance of information provided by the curve makes the comparison between biomarkers difficult, it is important to assess the discriminative capacity of the diagnostic test, and this is performed through the most important parameter useful in estimating efficiently the test, that is the area under the ROC curve (AUC, area under the curve), a unique and independent measure of the prevalence of the disease under study. The AUC reflects how

good the test is at discriminating patients with and without the disease across the entire range of possible cut-off points. The axes of the ROC curve graph adopt values between 0 and 1 (0% and 100%), delimiting a square of area = 1.00. Much the greater the AUC, the greater the discriminating power of the test. If a hypothetical test perfectly discriminated sick from healthy individuals, the area under the curve would have a value of 1, i.e. 100% accuracy. In case the test did not discriminate at all the sick from the healthy, the AUC would be 0.5 (or 50%). The classification proposed by Swets [286] was considered for the interpretation of the AUC values.

3.3.1.4.6 Pyrosequencing validation

Pyrosequencing is a sequencing-by-synthesis method based on the sequential addition and incorporation of nucleotides in a primer-directed polymerase extension. When the added nucleotide is complementary to the DNA template it could be incorporated by a DNA polymerase, during an event that is monitored in real time. The pyrosequencing technology uses the pyrophosphate molecules (PPi) released during the incorporation of unmodified nucleotides and that through an enzymatic cascade is quantitatively converted into a luminometric signal [287].

Thanks to its high reliability and flexibility, in the last decades, pyrosequencing technology has been widely used for various diagnostic applications, as well as for massively parallel sequencing. However, its key role in the quantitative nature of the results makes it useful for a variety of applications where the quantitative measure of the relative abundance of two individual nucleotides or short sequences is required. This is the case of DNA methylation where accurate quantitative analysis is very important, particularly in studies where the number of required CpGs within the primer binding sites might be insufficient for alternative methylation PCR approaches. In the last years, pyrosequencing has become a reference method for the validation and verification of newly developed methods for DNA methylation analysis, to check for demethylation or methylation levels of candidate biomarkers. In this case, the technology assesses quantitatively the methylation level of CpG sites in proximity after treating genomic DNA with bisulfite, as previously described. The protocol for DNA methylation pyrosequencing has been described in detail by Tost and Gut in 2007. Briefly, after treating the sample of interest with bisulfite, a target region of more than 350 bp is amplified by PCR using two primers complementary to the bisulfite-treated DNA sequence. One of two amplification primers shows a biotin label at its 5'-terminus, that is immobilized on streptavidin-coated beads

and then used to purify and provide a single-strand PCR. The pyrosequencing primer complementary to the single-stranded template is then hybridized to the template, and the pyrosequencing reaction is performed by adding single nucleotides in a predefined and sequential order. After this step, the template can be repurified and analyzed with a new sequencing primer for the serial pyrosequencing, a process that allows the sequencing of an entire PCR amplification product by successive annealing of different sequencing primers and washing away of the de novo synthesized template [288].

Pyrosequencing here was adopted to assess the methylation profile of some selected biomarkers in a validation cohort. The design of primers was performed using Qiagen's PyroMark Assay Design 2.0 software to hybridize to CpG- free sites to ensure methylation-independent amplification. After the extraction of genomic DNA, Pyrosequencing reactions and methylation quantification were performed in PyroMark Q96 System version 2.0.6 (Qiagen) using appropriate reagents and recommended protocols.

4
RESULTS

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

4.1 Assessing the role of host epigenetic changes after RSV infection in Respiratory Morbidity defined as wheezing and/or asthma

4.1.1 Preliminary analysis

The first part of the study focuses on interrogating the genome-wide DNA methylation profiles of 77 children infected by RSV to understand how epigenetics influences the different respiratory sequelae developed after the RSV infection.

During the preprocessing stages of data analysis, three samples belonging to the control group were removed for sex inconsistency, while one of the wheezing group was eliminated because of an extreme outlier; in addition, four samples of the wheezing group and one of the control group were discarded since they did not pass the initial quality control (Figure 12).

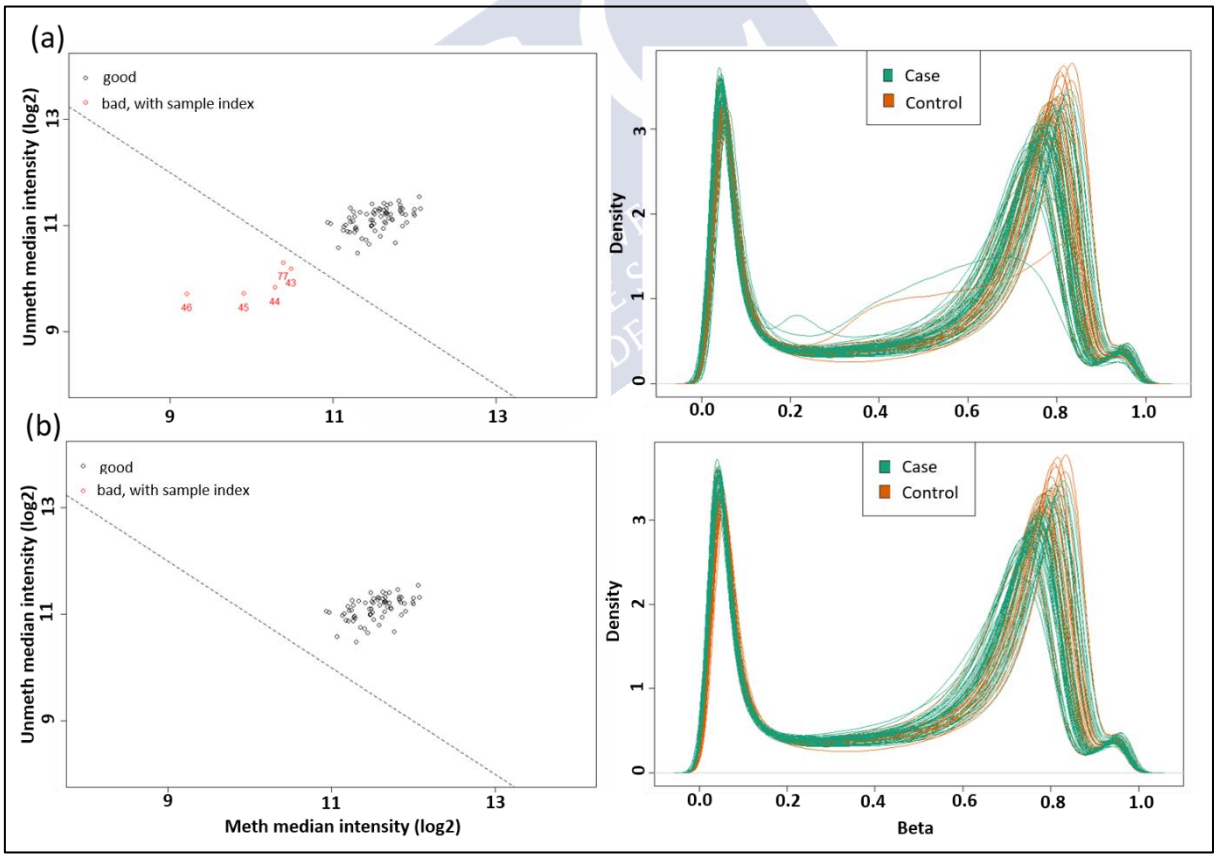


Figure 12: Qc plot and Density plot before and after correction. In the density plot, all samples exhibit two peaks: one corresponding to low or unmethylated probes with a Beta value close to 0 and the second one corresponding to highly or fully methylated probes with a Beta value close to 1, confirming the bimodal shape of the distribution of Beta values.

Of the 77 samples initially typed, 68 were left for the downstream analysis. After removing poorly performing probes, such as those that overlapped with SNPs or hybridized to multiple locations in the genome and those located on the X and Y chromosomes, a total of 811,035 probes were preserved for downstream analysis.

4.1.2 DMPs between controls (RSV+/normal recovery) and cases (RSV+ / recurrent wheezing and/or asthma)

The preparation of data for downstream analysis implicated the estimation of cell composition for both case and control group. After running the deconvolution function it was possible to observe that there was a significant difference (Student's T-Test) between natural killer cells in case and controls, as shown in figure 13.

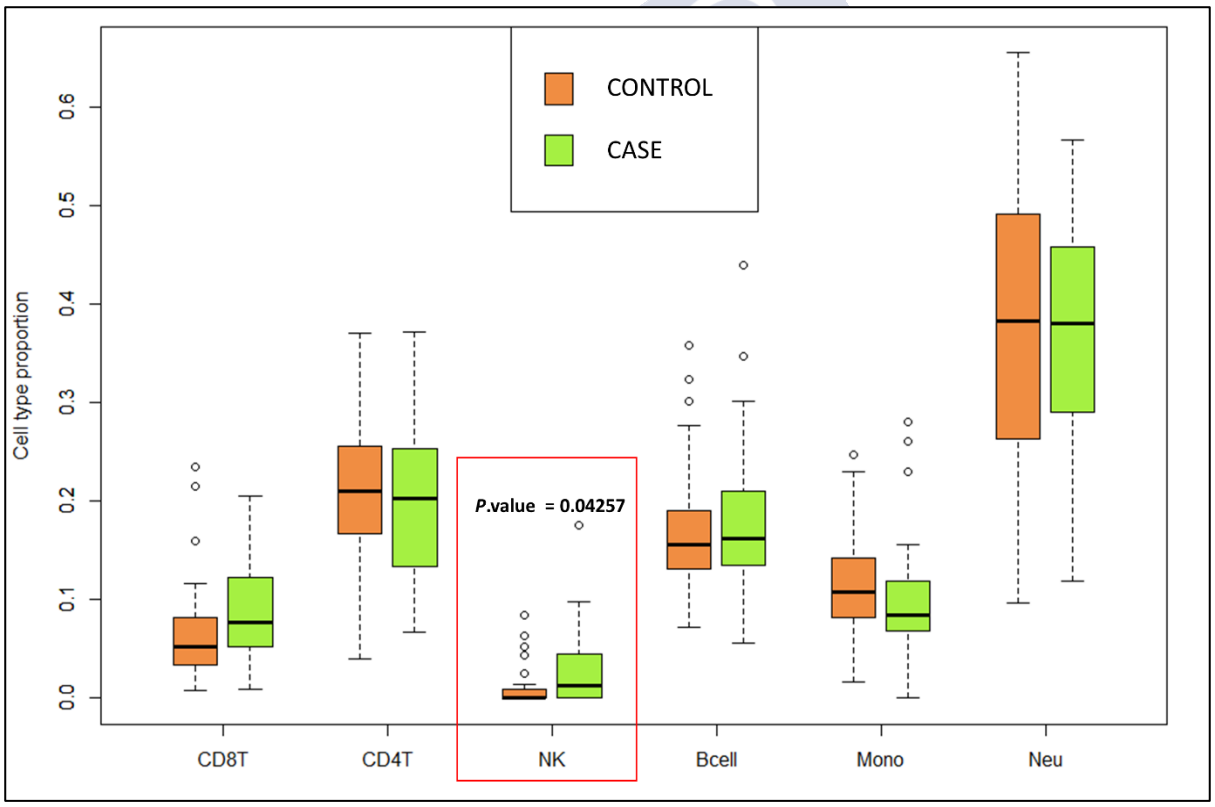


Figure 13: Boxplot showing the proportion of leukocyte cell type in case and control groups.

After adjusting the linear model for cell-type composition, age, and gender, 5,097 significantly DMPs were obtained between control and case groups (FDR *P*-adjusted < 0.01), which corresponded to 3,278 unique genes according to the Illumina Human Methylation EPIC manifest annotation file [262] (Table 5). Among these, 1,155 DMPs were hypomethylated

(22.7%) and 3,942 hypermethylated (77.3 %) in the control group. PCA of these significant DMPs showed a clear separation of samples along with the first component (accounting for 42.6% of the variance), and which moderately split case from control groups, suggesting that the dominant difference in DNA methylation across all samples was between RSV infected children with normal recovery and RSV infected children with subsequent development of recurrent wheezing and/or asthma (Figure 14).

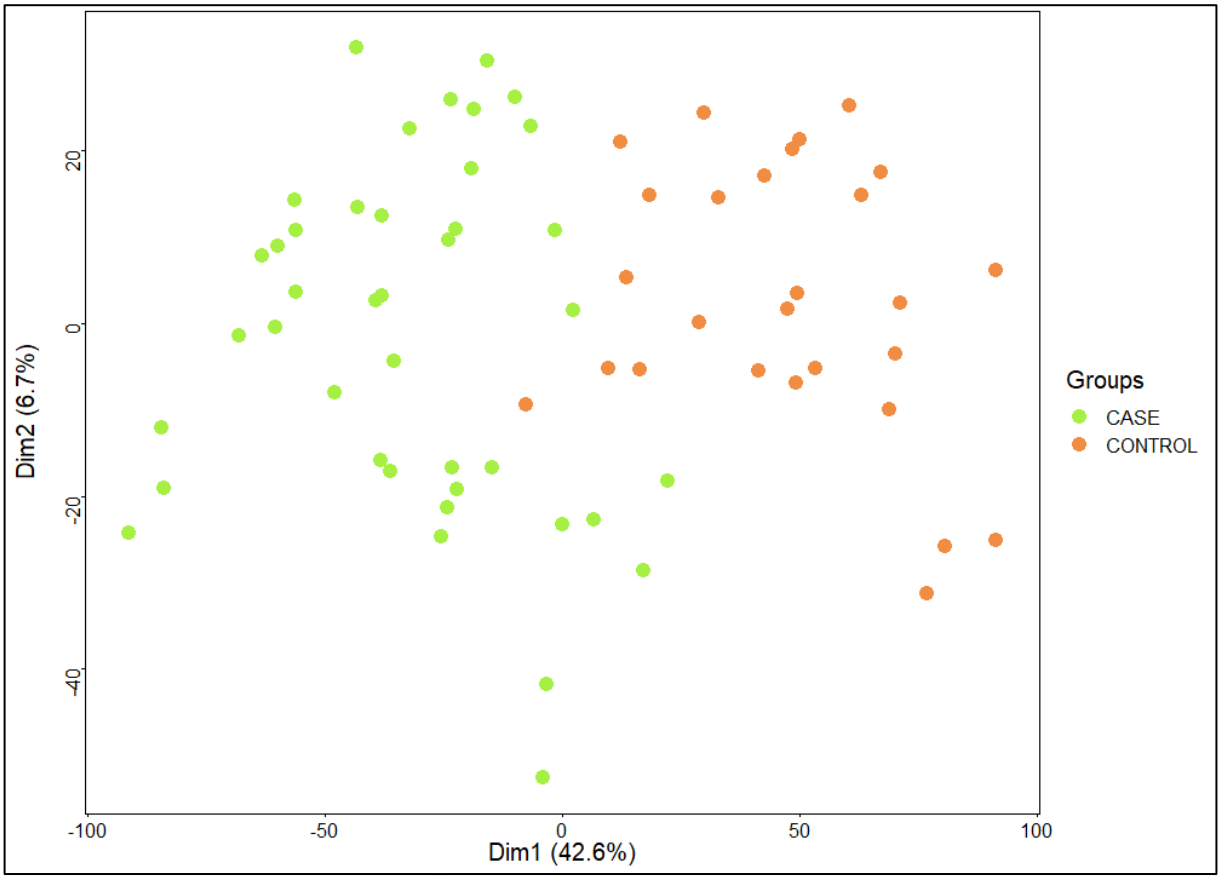


Figure 14: Principal Component Analysis of the significant DMPs (FDR *P*-value < 0.01) showed a clear separation between case and control groups.

To deeply study the significant DMPs, the distribution of the DMPs is observed in the genomic context and the CpG island context. There are significant differences between the distribution of hypomethylated and hypermethylated CpGs. Island's regions are overrepresented by hypomethylated sites (40%), while the OpenSea regions are characterized more by hypermethylated DMPs (61%); the distribution of shelf and shore is similar between the two groups (Figure 15a). Concerning the genomic distribution, hypermethylated positions are more concentrated in the gene body (59%), while the hypomethylated positions are distributed almost equally among the gene body (27%), TSS1500 (19%), TSS200 (21%), 'UTR (20%) and 1st

Exon (12%) (Figure 15b). Moreover, all chromosomes are more represented by hypermethylated DMPs than the hypomethylated ones (Figure 15c).

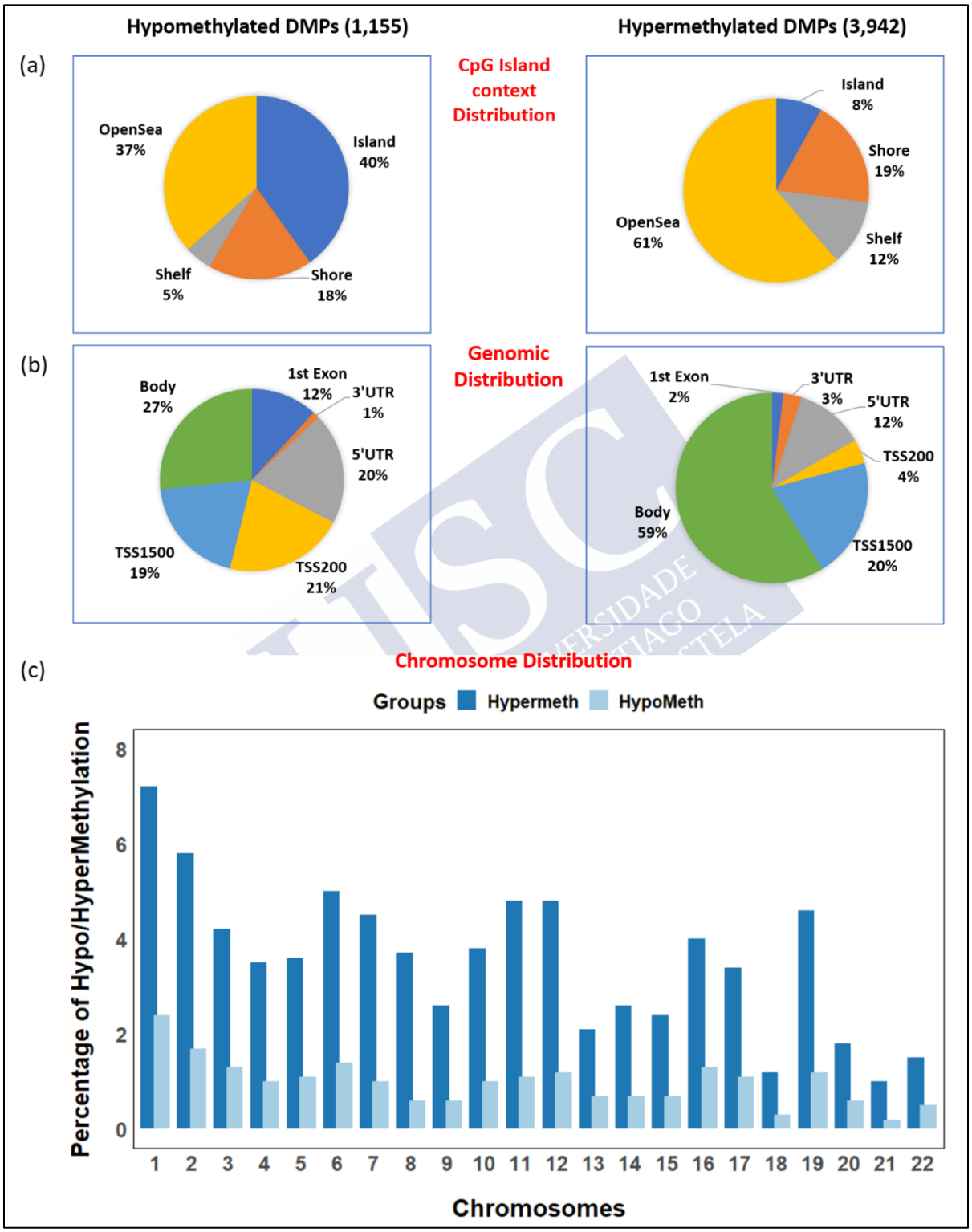


Figure 15: (a) Pie chart showing the percentage of the significant DMPs according to their distribution in the Island context and (b) in the genomic context. (c) BarPlot showing the distribution of hypomethylated and hypermethylated DMPs in the chromosomes.

Among the most significative DMPs, two interesting CpGs settled the start transcription sites (TSS1500) of the gene they belonged to were identified: the *EYA3* (FDR *P*-value = 2.77e-10), and *DDX27* (FDR *P*-value = 4.45e-10), that are shore CpGs and display a hypomethylation pattern in the case group when compared to the control one (Figure 16a). ROC curves were used to evaluate the ability of the methylation levels of these selected DMPs to discriminate cases from controls (Figure 16b).

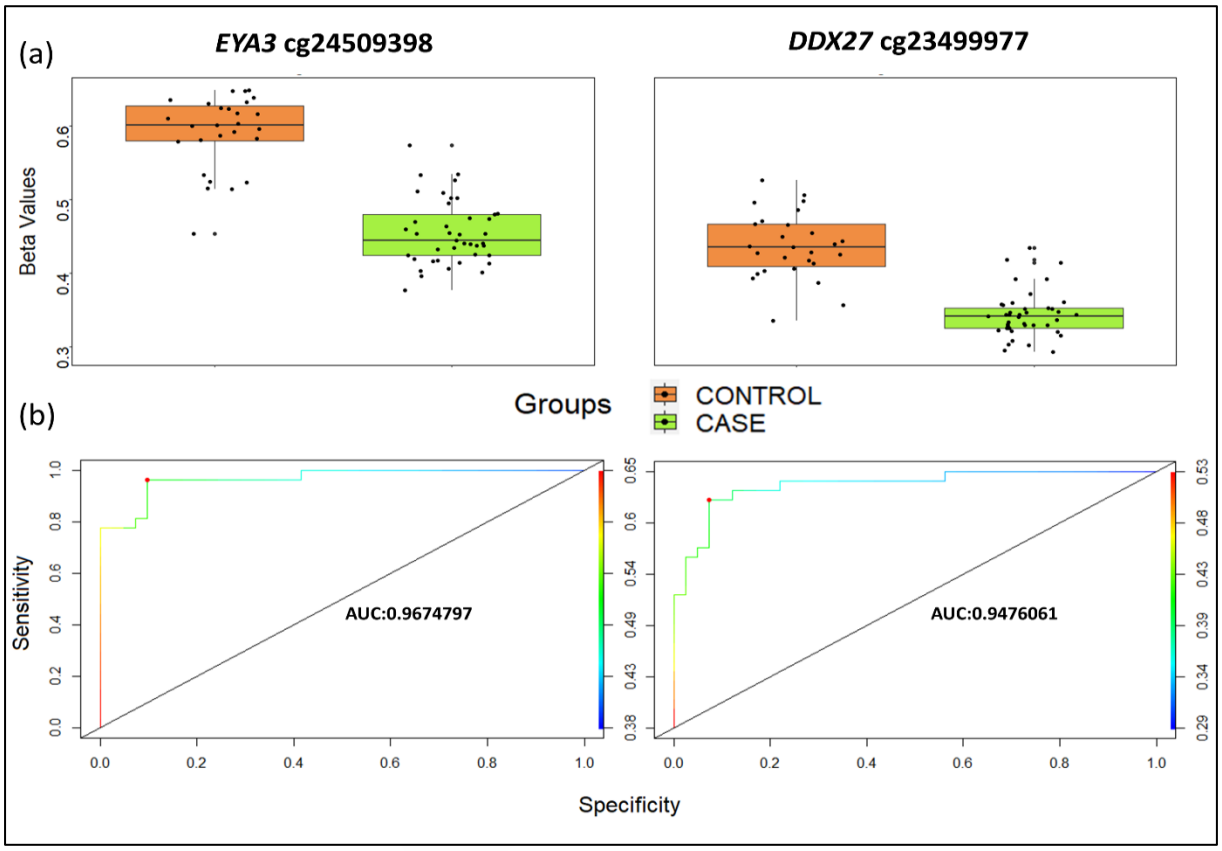


Figure 16: (a) Boxplot of the most significant differentially methylated positions: cg24509398 within EYA3, and cg23499977 within DDX27. (b) ROC curves indicating that the accuracy of the test based on the reported CpGs is very high (AUC > 94%) when comparing control with case groups.

To narrow down the analysis and to reduce the large number of significant methylated positions observed, the following steps were focused on finding the most consistent methylation changes by selecting among the significant CpGs only those with an absolute Delta Beta between control and case higher than 0.10. Doing this, it was possible to find 29 CpGs (Table 6), associated with 22 unique genes. Of these positions, only five were moderately hypomethylated in the control group versus the case group, the remaining 23 DMPs showed an opposite pattern with a higher methylation level for controls than for cases. One of the positions showing the highest difference in methylation among the two groups (Delta Beta = 0.136) is the cg24509398 within *EYA3*, found to be one of the most significant.

Table 6: Differentially methylated positions with *P*-value<0.01 and an absolute Delta Beta >0.10

CpG_ID	Chr	Position	Relation to Island	Gene Name	Gene Group	FDR *P*-Value	Delta Beta
cg06259441	chr3	23957589	N_Shore	*RPL15; RPL15; RPL15; RPL15; RPL15; NKIRAS1*	TSS1500; TSS1500; TSS1500; TSS1500; TSS1500; 5'UTR	1.48e-03	-0.115
cg20967739	chr1	50895827	S_Shelf			9.25e-04	-0.114
cg09728337	chr10	32668256	OpenSea	*EPC1*	TSS1500	5.79e-03	-0.109
cg25338438	chr5	161276341	OpenSea	*GABRA1; GABRA1; GABRA1; GABRA1; GABRA1*	5'UTR; 5'UTR; 5'UTR; 5'UTR; TSS1500	1.04e-03	-0.107
cg27552418	chr4	97598786	OpenSea			1.08e-03	-0.104
cg06583549	chr19	46387962	Island	*IRF2BP1*	1stExon	4.31e-04	0.101
cg04479860	chr4	190767364	Island			7.15e-03	0.101
cg16003687	chr6	168613889	S_Shore			3.27e-05	0.101
cg05992347	chr16	33964783	Island	*MIR1826*	TSS1500	4.53e-04	0.101
cg23621438	chr22	25850271	OpenSea	*MIR6817; CRYBB2P1; CRYBB2P1*	TSS1500; Body; Body	1.84e-09	0.102
cg14819618	chr8	8180214	S_Shelf	*PRAGMIN*	Body	2.26e-06	0.104
cg02737727	chr14	105561470	S_Shore	*LOC102723354*	Body	1.84e-09	0.105
cg02077481	chr16	33939020	Island			9.61e-03	0.105
cg13165070	chr11	2154113	Island	*INS-IGF2; IGF2; IGF2; IGF2*	Body; 3'UTR; 3'UTR; 3'UTR	1.87e-03	0.107
cg26900509	chr11	127514	Island	*LOC100133161*	Body	3.13e-04	0.108
cg19841649	chr5	4866322	Island			7.58e-03	0.108
cg15822108	chr4	140306311	OpenSea	*NAA15*	Body	4.28e-07	0.109
cg08938155	chr5	77043612	OpenSea	*TBCA*	Body	3.68e-03	0.109
cg05116443	chr20	62562680	Island	*DNAJC5*	Body	1.25e-04	0.109
cg19977004	chr11	1482563	Island	*BRSK2*	3'UTR	4.45e-04	0.111
cg06579481	chr7	104621597	N_Shelf			8.57e-06	0.112
cg12216772	chr10	46164062	N_Shelf	*ANUBL1; ANUBL1*	5'UTR;5'UTR	3.33e-08	0.115
cg05800416	chr8	19460097	Island	*CSGALNACT1; CSGALNACT1; CSGALNACT1*	TSS200; Body; 5'UTR	2.99e-04	0.116
cg15619333	chr12	133759653	S_Shore	*ZNF268; ZNF268; ZNF268; ZNF268; ZNF268; ZNF268; ZNF268; ZNF268; ZNF268; ZNF268*	5'UTR; Body; Body; Body; Body; Body; Body; Body; Body; Body	4.39e-09	0.116
cg00044440	chr5	138671449	OpenSea			6.77e-06	0.118
cg11702503	chr19	6215254	Island	*MLLT1*	Body	2.43e-03	0.118
cg24509398	chr1	28416532	S_Shore	*EYA3*	TSS1500	2.77e-10	0.136
cg21226224	chr8	55370171	Island	*SOX17*	TSS1500	3.29e-04	0.144
***Chr = Chromosome**							
***N_Shore/N_Shel = North Shore/Shelf**							
***S_Shore/S_Shelf = South Shore/Shelf**							

4.1.3 DMRs between recurrent case and control groups

100 significant DMRs (adjusted *P*-value<0.01) with at least 3 CpGs inside, were recognized when comparing patients with respiratory sequelae (case) against patients with normal recovery (control). Of the selected DMRs, 11 were hypomethylated in the control group, while the remaining 89 present a lower level of methylation in the case group (Table 7). The top DMRs showing an hypomethylation pattern for control group were the following genes: *VGLL4* represented by nine CpGs (*P*-value = 3.83e-03), *LINC00051* (*P*-value = 5.87e-03) represented by six positions, and *SLFN12L* (*P*-value = 2.30e-06) defined by 5 CpGs inside; among DMRs displaying an hypermethylation pattern for controls and a hypomethylation pattern for cases, there were *FDFT1* characterized by 10 CpGs (*P*-value = 6.34e-05), and *RNA5-8SP2* (*P*-value = 2.02e-03), *GUCY1B2* (*P*-value = 3.53e-03), *B9D1*(*P*-value = 5.21e-04), *LRRC47* (*P*-value = 8.31e-043), and *GGACT* (*P*-value = 4.36e-03) represented all by 8 CpGs sites.

Table 7: Differentially methylated regions (DMRs) associated with respiratory sequelae after RSV infection

Chr	Start	End	Width	Nº CpG	min smoot fdr	Stouffer	meandiff	Overlapping Genes
chr17	79022798	79023201	404	3	1.69e-05	5.87e-03	-0.044	*BAIAP2*
chr8	143279443	143279882	440	6	1.79e-11	5.87e-03	-0.021	*LINC00051*
chr17	33821622	33823690	2069	5	2.30e-06	6.59e-04	-0.019	*SLFN12L*
chr5	54472257	54472570	314	3	8.92e-06	5.98e-03	-0.009	NA
chr12	112451600	112452139	540	3	1.92e-10	5.97e-03	-0.009	*ERP29*
chr7	74072514	74073064	551	3	4.01e-08	1.73e-03	-0.009	*GTF2I*
chr6	130338360	130338938	579	3	3.89e-06	7.17e-03	-0.008	*L3MBTL3*
chr11	31828528	31829582	1055	4	1.79e-09	9.01e-03	-0.007	*PAX6*
chr6	64308512	64308590	79	3	1.69e-07	2.14e-03	-0.006	NA
chr20	33562131	33563009	879	4	1.41e-10	5.11e-03	-0.006	NA
chr3	11746183	11749113	2931	9	5.73e-06	3.83e-03	-0.005	*VGLL4*
chr3	136440459	136441137	679	4	1.03e-06	7.28e-03	0.001	*STAG1*
chr17	19247837	19248565	729	8	1.98e-11	5.21e-04	0.002	*B9D1, MIR1180*
chr3	52721831	52722452	622	3	7.93e-13	6.12e-03	0.003	*GNL3*
chr2	59400950	59401418	469	3	8.80e-06	9.50e-03	0.005	NA
chr16	221584	221811	228	3	3.69e-08	2.99e-03	0.008	NA
chr17	79870762	79871841	1080	3	8.72e-07	2.14e-03	0.009	*SIRT7*
chr12	56394243	56395439	1197	5	6.60e-10	8.17e-03	0.010	*SUOX*
chr10	134186822	134186967	146	3	9.05e-06	7.36e-03	0.011	*LRRC27*
chr6	46742620	46743572	953	3	3.59e-07	1.52e-03	0.011	NA
chr2	28680017	28680345	329	4	3.34e-07	5.73e-03	0.011	*PLB1*
chr6	86339171	86339817	647	3	1.07e-06	8.99e-03	0.012	*SYNCRIP*
chr2	63855048	63855118	71	3	1.72e-06	5.72e-03	0.014	*WDPCP*

Chr	Start	End	Width	Nº CpG	min smoot fdr	Stouffer	meandiff	Overlapping Genes
chr17	73517413	73518815	1403	6	2.11e-09	6.94e-03	0.014	*TSEN54*
chr7	95917902	95917945	44	3	4.94e-10	1.92e-04	0.015	*SLC25A13*
chr2	196411276	196412552	1277	5	4.82e-06	2.30e-03	0.015	NA
chr6	99358635	99358835	201	3	6.13e-08	1.99e-03	0.015	*FBXL4*
chr12	131159419	131159658	240	4	5.62e-06	5.86e-03	0.015	*RIMBP2, RP11-662M24.2*
chr4	3446682	3448150	1469	6	6.88e-11	2.56e-03	0.015	*HGFAC*
chr17	80138794	80139140	347	3	7.94e-08	5.50e-03	0.015	*CCDC57*
chr19	36102171	36103576	1406	10	1.33e-26	1.46e-03	0.015	*AC002115.9*
chr8	11652174	11653811	1638	9	2.99e-15	6.34e-05	0.016	*FDFT1*
chr10	130118500	130119293	794	3	1.03e-07	4.97e-03	0.016	NA
chr1	3703550	3705262	1713	8	3.35e-09	8.31e-03	0.017	*LRRC47*
chr11	41644411	41644529	119	3	6.38e-06	7.41e-03	0.017	*RP11-124G5.3*
chr12	295951	297831	1881	6	9.48e-10	9.88e-03	0.017	NA
chr6	111368770	111369778	1009	4	3.93e-07	5.84e-03	0.017	NA
chr7	44046988	44047593	606	3	1.61e-14	9.45e-03	0.018	*SPDYE1, POLR2J4, RP5-1165K10.2, AC004951.6*
chr2	240099221	240100522	1302	3	5.11e-11	3.37e-03	0.018	*HDAC4*
chr10	118153200	118154544	1345	4	1.05e-06	3.34e-03	0.018	NA
chr5	166210883	166211825	943	3	4.77e-07	5.87e-03	0.019	NA
chr6	12017547	12018635	1089	3	2.18e-05	4.40e-03	0.019	*HIVEP1*
chr15	63248230	63249401	1172	4	1.54e-11	7.66e-03	0.019	NA
chr7	5833275	5833444	170	3	2.45e-06	3.52e-03	0.019	NA
chr6	31545836	31547704	1869	7	1.51e-19	5.50e-03	0.019	*TNF*
chr13	101185600	101186537	938	8	5.58e-11	4.36e-03	0.019	*GGACT*
chr15	81624074	81624742	669	3	1.13e-08	6.75e-03	0.020	*RP11-761I4.3, TMC3*
chr11	116980464	116981386	923	4	6.44e-06	9.21e-03	0.020	NA
chr12	121812341	121813496	1156	3	2.50e-10	3.01e-03	0.020	*ANAPC5*
chr19	14019238	14019794	557	3	5.12e-10	9.19e-03	0.022	*CC2D1A*
chr12	100537139	100538115	977	4	1.50e-16	3.22e-03	0.022	*RP11-175P13.2*
chr1	40995880	40996360	481	3	1.37e-11	4.40e-04	0.023	NA
chr12	6651909	6652604	696	4	2.83e-20	2.49e-05	0.023	*IFFO1*
chr9	6046174	6046681	508	3	3.66e-07	9.44e-04	0.023	NA
chr9	139925395	139926494	1100	4	1.89e-05	9.86e-03	0.024	*C9orf139, FUT7*
chr1	165178556	165179845	1290	3	2.50e-10	3.70e-03	0.024	*LMX1A*
chr10	123748476	123748922	447	5	2.04e-12	1.48e-03	0.024	*TACC2*
chr1	167090618	167091161	544	7	1.76e-08	5.13e-03	0.024	*DUSP27*
chr13	51639953	51641069	1117	8	9.24e-09	3.53e-03	0.025	*GUCY1B2*
chr1	156162316	156163116	801	3	1.03e-07	4.92e-03	0.025	NA
chr9	79056558	79057230	673	3	2.93e-08	3.84e-03	0.025	*GCNT1*
chr1	1625395	1625702	308	4	1.11e-12	4.80e-03	0.025	NA
chr5	166353057	166354283	1227	5	5.60e-16	1.34e-03	0.025	*CTB-7E3.1*
chr6	71955181	71955317	137	3	1.03e-09	3.80e-04	0.025	*RP11-154D6.1*
chr3	126292582	126292738	157	3	1.39e-06	7.40e-03	0.025	*TXNRD3, TXNRD3NB*

Chr	Start	End	Width	Nº CpG	min smoot fdr	Stouffer	meandiff	Overlapping Genes
chr7	111594719	111594883	165	3	2.07e-06	1.75e-03	0.025	*DOCK4*
chr8	11946498	11947722	1225	3	5.08e-08	1.40e-03	0.025	NA
chr12	125474833	125475014	182	3	8.75e-07	9.29e-03	0.026	NA
chr5	141021866	141022405	540	3	3.00e-12	6.77e-04	0.026	*FCHSD1*
chr7	117287511	117287605	95	3	2.96e-07	8.72e-03	0.026	*CFTR*
chr3	150425182	150425714	533	3	2.54e-07	2.22e-03	0.026	*RP11-103G8.2*
chr13	91147296	91148429	1134	3	5.53e-07	8.18e-04	0.026	*LINC01049*
chr9	34127905	34128998	1094	3	3.63e-13	5.20e-03	0.027	NA
chr3	138326216	138326642	427	4	2.26e-13	1.08e-04	0.027	NA
chr11	67134607	67135897	1291	6	4.54e-10	1.47e-04	0.027	*AP003419.11, CLCF1*
chr2	198235557	198236295	739	3	1.78e-13	3.22e-05	0.027	NA
chr2	181966002	181966624	623	5	2.28e-11	5.16e-03	0.027	*AC068196.1*
chr18	43193929	43195041	1113	3	3.03e-10	9.24e-04	0.027	*SLC14A2*
chr15	34178460	34179329	870	3	3.35e-12	1.90e-04	0.029	*AVEN*
chr1	7705567	7706193	627	3	1.04e-06	5.22e-03	0.029	*CAMTA1*
chr6	18261541	18262523	983	3	1.46e-07	1.70e-04	0.029	*DEK*
chr1	3239139	3240386	1248	5	4.28e-07	1.49e-03	0.029	*PRDM16*
chr6	31373660	31374206	547	5	2.08e-14	6.87e-06	0.030	*HCP5, MICA*
chr1	205210699	205210953	255	3	6.36e-07	6.72e-03	0.030	*TMCC2*
chr17	41564457	41564846	390	3	3.38e-11	2.20e-04	0.031	*DHX8*
chr7	1364975	1365082	108	3	4.94e-06	8.86e-03	0.031	NA
chr20	36576591	36577384	794	3	9.77e-10	1.78e-03	0.032	NA
chr12	21167330	21167766	437	3	1.39e-07	4.14e-04	0.034	*SLCO1B3, LST3, SLCO1B7*
chr14	21397486	21398180	695	3	3.15e-22	6.56e-06	0.034	*RP11-219E7.4*
chr8	141903388	141903635	248	3	5.51e-10	3.82e-05	0.035	*PTK2*
chr13	57720225	57720245	21	3	8.49e-08	1.12e-03	0.035	NA
chr1	213176123	213176991	869	3	1.37e-06	2.94e-03	0.036	*ANGEL2*
chr19	6215016	6215556	541	4	3.99e-08	2.37e-03	0.037	*MLLT1*
chr4	49525597	49526636	1040	5	8.00e-07	8.15e-03	0.039	*RP11-241F15.5, RP11-1281K21.9*
chr19	940724	942270	1547	7	3.18e-14	6.39e-05	0.042	*ARID3A*
chr10	126850823	126852530	1708	6	8.21e-06	8.76e-03	0.043	NA
chr19	4040566	4040622	57	3	4.60e-05	8.29e-03	0.044	NA
chr16	33964200	33965879	1680	8	1.34e-12	2.02e-03	0.049	*RNA5-8SP2*
chr5	261389	261919	531	3	9.78e-09	3.20e-03	0.058	NA
chr19	53228974	53229766	793	3	1.26e-13	9.39e-05	0.058	*ZNF611*
***Chr: chromosome**								
***Width: base pair (bp) length of the DMR**								
***min_smoot_fdr: Minimum FDR of the smoothed estimate.**								
***Stouffer: Stouffer summary transform of the individual CpG FDRs.**								
***meandiff: Mean differential/coefficient across the DMR.**								

4.1.4 Identification of DMPs between control, wheezing, and asthma groups

The next step consists of investigating asthma and wheezing groups separately against controls. The same linear model and the same threshold for the significant sites detected in the case-control analysis (see above) were applied.

When the wheezing group was faced toward controls, 3,432 significant DMPs, of which 2,663 were hypermethylated, and 769 were hypomethylated in the control group, were detected. Likewise, considering asthma patients versus controls, it was possible to identify 1,358 DMPs with the vast majority being hypermethylated (1,059), against the 299 displaying a hypomethylated pattern in the control group. A total of 512 significant DMPs shared by the two contrasts corresponding to the most significant DMPs found when comparing cases versus controls (Figure 17). When plotting the principal component analysis of the three separate groups, a similar methylation profile for asthma and wheezing groups could be noticed (Figure 18).

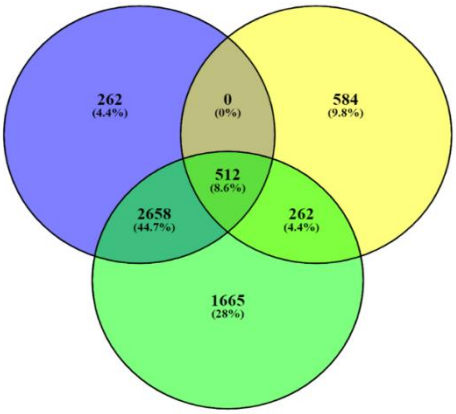


Figure 17: Venn Diagram with CpGs shared by three different contrast: Cases VS Controls (green), Wheezing VS Controls (blue) and Asthma VS Controls (yellow)

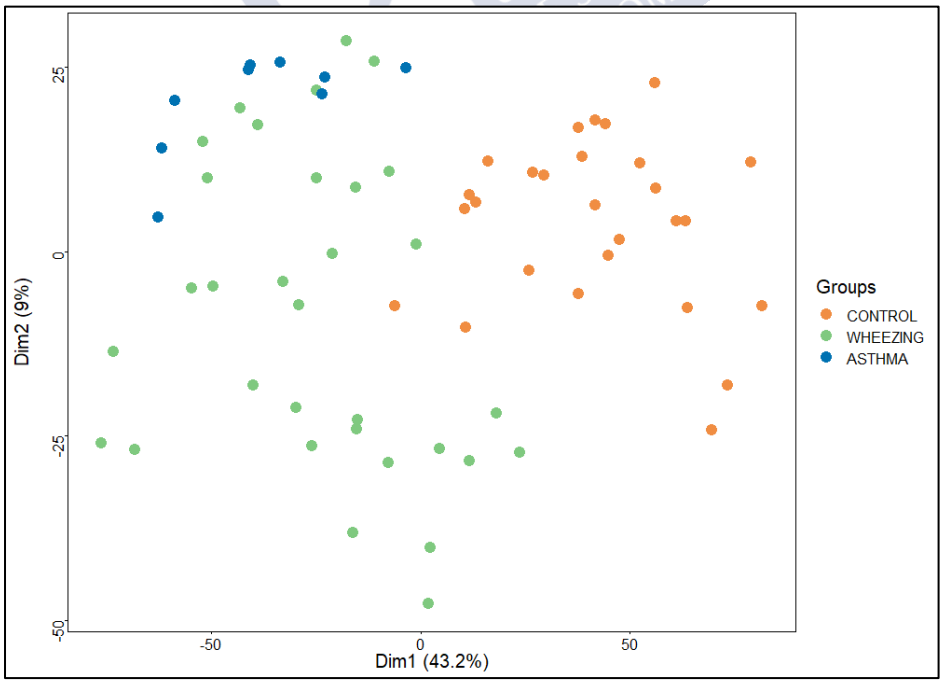


Figure 18: Principal Component Analysis of the significative DMPs (adjusted *P*-value < 0.01) showed the distribution of the three groups (recurrent wheezing RSV cases, asthma RSV cases, and not-wheezing/asthma RSV cases, here onwards referred as the control group)

However, by running the linear model with the contrast asthma-wheezing and using a less stringent threshold (FDR *P*-value <0.05), 47 CpG significantly differentially methylated associated with 41 genes were selected, which clearly separate the two groups in a PCA (Figure 19).

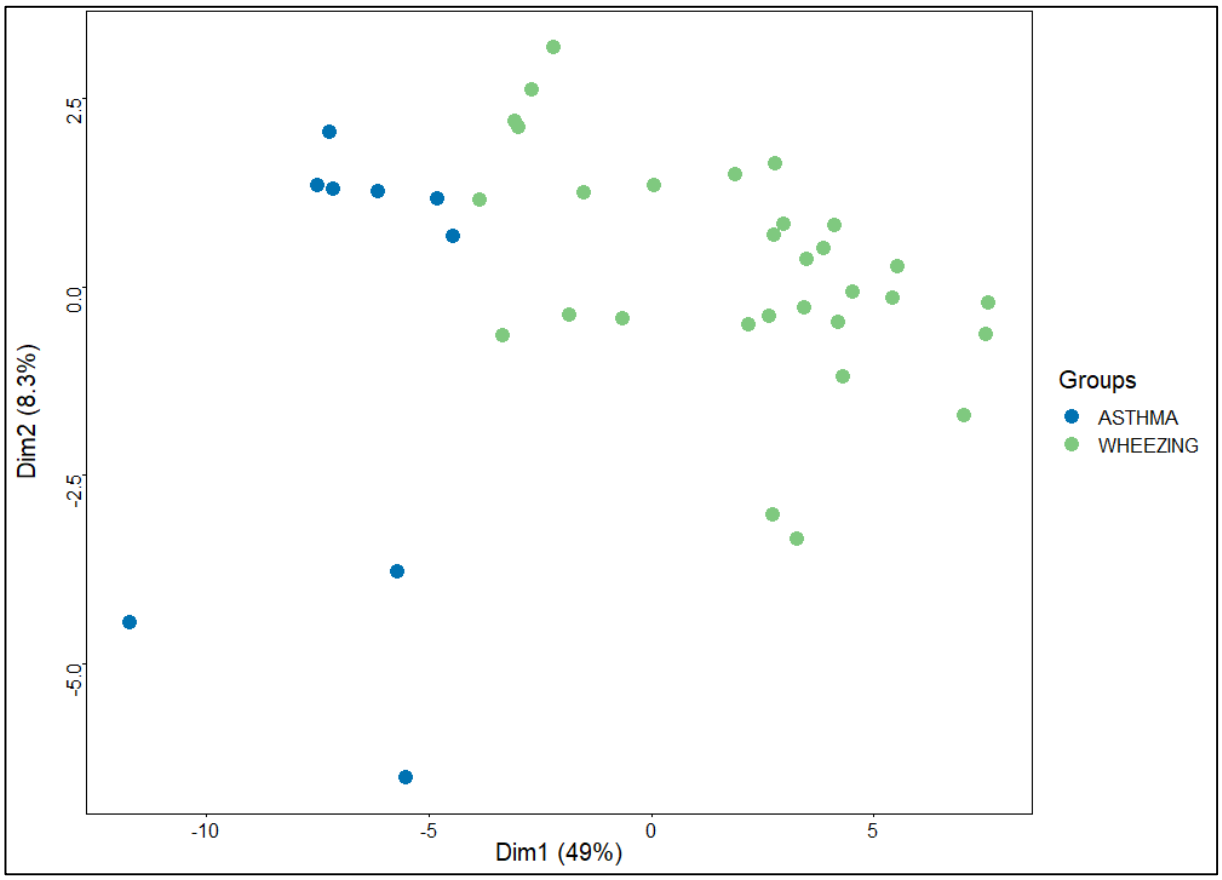


Figure 19: Principal Component Analysis of the most different DMPs (adjusted P-value < 0.05) shows an almost clear separation between wheezing and asthma groups.

Of the significant CpGs, 30 positions were found to be hypermethylated in wheezing group when compared to asthma, while the remaining 17 exhibits the opposite pattern (Table 8). One of the most significant and the most different positions (FDR *P*-value = 0.018 and Delta Beta >0.14) is the cg18873878 within the TP73 gene.

Table 8: Differentially methylated positions (DMPs) between asthma and wheezing groups ordered by FDR *P*-value

CpG_ID	Chr	Position	Relation to Island	Gene Name	Gene Group	FDR *P*-Value	Delta Beta
cg19332572	chr11	65321591	Island	*LTBP3; LTBP3; LTBP3*	Body; Body; Body	0.011	-0.117
cg14065121	chr9	77643271	Island	*C9orf41; C9orf41*	1stExon; 5'UTR	0.011	-0.109
cg05838113	chr10	135082349	Island	*ADAM8; ADAM8; ADAM8*	Body; Body; Body	0.011	-0.087
cg09519644	chr12	85401946	OpenSea			0.011	-0.075
cg11946459	chr6	29911558	S_Shore	*HLA-A*	Body	0.011	-0.075
cg23861120	chr12	67835705	OpenSea			0.014	-0.048
cg18873878	chr1	3607116	Island	*TP73; TP73; TP73; TP73*	Body; TSS200; TSS200; TSS200	0.018	-0.143
cg00664920	chr16	2664747	Island	*LOC652276*	Body	0.021	-0.069
cg08806408	chr16	51185001	Island	*SALL1; SALL1*	TSS1500; Body	0.021	0.016
cg01077616	chr6	42017945	N_Shore	*CCND3; TAF8; CCND3*	TSS1500; TSS1500; TSS1500	0.022	-0.056
cg05472874	chr22	44258179	Island	*SULT4A1*	1stExon	0.030	-0.077
cg27594116	chr9	100069897	Island	*CCDC180; LOC100499484-C9ORF174; LOC100499484-C9ORF174; LOC100499484-C9ORF174*	TSS200; Body; Body; Body	0.031	-0.108
cg01181415	chr12	16757954	OpenSea	*LMO3; LMO3*	5'UTR; 5'UTR	0.034	0.046
cg02478023	chr19	57351322	Island	*MIMT1; PEG3; PEG3; ZIM2; PEG3; PEG3; ZIM2; ZIM2*	TSS1500; 5'UTR; 5'UTR; 5'UTR; 5'UTR; 5'UTR; 5'UTR	0.036	-0.084
cg01997696	chr20	43374401	Island	*KCNK15*	TSS200	0.036	-0.067
cg09672082	chr5	271577	Island	*PDCD6*	TSS200	0.036	0.018
cg17748896	chr20	58940540	OpenSea			0.036	0.051
cg16740746	chr16	78134345	S_Shore	*WWOX; WWOX*	Body; Body	0.037	-0.061
cg11380640	chr18	57549413	OpenSea			0.037	0.046
cg25612428	chr5	10649867	Island	*ANKRD33B*	Body	0.040	-0.083
cg07367519	chr22	40075288	Island	*CACNA1I; CACNA1I*	Body; Body	0.041	-0.117
cg11507793	chr6	29856363	Island	*HLA-H*	Body	0.041	-0.113
cg10274606	chr14	73118334	OpenSea			0.041	-0.109
cg22371961	chr1	169132356	OpenSea	*NME7; NME7; NME7*	Body; Body; Body	0.041	-0.092
cg01848660	chr2	68269960	OpenSea	*C1D; C1D*	3'UTR; 3'UTR	0.041	-0.074
cg01979489	chr16	332603	Island	*ARHGDIG; PDIA2*	Body; TSS1500	0.041	-0.074
cg09046688	chr9	75621983	OpenSea			0.041	-0.067
cg14377711	chr16	1384369	Island	*BAIAP3; BAIAP3; BAIAP3; BAIAP3; BAIAP3*	5'UTR; 5'UTR; 5'UTR; 5'UTR; TSS200	0.041	-0.060
cg23215256	chr7	631862	Island	*PRKAR1B; PRKAR1B;*	Body; Body; Body; Body;	0.041	-0.060

CpG_ID	Chr	Position	Relation to Island	Gene Name	Gene Group	FDR *P*-Value	Delta Beta
				PRKAR1B; PRKAR1B; PRKAR1B; PRKAR1B	Body; Body		
cg14390580	chr17	35873008	Island	*DUSP14*	3'UTR	0.041	-0.042
cg18560442	chr1	39174410	Island			0.041	-0.040
cg15842722	chr20	23499644	OpenSea	*CSTT*	TSS200	0.041	-0.034
cg08103551	chr11	76777993	Island	*CAPN5; CAPN5*	1stExon; 5'UTR	0.041	0.007
cg10504753	chr13	100258763	Island	*CLYBL; CLYBL*	TSS200; TSS200	0.041	0.010
cg14181391	chr20	33265182	Island	*PIGU*	TSS200	0.041	0.017
cg25372335	chr10	82168065	Island	*C10orf58*	TSS200	0.041	0.020
cg21345913	chr18	61822270	OpenSea	*LOC284294*	Body	0.041	0.025
cg07648504	chr19	21262035	N_Shelf			0.041	0.035
cg18693345	chr5	2754148	Island	*C5orf38*	Body	0.041	0.036
cg07627452	chr13	44978069	OpenSea			0.041	0.041
cg18155210	chr20	3223112	S_Shelf			0.041	0.050
cg14606589	chr1	7750796	OpenSea	*CAMTA1*	Body	0.041	0.061
cg25550913	chr13	114783665	Island	*RASA3*	Body	0.042	-0.079
cg25326090	chr19	47197766	S_Shore	*PRKD2; PRKD2; PRKD2; PRKD2*	Body; Body; Body; Body	0.044	-0.034
cg05801818	chr22	23262424	OpenSea			0.044	-0.024
cg06787731	chr14	38069079	Island			0.044	0.012
cg00996588	chr8	143625200	N_Shore	*BAI1*	Body	0.044	0.048
***Chr = Chromosome**							
***N_Shore/N_Shel = North Shore/Shelf**							
***S_Shore/S_Shelf = South Shore/Shelf**							

4.1.5 Pathways Enrichment Analysis

To address the potential biological significance of the impact of differentially methylated CpG sites, a gene set enrichment and pathways analysis was carried out. For the case-control study, a large number of significantly enriched pathways (FDR *P*-value < 0.05) in the GO database was collected (Figure 20). Among the top overrepresented categories, there are those related to several cell cycle processes, such as the cell cycle checkpoint, the DNA damage, and the DNA integrity checkpoint.

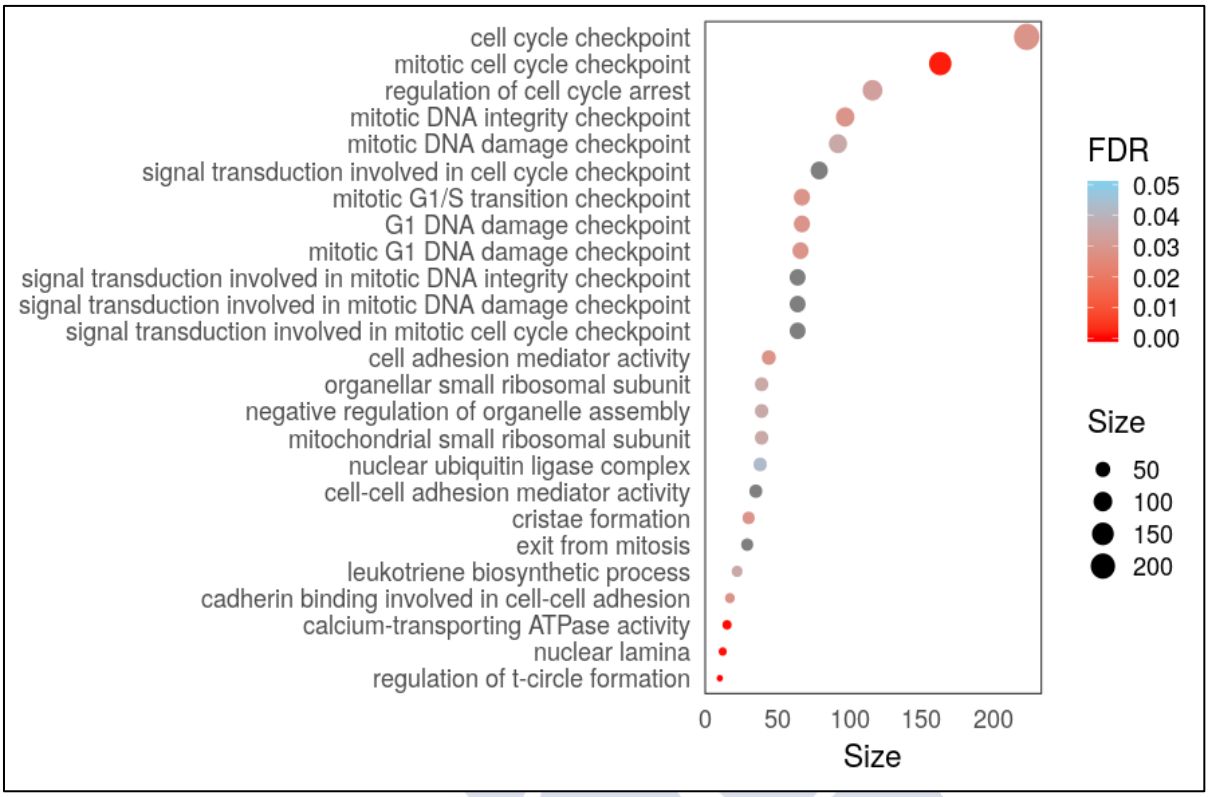


Figure 20: The top overrepresented categories of GO obtained with gene set pathways analyses performed considering all differentially methylated positions between the groups control and case.

4.1.6 Validation of candidate DMPs through Pyrosequencing in a new cohort

After analyzing the methylation levels of children with respiratory sequelae against children with normal recovery after RSV infection with the Infinium Human Methylation EPIC BeadChip array, the significant differences were corroborated by pyrosequencing, using a validation cohort of infants with the same clinical features and classification.

Pyrosequencing was used to evaluate the DNA methylation levels of one of the two most significant DMPs observed between children with normal recovery and those with respiratory sequelae after RSV the 24399977 within the *DDX27*. The methylation level of the CpG within *DDX27* was measured using the complete validation cohort of 43 patients, distributed into 22 cases and 21 controls.

No statistically significant differences were found in the methylation levels of DDX27 (*P*-value = 0.21; cg24399977) for the validation cohort. However, the differences between groups become statistically significant when both study and validation cohorts were considered together for that position (FDR *P*-value = 8.696e-06; cg24399977) (Figure 21).

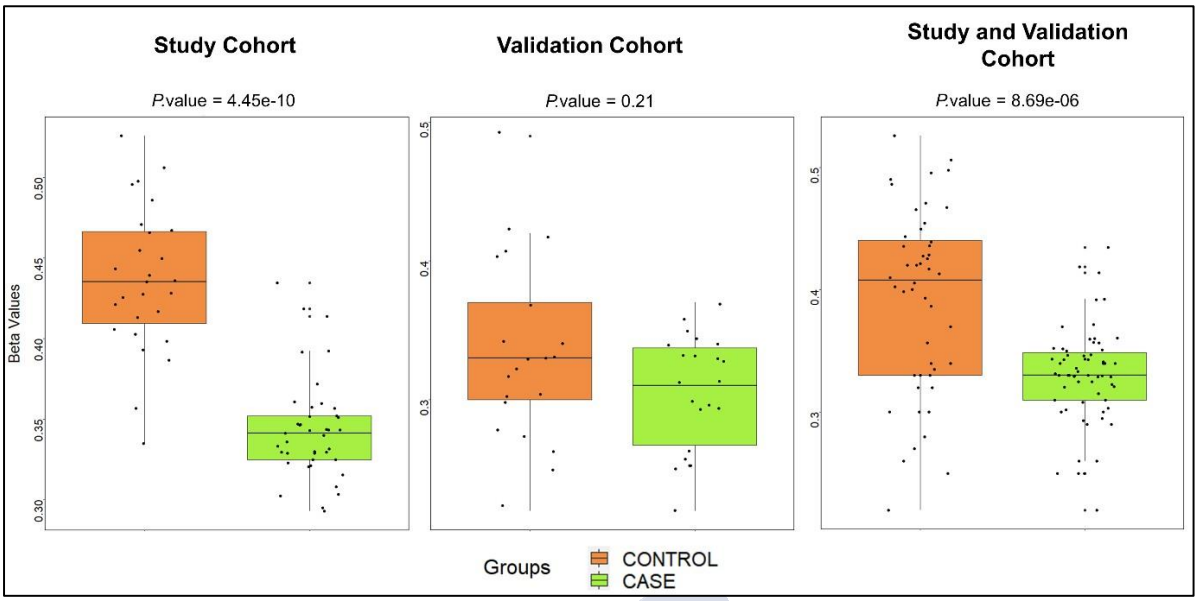


Figure 21: Boxplot showing DNA methylation differences of cg24399977 in study cohort, validation cohort and in both study and validation cohorts considered together.

4.2 CHANGES IN EPIGENETICS PROFILES THROUGHOUT EARLY CHILDHOOD AND THEIR RELATIONSHIP TO THE RESPONSE TO PNEUMOCOCCAL VACCINATION*

4.2.1 Preliminary analysis

For the second part of the study, twenty-four children (Table 2) who had received the 13-valent pneumococcal conjugate vaccine (PCV13) at 12 months of age, as part of a previously published clinical trial [241, 243, 244], were selected based on their aggregated IgG antibody response and categorized into high and low responders. Blood samples taken at 12 and 24 months of age were evaluated by DNA epigenetic microarray analysis.

All samples passed the quality control step as shown in the QC plot and the density plot in Figures 22a and 22b.

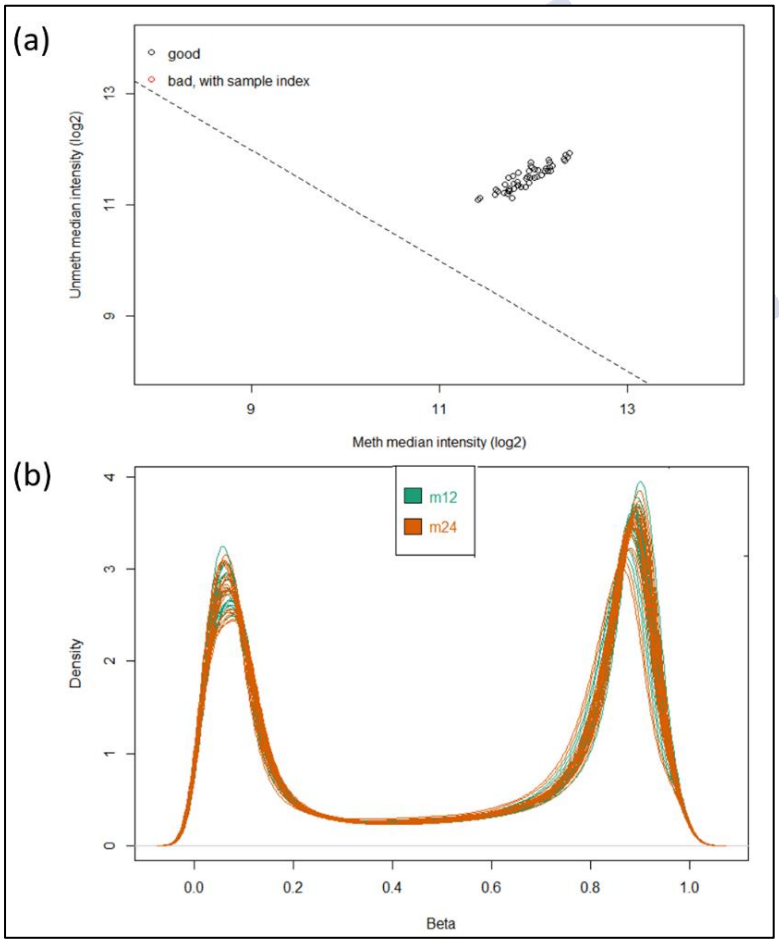


Figure 22: (a) Qc plot and (b) Density plot before showing the quality control of the raw data. In the density plot all samples exhibit two peaks: one corresponding to low or unmethylated probes with a b-value close to 0 and the second one corresponding to highly or fully methylated probes with a b-value close to 1, confirming the bimodal shape of the distribution of b-values.

* Pischedda, S., O'Connor, D., Fairfax, B.P. *et al.* Changes in epigenetic profiles throughout early childhood and their relationship to the response to pneumococcal vaccination. *Clin Epigenet* 13, 29 (2021). https://doi.org/10.1186/s13148-021-01012-w. (see Section 7. ANNEX)

After filtering preprocessing, 1,765 probes were removed for high *P*-value; 11,277 CpGs were eliminated because belonging to sex chromosomes; finally, other 16,717 probes were removed because of the presence of SNP or because of cross-reactive probes. 455,753 CpGs remained at the end for downstream analysis.

4.2.2 Widespread changes in the epigenome occur between 12 and 24 months of life.

The first step of the analysis focused on the methylation changes between the two timepoints (12 vs. 24 months). After correcting the linear model for cell-type composition, gender, and response to the vaccine, a paired-wise analysis was performed since the same subjects were compared at two different time points. The result of the linear model revealed 721 significant DMPs, associated with 421 unique genes, that were found statistically significant (FDR *P*-value <0.01) between 12 and 24 months of age. The QQPlot presented in Figure 23, shows quite large inflation that is not unexpected since it is due to the big differences in the epigenetic profiles of the children that are associated with age (epigenetic changes earlier in life), as it has been shown in previous studies [69]. For this reason, when correcting for age in the linear model, all the inflation is removed, as well as the significant DMPs observed. Therefore, this inflation is reflecting substantial age-related changes to the blood epigenome early in life.

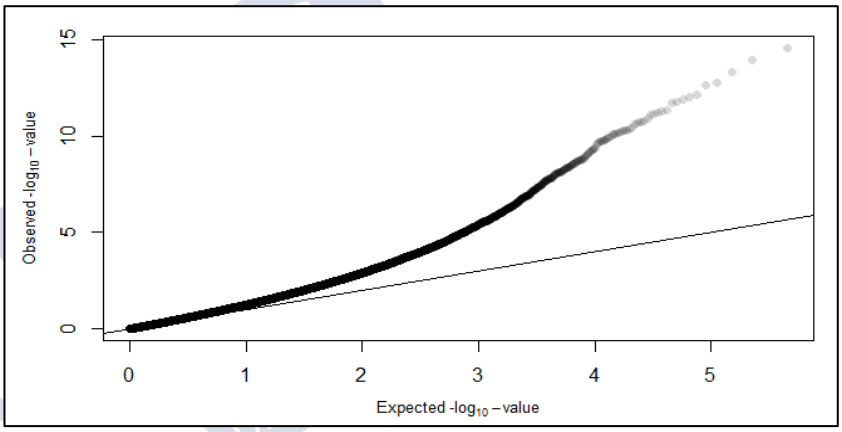


Figure 23: QQ plot of the observed p-values against the expected *P*-values

Of the significant DMPs, 314 CpGs (43.6%) were observed to be hypomethylated with the remaining 407 (56.4%) being hypermethylated, in samples collected at 12 months of age. Plotting the 721 positions it was possible to observe a clear separation between samples from 12 months and 24 months of age using principal component analysis and hierarchical clustering (Figure 24a and 24b). The first two principal components explain more than 40% of the total variation in the data.

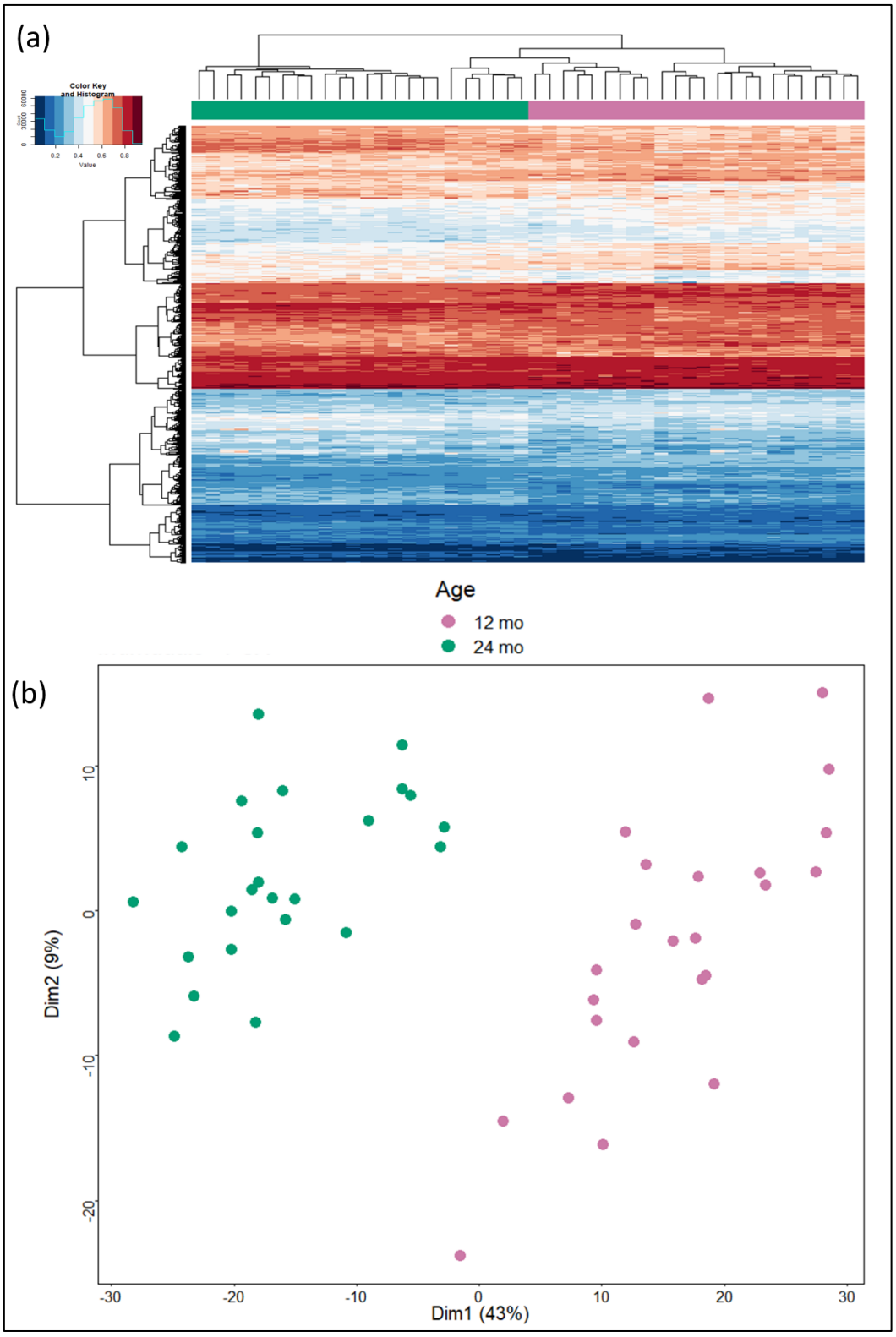


Figure 24: (a) Principal component analysis and (b) heatmap including unsupervised clustering of age (24 vs 12 months) using the 721 significant differentially methylated positions.

The following step in the analysis of the 721 DMPs, was the search of differentially methylated regions characterized by the presence of multiple CpGs inside. Here, looking at CpGs found together within the same Island region and showing differential methylation between the two different time points, several DMRs were identified. The largest DMRs were associated with the nuclear factor I X (NFIX) gene, characterized by 10 closed CpGs, and the proline-rich transmembrane protein 1 (PRRT1) gene depicted by 12 CpGs found within the same island region, chr6:32118101-32118544.

Five CpGs within the island region chr10:129534410-129537366 were identified associated with the Forkhead Box I2 (FOXI2) gene; other five DMPs in the island region chr17:7832532-7833164 were observed within the Potassium Voltage-Gated Channel Subfamily A Regulatory Beta Subunit 3 (KCNAB3) gene, and the same number of closed CpGs was detected within the island region chr1:55246742-55247519 of the Tetratricopeptide Repeat Domain 22 (TTC22) gene.

Looking at the distribution of the differentially methylated positions in the Island context, it was reported that most of the hypomethylated CpGs were mainly found in CpG islands whereas open sea, shore, and shelf regions were mostly hypermethylated (Figure 25a). The distribution of the 721 CpGs within the genomic regions was pretty different with most DMPs found in the gene body, followed by the TSS1500 and 5' UTRs; all the genomic regions were found almost dominated by hypomethylation patterns (Figure 25b).

Concerning chromosome distribution, it could be seen that chromosomes 1, 3, 6, 7, 11, 12, 15, and 16 are dominated by hypermethylated positions, while chromosomes 2, 17, and 9 are mainly characterized by CpGs with a lower level of methylation in 12 months than 24 months (Figure 25c).

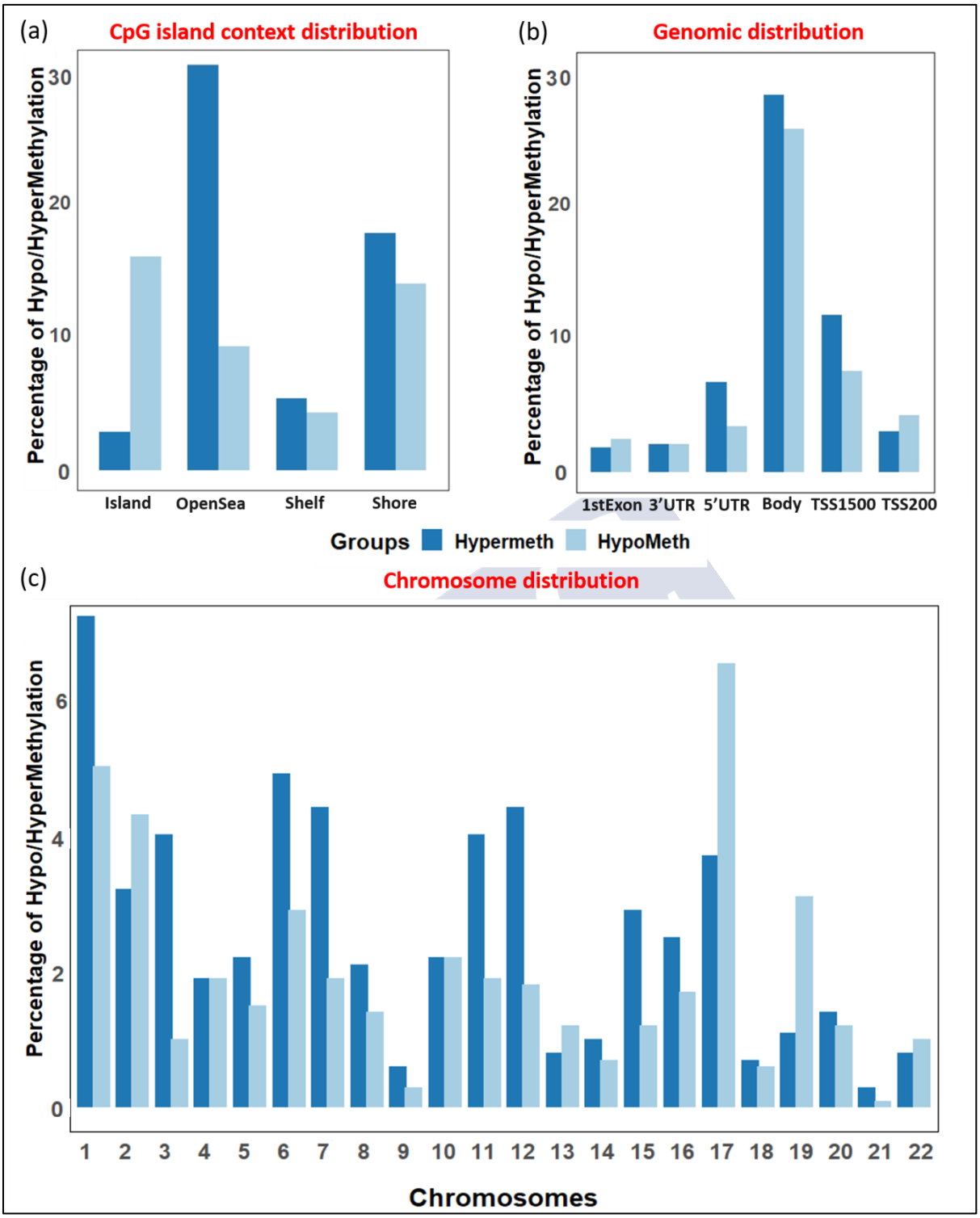


Figure 25: BarPlot showing the distribution of hypomethylated and hypermethylated DMPs (a) according to their distribution in the Island context and (b) in the genomic context and (c) in the chromosomes.

Further analyzing DMPs which showed the highest age-dependent differences in methylation by applying an absolute Delta Beta threshold of ≥0.10, it could be possible to detect 24 DMPs, considered as the DMPs most separating the two age groups (Table 9); of those, 18 CpGs exhibited a higher methylation level at 24 months such as three positions within the same island region of *PRRT1*, two positions within the *NFIX*, and two positions within the *KCNAB3*; the remaining 6 CpGs had a lower methylation pattern at 24 months, like the two positions within the *SLC22A8*.

Table 9: Differentially methylated positions with an adjusted *P*-value<0.01 and an absolute Delta Beta >0.10

CpG_ID	Chr	Position	Relation to Island	Gene Name	Gene Group	FDR *P*-value	Delta Beta
cg05825244	chr20	2730488	Island	*EBF4*	Body	2.96e-04	-0.220
cg25556035	chr19	13127873	S_Shelf*	*NFIX*	Body	4.27e-07	-0.193
cg14716990	chr10	129533731	N_Shore*			9.20e-08	-0.152
cg22268510	chr6	32118420	Island	*PRRT1*	Body	5.14e-07	-0.145
cg10767662	chr19	13127729	S_Shelf*	*NFIX*	Body	3.43e-07	-0.141
cg01323777	chr17	7832943	Island	*KCNAB3*	TSS200	4.09e-05	-0.137
cg00254681	chr6	32118457	Island	*PRRT1*	Body	1.01e-06	-0.132
cg13138089	chr2	233251770	Island	*ECEL1P2*	TSS200	7.75e-03	-0.131
cg13870494	chr9	72658358	N_Shore*	*MAMDC2*	TSS200	2.36e-05	-0.124
cg20471691	chr17	46681316	N_Shelf*	*LOC404266; LOC404266; HOXB6*	Body; Body; 5'UTR	4.05e-05	-0.116
cg23365801	chr17	7832909	Island	*KCNAB3*	TSS200	8.93e-04	-0.116
cg23491743	chr2	241989271	Island	*SNED1*	Body	2.15e-06	-0.114
cg27162435	chr17	7833163	Island	*KCNAB3*	TSS1500	8.86e-05	-0.109
cg00589520	chr7	23513039	N_Shore*			1.15e-09	-0.105
cg09490371	chr2	233253024	Island	*ECEL1P2*	TSS1500	2.54e-03	-0.105
cg11617964	chr6	32118399	Island	*PRRT1*	Body	1.82e-08	-0.103
cg25460807	chr8	21908022	S_Shelf*			6.25e-04	-0.101
cg11041817	chr17	46685327	Island	HOXB7	Body	4.88e-03	-0.101
cg16146033	chr11	62767323	OpenSea	*SLC22A8*	Body	5.40e-08	0.106
cg02481642	chr20	43343760	OpenSea	*WISP2*	TSS200	5.78e-06	0.109
cg25135018	chr1	154435948	OpenSea	*IL6R; IL6R*	Body; Body	1.16e-06	0.109
cg06688910	chr8	122466955	OpenSea			6.72e-08	0.110
cg17945323	chr11	62767406	OpenSea	*SLC22A8*	Body	6.66e-09	0.110
cg09978533	chr22	46465160	N_Shore*			9.39e-06	0.114
***Chr = Chromosome**							
***N_Shore/N_Shel = North Shore/Shelf**							
***S_Shore/S_Shelf = South Shore/Shelf**							

4.2.3 Age-associated changes in the epigenome are enriched for pathways linked to T cell regulation and activation

To investigate which molecular functions or biological processes were associated with the DMPs between 12 and 24 months of age, gene set enrichment analysis was performed. Significant enrichment (FDR *P*-value<0.05) was detected in 151 GO pathways, with the most significant pathways being related to the regulation of T cell activation, cell-cell adhesion, and regulation of cytokine production (Figure 26; Table 10).

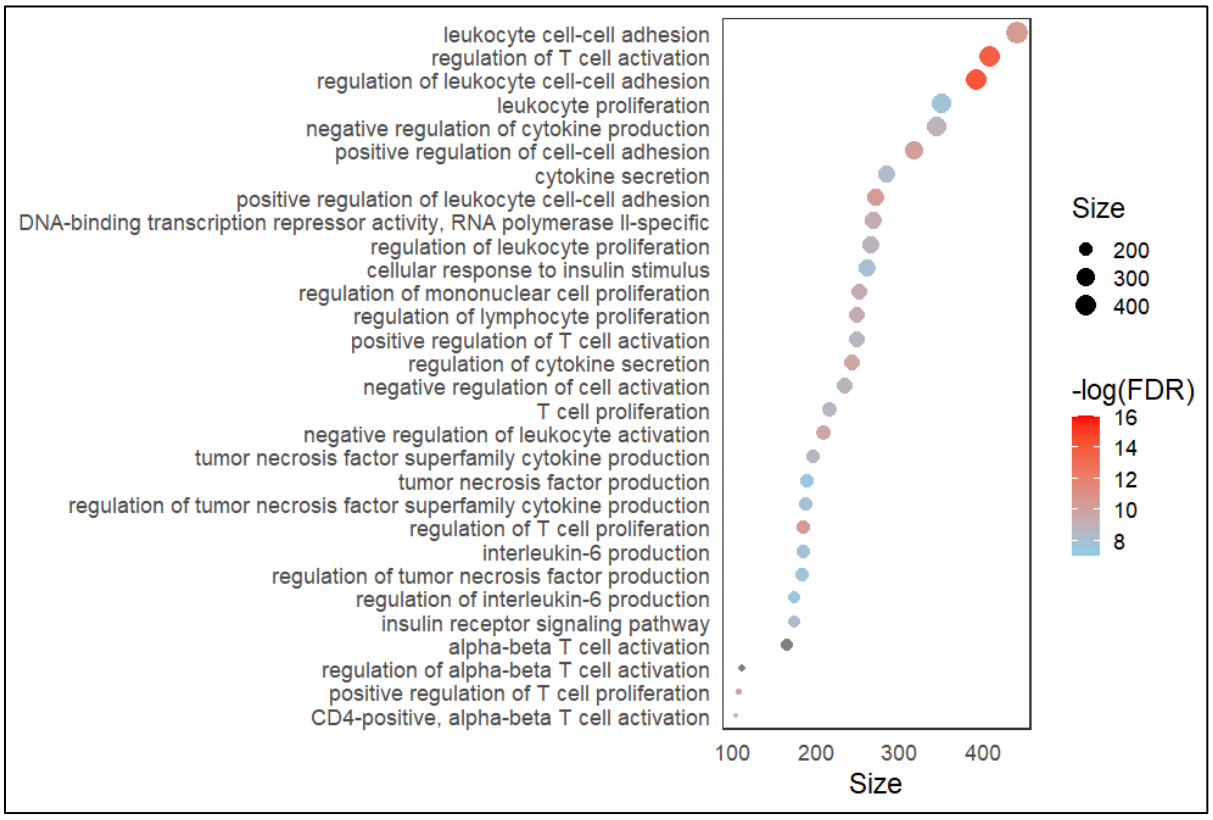


Figure 26: Dot plot of the top 30 GO pathways (FDR *P*-value<0.05). Size along the x-axis indicates the number of genes involved in each pathway. Dot colors correspond to the different FDR *P*-values associated with the pathways.

Table 10: List of GO pathways found to be significantly enriched

GO_ID	Description	Size	*P*-Value	FDR *P*-Value
GO:0046634	regulation of alpha-Beta T cell activation	112	2.92e-11	5.81e-08
GO:0046631	alpha-Beta T cell activation	166	8.38e-11	8.34e-08
GO:1903037	regulation of leukocyte cell-cell adhesion	391	1.12e-09	7.44e-07
GO:0050863	regulation of T cell activation	407	2.26e-09	1.12e-06
GO:0042129	regulation of T cell proliferation	185	6.90e-08	2.75e-05
GO:1903039	positive regulation of leukocyte cell-cell adhesion	272	1.04e-07	2.99e-05
GO:0007159	leukocyte cell-cell adhesion	440	1.05e-07	2.99e-05
GO:0022409	positive regulation of cell-cell adhesion	317	1.58e-07	3.93e-05
GO:0042102	positive regulation of T cell proliferation	107	2.80e-07	5.92e-05
GO:0050707	regulation of cytokine secretion	242	2.97e-07	5.92e-05
GO:0002695	negative regulation of leukocyte activation	209	3.54e-07	6.40e-05
GO:0032944	regulation of mononuclear cell proliferation	251	5.13e-07	8.51e-05
GO:0001227	DNA-binding transcription repressor activity. RNA polymerase II-specific	269	6.37e-07	9.46e-05
GO:0050670	regulation of lymphocyte proliferation	248	6.65e-07	9.46e-05
GO:0070663	regulation of leukocyte proliferation	266	1.12e-06	1.42e-04
GO:0050866	negative regulation of cell activation	234	1.21e-06	1.42e-04
GO:0001818	negative regulation of cytokine production	344	1.22e-06	1.42e-04
GO:0071706	tumor necrosis factor superfamily cytokine production	196	1.45e-06	1.57e-04
GO:0050870	positive regulation of T cell activation	248	1.50e-06	1.57e-04
GO:0042098	T cell proliferation	216	1.73e-06	1.73e-04
GO:0050663	cytokine secretion	285	2.64e-06	2.50e-04
GO:0008286	insulin receptor signaling pathway	174	3.01e-06	2.72e-04
GO:1903555	regulation of tumor necrosis factor superfamily cytokine production	188	4.13e-06	3.51e-04
GO:0032869	cellular response to insulin stimulus	262	4.23e-06	3.51e-04
GO:0032635	interleukin-6 production	184	4.87e-06	3.88e-04
GO:0070661	leukocyte proliferation	351	5.35e-06	4.10e-04
0	regulation of tumor necrosis factor production	183	6.05e-06	4.46e-04
GO:0032675	regulation of interleukin-6 production	174	7.48e-06	5.32e-04
GO:0035710	CD4-positive. alpha-Beta T cell activation	105	8.17e-06	5.54e-04
GO:0032640	tumor necrosis factor production	189	8.42e-06	5.54e-04
GO:0032755	positive regulation of interleukin-6 production	107	8.63e-06	5.54e-04
GO:0032946	positive regulation of mononuclear cell proliferation	153	1.02e-05	6.33e-04
GO:0070665	positive regulation of leukocyte proliferation	161	1.09e-05	6.56e-04
GO:0051048	negative regulation of secretion	265	1.15e-05	6.75e-04
GO:0046651	lymphocyte proliferation	320	1.26e-05	7.06e-04
GO:0002460	adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	446	1.28e-05	7.06e-04
GO:0050671	positive regulation of lymphocyte proliferation	150	1.35e-05	7.27e-04
GO:0032943	mononuclear cell proliferation	324	1.40e-05	7.31e-04
GO:0050868	negative regulation of T cell activation	133	1.45e-05	7.41e-04

GO_ID	Description	Size	*P*-Value	FDR *P*-Value
GO:1903531	negative regulation of secretion by cell	237	1.58e-05	7.84e-04
GO:0032868	response to insulin	323	1.71e-05	8.31e-04
GO:0005126	cytokine receptor binding	423	1.75e-05	8.31e-04
GO:1903038	negative regulation of leukocyte cell-cell adhesion	151	1.93e-05	8.92e-04
GO:0045785	positive regulation of cell adhesion	489	2.09e-05	9.33e-04
GO:0042107	cytokine metabolic process	150	2.11e-05	9.33e-04
GO:0008083	growth factor activity	183	2.93e-05	1.27e-03
GO:0071219	cellular response to molecule of bacterial origin	249	3.75e-05	1.59e-03
GO:0042089	cytokine biosynthetic process	149	3.87e-05	1.61e-03
GO:0055067	monovalent inorganic cation homeostasis	187	4.19e-05	1.70e-03
GO:0032649	regulation of interferon-gamma production	120	4.59e-05	1.83e-03
GO:0002367	cytokine production involved in immune response	111	5.09e-05	1.97e-03
GO:0022408	negative regulation of cell-cell adhesion	210	5.13e-05	1.97e-03
GO:0071222	cellular response to lipopolysaccharide	231	5.26e-05	1.97e-03
GO:0045165	cell fate commitment	313	5.99e-05	2.21e-03
GO:0051056	regulation of small GTPase mediated signal transduction	416	6.53e-05	2.36e-03
GO:0002699	positive regulation of immune effector process	258	6.64e-05	2.36e-03
GO:0072080	nephron tubule development	105	7.12e-05	2.49e-03
GO:0051250	negative regulation of lymphocyte activation	173	7.82e-05	2.68e-03
GO:0032609	interferon-gamma production	136	8.66e-05	2.92e-03
GO:0050709	negative regulation of protein secretion	148	9.29e-05	3.04e-03
GO:0030217	T cell differentiation	290	9.33e-05	3.04e-03
GO:0050673	epithelial cell proliferation	497	1.07e-04	3.45e-03
GO:0030667	secretory granule membrane	319	1.14e-04	3.61e-03
GO:0061326	renal tubule development	109	1.42e-04	4.39e-03
GO:1903901	negative regulation of viral life cycle	102	1.45e-04	4.39e-03
GO:0002696	positive regulation of leukocyte activation	448	1.45e-04	4.39e-03
GO:0046632	alpha-Beta T cell differentiation	116	1.64e-04	4.87e-03
GO:0071216	cellular response to biotic stimulus	276	1.74e-04	5.10e-03
GO:0050867	positive regulation of cell activation	465	1.89e-04	5.41e-03
GO:0002792	negative regulation of peptide secretion	157	1.90e-04	5.41e-03
GO:0014706	striated muscle tissue development	461	2.08e-04	5.84e-03
GO:0048525	negative regulation of viral process	120	2.12e-04	5.88e-03
GO:0050679	positive regulation of epithelial cell proliferation	225	2.27e-04	6.19e-03
GO:0060538	skeletal muscle organ development	192	2.31e-04	6.21e-03
GO:0030534	adult behavior	148	2.40e-04	6.36e-03
GO:0051251	positive regulation of lymphocyte activation	397	2.63e-04	6.89e-03
GO:0060537	muscle tissue development	484	3.00e-04	7.72e-03
GO:0042626	ATPase activity. coupled to transmembrane movement of substances	181	3.09e-04	7.72e-03
GO:0043492	ATPase activity. coupled to movement of substances	185	3.09e-04	7.72e-03
GO:0007519	skeletal muscle tissue development	182	3.12e-04	7.72e-03
GO:0032652	regulation of interleukin-1 production	124	3.14e-04	7.72e-03

GO_ID	Description	Size	*P*-Value	FDR *P*-Value
GO:0071887	leukocyte apoptotic process	109	3.34e-04	8.10e-03
GO:0015405	P-P-bond-hydrolysis-driven transmembrane transporter activity	190	3.97e-04	9.53e-03
GO:0097191	extrinsic apoptotic signaling pathway	291	4.15e-04	9.76e-03
GO:0002286	T cell activation involved in immune response	116	4.17e-04	9.76e-03
GO:0007517	muscle organ development	478	4.49e-04	1.04e-02
GO:0005125	cytokine activity	294	4.70e-04	1.08e-02
GO:0033002	muscle cell proliferation	278	5.06e-04	1.15e-02
GO:0007162	negative regulation of cell adhesion	335	5.42e-04	1.20e-02
GO:0015399	primary active transmembrane transporter activity	192	5.42e-04	1.20e-02
GO:0045444	fat cell differentiation	237	5.49e-04	1.20e-02
GO:0050777	negative regulation of immune response	169	5.59e-04	1.21e-02
GO:0050678	regulation of epithelial cell proliferation	427	5.97e-04	1.28e-02
GO:1901653	cellular response to peptide	469	6.92e-04	1.47e-02
GO:0030168	platelet activation	183	7.05e-04	1.47e-02
GO:0032612	interleukin-1 production	140	7.11e-04	1.47e-02
GO:0003231	cardiac ventricle development	139	7.16e-04	1.47e-02
GO:0001763	morphogenesis of a branching structure	214	7.26e-04	1.47e-02
GO:0050715	positive regulation of cytokine secretion	164	7.50e-04	1.51e-02
GO:0030098	lymphocyte differentiation	421	7.69e-04	1.51e-02
GO:0004896	cytokine receptor activity	179	7.71e-04	1.51e-02
GO:0046578	regulation of Ras protein signal transduction	261	7.83e-04	1.51e-02
GO:0002449	lymphocyte mediated immunity	434	7.83e-04	1.51e-02
GO:0002700	regulation of production of molecular mediator of immune response	159	7.87e-04	1.51e-02
GO:0050817	coagulation	439	8.11e-04	1.54e-02
GO:0003007	heart morphogenesis	316	8.76e-04	1.65e-02
GO:0042035	regulation of cytokine biosynthetic process	137	9.09e-04	1.69e-02
GO:1904018	positive regulation of vasculature development	277	9.65e-04	1.78e-02
GO:1902105	regulation of leukocyte differentiation	323	9.83e-04	1.79e-02
GO:0032147	activation of protein kinase activity	415	9.90e-04	1.79e-02
GO:0060562	epithelial tube morphogenesis	370	1.11e-03	1.97e-02
GO:0007596	blood coagulation	431	1.12e-03	1.97e-02
GO:0071375	cellular response to peptide hormone stimulus	394	1.12e-03	1.97e-02
GO:0051224	negative regulation of protein transport	202	1.16e-03	2.00e-02
GO:0030004	cellular monovalent inorganic cation homeostasis	132	1.17e-03	2.00e-02
GO:1901617	organic hydroxy compound biosynthetic process	373	1.17e-03	2.00e-02
GO:0007599	hemostasis	438	1.23e-03	2.09e-02
GO:0032103	positive regulation of response to external stimulus	380	1.27e-03	2.14e-02
GO:0022853	active ion transmembrane transporter activity	106	1.34e-03	2.21e-02
GO:0042625	ATPase coupled ion transmembrane transporter activity	106	1.34e-03	2.21e-02
GO:0006909	phagocytosis	448	1.34e-03	2.21e-02
GO:1904950	negative regulation of establishment of protein localization	206	1.50e-03	2.45e-02

GO_ID	Description	Size	*P*-Value	FDR *P*-Value
GO:0002822	regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	180	1.53e-03	2.48e-02
GO:0002698	negative regulation of immune effector process	133	1.62e-03	2.57e-02
GO:0045580	regulation of T cell differentiation	161	1.62e-03	2.57e-02
GO:0009952	anterior/posterior pattern specification	236	1.62e-03	2.57e-02
GO:0016323	basolateral plasma membrane	234	1.65e-03	2.59e-02
GO:0019216	regulation of lipid metabolic process	498	1.68e-03	2.62e-02
GO:0032651	regulation of interleukin-1 Beta production	104	1.74e-03	2.69e-02
GO:0003205	cardiac chamber development	190	2.08e-03	3.19e-02
GO:0032611	interleukin-1 Beta production	120	2.22e-03	3.37e-02
GO:1901222	regulation of NIK/NF-kappaB signaling	130	2.31e-03	3.46e-02
GO:0045765	regulation of angiogenesis	465	2.31e-03	3.46e-02
GO:0003279	cardiac septum development	122	2.34e-03	3.48e-02
GO:0071346	cellular response to interferon-gamma	202	2.40e-03	3.54e-02
GO:0002793	positive regulation of peptide secretion	337	2.48e-03	3.63e-02
GO:0002576	platelet degranulation	131	2.58e-03	3.74e-02
GO:0050714	positive regulation of protein secretion	308	2.63e-03	3.79e-02
GO:0001938	positive regulation of endothelial cell proliferation	125	2.77e-03	3.96e-02
GO:0009615	response to virus	422	2.80e-03	3.98e-02
GO:0045619	regulation of lymphocyte differentiation	195	2.89e-03	4.08e-02
GO:0050731	positive regulation of peptidyl-tyrosine phosphorylation	225	2.93e-03	4.09e-02
GO:0006959	humoral immune response	497	2.94e-03	4.09e-02
GO:0048754	branching morphogenesis of an epithelial tube	162	2.99e-03	4.10e-02
GO:0002703	regulation of leukocyte mediated immunity	241	2.99e-03	4.10e-02
GO:0032496	response to lipopolysaccharide	366	3.02e-03	4.12e-02
GO:0002824	positive regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	122	3.19e-03	4.32e-02
GO:0003206	cardiac chamber morphogenesis	142	3.21e-03	4.32e-02
GO:0062014	negative regulation of small molecule metabolic process	119	3.25e-03	4.35e-02
GO:0072009	nephron epithelium development	127	3.43e-03	4.55e-02
GO:1902106	negative regulation of leukocyte differentiation	109	3.58e-03	4.72e-02
GO:0014065	phosphatidylinositol 3-kinase signaling	176	3.62e-03	4.74e-02

In order to deeply understand the relationship between DNA methylation changes associated with age and the immune system, all CpGs settled immune genes according to Gene Ontology terms were investigated, among the significant positions. Thus, it was possible to observe 86 CpGs associated with 69 unique genes related to immune system processes. Of these positions, 53 CpGs exhibit a decrease in methylation with age, while the remaining 32 were characterized by an increase in methylation from 12 to 24 months (Figure 27). Some of the selected genes were represented by multiple DMPs, such as *HOXB6, NOD2*, and *LAG3*, where it was possible to observe both hypo- and hypermethylated CpGs. Among the hypermethylated positions observed in 12 months, most of them belong to genes involved in the control and regulation of T cell activation and proliferation, like CD*44, CHST3, CYP26B1, DGKZ, EFNB3, FCER1G, IL12A, ITGAX, LAG3, MAPK8IP1, PRKCQ, PTGER4, RASAL3, RUNX2, SOCS1*, and *ZC3H12A*. Another group of CpGs was found within genes that regulate differentiation and proliferation of B cells, such as *CLCF1, NTRK1, INSRR, RFTN1, SPNS2, BCL10*. In addition, CpG sites within genes that regulated the innate immune response were detected, like *TRIM15, PLD3, AIM2, HEXIM1, PML, IL1RAP, NOD1, FER, NLRC5, PRDM1*. Finally, positions within genes involved in leukocyte migration were identified like *AOC2, CX3CL1*, and *ITGA2B*, that display an opposite epigenetic pattern with a hypomethylation profile in 12 months.

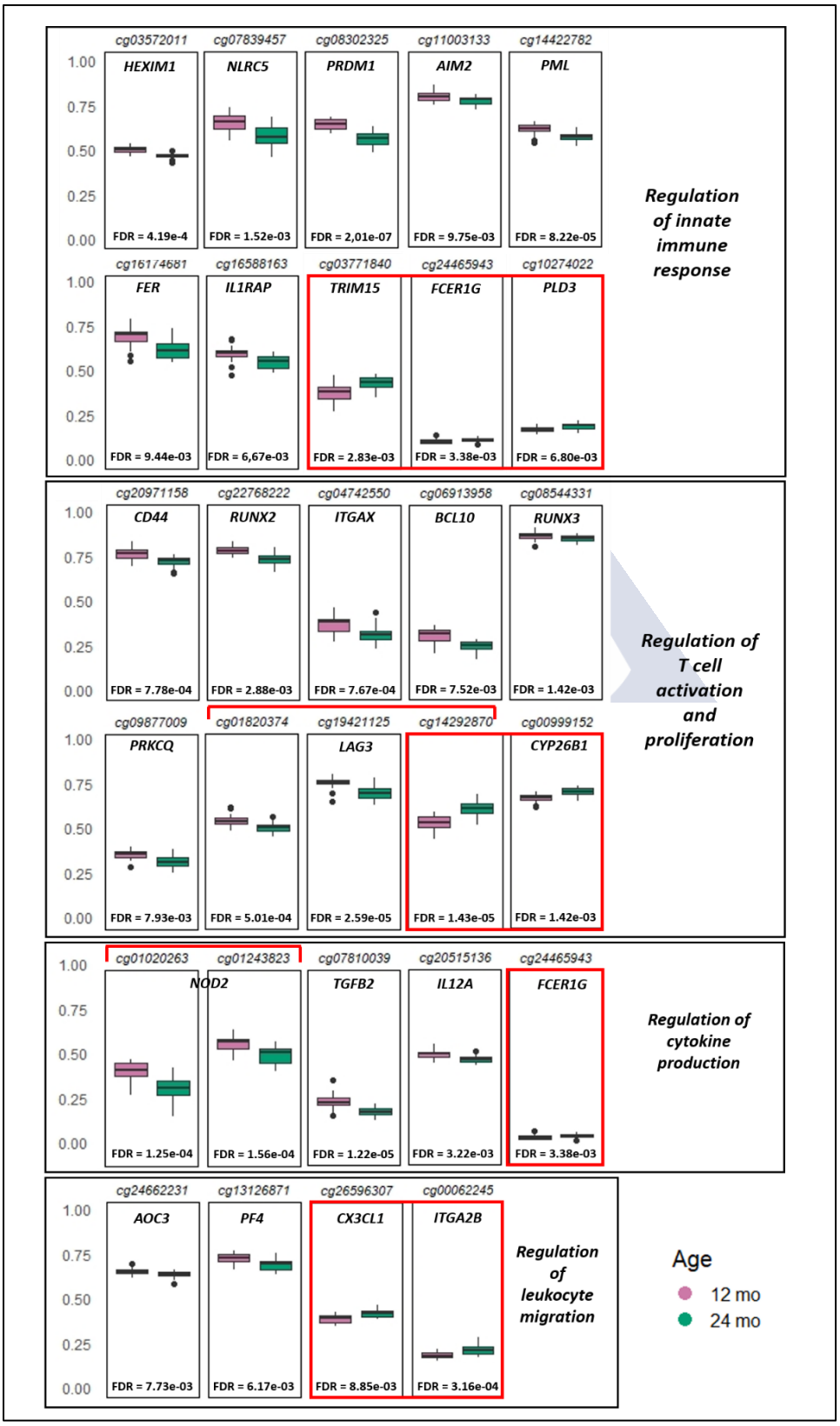


Figure 27: Thirty of the most significant DMPs associated with immune genes grouped according to their GO function. The y-axis represents the Beta value for each position. Red squares indicate CpGs with hypermethylation at 24 compared with 12 months samples, while for the remaining CpGs, hypomethylation was found at 24 compared with 12 months of age.

To understand if the different responsiveness to vaccination was associated with the "biological age" of the subjects, the epigenetic age estimator developed by Horvath in 2013 [63] was used to calculate biological age based on methylation patterns. Firstly, the age predictor was applied to the results obtained from samples at the two-time points revealing that chronological and predicted biological ages were correlated (Figure 28).

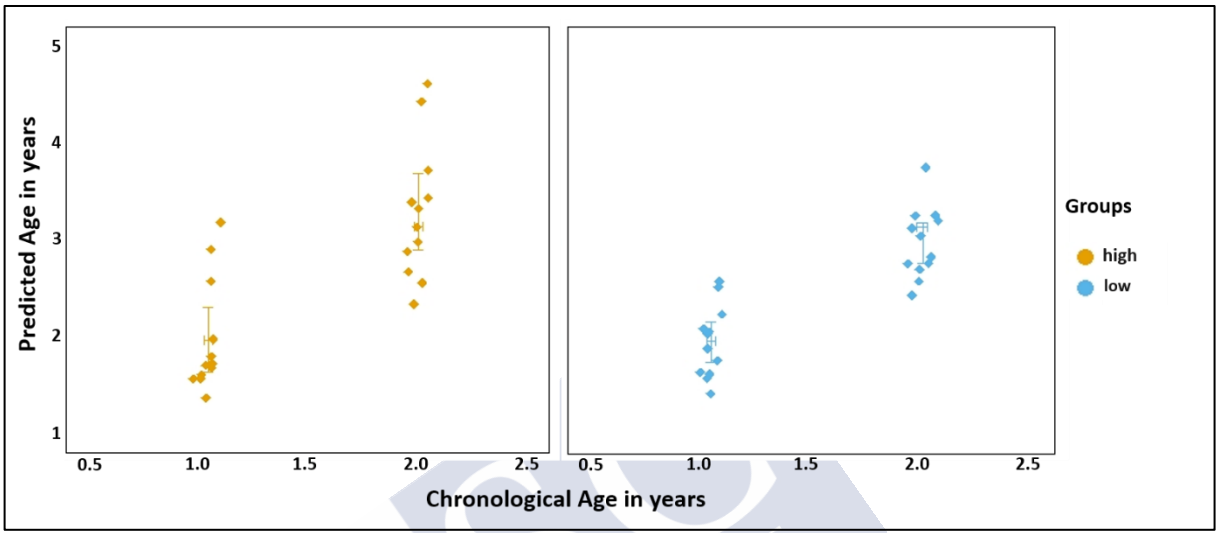


Figure 28: Correlation between chronological (x axis) and predicted biological age (y axis) as calculated with the Horvath epigenetic age estimator.

Then the differences between high and low responders in the two time-points were observed separately, but no significant differences in predicted biological age were detected (Figure 29).

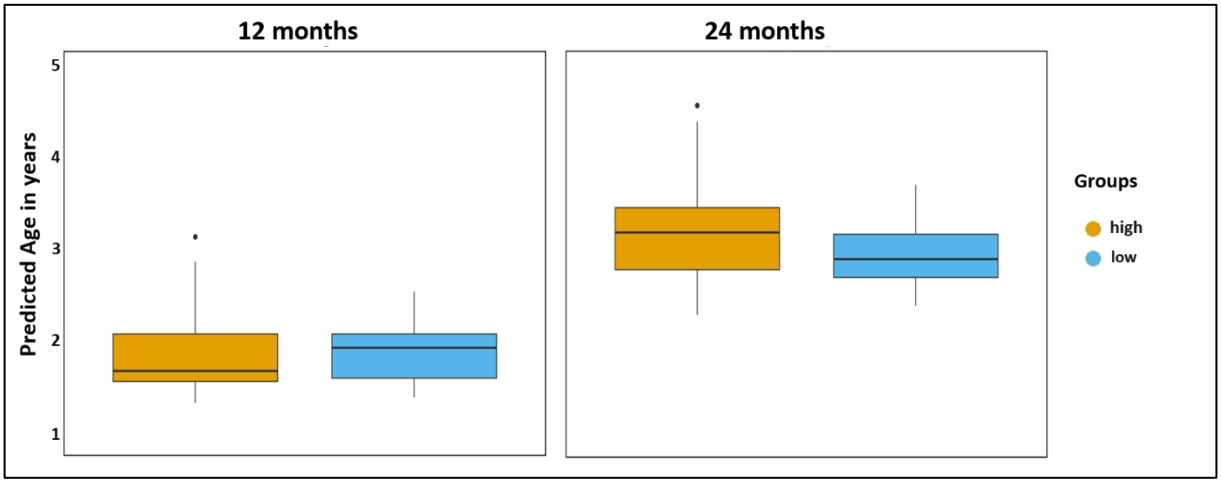


Figure 29: Comparison of predicted biological age between high- and low-responders at the two study time points.

4.2.4 Hypomethylation of HLA-DPB1 and hypermethylation of IL6 are associated with more robust responses to pneumococcal vaccination.

The next step of the analysis was performed to assess whether differences in antibody responses against vaccine antigens were associated with methylation patterns by comparing pre-vaccination blood epigenome profiles of high responders to those of low responders, at the basal time-point (12 months). After applying the same linear model described previously and an FDR *P*-value<0.01 correction, not statistically significant DMPs were detected. However, using a less stringent statistical threshold (uncorrected *P*-value <0.01), 4,067 differentially methylated CpG sites associated with 2,797 unique genes that distinguished high from low responders in a PCA (Figure 30a) and an unsupervised clustering method (Figure 30b) could be detected.

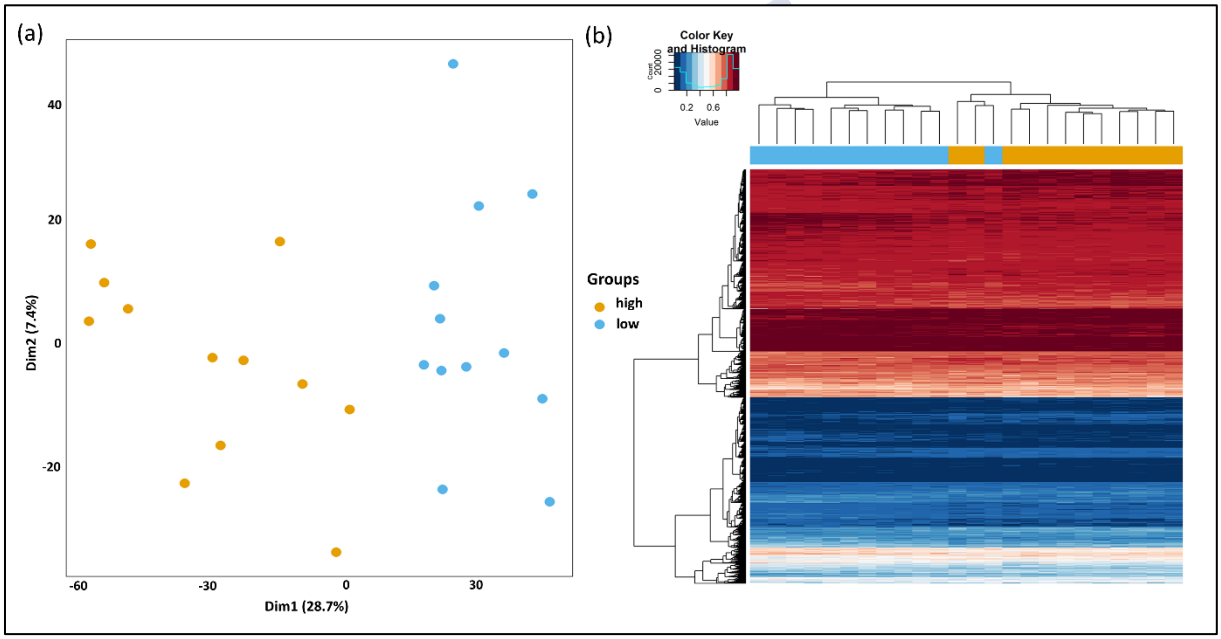


Figure 30: (a) Principal components analysis and (b) unsupervised clustering with heatmap of 4067 differentially methylated CpG sites (uncorrected *P*-value <0.01) at 12 months of age and grouping by high and low vaccine responders

Of these positions, 2,023 were found to be hypomethylated and 2,044 hypermethylated in high responders. Similar results were obtained when comparing the same groups at 24 months (Figure 31), where 4,680 CpGs, associated with 3,186 unique genes, among which 2,365 positions show a high level of methylation, while the remaining 2,315 show a low level of methylation in children with a most robust response, were identified.

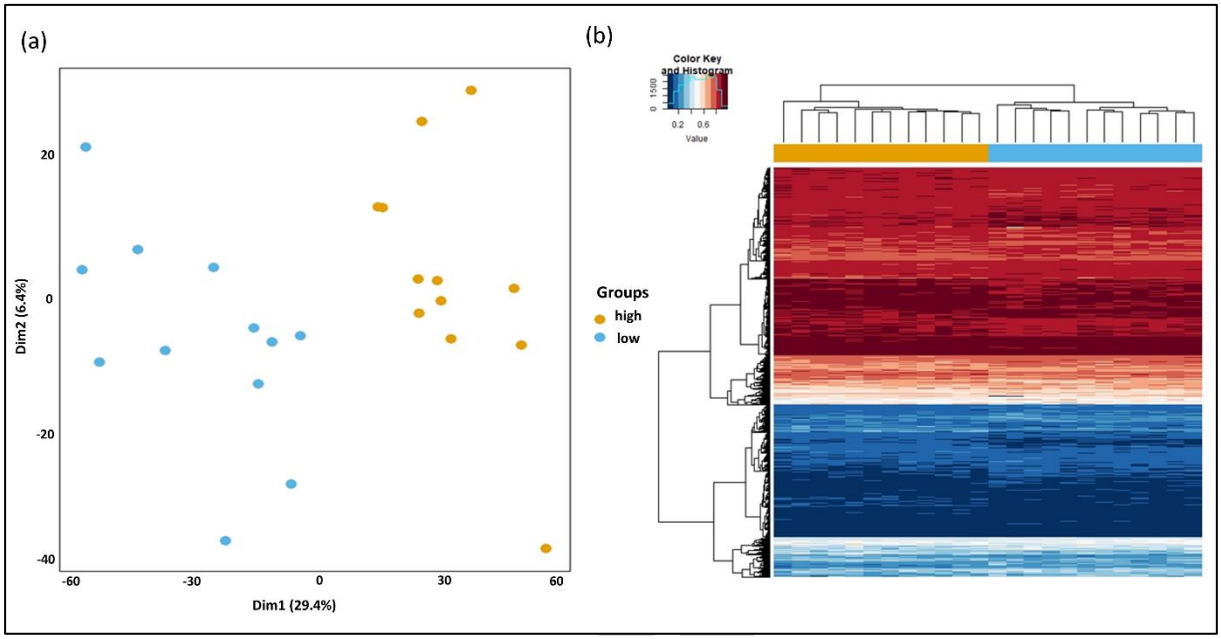


Figure 31: (a) Principal component analysis and (b) unsupervised clustering and heatmap of 5233 differentially methylated CpG sites detected using a threshold of an uncorrected *P*-value <0.01 at 24 months of age and grouping by high and low vaccine responders.

To gain greater insights on which DMPs may influence the different responses to the vaccine, among the 4067 DMPs, those positions which showed the highest difference in methylation between high and low vaccine responders (Delta Beta) and presented the lowest *P*-value, were collected. Of the selected DMPs, regions containing multiple DMPs in genes involved in immune system processes were assessed. Here, among the top DMPs identified, it was possible to observe 4 hypomethylated positions in high vaccine responders within the chr6:33048416-33048814 island region of the HLA class II histocompatibility antigen DP(W2) Beta chain (*HLA-DPB1*). In addition, there were also two CpGs associated with interleukin 6 (*IL6*) that showed hypermethylation in high with respect to low vaccine responders (Figure 32).

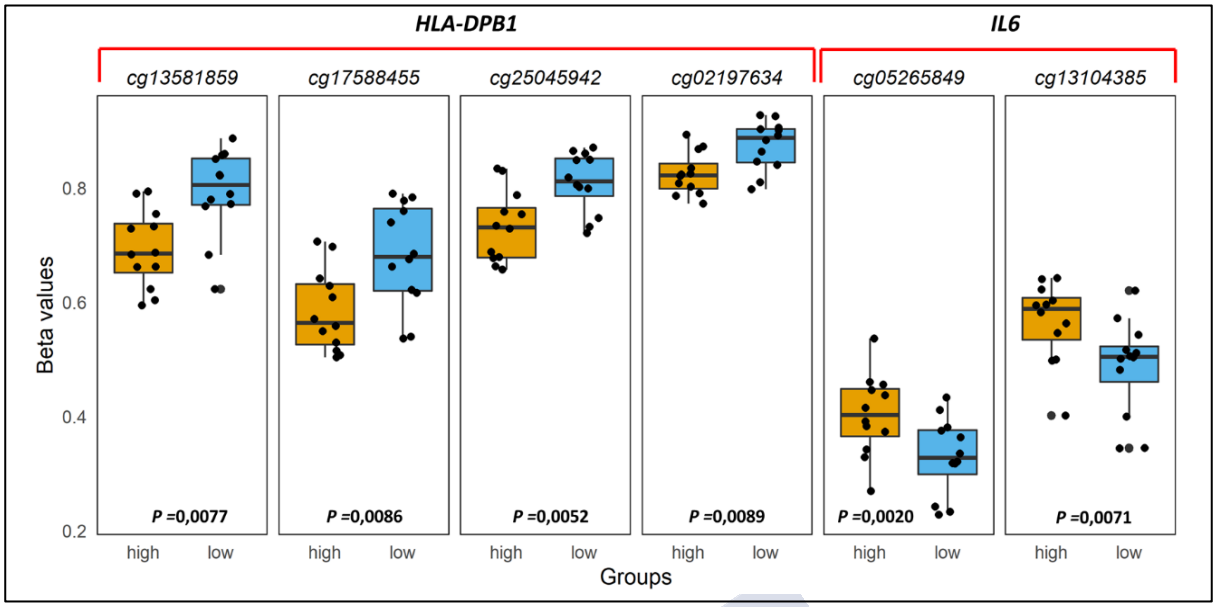


Figure 32: Boxplots showing 4 DMPs associated with hypomethylation of HLA-DPB1 and 2 DMPs associated with hypermethylation of IL6 in high vaccine responders. The y-axis represents the Beta value for each position in both groups (yellow for high responders, light blue for low responders).

5
DISCUSSION

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

In the present project, the high-throughput Illumina Infinium Methylation BeadChip technology was used to assess the role of epigenetics in the regulation and the modulation of the immune response in infants after infection and vaccination. Although some limitations could be pointed out, such as the low sample size and the lack of validation for some of the DMPs identified, the global outcomes of data analysis revealed that changes in DNA methylation are strongly implicated in the regulation of several immune processes which drive the differential immune response observed in children after foreign stimuli.

5.1 THE ROLE OF HOST EPIGENETIC CHANGES IN RESPIRATORY SEQUELAE AFTER RSV INFECTION

The first objectives of the present project focus on the search for a relationship between methylation changes after documented RSV infection, and posterior development of wheezing and/or asthma. Here the aim was to assess the association between RSV infection and wheezing and/or asthma development in infants since the underlying biological mechanism through which recurrent wheezing and asthma are prone to increase after RSV infection is not fully understood. The investigation of this tricky mechanism is fundamental for asthma prevention and treatment. Epigenetics is supposed to play an important role in the modulation of the interaction between host and pathogen, as well as in the regulation of the epithelial immune response after RSV infection [289]. It is clear that DNA methylation and gene expression are strictly correlated; several studies on cancer suggest that hypomethylation promote tumorigenesis, activating proto-oncogenes, while hypermethylation occurring specifically at the promoter regions is the leading cause of the expression silencing of key genes required for normal cell growth and differentiation [290, 291]. Although this mechanism in tumorigenesis and normal development is well accepted, the precise process describing how DNA methylation promotes misregulation of gene expression is still under investigation.

For this part of the project the analysis of DNA methylome of 77 children suffering different types of respiratory sequelae after an RSV infection was carried out and differential methylation patterns between children with normal recovery and those exhibiting subsequent recurrent wheezing and/or asthma, because of the RSV infection, was unveiled. Before performing the linear model to check for DMPs between case and control groups, the estimation of the relative proportion of cell types within the peripheral blood samples revealed a statistically significant difference in the number of natural killer cells between case and control,

in the specific children with respiratory sequelae show a higher proportion of NK cells than children with a normal recovery. It has been demonstrated that RSV infection can induce severe acute lung immune injury promoting the accumulation of lung natural killer cells at the early stage of infection in mice. The consequent activation of NK cells induces the production of IFN-γ, responsible for acute lung immune injury [188]. Concerning the role of NK cells in allergic lung inflammation, there is evidence suggesting that NK cells are involved in both the promotion and inhibition of allergic lung inflammation, however, the mechanisms underlying these contrasting roles are still not clear. Indeed, in 2010 it has been demonstrated that mice infected with RSV and deficient of NK cells during primary infection, showed suppression of IFN-γ production, leading to an allergic airway response and a subsequent increased production of IL-25 by airway epithelial cells [292]. Contrarily, a previous study demonstrated that the frequency of NK cell was higher in peripheral blood from asthmatic children during acute exacerbation when compared to stable asthmatics, suggesting that NK cells may influence multiple pathways during the progression of asthma [293].

After correcting the linear model for cell composition, age, and gender, the outcomes of the preliminary analysis revealed 5,097 CpGs significantly differentially methylated when comparing cases vs controls. Some of the positions detected in the analysis were already reported in different EWAS found in the literature, but no one of the reported DMPs was found to be associated with infectious disease outcomes or with the development of asthma and/or wheezing after infection for RSV so far.

The most significant DMP (Delta Beta > 0.13, FDR *P*-value = 2.77e-10) is the cg24509358 in the position 28416532 of the eye absent homolog 3 gene (EYA3) located in chromosome 1, which shows a loss in methylation in the case group. *EYA3* encodes for a tyrosine phosphatase involved in DNA repair and in distinguishing apoptotic and repair responses to genotoxicity. Recently, Wang et al., [294] suggested an over-expression for this gene in patients with pulmonary arterial hypertension (PAH), as they supposed that high levels of EYA3 gene products in the pulmonary arterial smooth muscle cells, might stimulate the survival of pulmonary vascular cells in the presence of DNA damage, causing vascular remodeling, a typical feature of PAH. These authors demonstrated that the inhibition of *EYA3* could protect from the development of vascular remodeling and substantially reverse established PAH. Although PAH and asthma are considered two different and unrelated phenomena, there are pieces of evidence demonstrating that they share some pathological features such as

inflammation, smooths muscle constriction, and proliferation [295]. The airway inflammation typical of asthma provokes tissue injury and related structural changes of airways walls as well as vascular changes in the mucosa, which may contribute to airway remodeling, which in turn is correlated with airflow obstruction and airway hyperresponsiveness in asthmatic patients [296-298]. Although DNA methylation in the CpG selected might not be exerting changes in the expression of the EYA3 gene, the putative involvement of the gene in the vascular remodeling (a key process in the asthma pathology), might indicate a possible connection between the altered methylation status of the gene and its uncontrolled gene expression. It has been reported that changes in the methylation pattern of CpG island shore were strongly related to gene expression [37]. Thus, the hypothesis is that the loss in methylation observed in the transcription start site of the shore area of the EYA3 gene that is observed in the case group might cause the over-activation of this gene, with the possible stimulation of the vascular remodeling in children which present recurrent wheezing and asthma after RSV infection.

The second position that shows a hypomethylation pattern for children with subsequent recurrent wheezing and/or asthma after RSV infection and was reported to be one of the most significant DMPs (FDR *P*-value = 4.45e-10) between the two groups is the cg23499977 in the position 47835458 of the TSS1500 of the *DDX27* in the chromosome 20. The position in this gene has not yet been described, even if the gene has been found to be one of the differentially expressed genes when comparing severe asthmatic with non-asthmatic patients [299]. The DEAD-Box Helicase 27 which is characterized by the conserved motif Asp-Glu-Ala-Asp (*DEAD*), is a putative RNA helicase. These genes are involved in the alteration of RNA secondary structure, in embryogenesis, spermatogenesis, and cellular growth and division. High levels of DDX27 mRNA are shown to increase cell proliferation in patients with colorectal cancer, while silencing of *DDX27* exerted an opposite effect inhibiting tumor growth and lung metastasis in vitro and in vivo. In addition, the NF-κB pathway has been identified as the major target of (DEAD)-box RNA helicases and *DDX27* in CRC. NF-κB transcription factor family regulates several genes involved in inflammation, cell proliferation, and metastasis, and up-regulation of DDX27 is observed to enhance and prolong TNF-α-mediated NF-κB signaling [300]. The NF-κB is an important transcription factor that plays a critical role in the production of many inflammatory cytokines. It is found to be associated with allergic airway diseases, and it is observed to be activated in bronchial asthmatic patient biopsies and the epithelium of the airways in mice [301]. Therefore, it is likely that the low methylation observed in the position

within DDX27, induces an up-regulation of the gene with possible consequent activation of the NF-κB signaling.

Another interesting position showing the highest difference in methylation among cases and controls is the cg21226224 (Delta Beta > 0.14, FDR *P*-value = 3.29e-04) within the TSS1500 of chromosome 8 of the SOX17 gene; the position shows a lower level of methylation for the control group suggesting a possible overexpression of the gene. The Sry-related high-mobility-group box 17 (SOX17) gene is shown to be regulated by the cytokine IL-33 during the development of asthma; recently it has been demonstrated that the genetic ablation of this gene in endothelial cells provoke an alleviation of some of the key clinical features of asthma [302].

Focusing on some interesting positions within genes that have been previously reported in the literature as associated with asthma risk, it could be possible to observe some genes that were involved in the transforming growth factor-Beta (TGF-β) signaling pathway and in pathways related to its regulation. TGF-β is one of the cytokines family that has been suggested to have a critical role in the development of airway inflammation and remodeling in asthma [303]. In asthmatic airways, TGF-β can induce an antiapoptotic effect in the airway epithelial cells through the SMAD signaling pathway and can trigger apoptosis activating MAPK signaling pathways [304]. Therefore SMAD proteins are the main signaling transducers of the TFG-β and work mediating signaling from cell surface receptors to nuclear target genes regulating various processes and cellular functions [305]. With this signaling cascade, SMAD proteins regulating the activity of TGF-β might play a role in the modulation of the airway remodeling process in asthma [306]. Among the significant DMPs within genes involved in the TFG-β related pathways, four CpGs within different SMAD genes that exhibited alterations in the methylation pattern between cases and controls were identified, in line with previous studies that reported findings on the altered methylation pattern of those genes as associated with increased asthma risk [307]. The methylation patterns observed in SMAD, as well as in other genes associated with the regulation of the TGF-β, could clinically be involved in the induction of a moderate or severe course of asthma, a common respiratory sequela observed in children after RSV infection.

Among the pathways found to be significantly enriched from the functional analysis, particular attention was focused on those involved in the cell cycle checkpoint that is observed

to be the most significantly enriched. These results could suggest that cell proliferation may play an important role in the pathogenesis of wheezing and asthma after RSV infection. Besides, significant enrichment in pathways involved in DNA damage and integrity checkpoint was also reported, probably because the cellular DNA damage machinery is activated and exploited by viruses which have evolved ways to manipulate the key regulators of these pathways, in order to promote their own replication.

When analyzing wheezing and asthma phenotypes separately significant differentially methylated positions (FDR *P*-value < 0.05) were detected, showing moderate changes in methylation between the two groups. Among these, the position with the highest difference in Delta Beta (0.148) is the cg18873878 observed within the TSS200 region of the TP73 gene which exhibits and hypermethylation pattern in children RSV+ with recurrent wheezing, in comparison with those with asthma. This Tumor Protein P73 gene encodes a member of the p53 family of transcription factors that are implicated in cellular responses to stress and development. The role of TP73 in airway epithelium is unknown, even if its homolog, TP63, is found to be essential for tracheobronchial epithelium development and homeostasis [308]. However, in 2016, it has been demonstrated that an isoform of *TP73*, called *TAp73* is a central transcriptional regulator governing airway multiciliogenesis; indeed, mice with *Tap73* deficiency developed impaired pulmonary function with chronic inflammation and macrophage infiltration. Through these analyses, it was possible to explain that *TP73* deficiency in mice causes a primary ciliary dyskinesia PCD-like airway phenotype of chronic inflammation and impaired pulmonary function [309].

Another interesting position with a lower methylation pattern in children with asthma is the cg05838113 within the body of the ADAM8 gene, an ADAM Metallopeptidase Domain which has been found to be linked to asthma. It seems that mice with allergic airway inflammation (AAI) show higher levels of ADAM8 in airway epithelium and airway inflammatory cells [310] and it has been demonstrated that ADAM8 mRNA expression is increased in the sputum and endobronchial biopsies of patients with moderate and severe asthma [311].

5.2 AGE-ASSOCIATED DNA METHYLATION AND RELATIONSHIP WITH VACCINATION RESPONSE *

Understanding the phenomenon of impaired response to vaccines, as well as immune system changes that occur in early childhood reflecting the variable susceptibility of children to develop a disease, is a real challenge. The interaction between the epigenome and immune system is not clear, but a growing body of evidence suggests an important role performed by epigenetic mechanisms, in the immune system response to infections as well as in the variability of response observed after vaccination in childhood [312].

In the second part of the present project age-associated DNA methylation changes between the first and second year of life that were largely related to T cell regulation and activation were described. Besides, using detailed antibody measurements following infant pneumococcal conjugate booster vaccination in a clinical vaccine trial, hypomethylation of HLA-DPB1 and hypermethylation of IL6 genes were found to be correlated with a stronger response to pneumococcal vaccination.

For this part of the study, a total of 721 CpGs were found to be significantly differentially methylated between 24 children at 12 and 24 months of age. Several of these within multiple DMRs showed similar methylation patterns suggesting a general epigenetic remodeling of these genetic loci. For example, within the chr6:32118101-32118544 locus, twelve DMPs exhibited a significantly higher methylation pattern at 24 compared with 12 months of age. These DMPs are observed within the body of *PRRT1* and could potentially alter the expression of *PRRT1*, as previously demonstrated for other hypermethylated positions in genes outside of the promoter area [313]. Little is known about the function of *PRRT1* although aberrant methylation of this gene has been related to neurodevelopmental disorders [314] and hepatic tumorigenesis [315]. Furthermore, a total of 118 CpG positions, previously reported to be altered between 3 and 60 months in blood leukocytes [69] were observed. In agreement with this previous work, 57 CpGs exhibiting an increase in methylation between 12 and 24 months of age, and 60 CpGs showing a decreased methylation with age were identified.

Through an enrichment pathway analysis of the age-dependent DMPs, 86 CpGs that were associated with 69 genes involved in several immune system processes were detected. Genes

* Pischedda, S., O'Connor, D., Fairfax, B.P. *et al.* Changes in epigenetic profiles throughout early childhood and their relationship to the response to pneumococcal vaccination. *Clin Epigenet* 13, 29 (2021). https://doi.org/10.1186/s13148-021-01012-w. (see Section 7. ANNEX)

that were hypomethylated at 24 compared with 12 months included *IL12A*, previously shown to be repressed at chromatin level in neonatal mononuclear compared with adult cells [316-318], as well as *BCL10* and *NOD2*, also previously demonstrated to decrease methylation with age [69]. In contrast, CpGs with increased methylation between 12 and 24 months included genes involved in the regulation of innate immune responses such as tripartite motif-containing 15 (*TRIM15*) [319], and the PML Nuclear Body Scaffold (*PML*) [320]. A particular pattern of methylation was observed for lymphocyte activating 3 (*LAG3*), also required for T cell regulation [321]. *LAG3* shows a hypomethylated status at two positions within the chr12:6882855-6883184 region and a hypermethylated profile in an open sea CpG at 24 months, suggesting that diverging methylation patterns may regulate expression of this gene.

A clear separation of predicted biological age between samples taken at the two study time points was detected. Predicted biological age was moderately higher than chronological age, most likely due to the small range of errors for whole blood samples that the age predictor exhibits [63]. However, predicted biological age was not found to be significantly different between high and low vaccine responders at any time point suggesting no differences attributable to "epigenetic age" between the groups.

To assess the relationship between the epigenome at 12 months of age before immunization and the strength of the antigen-specific IgG response to vaccination, a high number of CpGs separating high from low vaccine responders using a less stringent statistical threshold (uncorrected *P*-value <0.01), was identified. With the aim to investigate whether methylation changes of immune genes were modulating the different responsiveness to the vaccine, multiple probes mapping genes involved in immune system processes and with a large difference in methylation between the two groups were investigated. In this way, it was possible to observe two particularly interesting immune genes (*HLA-DPB1* and *IL6*) that contained multiple DMPs differing between high and low vaccine responders.

The four positions settled in the body of the *HLA-DPB1* gene, in the region chr6:33048416-33048814, exhibited a higher level of methylation in low responders when compared to high responders. The human leukocyte antigen system in chromosome 6 contains key genes involved in the regulation of immune response that encodes the major histocompatibility complex (MHC) in humans, essential for adaptive and innate immunity [322]. Among the different classes of genes encoding the MHC, there is one, belonging to the MHC-class II, that involves

the *HLA-DPB1* that encodes the β-chain of the *HLA-DP* [323]. The main role of these molecules is to bind and present on the surface of the CD4+T cells the extracellular viral antigen to stimulate B cells to produce specific antibodies. HLA genes exhibit several polymorphisms, that can regulate the expression of HLA region-encoded accessory proteins and may be crucial in the immune recognition and specific serological response to vaccines. A growing body of evidence demonstrating this was found in several papers collected in the literature: in 2013 it was reviewed the contribution of the HLA gene to host immunity associated with variable cellular immune responses to the Anthrax Vaccine Adsorbed [324]; in 2015, it was suggested that a polymorphism in the *HLA-DPB1* genetic region was associated with humoral immune response variations after rubella vaccination [325]; in 2017 it was revealed the association between amino acid residues in *HLA-DRB1* and the Hepatitis B vaccine response; what is more, recently, in 2019, it was evaluated the association of the class II genes *HLA-DRB1*, *HLA-DPB1*, and *HLA-DQB1* of the HLA with the humoral immune response to the inactivated Japanese encephalitis vaccine [231]. Finally, Dhiman et al., using a GeneChip technology confirmed the differential degree of HLA molecule expression in subjects seropositive and seronegative to the measles vaccine, as it was already reported in 1993 [326] and suggest that the activation of specific HLA molecules stimulate the different antibody response [327]. Following this, the hypothesis was that the hypomethylation pattern observed within the *HLA-DBP1* in the group of high responders could reflect a change in the expression of the gene for this group when compared to the other with a possible explanation of the different response to the vaccine.

Two CpGs in the body of *IL6* showed higher methylation in high compared with low vaccine responders. These CpGs in open sea positions (i.e. isolated in the genome) are considered to be "predicted enhancer elements" [328], therefore possibly impacting *IL6* expression. IL-6 is a proinflammatory cytokine involved in a variety of immune processes, and also plays an important role in mediating innate and adaptive immune responses [329], therefore a differential expression of *IL6* may impact vaccine immune responses. However, in contrast to our finding of hypermethylated (and possibly reduced expression of) *IL6* in high vaccine responders, previous work suggested that *IL6* promotes stimulation of humoral and cellular immune responses [330-333].

6
Conclusions

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

6.1 General Conclusion

The present study provides novel insights into the role of DNA methylation on the regulation of the immune system in early childhood in response to natural infection and vaccination, leading to disease sequelae or poor vaccine response, respectively, because of the immaturity of their immunity. The findings, included in this dissertation, furnish evidence on how epigenetics contributes to the different disease outcomes observed after infections as well as to the differential response to vaccination.

To address these questions, the thesis has been divided into two main parts. In the first part, the role of DNA methylation in respiratory morbidity after RSV infection was evaluated, while the second part was focused in studying how epigenetic mechanisms modulate the immune response to pneumococcal vaccine administration, suggesting an interesting role for DNA methylation in the variable response observed in children after vaccine doses.

The cutting-edge technology used in the present project has allowed the interrogation of the methylation status of more than 850,000 CpG sites located at various regulatory regions throughout the human genome, providing a deeper knowledge on the interaction between epigenetics and immunology and elucidating the molecular basis behind the regulation of different physiological processes involved.

The detection of specific CpGs located in genes potentially involved in the respiratory sequelae after infection, offers the possibility to identify specific epigenetic biomarkers that can be translated to the clinics to predict with more precision the different disease outcomes after RSV infection. In the same way, the discovery of specific CpG sites associated with the immune response to vaccination suggests that epigenetic patterns may influence antibody responses and could be used to evaluate the immune response to PCV vaccination.

Even if additional investigation using larger cohorts would help to further disentangle the role of epigenomics in immunity, this study, globally, provides strong evidence on the role of epigenetics modifications in the regulation of the immune response related to both natural infection and vaccination, in children.

6.2 SPECIFIC CONCLUSIONS

- A large number of DMPs associated with respiratory sequelae developed after RSV infection was detected, with interesting patterns of methylation profiles observed for genes involved in airway inflammation processes.

- Two interesting genes (*EYA3* and *DDX27*), both involved in inflammatory processes, were found to exhibit CpGs differentially methylated between children with normal recovery and children with respiratory sequelae after an RSV infection, that could reveal a possible misregulation of those genes in infants manifesting respiratory morbidity defined as recurrent wheezing and/or asthma.

- Functional analysis considering all CpGs observed between cases with respiratory sequelae and controls with normal recovery after RSV infection, revealed significant enrichment for pathways involved in cell cycle checkpoint, and DNA damage and integrity checkpoint.

- The evaluation of the differences between asthma and wheezing groups suggested an interesting role of *TP73* and *ADAM8* genes, two important genes that play a key role in cellular responses to stress and airway inflammation.

- A great number of genes with differential methylation patterns was revealed when comparing healthy children between 12 and 24 months of age, among which genes of the immune system are found to be the most representative.

- Pathway analysis showed significant enrichment for several immune pathways, with those involved in the regulation of T cell activation and proliferation, in cell-cell adhesion, and in regulation of cytokine production being the most representative.

- Several CpGs associated with the strength of the antigen-specific IgG antibody response to a pneumococcal conjugate booster vaccine were detected including differences in methylation patterns in *HLA-DPB1* and *IL6* genes, whose activation or inhibition is found to stimulate different antibody response to vaccines.

7
ANNEX

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Paper directly related with the present project.

Title: Changes in epigenetic profiles throughout early childhood and their relationship to the response to pneumococcal vaccination.

Authors: Sara Pischedda[1,2,3], Daniel O'Connor[4], Benjamin P. Fairfax[5], Antonio Salas[1,3], Federico Martinon-Torres[1,2], Andrew J. Pollard[4], Johannes Trück[4,6].

Authors Affiliations:

1 Genetics, Vaccines and Infections Research Group (GENVIP), Instituto de Investigación Sanitaria de Santiago, Santiago de Compostela, Spain

2 Translational Pediatrics and Infectious Diseases, Department of Pediatrics, Hospital Clínico Universitario de Santiago de Compostela, Santiago de Compostela, Spain

3 Unidade de Xenética, Instituto de Ciencias Forenses, Facultade de Medicina, Universidade de Santiago de Compostela, and GenPoB Research Group, Instituto de Investigaciones Sanitarias (IDIS), Hospital Clínico Universitario de Santiago (SERGAS), Galicia, Spain.

4 Oxford Vaccine Group, Department of Paediatrics, University of Oxford, and The NIHR Oxford Biomedical Research Centre, Oxford, United Kingdom.

5 MRC-Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK.

6 University Children's Hospital Zurich, Zurich, Switzerland.

Contributions: SP: data analysis and interpretation, manuscript writing; DOC: supervision of data analysis and writing; JT: conception and design of the study; SP, DOC, BPF, AS, FMT, AJP and JT: writing and approval of the final manuscript.

Date of publication: 04/02/2021

DOI: 10.1186/s13148-021-01012-w

ISSN: 1868-7083

Impact factor: 5.028

Quartile: Q1

8
REFERENCES

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

1. Waddington, C.H.J.E., *The epigenotype.* 1942. **1**: p. 18-20.

2. Goldberg, A.D., C.D. Allis, and E. Bernstein, *Epigenetics: a landscape takes shape.* Cell, 2007. **128**(4): p. 635-8.

3. Russo, V.E., R.A. Martienssen, and A.D. Riggs, *Epigenetic mechanisms of gene regulation*. 1996: Cold Spring Harbor Laboratory Press.

4. Kanherkar, R.R., N. Bhatia-Dey, and A.B. Csoka, *Epigenetics across the human lifespan.* Front Cell Dev Biol, 2014. **2**: p. 49.

5. Atlasi, Y. and H.G. Stunnenberg, *The interplay of epigenetic marks during stem cell differentiation and development.* Nat Rev Genet, 2017. **18**(11): p. 643-658.

6. Cheedipudi, S., O. Genolet, and G.D.J.F.i.g. Dobreva, *Epigenetic inheritance of cell fates during embryonic development.* Front Genet, 2014. **5**: p. 19.

7. Kornberg, R.D. and J.O.J.S. Thomas, *Chromatin structure: oligomers of the histones.* Science, 1974. **184**(4139): p. 865-868.

8. Kornberg, R.D., *Structure of chromatin.* Annu Rev Biochem 1977. **46**(1): p. 931-954.

9. Grunstein, M.J.S.A., *Histones as regulators of genes.* Scientific American, 1992. **267**(4): p. 68-75.

10. Lennartsson, A. and K.J.B.e.b.a.-g.s. Ekwall, *Histone modification patterns and epigenetic codes.* Biochimica et biophysica acta -general subjects, 2009. **1790**(9): p. 863-868.

11. Kouzarides, T.J.C., *Chromatin modifications and their function.* Cell, 2007. **128**(4): p. 693-705.

12. Li, B., M. Carey, and J.L.J.C. Workman, *The role of chromatin during transcription.* Cell, 2007. **128**(4): p. 707-719.

13. Greer, E.L. and Y. Shi, *Histone methylation: a dynamic mark in health, disease and inheritance.* Nat Rev Genet, 2012. **13**(5): p. 343-357.

14. Rossetto, D., N. Avvakumov, and J.J.E. Côté, *Histone phosphorylation: a chromatin modification involved in diverse nuclear events.* Epigenetics, 2012. **7**(10): p. 1098-1108.

15. Böhmdorfer, G. and A.T.J.T.i.c.b. Wierzbicki, *Control of chromatin structure by long noncoding RNA.* Trends Cell Biol, 2015. **25**(10): p. 623-632.

16. Xing, Z., et al., *lncRNA directs cooperative epigenetic regulation downstream of chemokine signals.* Cell, 2014. **159**(5): p. 1110-1125.

17. Carthew, R.W.J.S., *A new RNA dimension to genome control.* Science, 2006. **313**(5785): p. 305-306.

18. Goodrich, J.A. and J.F. Kugel, *Non-coding-RNA regulators of RNA polymerase II transcription.* Nat Rev Mol Cell Biol, 2006. **7**(8): p. 612-616.

19. Kim, V.N., *Small RNAs just got bigger: Piwi-interacting RNAs (piRNAs) in mammalian testes.* Genes Dev, 2006. **20**(15): p. 1993-1997.

20. Lewis, B.P., C.B. Burge, and D.P.J.c. Bartel, *Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets.* Cell, 2005. **120**(1): p. 15-20.

21. Li, L.-C., et al., *Small dsRNAs induce transcriptional activation in human cells.* Proc Natl Acad Sci, 2006. **103**(46): p. 17337-17342.

22. Smallwood, S.A. and G. Kelsey, *De novo DNA methylation: a germ cell perspective.* Trends Genet 2012. **28**(1): p. 33-42.

23. Messerschmidt, D.M., B.B. Knowles, and D. Solter, *DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos.* Genes Dev, 2014. **28**(8): p. 812-828.

24. Zeng, Y. and T.J.G. Chen, *DNA methylation reprogramming during mammalian development.* Genes, 2019. **10**(4): p. 257.

25. Okano, M., et al., *DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development.* Cell, 1999. **99**(3): p. 247-257.

26. Kaneda, M., et al., *Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting.* Nature, 2004. **429**(6994): p. 900-903.

27. Watanabe, D., et al., *Stage-and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis.* Mechanisms of development, 2002. **118**(1-2): p. 187-190.

28. Li, E., T.H. Bestor, and R.J.C. Jaenisch, *Targeted mutation of the DNA methyltransferase gene results in embryonic lethality.* Cell, 1992. **69**(6): p. 915-926.

29. Gowher, H., et al., *Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L.* J Biol Chem, 2005. **280**(14): p. 13341-13348.

30. Vertino, P.M., et al., *De novo methylation of CpG island sequences in human fibroblasts overexpressing DNA (cytosine-5-)-methyltransferase.* Mol Cell Biol 1996. **16**(8): p. 4555-4565.

31. Chen, C.-C., K.-Y. Wang, and C.-K.J. Shen, *DNA 5-methylcytosine demethylation activities of the mammalian DNA methyltransferases.* J Biol Chem, 2013. **288**(13): p. 9084-9091.

32. Ehrlich, M. and M.J.E. Lacey, *DNA methylation and differentiation: silencing, upregulation and modulation of gene expression.* Epigenomics, 2013. **5**(5): p. 553-568.

33. Bird, A.P., *CpG-rich islands and the function of DNA methylation.* Nature, 1986. **321**(6067): p. 209-13.

34. Saxonov, S., P. Berg, and D.L. Brutlag, *A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters.* Proc Natl Acad Sci, 2006. **103**(5): p. 1412-1417.

35. Gardiner-Garden, M. and M.J.J.o.m.b. Frommer, *CpG islands in vertebrate genomes.* J Mol Biol, 1987. **196**(2): p. 261-282.

36. Carninci, P., et al., *Genome-wide analysis of mammalian promoter architecture and evolution.* Nat Genet, 2006. **38**(6): p. 626-635.

37. Irizarry, R.A., et al., *Genome-wide methylation analysis of human colon cancer reveals similar hypo-and hypermethylation at conserved tissue-specific CpG island shores.* Nat Genet, 2009. **41**(2): p. 178.

38. Irizarry, R.A., et al., *The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores.* Nat Genet, 2009. **41**(2): p. 178-186.

39. Mohn, F., et al., *Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors.* Mol Cell, 2008. **30**(6): p. 755-766.

40. Bird, A.J.C., *The essentials of DNA methylation.* Cell, 1992. **70**(1): p. 5-8.

41. Shukla, S., et al., *CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing.* Nature, 2011. **479**(7371): p. 74-79.

42. Chodavarapu, R.K., et al., *Relationship between nucleosome positioning and DNA methylation.* Nature, 2010. **466**(7304): p. 388-392.

43. Hu, S., et al., *DNA methylation presents distinct binding sites for human transcription factors.* elife, 2013. **2**: p. e00726.

44. Day, J.J. and J.D.J.N.n. Sweatt, *DNA methylation and memory formation.* 2010. **13**(11): p. 1319-1323.

45. Shen, J., et al., *Exploring genome-wide DNA methylation profiles altered in hepatocellular carcinoma using Infinium HumanMethylation 450 BeadChips.* Epigenetics, 2013. **8**(1): p. 34-43.

46. Marzese, D.M., et al., *Epigenome-wide DNA methylation landscape of melanoma progression to brain metastasis reveals aberrations on homeobox D cluster associated with prognosis.* Hum Mol Genet, 2014. **23**(1): p. 226-238.

47. Doi, A., et al., *Differential methylation of tissue-and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts.* Nat Genet, 2009. **41**(12): p. 1350-1353.

48. Stadler, M.B., et al., *DNA-binding factors shape the mouse methylome at distal regulatory regions.* Nature, 2011. **480**(7378): p. 490-495.

49. Carone, B.R., et al., *Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals.* Cell, 2010. **143**(7): p. 1084-1096.

50. Sandovici, I., et al., *Maternal diet and aging alter the epigenetic control of a promoter–enhancer interaction at the Hnf4a gene in rat pancreatic islets.* Proc Natl Acad Sci, 2011. **108**(13): p. 5449-5454.

51. Aran, D., S. Sabato, and A. Hellman, *DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes.* Genome Biol, 2013. **14**(3): p. 1-14.

52. Aran, D. and A.J.C. Hellman, *DNA methylation of transcriptional enhancers and cancer predisposition.* Cell, 2013. **154**(1): p. 11-13.

53. Hellman, A. and A.J.s. Chess, *Gene body-specific methylation on the active X chromosome.* Science, 2007. **315**(5815): p. 1141-1143.

54. Ball, M.P., et al., *Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells.* Nat Biotechnol, 2009. **27**(4): p. 361-368.

55. Aran, D., et al., *Replication timing-related and gene body-specific methylation of active human genes.* Hum Mol Genet, 2011. **20**(4): p. 670-680.

56. Jones, P.A., *The DNA methylation paradox.* Trends Genet, 1999. **15**(1): p. 34-37.

57. Jjingo, D., et al., *On the presence and role of human gene-body DNA methylation.* Oncotarget, 2012. **3**(4): p. 462.

58. Fraga, M.F. and M. Esteller, *Epigenetics and aging: the targets and the marks.* Trends Genet, 2007. **23**(8): p. 413-8.

59. Sedivy, J.M., G. Banumathy, and P.D.J.E.c.r. Adams, *Aging by epigenetics—a consequence of chromatin damage?* Experimental cell research, 2008. **314**(9): p. 1909-1917.

60. Bjornsson, H.T., et al., *Intra-individual change over time in DNA methylation with familial clustering.* JAMA, 2008. **299**(24): p. 2877-2883.

61. Christensen, B.C., et al., *Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context.* PLoS Genet, 2009. **5**(8): p. e1000602.

62. Teschendorff, A.E., et al., *Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer.* Genome Res, 2010. **20**(4): p. 440-446.

63. Horvath, S., *DNA methylation age of human tissues and cell types.* Genome Biol, 2013. **14**(10): p. 3156.

64. Jylhävä, J., N.L. Pedersen, and S.J.E. Hägg, *Biological age predictors.* EBioMedicine, 2017. **21**: p. 29-36.

65. Horvath, S., et al., *Accelerated epigenetic aging in Down syndrome.* Aging Cell, 2015. **14**(3): p. 491-5.

66. Horvath, S. and A.J. Levine, *HIV-1 infection accelerates age according to the epigenetic clock.* J Infect Dis, 2015. **212**(10): p. 1563-1573.

67. McEwen, L.M., et al., *Systematic evaluation of DNA methylation age estimation with common preprocessing methods and the Infinium MethylationEPIC BeadChip array.* Clin Epigenetics, 2018. **10**(1): p. 123.

68. Martino, D.J., et al., *Evidence for age-related and individual-specific changes in DNA methylation profile of mononuclear cells during early immune development in humans.* Epigenetics, 2011. **6**(9): p. 1085-94.

69. Acevedo, N., et al., *Age-associated DNA methylation changes in immune genes, histone modifiers and chromatin remodeling factors within 5 years after birth in human blood leukocytes.* Clin Epigenetics, 2015. **7**: p. 34.

70. Pérez, R.F., et al., *Longitudinal genome-wide DNA methylation analysis uncovers persistent early-life DNA methylation changes.* J Transl Med, 2019. **17**(1): p. 1-16.

71. Heard, E. and R.A.J.C. Martienssen, *Transgenerational epigenetic inheritance: myths and mechanisms.* Cell, 2014. **157**(1): p. 95-109.

72. Miska, E.A. and A.C.J.S. Ferguson-Smith, *Transgenerational inheritance: models and mechanisms of non–DNA sequence–based inheritance.* Science, 2016. **354**(6308): p. 59-63.

73. van Otterdijk, S.D. and K.B.J.T.F.J. Michels, *Transgenerational epigenetic inheritance in mammals: how good is the evidence?* The FASEB Journal, 2016. **30**(7): p. 2457-2465.

74. McGrath, J. and D.J.C. Solter, *Completion of mouse embryogenesis requires both the maternal and paternal genomes.* Cell, 1984. **37**(1): p. 179-183.

75. Barton, S.C., M. Surani, and M.J.N. Norris, *Role of paternal and maternal genomes in mouse development.* Nature, 1984. **311**(5984): p. 374-376.

76. Thomsen, P.D.J.A.V.S., *Genomic imprinting–an epigenetic regulation of fetal development and loss.* Acta Veterinaria Scandinavica, 2007. **49**(S1): p. S7.

77. Reik, W. and J. Walter, *Genomic imprinting: parental influence on the genome.* Nat Rev Genet, 2001. **2**(1): p. 21-32.

78. Chotalia, M., et al., *Transcription is required for establishment of germline methylation marks at imprinted genes.* Genes Dev, 2009. **23**(1): p. 105-117.

79. Green, K., et al., *A developmental window of opportunity for imprinted gene silencing mediated by DNA methylation and the Kcnq1ot1 noncoding RNA.* Mamm Genome, 2007. **18**(1): p. 32-42.

80. Wutz, A., T.P. Rasmussen, and R. Jaenisch, *Chromosomal silencing and localization are mediated by different domains of Xist RNA.* Nat Genet, 2002. **30**(2): p. 167-174.

81. Ciccone, D.N., et al., *KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints.* Nature, 2009. **461**(7262): p. 415-418.

82. Wang, J., et al., *The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation.* Nat Genet, 2009. **41**(1): p. 125-129.

83. Rakyan, V.K., et al., *Transgenerational inheritance of epigenetic states at the murine AxinFu allele occurs after maternal and paternal transmission.* Proc Natl Acad Sci, 2003. **100**(5): p. 2538-2543.

84. Kaminsky, Z.A., et al., *DNA methylation profiles in monozygotic and dizygotic twins.* Nat Genet, 2009. **41**(2): p. 240-245.

85. Esteller, M., et al., *Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors.* J Natl Cancer Inst, 2000. **92**(7): p. 564-569.

86. Ellinger, J., et al., *DNA hypermethylation in papillary renal cell carcinoma.* BJU international, 2011. **107**(4): p. 664-669.

87. Pogribny, I. and S.J.J.C.l. James, *Reduction of p53 gene expression in human primary hepatocellular carcinoma is associated with promoter region methylation without coding region mutation.* Cancer letters, 2002. **176**(2): p. 169-174.

88. Hibi, K., et al., *Demethylation of the CDH3 gene is frequently detected in advanced colorectal cancer.* Anticancer research, 2009. **29**(6): p. 2215-2217.

89. Singh, U., et al., *Epigenetic regulation of human retinoblastoma.* Tumor Biology 2016. **37**(11): p. 14427-14441.

90. Ohta, T., et al., *Imprinting-mutation mechanisms in Prader-Willi syndrome.* Am J Hum Genet, 1999. **64**(2): p. 397-413.

91. Lalande, M., M.J.C. Calciano, and M.L. Sciences, *Molecular epigenetics of Angelman syndrome.* Cell Mol Life Sci, 2007. **64**(7-8): p. 947.

92. Garber, K.B., J. Visootsak, and S.T. Warren, *Fragile X syndrome.* Eur J Hum Genet 2008. **16**(6): p. 666-672.

93. Schenkel, L.C., et al., *Identification of epigenetic signature associated with alpha thalassemia/mental retardation X-linked syndrome.* 2017. **10**(1): p. 1-11.

94. Ling, C., et al., *Epigenetic regulation of PPARGC1A in human type 2 diabetic islets and effect on insulin secretion.* Diabetologia, 2008. **51**(4): p. 615-622.

95. Dick, K.J., et al., *DNA methylation and body-mass index: a genome-wide analysis.* The Lancet, 2014. **383**(9933): p. 1990-1998.

96. Abdolmaleky, H.M., et al., *Hypermethylation of the reelin (RELN) promoter in the brain of schizophrenic patients: a preliminary report.* Am J Hum Genet B Neuropsychiatr Genet, 2005. **134**(1): p. 60-66.

97. Williams-Karnesky, R.L., et al., *Epigenetic changes induced by adenosine augmentation therapy prevent epileptogenesis.* British journal of cancer, 2013. **123**(8): p. 3552-3563.

98. Pritchard, C., D. Cox, and R. Myers, *Methylation at the Huntington disease-linked D4S95 locus.* Am J Hum Genet, 1989. **45**(2): p. 335.

99. Rothstein, J.D., et al., *Selective loss of glial glutamate transporter GLT-1 in amyotrophic lateral sclerosis.* Annals of Neurology: Official Journal of the American Neurological Association the Child Neurology Society, 1995. **38**(1): p. 73-84.

100. Mastroeni, D., et al., *Epigenetic differences in cortical neurons from a pair of monozygotic twins discordant for Alzheimer's disease.* PLoS One, 2009. **4**(8): p. e6617.

101. Jowaed, A., et al., *Methylation regulates alpha-synuclein expression and is decreased in Parkinson's disease patients' brains.* J Neurosci, 2010. **30**(18): p. 6355-6359.

102. Richardson, B., et al., *Evidence for impaired T cell DNA methylation in systemic lupus erythematosus and rheumatoid arthritis.* Arthritis Rheumatism: Official Journal of the American College of Rheumatology, 1990. **33**(11): p. 1665-1673.

103. Lei, W., et al., *Abnormal DNA methylation in CD4+ T cells from patients with systemic lupus erythematosus, systemic sclerosis, and dermatomyositis.* Scandinavian journal of rheumatology, 2009. **38**(5): p. 369-374.

104. Yin, H., et al., *Hypomethylation and overexpression of CD70 (TNFSF7) in CD4+ T cells of patients with primary Sjögren's syndrome.* J Dermatol Sci, 2010. **59**(3): p. 198-203.

105. Arakawa, Y., et al., *Association of polymorphisms in DNMT1, DNMT3A, DNMT3B, MTHFR and MTRR genes with global DNA methylation levels and prognosis of autoimmune thyroid disease.* Epigenomics, 2012. **170**(2): p. 194-201.

106. Ventham, N.T., et al., *Beyond gene discovery in inflammatory bowel disease: the emerging role of epigenetics.* Gastroenterology, 2013. **145**(2): p. 293-308.

107. Li, Y., et al., *Abnormal DNA methylation in CD4+ T cells from people with latent autoimmune diabetes in adults.* Diabetes research clinical practice, 2011. **94**(2): p. 242-248.

108. Feinberg, A.P. and B.J.N. Vogelstein, *Hypomethylation distinguishes genes of some human cancers from their normal counterparts.* 1983. **301**(5895): p. 89-92.

109. Jones, P.A. and S.B. Baylin, *The fundamental role of epigenetic events in cancer.* Nat Rev Genet, 2002. **3**(6): p. 415-428.

110. Egger, G., et al., *Epigenetics in human disease and prospects for epigenetic therapy.* Nature, 2004. **429**(6990): p. 457-463.

111. Feinberg, A.P., R. Ohlsson, and S. Henikoff, *The epigenetic progenitor origin of human cancer.* Nat Rev Genet, 2006. **7**(1): p. 21-33.

112. Yoo, C.B. and P.A.J.N.r.D.d. Jones, *Epigenetic therapy of cancer: past, present and future.* Nat Rev Drug discovery, 2006. **5**(1): p. 37-50.

113. Qureshi, I.A., M.F.J.C.n. Mehler, and n. reports, *Advances in epigenetics and epigenomics for neurodegenerative diseases.* Curr Neurol Neurosci Rep, 2011. **11**(5): p. 464.

114. Telese, F., et al., *"Seq-ing" insights into the epigenetics of neuronal gene regulation.* Neuron, 2013. **77**(4): p. 606-623.

115. Urdinguio, R.G., J.V. Sanchez-Mut, and M.J.T.L.N. Esteller, *Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies.* The Lancet Neurol 2009. **8**(11): p. 1056-1072.

116. Lunnon, K., et al., *Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease.* Nat Rev Neurosci, 2014. **17**(9): p. 1164-1170.

117. De Jager, P.L., et al., *Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci.* Nat Rev Neurosci, 2014. **17**(9): p. 1156-1163.

118. Masliah, E., et al., *Distinctive patterns of DNA methylation associated with Parkinson disease: identification of concordant epigenetic changes in brain and peripheral blood leukocytes.* Epigenetics, 2013. **8**(10): p. 1030-1038.

119. Heman-Ackah, S.M., et al., *RISC in PD: the impact of microRNAs in Parkinson's disease cellular and molecular pathogenesis.* Front Mol Neurosci, 2013. **6**: p. 40.

120. Valor, L.M. and D.J.N. Guiretti, *What's wrong with epigenetics in Huntington's disease?* Neuropharmacology, 2014. **80**: p. 103-114.

121. Huynh, J.L. and P.J.T.L.N. Casaccia, *Epigenetic mechanisms in multiple sclerosis: implications for pathogenesis and treatment.* The Lancet Neurol, 2013. **12**(2): p. 195-206.

122. Milagro, F.I., et al., *A dual epigenomic approach for the search of obesity biomarkers: DNA methylation in relation to diet-induced weight loss.* The FASEB Journal 2011. **25**(4): p. 1378-1389.

123. Nile, C.J., et al., *Methylation status of a single CpG site in the IL6 promoter is related to IL6 messenger RNA levels and rheumatoid arthritis.* Arthritis Rheumatism 2008. **58**(9): p. 2686-2693.

124. Zhang, Z., et al., *Global H4 acetylation analysis by ChIP-chip in systemic lupus erythematosus monocytes.* Genes, 2010. **11**(2): p. 124-133.

125. Deans, C. and K.A.J.G. Maggert, *What do you mean,"epigenetic"?* Genetics, 2015. **199**(4): p. 887-896.

126. Berger, S.L., et al., *An operational definition of epigenetics.* Genes Dev, 2009. **23**(7): p. 781-783.

127. Jablonka, E. and M.J.J.A.o.t.N.Y.A.o.S. Lamb, *The changing concept of epigenetics.* Annals of the New York Academy of Sciences, 2002. **981**(1): p. 82-96.

128. MacArthur, D., *Why do genome-wide scans fail? Genetic Future. 2008*.

129. Rakyan, V.K., et al., *Epigenome-wide association studies for common human diseases.* Nat Rev Genet, 2011. **12**(8): p. 529-41.

130. Lappalainen, T. and J.M. Greally, *Associating cellular epigenetic models with human phenotypes.* Nat Rev Genet, 2017. **18**(7): p. 441.

131. Herceg, Z. and P.J.M.o. Hainaut, *Genetic and epigenetic alterations as biomarkers for cancer detection, diagnosis and prognosis.* Mol Oncology, 2007. **1**(1): p. 26-41.

132. Baylin, S. and T.H.J.C.c. Bestor, *Altered methylation patterns in cancer cell genomes: cause or consequence?* Cancer cell, 2002. **1**(4): p. 299-305.

133. Delves, P.J. and I.M. Roitt, *The immune system.* N Engl J Med, 2000. **343**(1): p. 37-49.

134. Parkin, J. and B.J.T.L. Cohen, *An overview of the immune system.* The Lancet, 2001. **357**(9270): p. 1777-1789.

135. Akashi, K., et al., *A clonogenic common myeloid progenitor that gives rise to all myeloid lineages.* Nature, 2000. **404**(6774): p. 193-197.

136. CARUSO, C., et al., *Major histocompatibility complex regulation of cytokine production.* J Interferon Cytokine Res, 1996. **16**(12): p. 983-988.

137. Neefjes, J., et al., *Towards a systems understanding of MHC class I and MHC class II antigen presentation.* Nat Rev Immunol, 2011. **11**(12): p. 823-836.

138. Akkaya, M., K. Kwak, and S.K. Pierce, *B cell memory: building two walls of protection against pathogens.* Nat Rev Immunol, 2020. **20**(4): p. 229-238.

139. Netea, M.G., et al., *Trained immunity: a program of innate immune memory in health and disease.* Science, 2016. **352**(6284).

140. Bröske, A.-M., et al., *DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction.* Nat Genet, 2009. **41**(11): p. 1207-1215.

141. Moran-Crusio, K., et al., *Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation.* Cancer cell, 2011. **20**(1): p. 11-24.

142. Ivashkiv, L.B.J.T.i.i., *Epigenetic regulation of macrophage polarization and function.* Trends Immunol, 2013. **34**(5): p. 216-223.

143. Schoenborn, J.R., et al., *Comprehensive epigenetic profiling identifies multiple distal regulatory elements directing transcription of the gene encoding interferon-γ.* Nat Immunol, 2007. **8**(7): p. 732-742.

144. Trowbridge, J.J., et al., *DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells.* Cell Stem Cell, 2009. **5**(4): p. 442-449.

145. Ladle, B.H., et al., *De novo DNA methylation by DNA methyltransferase 3a controls early effector CD8+ T-cell fate decisions following activation.* Proc Natl Acad Sci, 2016. **113**(38): p. 10631-10636.

146. Saeed, S., et al., *Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity.* Science, 2014. **345**(6204).

147. Chen, N. and E.H.J.T. Field, *Enhanced type 2 and diminished type 1 cytokines in neonatal tolerance.* Transplantation, 1995. **59**(7): p. 933-941.

148. Klein Klouwenberg, P., L.J.C. Bont, and D. Immunology, *Neonatal and infantile immune responses to encapsulated bacteria and conjugate vaccines.* Clinical Developmental Immunology, 2008.

149. Kollmann, T.R., et al., *Protecting the Newborn and Young Infant from Infectious Diseases: Lessons from Immune Ontogeny.* Immunity, 2017. **46**(3): p. 350-363.

150. Wang, L., et al., *Expression of Nodal on Bronchial Epithelial Cells Influenced by Lung Microbes Through DNA Methylation Modulates the Differentiation of T-Helper Cells.* Cell Physiol Biochem, 2015. **37**(5): p. 2012-22.

151. Syn, G., J.M. Blackwell, and S.E. Jamieson, *Epigenetics in Infectious Diseases*, in *Epigenetic Biomarkers and Diagnostics*, 2016. p. 377-400.

152. Hamon, M.A. and P. Cossart, *Histone modifications and chromatin remodeling during bacterial infections.* Cell Host Microbe, 2008. **4**(2): p. 100-9.

153. Balakrishnan, L. and B. Milavetz, *Epigenetic Regulation of Viral Biological Processes.* Viruses, 2017. **9**(11).

154. Chen, H.S., F. Lu, and P.M. Lieberman, *Epigenetic regulation of EBV and KSHV latency.* Curr Opin Virol, 2013. **3**(3): p. 251-9.

155. Paschos, K. and M.J. Allday, *Epigenetic reprogramming of host genes in viral and microbial pathogenesis.* Trends Microbiol, 2010. **18**(10): p. 439-47.

156. Scott, R.S., *Epstein-Barr virus: a master epigenetic manipulator.* Curr Opin Virol, 2017. **26**: p. 74-80.

157. Hammerschmidt, W., *The epigenetic life cycle of Epstein–Barr virus*, in *Epstein Barr Virus Volume 1*. 2015, Springer. p. 103-117.

158. Kalla, M., C. Göbel, and W. Hammerschmidt, *The lytic phase of Epstein-Barr virus requires a viral genome with 5-methylcytosine residues in CpG sites.* J Virol, 2012. **86**(1): p. 447-458.

159. Ferrari, R., et al., *Adenovirus small E1A employs the lysine acetylases p300/CBP and tumor suppressor Rb to repress select host genes and promote productive virus infection.* Cell host, 2014. **16**(5): p. 663-676.

160. Durzynska, J., K. Lesniewicz, and E. Poreba, *Human papillomaviruses in epigenetic regulations.* Mutat Res Rev Mutat Res, 2017. **772**: p. 36-50.

161. Wei, X., et al., *miR-101 is down-regulated by the hepatitis B virus x protein and induces aberrant DNA methylation by targeting DNA methyltransferase 3A.* Cell Signal 2013. **25**(2): p. 439-446.

162. Schafer, A. and R.S. Baric, *Epigenetic Landscape during Coronavirus Infection.* Pathogens, 2017. **6**(1).

163. Marazzi, I., et al., *Suppression of the antiviral response by an influenza histone mimic.* Nature, 2012. **483**(7390): p. 428-433.

164. Menachery, V.D., et al., *Attenuation and restoration of severe acute respiratory syndrome coronavirus mutant lacking 2'-o-methyltransferase activity.* J Virol, 2014. **88**(8): p. 4251-64.

165. Corley, M.J., et al., *Genome-wide DNA methylation profiling of peripheral blood reveals an epigenetic signature associated with severe COVID-19.* 2021.

166. Morris, J., et al., *Recovery of cytopathogenic agent from chimpanzees with goryza.* Proc Soc Exp Biol, 1956. **92**(3): p. 544-549.

167. Liu, L., et al., *Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals.* The Lancet, 2016. **388**(10063): p. 3027-3035.

168. Shi, T., et al., *Risk factors for respiratory syncytial virus associated with acute lower respiratory infection in children under five years: Systematic review and meta–analysis.* J Glob Health 2015. **5**(2).

169. Shi, T., et al., *Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in young children in 2015: a systematic review and modelling study*. The Lancet, 2017. **390**(10098): p. 946-958.

170. Obando-Pacheco, P., et al., *Respiratory Syncytial Virus Seasonality: A Global Overview.* J Infect Dis, 2018. **217**(9): p. 1356-1364.

171. McLellan, J.S., W.C. Ray, and M.E. Peeples, *Structure and function of respiratory syncytial virus surface glycoproteins*, in *Challenges and opportunities for respiratory syncytial virus vaccines*. Springer, 2013. p. 83-104.

172. Johnson, P.R., et al., *The G glycoprotein of human respiratory syncytial viruses of subgroups A and B: extensive sequence divergence between antigenically related proteins.* Proc Natl Acad Sci, 1987. **84**(16): p. 5625-5629.

173. Bukreyev, A., et al., *The secreted form of respiratory syncytial virus G glycoprotein helps the virus evade antibody-mediated restriction of replication by acting as an antigen decoy and through effects on Fc receptor-bearing leukocytes.* J Virol, 2008. **82**(24): p. 12191-12204.

174. Lamb, R.A.J.V., *Paramyxovirus fusion: a hypothesis for changes.* Virology, 1993. **197**(1): p. 1-11.

175. Gower, T.L., et al., *RhoA signaling is required for respiratory syncytial virus-induced syncytium formation and filamentous virion morphology.* J Virol, 2005. **79**(9): p. 5326-5336.

176. Heminway, B., et al., *Analysis of respiratory syncytial virus F, G, and SH proteins in cell fusion.* Virology, 1994. **200**(2): p. 801-805.

177. McNamara, P.S. and R.L.J.B.m.b. Smyth, *The pathogenesis of respiratory syncytial virus disease in childhood.* British medical bulletin, 2002. **61**(1): p. 13-28.

178. Blanken, M.O., et al., *Respiratory syncytial virus and recurrent wheeze in healthy preterm infants.* N Engl J Med, 2013. **368**(19): p. 1791-9.

179. Kurt-Jones, E.A., et al., *Pattern recognition receptors TLR4 and CD14 mediate response to respiratory syncytial virus.* Nat Immunol, 2000. **1**(5): p. 398-401.

180. Everard, M.L., et al., *Analysis of cells obtained by bronchial lavage of infants with respiratory syncytial virus infection.* Arch Dis Child, 1994. **71**(5): p. 428-432.

181. Wang, S.Z. and K.D.J.R. Forsyth, *The interaction of neutrophils with respiratory epithelial cells in viral infection.* Respirology, 2000. **5**(1): p. 1-9.

182. Sebina, I. and S.J.V. Phipps, *The Contribution of Neutrophils to the Pathogenesis of RSV Bronchiolitis.* Viruses, 2020. **12**(8): p. 808.

183. Stokes, K.L., et al., *The respiratory syncytial virus fusion protein and neutrophils mediate the airway mucin response to pathogenic respiratory syncytial virus infection.* J Virol, 2013. **87**(18): p. 10070-10082.

184. Kimpen, J.L., *Respiratory syncytial virus and asthma: the role of monocytes.* Am J Respir Crit Care Med, 2001. **163**(supplement_1): p. S7-S9.

185. Domachowske, J.B. and H.F. Rosenberg, *Respiratory syncytial virus infection: immune response, immunopathogenesis, and treatment.* Clin Microbiol Rev, 1999. **12**(2): p. 298-309.

186. Rosenberg, H.F., K.D. Dyer, and J.B.J.A.r. Domachowske, *Respiratory viruses and eosinophils: exploring the connections.* Antiviral research, 2009. **83**(1): p. 1-9.

187. Hussell, T. and P.J.J.o.G.V. Openshaw, *Intracellular IFN-gamma expression in natural killer cells precedes lung CD8+ T cell recruitment during respiratory syncytial virus infection.* J Gen Virol, 1998. **79**(11): p. 2593-2601.

188. Li, F., et al., *Natural killer cells are involved in acute lung immune injury caused by respiratory syncytial virus infection.* J Virol, 2012. **86**(4): p. 2251-2258.

189. Bhat, R., M.A. Farrag, and F.N.J.I.R.o.I. Almajhdi, *Double-edged role of natural killer cells during RSV infection.* Int Rev Immunol, 2020. **39**(5): p. 233-244.

190. van Erp, E.A., et al., *Respiratory syncytial virus infects primary neonatal and adult natural killer cells and affects their antiviral effector function.* J Infect Dis, 2019. **219**(5): p. 723-733.

191. Steff, A.-M., et al., *Pre-fusion RSV F strongly boosts pre-fusion specific neutralizing responses in cattle pre-exposed to bovine RSV.* Nat Commun, 2017. **8**(1): p. 1-10.

192. McLellan, J.S., et al., *Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody.* Science, 2013. **340**(6136): p. 1113-1117.

193. Collins, P.L. and B.S. Graham, *Viral and host factors in human respiratory syncytial virus pathogenesis.* J Virol, 2008. **82**(5): p. 2040-55.

194. McNamara, P.S., et al., *Interleukin 9 production in the lungs of infants with severe respiratory syncytial virus bronchiolitis.* The Lancet, 2004. **363**(9414): p. 1031-1037.

195. Barry, W., et al., *Ribavirin aerosol for acute bronchiolitis.* Arch Dis Child, 1986. **61**(6): p. 593-597.

196. Halsey, N.A., et al., *Reassessment of the indications for ribavirin therapy in respiratory syncytial virus infections.* Pediatrics, 1996. **97**(1): p. 137-140.

197. Olchanski, N., et al., *Palivizumab Prophylaxis for Respiratory Syncytial Virus: Examining the Evidence Around Value.* Open Forum Infect Dis, 2018. **5**(3): p. ofy031.

198. Higgins, D., C. Trujillo, and C.J.V. Keech, *Advances in RSV vaccine research and development–A global agenda.* Vaccine, 2016. **34**(26): p. 2870-2875.

199. Mammas, I.N., et al., *Update on current views and advances on RSV infection.* Int J Mol Med, 2020. **46**(2): p. 509-520.

200. Sigurs, N., et al., *Respiratory syncytial virus bronchiolitis in infancy is an important risk factor for asthma and allergy at age 7.* Am J Respir Crit Care Med, 2000. **161**(5): p. 1501-1507.

201. Sigurs, N., et al., *Severe respiratory syncytial virus bronchiolitis in infancy and asthma and allergy at age 13.* Am J Respir Crit Care Med, 2005. **171**(2): p. 137-41.

202. Holtzman, M.J., et al., *"Hit-and-run" effects of paramyxoviruses as a basis for chronic respiratory disease.* Pediatr Infect Dis J, 2004. **23**(11 Suppl): p. S235-45.

203. Taussig, L.M., et al., *Tucson Children's Respiratory Study: 1980 to present.* J Allergy Clin Immunol, 2003. **111**(4): p. 661-75; quiz 676.

204. James, A.L. and S. Wenzel, *Clinical relevance of airway remodelling in airway diseases.* Eur Respir J, 2007. **30**(1): p. 134-55.

205. Clough, J.B., et al., *Can we predict which wheezy infants will continue to wheeze?* Am J Respir Crit Care Med, 1999. **160**(5): p. 1473-1480.

206. Everard, M.L., *The relationship between respiratory syncytial virus infections and the development of wheezing and asthma in children.* Curr Opin Allergy Clin Immunol, 2006. **6**(1): p. 56-61.

207. Martinez, F.D.J.P., *Development of wheezing disorders and asthma in preschool children.* Pediatrics, 2002. **109**(Supplement E1): p. 362-367.

208. Shi, T., et al., *Association Between Respiratory Syncytial Virus-Associated Acute Lower Respiratory Infection in Early Life and Recurrent Wheeze and Asthma in Later Childhood.* J Infect Dis, 2020. **222**(Supplement_7): p. S628-S633.

209. Busse, W.W., S.P. Banks-Schlegel, and G.L. Larsen, *Effects of growth and development on lung function: models for study of childhood asthma.* Am J Respir Crit Care Med, 1997. **156**(1): p. 314-319.

210. Garcia, C.G., et al., *Risk factors in children hospitalized with RSV bronchiolitis versus non-RSV bronchiolitis.* Pediatrics, 2010. **126**(6): p. e1453-60.

211. Sherrill, D.L., et al., *Total serum IgE and its association with asthma symptoms and allergic sensitization among children.* J Allergy Clin Immunol, 1999. **104**(1): p. 28-36.

212. Puthothu, B., et al., *Impact of IL8 and IL8-receptor alpha polymorphisms on the genetics of bronchial asthma and severe RSV infections.* Clinical Molecular Allergy, 2006. **4**(1): p. 1-6.

213. Kabesch, M., et al., *IL-4/IL-13 pathway genetics strongly influence serum IgE levels and childhood asthma.* J Allergy Clin Immunol, 2006. **117**(2): p. 269-274.

214. Wu, P., et al., *Evidence of a causal role of winter virus infection during infancy in early childhood asthma.* Am J Respir Crit Care Med, 2008. **178**(11): p. 1123-1129.

215. Thomsen, S.F., et al., *Exploring the association between severe respiratory syncytial virus infection and asthma: a registry-based twin study.* Am J Respir Crit Care Med, 2009. **179**(12): p. 1091-1097.

216. Lee, G.R., et al., *T helper cell differentiation: regulation by cis elements and epigenetics.* Immunity, 2006. **24**(4): p. 369-79.

217. Elgizouli, M., et al., *Reduced PRF1 enhancer methylation in children with a history of severe RSV bronchiolitis in infancy: an association study.* BMC Pediatr, 2017. **17**(1): p. 65.

218. Elgizouli, M., et al., *Cord blood PRF1 methylation patterns and risk of lower respiratory tract infections in infants: findings from the Ulm Birth Cohort.* Medicine (Baltimore), 2015. **94**(1): p. e332.

219. Ptaschinski, C., et al., *RSV-Induced H3K4 Demethylase KDM5B Leads to Regulation of Dendritic Cell-Derived Innate Cytokines and Exacerbates Pathogenesis In Vivo.* PLoS Pathog, 2015. **11**(6): p. e1004978.

220. Nagata, D.E., et al., *Epigenetic control of Foxp3 by SMYD3 H3K4 histone methyltransferase controls iTreg development and regulates pathogenic T-cell responses during pulmonary viral infection.* Mucosal Immunol, 2015. **8**(5): p. 1131-43.

221. Blanchard-Rohner, G. and A.J.J.E.r.o.v. Pollard, *Long-term protection after immunization with protein–polysaccharide conjugate vaccines in infancy.* Expert Rev Vaccines, 2011. **10**(5): p. 673-684.

222. Westerink, M.J., H.W. Schroeder Jr, and M.H. Nahm, *Immune responses to pneumococcal vaccines in children and adults: rationale for age-specific vaccination.* Aging and disease, 2012. **3**(1): p. 51.

223. Zimmermann, P. and N. Curtis, *Factors that influence the immune response to vaccination.* Clin Microbiol Rev, 2019. **32**(2).

224. Mentzer, A.J., et al., *Searching for the human genetic factors standing in the way of universally effective vaccines.* Philos Trans R Soc Lond B Biol Sci, 2015. **370**(1671).

225. O'Connor, D., et al., *Common Genetic Variations Associated with the Persistence of Immunity following Childhood Immunization.* Cell Rep, 2019. **27**(11): p. 3241-3253 e4.

226. Höhler, T., et al., *Differential genetic determination of immune responsiveness to hepatitis B surface antigen and to hepatitis A virus: a vaccination study in twins.* The Lancet, 2002. **360**(9338): p. 991-995.

227. Tan, P.-L., et al., *Twin studies of immunogenicity—determining the genetic contribution to vaccine failure.* Vaccine, 2001. **19**(17-19): p. 2434-2439.

228. Kennedy, R.B., et al., *Multigenic control of measles vaccine immunity mediated by polymorphisms in measles receptor, innate pathway, and cytokine genes.* Vaccine, 2012. **30**(12): p. 2159-2167.

229. Crux, N.B. and S. Elahi, *Human Leukocyte Antigen (HLA) and Immune Regulation: How Do Classical and Non-Classical HLA Alleles Modulate Immune Response to Human Immunodeficiency Virus and Hepatitis C Virus Infections?* Front Immunol, 2017. **8**: p. 832.

230. Sakai, A., et al., *Identification of amino acids in antigen-binding site of class II HLA proteins independently associated with hepatitis B vaccine response.* Vaccine, 2017. **35**(4): p. 703-710.

231. Yao, Y., et al., *HLA Class II Genes HLA-DRB1, HLA-DPB1, and HLA-DQB1 Are Associated With the Antibody Response to Inactivated Japanese Encephalitis Vaccine.* Front Immunol, 2019. **10**: p. 428.

232. Jasiulionis, M.G., *Abnormal Epigenetic Regulation of Immune System during Aging.* Front Immunol, 2018. **9**: p. 197.

233. Zimmermann, M.T., et al., *System-wide associations between DNA-methylation, gene expression, and humoral immune response to influenza vaccination.* PLoS One, 2016. **11**(3): p. e0152034.

234. Gensous, N., et al., *Responders and non-responders to influenza vaccination: A DNA methylation approach on blood cells.* Exp Gerontol, 2018. **105**: p. 94-100.

235. Lu, Y., et al., *Exploring the molecular causes of hepatitis B virus vaccination response: an approach with epigenomic and transcriptomic data.* BMC Medical Genom, 2014. **7**(1): p. 1-8.

236. O'Brien, K.L., et al., *Burden of disease caused by Streptococcus pneumoniae in children younger than 5 years: global estimates.* The Lancet, 2009. **374**(9693): p. 893-902.

237. Organization, W.H., *Pneumococcal conjugate vaccine for childhood immunization—WHO position paper.* Weekly Epidemiological Record= Relevé épidémiologique hebdomadaire, 2007. **82**(12): p. 93-104.

238. Trück, J. and A.J.J.E.h.d. Pollard, *Challenges in immunisation against bacterial infection in children.* Early human development, 2010. **86**(11): p. 695-701.

239. Kelly, D.F., E.R. Moxon, and A.J.J.I. Pollard, *Haemophilus influenzae type b conjugate vaccines.* Immunology, 2004. **113**(2): p. 163-174.

240. Snape, M.D., et al., *Immunogenicity and reactogenicity of a 13-valent-pneumococcal conjugate vaccine administered at 2, 4, and 12 months of age: a double-blind randomized active-controlled trial.* Pediatr Pulmonol, 2010. **29**(12): p. e80-e90.

241. Truck, J., et al., *The Antibody Response Following a Booster With Either a 10- or 13-valent Pneumococcal Conjugate Vaccine in Toddlers Primed With a 13-valent Pneumococcal Conjugate Vaccine in Early Infancy.* Pediatr Infect Dis J, 2016. **35**(7): p. 787-93.

242. Lipsitch, M., et al., *Serotype-specific immune responses to pneumococcal conjugate vaccine among children are significantly correlated by individual: Analysis of randomized controlled trial data.* Vaccine, 2018. **36**(4): p. 473-478.

243. Truck, J., et al., *Divergent Memory B Cell Responses in a Mixed Infant Pneumococcal Conjugate Vaccine Schedule.* Pediatr Infect Dis J, 2017. **36**(5): p. e130-e135.

244. Truck, J., et al., *Pneumococcal serotype-specific antibodies persist through early childhood after infant immunization: follow-up from a randomized controlled trial.* PLoS One, 2014. **9**(3): p. e91413.

245. Bock, C., et al., *Quantitative comparison of genome-wide DNA methylation mapping technologies.* Nat Biotechnol, 2010. **28**(10): p. 1106-1114.

246. Bibikova, M. and J.-B. Fan, *GoldenGate® assay for DNA methylation profiling*, in *DNA Methylation*. 2009, Springer. p. 149-163.

247. Bibikova, M., et al., *Genome-wide DNA methylation profiling using Infinium® assay.* Epigenomics, 2009. **1**(1): p. 177-200.

248. Bibikova, M., et al., *High-throughput DNA methylation profiling using universal bead arrays.* Genome Res, 2006. **16**(3): p. 383-93.

249. Steemers, F.J., et al., *Whole-genome genotyping with the single-base extension assay.* Nat Methods, 2006. **3**(1): p. 31-33.

250. Weisenberger, D., et al., *Comprehensive DNA methylation analysis on the Illumina Infinium assay platform.* 2008.

251. Bibikova, M., et al., *High density DNA methylation array with single CpG site resolution.* Genomics, 2011. **98**(4): p. 288-95.

252. Dedeurwaerder, S., et al., *A comprehensive overview of Infinium HumanMethylation450 data processing.* Brief Bioinform, 2014. **15**(6): p. 929-941.

253. Sandoval, J., et al., *Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome.* Epigenetics, 2011. **6**(6): p. 692-702.

254. Moran, S., C. Arribas, and M.J.E. Esteller, *Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences.* Epigenomics, 2016. **8**(3): p. 389-399.

255. Takai, D. and P.A. Jones, *Comprehensive analysis of CpG islands in human chromosomes 21 and 22.* Proc Natl Acad Sci, 2002. **99**(6): p. 3740-3745.

256. Solomon, O., et al., *Comparison of DNA methylation measured by Illumina 450K and EPIC BeadChips in blood of newborns and 14-year-old children.* Epigenetics, 2018. **13**(6): p. 655-664.

257. Du, P., et al., *Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis.* BMC bioinformatics, 2010. **11**(1): p. 587.

258. Durbin, B.P., et al., *A variance-stabilizing transformation for gene-expression microarray data.* Bioinformatics, 2002. **18**(suppl_1): p. S105-S110.

259. Wilhelm-Benartzi, C.S., et al., *Review of processing and analysis methods for DNA methylation array data.* British journal of cancer, 2013. **109**(6): p. 1394-1402.

260. Aryee, M.J., et al., *Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays.* Bioinformatics, 2014. **30**(10): p. 1363-9.

261. Fortin, J.P., T.J. Triche, Jr., and K.D. Hansen, *Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi.* Bioinformatics, 2017. **33**(4): p. 558-560.

262. Hansen, K. and M. Aryee, *IlluminaHumanMethylationEPICmanifest: Manifest for Illumina's EPIC methylation arrays.* R package version 0.3. 0, 2016.

263. 0, H.K.J.R.p.v., *IlluminaHumanMethylationEPICanno.ilm10b4.hg19: Annotation for Illumina's EPIC methylation arrays.* 2017.

264. Daca-Roszak, P., et al., *Impact of SNPs on methylation readouts by Illumina Infinium HumanMethylation450 BeadChip Array: implications for comparative population studies.* BMC Genomics, 2015. **16**: p. 1003.

265. Naeem, H., et al., *Reducing the risk of false discovery enabling identification of biologically significant genome-wide methylation status using the HumanMethylation450 array.* BMC Genomics, 2014. **15**: p. 51.

266. Chen, Y.-a., et al., *Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray.* Epigenetics, 2013. **8**(2): p. 203-209.

267. Chen, Y.A., et al., *Sequence overlap between autosomal and sex-linked probes on the Illumina HumanMethylation27 microarray.* Genomics, 2011. **97**(4): p. 214-22.

268. McCartney, D.L., et al., *Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip.* Genom Data, 2016. **9**: p. 22-4.

269. Fortin, J.-P., et al., *Functional normalization of 450k methylation array data improves replication in large cancer studies.* 2014. **15**(11): p. 1-17.

270. Wu, Z. and M.J. Aryee, *Subset quantile normalization using negative control features.* J Comput Biol, 2010. **17**(10): p. 1385-1395.

271. Touleimat, N. and J.J.E. Tost, *Complete pipeline for Infinium® Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation.* Epigenomics, 2012. **4**(3): p. 325-341.

272. Ziller, M.J., et al., *Charting a dynamic DNA methylation landscape of the human genome.* Nature, 2013. **500**(7463): p. 477-81.

273. Jaffe, A.E. and R.A. Irizarry, *Accounting for cellular heterogeneity is critical in epigenome-wide association studies.* Genome Biol, 2014. **15**(2): p. 1-9.

274. Houseman, E.A., et al., *Cell-composition effects in the analysis of DNA methylation array data: a mathematical perspective.* BMC Bioinformatics, 2015. **16**: p. 95.

275. Sehouli, J., et al., *Epigenetic quantification of tumor-infiltrating T-lymphocytes.* Epigenetics, 2011. **6**(2): p. 236-246.

276. Houseman, E.A., et al., *DNA methylation arrays as surrogate measures of cell mixture distribution.* BMC Bioinformatics, 2012. **13**: p. 86.

277. Chen, C., et al., *Removing batch effects in analysis of expression microarray data: an evaluation of six batch adjustment methods.* PLoS One, 2011. **6**(2): p. e17238.

278. Lazar, C., et al., *Batch effect removal methods for microarray gene expression data integration: a survey.* Brief Bioinform, 2013. **14**(4): p. 469-490.

279. Leek, J.T., et al., *The sva package for removing batch effects and other unwanted variation in high-throughput experiments.* Bioinformatics, 2012. **28**(6): p. 882-883.

280. Smyth, G.K., *Limma: linear models for microarray data*, in *Bioinformatics and computational biology solutions using R and Bioconductor*. 2005, Springer. p. 397-420.

281. Benjamini, Y. and Y. Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiple testing.* J R Stat Soc Series B, 1995. **57**(1): p. 289-300.

282. Novak, P., et al., *Agglomerative epigenetic aberrations are a common event in human breast cancer.* Cancer research, 2008. **68**(20): p. 8616-8625.

283. Peters, T.J., et al., *De novo identification of differentially methylated regions in the human genome.* Epigenetics Chromatin, 2015. **8**: p. 6.

284. Ren, X. and P.F. Kuan, *methylGSA: a Bioconductor package and Shiny app for DNA methylation data length bias adjustment in gene set testing.* Bioinformatics, 2019. **35**(11): p. 1958-1959.

285. Mi, G., et al., *Length bias correction in gene ontology enrichment analysis using logistic regression.* PLoS One, 2012. **7**(10): p. e46128.

286. Swets, J.A.J.S., *Measuring the accuracy of diagnostic systems.* Science, 1988. **240**(4857): p. 1285-1293.

287. Ronaghi, M., M. Uhlén, and P.J.S. Nyrén, *A sequencing method based on real-time pyrophosphate.* Science, 1998. **281**(5375): p. 363-365.

288. Tost, J. and I.G.J.N.p. Gut, *DNA methylation analysis by pyrosequencing.* Nature protocols, 2007. **2**(9): p. 2265.

289. Caixia, L., et al., *Involvement of epigenetic modification in epithelial immune responses during respiratory syncytial virus infection.* Microb Pathog, 2019. **130**: p. 186-189.

290. Deaton, A.M. and A. Bird, *CpG islands and the regulation of transcription.* Genes Dev, 2011. **25**(10): p. 1010-22.

291. Eden, S. and H. Cedar, *Role of DNA methylation in the regulation of transcription.* Curr Opin Gen Dev, 1994. **4**(2): p. 255-259.

292. Kaiko, G.E., et al., *NK cell deficiency predisposes to viral-induced Th2-type allergic inflammation via epithelial-derived IL-25.* J Immunol, 2010. **185**(8): p. 4681-4690.

293. Lin, S.J., et al., *Decreased intercellular adhesion molecule-1 (CD54) and L-selectin (CD62L) expression on peripheral blood natural killer cells in asthmatic children with acute exacerbation.* Allergy, 2003. **58**(1): p. 67-71.

294. Wang, Y., et al., *The EYA3 tyrosine phosphatase activity promotes pulmonary vascular remodeling in pulmonary arterial hypertension.* Nat Commun, 2019. **10**(1): p. 4143.

295. Said, S.I., S.A. Hamidi, and L. Gonzalez Bosc, *Asthma and pulmonary arterial hypertension: do they share a key mechanism of pathogenesis?* Eur Respir J, 2010. **35**(4): p. 730-4.

296. Fehrenbach, H., C. Wagner, and M. Wegmann, *Airway remodeling in asthma: what really matters.* Cell Tissue Res, 2017. **367**(3): p. 551-569.

297. Shifren, A., et al., *Mechanisms of remodeling in asthmatic airways.* J Allergy (Cairo), 2012. **2012**: p. 316049.

298. Wilson, J.W. and T. Kotsimbos, *Airway vascular remodeling in asthma.* Curr Allergy Asthma Rep, 2003. **3**(2): p. 153-158.

299. Bigler, J., et al., *A severe asthma disease signature from gene expression profiling of peripheral blood from U-BIOPRED cohorts.* Am J Respir Crit Care Med, 2017. **195**(10): p. 1311-1320.

300. Tang, J., et al., *DEAD-box helicase 27 promotes colorectal cancer growth and metastasis and predicts poor survival in CRC patients.* Oncogene, 2018. **37**(22): p. 3006-3021.

301. Rico-Rosillo, G. and G.B.J.R.A.d.M. Vega-Robledo, *The involvement of NF-κB Transcription factor in asthma.* Revista Alergia de Mexico, 2011. **58**(2).

302. Ha, E.H., et al., *Endothelial Sox17 promotes allergic airway inflammation.* J Allergy Clin Immunol, 2019. **144**(2): p. 561-573 e6.

303. Ohno, I., et al., *Transforming growth factor beta 1 (TGF beta 1) gene expression by eosinophils in asthmatic airway inflammation.* Am J Respir Cell Mol Biol, 1996. **15**(3): p. 404-409.

304. Aschner, Y. and G.P. Downey, *Transforming Growth Factor-beta: Master Regulator of the Respiratory System in Health and Disease.* Am J Respir Cell Mol Biol, 2016. **54**(5): p. 647-55.

305. Kocwin, M., et al., *The role of the TGF-SMAD signalling pathway in the etiopathogenesis of severe asthma.* Pneumonol Alergol Pol, 2016. **84**(5): p. 290-301.

306. Groneberg, D.A., et al., *Smads as intracellular mediators of airway inflammation.* Exp Lung Res, 2004. **30**(3): p. 223-50.

307. Lund, R.J., et al., *Atopic asthma after rhinovirus-induced wheezing is associated with DNA methylation change in the SMAD3 gene promoter.* Allergy, 2018. **73**(8): p. 1735-1740.

308. Hogan, B.L., et al., *Repair and regeneration of the respiratory system: complexity, plasticity, and mechanisms of lung stem cell function.* Cell Stem Cell, 2014. **15**(2): p. 123-138.

309. Nemajerova, A., et al., *TAp73 is a central transcriptional regulator of airway multiciliogenesis.* Genes Dev, 2016. **30**(11): p. 1300-1312.

310. Knolle, M.D. and C.A.J.E.o.o.t.t. Owen, *ADAM8: a new therapeutic target for asthma.* Expert opinion on therapeutic targets, 2009. **13**(5): p. 523-540.

311. Foley, S.C., et al., *Increased expression of ADAM33 and ADAM8 with disease progression in asthma.* J Allergy Clin Immunol, 2007. **119**(4): p. 863-871.

312. Fernandez-Morera, J.L., et al., *Epigenetic regulation of the immune system in health and disease.* Tissue Antigens, 2010. **76**(6): p. 431-9.

313. Arechederra, M., et al., *Hypermethylation of gene body CpG islands predicts high dosage of functional oncogenes in liver cancer.* Nat Commun, 2018. **9**(1): p. 3164.

314. Ladd-Acosta, C., et al., *Common DNA methylation alterations in multiple brain regions in autism.* Mol Psychiatry, 2014. **19**(8): p. 862-71.

315. Maschietto, M., et al., *DNA methylation landscape of hepatoblastomas reveals arrest at early stages of liver differentiation and cancer-related alterations.* Oncotarget, 2017. **8**(58): p. 97871.

316. Henry, C.J., et al., *IL-12 produced by dendritic cells augments CD8+ T cell activation through the production of the chemokines CCL1 and CCL17.* J Immunol, 2008. **181**(12): p. 8576-8584.

317. Martinez, G.J., et al., *Smad3 differentially regulates the induction of regulatory and inflammatory T cell differentiation.* J Biol Chem, 2009. **284**(51): p. 35283-35286.

318. Scharschmidt, E., et al., *Degradation of Bcl10 induced by T-cell activation negatively regulates NF-κB signaling.* Mol Cell Biol, 2004. **24**(9): p. 3860-3873.

319. Kawai, T. and S. Akira, *Regulation of innate immune signalling pathways by the tripartite motif (TRIM) family proteins.* EMBO molecular medicine, 2011. **3**(9): p. 513-527.

320. Lunardi, A., et al., *A role for PML in innate immunity.* Genes, 2011. **2**(1): p. 10-19.

321. Huang, C.T., et al., *Role of LAG-3 in regulatory T cells.* Immunity, 2004. **21**(4): p. 503-13.

322. Janeway Jr, C.A., et al., *The major histocompatibility complex and its functions*, in *Immunobiology: The Immune System in Health and Disease. 5th edition*. 2001, Garland Science.

323. Klein, J. and A. Sato, *The HLA system.* N Engl J Med, 2000. **343**(10): p. 702-709.

324. Ovsyannikova, I.G., et al., *Human leukocyte antigens and cellular immune responses to anthrax vaccine adsorbed.* Infect Immun, 2013. **81**(7): p. 2584-91.

325. Lambert, N.D., et al., *Polymorphisms in HLA-DPB1 are associated with differences in rubella virus-specific humoral immunity after vaccination.* J Infect Dis, 2015. **211**(6): p. 898-905.

326. Leopardi, R., et al., *Effect of measles virus infection on MHC class II expression and antigen presentation in human monocytes.* Cell Immunol, 1993. **147**(2): p. 388-396.

327. Dhiman, N., et al., *Differential HLA gene expression in measles vaccine seropositive and seronegative subjects: a pilot study.* Scand J Infect Dis, 2003. **35**(5): p. 332-6.

328. Consortium, E.P., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.

329. Jones, S.A., *Directing transition from innate to acquired immunity: defining a role for IL-6.* J Immunol, 2005. **175**(6): p. 3463-3468.

330. Su, B., et al., *The effects of IL-6 and TNF-α as molecular adjuvants on immune responses to FMDV and maturation of dendritic cells by DNA vaccination.* Vaccine, 2008. **26**(40): p. 5111-5122.

331. Kurtz, S.L., et al., *Interleukin-6 is essential for primary resistance to Francisella tularensis live vaccine strain infection.* Infect Immun, 2013. **81**(2): p. 585-97.

332. Velazquez-Salinas, L., et al., *The role of interleukin 6 during viral infections.* Front Microbiol, 2019. **10**: p. 1057.

333. Su, L.-k., et al., *Intranasal co-delivery of IL-6 gene enhances the immunogenicity of anti-caries DNA vaccine.* Acta Pharmacologica Sinica, 2014. **35**(5): p. 592-598.