Leverage single cell genomics approaches to decipher early cellular mechanisms involved in the development of adult chronic diseases

Alexandre Pelletier

UNIVERSITY OF SCIENCE AND TECHNOLOGY OF LILLE

HEALTH-BIOLOGY DOCTORAL SCHOOL OF LILLE

Year 2022

Leverage single-cell genomics approach to decipher early cellular mechanisms involved in the development of adult chronic diseases

PhD Thesis

Presented and publicly defended by

Alexandre PELLETIER

The 2nd of December 2022 to the following jury

Reporters	Pr Susan OZANNE, University of Cambridge, UK
	Dr Pilib O BROIN, University of Galway, Ireland
Reviewers	Pr Didier VIEAU, Lille University of Science and Technology
	Dr Marco Antonio MENDOZA, Genoscope, Paris
Guest of honor	Dr Delphine Eberlé, Lille University of Science and Technology
Director	Dr Fabien DELAHAYE, Pasteur Institut of Lille
Co-director	Pr Philippe FROGUEL, Lille University-Pasteur-CHU

UNIVERSITE DES SCIENCES ET TECHNOLOGIES DE LILLE

ECOLE DOCTORALE BIOLOGIE-SANTE DE LILLE

Année 2022

Approches de génomique intégrative à l'échelle de la cellule unique pour étudier les mécanismes précoces de développement de maladies chroniques de l'adulte

THESE DE DOCTORAT

Présentée

et soutenue publiquement par

Alexandre PELLETIER

Le 2 Décembre 2022 devant le jury ci-dessous

Rapporteurs	Pr Susan OZANNE, University of Cambridge, UK
	Dr Pilib O BROIN, University of Galway, Ireland
Examinateurs	Pr Didier VIEAU, Université des sciences et technologies de Lille
	Dr Marco Antonio MENDOZA, Genoscope, Paris
Invité d'honneur	Dr Delphine Eberlé, Université des sciences et technologies de Lille

Directeur	Dr Fabien DELAHAYE, Institut Pasteur de Lille
Co-directeur	Pr Philippe FROGUEL, Université de Lille-Pasteur-CHU

Table of Content

Abstract

The emergence of genomics technologies, especially those combining single-cell resolution and high throughput sequencing, is enabling us to characterize physiopathological mechanisms at a resolution never achieved before. Single-cell genomics sequencing allows researcher to highlight previously hidden intra-tissue cellular heterogeneity and its influences on diseases development. Our ability to generate high-throughput and single-cell data opens new perspectives but also brings new challenges to face. In my thesis, I focused on implementing new strategies to address some of these limitations providing with new evidence toward a better understanding of early disease mechanisms, focusing on two models: the early epigenetics programming of hematopoietic stem and progenitor cells (HSPCs) and the Alzheimer's Diseases (AD) susceptibility gene BIN1 function on study.

In my first model, I leveraged single-cell genomics to decipher early influence of being large for gestational age (LGA) on HSPCs plasticity. We characterized the transcriptional and functional consequences of DNA methylation alterations observed in LGA HSPCs compared to appropriately grown neonates (CTRL) combining single-cell epigenomics, single-cell transcriptomics, and *in vitro* analysis. We found that DNA hypermethylation is associated with hematopoietic stem cells (HSC) specific chromatin rearrangement in the regulatory network of EGR1, KLF2, and KLF4 transcription factors, affecting downstream genes known to sustain HSCs quiescence like SOCS3, JUNB, and DUSP2. Furthermore, we found that this network was enriched for genes with decrease expression in LGA compared to CTRL, supporting transcriptional consequences of these epigenetics alterations. Finally, leveraging both single-cell resolution of our transcriptomics data and *in vitro* differentiation analysis, we found a reduce ability for LGA HSC to stay quiescent/undifferentiated in response to stimulations. Together, this single-cell genomics integrative approach supports that human fetal overgrowth affects HSC quiescence signaling via epigenetic remodeling.

With my second model, I leveraged single-cell transcriptomics to investigate the role of BIN1, the 2nd most AD associated gene, on human brain models. We investigated the cellular effect of BIN1 deletion on both 2D neuronal culture and 3D cerebral organoid derived from human iPSC. We found that BIN1 loss-of-function leads to specific transcriptional alterations in glutamatergic neurons, resembling to the one found in AD brains, involving several genes associated with calcium homeostasis, ion transport and synapse function. Using functional assay, we found that calcium homeostasis and neural networks activity were dysregulated in BIN1 deleted brain models, and that BIN1 was able to interact with voltage-gated calcium channel Cav1. Pharmacological calcium channel blocker was able to partially rescue BIN1 mediated neuronal activity dysregulation supporting important BIN1 role in calcium channel regulation. These single-cell approaches have allowed to show neuronal specific

alteration of BIN1 deletion, and functionally validate role of BIN1 in calcium homeostasis related neuronal activity, while highlight its potential role in AD pathogenesis.

Résumé

L'émergence des approches de génomiques, en particulier depuis l'apparition des technologies de séquençage à l'échelle de la cellule unique (SGC), nous a permis de caractériser des mécanismes physiopathologiques à une résolution jamais atteinte auparavant. Le SGC a mis en évidence l'hétérogénéité cellulaire intra-tissulaire ainsi que son influence sur le développement des maladies, et ouvert de nouvelles perspectives vers une meilleure compréhension de ces mécanismes. Dans ma thèse, j'ai mis en œuvre de nouvelles stratégies d'analyse permettant l'intégration de ces données à l'échelle de la cellule unique dans deux modèles : la programmation épigénétique précoce des cellules souches et progéniteurs hématopoïétiques (CSPH), et l'étude du gène BIN1 dans la maladie d'Alzheimer.

Dans mon premier modèle, j'ai exploité la technologie de SGC pour étudier l'influence d'un excès de croissance gestationnelle (macrosomie) sur la plasticité des CSPH. Nous avons caractérisé les conséquences transcriptionnelles et fonctionnelles de l'altération de la méthylation de l'ADN observées chez les nouveau-nés macrosomes (NNM). Pour cela, nous avons intégré des données épigénétiques transcriptionnelles, et fonctionnelles. Nous avons découvert que l'hyperméthylation de l'ADN chez les NNM était associée à un réarrangement de la chromatine dans les cellules souches hématopoïétiques (CSH), touchant spécifiquement les facteurs de transcription EGR1, KLF2, et KLF4 connus pour soutenir la quiescence des CSH et réguler leur activation. Ce réseau de facteurs de transcriptions inclut notamment SOCS3, JUNB et DUSP2, et est enrichi en gènes dont l'expression est réduite chez les NNM, suggérant que les altérations épigénétiques ont des conséquences sur l'expression de ces gènes. Enfin, grâce à la résolution à la cellule unique de nos données, et à l'analyse de la différenciation *in vitro* des CSH, nous avons constaté une capacité réduite des CSH à rester quiescentes/indifférenciées en réponse aux stimulations chez les nouveau-nés macrosomes. Notre approche intégrative s'appuyant sur l'étude des différents facteurs régulant l'expression génique à l'échelle de la cellule unique nous a permis de confirmer que l'excès de croissance fœtale affectait la signalisation régulant la quiescence des CSH par le biais d'un remodelage épigénétique.

Dans mon deuxième modèle, j'ai exploité le SGC pour étudier le rôle de BIN1, le deuxième gène le plus associé à la maladie d'Alzheimer, sur des modèles de cerveau humain. Nous avons étudié l'effet de la délétion de BIN1 dans une culture neuronale en 2D et sur un organoïde cérébral en 3D tout deux dérivés de cellules souches pluripotentes induites (iPSC) humains. Nous avons découvert que la perte de fonction de BIN1 entraînait des altérations transcriptionnelles spécifiquement dans les neurones glutaminergiques, ressemblant au changement d'expression trouvé chez les individus affectés par la maladie d'Alzheimer. Ces gènes sont fortement associés à la régulation du calcium dans les neurones,

ainsi qu'au transport ionique et à la fonction des synapses. En utilisant des tests fonctionnels *in vitro*, nous avons découvert que la régulation du calcium et l'activité des réseaux neuronaux étaient altérés dans les modèles cérébraux ou BIN1 était inactivé, et que BIN1 était capable d'interagir avec le canal calcique Cav1. Le blocage de ce canal par un inhibiteur pharmacologique permet d'empêcher partiellement l'altération de l'activité neuronale médiée par BIN1, soutenant son rôle important dans la régulation des canaux calciques. Cette approche de SGC a permis de mettre en avant une altération spécifique des neurones due à la délétion de BIN1, et valider fonctionnellement le rôle de BIN1 dans l'activité neuronale liée à la régulation du calcium intracellulaire, tout en soulignant son rôle potentiel dans la pathogenèse de la maladie d'Alzheimer.

Remerciements/ *Acknowledgement*

Ces trois années de thèse ont été une expérience exceptionnelle pour moi, autant d'un point de vue scientifique que d'un point de vue humain. J'ai appris énormément sur le plan technique, scientifique, mais aussi personnel, et cela grâce aux diverses et multiples interactions que j'ai pu avoir avec chacun d'entre vous tout au long de ces 3 années. Pour cela, je vous en serais eternellement reconnaissant.

Tout d'abord, un grand (grand) merci au Dr Fabien Delahaye, qui été le directeur de thèse que beaucoup aimerait avoir, d'un soutien indéfectible, arrangeant, à l'écoute et authentique. Nos interactions, riches et variées, scientifiques ou non, ont largement contribué à mon épanouissement et au fait que je n'ai pas vu passer ces 3 années. A l'heure où j'écris ces lignes, je ne pense ne pas avoir totalement réalisé à quel point cette relation était unique, j'espère qu'elle continuera pour de nouveaux projets scientifiques. Merci pour tous.

Un grand merci également au Pr. Philippe Froguel de m'avoir permis d'intégrer son laboratoire puis partagé son expérience, expertise et vision de la recherche. Vous m'avez toujours donné de bon conseils pour la thèse ou l'après thèse. L'apport de votre recul et encouragements à certain temps clé de la thèse ont été très important pour moi. Merci également au Dr Amélie Bonnefond, qui a su m'apporter son soutien dans les temps instables de ma thèse, mais aussi sa confiance pour installer la technologie single-cell dans le laboratoire NGS/ Plateforme LIGAN.

Je tiens également à remercier Dr Arnaud Carrier, mon coéquipier des premières heures. Merci pour tout ce que tu as fait pour la plateforme single-cell et le projet LGA. Merci de m'avoir formé aux manipulations single-cell et pour tout tes apports/protocoles qui ont été précieux et m'ont permis de continuer les travaux de recherches dans de bonnes conditions même une fois que tu étais parti.

Une attention particulière au Dr Marcos Costa et son équipe, pour cette collaboration fructueuse et ces riches interactions scientifiques. En plus de m'avoir fait murir scientifiquement (et fait murir mon anglais), vous m'avez confirmé que la vision que j'ai de la recherche n'est pas une utopie.

Je tiens également à remercier chaudement l'équipe Bio-informatique de l'UMR, Mehdi, Alaa, Souhila et Lionel, pour leur aide sur le développement bio-informatique et le pipeline de single cell. Merci également à Mickaël, pour ses nombreuses aides sur R ou son environnement, et Lijao et Mathilde pour avoir toujours était présentes pour répondre à mes questions statistiques ou systèmes.

Un grand merci également à Delphine et toute l'équipe TAG menée par le Dr David Hot, pour leur sympathie et soutien technique pour l'utilisation du Nanodrop, Bioanalyzer et le thermocycleur à

Pasteur. Sans vous, la plateforme de single-cell n'aurait pas eu la même convenue. Merci également au Dr Sylvianne Pied et le reste de son équipe pour son aide technique et matériel notamment pour le partage de la salle de culture.

Une thèse serait dure à mener sans le soutien de bons amis qui comprennent ce que tu vis.

Lucas mon compère des « presques » premières heures, tu as été un soutient énorme et devenu un ami, très rapidement. Cette connexion unique que l'on a et nos diverses discussions scientifiques, (géo)politiques ou philosophiques (et même économique !) y ont largement contribué et m'ont apporté recul et légèreté nécessaire pour apprécier cette aventure. Merci d'être entré dans ma vie.

Vincent, tu es arrivé il y a seulement un an, et pourtant j'ai l'impression qu'on se connait depuis 10 ans. Merci énormément pour tous ces moments partagés avec toi et Lucas, à l'open space étudiants ou ailleurs (surtout ailleurs), pour ta fraicheur, finesse d'esprit et d'humour, mais également pour ton soutient dans les moments plus stressant. Je suis fier de te compter parmi mes amis.

Erwan, Dolores, vous êtes les suivants, évidemment. Merci à vous aussi, pour votre gentillesse, sympathie et folie qui nous a si facilement réunis, pour ces moments où l'on a chanté à tue tête au karaoké, ou où l'on a escaladé des montagnes avec plus ou moins de succès.

Une attention également à toi Nawel, la battante au grand cœur, même si l'on ne se voyait pas tous les jours, moi qui était à Pasteur la plupart du temps, tu m'as toujours apporté ton soutient et écoute dès les premiers mois de thèse, et je t'en remercie.

Ma thèse n'aurait également pas eu la même saveur sans mes compères de Pasteur, alors merci Inès, Jérémy, pour votre soutien et pour les nombreux moments de détentes que l'on a pu partager à l'Alchimiste ou ailleurs.

Merci également à toi Théo, copain de thèse en bioinformatique devenu «el presidente » de Bioaddoct. On n'a pas eu beaucoup de temps pour se voir la dernière année, mais ton soutien lors des deux premières années m'a été très précieux.

La famille c'est très important pour surmonter cette épreuve, et heureusement, j'en ai une formidable. Merci du fond cœur Maman, Papa, Amandine, pour votre soutien indéfectible même dans les moments les plus durs, je suis fier d'avoir une famille comme vous et tellement reconnaissant. Je vous aime.

Finally, I would like to thanks Prof Susan Ozanne and Dr Pilib Ó Broin to accept being reporter of this thesis, and Pr. Didier Vieau and Dr Marco Antonio Mendoza to be reviewers for my PhD Defense.

Thank you also Dr Delphine Eberle to accept my special invitation for my PhD defense. Another thank-you to Dr Pilib O Broin and Pr. Didier Vieau to have followed me and give feedback during these 3 years for my CSI.

Abbreviations

Abeta	Amyloid beta
ACDs	Adult chronic diseases
AD	Alzheimer's diseases
APOE	Apolipoprotein E
APOE4	allelic version ε4 of APOE
APP	Amyloid Precursors protein
ATAC-seq	Assay for Transposase-Accessible Chromatin using sequencing
Bisulfite-Seq	Bisulfited DNA Sequencing
cDNA	complementary DNA
ChIP	Chromatin Immunorecipitation
ChIP-seq	ChIP-sequencing
ChIP-seq	Chromatin Immunorecipitation sequencing
CpG	Cytosine-phospate-Guanine
CREs	cis-regulatory elements
CRISPR	clustered regularly interspaced short palindromic repeats
CRISPRi	CRISPR interference
CTRL	appropriately grown neonates
CVD	cardiovascular diseases
dATP	deoxyadenine triphosphate
dCTP	deoxycytosine triphosphate
ddTTP	deoxythymine triphosphate
DEGs	differentially expressed genes
dGTP	deoxyguanine triphosphate
DMCs	differentially methylated CpGs
DNA	**Deoxyribonucleic acid**
DNMT	DNA methyltransferase
dNTP	deoxynucleotides triphosphate
dNTP	dideoxynucleotides triphosphate
DOHaD	developmental origins of health and diseases

DSB	DNA double strand break
EMP	erythro-myeloid progenitors
EOAD	early onset AD
eQTLs	expression quantitative traits loci
eQTM	expression Quantitative Traits Methylation
EWAS	epigenome wide association studies
FACS	Fluorescence activated cell sorting
FISH	RNA fluorescence in situ hybridization
GABA	gamma-aminobutyric acid
GDM	gestational diabetes mellitus
GO	Gene Ontology database
GSEA	Gene Set Enrichment Analysis
GTEx	Genotype-Tissue Expression project
GWAS	genome wide association studies
H3K27ac	acetylation of the Lysine in the 27th position of the Histone H3
H3K36Me3	tri-methylation of the Lysine in the 36th position of the Histone H3
H3K4Me3	tri-methylation of the Lysine in the 4th position of the Histone H3
H3K9Me3	tri-methylation of the Lysine in the 9th position of the Histone H3
HDAC	Histone Deacetylase
HDL	high density lipoprotein
HELP	HpaII tiny fragment Enrichment by Ligation-mediated PCR
HFD	high fat diet
hiNs	hiPSC derived neuronal culture
hiPSC	human iPSC
HMT	Histone methyltransferase
HPA	hypothalamic–pituitary–adrenal
HSC	hematopoetic stem cells
HSPCs	hematopoietic stem and progenitors cells
HTO	hashtag oligonucleotides for sample multiplexing before scRNA-seq
IEGs	immediate early response genes

IGF1	Insulin like Growth Factor 1
iPSC	induced pluripotent stem cells
KEGG	Kyoto Encyclopedia of Genes and Genomes
KI	knock-in
KD	knock-down
KO	knock-out
LGA	large for gestational age
LMPPs	Lymphoid-primed multipotential progenitors
lncRNA	long non coding RNA
LSI	latent semantic indexing
LT-HSC	Long term HSC
LTP	long term potentiation
LVGCCs	L-type voltage-gated calcium channels
MAF	minor allele frequency
MBDs	methyl-CpG-binding domain proteins
miRNA	micro RNA
MNC	mononuclear cells
MNN	mutual nearest neighbors
MPP	hematopoietic multipotent progenitors
mQTLs	methylation Quantitative Traits Loci
mRNA	messenger RNA
MSCs	mesenchymal stem cells
NFTs	neurofribrillary tangles
NGS	new generation sequencing
NPCs	neural progenitor cells
PCA	principal component analysis
PCR	Polymerase chain reaction
POMC	proopiomelanocortin
pre-mRNA	premature messenger RNA
PRS	polygenic risks score

QTLs	quantitative traits loci
RAGE	Receptor for Advanced Glycosylation End
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNA-seq	RNA sequencing
SAM	S-Adenosyl methionine
scATAC-seq	single-cell ATAC sequencing
SCENIC	Single-Cell rEgulatory Network Inference and Clustering
scRNA-seq	Single-cell RNA sequencing
SGA	small for gestational age
sgRNA	single guide RNA
shRNA	short hairpin RNA
siCTRL	siRNAs control
siKLF2	siRNAs targetting KLF2
siRNAs	small interfering RNAs
T2D	type 2 diabetes
TALEs	transcription activator-like effectors
TF	transcription factor
UMI	unique molecular identifier
US	United States
UV	Ultra-violet
WES	whole-exome sequencing
WGS	whole genome sequencing
WHO	World Human Health Organization
WT	wild type
ZFNs	Zinc finger nucleases

Preface

Adult chronic diseases (ACDs) are the leading cause of death worldwide, accounting for 90% of mortality in developed countries [1]. ACDs occurrences increase exponentially with age (figure1) thus the growing aging of the population predict high health and societal consequences in the next decades. Such prevalence calls for new strategies to manage and prevent ACDs. Research on these domains is intense but early mechanism involved in their development are still not completely understood. ACDs are multifactorial relying on complex interactions between environmental and (epi)genetics factors making it difficult to pinpoint specific targets and causal mechanisms. However, the emergence of genomics technologies these recent years, especially those at single-cell resolution, enable us to characterize these mechanisms at a resolution never achieved before. Single-cell genomics sequencing allows researcher to highlight previously hidden cellular heterogeneity and its influences on diseases development. Our ability to generate high-throughput and single-cell data opens new perspectives but also brings new challenges to face. In my thesis, I focused on implementing new strategies to address some of these limitations providing with new evidence toward better understanding of early disease mechanisms, focusing on two models: the fetal programming of ACDs and the Alzheimer's Diseases (AD) susceptibility gene BIN1 function on study.

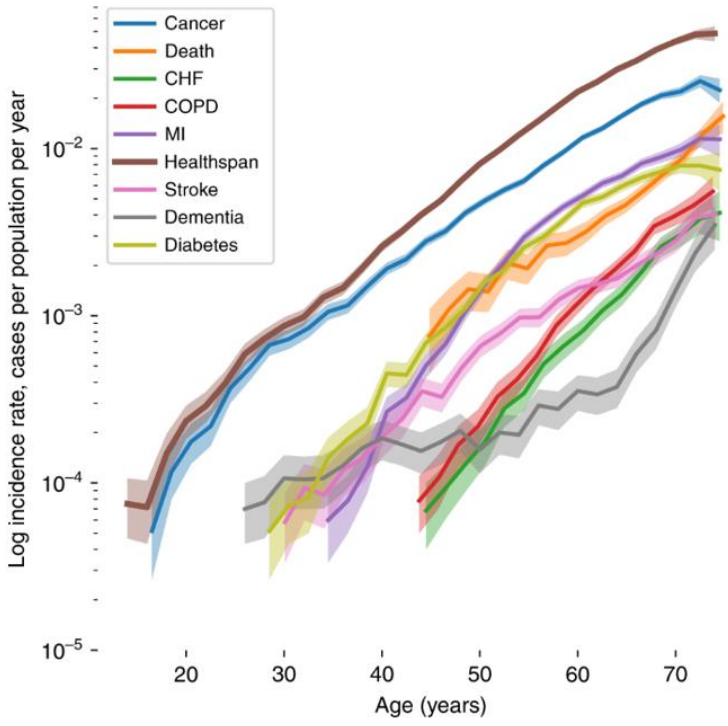


Figure 1 : Exponentional increased of ACDs incidence with age. Data collected from the UK Biobank. Reprinted from Zenin *et al*, communications biology, 2019

INTRODUCTION

I. Genomics approaches to understand adult chronic diseases etiology

I.1. Adult chronic diseases

Adult chronic diseases (ACDs) are chronic diseases whose onset increase with age, leading to progressive and permanent consequences. ACDs are multifactorial, relying on a detrimental interplay between (epi)genetics and environmental factors in the context of aging. Not restricted to genetic heritage and current environment, factors of ACDs also include epigenetics mechanisms and past exposure. ACDs encompass more than 90 diseases with the most important in term of incidence and mortality being cardiovascular diseases (CVD), cancer, Alzheimer's diseases (AD), and type 2 diabetes (T2D)[2,3]. ACDs are interconnected, i.e. one can be a risk factor of another (for example type 2 diabetes with cardiovascular diseases[4]) and share common features still important gaps remain in our understanding of ACDs etiology. I will present in this thesis how genomics approaches help us identify factors involved disease etiology focusing on T2D and AD, two major ACDs that I have been studying.

I.1.a. Type 2 diabetes

Type 2 diabetes (T2D) is a form of diabetes characterized by a hyperglycemia caused by a relative lack of insulin secretion by pancreas on a context of age and obesity associated insulin resistance. It is diagnosed by blood test if during two occasions fasting plasma glucose is over 7mmol/L or if plasma glucose is over 11.1 mmol/L two hours after a glucose tolerance test. One major risk factor of T2D is obesity, still evidence support strong genetics and epigenetics influences. T2D is responsible for severe complications such as heart disease and stroke, with half of diabetic people dying from CVD but will also impact eyes, kidneys, and nerves leading blindness or amputation[5]. With the increasing aging population in both wealthy and low income/middle-income countries, according to WHO diabetes will be the seventh leading cause of death in 2030[5,6].

I.1.b. Alzeihmer's disease

Alzheimer's disease (AD) is a neurodegenerative disease characterized by progressive neuronal degradation in the brain associated with memory and cognitive loss. Even if the definitive diagnosis can be performed only after brain autopsy, cognitive test and PET scan help to have a clinical diagnosis of the disease. It is the cause of 60–70% of cases of dementia and are one of the major causes of disability and dependency among older people globally. First symptoms are short term memory loss and inability to acquire new information as the results of the reduce neuronal plasticity[7]. Cognitive and motor function progressively decline after years as the consequence of neurotoxic aggregates

spreading across the brain. While the causal mechanisms remain poorly understood, several environmental and genetics factors increase the risk to develop AD. A major genetics risk is APOE4, found in ~60% of AD career, while in 15% of the global population[8,9].

I.2. What is Genomics?

Genomics is a recent field in biology corresponding to the study of biological mechanisms at genome[A] wide level rather than at gene level. Genomics study the genetics information encoded in DNA of an organism, their interrelations and influence on the organism. Compared to previous targeted approaches, genomics allows the unsupervised discovery of genes or other molecular elements involved in a physiological or pathological condition. In the context of ACDs, which are polygenic diseases, i.e. their development is dependent on several genetics risks, genomic research comes into its own to decipher these different genetics factors and interaction. It initiates through the development of high-density DNA micro-arrays 25 years ago allowing the simultaneous interrogation of thousands of genes[10]. Then, high-throughput sequencing rapidly emerged enabling measurement of millions of genetics element in one assay [11]. Several 'omics' field are derived from genomics focusing on a downstream or parallel molecular layer, including transcriptomics, studying RNAs, epigenomics, studying the epigenome, or proteomics, studying proteins. However, advance in the specific field largely depend on the technology available and our ability to analyze them. In this thesis, I will mainly focus on genomics, transcriptomics and epigenomics approaches.

I.2.a. Transcriptomics

Transcriptomics study the set of RNAs produce by a cell or a population of cells in a given condition. The gene expression profile, *i.e.* the transcriptome, is tissue or cell type specific allowing specific proteome expression and therefore cellular activity. Because RNAs, compared to proteins, can be easily isolated and sequenced, transcriptomics analysis is a method of choice to study these cellular activities or assess role of genes in physiological or pathological conditions. This is one of the most used approaches in functional genomics, studying the role of genes in an organism, because of its mature technologies and wide application in fundamental or translational studies. They are essential to study the functional impact of a disease in a tissue/ cell type, as well as deciphering the cellular response to a treatment, useful notably for drug screening or understand drug resistance mechanism[12–14]. They are also commonly used in combination with microarray-based genotyping to study tissue specific impact of genetics variants on gene expression (known as eQTL for expression Quantitative Traits Locis). The latter are largely used to investigate the putative impact of risk loci on

[A] The genome is all genetics information of an organism or a population.

gene expression at tissue / cell type level and thus give biological insights on genetics risk and diseases mechanisms [15]. The main transcriptomic approach was DNA microarray, but was now largely outperformed by RNA sequencing (RNA-seq), allowing unsupervised assessment of gene expression profile in a tissue or in immune-phenotypically defined cell type.

I.2.b. Epigenomics

The epigenomics approaches allow the study of the epigenome. Epigenetics mechanisms are critical aspect in multicellular organism to allow cell specialization and identity. Epigenetics mean 'above' genetics, and was first conceptualized by Conrad Waddington in 1956, when he succeeded to demonstrate inheritance of characteristic in a population in response to an environmental stimulus[16], showing existence of mechanism of inheritance 'above' standard genetics. What is consider as "epigenetics" has largely evolved since then. Even if is still debated, it can be defined as "all molecular or structural change that stably regulate expression of genes without altering DNA sequence". They are for the most part stable across cell division, allowing cell identity, differentiation and related specific gene expression profile. They allow also cells to adapt to a specific environment and can therefore testify about past environmental exposure with long term functional consequences on cell activity. Therefore, they have a central role in disease susceptibility by mediating the long-term consequences of current or past exposure.

In the DNA landscape, epigenetics mechanisms mainly include the CpG methylation and histone marks modification (e.g. specific lysine acetylation or methylation of histone tails) which, by their coordinated remodeling regulate chromatin accessibility to DNA-binding protein like transcription factors (TFs), and activity of the transcriptional machinery leading to the control of gene transcription. Epigenetics mechanisms can also, depending on the definition, include non-coding RNA activity like miRNAs, which targets specific coding transcripts, and regulate their expression by reducing their stability (figure 2).

I.2.b.i. DNA methylation

Cytosine, by their chemical structure, can be methylated in the 5th position of the pyrimidine ring, forming 5-Methylcytosine, or 5mC. In mammals, this methylation mainly occurs in the CpG context, a cytosine followed by a guanine. Transfer of methyl group to Cytosine is catalyzed by DNA methyltransferase (DNMT) enzymes DNMT1, and DNMT3 family. DNMT1 is responsible of maintaining DNA methylation pattern across cell division. DNMT3 family, including DNMT3A and DNMT3B are involved in *de novo* DNA methylation. Around 75% of CpG are methylated across the human genome[17] . Due to their chemical resemblance with Thymine, they can mutate by error during replication, which have led during evolution to a global CpG depletion. However, certain CpGs rich region have been

conserved, called CpG island which have an important role in regulating transcription. They are mainly un-methylated and located in gene promoter regions. Methylation of CpG island has been shown to repress transcriptional expression through two putative independent mechanisms: i) the physical constraint to the transcriptional machinery[18], and ii) the recruitment of methyl-CpG-binding domain proteins (MBDs) which can recruits others epigenetics modifiers like histone modifiers, leading to chromatin conformation change[19]. DNA methylation also plays a key role in repressing transposable elements and are found in gene body of highly transcribed genes where they could regulate splicing and repress activity of cryptic intragenic transcriptional units[20–22] . Outside of promoter and gene body region, DNA methylation could have a TF and context dependent role in promoting or repressing the TF binding in cis regulatory element like enhancer[23] . The main approach to study DNA methylation is genome wide DNA methylation microarray, assessing methylation in hundreds of thousands of CpG sites at genome scale, but approach based on sequencing are also widely used especially to study more targeted regions or focus on a specific population or disease model.

I.2.b.ii. Histone modification

DNA roll up on nucleosomes formed by complexes of proteins called histones, that allows for compaction of the DNA pellet. Several posttranslational histone modifications can alter their affinity to DNA modulating its accessibility to transcriptional machinery. Epigenetics modification of histones are cell type specific and are mainly involved the acetylation or methylation of lysine located in the histone tails. Histone marks are regulated by different epigenetic modifiers enzyme and structure the epigenetics landscape. Each histone modification has a specific role, allowing specific epigenetics readers recruitment and thus regulatory activity[24–26]. For example, H3K27ac (acetylation of the Lysine in the 27th position of the Histone H3) are enriched in enhancer allowing their activation. H3K4Me3 are enriched in promoter and activate transcription, while H3K9Me3 repress them. H3K36Me3 are enriched in gene body of transcriptionally active gene. These histone marks are remodeling through differentiation allowing cell specific chromatin profile and associated transcriptional program. Different consortia were created aiming to generate cell type specific epigenomics data to better understand genomics context, like the Roadmap Epigenomics Program which regroup more than 2800 cell type specific epigenomics data[27]. The main approach to study histone marks profile is ChIP-sequencing (ChIP-seq) combining Chromatin ImmunoPrecipitation (ChIP) and high throughput sequencing.

I.2.b.iii. Non coding RNAs

Non-coding RNAs include all RNA, which are not translate into protein. They principally regulate or modify expression of protein coding RNAs but can also have a direct role in cells activity. The main non-coding RNAs regulating protein-coding gene expression are miRNAs and long non-coding

RNA (lncRNA). Micro-RNAs (miRNAs) are single stranded RNA of about 22 nucleotides that regulate gene expression through RNA silencing. MiRNA is complementary to a sequence of the targeted gene allowing its binding to the mRNAs and that silenced its translation. About 2000 miRNAs were identified in the human genome and have been shown to regulate 60% of the human coding genes[28–30] . lncRNAs are non-coding RNA with size greater than 200pb and have diverse activities, including regulation of gene transcription, epigenetic marks regulation, and post transcriptional regulation. The most known lncRNA is Xist, which inactivate the 2nd X chromosome in female placental mammals through irreversible chromatin modifications[31]. Non-coding RNAs can be studied using (deep) RNA sequencing, but microarrays also exists, allowing notably to profile miRNAs in large cohorts, even if suffer from quantification issues[32].

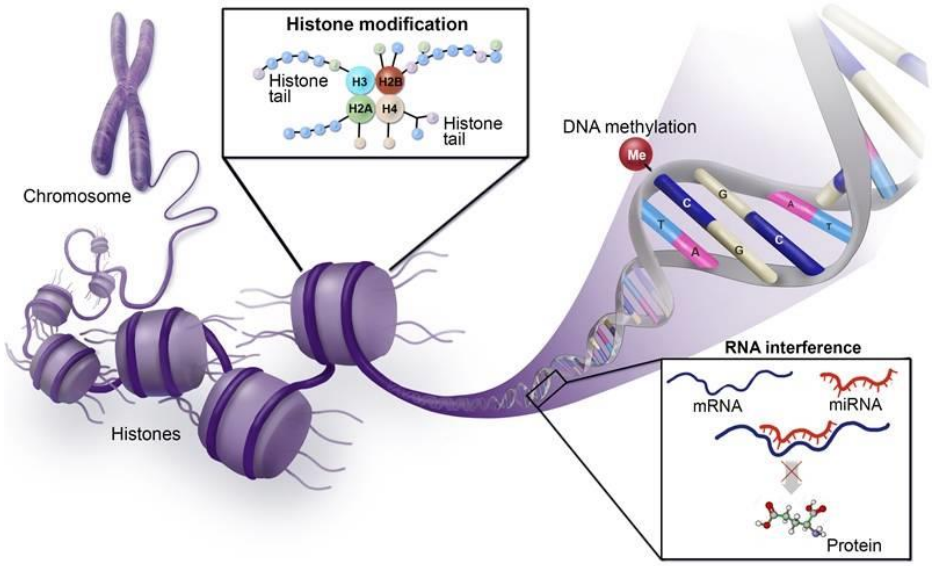


Figure 2 : Different type of epigenetics mechanisms. Source[B]

I.3. Genomics technologies and studies

I.3.a. DNA microarray to perform genomics studies in large cohort

I.3.a.i. Principle

First prominent genome wide assay was based on DNA microarray, allowing interrogating of thousands of genetics elements in one assay. DNA microarray has largely evolved since its birth in the 90's[33], but principle remain similar: thousands of oligonucleotides are bound to a surface and used to measure relative concentration of labeled nucleic acids. These measures are allowed thanks to the complementary sequence hybridization and subsequent quantitative detection of hybridization

[B] https://www.hematology.org/research/ash-agenda-for-hematology-research/epigenetic-mechanisms

events. Recent microarrays are based on microbeads cover by multiple copies of an oligonucleotide probe, which span hundreds of thousands of microwells, link to optic fibers allowing fluorescence detection. Today, microarrays are still largely used for genotyping allowing genome wide association studies (GWAS) or to profile DNA methylome in large cohort allowing epigenome wide association studies (EWAS).

I.3.a.ii. GWAS and genetics risk

Genome wide association studies (GWAS), study association between genetics variant and a phenotypic trait. GWAS used large cohort of case (possessing the certain phenotype/disease) and control individuals with genome wide assay to associate genetics variants to the specific trait. They allow identifying variants, and by extension genes, involved in a multifactorial /polygenic disease like ACDs.

For example for T2D, 245 independents variants and 18 putative causal genes have been identified in the most recent meta-analysis of 32 GWAS[34]. This study has highlighted how increases sample size and variants diversity affects discovery of causal T2D risk alleles, giving then new insight into the T2D genetics mechanisms as well as clinical benefits. For AD, the most recent meta-analysis have identified a total of 75 risk loci and around 100 putative causal genes, with strong enrichment for genes express in immune related tissues, and lipid related processes[35] .

While GWAS are important to identify genes involved in the disease and to estimate individual genetics susceptibilities, they do not, or only partially, take into account the environmental and epigenetics component in diseases development. Age, BMI, and sex are often the only environmental factors integrated in GWAS model limiting ability to find gene-environment interactions. Other limitations are that GWAS discovery is generally limited to frequent variants (>5% minor allele frequency; MAF) present in genome wide array (typically 1.8M variants with the 6.0 affymetrix array), but methods are used to impute genotype for variant with <1% frequency in population. Furthermore, the power of GWAS is highly dependent to cohort size. To tackle this limitation, meta-analysis of several GWAS is frequently performed, and more and more studies tend also to assess the variant-variant or gene-gene interaction on disease risk using notably polygenic risks score (PRS), which can considerably increase discovery power and highlight multigenics effect on diseases development[36–38].

In addition to identify genetics risk and genes involved in diseases, GWAS allow also to assess causality between two traits using mendelian randomization. Mendelian randomization studies leverage the fact that genetics variants are randomly distributed across the population to assess effect of one trait (*e.g.* blood cholesterol level) on another (*e.g.* having T2D). They used for that genetics variants influencing the first trait (*e.g.* genetics variants reducing blood cholesterol) to assess if

individuals carrying such genetics variants have less or more risk to have the second trait (*e.g.* having T2D). For example, for AD, GWAS have allow the study of the causal relationship between cognitive related traits and AD, as well as identified a protective effect of cognitive ability and educational attainment on AD risk[39]. However, such studies are limited by the availability of GWAS of the putative causal traits as well as presence of enough independent causal genetics variants in the population. Yet, to validate the causality of an environmental or (epi)genetics variable, interventional studies are needed such as randomized controlled clinical trial, in vitro studies and/or in vivo models.

I.3.a.iii. EWAS and epigenetics mechanism

Epigenome wide association studies (EWAS), accordingly to GWAS, study the associations between epigenetics factors and a phenotypic trait. Most of the EWAS are based on genome wide DNA methylation assay, interrogating association between CpG methylation site and disease. This is typically performed thanks to methylation microarrays, like the Infinium MethylationEPIC interrogating 850k methylation sites. These studies aim to identify disease related epigenetics biomarkers and associated biological pathways. For example, in T2D, EWAS on $CD4^+$ T cells has allowed the identification of a CpG methylation within the ABCG1 gene associated with blood insulin and insulin resistance[40] . In peripheral white blood cells, 798 CpGs were associated with insulin resistance[41]. In a study of obesity, EWAS have found HIF3A methylation in adipose tissue and blood cells, and SOCS3 methylation in blood cells, as the most obesity associated genes methylation[42,43]. In AD brain, cross cortex meta-analysis of EWAS study have found 220 CpGs methylation associated with neuropathology targeting 121 genes[44,45] . Some of these association are brain region specific with notably CpG sites link to ABCA7 gene and HOXA5/HOXA3/HOX-AS3 cluster in the superior temporal gyrus region, while CpG sites link to MCF2L gene in the inferior frontal gyrus region.

I.3.b. High throughput sequencing for unsupervised discovery

I.3.b.i. History of DNA sequencing

The first method to sequence DNA was developed by Frederick Sanger and colleagues in 1977[46]. The original Sanger method relied on an *in vitro* targeted DNA replication of a DNAs sample, using an oligonucleotide primer, a DNA polymerase, classical deoxynucleotides (dNTPs), and a small amount of a modified dideoxynucleotide (ddNTP). Before the DNA replication, DNA sample is divided into 4 separate reactions to incorporate one of the 4 ddNTPs (ddATP, ddGTP, ddCTP, or ddTTP) that when incorporated into the nascent DNA strand block the new strand synthesis and generate a panel of single strand DNA fragments of different size, which can then be separated by electrophoresis. The DNA bands are then visualized using classical autoradiography or UV methods and the DNA sequence of the interrogated genomic region is determined based on these bands (figure 3). Derivatives of this method are still used today because of its very low error rate [47].

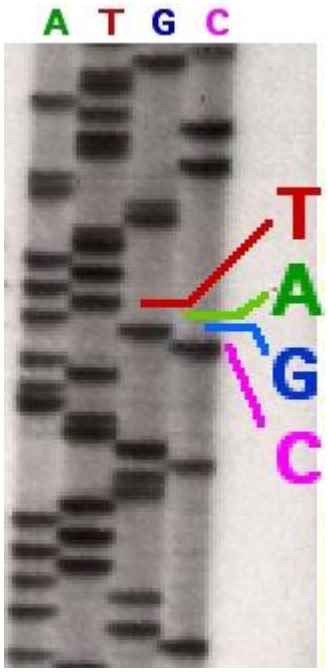


Figure 3 : Revelation of DNA sequence using the Sanger method. Radioactively labelled gel electrophoresis of the four reactions containing either ddATPs, ddTTPs, ddGTPs, or ddCTPs blocking DNA elongation when integrated. Source[C]

This method was used through the Human Genome Project. This unique initiative launched in October 1990 and completed in April 2003, aimed to sequence the entire human genome costing around 3 billions of dollars[48]. Currently, a whole genome can be sequenced in one day for an approximate cost of ~1000$ [49]. This new generation sequencing (NGS) technology, also called high-throughput sequencing, are now largely used across scientific community with various application in genomics, epigenomics and transcriptomics research.

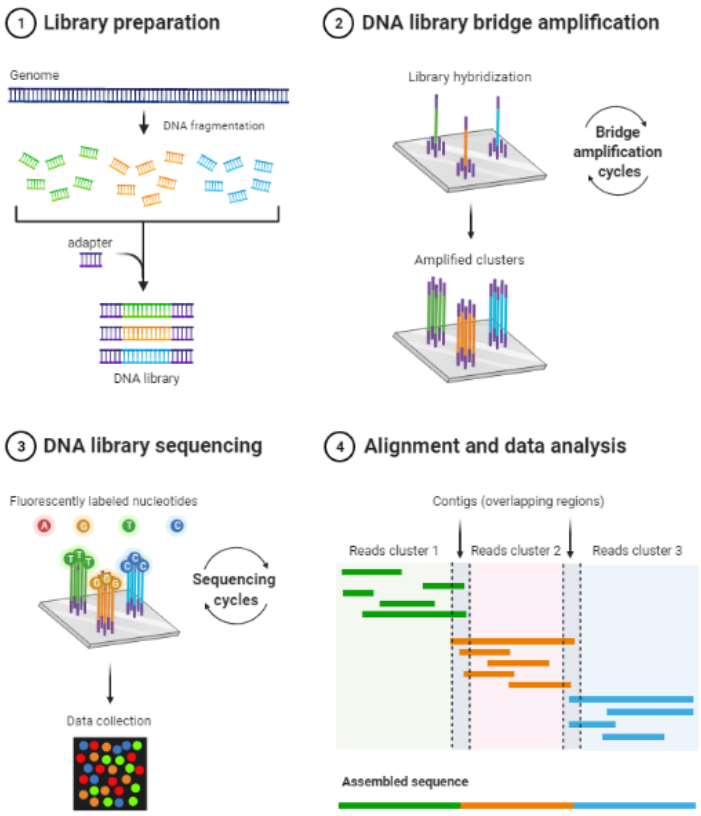
I.3.b.ii. Principle of high throughput sequencing

Several high-throughput sequencing methods exist but the widely used are synthesis-based method, as the one developed by Illumina [50]. They allow sequencing of hundreds of millions of DNA fragments of 100-300 bp in a massively parallel way. NGS is used for a variety of applications including the whole genome sequencing (WGS), the transcriptome sequencing (RNA-seq) and the epigenome sequencing (Bisulfite-Seq or ChIP-seq). For RNA-seq, a preliminary step of retro-transcription is necessary to convert RNA into DNAs.

The method consists of 4 steps: the DNA preparation, the clusters generation in a flow cell, the sequencing by synthesis, and the sequence mapping on a reference (figure 4). The DNA preparation include the DNA fragmentation into 100-300pb fragment and ligation with adaptors including

[C] https://upload.wikimedia.org/wikipedia/commons/c/cb/Sequencing.jpg

sequence allowing hybridization on the flow cell and sample indexing. These fragments are hybridized on a flow cell coated with billions of oligonucleotides primers and amplified by polymerase chain reaction (PCR) generating clusters of copy of a same DNA fragment. Similarly to what is done by Sanger the sequencing by synthesis relies on nucleotide-by-nucleotide synthesis of the complementary strand of the amplified DNA fragments. This synthesis is based on cycle of 3 steps. A mix of dATPs, dGTPs, dCTPs and dTTPs chemically modified to contain a specific fluorescent tag, which block integration of subsequent dNTPs. The polymerase adds one of this modified dNTPs to the nascent strand, which block the synthesis (i). Then, the nature of the nucleotide newly integrated is determined by reading the fluorescence (ii). Finally, the fluorescent tag is removed from the new nucleotide (iii), and the synthesis can continue (step i), until the defined number of cycles wanted (typically 100 or 250pb). The fluorescence is read simultaneously for all clusters at each cycle generating finally the composition of nucleotides of hundreds of millions of fragments simultaneously. These millions of sequences, called "reads" are then aligned on a reference genome or transcriptome depending on the usage, and allow variety of downstream analysis including mutation analysis, gene expression measurement, DNA methylation and histone marks profiling according to the assay/ starting material.



25

Figure 4: Workflow of sequencing by synthesis method. Reprinted from "Next Generation Sequencing (Illumina)", by BioRender, June 2020[D]

The main advantage of NGS compared to microarray-based assay are their ability to measure a genetic information in an unsupervised way. This allows the identification of new biological features in genomics and related research.

I.3.b.iii. Whole genome sequencing to discover new mutations

Whole-genome sequencing (WGS) allow the detection of new or rare variants that are not present in microarrays. An alternative of it is whole-exome sequencing (WES), targeting coding region to increase discovery power of diseases-associated rare variants, which are likely to have greater impact on genes function.

For GWAS needed large cohort, this is quite rare to use WGS/WES because of their cost and amount of data produced. However, some studies have done this effort to understand contribution of rare variant in ACD inheritability. In a study focusing on its question for T2D, no evidence of significant contribution of rare variants was found explaining T2D inheritability[51] . To contrast, a recent WGS based GWAS have found two novel AD associated genes region, *DLG2* and *DNTB*. These AD associated genes region was found thanks to association with rare variants using a sliding window approach across the genome[52] .

Beyond inherited genetics variants, WGS can also be used to identify somatic mutations. Somatic mutations are DNA mutation within non-germinal cells due to DNA replication error or incorrect DNA repair. It's estimated to occur at a rate of 2-10 mutations per diploid genome per cell division[53] leading to mosaicism within each individual, i.e. cells of an organism will not have the exact same genomic information. Using a greater sequencing depth, somatic mutation can detected using WGS, which can be helpful for ACDs research. Indeed, somatic mutation is a well-recognized factor in cancer[53] but play also an important role in others ACD[54,55]. For example, somatic-mutation-driven clonal hematopoiesis and clonally expanded chromosomal alterations in blood was associated with an increased incidence of CVD and T2D[54,56,57] . The causal role of somatic-mutation-driven clonal hematopoiesis has been functionally validated for one of them, the TET2 loss-of-Function mutation, which drive clonal hematopoiesis while aggravate insulin resistance in aged mice and in obese mice[58]. In brain it was shown that neuronal somatic mutation increase with age and are associated with neurodegenerative disorders[55,59]. Putative deleterious somatic mutations was found in 27% of AD brains and particularly enriched for genes contributing to AD pathogenesis[59].

[D] https://app.biorender.com/biorender-templates/figures/all/t-5ef134a11c72b100ad8d13ac-next-

I.3.b.iv. Functional studies using bulk RNA-seq

Bulk RNA-sequencing (RNA-seq) can assess the entire genes expression profile (i.e. the transcriptome), of a tissue or a cell type, permitting to characterize tissue specific genes activity but also disease associated gene expression. To do so, RNAs are isolated from a mix of cells from a tissue or a cell line, and retro transcripts into cDNAs for NGS based sequencing. It was used to build atlas of tissue expression in differents species[60–62] , notably in human with the Genotype-Tissue Expression (GTEx) project[63,64] or with the Human Protein Atlas program[65]. Furthermore, RNA-seq is also widely used as first approach to assess impact of a disease in a tissue. It allow to identify differentially expressed genes (DEGs) in the diseases condition compared to an healthy control condition, highlighting putative biological process or signaling pathways altered or involved in the disease. It is also the primary analysis to identify function of gene in physiological or pathological conditions. Indeed, RNA-seq is also used in first approach to assess the functional impact of modifying or deleting a studied gene in a specific tissue. In ACDs research, RNA-seq data are also useful to assess impact of a genetics variant on gene expression, either through eQTL studies[66] (see part I.3.a) or direct *in vitro*/*in vivo* characterization using genome editing methods[67] (see part I.4). RNA-seq can also be used as readout in drug screening study[14,68]. Drug screening employ a large bank of drugs or molecular compound to characterize impact of these compounds on cells, which can be used for drug discovery or drug repositioning. This approach was largely used for personalized cancer treatment[69] , but similar strategies was also deploy for AD [70]. Then, RNA-seq have broad application and was therefore a major advance in functional genomics research. However, the major drawback of this technique if the lack of the cellular heterogeneity consideration that we will discuss in next parts.

To note, RNA-seq is also used for epigenomics studies, because having the ability to detect non coding RNA, notably miRNAs that are important in diseases mechanisms and are promising circulating biomarkers[71–73].

I.3.b.v. Sequencing based epigenomics studies

High throughput sequencing is also largely used to characterize the epigenome, its dynamics across tissue and diseases and its ability to regulate gene expression. The main methods to investigate epigenome using NGS is Bisulfite seq, ChIP seq, and ATAC-seq.

Bisulfite-Seq allow whole genome DNA methylation profiling in a tissue. It is based on bisulfite treatment of DNA followed by sequencing. Bisulfite treatment convert non-methylated cytosine into thymine, allowing segregation of methylated Cytosine from unmethylated Cytosine by sequencing. However, this technic requires a large amount of DNA material so is difficult to perform if the starting cells number are limited[74].

ChIP-seq allows genome wide Histone marks profiling but also profiling of DNA binding proteins on DNA. It is based on the immunoprecipation of chromatin using specific antibody followed by sequencing. Briefly, DNA and DNA binding proteins are first cross-linked using formaldehyde and DNA are randomly fragmented using restriction enzyme or DNA sonication. Then, DNA fragments containing histone marks or TF are pull down using an antibody recognizing this specific histone marks/TF (this is, chromatin immunoprecipation, or ChIP). These fragments are then sequenced and mapped onto a reference genome similarly to others high throughput sequencing based assay. As for Bisulfite-seq, ChIP-seq requires a large amount of DNA material (i.e. cells) to conserve enough material after treatment with formaldehyde, DNA fragmentation, and Immunoprecipation.

ATAC-seq is focused on profiling open chromatin regions, which are marks of active genomic regions. ATAC is for Assay for Transposase-Accessible Chromatin. ATAC-seq assay use a hyperactive mutant transposase Tn5 enzyme, which has the ability to cut and tag accessible DNAs, producing labeled DNA fragments which can then be amplified and sequenced.

Epigenomics assay using sequencing has allowed the discovery of several epigenetics mechanisms involved in ACDs. Bisulfite-seq has notably highlighted that aberant DNA methylation often occurs before cancer development, induced by different life events including acute infection, chronic alcohol consumption, or dysregulated inflammation, and can persist throughout all the lifetime of an individual, even if the carcinogenic factor is no longer present[75–78]. This approach has also showed the important role of DNA methylation as an alternative way to silence tumor suppressor genes, similarly to genetics. Such genome wide methylated DNA sequencing has also allowed to identify important methylation differences within T2D discordant monozygotic twins, with the stronger change being in *MALT1* locus, a gene regulating insulin and glycemic pathway[79]. ChIP-seq has allowed the discovery of “super-enhancer” regions in cancer, regions enriched in H3K27ac histone mark allowing stable oncogene activation[80,81]. In AD, such analysis has identified that H3K27ac and H3K9ac marks correlate with upregulation of chromatin and transcription related genes and contribute to amyloid-β42-driven neurodegeneration[82]. H3K9me3, mediating heterochromatin condensation, was also found enriched in AD brains, leading to downregulation of proximal genes mainly involved in synaptic transmission and plasticity[83]. This assay has also identified that hyperglycemia led to important histone acetylation changes in specific genomics regions, which are associated to persistent expression of proinflammatory genes[84,85]. Such epigenomics tools give us a new understanding of diseases development. However, to assess functional consequences of such epigenetics remodeling, integration with different omics layer is needed.

I.3.c. Multi-omics integration for functional characterization

Studies focusing only at one specific biological layer (genetics, RNA, DNA methylation, histone marks) limit considerably the understanding of biological or diseases mechanisms. Disease development relies on a complex biological system, necessitating the integration of multiple layers (multi-omics) to highlight interactions between them bringing substantial new insights in our understanding of the molecular mechanisms involved.

I.3.c.i. Quantitative trait loci (link genomics with other layers)

To measure the influence of genetics variants on gene expression or other biological variables, quantitative trait loci (QTLs) studies are performed. Expression Quantitative Trait Loci (eQTLs) studies look for associations between genetics variant and expression levels of mRNAs in a tissue. They are mainly cis-eQTLs studies, *i.e.* considering genes in relatively close proximity of the genetic variant (typically in a 1 Mb window). They allow to identify in some extent the 'proximal' regions able to regulate expression of genes, *i.e.* cis-regulatory elements (CREs) such as promoter and enhancer. To contrast, trans-eQTL studies looked at association between genetics variants and gene expression with more than 1 Mb distance or from different chromosome. They are performed to identify distant association, which can reflect indirect influence of a variant on gene expression, for example the impact of a TF variant on TF downstream target genes expression. However, these analyses require significant analytical resources to perform extremely large number of test (1.8M variant on classical array multiply by 20k human genes = 36 billions of test), increasing considerably false discovery rate and power discovery. Targeted or variable selection methods are thus required to effectively identify significant association. The tissue specific human eQTLs reference is the Genotype-Tissue Expression (GTEx) project, which aim to build a reference database of human tissue specific gene expression and regulation regrouping transcriptomic data and eQTLs analysis from nearly 1000 individuals across 54 healthy tissues. However they suffer from some bias of sampling because mostly based on postmortem tissue expression of aged or intoxicated individuals so others initiative have been led notably the Roadmap Epigenomics Project[27], aiming to have more representative tissue wide datasets.

Cis-eQTLs are used in complement of GWAS to functionally characterize disease risk loci, linking them to a gene with putative tissue specific impact. Some eQTLs studies have been designed to specifically characterized ACD related regulatory link and involvement of genetics modifications in T2D and AD molecular mechanims[70,86]. In T2D, a study analyzed genomics and transcriptomics data from 112 islet samples associated with ATAC-seq based chromatin profile and showed that T2D risk alleles were enriched in islet specific enhancer and disturb the islet Regulatory Factor X (RFX) activity[86] . In AD, a study regrouping 364 donors has characterized thousands of molecular changes and neuronal gene subnetwork associated to AD neuropathology or severity[70]. While having great interest to discover tissue specific role of non coding region in gene expression regulation, eQTL discovery can be

limited because it requires a large number of statistical tests (one by putative variant-gene association). This number of tests reduces the statistical power of discovery, then a large population is needed to identify significant association after p-value correction for multi testing.

Other QTLs study can be performed, including methylation Quantitative Traits Loci (mQTLs), deciphering role of genetics variants in DNA methylation landscape[87].

I.3.c.ii. Others multi-omics integration (associate epigenomics with transcriptomics..)

Omics integration is not limited to QTL studies. Others association between different regulatory layers can be study to decipher molecular mechanisms behind a phenotype or a diseases development. Different methods exist to integrate multiple omics layer. These methods try to find correlation between these layers either with an unsupervised or supervised approach[88] . Most frequent omics integration are DNA methylation with gene expression, or chromatin accessibility with gene expression, giving insight into epigenetics influences on gene expression but also gene regulatory network involved in physiological or pathological processes.

Epigenomic and transcriptomic data integration allows to correlate epigenetics change to gene expression to better understand functional consequences associated to epigenetic changes. This analysis can be named expression Quantitative Traits Methylation (eQTM) analysis in reference to eQTLs. Several DNA methylation and gene expression correlation have been found within T2D. For examples, DNA methylation of the insulin promoter and the PDX-1 gene were shown to be associated with reduced insulin expression and increased HbA_{1c} levels in pancreatic islet of T2D patients[89] . In AD, integration of DNA methylation and gene expression in multiple brain regions has allowed the discovery of genes epigenetically regulated in AD including ANKRD30B as well as several genes related to immunity and calcium homeostasis (figure)[90].

Others multi-omics integration can give insight into diseases development, including integration of gene expression with miRNA profile, histone marks, but also metabolomics or microbiota (i.e. metagenomics)[91,92]. This multi layers integration allows to characterize the whole molecular network involved in a specific phenotype (Figure 5). In AD brains, integration of transcriptomic, proteomic and epigenomics data have contributed to the identification of major epigenome reconfiguration including increase H3K27ac and H3K9ac associated to upregulation of concordant genes regulating transcription and histone marks as well as AD related pathways[82]. In addition to better understand disease mechanism, it was shown that multi-omics integration allows better patient stratification, diseases subtyping, and give insight into diseases subgroup specific molecular signatures[93,94] . For examples, integration of methylation, gene expression, and miRNAs has allowed to identify a set of multi-omics biomarkers associated with subtype of prostate adenocarcinoma with high risk of recurrence [91]. In AD,

proteomics, metabolomics, and lipidomics integration of cerebrospinal fluid in healthy and AD brain with different cognitive severity has identified new central nervous system pathway alteration in AD and contributed to a better AD prediction and associated cognitive decline based on four multi-omics molecular markers [95].

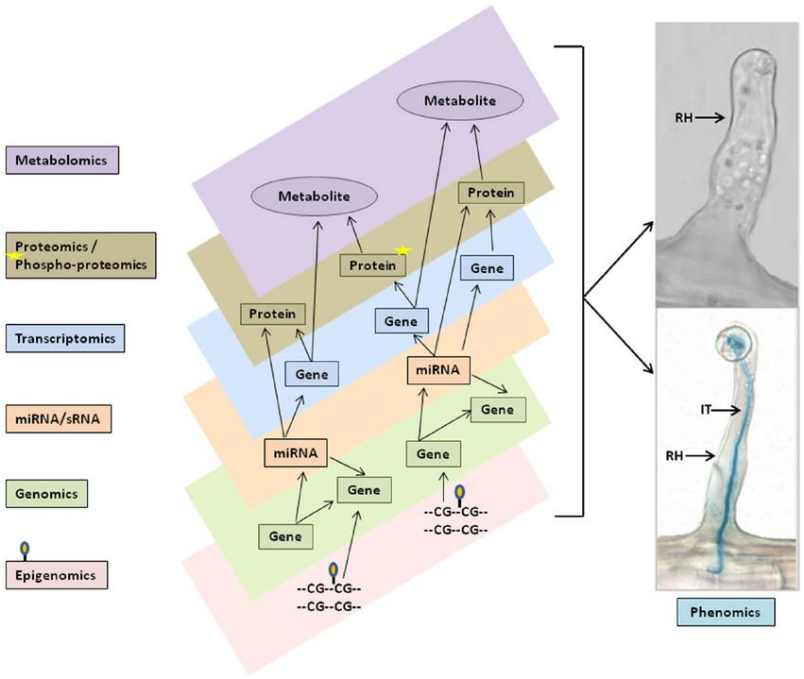


Figure 5 : Advantage of integrate multi-omics layer to characterize molecular network involved in a phenotype. Reprinted from Hossain *et al*, Frontiers in Plant Science, 2015[E]

I.4. (Epi)genetics editing tools

Most of the findings in genomics studies are associations between two variables (e.g. a CpG methylation and a disease). However, association does not imply causality. Even if some statistical approaches are being developed trying to infer causality, including mendelian randomization discussed above, they are limited on the availability of genetics influences on the studied variable. Then, to help decipher causality of gene or (epi)genetics element, we used interventional studies using molecular biology tools. (Epi)genome editing tools are widely used in functional genomics for this purpose. They allow to specifically target a gene or an (epi)genetics element in order to characterize its function. In combination with RNA-seq or others genomics approaches cited above, these are un unsupervised way to assess the role of a gene or an (epi)genetics factors in biological or diseases related processes.

[E] https://www.frontiersin.org/articles/10.3389/fpls.2015.00363/full

I.4.a. Gene perturbation

Since emergence of molecular biology, several tools have been developed allowing us to perturb expression of a gene. Gene silencing using small interfering RNAs (siRNAs), pharmacological inhibitors, gene transfection using plasmid vector, and CRISPR based gene editing, are the main approaches to study the role of a gene and the associated downstream mechanisms and functional consequences.

siRNAs allow knock down (KD) of specific gene expression using RNA-induced silencing complex (RISC) cellular machinery. siRNAs, designed to be complementary to a specific mRNA region (mostly the 3' end untranslated region), is integrated in the cell by RISC and allow specific mRNA binding and cleaving. It is mostly used for transitory downregulation of expression but can also be stably transfected if using short hairpin RNA (shRNA) system. Pharmacological inhibitors are synthetized compound with a specific molecular structure, which, by resembling to natural substrate or ligand of an enzyme or receptor will interfere with the protein activity. Specific pharmacological inhibitors can be difficult to synthetized and can have off target effect limiting their application. Gene transfection using plasmid vector or lentivirus is used to induce expression of an exogenous gene in a cell. This approach is interesting to study the impact of gene overexpression or a specific mutation but can have limited physiological relevance.

To increase physiological relevance, genome editing methods are used[96]. Rather than adding an exogeneous gene, they allow to modify sequence of the endogenous gene or genetic element. These methods mostly used targeted DNA double strand break (DSB) and endogenic homologous recombination-based DNA repair processes to edit genome(Figure 6). First methods developed used fusion protein composed of nonspecific DNA cutting domain coupled with specific DNA sequence recognizing peptides like Zinc finger nucleases (ZFNs) or transcription activator-like effectors (TALEs) allowing targeting DNA cutting. Because they are based on specific fusion protein design, these methods are relatively complex to set up requiring significant molecular biology skills and time. In the past 10 years, a new genome editing tool emerged, bypassing this limitation using only a small RNA (called sgRNA for single guide RNA) to guide the DNA breaking by endonuclease. This tool is called CRISPR-cas9 and became rapidly the gold standard method. This method is based on CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats), a genetics element used by bacteria to fight against viruses. CRISPR technics allow targeting of specific DNA element thanks to a DNA endonuclease enzyme, most of the time Cas9, guided by an easily customizable sgRNA. Cas9 can be catalytically active allowing DSB but also partially or completely inactived (dead Cas9, or dCas9) depending on the DNA modification desired. CRISPR based system allows specific modification of

genomics sequences for gene knock-out (KO) or knock-in (KI) but can also be used for various other application when using a modified version of the system (Figure 7). For example, gene silencing (called CRISPR interference, or CRISPRi) can be performed with such assay by using Cas13 instead of Cas9, an enzyme that targets RNA instead of DNA.

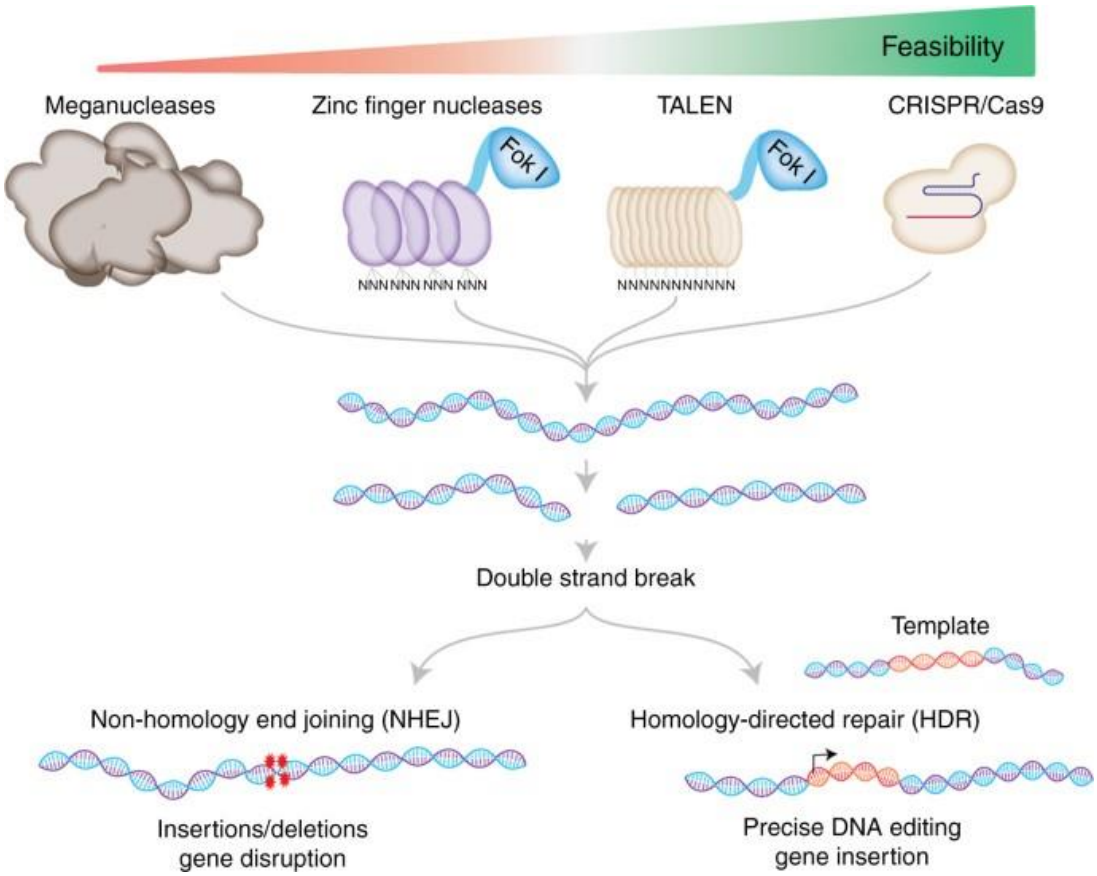


Figure 6 : Principle of major genome editing methods. Reprinted from Adli, Nature Communications, 2018[F]

I.4.b. Epigenetics editing

In the same way than genome editing, some epigenetics editing method have been developed to study role of epigenetics elements. For DNA methylation, epigenetics drugs like S-Adenosyl methionine (SAM), the principal substrate of methyl group transfer can be used to modify DNA methylation but does not allow for targeted epigenetics modification . To get a targeted epigenetics editing, most promising approaches used the CRISPR system. They used inactivated Caspase like dCas9 coupled with epigenetics modifiers catalytic domain to edit specific chromatin marks. Notably, several

[F] https://www.nature.com/articles/s41467-018-04252-2/figures/3

studies have shown ability of such system to edit specific histone acetylation or methylation but also edit DNA methylation. However, these methods still lack robustness requiring further development.

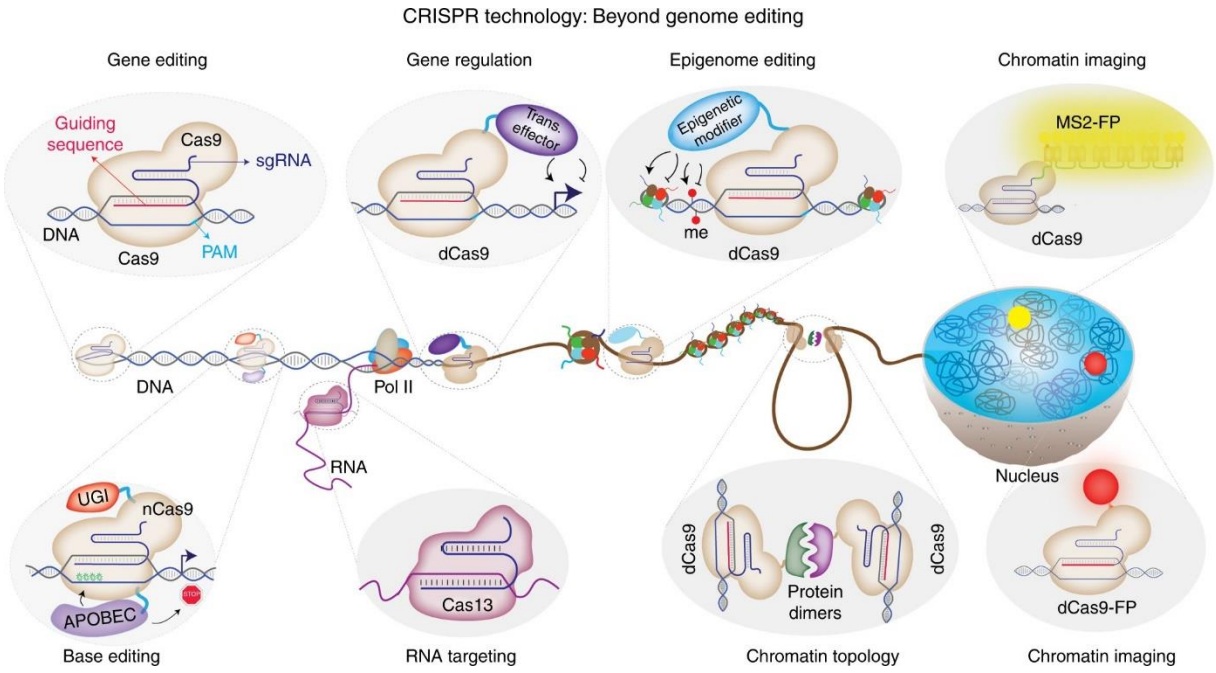


Figure 7 : Differents application of CRISPR based systems. Reprinted from Adli, Nature Communications, 2018[G]

I.5. Main limit of classical genomics approaches

These genomics approaches cited above still suffered from a major gap: the consideration of cellular heterogeneity within a tissue. Indeed, these genomics assays are performed in bulk, i.e. with a RNAs/DNAs mix from several cells. This mix of cells is performed at tissue level or using presorted cells based on cell surface markers. Bulk approach was required to have enough genomics material to perform genomics library preparation for sequencing, but lead to the loss of crucial heterogeneity within this cell population. Notably, the biological insight of RNA-seq based studies is limited because it fails to explain from which cells (subpopulation) the effect observed come from. Similar objection can be raised for EWAS or epigenomics study. Epigenetics influence is cell type specific, then epigenetic change in a cell type cannot be generalize to others but rather highlight cell type specific epigenetics mechanism and role in disease development. We will see in the next parts how the transcriptomic and epigenomics cellular heterogeneity is important in health and diseases and how its consideration in genomics research can give new insight into disease development, mainly focusing on ACDs.

[G] https://www.nature.com/articles/s41467-018-04252-2/figures/3

II. Importance of cellular heterogeneity

II.1. Cellular heterogeneity in multicellular organism

Cellular heterogeneity is a key feature of multicellular organism. It allows asymmetric cell-to-cell interactions, and emergence of complex functions and behaviors. The cell specialization relies on epigenetics remodeling, which give rise to a specific cell structure and activity. During development, this mechanism allows the formation of tissue, *i.e.* grouping of cells specialized in a specific task. However, this cellular heterogeneity goes beyond the tissue level. Indeed, there is different cell type and cell states in each tissue allowing regulation of tissue functions. Cellular heterogeneity is even found within a same cell type according the cell cycle phase, the micro-environment, the cell to cell communication, but also according to somatic mutation and epigenetics mosaicism as seen in normal aging. Furthermore, the cellular plasticity, which allow organism to adapt to environmental change, is also an important factor of cellular heterogeneity. Cellular plasticity is the ability for cells, to change their activity, or to differentiate, in response to environmental cues. Adult stem cells, located in cellular niche across each tissue, can differentiate to regenerate tissue following damaged or for physiological turnover of cells. Differentiated cells itself can be reprogrammed into another cell type, a process called transdifferentation, allowing further organism plasticity to environmental exposure. For example, astrocyte can differentiate in neurons after brain injury[97,98] and white adipocytes can differentiate in brown adipocyte following cold exposure[99]. Then, accounting for intra-tissue heterogeneity in our genomics approaches appears crucial to better understand the biological system and how its dysregulation can lead to diseases development (Figure 8).

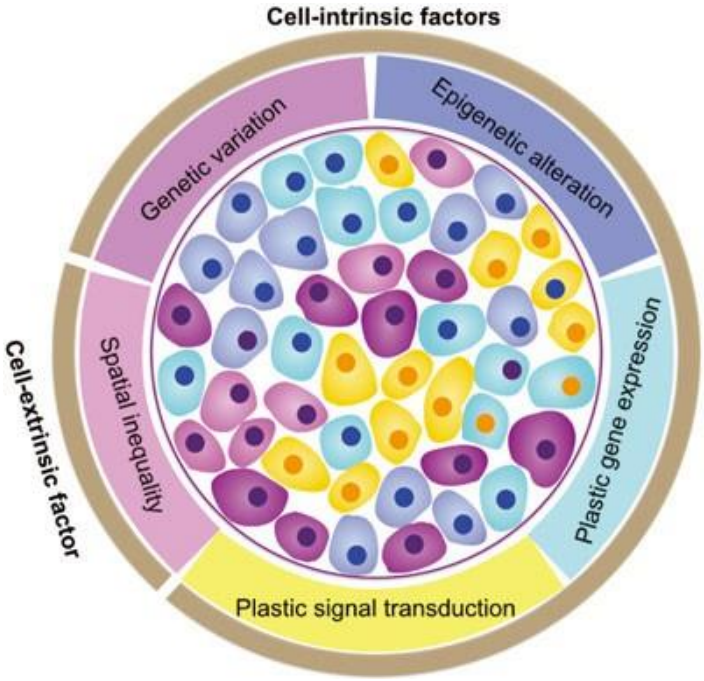


Figure 8: sources of cellular heterogeneity within a tissue. Reprinted from Sun *et al*, Acta Pharmacologica Sinica, 2015[H]

II.2. Cells heterogeneity alteration in diseases

Several evidence support the role of cellular heterogeneity or plasticity alteration in the early development of ACDs. Notably, the stem cells heterogeneity alteration is an important hallmark of aging and contribute to ACDs risk . This stem cells heterogeneity alteration is mainly driven by the clonal expansion of defective stem cells across time. Indeed, as exemplified in hematopoiesis, somatic mutations and epi-mutations[I] accumulate with age according a fitness advantage to certain clone compared to others. Such clonality directly affects the heterogeneity and plasticity of hematopoietic stem cells (HSC) niches driving hematopoietic dysfunction and increasing ACDs susceptibilities[54,56].

In T2D, imbalance between Beta cells and alpha cells in pancreas as well as Beta cells dedifferentiation lead to decrease insulin secretion and diabetes development [100–102]. Similarly, excess of large size white adipocytes in adipose tissue led to obesity associated inflammation and insulin resistance [103]. In addition to adipocytes, adipose tissue is composed of other cell types including stem cells, pre-adipocytes, endothelial cells, neutrophils, lymphocytes, and macrophages[104] . A balanced proportion of these cells is closely related to the maintenance of energy homeostasis, while

[H] https://www.nature.com/articles/aps201592
[I] Stochastic epigenetics alteration, like DNA hypometylation observed in aging

dysregulation of this equilibrium is associated with the metabolic syndrome[J]. It was shown that an increase of adipocytes sizes, types, as well as increase in number of lymphocytes and macrophages infiltration contributes to the metabolic syndrome, associating a low-grade inflammatory state and peripheral insulin resistance[104]. For AD, change in brain cellular composition has been observed. Indeed, it was recently shown through scRNA-seq analysis that a new type of microglia appears in AD brain compared to normal brain and was shown to allow Abeta clearance once activated[105]. It was difficult to assess a loss or a gain of cellular heterogeneity prior to the emergence of single-cell genomics, as cell type identification relied on known cell surface markers.

An important implication of cellular heterogeneity is the ability to develop asymmetrical cell to cell communication essential for the organism homeostasis and tissue synchronization. This cross talk can be altered and played an important role in ACDs. During atherosclerosis leading to CVD, the inflammatory cross talk between macrophage and endothelial cells failed to resolve and lead to accumulation of senescent lipid rich macrophages on the subendothelial space upon rupture[106]. In AD, defective cross talk between neurons, astrocytes and microglia appears primordial in AD pathogenesis especially in the Ab plaques maturation and propagation[107]. This cell-to-cell interaction further highlights the importance to consider cellular heterogeneity when studying ACDs. Considering cell to cell interaction adds complexity in our models but will improve identification of actionable targets. For example, all AD clinical trials focusing on treatment of astrocytes and microglial-mediated neuroinflammation have failed[108] suggesting that targeting only neuroinflammation is not sufficient and better understanding of the cells crosstalk involved in immune cells activation is needed. Recent studies suggest that microglial activation could be necessary to decrease neurons dysfunction and amyloid beta (Abeta) accumulation[105,109]. Complex cellular models integrating this cellular heterogeneity like organoid coupled with single-cell genomics assay appear then a first choice strategy to decipher these cross talks.

III. Single-cell genomics approaches, their interests and how to manage them

The intra-tissue heterogeneity plays a key role in cellular homeostasis, tissue function regulation, and disease development but was often disregarded due to the lack of available tools to characterize it. The recent emergence of single-cell genomics technologies allows us to tackle this limitation.

[J] The metabolic syndrome is an obesity, T2D, CVD risk associated medical condition regrouping at least 3 over the 5 following strinkingly correlated metabolic phenotypes: abdominal obesity, high blood pressure, high blood sugar, high serum triglycerides, and low serum high-density lipoprotein (HDL)

III.1. History

Even if we can speak about single-cell "genomics" only recently with the rise of high-density microarray and high-throughput sequencing, technologies allowing analysis of samples at single-cell resolution exist since decades. In combination with staining methods microscope allows exploration and analyze of cells and tissues at single-cell resolution. After some centuries, modern technics combining molecular labelling and computational analysis have enable single-cells analysis in a more quantitative way. Fluorescence activated cell sorting (FACS), immunofluorescence, and RNA fluorescence in situ hybridization (FISH) as well as fluorescent fusion proteins are methods allowing quantification of gene expression at cellular level. However, these methods are limited in scale either on number of cells or in the number of targets explored simultaneously. The first massive parallel sequencer was commercialized in 2006 allowing the sequencing of the whole genome in a day. The genomics field then emerged and soon enables single-cell approaches.

In 2009 was published the first whole transcriptome at single-cell resolution[110]. Few years after, different protocols arised to isolate cells including well-based, FACS-based, and droplet-based assays. All these protocols are based on a similar process: a cell isolation (i), RNA retrotranscription adding cell specific barcodes (ii), and finally cDNA library amplification and preparation for Illumina related sequencing (iii).

III.1. Principle

Well-based assay used limiting dilution or micromanipulation to isolate one cell by well allowing specific barcoding of their RNAs. They allow for visualization and confirmation of cell isolation process, and are classically associate to Smart-seq protocol, a RNA-seq library preparation method, allowing full-length whole transcriptome sequencing[111]. However, such methods suffer from a limited number of cells that can be processed simultaneously.

Rapidly, FACS methods have been developed to reduce manipulation and time while allowing sorting of thousands of cells in plate. Such methods are classically associated with SMART-seq2 protocol, an improve protocol of Smart-seq allowing also full length transcriptome sequencing but with greater sensitivity[111,112]. A specificity that allows the interrogation of gene isoforms, or alternative splicing. This workflow is also useful when studying exome based somatic mutation, and can be used to perform lineage tracing experiments, based on spontaneous mitochondrial DNA mutation monitoring[113]. However, FACS based technics still have a limited cells throughput (of about 1000 cells) while requiring a large number of starting materials (>100k cells to good recovery). Furthermore, such assay is also expensive according to the large reagent volume it requires. Indeed, the critical steps of

such assay, the retro transcription, the tagmentation and the PCR, are performed in each individual well requiring large amount of associated enzymes and buffers.

Since 2015, a new isolation method has emerged allowing sequencing of thousands of cells in one assay[114]. This method is a microdroplet-based microfluidics approach. It allows cell isolation through an oil-based emulsion of water micro droplets containing all necessary materials to perform retro transcription and cell resolution RNA barcoding. This design reduces dramatically reagent volume and cost while increasing number of cells that can be simultaneously profiled. This technic, called Drop-seq, was then commercialized by 10X Genomics and widely used across scientific community. This assay is not designed to interrogate full length RNA like Smart-seq2 methods, limiting splicing analysis / isoforms profiling as well as (somatic) mutation analysis.

Regardless of the cell isolation methods applied, the principle of single genomics methods remains the same (Figure 9). Cells are dissociated from a tissue, and isolated to allow the cell level barcoding of the genomics element of interest (RNA or DNA). If single-cell RNA-seq, is performed, the RNA is retrotranscripted into DNA, and the resulting barcoded DNA is sequenced according to classical NGS workflow. After mapping reads on the reference genome, we obtain a gene expression (or epigenetic) profile for every cell sequenced. These profiles are then compared between sequenced cells to identify the different cell subpopulations present in our original tissue.

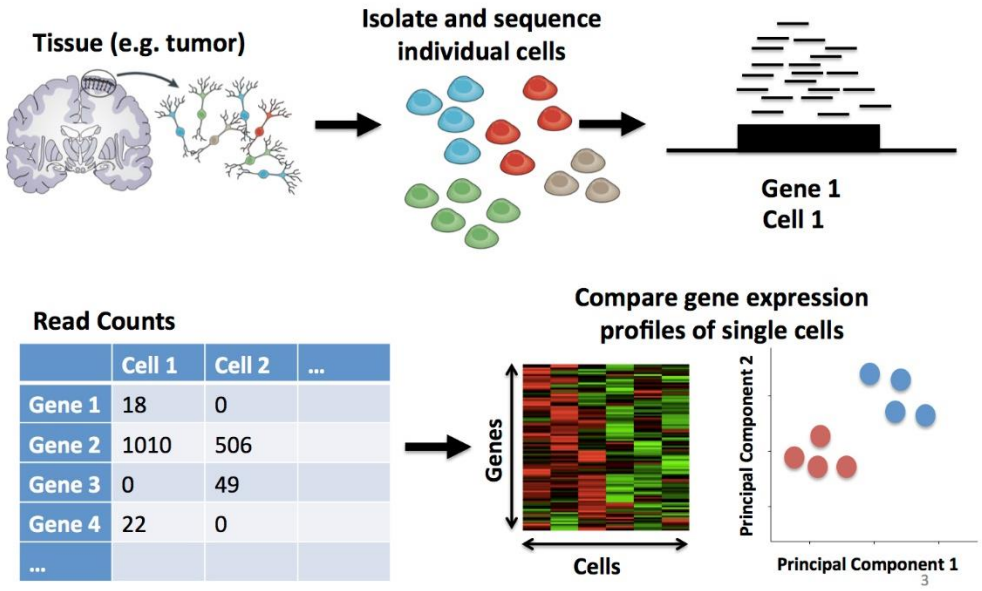
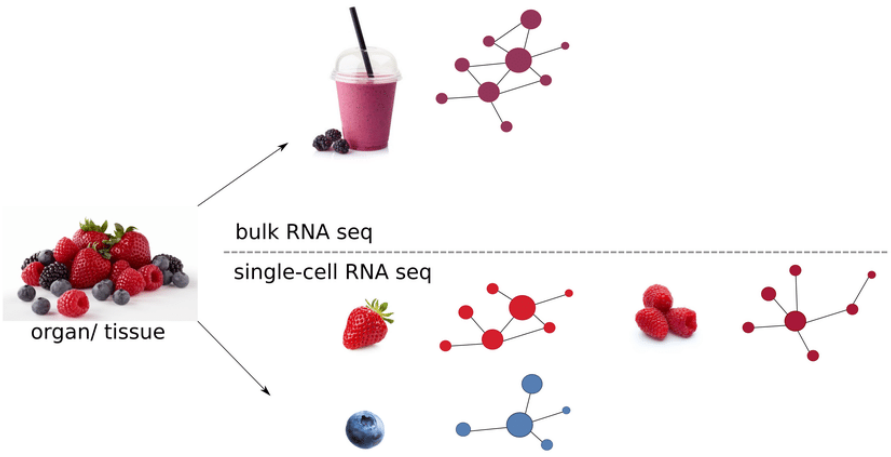


Figure 9: Single-cell RNA-sequencing workflow. Source[K]

[K] https://learn.gencore.bio.nyu.edu/single-cell-rnaseq/

III.2. Gain of resolution associated to single-cell approaches

As seen previously, genomics assays were usually performed in bulk, i.e. from a mix of cells leading to a loss of cell specific information. Therefore, only giving us access to an estimate of the predominant features from the whole cell population failing to highlight the cell type specific features and associated regulatory network (Figure 10). Bulk approaches also failed to assess if difference between two conditions reflects a change in cell population or a feature change within a same cell subpopulation which are two separate mechanisms essential to adequately understand disease etiology. Before the emergence of single cell technologies, the study of tissue heterogeneity was limited to cell surface markers-based phenotyping. Such studies are supervised relying on a limited number of cell surface markers chosen by researchers which is likely to affect the resolution of the study. With single cell genomics assay, gene or whole genome features are assessed in an unsupervised way allowing objective cell subpopulation definition and heterogeneity studies. Furthermore, molecular characterization of cell fate decision, an inherent unicellular process, was not possible with bulk genomics approach because only average cells states were captured. With single cell genomics, it is now possible to capture every cell state in the process of differentiation, allowing differentiation program characterization at transcriptomic and epigenomics level. Such gain of objectivity and resolution has already revolutionized our understanding of developmental processes and cell fate decisions [113,115] , but also of disease related cellular and molecular mechanisms[116,117]. For example, in a pioneer study in AD brains using single cell genomics, an AD specific microglial cell type was discovered (and molecularly characterized),a discovery that would not have been possible with previous supervised and/or bulk analysis[105].



40

Figure 10: Gain of resolution obtain with single-cell RNA-sequencing compared to bulk RNA-seq exponentional increased of ACDs incidence with age. Reprinted from Steinheuer et al, biorxiv, 2021[L]

III.2.a. Single-cell transcriptomic

Single-cell transcriptomic assay (scRNA-seq) follows an exponential trend of usage since few years, as illustrated by pubmed results: at the date of writing (15/08/2022) 8021 PubMed articles contain “single-cell RNA-seq” in their title or abstract with 2420 articles only for the year 2021 (more than 6 per day!).

III.2.a.i. Tissue heterogeneity

scRNA-seq allows the studies of cellular heterogeneity within tissue at transcriptomic resolution with the all transcriptome to be interrogated simultaneously in each cell increasing considerably the resolution in comparison of what was previously used to classify cells i.e. morphology, physical and chemical property, and cell surface markers where only about 40 markers can be interrogated simultaneously. This gain in resolution required the establishment of new reference for cell type definition. The Human Cell Atlas consortium (HCA) is one of these initiatives that aims to build a high-quality human cell atlas referring every human tissue and heterhogeneity[118] at molecular level. These atlases could also provide insight about transcriptional program and molecular pathways activated in these different cell types. For example, scRNA-seq study of pancreatic islet has highlighted a rare population of alpha cells able to proliferate through activation of the Sonic hedgehog-signaling[119]. Such finding would not have been possible with bulk RNA-seq. Furthermore, it highlighted putative pathway to target for alpha cells proliferation regulation. Numerous scRNA-seq studies were performed in brain, allowing the identification of novel neuronal or non-neuronal subtypes, with specific neuropeptide or receptors expression, as well as distinct transcriptional program driving central nervous system development[120,121].

III.2.a.ii. Differentiation process

scRNA-seq captures cells at various differentiation levels offering the possibility to analyze the differentiation process at a resolution never reached before. In hematopoiesis, it was shown that the differentiation process from multipotent HSC to lineage restricted progenitors doesn’t followed a step by step process as previously modeled but rather followed a continuous process with high transcriptomic variability within HSC and Multipotent Progenitors (MPP) cells highlighting gene expression stochasticity in these cells [122,123]. These scRNA-seq results further confirm previous studies

[L] https://www.biorxiv.org/content/10.1101/2021.04.02.438193v1.full

which demonstrate the role of gene expression stochasticity in cell's fate dynamics in multipotent stem cells differentiation and self-renewal balance[124,125]. Furthermore, scRNA-seq analysis allows differentiation trajectories analysis. For example, using scRNA-seq and a graph based trajectory construction tool, a recent article has highlighted 56 different cell differentiation trajectories involved in mammalian organ development [126]. Cell differentiation trajectory can also be estimated thanks to transcriptional dynamics across cells based on ratio between pre-mRNA and mature mRNA in each cell, i.e. the RNA velocity. RNA velocity analysis has notably shown unexpected transition between two immune cell types following severe COVID-19[127] .

III.2.a.iii. Cell to cell communication

As described in part II, cellular heterogeneity contributes to an asymmetrical cellular cross talk, crucial in physiological functions regulation. scRNA-seq enable the study of this cell to cell communication thanks to its ability to catch Ligand and Receptor co-expression in the different cells. This analysis has notably shown a disease specific cross talk between choroid plexus epithelium and brain astrocytes as well as oligodendrocytes and microglia in the brain of severe COVID19 patient[128]. Similarly, a disease related crosstalk between smooth muscle cells and fibroblast was observed in coronary artery disease[129]. Furthermore, a TNF-α mediated autocrine microglia activation as well as a TGFβ2 mediated regulation of microglial activation by neurons was demonstrated at early stage of diabetic retinopathy [130]. In AD, researchers have also identified a TGF-β mediated overstimulation of perivascular fibroblast driven by other cerebrovascular cells[131].

III.2.a.iv. Functional studies at single-cell resolution

Genetic editing coupled with scRNA-seq can decipher the role of genetic element in tissue function and disease development at cellular level, highlighting cell type specific transcriptional alteration and putative functional consequences. In a recent remarkable contribution, scRNA-seq was used both to identify T2D specific regulatory networks in pancreatic islet and a master regulator, BACH1, driving the metabolic inflexibility and endocrine progenitor/stem cell features of a T2D-specific subpopulation[132]. Authors then showed that a knockout of this master regulator reverse the T2D specific cellular features up to a non-diabetic phenotype. In another outstanding article[133], APOE4, the main genetic risk of AD, has been induced in a mouse model of AD, and reveal that its selective removal in astrocytes was able to decrease AD signature in astrocytes but also in neurons, oligodendrocytes, and microglia.

Other advantage of scRNA-seq analysis is that we can leverage the single cell resolution to perturb several gene in one assay. Perturb-seq allows to study the function of several genes at the same time in a tissue and cell type specific manner[134] . Perturb-seq is based on pooled CRISPR screen

with scRNA-seq read out. Dozens of sgRNA can be designed to target specific genes and transduced in limited dilution to have one or two sgRNA by cells, allowing single perturbation at single-cell resolution. These transduced cells can then be processed through any high throughput scRNA-seq methods, but droplet based is more appropriate because it allows the sequencing of a greater number of cells increasing the resolution of the approach. This method was used to study the impact of 200 oncogenic variants on lung cancer cells[135]. The authors were able to classify variants into gain of function, loss of function or dominant negative variants and discovered that KRAS variants span a continuum of gain of function phenotype rather than a discrete functional alteration. Considering TFs perturbation, perturb-seq enable the study of TFs downstream target genes in a cell type specific manner as demonstrated by a pilot study focused on TF regulating dendritic cells to lipopolysaccharides[134]. They also demonstrated the ability of this method to infer cell type specific TFs associated regulatory network.

scRNA-seq can also be used to assess the cell specific effect of a drug or a targeted therapy. In addition to assess the effectiveness on a specific cell type, scRNA-seq has the key advantage to also measure putative side effect or off target effect on other cell types. For example, drug use to mediate FOXO inhibition in pancreatic islet was shown to also induce dedifferentiation of both alpha and Beta cells[136]. In addition, effect of morphine in brain cell type was assessed showing an oligodendrocytes specific cell response never observed before[137].

III.2.b. Single-cell epigenomics

In parallel of scRNA-seq, other single-cell genomics assays were developed focusing on other layer of the regulatory landscape, ranging from whole DNA sequencing to epigenomics assay. Even if some single-cell DNA methylation and Histone marks profiling were developed, because of their single molecules level, loss of material and contamination or measurement error can have a strong impact. Then generating such data produce typically high noise and dropout rates (zero inflated data due to missing value) limiting their application[138]. Still, considerable effort has been made in assessing chromatin accessibility at single-cell resolution[139]. Remodeling of chromatin accessibility is a key epigenetics mechanism regulating cell type specific gene expression and cell differentiation[140]. Epigenetics modifiers like DNA Methyltransferase (DNMT), Histone Deacetylase (HDAC), and Histone methyltransferase (HMT) influence chromatin accessibility. These epigenetics modifications typically took place on enhancer or promoter region, to regulate the binding of TF or the transcriptional machinery. Then, assessment of open chromatin region at single-cell resolution (scATAC-seq) is a great opportunity to decipher the cell specific regulatory landscape. For example, based on cell specific chromatin accessibility, 12 different cell clusters were found in pancreatic islet, including several alpha, beta, and delta cell states[141]. It is also very useful to better understand the gene regulatory network

governing cell transcription/activity and programing. Furthermore, through its key role in gene expression regulation, chromatin accessibility assessment can be used as an alternative to estimate gene expression in tissue where RNA collection is very challenging because instable and degraded, while DNA is stable and easier to collect[142].

ScATAC-seq droplet-based method are similar to the one for scRNA-seq assay except that accessible DNA are captured instead of RNA (Figure 11). Before encapsulation, cells (or nuclei) are transposed using the mutant transposase tn5, which, as for classical bulk ATAC-seq method, cuts DNA in accessible region while adding a tag. After cell/nuclei encapsulation, this tagged accessible DNA fragments can be amplified with addition of a cell specific barcode.

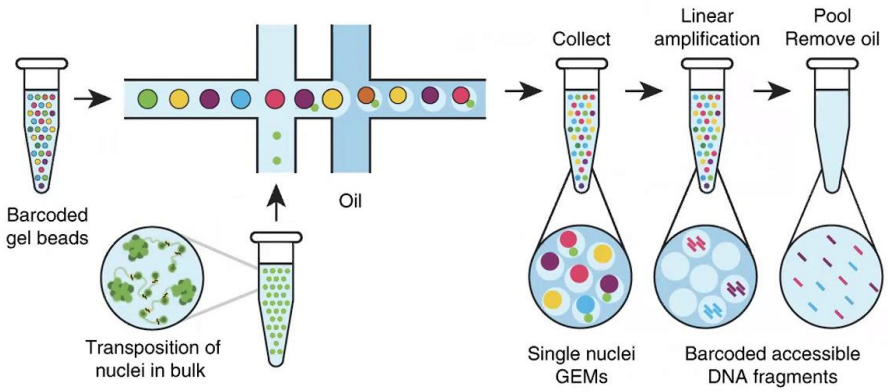


Figure 11 : Droplet based single-cell ATAC-sequencing workflow. Adapted from Satpathy *et al*, nature biotechnology, 2019[M]

Through their ability to bind with TFs, open chromatin regions are putative cis-regulatory elements (CREs), e.g. promoter or enhancer, regulating expression of neighborhood genes. Then, cell type specific open chromatin regions identification allows the discovery of cell type specific CREs. Mapping these CREs across cell type is then crucial to refine functional consequences of a variant or epigenetics alterations on gene expression. Indeed, a variant or epigenetic alteration like DNA methylation falling in pancreatic Beta cells specific CREs are likely to alter specifically Beta cells expression. This can be very useful for functional interpretation of GWAS variants that often fall in non-coding region[143,144]. For example, in T2D, scATAC seq has allowed to demonstrate that the causal T2D variant at the *KCNQ1* locus targets specifically Beta cells specific enhancer[141]. Single-cell chromatin accessibility landscape was assessed on healthy brain allowing functional characterization of AD and

[M] https://www.nature.com/articles/s41587-019-0206-z

Parkinson diseases associated non-coding SNPs[145]. They were able to link them to cell type specific target genes and predicts regulatory effect of MAPT (encoding tau) risks variants.

III.2.c. Single-cell multimodality

Recently new single-cell assays have been developed to simultaneously measure different modalities in the same cell allowing the investigation of interaction between different regulatory layers. Interrogating protein level as well as RNA level at cell level highlight correlation between gene transcription and protein expression. This is possible with CITE-seq technics, which, in addition to capture mRNA, label cell surface protein with oligo linked antibody. The oligo will be subsequently sequenced to estimate protein level. However, this technic still relies on a limited number of proteins (~200s for cell immune profiling), and only cell surface markers, limiting their scale.

Another important development is the assessment of chromatin accessibility and RNA expression within the same cell. This approach allows unbiased assessment of interaction between chromatin remodeling and gene expression at genome-wide resolution. In particular, we can better identify genomics region influencing gene expression, i.e. CREs. Indeed, if chromatin accessibility at one genomics region correlates with expression of neighbor gene at cell level, it is likely that this region influences its expression. Then, we can identify which TFs are likely to bind these CREs using TF motif analysis and TF footprint analysis. By this way, we can estimate the TFs regulatory activity on downstream target gene and infer a TF-gene regulatory network at subpopulation level. This approach can be very useful to demystify the gene regulatory network involved in differentiation process or in disease development. Few studies have been performed to date because of their recent accessibility, but they already impacted our understanding of cell type or diseases specific gene regulatory network identification[146,147]. Thanks to this approach, it has been shown that TF expression and TF motif accessibility correlate and their activation precede transcriptional expression of targeted genes, highlighting the key role of chromatin accessibility in programming differentiation of cells[140,148]. Single-cell multiomics analysis of AD brain has shown a SREBF1 regulatory network alteration in oligodendrocytes of late stage AD[146]. Single-cell multimodal analysis on arthritic fibroblast shows conserved disease specific gene regulatory networks regulated by NFkB and new candidates, including Runx1[147].

III.2.d. Other single-cell genomics related approaches

Others single-cell genomics approach exist like single-cell whole genome[149], single-cell immune profiling with V(D)J B cells or T cells receptors screening [150], allowing diverse others applications. Until recently, protein binding and DNA methylation profiling at single-cell resolution lacked robust assay.

However, recent new technologies emerged trying to overcome these issues[151,152]. scCut&Tag assay, as an extension of scATAC-seq assay, allows to map the location of specific DNA binding proteins at cell level. To do that the ATAC transposase is modified to bind antibody prealably fixed to specific histone marks or DNA binding protein. Transposase will then preferentially cut and tag DNA regions specific to those binding proteins.

Even if a commercialized kits are not yet available and dropout rates are still an issue, several methods exist to profile DNA methylation at single-cell resolution[153–155]. One of the most promising methods is scNMT-seq that enables joint profiling of chromatin accessibility, DNA methylation and gene expression in single-cells. Following cell isolation and lysis in single well, cytoplasmic RNA is isolated from nucleic DNA to assess cell transcriptome profile using conventional Smartseq2. Then remaining nuclei DNA is treated with a GpC methyltransferase, which catalyze cytosine methylation only in accessible DNA region. Because in mammal, most of the cytosines are not methylated excepted in CpG context, this step allows the labeling of accessible region by methylating accessible cytosine. Then, this labeled DNA is bisulfite converted allowing both assessment of GpC converted accessible region and endogeneous CpG methylation in parallel. This assay has been implemented in mouse embryonic stem cells, and further apply to a study of mouse gastrulation[156]. In this last paper, Arguelaget *et al* have shown that first exit of stem cell pluripotency coincides with establishment of repressive epigenetics marks and followed by lineage specific epigenetics pattern. Then, even if such assays are still in their infancy, these results promise great future in the understanding of the role of DNA methylation in developmental processes but also in cell type specific epigenetics mechanisms.

Other single-cell omics methods are still in development, like single-cell proteomics[157] and metabolomics[158] with exciting perspectives for our understanding of cellular mechanisms involved in diseases development.

III.2.e. Spatial transcriptomics

An important emerging single-cell related assay is spatial transcriptomics[159]. Spatial transcriptomics allows the interrogation of *in situ* cellular heterogeneity by measuring thousands of gene expression at cell resolution without loss of the tissue structure. Different methods exist based on imaging or on high throughput sequencing. First methods developed were imaging based, relying on fluorescence *in situ* hybridization (FISH) where fluorescent RNA probe allow targeting of cellular mRNA [160] or in situ sequencing where amplification of retrotranscripts mRNA and sequencing are performed directly on sliced tissue[161]. These methods have several advantages including higher resolution (100 nm) and sensitivity but have a limited gene throughput (even if last methods can access to 10k genes[162]) as well as limited feasibility due to single-molecule imaging.

High-throughput sequencing assay are based on array covered by geolocalizable oligo barcode (spatial barcode), which retain RNA location within the tissue prior to tissue dissociation and library preparation for standard illumina based sequencing[163]. These methods are unbiased, capturing all polyadenylated transcripts and giving the whole cDNA sequence information, interesting for splice isoforms, single nucleotides variants or mutations detection. However, there still have a limited resolution and sensitity, even it most recent in-development method argue to reached a spatial barcoding of 1µM resolution and about 100 unique transcripts per µm²[164,165]. A commercialized version of this method exist, developed by 10X Genomics with their Visium, which have a spot resolution of 55µM diameter[166].

Spatial transcriptomic technologies have the potential to generate unbiased picture of tissue composition, allowing the establishment of tissue atlases and reference maps. They have already revolutionized the analysis of nervous system with several studies highlighting spatial transcriptomics maps of the entire brain or of specific regions[167–172], with specific insight on neurological disorders like autism or schizophrenia[172] .

Others biological fields have largely benefited from this enhance technology including developmental biology to elucidate spatial dynamics of heart development, spermatogenesis, and intestinal development[173,174] but also for studying tissue disorganization in disease[175,176]. In AD, this technology has already revealed that genes modulating stress response are spatially differentially regulated in hippocampi and olfactory bulbs, with notably Bok, being spatially downregulated in hippocampus of mouse and human AD brains[177]. Another study has found early alterations in a network enriched for myelin and oligodendrocytes genes around amyloid plaques, while a network enriched for plaques induced genes related to oxidative stress, lysosomes and inflammation in later phase[178].

III.3. Computational challenges

Analysis of single-cell genomics data requires important computational skills and remains quite challenging.

III.3.a. Single-cell data analysis pipeline

High-througphut sequencing generates thousands of reads per cells that need to be mapped to a reference genome to generate a gene-cell count matrix, representing the number of transcripts detected for each gene in each cell. Excepted the need for computational resources, generating this count matrix is trivial because it relies on ready to use bioinformatics pipelines.

The next step after generating this count matrix is the cell cluster generation based on transcriptome similarities (Figure 12). Because data are at genome wide and cell level, the associated matrix is of high dimensionality (typically ~30k genes for human transcriptome) with numerous observations (n= the number of cells). High dimensional data are challenging to appropriately cluster. Furthermore, because data generation relies on little amount of genomics material, single-cell genomics data represents a sparse matrix, with lot of zero, making it even more difficult to cluster properly. Therefore, there is a need to reduce the dimensionality by looking at common variability between genes. Several methods exist to reduce dimensionality including the classical principal component analysis (PCA) or more sophisticated one including latent semantic indexing (LSI).

PCA is a widely used reduction dimension method trying to maximize variability explanation in a limited number of dimensions, the principal components, using orthogonal vectors. It is quick to compute and give a linear projection of cells on principal components that explain the greatest covariance of genes across cells. However, PCA is a linear reduction method so does not catch nonlinear variability/pattern. PCA reduction is mostly used for scRNA-seq data as a first step reduction method. LSI is a dimensionality method allowing to give higher weight to rare feature mainly used to reduce scATAC-seq data[179,180]. Such reduction method typically reduces the data to ~50 dimensions, keeping the linear structure of the data while removing the zero biased, thus facilitating cluster identification. Clustering usually implements a graph-based method relying on shared nearest neighbor graph. It links cells according to their proximity in the dimension reduced space and further refines their link weight based on their mutual cell neighbors. Once this graph is produced, graph-based clustering algorithm, like the Louvain algorithm[181] aiming to optimize the modulatory (i.e. module/cluster of highly connected cells), is performed to produce a clustering of cells reflecting their transcriptomic similarities relative to others cells. After this clustering step, cells cluster can be annotated thanks to identification of cell type specific markers. This step requires manual curation and knowledge about the interrogated tissue/sample. It is now also possible to annotate your cells based on scRNA-seq references, highlighting the interest of the human cell atlas and similar initiative[182]. To note, because cell type definition is dependent of the transcriptional profile in scRNA-seq, it is sometime hard to estimate if change between two conditions rely on cell subpopulation difference or rather reflect cell activity change. For this reason, it is important to define the cell type prior to perform differential expression analysis.

Once subpopulations are identified, it is then possible to compare different conditions at subpopulation level. Two main analyses are classically done: differential expression and differential cell-type abundance. Differential expression can be performed within each subpopulation to identify cell-type specific transcriptional alterations. This step is not trivial as gold standard has not yet emerge

for the statistical part and the analysis will be highly dependent on your sample design. Indeed, to avoid technical bias or inflated p-value due to large number of cells, it is recommended to do this analysis at pseudo bulk level. To do that, single-cell count is summed for each sample replicates within each subpopulation producing a sample gene count matrix by subpopulation of interest. However, we need to have enough sample replicates to perform such analysis (at least N>3), which is often not the case due to the experiment cost and time. In the other case, differential expression analysis can be performed at cell level. In this case, the first step is to normalize for cellular sequencing depth and stabilized for variance using a regularized negative binomial model, to reduce technics dependent bias[183]. Then, a standard Wilcoxon rank-sum test can be used to highlight genes differentially expressed in your tested condition without making assumption on the sample distribution.

Difference of cell type abundance between conditions can also be analyzed. Accordingly, to previous part, depending on the number of biological replicates, the tests to use will be different. Chi-squared test can be used if comparing two proportions with just one replicate by conditions, assuming that these proportions represent the whole population. Otherwise, if multiple replicates are available, a Wilcoxon test should be performed to compare cell type proportions between conditions. There are tests that have been developed specifically for single-cell data[184,185], including Milo, a statistical framework that used cell-cell similarity k-nearest neighbor graphs, which have shown better performances than alternative methods to perform differential abundance testing. This method enable the identification of perturbation in cell composition that are hidden when discretizing cells into clusters, identifying notably the decline of fate-biased epithelial progenitors in aging mouse thymus.

Then, several others downstream analysis can be performed as explain in previous parts (Applications parts), including pseudotime analysis, RNA velocity, and TF activity measurement, each with specific set of statistical considerations and challenges that I won't develop here.

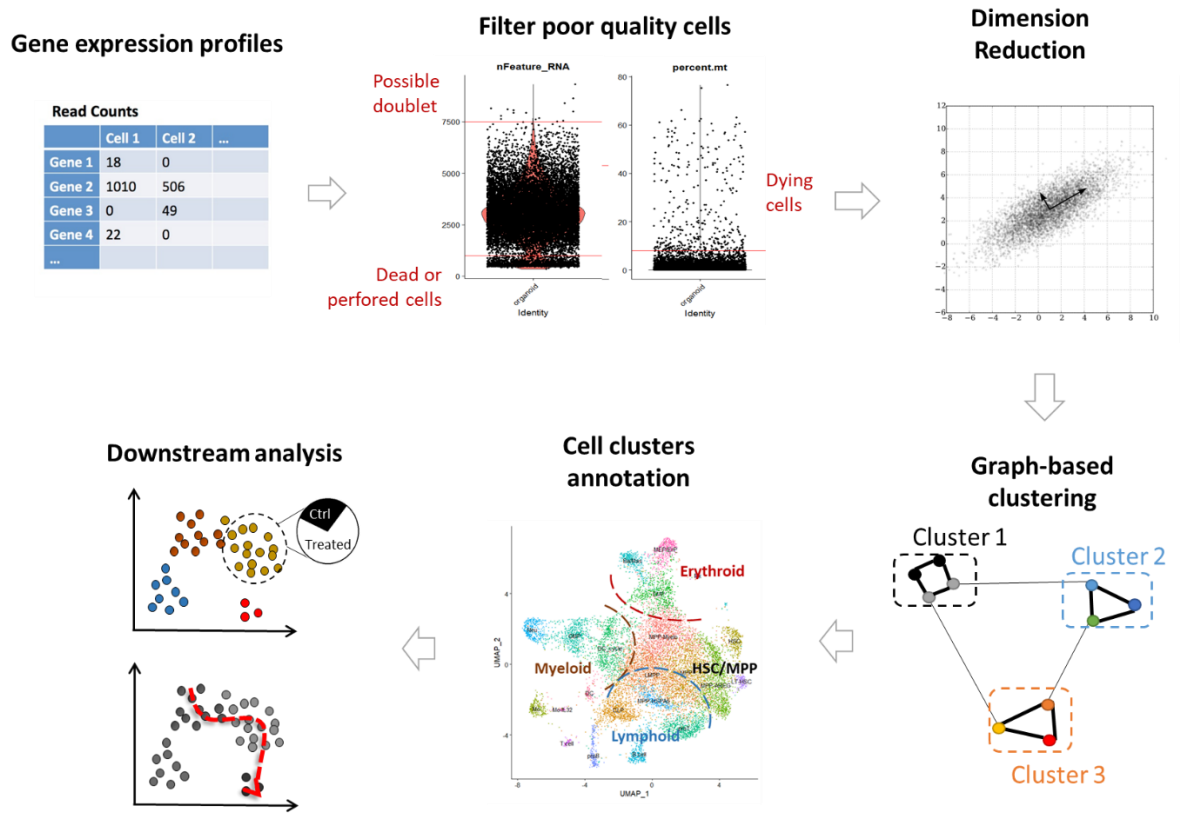


Figure 12 : Classical single-cell data analysis pipeline

III.3.b. Integrative analysis

An important other challenge, maybe the major one, is data integration.

III.3.b.i. Batch integration

First level of integration is considering same modality datasets from different experiments / laboratory / sample / species. The challenge is then to remove batch dependent technical bias while conserving the biological cell type specific variation. Several methods have been developed to perform this task with different efficiency depending on the complexity of the datasets to integrate[186]. Most popular tools are based on mutual nearest neighbors (MNN) graph which will link cells from different dataset according to their common cell neighbors in common low dimensional space[187,188]. Another widely used method is Harmony, which corrects directly in low dimensional space the cell position to remove batch specific variability[189]. For more complex integration task (e.g. multiple species integration) deep learning based method like scVI or scGen are more efficient, because they can adapt for nonlinear variation[186] .

III.3.b.ii. Single-cell multi omics integration

A second level of integration is the integration across multiple modalities. First of all, between scRNA-seq and scATAC-seq data. Similar tool can be used if scATAC-seq data is also considered at gene level but this will result in a loss of scATAC-seq specific information, which reduce biological significance. Further methods have been developed to perform this specific task based on gene and chromatin accessibility co-variability[142] , or on reference multimodal data[190] but a lot of work remains to be done to effectively integrate both modalities.

The single-cell multimodal assay allows simultaneous profiling of epigenetic and transcriptional landscape in the same cells, which avoid the integration step and enable new applications and challenges. The main challenge is then to correctly associate an open chromatin region (a peak) to a neighborhood gene expression. In other words, the interest is to find which peak (putative regulatory region) regulates which gene, and in which cell type. If these links can be found, then we could define cell type specific CREs, and predict gene expression only based on open chromatin data, and inversely. This problem is still challenging as I am writing this thesis[191]. For the moment, only few methods have been proposed and are still trivial. SHARE-seq article proposes to simply link peak to gene based on the spearman correlation and compared against randomly selected features matching genomics region, to evaluate its statistical significance thanks to a ground truth[148]. This approach can evaluate if the peak accessibility explains gene expression but miss the co-variation with others peaks and the sparsity of the chromatin accessibility data, which can hide some complex associations. ArchR a software design for single-cell analysis of regulatory chromatin used nearest neighbor method to group resembling cells and then merge count by group to reduce the sparsity problem[192]. However, more sophisticated or clever methods should emerge to better associate these two modalities. Finally, another promising avenue with this single-cell multimodal data is to infer cell type specific as well as disease relevant gene regulatory network[193–196] , but we are still in the very beginning.

III.3.b.iii. Integrate single-cell with bulk data

Another challenge in the genomics field is to integrate single-cell data with bulk data. Deconvolution methods exist for bulk transcriptome in order to find single-cell composition[197] but they rely on cell specific datasets references (scRNA-seq or bulk RNAseq on isolated subpopulation). In addition, integration of different modalities from different resolutions appears more challenging. Typically, genome wide DNA methylation are performed in bulk due to the limited recovery of actual single-cell methylation assay[155]. If cell type composition is known, deconvolution can be performed to fit with scRNA-seq data. Otherwise, linear comparison with every subpopulation expression could be done to evaluate cell type specific DNA methylation impact. In any case, a critical step is the need to

integrate DNA methylation with transcriptomic data. DNA methylation is at CpG level and need to be compared to transcriptomic, which is at gene level. Several approaches can be used to do that but are still biased. In my first model I will develop a method trying to improve this link, integrating both TSS distance, tissue specific chromatin profile, and eQTL information.

III.3.c. New tools

To help overcome all these challenges, new computational tools are emerging. Most specifically, deep learning-based framework are very promising to perform complex task[186]. Deep learning is based on *in silico* neural networks resembling to the functioning of the biological one for lot of aspect. They allow complex task integration better than standard statistical tool because they can catch nonlinear pattern in a semi-automated way, and at high scale[198,199]. Most of these methods aim to reduce the dimensionality of the data like PCA or LSI but this time by extracting more abstract features, that could be shared by different omics layers. Like for feature extraction from pictures, where deep learning tools have allowed huge advances in identifying objects, animals, or human faces in very different context, this tool start to be applied for biological feature extraction and promise great advance in this field. scVI is a widely used deep learning tool that reduces the technical bias inherent to single-cell data while accounting for batch effect, allowing to efficiently integrate heterogeneous single-cell datasets[200]. Other deep learning-based methods have been proposed then integrating other neural network architecture including generative adversarial network, claiming to improve discrimination of batch effect[198,201,202].

In this thesis, I will present how I took advantage of single-cell genomics data using preexisting tool and developing new approaches in order to decipher cell type specific molecular mechanisms involved in adult chronic diseases development. To do that, I focus on two models: the epigenetics programming of hematopoietic stem and progenitor cells (HSPC) in the context of early programming of adult chronic diseases, and the Alzheimer's Diseases susceptibility genes BIN1 function.

IV. Model 1: Epigenetics programming of hematopoietic stem and progenitor cell (HSPC)

IV.1. Early Programming of chronic metabolic diseases

Chronic metabolic diseases, including type 2 diabetes (T2D), obesity and cardiovascular diseases (CVD) constitute approximately 70% of deaths worldwide (WHO, 2017), thereby becoming the most significant burden to healthcare systems. Although inherited genetic risk and lifelong environmental exposure contribute to their development, they cannot explain alone the distressing rise in obesity and diabetes of these recent years[6,203]. Several epidemiological and experimental studies indicate that perinatal exposure (fetal and early development, up to 1000 days after birth) to a metabolic stress increase susceptibility to the development of chronic complex diseases several decades later. Perinatal development is a critical period of rapid growth and differentiation when organs shape and acquire their function. It is then a period of intense epigenetics remodeling and sensibility to environments which can have durable impact on organ structure and function[204].

Our society have known radical changes this last century, notably in food industry, robotization and tertiary deployment, which have significantly modify our lifestyle and influence our exposure to nutrient as early as *in utero*. In 2010, more than half of pregnant women were considered obese in US [205]. In France, a 90% increase of obesity rate among adult woman have been observed between 1997 and 2012[206]. Incidence of gestational diabetes mellitus (GDM), corresponding to high blood sugar that develop during pregnancy but usually disappearing after giving birth, have also considerably increase. The past decade, the incidence increases by 30% in young US women, while doubling for some population including Asian Indian[207]. In Europe, prevalence of GDM reaches 11% of total pregnant women. However, impact of this recent change on fetus development and long-term consequences remain poorly studied.

Yet, Barker and colleagues were the first to demonstrate in 1986 that an early (fetal) exposure to a nutrient stress increases the risk to develop diseases decades' later [208]. They first observed that English regions which were the most impacted by starvation and infant mortality in 20th century were also the regions the most affected by CVD decades later[208]. Following this first observation, two large studies were led by Barker *et al.* to investigate this link and found concordantly a strong association between low birth weight, head circumferences or ponderal index, and death from coronary heart diseases and T2D decades later[209–211]. This observation was replicated in 3 others countries [211–213] and show that maternal undernutrition conditioned progeny to future environmental fitness and diseases susceptibility. These observations have opened the field of the developmental origins of health and diseases (DOHaD), which study how early exposures conditioned people to adapt, or mis-adapt, to

future environments. Numerous others epidemiological, animal, and transversal studies succeeded to better understand this association and the mechanism behind the apparent early programming of ACDs.

IV.1.a. Epidemiological evidence

IV.1.a.i. Birth weight

Several evidence have shown that both extreme of fetal growth increase the risk of ACD. To show that, researchers have studied diseases rate in small for gestational age (SGA) or, large for gestational age (LGA), compared to appropriately grown neonates. SGA and LGA being defined respectively as neonates under the 10th or over the 90th percentile birthweight and ponderal index (PI= weight / height^3) adjusted for gestational age and sex. Epidemiological studies have found that in addition to coronary heart diseases, SGA are associated to an increased susceptibility to numerous diseases, including hypertension [214], type 2 diabetes[215], stroke[216,217] , dyslipidemia[218] , and impaired neurodevelopment [219]. On the other sides of the birth weight spectrum, LGA have an increased risk of ACDs and related comorbidities including cancer[220] , obesity[221], metabolic syndrome[222], T2D[223] and CVD[224] as well as increased risk for neurologic disorders including depression, anxiety, autism, and cognitive delay[225–228] . Then, these results show that both restriction and excess of fetal growth is associated with increased ACD susceptibilities with similar outcomes(Figure 13), suggesting converging mechanisms.

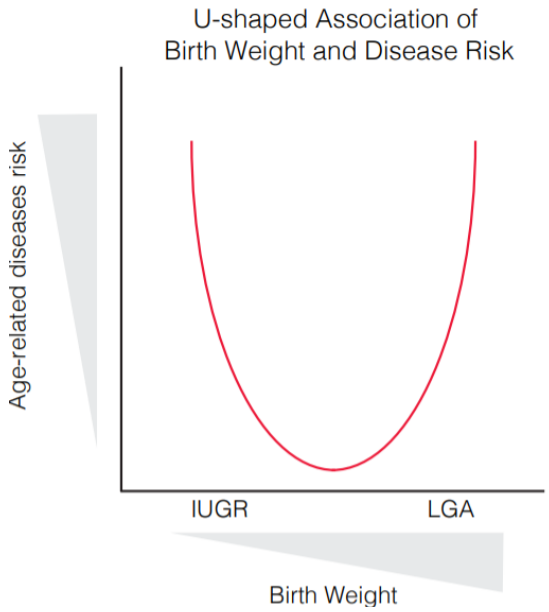


Figure 13: U shaped association between birth weight and ACD risk

Several theories emerged to explain these links. The thrifty phenotype hypothesis, first formulated by Barker *et al.* in 1992[229], stipulates that association between SGA and ACD risk results

from irreversible alteration in the glucose-insulin metabolism during development. Maternal undernutrition led to a decreased insulin secretion by the progeny and to an increased peripheral insulin resistance[230] , leading to a greater glucose availability for brain and heart. If nutrients become abundant in postnatal life, the pancreatic Beta cells will defect and insulin resistance of peripheral tissue could then predispose to glucose intolerance and diabetes, as observed in SGA neonates with rapid catch-up growth who have more chance to develop insulin resistance and T2D later in life[231,232]. A competitive hypothesis based on genetic influence was developed by Hattersley [233]. Indeed, the link between SGA and glucose intolerance/diabetes in adulthood could be explained in part by genetics factors, for example variants influencing insulin secretion, that can contribute to a decrease in birth size and glucose tolerance[234–236] . However, genetics variants associated to birth weight through GWAS explain only about 7% of the birth weight variation, still supporting an independent influence of prenatal exposure affecting birth weight and ACDs susceptibilities[237] .

Gluckman *et al* tried to generalize the Barker's hypothesis bringing an evolutionary point of view. They argue that these associations between prenatal exposure and ACD susceptibilities could reflect a (failed) predictive adaptation to future environment, i.e a way to have a fitness advantage on expected future environment[238]. To support this hypothesis, evidence of *in utero* adaptation to expected similar future environments exists[239–242] . Notably, it was shown that offspring from rat fed with HFD develop hypertension and endothelial dysfunction. . However, the endothelial dysfunction was prevented if offspring were kept on HFD during early life[239]. In human, poor *in utero* environment induces morphological and physiological changes like fat deposition, which promote future survival in deleterious environment[240] . However, if the structural/functional choice made during development ended up inappropriate in regard to the future environment, there is a mismatch between tissue adaption and reality, which increase the disease susceptibility. Several observations have been made in this sense, notably when *in utero* growth restricted individuals like SGA have postnatal environment favoring overconsumption, leading to further glucose intolerance, insulin resistance, and reduced lifespan in human or animal models[238,241,242].

IV.1.a.ii. Maternal hyperglycemia

Gestational Diabetes Mellitus (GDM) or maternal hyperglycemia during pregnancy is a current common complication (prevalence was about 17% in 2013[243,244]) . GDM is strongly associated with increased birth weight [245–248] with persistent elevated glycemia (HbA1c ≥ 5.6%) at 3-month pregnancy leading to LGA in 26% of case[248]. The association is even more pronounced for type 1 diabetes mothers where the occurrence of LGA is about 56% in a Lille hospital retrospective study[249]. The *in utero* exposure to hyperglycemia is associated to increase childhood cardiometabolic risk in offspring including higher rates of impaired glucose tolerance, obesity, and higher blood pressure. Importantly,

these associations are independent of BMI before pregnancy, being born large for gestational age, and childhood obesity, highlighting the direct effect of the *in utero* exposure to hyperglycemia[250] . Similarly, a previous study has shown in Pima Indian cohort, that T2D mothers lead to 45% of T2D in their offspring at age 20-24years old, while only 1.4% for non-diabetic mothers[251] . The risk persists even after correcting for paternal diabetes, age of diabetic onset, and offspring BMI. Another study showed that offspring exposed to T2D during gestation have a higher risk to develop T2D than their siblings born before maternal T2D onset[252].

IV.1.a.iii. Maternal Obesity

In western countries, obesity prevalence is estimated around 30%, while 40 % of women are overweight during pregnancy[253,254]. Growing evidence show that this deleterious metabolic status has long-term consequences on offspring especially on adiposity, cardiovascular and metabolic risk. Maternal pre-pregnancy obesity is associated with significant increase risk of LGA [255,256] as well as with a 3-fold risk of childhood obesity[257], while maternal over weight gain during pregnancy is associated with increase BMI[258–260] . Furthermore, these evidence are associated with higher blood pressure, adverse lipid profile, insulin resistance and higher inflammatory markers in childhood even if those cardiometabolic risks can be partially attributed to the increase BMI of the child[260–264] . Finally, a follow-up study on 37,709 individual has shown that higher maternal BMI detected at first prenatal visit is associated with increased risk of premature all-cause mortality and hospitalization for CVD[265].

IV.1.a.iv. Glucocorticoids

Glucocorticoids exposure during development has been shown to be associated with an increased CVD risk and insulin resistance[266]. Glucocorticoids are key factors within the hypothalamic–pituitary–adrenal (HPA) axis contributing to stress response. Glucocorticoids have a key role in fetal development, particularly during the 3rd semester, where maternal glucocorticoids secretion regulates fetal growth, brain development, and organ maturation, allowing the fetus to prepare for extra uterine life. Glucocorticoids treatment during pregnancy reduces birth weight and lead to SGA associated ACDs risk as well as higher blood pressure in adolescent and altered neurological functions[267,268]. Furthermore, adult born SGA have altered control of cortisol expression and increased activity of the hypothalamic-pituitary-adrenal suggesting a programming of the HPA axis activity and regulation[269,270]. The altered cortisol level is associated with adverse metabolic profile (higher glucose, blood pressure, and dyslipidemia) in adult [270].

IV.1.a.v. Early life exposure

Not restricted to in utero exposure, evidence have also shown the influence of early postnatal life when differentiation and maturation of the tissues and cells are still intense. Psychological stress

during childhood increases the risk of ACD and are linked to macrophage pro-inflammatory tendencies [271]. Early life nutrition and weight gain have also an important role in the programming of ACDs. Fast weight gain has been associated to later obesity, cardiovascular diseases, while poor weight gain was associated with metabolic syndrome, impaired glucose tolerance and T2D later [231,272–277]. Breast feeding could also be associated to decrease risk of later metabolic disorders, but results are controversial because of important confounding factors [278–280]. The first 1000 days of life, are a critical period for tissue development especially for brain[281,282]. It is then a period of vulnerability, where deleterious exposure, including poor nutrition, exposure to toxicants and microbiota imbalance can have long lasting consequence on adult [283]. Improved nutritional income during these days improve cognitive function and school results. At the contrary, deficit in iodine during this period, which is a critical nutrient for brain development, impact future cognitive function even if iodine deficit is moderate[284] . Another important early life factor seems to be the microbiota colonization[285]. Main gut bacterial colonization occur in the first year of life driving by breastfeeding and other maternal transfer and is critical for immune system development[286,287]. Exploding evidence have highlighted the role of microbiota imbalance ("dysbiosis") on immune related diseases, including inflammatory bowel disease but also on metabolic disorder including obesity and T2D, and neurological disorders, including depression and anxiety, suggesting a role of early microbiota dysregulation in the programming of chronic diseases[286–288] .

In light of the impact of early development, clinical initiatives have been launched to better understand and inform on this critical period. In Lille, the program "1000 jours pour la santé" held by Laurent Storme aims to promote fundamental, clinical and technical research to better identify the critical early life factors influencing ACD development and better prevent them[289].

IV.1.b. Animal models to understand the physiological mechanisms

Several animal models have been developed to validate the influence of perinatal environment on future ACDs risk and better understand the biological mechanisms behind.

In rats, global maternal undernutrition, or specific protein restriction, result in reduce birthweight[290] , increased blood pressure[291], and impaired glucose tolerance[292] in the offspring in adulthood [293,294]. These results were reproduced in Guinea pig and sheep[295,296] . Putative mechanism behind the long term programming of glucose tolerance have been investigated in protein restricted pregnant rat model[293] . Such diet appears to reduce fetal pancreatic islet expansion leading at birth to reduce endocrine and Beta cells mass as well as reduced insulin secretion. Further studies have shown

that islet cells exposed to this *in utero* restricted environment have a decrease of replication rate *in vitro* , and Beta cells an even lower proliferation rate[297,298]. Interestingly, the alteration of insulin secretion and islet expansion are still apparent after 7 days of culture in normal metabolic environment, demonstrating the programming of these cells [294,299]. Long term consequences of such alteration appear influenced by several parameters, including the sex of the offspring, the developmental window targeted by the exposure, and the postnatal nutrition[300,301]. If the protein-restricted diet is present just during pregnancy, only female have reduced insulin response at 3 months old[294]. If restriction remains during lactation, plasma insulin is reduced even in adulthood and insulin response is greatly reduced in both sex[301]. Interestingly, such alterations are not associated with glucose intolerance, but at contrary, with greater glucose tolerances[301]. Further studies have shown that this appearing contradiction can be explained in part by the fact that peripheral tissue like liver, adipose tissue and muscle express more insulin receptors to compensate[302–305]. These evidence confirm the impact of maternal malnutrition on future cardiometabolic health with programming of metabolic circuits aspecially the insulin pathway. With the example of the endogeneous pancreas, these animal models studies also show that both tissue structure alteration (decrease Beta cells mass) and cell intrinsic factor (reduce Beta cell proliferation and insulin secretion ability) can mediate the long term consequences of early exposure to detrimental environment.

Further studies have assess the role of glucocorticoids signaling in these programming of metabolic risk[306,307] . Fetal exposure to glucocorticoid lead to decrease birth weight and increase blood pressure in sheep[308,309], while maternal dexamethasone intake (a synthetic glucocorticoid) in rats lead to a reduced progeny birth weight as well as to hypertension and glucose intolerance with possible insulin resistance in adulthood[310,311]. Furthermore, it was shown that maternal undernutrition increases maternal glucocorticoid secretion [312], while adrenalectomy, abolished effect of maternal low protein intake in the offspring, highlighting the key role of stress related glucocorticoid signaling in fetal programming of metabolic circuits[313] .

Interestingly, in both maternal low protein and caloric restricted diets rat models, impairment in glucose tolerance appear only following subsequent adverse life events or mis-adapted environment. This effect was notably demonstrated in the context of rapid catch up growth/ high food consumption in childhood leading to obesity[242] or during aging[314,315]. Indeed, offspring of rat fed with low protein diet show impairment in glucose tolerance only at around 15 months, and diabetes few months after[316]. Mechanistically, this can be explained because of the age dependent development of insulin resistance in peripheral tissue including adipocytes and skeletal muscle. This insulin resistance seems to develop from a molecular defect downstream of the insulin receptor which impaired PI3K kinase pathway activation[317]. These evidence show that *in utero* detrimental exposure like protein

restriction can predispose individual to have ACDs following future adverse life events, further supporting a reduced cellular plasticity to adapt to future environment.

Not restricted to peripheral tissue, exposure to maternal undernutrition appear also to program behavior related to the central nervous system. Indeed, Delahaye *et al* have previously shown that rat maternal undernutrition can durably program hypothalamic appetite regulatory system through a drastic decreased of leptin surge involved in the development of this system and a reduced responsiveness of anorexigenic POMC neurons[318,319]. In an opposite way, offspring of obese rat show an amplified and prolonged neonatal leptin surge and lead to a long term leptin resistance, which could explain the programming of hyperphagia and obesity in its animals, but also in human[320].

To remain on the other side of the exposure spectrum, the impact of over exposure to nutrients have also been shown to predispose to metabolic disorders. In rat, mild diabetic mothers lead to macrosomic progeny with increase pancreatic islets development and Beta cells mass due to hyperplasia and hypertrophy, while declare a glucose impaired tolerance later[321]. Similarly, maternal overeating lead to glucose intolerance on the offspring at 3month old[322], while maternal high fat diet lead to hypertension, leptin, and insulin resistance, as well as fat accumulation [323–325]. Furthermore, a recent study has shown that parental HFD or High sugar diet program inflammatory and oxidative parameters in reproductive tissue of rats offspring, highlighting putative mechanism of transgenerational transmission[326]. Interestingly, some of those impacts were sex specific with, for example, female offspring being more affected by hypertension[327] . Together, these findings strongly show that mother diets or metabolic status impact future offspring metabolic health with impact on central and peripheral tissues.

Thus, several *in utero* and early life factors can influence future adult disease risk, influencing metabolic parameters in early life. However, the lifelong molecular and cellular mechanism behind remain poorly understood.

IV.1.c. Epigenetics memory of early exposure

There is several decades between the fetal exposure and the associated ACDs onset, suggesting that early exposure results in long term tissue development alteration and/or decreased cellular plasticity. Such programming mechanism often relies, at least in part, on epigenetics modifications. Indeed, epigenetics mechanisms play a key role in mediating the influence of environmental exposures at cellular and molecular level. Diet, living place, drug treatments, or unhealthy habits are environmental factors known to influence epigenetics status [328]. Dietary restriction protects from age associated DNA methylation and induces epigenetics reprogramming of

lipid metabolism[329]. Work out influences DNA methylation, but also histone acetylation and miRNAs expression[330]. Six month aerobic exercise reshapes the whole genome DNA methylation in skeletal muscle and adipose tissue influencing lipogenesis, while miRNAs expression profiles allow discrimination between low and high responders to resistance exercise [330]. Chronic alcohol consumption leads to significant reductions in S-adenosylmethionine (SAM) levels, the metabolite substrate of DNMT enzyme, thereby contributing to DNA hypomethylation[331]. Smoking alters DNA methylation of numerous genes and disturb several miRNAs expressions [332]. In adult, impact of methyl donor deficient diets on methylation status and gene expression is partially reversible when the methyl donor is added back into the diet[333,334]. Several durable epigenetics alterations have been shown in individual exposed to *in utero* stress, when the epigenome is established supporting important role of epigenetics in the early programming of ACDs [335,336].

Evidence for an epigenetic programming of the metabolic syndrome have been shown in SGA. A global DNA methylation alteration was observed in adult offspring exposed to prenatal famine in Dutch Hunger Winter[337], with a decrease DNA methylation targeting the IGF2 gene [338]. An increase methylation and decrease expression of proopiomelanocortin (POMC) in cord blood, precursors of many metabolic hormones, have been associated with lower birth weight and with a higher triglyceride and insulin blood level during childhood, exposing a predictive epigenetic biomarker of future metabolic condition[339] . Abnormal birth weight is also associated with several durable epigenetics modifications in energy homeostasis genes, including DNA hyper methylation and reduce expressions of ATG2B, NKX6.1, and SLC13A5, related respectively to autophagy, Beta cells development and lipid metabolism; and hypo methylation and increase expression of GPR120 gene, regulating free fatty acid[340]. In rat, maternal dietary restriction led to decrease promoter methylation of glucocorticoid receptor (GR) and peroxisomal proliferator-activated receptor (PPAR), involved in stress response and lipid metabolism, and are associated to an increase expression of these genes in the offspring liver[341]. Epigenetics change have also been shown in kidney and adrenal gland from in utero diet restricted offspring respectively in p53 and in angiotensin II type 1b receptor genes, both playing a role in hypertension programming[342,343].

Epigenetics programming also occurs in LGA or related *in utero* excess nutrient exposure. In whole cord blood of LGA neonates, DNA hypermethylation of the FGFR2 gene have been observed[344]. In placental, an hypermethylation has been found in repetitive elements LINE-1 and AluYb8 and was associated with the methylation of polycomb group targeted genes as well as developmentally related transcription factor binding sites[345]. Interestingly, similarly to SGA from Dutch Hunger Winter, GDM induced LGA display a change in IGF2 methylation pattern in cord blood and placental tissue [346–348]. IGF2 change in methylation was shown to be associated with neonatal adiposity[338]. The common

epigenetics signature between LGA and SGA was further demonstrated by Delahaye et al focusing on CD34+ cord blood progenitors identifying several DNA hypermethylation targeting stem cells and metabolic pathways[349], that I will further develop later.

Epigenetics alterations could also program insulin resistance as a global DNA methylation pattern in cord blood was shown associated with insulin sensitivity in childhood[350]. GDM induces hypermethylation and decreases expression of lipoprotein lipase gene in placenta and was associated with 5 years offspring body fat composition[351]. DNA methylation alteration persists throughout life as observed in blood leukocytes of in utero exposed children and adult, on genes known to contribute to T2D and pancreatic Beta cells function[352–354].

The expression of leptin, the major regulator of food intake and body mass, found altered in obese peoples[355], can also be epigenetically programmed *in utero*, as its methylation status in cord blood offspring is associated with maternal glucose intolerance, GDM, and maternal obesity[356–358]. In mice, a global DNA methylation was shown in liver, muscle, and adipose tissue of offspring of mother fed with high fat diet (HFD) [359]. In primates, obesity and maternal HFD even prior to pregnancy and obese mother were shown to modify the chromatin structure of fetal liver through histone modification and were associated with dysregulated fetal lipid accumulation[360,361]. In female hypothalamus, perinatal maternal high-fat diet environment induced decreased melanocortin 4 receptor (Mc4r) and increase H3K27ac in its promoter and were associated with increased food intake and obesity in offspring[362].

Other prenatal stress, not dietary related, can also lead to long-term epigenetics change. In mouse brains, prenatal stress (mother subjected to daily physical constraint) induces important long term epigenetics alterations including aberrant DNA methylation and persistent DNMT expression while program for hyperactivity and for altered social interaction in adulthood[363]. These alterations were corrected upon the administration of an histone deacetylate inhibitor (valproic acid) and by antipsychotic agent with DNA demethylation activity (clozapine), highlighting a causative effect of DNA methylation alteration in cognitive disorders[363] . Such epigenetics programming was also shown in both human and animal models for others neurological disorders including anxiety, depression, attention deficit, and autism[364]. For AD, mouse models of AD exposed to prenatal or early life stress accelerate the impaired cognitive function including deficit in object location memory and impaired spatial learning[365,366]. Furthermore, chronic early life stress in these mice increases defective Abeta levels in middle age and correlates with reduce cognitive flexibility, while short treatment with glucocorticoids receptor antagonist rescues cognitive deficit and Abeta load[367] .

Then, different in utero stress can led to durable epigenetics change. Even if some targeted approaches have shown correlation between epigenetics alteration and subsequent gene expression alteration, as the study observing hypermethylation of lipoprotein lipase promoter associated with reduction of its expression in pancreas[351], the impact of the global epigenetics memory on transcription and signaling pathway remain poorly studied. Yet, this is critical to have a comprehensive view of the tissue dysfunction. Different integrative genomics analysis should be performed to understand more clearly the cellular and molecular signaling affected by these epigenomics alterations. Dr Fabien Delahaye and colleagues have worked on that these last years and I had the opportunity to continue along this effort during this thesis.

IV.2. HSPCs model to study early influences

To study the influence of early exposure on regulatory landscape and cell signaling, Dr Fabien Delahaye and colleagues focused on the specific model of human hematopoietic stem and progenitors cells (HSPCs).

HSPCs contain hematopoetic stem cells (HSC) and more differentiated progenitors including the multipotent progenitors (MPP), and three main lineage progenitors the erythroid (MEPs), lymphoid (CLPs) and myeloid (GMPs) progenitors (Figure 14). The erythroid progenitors give rise to megakaryocytes derived platelets, involved in blood clotting process, and erythrocytes, or red blood cells ensuring mainly O2 and CO2 transport for cellular respiration. The myeloid lineage give rise to monocytes, macrophage, granulocytes, and dendritic cells governing the innate immune response and inflammatory process. Finally, the lymphoid lineage, giving rise mainly to B cells and T cells, are involved in the acquired immune response. HSC can self renew and differentiate to produce these different progenies, a process called hematopoiesis.

Hematopoiesis occurs in different places in the body during development. Primitive hematopoiesis take place in the yolk sac approximately at day 7 of embryonic development. Immature precursors allow the generations of erythrocytes for embryonic O2 supply[368,369]. Placenta is the first reservoir of mature HSC (which can give all blood cell types) during development. Once vasculature developed (at embryonic day 12) HSC migrate to fetal liver, where they actively cycled (in contrast to bone marrow). During this HSC expansion in fetal liver, cartilage and bone are generating during mesenchymal condensations and are associated with bone vascularization (embryonic day 17.5) allowing finally HSC colonization of bone marrow. During life, HSC remain in bone marrow in a quiescent state. The cellular niche is an hypoxic environment around arterioles, where perivascular, endothelial, Schwann, and sympathetic neuronal cells secrete quiescence promoting cytokines such as CXCL12 and SCF [368]. Differentiated hematopoietic cells like macrophages or megakaryocytes are also

able to feed back in the niche to regulate HSC dormancy, either promoting or regulating HSC proliferation or migration depending on the context[370–372] . In basal state, macrophages promote the retention of hematopoietic stem cells by regulating CXCL12 production in the bone marrow and megakaryocytes localized with HSC promoting their quiescence through CXCL4 and TGF-β1 production[370–372] . Under hematopoietic demand, depletion of macrophage mediated CXCL12 production allow HSC mobilization, while FGF1 production by megakaryocytes under stress promote HSC expansion[371].

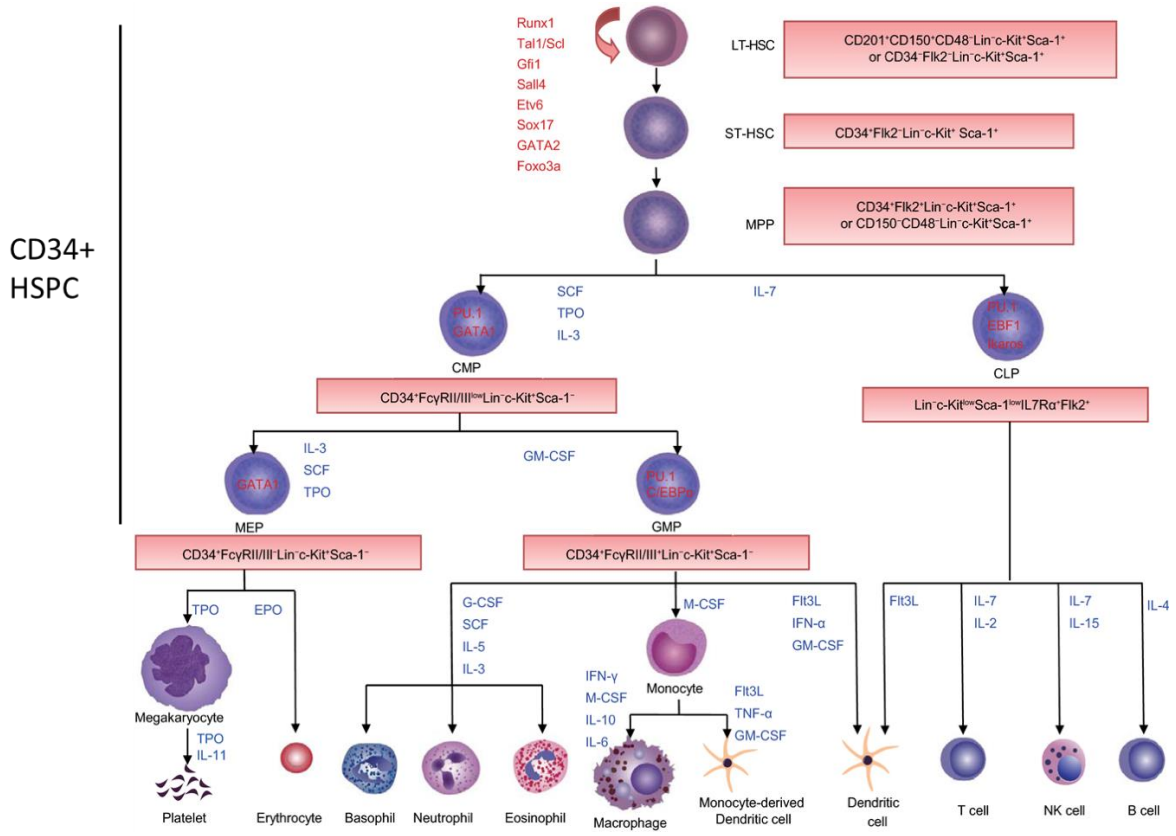


Figure 14 : Classical hematopoetic hierarchy and main regulators. Adapted from Cheng et al, Protein & Cell, 2020[N]

HSPCs are a relevant model to study influences of early environment on ACDs risk. They are easily isolable at birth from cord blood while being able to self-renew throughout life, conserving epigenetic memory of past exposure. Furthermore, the hematopoietic system play an important role in ACDs development as I will develop in next sections.

IV.2.a. Cord Blood HS(P)C

[N] https://link.springer.com/article/10.1007/s13238-019-0633-0

Cord blood HSPC are defined through CD34 cell surface marker. CD34+ cells represent around 0.4% of cord blood mononuclear cells (MNC) with high variability between samples typically from 0.2% to 1.4%[373]. By contrast, less than 0.01% of CD34+ cells are found in adult peripheral MNC and around 1.5% in bone marrow MNC. Among the CD34+ population, cells negative for CD38 cell surface marker phenotype are the HSCs with the greater long-term repopulation ability[373] and can be used for transplantation[374,375] . They represent around 0.05% of whole cord blood MNC.

Cord blood HSC are more responsive to stimulation than adult bone marrow HSC. Indeed, their quiescent form have a greater proliferative response to cytokines with lower dependence on stromal cells than bone marrow or adult blood HSC [376,377]. They also give rise to relatively different progeny compared to adult HSC. Notably, they give rise to less NK cells, produce a specific T cell progenitors (with the phenotype CD3-/CD8-), and different number of B cell subpopulation [377]. Several cytokines can stimulate HSC to proliferate notably SCF, Flt3, IL-11, IL-3, IL-6, GM-CSF while others can influence their differentiation, notably M-CSF, G-CSF, Epo, and Tpo.[377,378].

IV.2.b. Stem cells epigenetic memory

HSPCs, as for others stem cells, are present throughout life and thus have the ability to conserve putative cell memory of past exposure. Evidence of stem cells epigenetics programming exist. For example, it was recently shown that stem cells of follicle hairs can reprogram following wound damages to repair the epidermis through long-term epigenetics memories [379]. Adipose derived mesenchymal stem cells (MSCs) are functionally reprogrammed in obesity which lead to loss of stemness capacity and multipotency, change in their metabolism, and reduced immunomodulation and angiogenic capacity[380,381]. Dynamic regulation of DNA methylation plays an important role in orchestrating stem cell function [382]. Decreased expression of DNMT1, the enzyme maintaining DNA methylation pattern throughout division, reduces cell renewal and induces premature differentiation of epidermal progenitor cells, leading finally to tissue loss[383]. In HSC, reduced DNMT1 activity lead to defect in self renewal but also in decreased differentiation potential mirroring the defect observed in aging[383–385]. Then, the stem cells epigenetic memory appears as an important player in regulating stem cells function with important consequences on tissue regeneration capacities and tissue long-term function.

IV.2.c. Hematopoiesis and ACDs

The choice to study HSPC is also relevant because of its role in ACDs. The hematopoietic system plays a critical role in processes like inflammation, angiogenesis, and cardiovascular repair throughout life, making its progressive alteration a candidate mechanism in the development of ACD.

IV.2.c.i. Hematopoiesis regulation

HSC multipotency requires fine control of differentiation in order to give appropriate progeny and ensure cellular homeostasis. To ensure this control, it was recently demonstrated that epigenetics remodeling and TF activity act in concert. The role of DNA methylation in HSC differentiation has been pinpointed after the identification that the principal somatic mutations driving HSC clonal expansion and defective hematopoeisis in aging are in *DNMT3A* and *TET2* genes, the writer and eraser of DNA methylation[386]. Further functional studies have validated DNMT3A and TET2 roles in regulating HSC self renewal and differentiation in blood progeny[387–390]. Notably, the loss of TET2 leads to profound increase of HSC self-renewal while responsible for a myeloid bias differentiation[388–390]. TET1 loss leads to increase HSC self-renewal but was associated with lineage bias toward B cell production[391]. Ablation of DNMT3A impairs HSC differentiation while promoting HSC expansion and is associated with substantial changes in CpG island methylation with upregulation of HSC multipotency genes including GATA3, RUNX1, PBX1, and CDKN1A, while downregulation of differentiation factors including FLK2, SPI1, and MEF2C. Further studies have further highlighted the important role of DNMT1, TET1 and TET2 in regulating methylation of HSC differentiation program[385]. Together, these evidence show the crucial role of DNA methylation in the control of HSC differentiation.

Further studies have investigated other component of the regulatory landscape of hematopoiesis, including the chromatin dynamics and the TF activities. In the first study leveraging single-cell ATAC-seq to understand the chromatin changes governing hematopoiesis, Buenrostro *et al* have demonstrated that the regulatory landscape of HSPC is governed by modulation of lineage specifying TF motif accessibility [392]. Notably, they showed critical chromatin changes of GATA2 and MESP1 motif accessibility in HSC, respectively involved in erythroid and lymphoid lineage differentiation. Such rearrangement were also observed in more restricted hematopoeitic progenitors notably in TCF4 for the lymphoid-primed multipotential progenitors (LMPPs), STAT1 accessiblity for peripheral dendritic cells, and CEBPE for GMP differentiation. In a recent study assessing the impact of DNMT3A and TET2 mutations on chromatin landscape using single-cell sequencing, Izzo *et al* have shown that these mutations disrupt hematopoietic differentiation landscape, with opposite effects on erythroid and myeloid progenitors distribution[393]. They further show that Tet2 or Dnmt3a knockout (KO) induced opposite DNA methylation changes (hypermethylation for Tet2 KO and Hypomethyation for Dnmt3a KO) but both occur in same CpG rich accessible regions. These regions are enriched for erythroid TF motif, including Tal1 and Klf1, reconciling the opposite lineage priming (erythroid primed for Dnmt3a KO, and myeloid primed for Tet2 KO) observed in these two mutants. This opposite methylation change is associated with a shift in TF motif accessibility with decrease accessibility in Tet2 KO HSC while an increase accessibility for Dnmt3a KO HSC as demonstrated using single nuclei ATAC-

seq. These effects were studied in mice but were further validated in *DNMT3A*-mutated human samples supporting the role of the DNA methylation mediated chromatin change in the erythroid skew observed in these mutants. These results strongly support the crucial role of DNA methylation in remodeling chromatin accessibility and controlling activity of lineage defining TFs for appropriate HSC differentiation and multipotency.

Together, these studies have shown the importance of DNA methylation and chromatin accessibility remodeling in the control of hematopoiesis and emphasize the putative impact of their alterations on hematopoietic function.

IV.2.c.ii. Importance in aging and diseases

Defective hematopoiesis, chronic inflammation, and oxidative stress are key interconnected mechanism observed in aging and contributing to ACD risk. Clonal hematopoiesis, defined as a clonal expansion of dysfunctional immune cells occurring during normal aging, doubles the risk of coronary heart diseases in human[394]. Such clonality directly affects the heterogeneity and plasticity of HSC niches driving hematopoietic dysfunction and increasing ACDs susceptibilities[54,56]. Indeed, a relation between HSC clonal expansion and bias toward proliferation was observed in aging[395,396]. Furthermore, myeloid biased HSCs concomitant to an increase in myeloid cells in blood are observed in aging and strongly associated with CVD risk [395,397,398]. This increased myeloid cells, such as macrophages are known to worsen chronic inflammation by increasing levels of inflammatory cytokines[394,399]. Considering the hematopoeitic compartment, this disturbed homeostasis (balance between differentiation and proliferation) can have deleterious consequences on regulation of inflammation and therefore was shown to contribute to the inflamm-aging phenotype, a chronic low-grade inflammation observed in aging[56,395,400]. Alone, the dysregulated inflammation is a major contributors to the vicious circle of obesity, T2D and CVD development[395,401,402]. Together, these evidence emphasize the role HSC heterogeneity or plasticity alteration in ACDs programming, highlighting the relevance of the hematopoietic system as a model to study the early programming of ACDs.

IV.2.c.iii. Evidence of early programming of hematopoiesis

Several detrimental exposures can affect hematopoietic compartment plasticity and exposes to diseases risk. For example, short term hyperglycemic spikes, as observed in prediabetic or T2D patient, increase myeloid cells production in bone marrow, which accelerates atherosclerosis[403]. It was also shown in a mouse models that hyperglycemic environment lead to a reduced HSCs mobilization capacity in response to G-CSF[404]. However, few studies have been done in the context of *in utero* exposure. In human, the concentration of circulating CD34+ cells (HSPCs) was shown to be associated with extreme fetal growth[405–407], suggesting impact on HSPCs mobilization. More recently, a study led

by Kamimae-Lanning *et al* have assess the impact of maternal obesity and/or high fat diet on the fetal hematopoieitic system development [408]. They collected HSPCs from fetal liver to compare their expansion and repopulation ability depending on the nutritional exposure. They showed that HSPCs from fetus exposed to HFD, have a restricted physiological expansion and repopulating capacity while having an increased differentiation shifted toward myeloid lineage. These alterations are associated with changes in the expression of several genes involved in metabolism, immune and inflammatory processes, as well as stress response, and of key genes involved in self-renewal and HSC maintenance like *Egr-1* and *Bmi1*. This evidence support the early influence onf the hematopoetics system, however, if such alterations are conserved after birth remain to be explored.

IV.2.d. First study of early epigenetics programming of HSPCs

To answer these questions, Dr Fabien Delahaye *et al* have performed epigenomics analysis on HSPCs collected from neonates exposed to extreme fetal growth (both restricted growth, i.e. SGA, and overgrowth, i.e. LGA) and from appropriately grown neonates (CTRL).

They collected HSPCs from 60 samples (20 SGA, 20 LGA, and 20 CTRL) isolated from cord blood thanks to the CD34+ cell surface markers and performed on DNAs genome wide CpG methylation assay using the HELP tagging method[O][409]. They compared either SGA or LGA samples to appropriately grown neonates (CTRL) samples. They showed that both SGA and LGA neonates present a global increase of DNA methylation close to genes regulating stem cells function [349]. They also found an interesting sexual dimorphism, with female LGA neonates being more affected by DNA hyper methylation than male. To study putative influence of these epigenetics alterations on gene expression, they integrated these results with histone marks profiling of CD34+ cells. They evaluated the putative regulatory landscape of HSPC defining promoter, enhancer, as well as heterochromatin region using these histone marks profiling. They observed an enrichment for differentially methylated CpGs in promoter and active enhancer regions suggesting putative transcriptional consequences associated to this epigenetics alteration[349,410]. Together, these results have highlighted evidence of epigenetics memory/programming of HSPC both in SGA and LGA.

IV.3. Remaining challenges

This previous study by demonstrating extreme fetal growth associated epigenetics programming of HSPCs rised important biological questions i) What is the impact of these methylation changes on gene expression; ii) What is the consequences on HSPC homestasis and function?

[O] the HELP (HpaII tiny fragment Enrichment by Ligation-mediated PCR) Tagging method is an enzymatic method based on the DNA cutting by the restriction enzyme HpaII of unmethylated CCGG site

Indeed, the relationship between change in DNA methylation and change in gene expression remained mostly uncovered. Evidence suggest that hypermethylation of CpG islands close to promoter lead to transcriptional repression[18,19], but the effect of the methylation respond to far more complex model largely influenced by the genomic context[411]. Then, integrate epigenomics and transcriptomic data is still necessary to assess the impact of epigenetics alterations on gene expression. This is integration will help us better understand the mechanism behind early programming of ACDs. Such integrative approaches have shown their interest in animal model. Notably, a study has shown that intrauterine growth restriction induce persistent DNA methylation alterations in pancreatic islet of rat with concordant changes in expression of nearby genes prediposing to T2D[412]. However, in human, such studies are lacking. This is why I dedicated part of my thesis works to the integration of epigenomic and transcriptomic data from exposed and control neonates HSPCs.

Hematopoiesis is finely regulated by epigenetics mechanisms, then epigenetics alteration observed in HSPCs exposed to extreme fetal growth could have a direct impact on HSPCs heterogeinety, or plasticity/differentiation capacity, and explain future tissue dysfunction and diseases susceptibility. Challenging exposure, like infection, tissue damaged, or blood loss can lead to durable change in tissue composition and function as well as regeneration capacity, impacting risk of developing future diseases[413,414]. Such impact on stem cells plasticity was already observed following detrimental exposure but also occur 'naturally' altered in aging. Indeed, a stochastic epigenetics alterations is observed in aging[382], mostly reflecting imperfect tissue maintenance of epigenetics marks, creating an epigenetics mosaicism between cells of a same individual, aspecially for stem cells wich self-renew across life. This phenomenon have the ability to restrict stem cells plasticity and associated function but also lead to clonal expansion of defective stem cells[415]. Ultimately, that can lead to stem cells exhaustions, an other hallmarks of aging, but also tissue dysfunction and development of ACDs[385,415].

Recent studies have shown that HSC differentiation is a continuous process rather than have discrete steps of differention as previously stated (Figure 15). Cells in differentiation are not synchronized forming a spectrum of differentiation, *i.e.* a continuum of cellular heterogeneity. It was notably shown, that hematopoietic stem cells (HSCs) rather than being a highly homogeneous cell population are in fact an 'HSCs cloud' of cells at different differentiation level thus challenging a long-standing differentiation model, highlighting the interest to consider cell heterogeneity[122]. Similarly, a study in bone marrow identified a subset of HSC primed to become megakaryocytes [416]. Other studies of HSC niches during development or during aging have also reported HSC heterogeneity and a biased toward a specific lineage[417]. Similar increase heterogeneity has been observed in multipotent progenitors (MPPs) that express lineage biases gene even at early stage of differenciation [418].

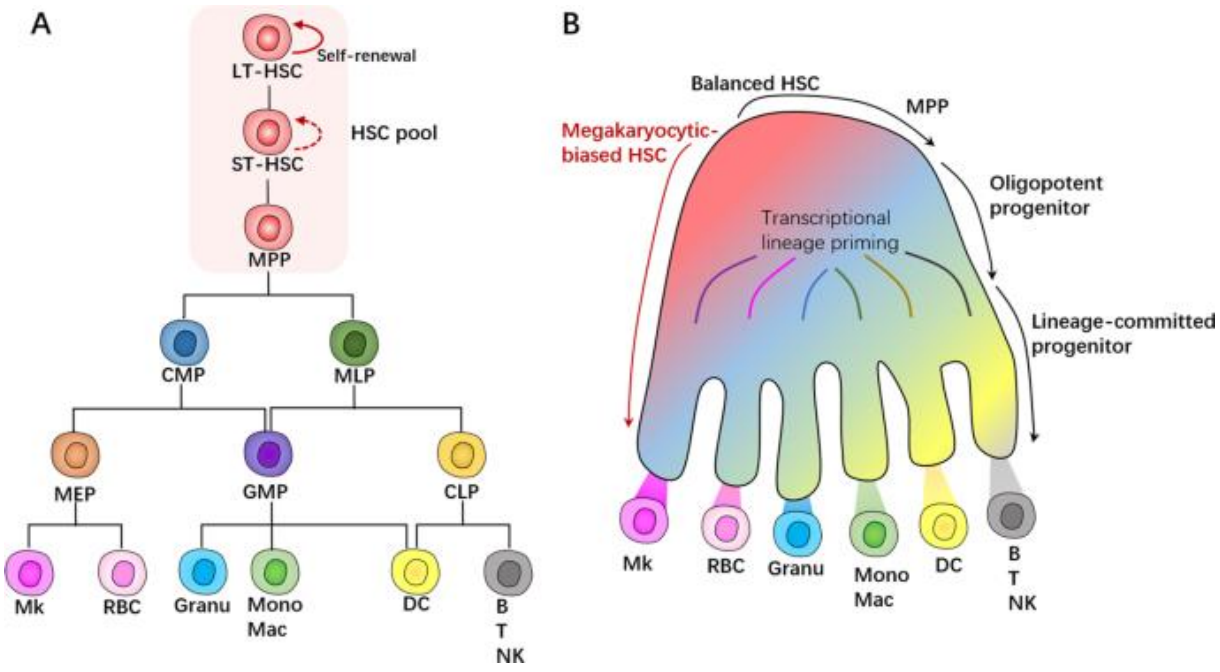


Figure 15 : Classical hematopoetic hierarchy (A) versus revised view of hematopoesis based on single-cell RNA-sequencing evidences (B). Reprinted from Zhang et al, Stem Cell Research & Therapy, 2022[P]

In regard of these evidence, we hypothesized that early epigenetics alteration of stem cells as observed in SGA and LGA could reduce HSPCs heterogeneity or plasticity. In this thesis work, I leveraged single-cell genomics data to assess such impacts, focusing on the LGA model.

IV.4. Large for gestational age model

LGA are characterized by an excessive fetal growth leading to a birth weight and ponderal index over the 90th percentile, and have an increase susceptibilities to develop ACDs including obesity[221], Type 2 diabetes[223] and cardiovascular diseases[224]. These increase susceptibilities are associated with an increased risk of impaired glucose tolerance and insulin resistance during childhood suggesting an early cellular reprogramming in these neonates[419–423]. Compared to SGA which have been intensively studied aspecially during hunger period, LGA have been less studied. Yet, LGA onset have increased during the 20-30 years, about 15-25% in developed country[424] and even larger in developing country[425]. However, the physiopathological consequences of being born LGA remain poorly understood. While being born LGA have also a genetic component, the recent raise of LGA clearly suggest that our modern environment contribute largely to this phenotype. Different *in utero* exposure can explain the fetal overgrowth. Maternal hyperglycemia, is associated with fetal hyperinsulinemia and increase by two the risk to give birth to LGA neonate[245,426,427]. Maternal obesity increases also by two the risk of LGA[255].

[P] https://stemcellres.biomedcentral.com/articles/10.1186/s13287-022-02718-1

Similarly, maternal dyslipidemia, including high triglycerides and high density lipoprotein (HDL) are associated with increased birthweight and LGA risk, and was consistent across different populations[428]. Maternal dyslipidemia is also associated to over secretion of placental Insulin like Growth Factor 1 (IGF1) highlighting a putative molecular pathway linking maternal dyslipidemia and excessive fetal growth[429]. Then, fetal overgrowth can be explained by overexposure to growth factor including glucose, insulin or IGF1. However, the long-term cellular and molecular consequences of this fetal exposures leading to LGA remains poorly understood. In these thesis works, I studied influence of this excessive fetal growth on HSPC regulatory landscape and plasticity, integrating both single-cell epigenomics and transcriptomics.

V. Model 2: Alzheimer's Diseases (AD) susceptibility gene BIN1 study

V.1. Alzeihmer's disease

Alzheimer's disease (AD) is a neurodegenerative disease responsible of 70% of dementia, affecting more than 20% of elderly people (> 75 years old). It is the 7th leading cause of death worldwide and is the major causes of disability and dependency among older people globally [8,430]. It is characterized by progressive neuronal degeneration in the brain associated with memory and cognitive loss. Even if the definitive diagnosis can be performed only after brain autopsy, cognitive test and PET scan help to have a clinical diagnosis of the disease. First symptoms are short term memory loss and inability to acquire new information as a result of the reduced neuronal plasticity[7]. Cognitive and motor function progressively decline after years as the consequence of neurotoxic aggregates spreading across the brain. Many unresolved questions regarding the AD pathophysiological process remain, so that no efficient treatment is yet available to prevent or cure this disease. AD is estimated having a 70% heritability suggesting strong genetics influences. A major genetic risk is APOE4, found in ~60% of AD carrier, while in 15% of the global population but not explain every diseases onset, as at least 1 third of AD patient do not have this variant[Q431]. Then, it is critical to understand others genetics influences as well as the role of these genes on brain and AD pathogenesis.

V.2. Brain complexity

The brain is a complex tissue displaying diverse cell types and complex cell to cell interactions. It can be broadly described as a neuronal network supported by glial cells, working in concert to efficiently convey and process information under the form of electrochemical pulses, giving rise to appropriate behavior and body homeostasis. Glial cells include astrocytes, oligodendrocytes as well

[Q] In a recent study, APOE related variants was able to explain 23% of all AD cases[39]

as microglia, bringing structural, metabolic and immune supports. Information are transfer across neurons thanks to “all or nothing” electrochemical pulse, called ‘action potential’, generated by neurons and transferred through their axon to synapses, where the information can be transmitted to other neurons through chemical signal. There is excitatory and inhibitory signal, governed mostly by Glutamatergic and GABAergic neurons, which respectively excite and inhibit downstream neurons through Glutamate and GABA neurotransmitters. The sum of excitatory and inhibitory signals transmitted to a neuron will be integrated to give rise or not to a new action potential transmit to downstream neurons. Furthermore, neurons are not only a passive information relay, but they are also able to retain information through biochemical and structural changes and therefore adapt their excitability and synapses connectivity. This feature allows plasticity of the neural network, and thus continuous adaptation and learning. That’s how the brain allow complex treatment and modeling of the information giving rise to learning, analytical thinking, and other complex behaviors.

Even if all neurons have similar structures with dendrites and cellular body receipting and integrating electrochemical signal from others and one axon transmitting signal to others neurons, there are different neuronal subtypes depending on the neurotransmitter they produce/release at synapses. Most of the neurons in brain are glutamatergic (around 40%), i.e. they produce glutamate as main neurotransmitter, but others can be GABAergic (producing GABA), dopaminergic (producing dopamine) or cholinergic (producing acetylcholine). This diversity allow specific neural system, complex interactions and independent cognitive process. Glutamatergic synapses are mainly excitatory, i.e. they produce depolarization inpost synaptic neurons and are linked to many other neurotransmitter pathways, with glutamate receptors being found in more than 90% of all neurons[432]. GABAergic neurons are the counter side of glutamatergic neurons producing mainly inhibitory signal, *i.e.* their principal role is to reduce excitability of postsynaptic neurons. Dopaminergic and cholinergic neurons take part of independent neurocircuits in brain involved in specific cognitive process. Cholinergic system is mainly involved in memory and learning in hippocampus[432], while dopaminergic neurons are mainly involved in the rewarding system[433].

Glial cells have also an important diversity in term of structure and functions than neurons. They include astrocytes, microglia, and oligodendrocytes. Astrocytes have generally a star shape and bring metabolic and structural support to the neurons. They derive from same progenitors cells than neurons (neural progenitor cells; NPCs) and represent around 20-40% of glial cells. They have many function including neurotransmitter recycling, provisions of nutrient to neurons, control of extracellular ion balance, but also brain repair processes and immune clearance in response to injury, neurotoxic agents, or infection [434]. Microglia are resident macrophage cells in the brain involved in maintenance of brain homeostasis, pruning of synapses and active immune defense. For a long time

thought to derive from hematopoietic stem cells like others macrophages, they in fact derived from erythro-myeloid precursors in the yolk sac, having characteristic gene expression profile compared to others myeloid cells[435] . They are primary involved in the tight regulation of brain homeostasis, being very sensitive to any pathological change and allowing clearance of unnecessary or neurotoxic components including plaques, damaged or unnecessary neurons and synapse, and infectious agents [435]. Oligodendrocytes mainly provide support and insulation of neuronal axons, producing the myelin sheath which wrap around axons allowing fast electrochemical action potential transmission to synapses. They also give metabolic and trophic support to neurons producing neurotrophic factor and insulin like growth factor-1 [436]. Like for neurons and astrocytes, they derive from NPCs. Besides these glial cells, there are also other cell types in brain that form the vascular system, maintaining the blood brain barriers including the vascular endothelial cells and pericytes. All of these non-neuronal cell types are affected in AD, but their contribution to AD-associated neurodegeneration remains incompletely understood [437]. Overall this cellular heterogeneity and tight cell-cell interactions allow integrity of the brain and its complex functions including synaptic plasticity and learning. A drawback of this complex and plastic interactions is that deregulation of one of this cellular interaction can be defective for the whole system. This also highlights the importance to take into account this cellular heterogeneity and interactions when modeling influence of genetics risk on brain functions and on AD pathogenesis.

V.3. How Alzeihmer's disease alter brain function ?

In AD brain, there is a progressive neuronal loss (neurodegeneration) characterized by toxic amyloid deposition in synapses and tau tangles accumulation in neurons. This leads to neuronal and synaptic loss, coupled with inflammation triggered by the proteotoxicity, leading to progressive memory loss and cognitive function impairment. The early mechanisms driving this defective accumulation is not fully understood, but key pathophysiological mechanisms are related to Abeta oligomers and tau tangles accumulation, as well as the associated neuroinflammation.

V.3.a. Amyloid beta accumulation

Probably the most important mechanism of AD development is defective Amyloid Precursors protein (APP) processing. Amyloid beta (Abeta) are byproduct of the APP cleavage in synapse. APP is a transmembrane protein playing essential role in nervous system development, synaptogenesis as well as modulation of synapse plasticity, and thus in learning and memory[438]. In normal APP processing, APP are cleaved by catalytic enzymes namely Alpha-secretase, Beta-secretase and Gamma-secretase following 2 different pathway. First processing pathway involve the Alpha-secretase and are non amyloidogenic. In the second pathway, APP are subsequently cleaved by Beta-secretase and the Gamma-secretase which lead to Abeta peptide formation. Abeta structure have the particularity to

oligomerize which in high concentration are toxic for synapses and neurons activity producing ROS production, inflammation, and synapse loss[439–442]. Most specifically, the suboptimal cleaving by Gamma-secretase can produce elongated form of Abeta (Abeta42) are particularly important in AD development. Ab42 is found enriched in early onset AD patients, who carrier defective mutation of APP or catalytic subunits of the Gamma-secretase PSEN1 and PSEN2[443–445]. Abeta42 have a particular structure and conformation decreasing its solubility, leading to significant increase oligomerization, and fibril (Abeta plaque) formation. More than being toxic for synapse, Abeta plaque appear to be also a major catalyzer of soluble Abeta oligomers, which can spread across the brain[446,447].While Abeta have important role in synapse signaling and neurotransmission regulation at low concentration, their increase concentration and aggregation into soluble oligomers are deleterious for synapse. Indeed, Abeta oligomers bind to several synapse receptors, notably in post synaptic regions, impairing synapse conformation and composition, as well as neurotransmission. Local increase of Abeta concentration can be beneficial for synaptic plasticity regulation, but become detrimental when the Abeta accumulation is systemic like in AD[442,448]. Abeta level is not strikingly correlated to AD severity, but rather a prognostic of future AD onset. Some healthy elderly have amyloid plaque without have dementia or significant AD symptom [448]. Ab deposition start decades before first symptoms and are not always associated to AD[449]. Then, Abeta accumulation appear to be necessary but not sufficient to explain AD pathogenesis. Further diseases development are likely to be driven by the pathological Tau hyperphosphorylation and aggregation, as well as the induced neuroinflammation.

V.3.b.Tau tangles formation

The other important pathophysiological mechanism recognized in AD is the accumulation of pathological Tau conformation. Tau is encoding by the gene MAPT, and are a microtubule binding protein stabilizing axonal microtubule and playing important role in intracellular transport notably mitochondrial and neurotransmitter transport to synaptic regions[450]. Tau is also found in postsynaptic region, in dendritic spines, where it can play important role in glutamate receptor scaffolding regulation. Once phosphorylated, Tau loose affinity to microtubule generating free Tau able to twist around each other's in paired helical fragments and form neurofribrillary tangles (NFTs). Loss of microtubule affinity disrupt cytoskeleton and intracellular transport, impacting notably the transport of mitochondria and glutamate receptors to post synaptic region, reducing ATP production, calcium buffering and thus neurotransmitter containing vesicles release. Furthermore, pathological tau can accumulate in dendritic spines, where they can directly affect synaptic transportation and post synaptic excitability[451]. The progressive formation of NFTs next aggregates into paired helical filaments, and spread through the peripheric part of stroma. Finally, NFT lead to proteotoxicity and neuronal loss, defect in axonal transport, mitochondrial damage and stress, and microglial activation /

inflammation[451] . This NFT accumulation are good indicator of neurodegeneration and strongly correlate with AD cognitive defects[452] . However, the causal mechanism leading to Tau hyperphosphorylation and aggregation remain not well understood. Abeta oligomers have been shown to promote Tau pathological conformation[453,454] but this element is not sufficient to explain Tau hyperphosphorylation as Ab accumulation have found in healthy elderly people without tau associated defects. Others factors have been identified promoting Tau hyperphosphorylation related to the imbalanced regulation of protein kinases and phosphatases notably the glycogen synthase kinase-3β (GSK-3β) and protein phosphatase 2A (PP2A)[455,456]. Then, understand the critical mechanism leading to Phospho-Tau accumulation appear crucial to better understand AD development.

V.3.c. Neuro inflammation

In addition to amyloid plaques and NFTs, neuroinflammation also plays an important role in AD. Like for tau tangles, this chronic inflammatory state in brain is strongly correlates with AD severity[457]. Neuroinflammation relies on microglial and astrocytes activation in response to neurotoxic agent. They allow phagocytosis of toxic product and release of cytotoxic factors to clean injured sites. Astrocytes and microglia play a major role in preventing Abeta mediated neurotoxicity. Indeed they are both involved in Abeta clearance through their internalization and degradation capacity. They notably express high levels of the Receptor for Advanced Glycosylation End (RAGE) which recognizes Abeta to degrade it through the endolysosomal process. However, after a certain concentration and specific AD related conditions, glial cells are overburdened, their benefical role switch off and they may even show a pernicious role in Abeta processing, contributing to the Abeta fibril formation and chronic neuroinflammation[107] . Microglial can remain long time in its activated form specially if neurotoxic agent is still present as for Abeta deposits, producing proinflammatory cytokines and toxic molecules which can worsen neurodegeneration and brain damage. As exposed previously, the pathological neuroinflammation can be mediated by the microglial activation in response to APP processing[458] or Abeta deposits accumulation[459] but others studies suggest others factors influencing neuroinflammation in AD like obesity and traumatic brain injuries[460]. GWAS studies have identified several genetics risk of AD falling genes express in immune cells and related to microglial functions including TREM2, SCIMP, MS4A3, HLA-DRA, HLA-DRB1 confirming important role of microglia in AD[39,461–463]. Futhermore, the low grade systemic inflammation seen in normal aging or associated to obesity, could also play an important role in the priming of microglial activation through proinflammatory cytokines production and diffusion through the blood brain barrier[460,464,465]. Both inflammaging condition, and peripheral chronic inflammatory diseases are associated with accelerated neuroinflammation and higher AD risk[464,466,467]. Then, neuroinflammation as well as microglial and astrocytes activation appear to be important actors in AD pathogenesis.

V.3.d.Calcium signaling and AD

Another important early molecular mechanism found altered in AD is the calcium homeostasis. Calcium homeostasis and calcium related intracellular signaling are key mechanisms to regulate neuronal activity. Indeed, Ca2+ participate to membrane depolarization and thus electrochemical signal transmission in neurons through plasma membrane receptors and voltage-dependent ion channels[468]. One important process for neurons after its depolarization is its rapid repolarization to be mobilized again and synchronized with others neurons[468] . This is allow through delivery of Ca2+ in extracellular matrix through ATP dependant pumps or Na+/Ca2+ exchanger. Fine tune regulation of calcium channel and exchanger is crucial for neurons to have this appropriate equilibrium between neuronal excitability (depolarization) and fast repolarization / return to basal level. Neurons have important Ca2+ dependent signaling pathways to couple Ca2+ with their biochemical machinery. Notably, Ca2+ dependent kinases activation allow long term potentiation (LTP) of neurons, a key mechanism of memory and learning[468,469]. Moreover, Ca2+ signal allow also to communicate between the synapse and the nucleus through gradient diffusion into nucleus, where Ca2+ can activate the TF CREB and control number of target genes. Then, calcium homeostasis and related signaling are crucial actors for appropriate neuronal activity and brains function including memory and learning.

Nevertheless, both calcium homeostasis and calcium related pathway appear altered at early stage in AD interacting with the amyloidogenic pathway[470]. Indeed, several studies have found that basal Ca2+ level in neurons close to Abeta deposits are increase compared ton non pathological neurons[471,472]. Mechanisms behind this phenomenon are unclear but predominant studies have found that Abeta oligomers could affect calcium homeostasis through its interactions with synaptic calcium channels, or by increasing Ca2+ permeability of the membrane[473–476] . But others evidence support an invert relationship, where calcium dyshomeostasis arise independently of Abeta but promote Abeta accumulation. Dreses-Werringloer *et al* have shown that a genetic variant in the Ca^{2+} homeostasis modulator 1 (CALHM1), a voltage-gated ion channel which promotes Ca^{2+} entry from the extracellular ambient, may increase the risk of AD while increase Abeta levels by interfering with Ca2+ permeability[477] . Furthermore, increase of Ca2+ can stimulate metabolism of APP[478–480] . Several studies have also shown that an increased calcium influx/signaling increase tau phosphorylation in neurons, further highlighting implication of calcium deregulation in AD pathogenesis[479,481–483].

One of the main consequences of disrupted Ca2+ signaling is alteration of synaptic plasticity and cell death[484]. Neurons near Abeta plaques have an increase activity (hyperexcitability) resulting from the reduced control of intracellular Ca2+ level[485] . Concordant with a role of Abeta in calcium mediated cognitive functions, Abeta oligomers emergence is correlated with reduced LTP [486–488], and

LTP defect can be reversed by antibody against Abeta[488] . Excess Ca2+ basal level can also activate mitochondrial released of cytochrome C inducing apoptosis cascade[484]. In support to this hypothesis, over expression of the antiapoptotic Bcl2 (reducing Ca2+ release by mitochondria), improve cognition in a mouse model of AD, before any sign of neurodegeneration [489].

What are the cellular mechanisms contributing to this calcium deregulation in AD remain unclear. Even if Abeta itself can alter calcium homeostaisis as seen previously, some studies have shown calcium deregulation before amyloid beta plaque detections [490–492]. Overall, these evidences shows the great importance of calcium dyshomeostasis and signaling in AD onset/progression.

V.4. Non genetics factors of AD

While genetics risk contribute to a major part of AD, it fails to explain around 30% of AD cases. Indeed, AD is a multifactorial diseases with sporadic onset, indicating that others factors influences their development including environmental, age related and epigenetics factors.

Different environmental exposure have been shown to modulate risk of AD. It was notably shown that repeated exposure to chemical agent including several heavy metals (including aluminum, arsenic, and mercury), particulate air, and some pesticides contribute to AD development [493]. Sleep deprivation appear also an environmental factors influencing Abeta accumulation, and could then increase AD risk[494,495] . On the other hand, cognitive ability and educational attainment appear protective factors against AD development [39]. These evidences of environmental influence highlight the importance to consider environmental exposure for AD prevention.

Other pathological conditions can also contribute to AD risk. Having had a stroke increase by 50% the risk to have AD [496]. T2D, which increase stroke probability, increase also risk of developing AD[497]. Systemic inflammation induced by obesity or rheumatoid arthritis, or low grade chronic inflammation observed in normal aging (inflamm-aging), increase risk of AD development, and further support importance of inflammation in AD[465,498–501]. To support that, anti-inflammatory treatment TNF-alpha inhibitor reduces AD onset in arthritis rheumatoid patients[502]. In addition to inflamm-aging, several others aging related conditions have shown to contribute to AD development: Increase oxydative stress[503], DNA damage[504,505], but also increase senenescence[506] are others age related factors contributing to AD. Notably, senolytic drug-mediated removal of senescent Abeta-associated oligodendrocytes progenitors have been shown to improves cognitive functions while reducing neuroinflammation and amyloid beta load in an AD mouse model[506].

In other conditions linked to aging, the somatic mutation take attention these recent years, because somatic mutation in neurons was shown increase with age and sensitive to AD development[55,59]. In a recent study, putative deleterious somatic mutation was differentially found in 27% of AD brains compared age matching healthy brain, with specific enrichment for genes contributing to tau related AD pathogenesis, including PIN1[59]. These results reconciling the strong genetics effect in AD susceptibilities and the late onset and sporadic feature of AD.

Epigenetic alteration is also a hallmark of aging and could have important role in AD. In aging brain, a DNA hypomethylation has been shown in the APP promoter[82]. In AD brains, H3K27ac and H3K9ac marks correlate with upregulation of chromatin and transcription related genes that contributes to Abeta42-driven neurodegeneration[82]. H3K9me3, mediating heterochromatin condensation, are also enriched in AD brains and lead to downregulation of associated genes involved in synaptic transmission and plasticity[83]. AD synaptic dysfunctions are strikingly associated with miRNAs expression profile alteration. The best characterized miRNA biomarker is probably miR-125b upregulation in AD[71] and influencing the essential synaptic glycoprotein synapsin-2 (Syn-2) contributing to APP processing and Abeta degradation[507,508], as well as tau Hyper-phosphorylation [509]. Too few studies have been made to appreciate the putative causal role of this epigenetics alteration in AD pathogenesis but require more attention regarding importance of epigenetics in others age related diseases. In any case, the epigenetic mechanisms identified in AD related features could be critical targets to better consider in AD diagnosis or management.

Together, these evidences of environmental, epigenetics, or age related factor give us insights about the critical non-genetics influences in AD pathogenesis and highlight important biological processes involved in AD, including inflammation/immune system deregulation, senescence, somatic mutation emergence and epigenetics alteration, which should be consider when modeling AD.

V.5. Genes involved in AD

The genetics influence in AD is high. The heritability is estimated at 60-80 % of all AD cases based on parental diagnosis linkage[510] . The early onset AD (EOAD, before 65 years old), representing 10% of all AD onset, have even more heritability ranging between 92% to 100%[444] . The known mutations involved in EOAD have high penetrance and are located in genes contributing to APP processing/ Abeta formation (PSEN1, PSEN2, and APP itself), but explain only 5-10% of EOAD case. The 90-95% remaining are still unexplained suggesting non-mendelian genetics or epigenetics transmission[444].

To explain influence of genetics in late onset sporadic AD, GWAS are used allowing to identify loci associated to AD but also genes and biological process potentially implicated in AD[35,39,39,511,512]. In

2019, a meta-analysis have identified 29 risk loci, implicating 215 potential causative genes and shown the important role of genes express in immune related cell types and regulating lipid related processes and processing of APP[39] . Then, two subsequent studies increasing meta-cohort size identified respectively 37 and 38 risk locis still highlighting strong role of microglia, immune cells and protein catabolism while discovering new genes candidates including CCDC6, TSPAN14, NCK2 and SPRED2[66,511]. The most recent meta analysis led by Bellenguez *et al,* found 75 risk loci associated with AD, of which 42 newly identified[35]. Pathways enrichment analysis on genes associated to these variants further validate importance of amyloid and tau pathways as well as the role of microglia related process in AD pathogenesis. Importantly, the vast majority of risk loci fall in noncoding regions mostly in intronic or intergenic region[39]. Only 2% are in exonic region, and 1% as non-synonymous mutation. Those noncoding variants are enriched for active regulatory regions and tissue specific eQTLs, highlighting their role in regulating transcriptional activity, but ask further consideration before assigning them to a gene [39]. It is for this reason that gene prioritization analysis have done integrating several tissue specific regulatory information to prioritize genes in order to link risk loci to putative causal gene. In the last meta-analysis, gene prioritization analysis have highlighted 31 newly associated genes involved in putative new AD associated processes including tumor necrosis factor alpha pathway. Furthermore, they construct a genetics risk score integrating all this newly identified risk loci to predict AD onset, and show that this score was able to predict 1.6 to 1.9 fold increase risk of AD from the lowest to the highest decile. Together these GWAS meta-analysis highlight around 40-80 putative genes involved in AD with important role in APP processing, tau related process, lipids metabolism, endocytosis and immune system process including microglial activation.

The first genetics risk locus is *APOE,* well characterized since its discovery in 1993[513]. Its allelic version ε4 (APOE4) is present in 14% of the total population[431], while found in between 40 to 80% of AD cases[514]. It increase by 3- to 15-fold the AD risk depending of the zygotic status[431]. APOE is an apolipoprotein able to transport lipids within or between cells and is involved in the clearance of Abeta by glial cells[515]. Furthermore, The APOE4 isoform have been shown to promote the lysosomal cholesterol accumulation in glial cells while impairing extracellular matrix homeostasis[67]. It has also been shown to promote accumulation of extracellular and intraneural Abeta[516].

The second most associated risk loci is *BIN1*, encoding for an adaptator protein mostly involved in lipid membrane dynamics including endocytosis regulation, but have also a role in intracellular trafficking as well as cellular excitability[517–519] . Some studies have been made since its discovery to decipher its role in AD but no consensus are still emerged. I will further develop in next part our understanding of BIN1 and interest to better characterize its role in brain and AD.

Others important risk loci have been robustly associated to AD including CLU, TREM2, and PTK2B, involved in previously mentioned AD related process. CLU is an extracellular chaperone inhibiting amyloid fibrils formation by sequestering the oligomeric forms[520,521]. TREM2 is a membrane receptor express in microglial recognizing Abeta42, which mediate its uptake and degradation by microglial[461,522]. For PTK2B, no obvious link with AD process was identified and need then further studies to its role in pathogenesis. This remark can be done for a majority of newly AD associated genes offering new perspective on our AD understanding.

V.6. Understand early mechanism of AD through BIN1 gene function study

Amyloid plaque, NFTs, neuroinflammation and calcium deregulation are central mechanism behind AD pathogenesis but the initial factor leading to this pathophysiological mechanism remain unclear. Even if amyloid plaques formation start decades before AD symptoms, suggesting critical causal mechanism, what predispose certain individuals to develop Abeta plaques mediated neurotoxicity remain to be determined. As shown in previous section, studies of genetics risk give us cues about what genes or biological process are crucially involved in AD development. GWAS studies clearly confirms the importance of gene regulating APP processing (ADAM10, PSEN1, PSEN2) and Abeta management by microglia (APOE [515,516], CLU [520,522,523], TREM2[461]), but some genes identified through GWAS do not directly links with Abeta pathway, indicating either independent mechanisms, or upstream non-direct regulators of Abeta pathway[39] . This is the case for *BIN1* associated variants, which were significantly associated with total Tau level and phospho-Tau in the cereobrospinal fluid but not with Abeta level [524]. *BIN1,* similarly to other AD risk genes including PICALM, CD2AP, CD33, EPHA, RIN3, MEF2C, and PTK2B, are involved in endosomal/membrane trafficking, suggesting important role of this biological process in AD pathogenesis in an Abeta independent manner[525–528]

While *BIN1* is the 2nd most associated genetics risk locus, little is known about its role in AD. Main variant associated to BIN1 increase by 17% the risk to have AD and are found across 40% of the population[35]. Because AD associated variants fall directly in BIN1 intronic region or colocalize with cis-eQTLs regulating BIN1 expression[66] , it appears evident than AD associated variants impact BIN1 gene expression directly, suggesting important role of this gene in AD development. BIN1 is ubiquitously express in the body, with highest expression in skeletal muscle and the brain [529]. BIN1 gene is composed of at least 20 exons subject to intense splice events, having then several isoforms and tissue specific pattern of expression. Isoform 1 to 7 are specifically express in the brain, isoform 8 in skeletal muscles, while isoforms 9 and 10 are ubiquitous[530–536] .

BIN1 is an adaptator protein of the Bin/Amphiphysin/Rvs (BAR) family regulating lipid membrane dynamics[537,538]. All isoforms are able to induce or sense membrane curvature through their

BAR domain, giving them a wide range of cellular functions in the control of cell membrane curving, shaping and remodeling[539]. However, only the neuronal isoforms have the CLAP domain mediating interaction with clathrin and AP2, two important protein involved in clathrin-mediated endocytosis[535,538,540,541]. Then, this information suggest specific role for BIN1 regulating endocytosis in brain.

Relatively few studies have been made since its discovery to decipher the role of BIN1 in AD. In brain, BIN1 is express mostly in oligodendrocytes, glutamatergic neurons, microglia and GABAergic neurons[530,542,543]. In neuron specifically, BIN1 have been shown to negatively modulate endocytic flux[517], and we recently found that BIN1 isoform 1 was able to regulate early endosome maturation and trafficking[517,544]. It has also been shown to participate to neuronal excitability by interacting with L-type voltage-gated calcium channels (LVGCCs)[545] . Recently, it have been shown regulating presynaptic neurotransmitter release with a role in memory consolidation[542,546]. In microglia, BIN1 could be involved in exosomes secretion[547] and have a role in regulating inflammation[548]. In oligodendrocytes, BIN1 could have role in membrane remodeling contributing with the process of myelination[530]. However, studies of the role of BIN1 in non -neuronal cell types of the brain remain elusive.

Role of BIN1 in AD remain unclear. In AD brains, BIN expression has controversially been shown to increase and decrease[543,549,550] . Indeed, one study have shown an increase total mRNA expression in frontal cortex of AD brains while others have found a decrease of mRNA in AD brains from multiple datasets and brain region[543] . To explain this putative contradiction, another study have shown that the protein level of the neuronal specific BIN1 isoform 1 was decrease in AD brain, but compensate by an increase of the ubiquitously express BIN1 isoform 9[551]. Furthermore, we do not know if this expression change occur similarly in all brain cell types or show cell specific expression alteration. Then, further analysis on clinical cell type specific transcriptomic data is needed to really decipher the cell type specific expression of BIN1 in brain and its alteration in AD.

Nevertheless, some preliminary studies have tried to decipher role of BIN1 in AD development. Studies related to role of BIN1 in Abeta accumulation appears inconclusive. In primary works, researchers have studied the role of BIN1 in regulated APP processing and show that a BIN1 knockdown in neuronal cultures can increase Abeta generation through β-secretase (BACE1) sequestrating in early endosomes[518,552] . However, in our previous work in drosophila model of AD, we found that human BIN1 deletion do not alter significantly APP processing while BIN1 dysregulation affect endocytosis and promote neurotoxicity[544]. Furthermore, another study in AD mouse model have also found that reduction of BIN1 expression, do not affect amyloid pathology[553].

More evidence have been shown for role of BIN1 in tau pathology. BIN1 can interact with Tau and its expression correlate with Tau level and tau tangles pathology in AD brains [549,551]. In a co-cultured model of tau propagation, Calafate *et al* have shown that knockdown or overexpression of neuronal specific BIN1 isoform 1 respectively promotes or inhibits Tau propagation in neurons[517]. They further show that loss of BIN1 increase endosomic flux, which increase Tau aggregate internalization and release into cytoplasm inducing Tau propagation. These results suggest that neuronal BIN1 have a protective role in preventing Tau pathology, and its downregulation as observed in AD brains could induce Tau pathology. Concordant with that, a recent study have shown that human BIN1 is able to recover human Tau induced cognitive defect in transgenic mice, preventing Tau mislocalization and somatic inclusions in the hippoccampus[546]. Authors further show that BIN1 can dynamically modulate its interaction with Tau through phosphorylation to compensate AD related Tau accumulation[546].

In another study regarding effect of overexpressing neuronal BIN1 isoform 1 in cultured rat hippocampal neurons, authors shows that upregulating BIN1 induced neuron hyperexcitability, increasing frequency of synaptic transmission, and increasing calcium transients showing neuronal hyperactivity[545]. They further suggested that over-expressed BIN1 could indirectly interact with L-type voltage-gated calcium channels (LVGCCs) through Tau. These calcium channels are involved in calcium transients and are known to be stabilized in membrane of cardiomyocytes thanks to BIN1[519,554], which could then explain this BIN1 mediated neuronal hyperactivity. They finally show that Tau mediate the BIN1-LVGCC interaction being bound by both proteins, and that is reduction prevent network hyperexcitability operate by BIN1 overexpression[545].

Role of BIN1 in neuronal excitability was further shown recently but regulating another pathway. Indeed, by conditionally deleting BIN1 in neurons, De Rossi *et al* have shown that BIN1 localize preferably in presynaptic sites of glutamatergic synapse in hippocampus, and that BIN1 are able to regulate the releases of neurotransmitter vesicles. Neuronal BIN1 deletion reduce synapses density and presynaptic protein cluster formation, while increase synaptic vesicles. Furthermore, these alterations in synaptic transmission in the center of memory consolidation (hippocampus) are concordantly associated with defective spatial learning and memory ability of Bin1 null mice. This evidence support a new role of BIN1 in synaptic function with putative effect on AD relevant coginitive process.

Taken together, these previous studies highlight the broad range of possible roles of BIN1 in brain/neurons but fails to define a clear mechanism behind BIN1 related AD susceptibility. All these previous studies are biased toward a certain hypothesis, which exclude discovery of putative unrecognized role. Furthermore, they mostly consider study of neurons, in 2D culture or non-human

model, which could miss the real BIN1 influence in human brain and associated cellular heterogeneity and interaction. Finally, studies overexpressing BIN1 can exaggerate BIN1 activity failing to highlight patho physiologically relevant mechanism. The inconsistency of the results exposed above, notably in the controversial role of BIN1 in APP processing, further highlight the need to study BIN1 in an unsupervised way. Then, to decipher the role of BIN1 in AD related brain function while being liberated of previous biased, we implemented a new modeling approach, integrating different human relevant brain models, unsupervised investigation and functional characterization. This approach leveraged single-cell transcriptomics to investigate the cellular effect of BIN1 deletion on both neuronal culture and cerebral organoid derived from human induced pluripotent stem cells (iPSC).

V.7. iPSC derived neuronal models

Transgenic mice have been largely used since decades to model AD, and have allowed clarifying several pathophysiological mechanisms. First mouse model was developed in 1995 based on the overexpression of human APP harboring familial AD mutation. Since then, varieties of mouse AD models have been developed allowing demystify or consolidate role of Abeta deposit and pathological Tau conformation in neurodegeneration but also the important role of microglia in pathological Tau spreading or the role of astrocyte activation and neuro-inflammation in synaptic dysfunction[555]. However, these in vivo models based on familial AD mutations or AD genes overexpression, failed to accurately mimic biology of AD, especially the sporadic form which not rely on severe AD gene mutation, but rather at a progressive development depending of several early cellular or molecular mechanisms. These shortcomings of current animal models are illustrated by the fact that all promising preclinical therapies failed when translated into clinics. It is then necessary to develop new models of AD, which better integrate human and sporadic AD specificities.

Here we used neural progenitors cells (NPCs) derived from human iPSC (hiPSC) to generate neuronal bidimensional culture and brain organoid that have their own set of advantages compared to animal models. First, human iPSC can be easily manipulated and edit genetically allowing assessing impact of risk loci or role of gene in AD. Furthermore, patient derived iPSC can be used directly to assess the effect of population-relevant genotype in cell activity. Second, iPSC can be reprogrammed in a variety of brain cell types including neurons, astrocytes, microglia and blood barrier endothelial cells, allowing deciphering role of a gene or variant in each specific cell type. Such human derived cellular model better model human susceptibilities, exemplified recently on the study APOE4 AD risk allele[67] . Indeed, authors of this study have shown that APOE4 have human specific effect on gene expression on microglia and astrocytes when compared to same cell types derived from APOE4 engineered mouse model. Finally, used of human derived brain organoid can recapitulate both cellular heterogeneity and interactions present in brain while kept human specific genetics influences[437].

We studied thanks to these models effect of the loss of function of one or two copies of the BIN1 gene (BIN1+/- and BIN1-/- respectively) on neuronal functions in the context of brain relevant cellular heterogeneity and structure (the cerebral organoid model), and in the context of cell autonomous activity, using pure hiPSC derived neuronal culture (hiNs). For this 2nd study model, we leveraged scRNA-seq to assess cellular heterogeneity and cell type specific effect of this BIN1 deletion in these two different models, and integrate results with AD brain scRNA-seq data to assess resemblance with AD related cellular activity and signaling.

RESULTS

I. EPIGENETICS PROGRAMMING of HSPCs

The prenatal period is a critical period of rapid growth and differentiation where tissue acquire their structure and functions. Environmental challenge, during this period, like over exposure to nutrient leading to LGA neonates, predisposes individual to develop age related and metabolic diseases, but the molecular mechanisms remain unclear. Several durable epigenetics alterations are found in early exposed tissue suggesting epigenetics programming of these diseases. Notably, my PhD supervisors are previously found a global DNA hypermethylation in LGA HSPC targeting genes regulating stem cell function suggesting early alteration of the hematopoiesis. Alteration of the hematopoietic system play an important role in the development of ACDs, and can be program as early as *in utero* following nutrient stress with long term consequences on hematopoietic functions. Then, based on these evidence, we hypothesized that early environment leading to excessive fetal growth affect hematopoietic compartment plasticity through epigenetics remodeling, thus modifying core functions of hematopoiesis and linking excessive fetal growth to increase sensitivity to metabolic and age-related diseases later in life. To challenge this hypothesis, we performed integrative analysis of DNA methylome, transcriptome and chromatin accessibility at single-cell resolution in early exposed HSPC in order to validate the DNA epigenetics alterations observed previously and evaluate their impact on HSPCs functions and plasticity.

We conducted a first set of analysis highlighting an epigenetics programming of the quiescence signaling in LGA HSC that we published recently. As a first step, using DNA methylation and scATAC-seq data, we further characterized the epigenetic memory of LGA HSPCs and predict impact on HSPCs function. For that we increase our number of cord blood derived HSPC DNA methylation data adding 16 CTRL and 16 LGA samples. To better highlight the influence of CpG methylation on gene expression, we implemented a new strategy integrating tissue specific regulatory annotation to weight each CpGs and infer regulatory link between CpG and neighbor genes. We confirmed that the DNA hypermethylation in LGA HSPCs target stem cells and growth signaling pathway. We also performed scATAC-seq on 5 LGA and 6 CTRL samples to validate the epigenetics alteration at cell level with another epigenetics layer; the chromatin accessibility. It also allowed us to validate the regulatory potential of differentially methylated CpGs (DMCs). Thanks to this integration, we found that both DNA methylation change and chromatin rearrangement occur in LGA in HSC-specific open chromatin region. As a second step, using scRNA-seq on 6 LGA and 7 CTRL samples, we assessed the impact on transcription at subpopulation level. We further integrated with the epigenomics data to decipher if direct link can been found between epigenetic alterations and gene expression change. We observed

a concordant decreased expression of genes of an epigenetically altered regulatory network governing by EGR1, KLF2 and KLF4 TFs regulating activation/differentiation of HSC. Finally, using both scRNA-seq data and in vitro differentiation assay, we deciphered the functional impact on HSPCs observing both a decreased HSC abundance in LGA neonates and decreased HSC-derived colonies, suggesting a reduced ability for HSC to remain undifferentiated.

To further validate these results and challenge the hypothesis that LGA HSCs have a decrease regulation of their differentiation, I performed another set of analysis leveraging supplemental information of already generated scRNA-seq data but also generated new one using gene silencing experiments, cytokines stimulation, and single-cell multi-omics assay. Notably, I tested using these others approaches the hypothesis that LGA HSCs have a differentiation bias in response to stimulation(i), and challenged our scRNA-seq and scATAC-seq results using the recent single-cell multimodal assay with new samples(ii). These complementary analyses confirm the HSC differentiation bias in LGA by using two independent approaches. They also validate the KLF2 influence of downstream targets genes based on gene silencing experiment and single-cell multimodal analysis, and support the role of KLF2 and EGR1 related regulatory network on regulating response to cytokines stimulation and HSC activation.

I.1. Published results

Pelletier, A.; Carrier, A.; Zhao, Y.; Canouil, M.; Derhourhi, M.; Durand, E.; Berberian-Ferrato, L.; Greally, J.; Hughes, F.; Froguel, P.; Bonnefond, A.; Delahaye, F. **Epigenetic and Transcriptomic Programming of HSC Quiescence Signaling in Large for Gestational Age Neonates**. *Int. J. Mol. Sci.* 2022, *23*, 7323. https://doi.org/10.3390/ijms23137323

International Journal of *Molecular Sciences*

MDPI

Article

Epigenetic and Transcriptomic Programming of HSC Quiescence Signaling in Large for Gestational Age Neonates

Alexandre Pelletier [1,2,†], Arnaud Carrier [1,2,†], Yongmei Zhao [3] Mickaël Canouil [1,2], Mehdi Derhourhi [1], Emmanuelle Durand [1], Lionel Berberian-Ferrato [1], John Greally [4], Francine Hughes [5], Philippe Froguel [1,2,6], Amélie Bonnefond [1,2,6] and Fabien Delahaye* [1,2]

[1] Inserm U1283, CNRS UMR 8199, European Genomic Institute for Diabetes, Institut Pasteur de Lille, 59000 Lille, France; mickael.canouil@cnrs.fr (M.C.); mehdi.derhourhi@cnrs.fr (M.D.); emmanuelle.durand@cnrs.fr (E.D.); lionel.berberian@cnrs.fr (L.B.-F.); philippe.froguel@cnrs.fr (P.F.); amelie.bonnefond@cnrs.fr (A.B.); fabien.delahaye@pasteur-lille.fr (F.D.)
[2] Lille University Hospital, University of Lille, 59000 Lille, France
[3] Department of Obstetrics & Gynecology and Women's Health, Albert Einstein College of Medicine, 1300 Morris Park Ave, Bronx, NY 10461, USA; yongmei.zhao@einsteinmed.org
[4] Department of Genetics, Albert Einstein College of Medicine, 1301 Morris Park Avenue, Price Building, Room 322, Bronx, NY 10461, USA; john.greally@einsteinmed.org
[5] Obstetrics & Gynecology and Women's Health, Division of Maternal-Fetal Medicine, Montefiore Medical Center/Albert Einstein College of Medicine, Bronx, NY 10461, USA; fhughes@montefiore.org
[6] Department of Metabolism, Digestion and Reproduction, Imperial College London, Exhibition Rd, South Kensington; London SW7 2BX, UK
* Correspondence: fabien.delahaye@pasteur-lille.fr (F.D.); (+33) 320877258 (F.D.); philippe.froguel@cnrs.fr (P.F.); (+33) 374008101(P.F.)

† These authors contributed equally to this work.

Citation: Pelletier, A.; Carrier, A.; Zhao, Y.M.; Canouil, M.; Derhourhi, M.; Durand, E.; Berberian-Ferrato, L.; Greally, J.; Hughes, F.; Froguel, P.; et al. Epigenetic and Transcriptomic Programming of HSC Quiescence Signaling in Large for Gestational Age Neonates. *Int. J. Mol. Sci.* **2022**, *23*, x. https://doi.org/10.3390/xxxxx

Academic Editor(s):

Received: date
Accepted: date
Published: date

Abstract: Excessive fetal growth is associated with DNA methylation alterations in human hematopoietic stem and progenitor cells (HSPC), but their functional impact remains elusive. We implemented an integrative analysis combining single-cell epigenomics, single-cell transcriptomics, and in vitro analyses to functionally link DNA methylation changes to putative alterations of HSPC functions. We showed in hematopoietic stem cells (HSC) from large for gestational age neonates that both DNA hypermethylation and chromatin rearrangements target a specific network of transcription factors known to sustain stem cell quiescence. In parallel, we found a decreased expression of key genes regulating HSC differentiation including *EGR1*, *KLF2*, *SOCS3*, and *JUNB*. Our functional analyses showed that this epigenetic programming was associated with a decreased ability for HSCs to remain quiescent. Taken together, our multimodal approach using single-cell (epi)genomics showed that human fetal overgrowth affects hematopoietic stem cells' quiescence signaling via epigenetic programming.

Keywords: Epigenomics; Single-cell; Stem-cells; Fetal programming; Hematopoiesis

1. Introduction

Hematopoietic stem cells (HSC) are involved in essential processes such as inflammation, cardiovascular repair, and immunity throughout the entire lifespan [1,2]. Thus, alterations in HSC's ability to self-renew and to adequately produce differentiated progeny have been suggested to contribute to the onset and progression of age-related diseases such as cancer and cardiovascular diseases [3,4]. Systemic alterations or the action of various stressors like aging [5,6] can result in alteration of HSC destiny, and ultimately hematopoietic functions. The early mechanisms that control their long-term

functions in humans are not well understood, in part due to the diversity of phenotypes and behaviors of HSCs [7].

In mice, a maternal high-fat diet during gestation limits fetal hematopoietic stem and progenitor cells (HSPC) expansion and ability to repopulate while inducing myeloid-biased differentiation [8]. In humans, a limited number of studies have been conducted. Fetal growth was shown to alter the number of circulating CD34+ HSCs [9,10]. We previously described a global increase of DNA methylation in cord blood-derived CD34+ HSPCs from large for gestational age (LGA) infants compared to neonates with normal birth weight [11]. Still, the functional impacts of these early epigenetic alterations remain to be elucidated. Such an effort is essential to determine how these epigenetic modifications could mediate the association between early-life exposures and the induction of persistent life-long functional changes within the hematopoietic system.

We conducted a multimodal analysis combining single-cell epigenomics, single-cell transcriptomics, and in vitro analyses to link the DNA methylation alterations observed in LGA neonates with functional alterations in human cord blood-derived HSPCs. We developed novel analytical approaches to improve the integration of epigenomic and transcriptomic data. We found that the DNA hyper-methylation observed in LGA HSPC is associated with an HSC-specific decreased chromatin accessibility and gene expression of key genes involved in the HSC quiescence signaling as well as an alteration of the HSC colony-forming capacity.

2. Results

2.1. Optimized Methylation Gene Set Analysis Reveals Association between LGA DNA Hypermethylation and Stem Cell Differentiation Pathways

To confirm the LGA-associated DNA hypermethylation we previously observed, we significantly increased the power of our analysis. We expanded our original study through additional patient inclusions, thereby doubling the size of our cohort [11]. Using the HELP-tagging assay (HpaII tiny fragment Enrichment by Ligation-mediated PCR), we generated genome-wide DNA methylation data on 16 CTRL and 16 LGA cord-blood derived human CD34+ HSPC samples. We independently retrieved in this new dataset the global DNA hypermethylation initially found in LGA compared to controls [11] (Figure 1A). Then, to increase our detection power, we pooled both datasets and detected a total of 4815 differentially methylated CpGs (DMC) with 4787 CpGs hypermethylated and 28 CpGs hypomethylated in LGA (n = 36) compared to CTRL (n = 34, p-value < 0.001 and |methylation difference| > 25%; Figure 1A, Supplemental Table S1). This new set of DMCs was then used throughout the following analysis.

As the functional interpretation is performed at the gene level, each CpG (or DMC) must be linked to a specific gene. Thus, our ability to adequately infer the regulatory effect of a CpG and its target gene will affect our ability to identify relevant pathways. Standard analytical approaches usually rely on the distance between CpG and transcription start site (TSS) of the targeted gene and often only consider the top candidate DMC per gene, not taking into account the cell specific genomic context. Therefore, we refined the CpG-gene association to optimally assess the influence of DNA methylation changes on gene expression and enhance functional interpretation. We built a novel gene-methylation score considering (1) the distance between TSS and CpG; (2) the CpG overlap with expression quantitative trait loci (eQTL) annotation, as eQTL information allows us to identify tissue-specific genomic region links to gene expression changes; and (3) the regulatory annotation (e.g., Promoter, Enhancer) based on cell-specific histone marks [12] and on the Ensembl Regulatory database, as we know that the relationship between change in DNA methylation and change in gene expression will depend on a cell-specific genomic context (Figure 1B). We established 756,470 CpG-gene associations including 34% of them found through eQTL annotation. We then summarized the CpG information at the gene level, generating a gene-methylation score for each gene (n = 24,857,

Supplemental Table S2). We first confirmed that the gene-methylation score properly recapitulates the influence of key parameters in DMC analysis such as significance and effect size of the methylation change, number of DMCs per gene, and distance from TSS, as well as promoter and enhancer localization (Figure 1C). We also confirmed that while preserving key information from standard methylation metrics, the gene-methylation score presented a better association with DEGs than significance or methylation change alone. Thus, the gene-methylation score appears to be a better predictor of the methylation influence on gene expression (Figure 1C). We then used our gene-methylation score to perform pathway enrichment analysis and data integration, especially considering integration with gene expression data.

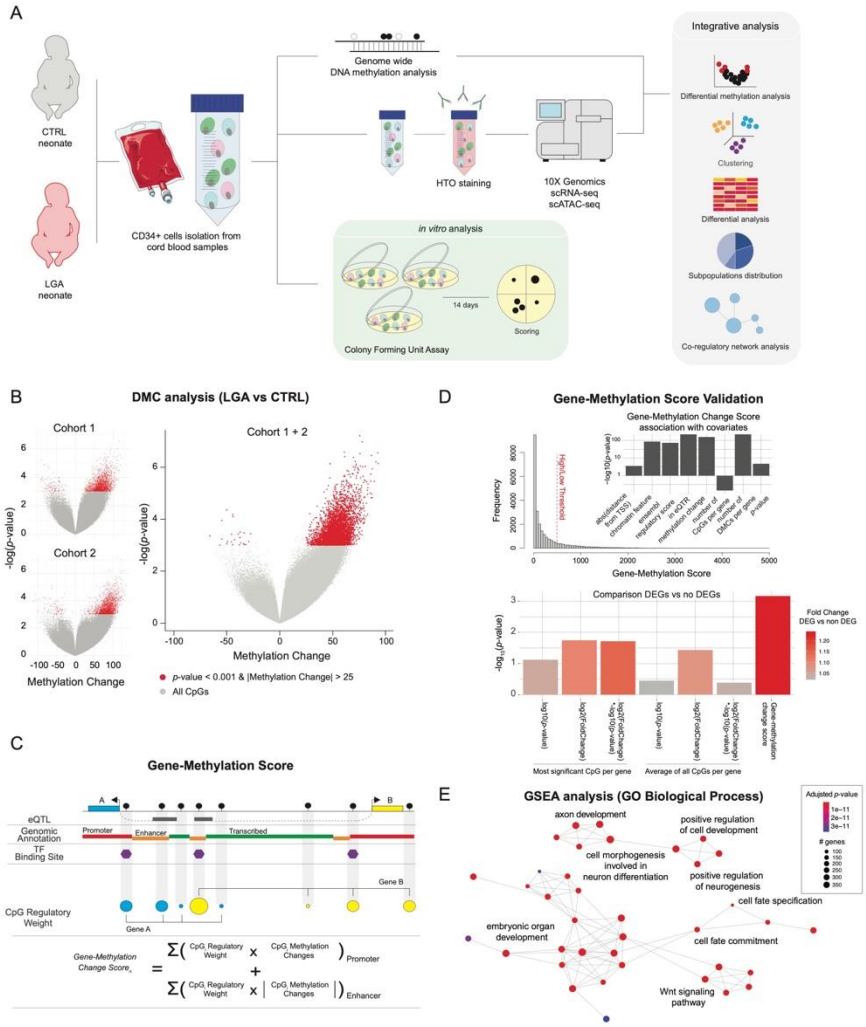


Figure 1. LGA is associated with DNA hypermethylation targeting key stem cell signaling pathways. (**A**) Overview of study design (**B**) Volcano plot of DNA methylation score differences for LGA compared to CTRL in cohort 1, cohort 2, and cohort 1 + 2. Differentially methylated loci with *p*-value < 0.001 and |methylation difference| > 25% are shown in red. (**C**) Summary of calculation for the gene-methylation score. (**D**) Validation of the gene-methylation score. Gene-methylation score distribution. Bar plot of the association between gene-methylation score and genomic or methylation-related features. Bar plots representing the significance of the difference in gene-methylation score of DEGs compared to non-DEGs considering different metrics. eQTR, region with expression quantitative traits loci; DMC, differentially methylated CpGs. (**E**) Network representation of GO Biological Process enriched in hypermethylated genes. Significantly enriched GO terms were identified using GSEA based on the gene-methylation score. Edges represent interactions (gene overlap) between pathways.

Using the gene ontology (GO) reference database, we performed methylation gene-set enrichment analysis (GSEA) based on the gene-methylation score. We found that change in DNA methylation in LGA HSPC samples targeted genes involved in signaling regulating fetal development as well as in key stem cell pathways such as Wnt signaling, cell fate specification, and cell fate commitment pathways (adjusted *p*-value < 0.01, Figure 1D) confirming previous findings [11].

2.2. Single-Cell Transcriptomic Analysis Confirms Alteration of Hyper-Methylated Genes in Pathways Regulating Stem Cell Differentiation among LGA HSCs

To identify genes altered in LGA and to obtain further biological insight into the functional consequences of the DNA methylation modifications observed in LGA, we performed a single-cell transcriptomic analysis comparing CTRL and LGA HSPCs.

To enable lineage-specific transcriptomic analysis, we created a hematopoietic reference map (i.e., hematomap) by integrating data generated from cord blood-derived CD34+ HSPC cells (*n* = 18520) from 7 control neonates (Figure 2A). Based on cluster-specific gene expression, we identified 18 distinct clusters representative of major lineages (Long-Term HSC, HSC, Multi-Potent progenitor, Lymphoid, Myeloid, and Erythroid) of the hematopoietic compartment (Figure 2B, Supplemental Figure S2). Each cluster was annotated using cell-type-specific markers. Markers were then ranked based on their expression fold change and the specificity of the cluster. Top cluster-specific markers were compared with published cell-type-specific genes [13–16] (Supplemental Table S3). Candidate cell subpopulations were distributed as follows: 1% LT-HSC (*ID1*); 24% HSC (*AVP*); 45% MPP/LMPP (*CDK6*); Lymphoid (*CD99, LTB*); 1% B cell (*IGHM*); 1% T cell (*CD7*); 14% Erythro-Mas (*GATA1*); <1%Mk/Er (*PLEK, HBD*); 8% Myeloid (*MPO*); <1% DC (*CST3*, *CD83*).

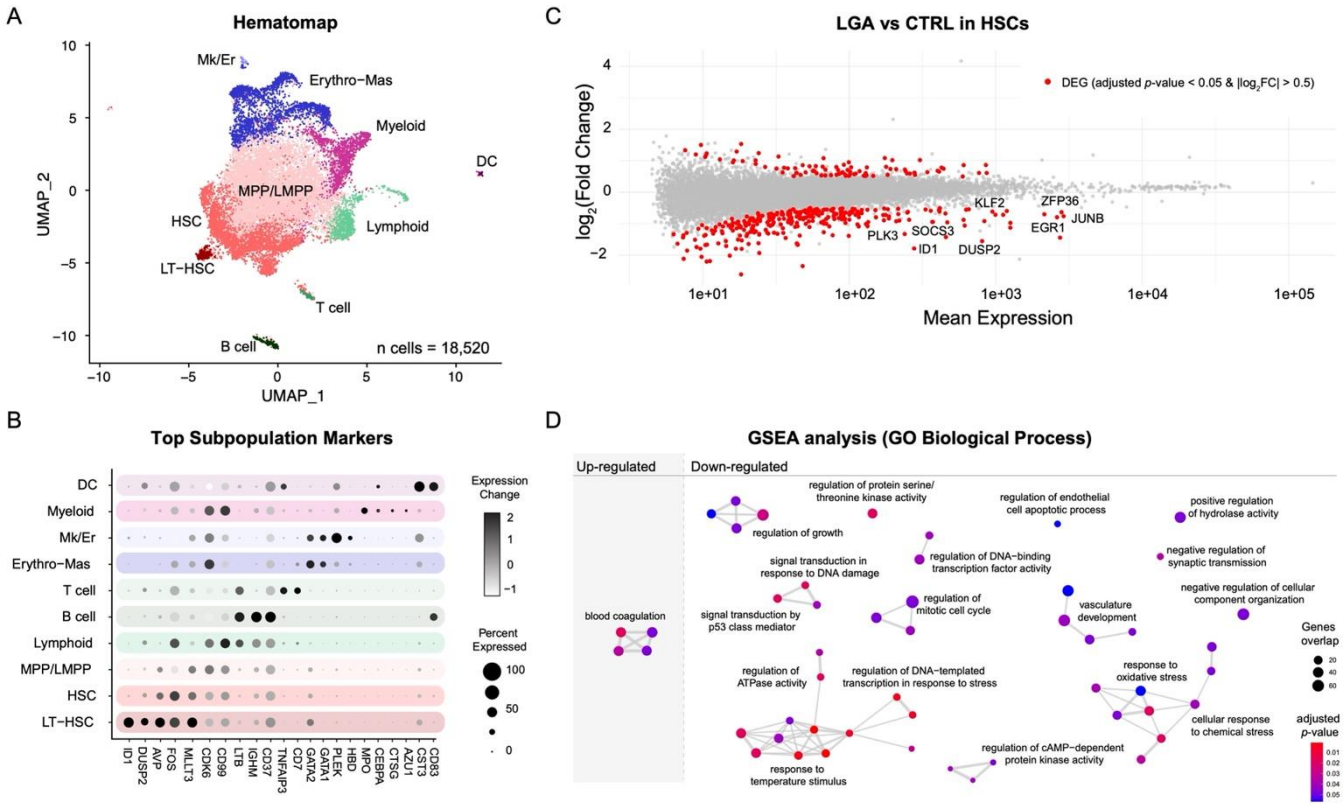


Figure 2. Lineage-specific transcriptomic analysis. (**A**) Hematomap, UMAP representation of distinct HSPC lineages. (**B**) Dot plot representing key markers used to annotate cell populations. LT-HSC, long-term hematopoietic stem cell; HSC, hematopoietic stem cell; MPP, multipotent

progenitor; LMPP, lymphoid-primed multipotent progenitors; Erythro-Mas, erythroid and mast precursor; Mk/Er, megakaryocyte and erythrocyte; DC, dendritic cell. (**C**) MA plots representing gene expression analysis in HSCs comparing LGA vs. CTRL. Differentially expressed genes with adjusted *p*-value < 0.05 and $|\log_2 FC| > 0.5$ are shown in red. (**D**) Network representation of significantly enriched pathways identified through GO GSEA analysis comparing LGA vs. CTRL. Non-redundant pathway annotations have been used. Edges represent interactions between pathways.

To identify differentially expressed genes (DEG) between CTRL and LGA samples, we implemented the Hash Tag Oligonucleotide (HTO) multiplexing strategy [17] allowing simultaneous processing of CTRL and LGA samples. Multiplexing is a means to limit the influence of technique-driven batch effects at every stage of the analysis to improve the biological relevance of the finding. We generated multiplexed single-cell transcriptomic data from 6 LGA (n = 6861 cells) and 7 CTRL (n = 5823 cells) samples. In LGA samples, we observed a shift toward downregulated genes (Supplemental Figure S3) especially in the HSC subpopulation (n = 285 downregulated genes over 373 DEGs, adjusted p-value < 0.05 and $\log_2 FC < (-0.5)$, Figure 2C; Supplemental Table S4). Notably, the well-known *EGR1*, *JUNB*, and *KLF2* genes were among the top affected genes. Using GO enrichment analysis, we found that downregulated genes were enriched in growth-related pathways (e.g., regulation of growth) as well as in stress-related biological processes (e.g., response to temperature stimulus, cellular response to chemical stress; Figure 2D, adjusted p-value < 0.05).

To assess if these HSC-specific transcriptomic changes may be associated with epigenetic changes, we integrated bulk DNA methylation with single-cell gene expression data using the gene-methylation score. We found that DEGs, and particularly the down-regulated genes, mostly showed high gene-methylation scores (Figure 3A,B). We then assessed the association between changes in DNA methylation and gene expression at the pathway level. We looked for enrichment for differentially methylated genes considering pathways that were identified based on DEGs. We found a significant overlap between GO terms enriched in LGA HSC downregulated genes and GO terms enriched in hypermethylated genes (10 out of 46; p-value < 0.05, hypergeometric test). The most co-enriched term is "regulation of growth" including notably *SOCS3*, *SIRT1* and *SESN2* genes that are both downregulated and within the top 10% of hypermethylated genes (Figure 3C). These results suggest that the epigenetic change in LGA could lead to an HSC-specific alteration of the regulation of growth signaling.

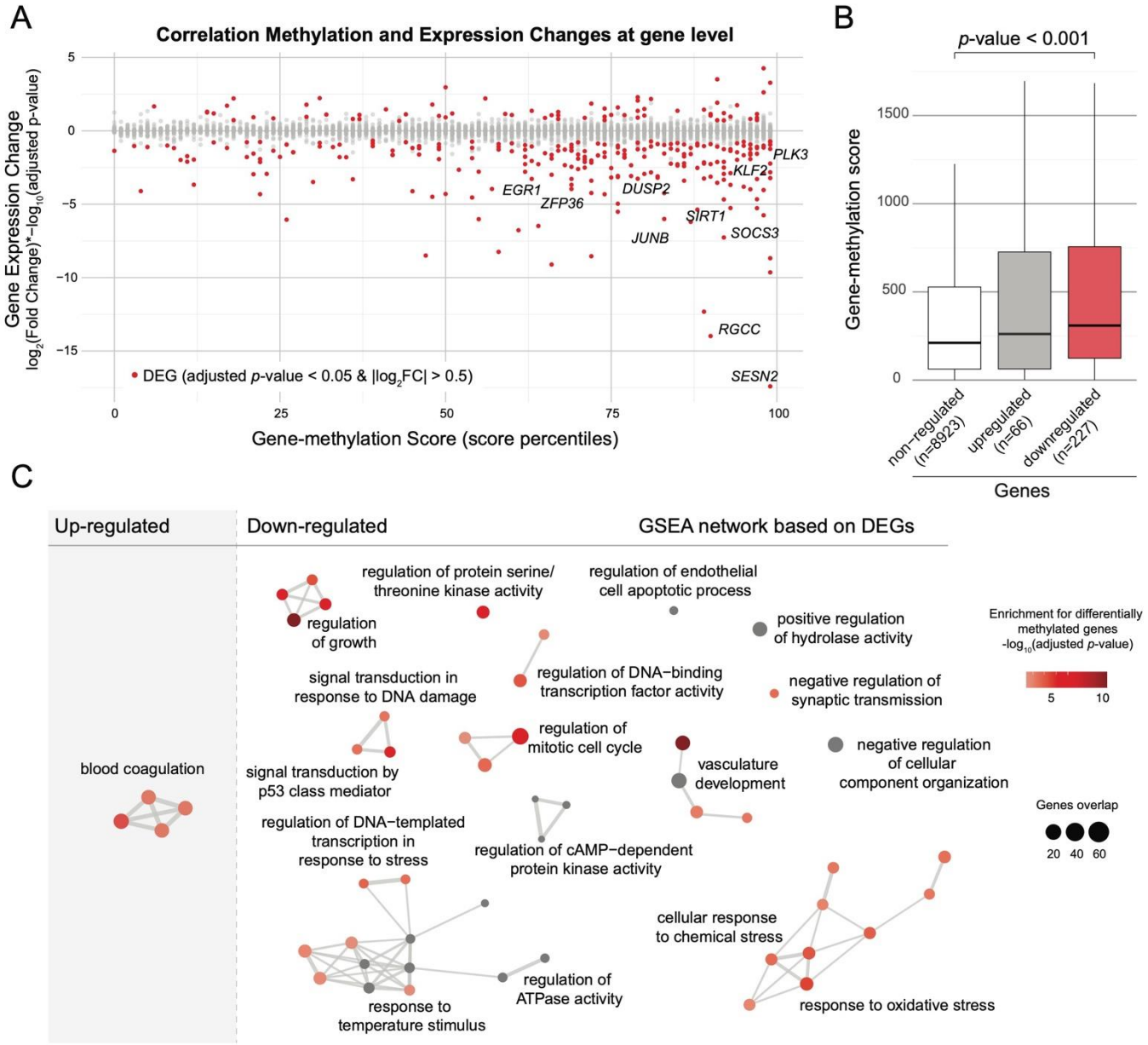


Figure 3. Association between changes in DNA methylation and in gene expression. (**A**) Dot plot representing the correlation between DNA methylation and gene expression changes. Differentially expressed genes with adjusted *p*-value < 0.05 and $|\log_2 FC| > 0.5$ are shown in red. (**B**) Boxplots representing gene-methylation score distribution associated with non-DEG, up-regulated, and down-regulated genes (Wilcoxon test). (**C**) Network representation of significantly enriched pathways identified through GO GSEA analysis based on DEG identified comparing LGA vs. CTRL. Nodes are color-coded based on enrichment for differentially methylated genes using the gene-methylation score. Edges represent interactions between pathways.

2.3. DNA Methylation Changes Occurs in HSCs and DEGs Associated Open Chromatin Regions

To assess if the HSC-specific transcriptional alteration could be due to HSC-specific epigenetic change, we profiled chromatin accessibility at the single-cell level (i.e., single-cell ATAC-seq). We generated open chromatin data across 8733 cells in HSPCs from 6 CTRL and 5 LGA neonates. We first annotated subpopulations using the label transfer approach between ATAC-seq data and the lineage labels from the Hematomap (Figure 4A, Supplemental Figure S4A). To validate the relevance of our lineage annotation, we

performed TF motif enrichment and observed that lineage-specific peaks were effectively associated with well-known lineage-specific TF (Supplemental Figure S4B).

We then integrated our bulk DNA methylation data with our single-cell ATAC-seq data to assess DMCs distribution within open chromatin regions (OCRs). Overall, 31% of the 211,479 peaks contain CpGs queried by our genome-wide methylation assay. We first observed a strong enrichment for DMCs in OCRs with 74% of them located in OCRs compare to only 34% of overall queried CpGs (p-value < 0.001, hypergeometric test). Such enrichment further supports the putative regulatory influence of our DMCs. By performing lineage-specific analysis, we observed DMCs enrichment in HSC-specific open chromatin region with a total of 11% of HSC-specific peaks containing DMCs (adjusted p-value < 0.001, Figure 4B), while no enrichment was observed for the other lineages. This result corroborates the HSC-specific transcriptional impact of the DNA methylation changes observed in LGA. Furthermore, we observed that DEGs in LGA HSC and especially down-regulated genes were enriched for OCRs containing DMCs (Figure 4C).

Not limiting our analysis to the regulatory role of DMCs within open chromatin regions, we then assessed the change in chromatin accessibility in LGA HSCs. We identified 278 open chromatin regions that significantly differ between LGA and CTRL HSCs, with 215 showing decreased and 63 showing increased accessibility (adjusted p-value < 0.001 and $|\log_2 FC| > 0.25$, Supplemental Figure S4C). By performing TF Motif analysis on regions with decreased accessibility, we identified that the motif of the transcriptionally downregulated TFs EGR1 and KLF2 are highly enriched (p-value < 1.10^{-40}) and among the top 6 enriched motifs (Figure 4D, Supplemental Figure S4D).

We then assessed the interaction between DNA methylation, gene expression, and chromatin accessibility. Regions with decreased accessibility were also strongly enriched in peaks including DMCs and peaks associated with DEGs (Figure 4E), with 3-fold and a 2.5-fold enrichment, respectively. Furthermore, these regions were strongly enriched for peaks containing both DMCs and associated with DEGs (23-fold enrichment) illustrating that early epigenetic programming is actually not limited to changes in DNA methylation but also involves chromatin rearrangement targeting altered genes.

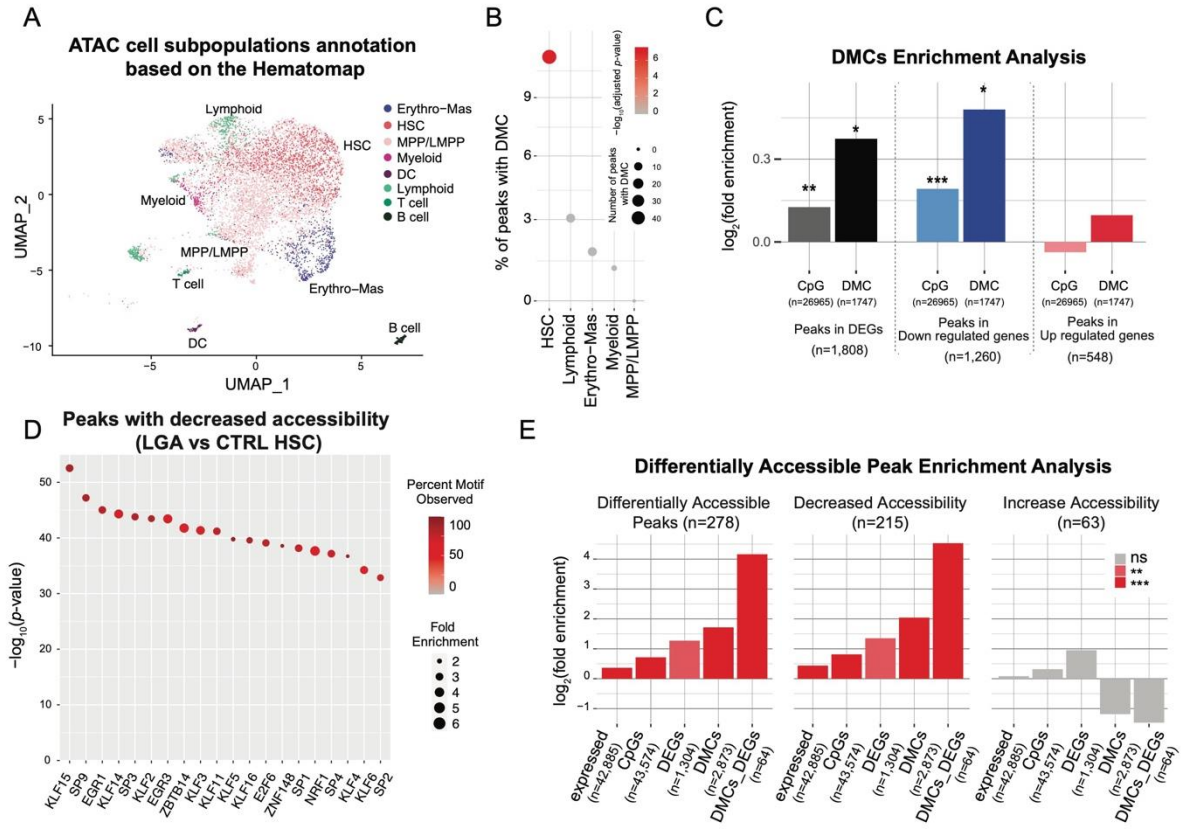


Figure 4. Chromatin accessibility analysis. (**A**) UMAP representing HSPCs lineage based on chromatin accessibility. Annotations are based on the Hematomap using the transfer label approach. (**B**) Dot plot representing enrichment for DMC within lineage-specific peaks. (**C**) Bar plots representing enrichment for peaks containing CpG or peaks containing DMC associated to DEGs, up-regulated and down-regulated genes (* p-value < 0.05; ** p-value < 0.01, *** p-value < 0.001, hypergeometric test). (**D**) Dot plot representing enrichment for transcription factor motif within Down peaks identified comparing chromatin accessibility between LGA and CTRL. Dots are color-coded based on percentage of peaks with motif and y-axis represents the significance of the enrichment. (**E**) Bar plots representing enrichment analysis considering accessible, down, and up peaks. Enrichment is performed using peaks in expressed genes (expressed), peaks with CpGs (CpGs), peaks in DEG (DEGs), peaks with DMC (DMCs), and peaks in DEG with DMC (DMCs_DEGs) as reference gene sets (** p-value < 0.01, *** p-value < 0.001, hypergeometric test).

2.4. EGR1, KLF2, and KLF4 Are Key Upstream Regulators Influenced by Early Epigenetic Programming in LGA

To further characterize the molecular mechanisms affected in LGA HSCs and identified master regulators, we leveraged the single-cell resolution of our approaches to perform a co-regulatory network analysis. This approach allowed us to model the influence of upstream transcription factors (TF) on expression changes of downstream target genes. We performed co-expression analysis to identify genes co-regulated by the same TF, i.e., regulons, and filter each regulon based on the presence of TF motif within a cis-regulatory region (SCENIC). We identified a total of 250 regulons but only considered for further analyses the 106 regulons identified based on high confidence cis-regulatory motif. These regulons only rely on associations for which the presence of the TF motif was experimentally validated. We then scored the regulons activity in each cell using gene expression profiles of the entire regulons (AUCell). We observed that lineage-specific regulons are associated with concordant lineage determining hematopoietic TFs such as GATA2, GATA3, MEIS1, TAL1, TCF3, EGR1, CEBPB, HOXB4, SPI1, and STAT1/3 further supporting our subpopulation annotation and the SCENIC approach (Supplemental Figure S5A, Supplemental Table S5) [18].

To identify TF associated with the changes in gene expression observed in LGA HSC, we compared the regulon activity between CTRL and LGA. We found seven regulons with a significant decrease in activity in the LGA HSC population (adjusted *p*-value < 0.001 and |activity score fold change| > 10%, Supplemental Table S6). No regulons were upregulated. These regulons were associated with ARID5A, EGR1, KLF2, KLF4, KLF10, FOSB, and JUN (Figure 5A). Among them, ARID5A, EGR1, KLF2, FOSB, and JUN were part of the 10 top active regulons in HSCs (Supplemental Table S7). Based on functional enrichment analysis using as reference GO:BP gene sets, and HSC signatures of quiescence or proliferative state [19], we showed that these regulons were enriched in genes regulating stress response, proliferation, and HSC differentiation (Figure 5B).

To further support the association between change in DNA methylation and change in gene expression previously identified at the gene level, we performed GSEA analysis to identify regulon enriched for both differentially methylated and differentially expressed genes. We found 9 regulons enriched in both hypermethylated and downregulated genes (adjusted *p*-value < 0.01 and NES < −1.6), including the differentially active regulons ARID5A, EGR1, FOSB, JUN, KLF2, and KLF4 (Figure 5C). We also found 9 regulons enriched in hypermethylated and upregulated genes (adjusted *p*-value < 0.01 and NES > 1.6) with key HSPC-specific regulons such as SPI1 promoting myeloid differentiation [20] and HOX family (HOXA9, HOXA10, HOXB4) promoting HSPC expansion (Figure 5C) [21–23].

To confirm the putative influence of methylation change on TF activity, we performed TF motif analysis considering the proximal regions surrounding each DMCs (±20 bp). We found significant enrichment for 23 TF motifs (adjusted *p*-value < 0.05, Figure 5D). Among them, we found EGR1 and several members of the Kruppel-like factors (KLF) family: KLF14, KLF5, KLF1, and KLF6. Furthermore, by taking advantage of our single-cell ATAC-seq data, we looked at the enrichment of the TF motif in open chromatin regions of HSC containing DMCs. We found a strong enrichment in EGR1, KLF2, and KLF4 motifs indicating that DNA methylation change occurred in active regions of the EGR1/KLF2/KLF4 TF network (Figure 5E).

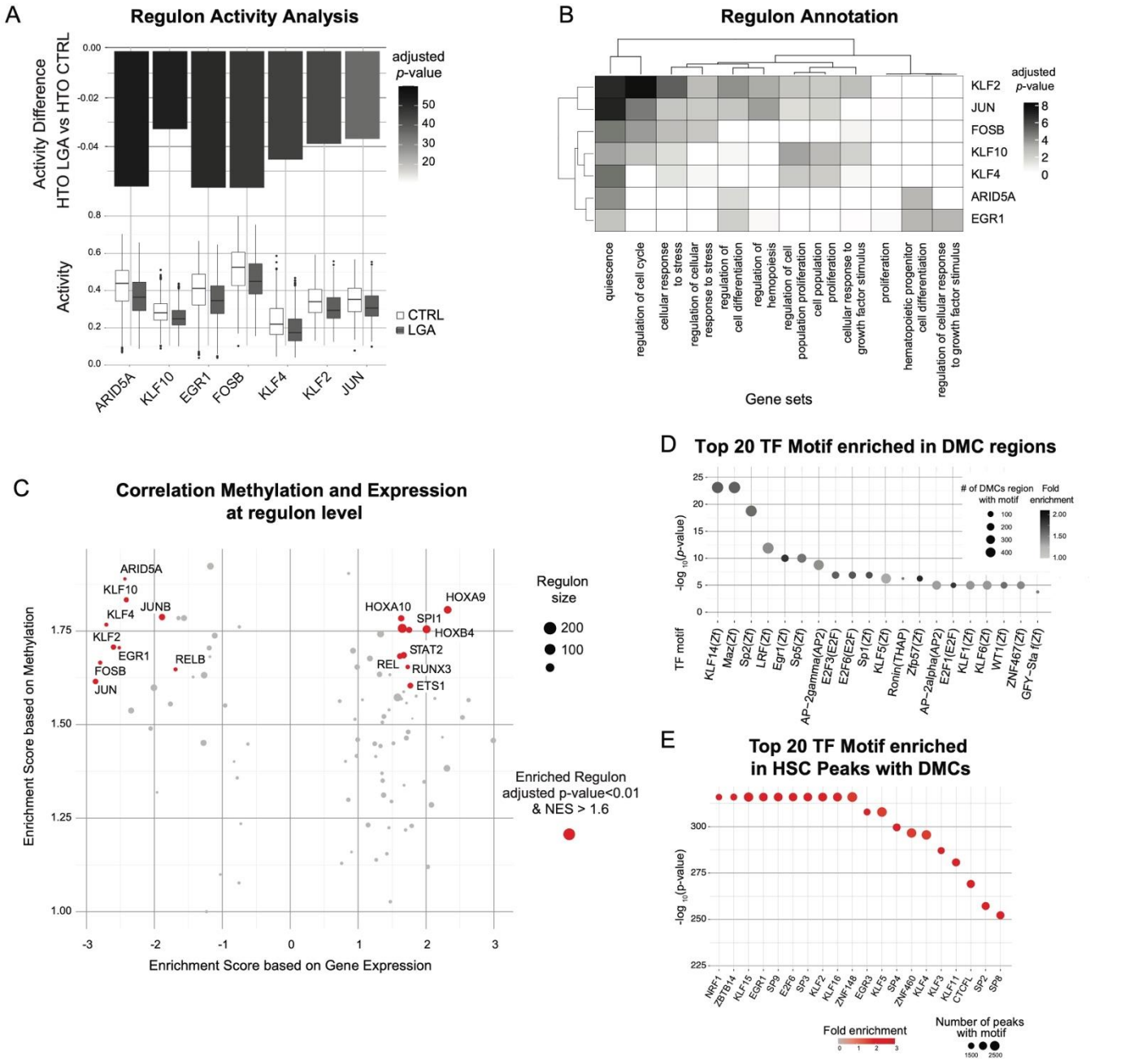


Figure 5. Epigenetic programming of HSC-specific regulons altered in LGA neonates. Regulons and TF target information were obtained through the SCENIC workflow. (**A**) Boxplots representing regulon activity score in CTRL and LGA HSC lineage. Barplot representing the change in regulon activity and significance comparing LGA vs. CTRL. Only significantly affected regulons are represented (adjusted *p*-value < 0.001 and |activity score fold change| > 10%). (**B**) Heatmap representing association between altered regulons and selected gene sets annotation. (**C**) Volcano plot representing enrichment in the change in expression and DNA methylation in regulons. Regulons enriched considering both expression and methylation (adjusted *p*-value < 0.01 and NES > 1.6) are in red. (**D**) Dot plot representing enrichment for TF binding motifs using HOMER considering a ±20 bp region around DMCs. Dots are color-coded based on the significance of the enrichment and y-axis represent the number of regions with binding motif among DMCs. (**E**) Dot plot representing enrichment for TF binding motifs using HOMER considering peaks with DMCs. Dots are color-coded based on the fold-enrichment and y-axis represents the significance of the enrichment.

2.5. Multimodal Co-Regulatory Network Recapitulating TF-Gene Interactions Influenced by Early Epigenetic Programming in LGA

Based on the integration of the DNA methylation, single-cell ATAC-seq, and single-cell RNA-seq data, we built a network recapitulating interaction between main TFs and downstream target genes within the principal regulons altered in LGA neonates: EGR1, KLF2, and KLF4 (Figure 6). EGR1, KLF2, and KLF4 regulons rely on highly interconnected (co-regulated) genes (Figure 6A). For each target gene, we confirmed the presence of a unique or shared upstream TF binding motif within the open chromatin regions. We observed a high concordance between the regulons and open chromatin motif analysis: 96%, 91%, and 95% of genes included in EGR1, KLF2, and KLF4 regulons, respectively, were associated with at least one peak containing the corresponding TF motif supporting the association between genes and TFs. We then looked for evidence of epigenetic modifications that may alter TF-target interactions. We annotated genes with associated open chromatin regions containing at least one DMC (middle area) or identified as differentially accessible between CTRL and LGA (inside area) (Figure 6B). Overall, 23% (n = 27) of genes targeted by these TFs networks have epigenetic alteration (DMCs or decrease accessibility) in open chromatin regions while 22% (n = 26 genes) appear downregulated in LGA. Finally, we highlight KLF2 as possible master regulators influenced by early programming. Indeed, we identified KLF2 as a hypermethylated and downregulated gene that interacts directly with EGR1 and KLF4 suggesting the downstream influence of KLF2 on these TFs. Conversely, KLF2 was not identified as part of EGR1 and KLF4 regulons suggesting that KLF2 is not a target of these TFs. This network also further validated *JUNB* and *SOCS3* being highly epigenetically altered in cis-regulatory regions (Figure 6C), as well as *ID1*, *CDKN1A*, *IER2*, *IER3*, and *IER5* as key downstream altered targets of KLF2, EGR1, and/or KLF4, again highlighting how early programming alters signaling involved in the regulation of cell proliferation and differentiation.

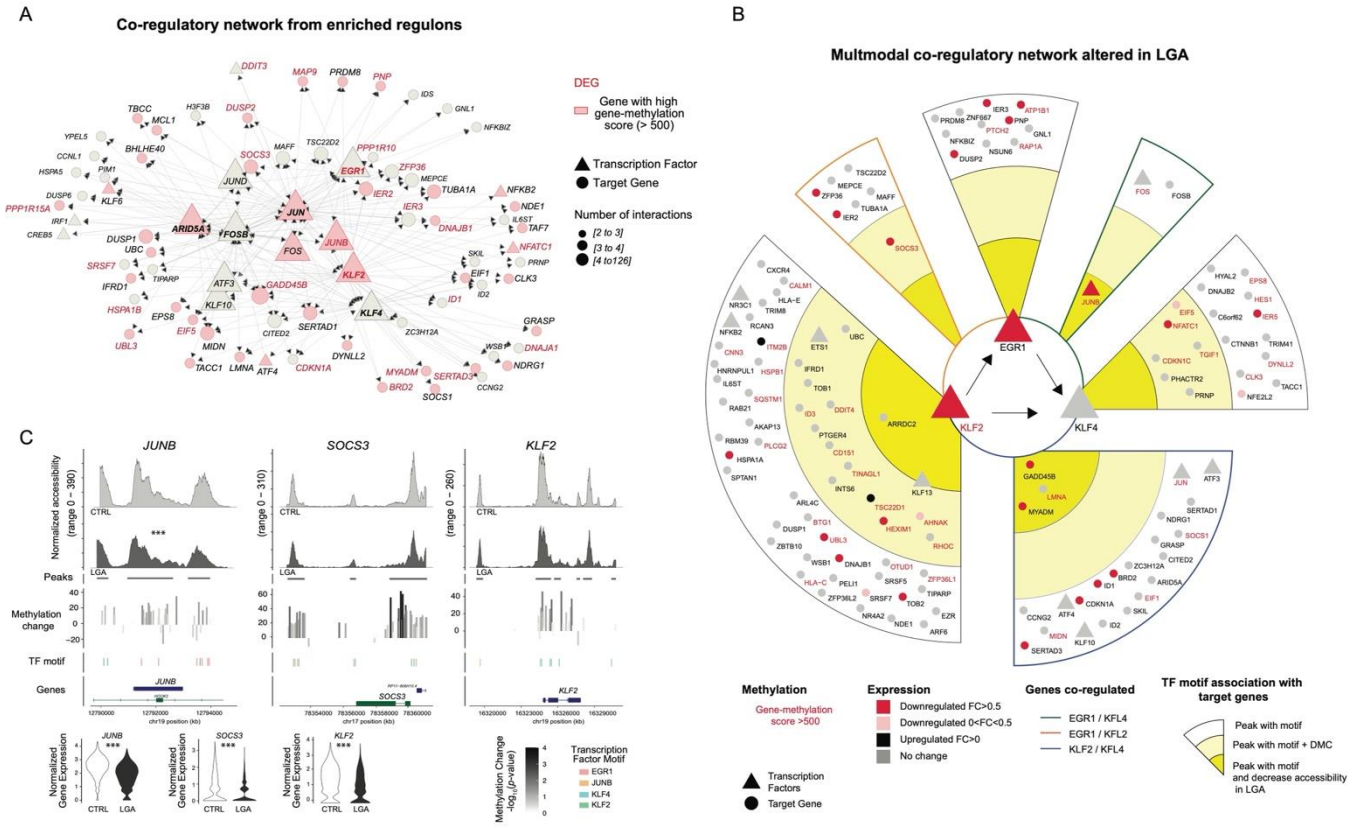


Figure 6. Network recapitulating interaction between the epigenomic and transcriptomic alterations in LGA. (**A**) Network representing interactions between target genes and transcription factors considering our top affected regulons ARID5A, EGR1, KLF2, KLF4, FOSB, and JUN. Each dot represents a gene within the network, the triangle represents a transcription factor, the arrow represents the interaction between the transcription factor and target genes, shapes are color-coded to reflect the change in gene-methylation score, and DEGs are labeled in red. Size of the shape represents the number of interactions. Only genes with two or more interactions are represented. (**B**) Tracks representing DNA methylation and chromatin accessibility for selected representative regions**.** Histogram representing change in DNA methylation at CpG level comparing LGA vs CTRL. Violin plot representing gene expression for selected genes (**C**) Network-based on the integration of DNA methylation, gene expression, and chromatin accessibility representing transcription factors and downstream target interactions within EGR1, KLF2, and KLF4 regulons. Only genes associated with peaks with TF motifs of interest are annotated. Donuts represent different levels of interactions. ***: significant change of peak accessibility (logistic regression) or gene expression (Wilcoxon test) in LGA compared to Control HSCs, adjusted *p*-value < 0.001.

2.6. In Vitro Analysis Confirms the Alteration of HSPCs Differentiation Capacities in LGA

Our integrative analyses highlighted epigenetic and transcriptomic alterations targeting signaling pathways involved in the regulation of HSC differentiation and proliferation. Thus, we decided to challenge HSPC differentiation and proliferation potential in vitro using colony-forming unit (CFU) assays. After 14 days of expansion, colonies from 4 CTRL and 4 LGA samples were classified into three categories: those derived from common myeloid progenitors (CFU_GEMM), erythroid progenitors (BFU-E), and granulocyte-macrophage progenitors (CFU_GM) based on the morphology of each colony. We observed a significant decrease in the number of common myeloid progenitor colonies in LGA samples (p-value < 0.05; Figure 7A) as well as striking differences in shape and size of more differentiated colonies (Figure 7B). CFU_GEMM colonies are the product of a non-committed hematopoietic progenitor able to differentiate in both erythroid and myeloid lineage. In our samples, only HSC and MPP have these features, suggesting that the decreased CFU GEMM proportion in LGA reflects either fewer HSC/MPP in starting cell subpopulation composition or a decreased proliferation and differentiation capacity of these cells.

To evaluate these two possibilities, we monitored cell population distribution across conditions at molecular resolution using our single-cell expression dataset. We observed a decrease in HSC cells (p-value = 0.015) and a trend toward increased MPP cells (p-value = 0.13, Figure 7C) in LGA compared to CTRL. Another way to look at population shift is to use pseudotime, i.e., a measure that reflects how far an individual cell is in a differentiation process. Indeed, cord-blood-derived CD34+ HSPCs represent a heterogeneous population of cells ranging from progenitors to progressively restricted cells of the erythroid, myeloid, or lymphoid lineages as confirmed by our single-cell transcriptomic analysis. To follow cell distribution through these levels of differentiation and assess the influence of the LGA environment we used the pseudotime tool from Monocle [24] Collecting the pseudotimes across our different cell populations, we observed a positive correlation between pseudotime and lineage differentiation as expected (r = 0.99, Pearson correlation, Figure 7D, Supplemental Figure S6A). We then compared the distribution of the pseudotime between LGA and CTRL using the least differentiated cells as roots, i.e., the long-term HSCs. At the population level, we observed an increase in pseudotime in LGA (p-value < 0.001, Figure 7E). Indeed, we observed a decrease in the number of cells presenting pseudotime associated with the HSC state in LGA samples (p-value < 0.05) and a shift toward cells presenting elevated pseudotime suggesting that LGA HSCs exit quiescence and differentiate more quickly compared to CTRL HSCs (Figure 7E). Altogether, our analysis supports the association between LGA exposure and cell growth signaling targeted by DNA methylation and gene expression changes with alteration of differentiation and proliferation capacities.

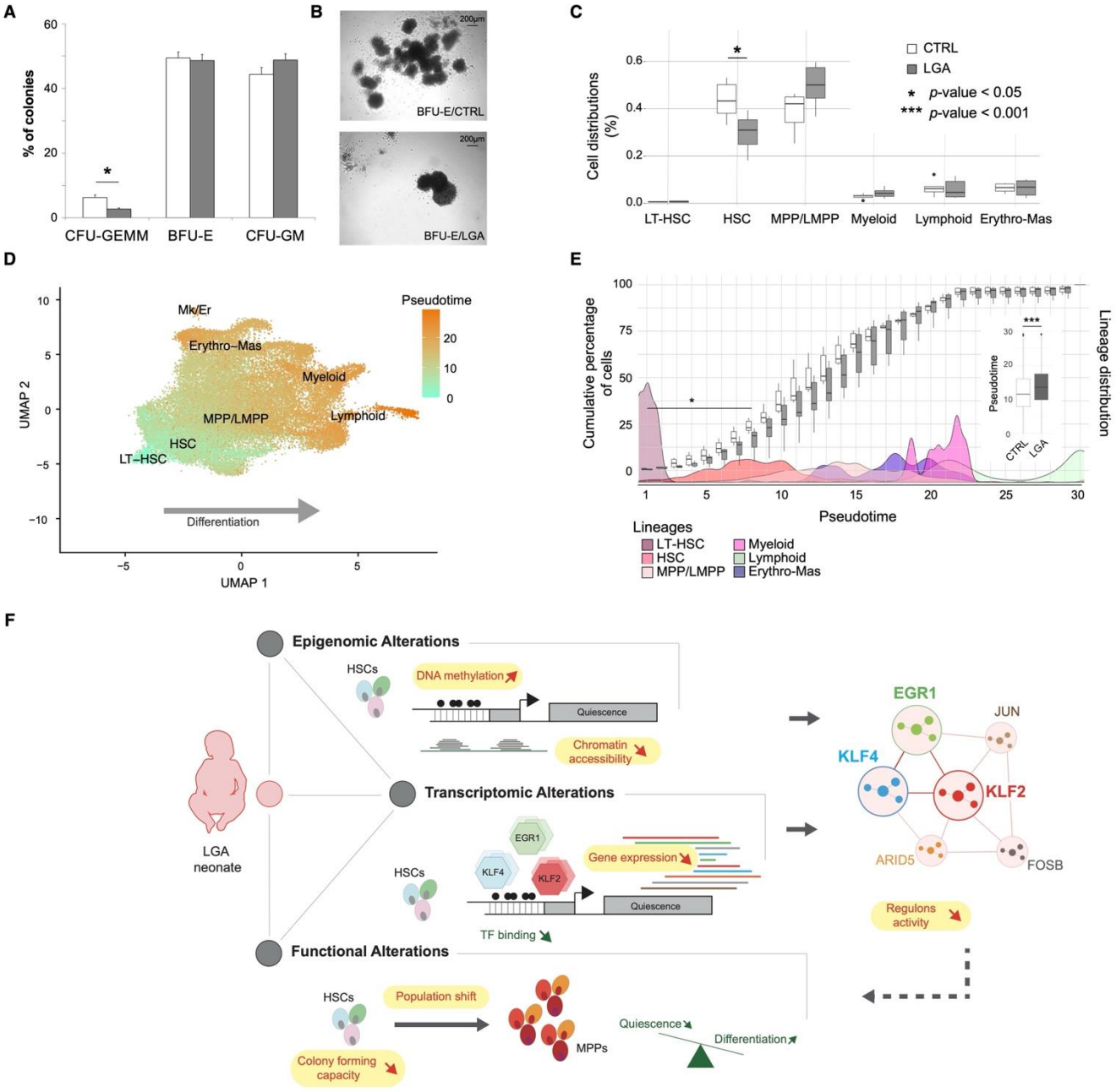


Figure 7. LGA is associated with decreased expansion capacity and an HSC shift toward more differentiated cells. (**A**) Bar plot representing colonies distribution after CFU assays. (**B**) Representative capture of colonies' morphological differences found in CTRL and LGA. (**C**) Boxplots representing the cell distribution across hematopoietic main lineages in CTRL and LGA. (**D**) UMAPs representing pseudotimes across lineages. (**E**) Box plots representing the cumulative percentage of cells per pseudotime in CTRL and LGA. Boxplots in the vignette represent overall pseudotime distribution in CTRL and LGA. Density plots correspond to cell populations distribution across pseudotimes. (**F**) Model recapitulating the influence of LGA on the hematopoietic compartment. (LT-HSC, long-term hematopoietic stem cell; HSC, hematopoietic stem cell; MPP, multipotent progenitor; LMPP, lymphoid-primed multipotent progenitors; Erythro-Mas, erythroid and mast precursor; Mk/Er, megakaryocyte and erythrocyte; DC, dendritic cell; CFU-GEMM, common myeloid progenitors; BFU-E, erythroid progenitors; CFU-GM, granulocyte-macrophage progenitors.

3. Discussion

Here, we interrogated three major layers of the regulatory landscape in cord-blood-derived CD34+ HPSCs, DNA methylation, chromatin conformation, and gene expression. We characterized, in-depth and at single-cell resolution, the functional consequences associated with early DNA methylation changes observed in LGA neonates. Through, the integration of multiple datasets and the development of novel analytical approaches, we addressed a very challenging aspect of functional (epi)genomics, the interpretation of DNA methylation changes. Focusing on HSPCs, we believe that we contributed to a better understanding of how early environment shapes the hematopoietic compartment development and long-term function.

We demonstrated in LGA neonates a correlated increase in DNA methylation and change in chromatin accessibility associated with decreased expression of downstream target genes under the influence of key HSC transcription factors EGR1, KLF2, and KLF4. EGR1, KLF2, and KLF4 are zinc-finger transcription factors involved in HSC quiescence signaling. EGR1 has a known role in regulating cell growth, development, and stress response in many tissues. In HSPC, EGR1 plays a role in the homeostasis of HSCs regulating proliferation [25]. EGR1 promotes quiescence and decreases through differentiation. Interestingly, EGR1 has also been shown to interact with epigenetic regulators forming a complex with DNMT3 and HDAC1 [26] suggesting a possible role in the epigenetic remodeling observed in LGA HSC. The KLF family is implicated in key stem cell functions. KLF4 is the most well-known factor in this family due to its role in reprogramming somatic cells into induced pluripotent stem cells [27]. KLF4 has been identified as a target for PU.1 transcription factor required for lineage commitment in HSPCs [28]. KLF2 and KLF4 promote self-renewal in embryonic stem cells [29] but no study has looked specifically at KLF2 and KLF2/KLF4 interactions in HSPCs. Our data suggest direct and indirect (shared downstream target) interactions between these three transcription factors in HSPCs. EGR1, KLF2, and KLF4 represent targets to be further explored in order to challenge causality. Still, our findings lead to a better understanding of how early exposure can affect long-term hematopoietic maintenance in humans via epigenetic programming of the EGR1, KLF2, and KLF4 signaling. Furthermore, these coordinated epigenetic and transcriptomic changes target genes regulating growth signaling, such as *SOCS3*, *SIRT1,* and *SESN2* [30–32]. Alteration of growth signaling highlights the tight correlation between in utero environment and the epigenetic programming. Indeed, excessive fetal growth observed in LGA neonates results in part from gestational hyperglycemia, dyslipidemia, or over secretion of placental insulin-like growth factors [33–35]. Altogether, these results further illustrate how DNA methylation and chromatin accessibility are key co-epigenetics actors regulating TF activity. Such interplay was already observed in the context of lineage commitment [36,37], but not yet in the context of developmental programming of HSCs. This highlights the interest in considering both methylation and chromatin rearrangement in fetal programming studies to decipher putative epigenetic imprinting and functional consequences.

Interestingly, EGR1, KLF2, and KLF4 are not only involved in the regulation of proliferation and differentiation *per se* but are key factors of the immediate, early response involved in stimulation-related cell activation. *EGR1* and *KLF2* expression increase in response to extrinsic stimulation. Elevated *EGR1* and *KLF2* expression promote self-renewal and quiescence in HSC [25,29]. Our transcriptomic data suggests that such activation may be occurring in our samples with the activation of stress-related signaling. The primary scope of our study was not to characterize the environmental exposure that would trigger such responses. However, one can speculate that the activation could result from stress due to cold exposure or handling time inherent to sample preparation. Still, the decreased activity observed in LGA suggests that LGA HSCs' capacities to respond to environmental challenges are diminished. This hypothesis fits with the concept of early programming in which disease susceptibility relies not only on early impairment of organ development but also on a decreased adaptability to further environmental challenges to

trigger disease [38]. Indeed, fine-tuning HSC quiescence mechanisms is of crucial relevance for optimal hematopoiesis. Not responsive dormant HSC would lead to hematopoietic failure due to a lack of differentiated blood cells. Although highly responsive HSC would lead to exhaustion of the population and a lack of long-term maintenance of the hematopoietic system [39].

To validate findings from our integrative approach, we challenged HSPCs in vitro and found a significant decrease in the number of CFU-GEMM colonies, colonies containing both erythroid and myeloid cells. These colonies are likely to originate from HSC or MPP cells, as only these cells have this multi-potential. These alterations could result from the decreased differentiation and proliferation capacities of these CD34+ cells or a decrease in their initial proportion in LGA cord blood. Our data suggest that both are altered in LGA. Indeed, the cell population analysis at the transcriptomic level revealed a decrease in HSCs in LGA neonates but a tendency to an increase in MPPs. We also observed epigenomic and transcriptomic alterations in signaling pathways and transcription factors regulating differentiation and proliferations of HSCs. Yet, this loss of stemness capacities in HSC is likely to drive the decrease in HSC subpopulations observed in our data and the decreased colony-forming capacity.

These findings corroborate previous studies on the developmental programming of the hematopoietic system [9,10]. A reduction in self-renewal of HSPCs and increased differentiation in both lymphoid and myeloid lineages have been observed in a mouse model of maternal obesity [8]. These effects may drive long-term consequences in human health as illustrated by the study performed by Kotowski et al. in which the integrity of the hematopoietic system in neonates was associated with susceptibility to onset of hematopoietic pathologies [40].

Hematopoietic stem cell differentiation and self-renew rely on a synergic interplay between genetically encoded signaling, cell-intrinsic, and cell-extrinsic factors as well as epigenetic modifiers [41]. This interplay appears altered in LGA neonates. We here provide a comprehensive model recapitulating the functional influence of the epigenetic early programming on HSPCs fitness to later environmental exposure (Figure 7F). We also linked LGA-associated epigenetic modifications to gene expression and functional alterations through a novel integrative approach. In this regard, we identified targets to be further explored. We also brought a better understanding of how early exposure can affect long-term tissue maintenance via epigenetic programming of EGR1, KLF2, and KLF4 associated regulation of growth signaling.

4. Methods

See the Supplemental Methods for additional information.

4.1. Clinical Sample Collection

Cord blood samples were obtained from CTRL and LGA neonates. LGA were defined by birth weight and ponderal index values greater than the 90th percentile for gestational age and sex. Control infants had normal parameters (between 10th and 90th percentiles) for both birth weight and ponderal index. Maternal and infant characteristics are shown in Supplemental Table S8.

4.2. Isolation of CD34+ HSPCs

Mononuclear cells were separated using PrepaCyte-WBC following which CD34+ cells were obtained by positive immunomagnetic bead selection, using the AutoMACS Separator (Miltenyi Biotech, Cologne, Germany). Cells were cryopreserved in 10% dimethyl sulfoxide using controlled rate freezing upon analysis.

4.3. Genome-Wide DNA Methylation Assay

DNA methylation levels for >1.7 M CpGs were obtained using the HELP-tagging assay as previously described [42].

4.4. Single-Cell RNA Sequencing Libraries Preparation

After cell count and viability check, the cell suspension was loaded into the Chromium controller (10x Genomics, Pleasanton, California, US) and library was generated using the chromium single-cell v3 chemistry following manufacturer recommendations. Gene expression library was sequenced using 100 bp paired-end reads on the Illumina NovaSeq 6000 system (Illumina, San Diego, California, US).

4.5. Single-Cell ATAC Sequen1cing Libraries Preparation

After cell count and viability check, nuclei were isolated from cell suspension and incubated with transposase. Transposed nuclei were then loaded into the Chromium 10x Genomics controller and library was generated using the chromium single-cell ATAC v1.1 chemistry following manufacturer recommendations. Gene expression library was sequenced using 150 bp paired-end reads on the Illumina NovaSeq 6000 system.

4.6. HTO Protocol

After cell counting and viability check and prior to cell suspension loading on the Chromium controller, cell hashtag (HTO) staining (Biolegend, San Diego, California, US) was used following the cell-hashing protocol [17].

4.7. Colony Forming Unit Assay

To assess clonogenic progenitor frequencies, 3×10^3 CD34+ HSPC cells were plated in methylcellulose containing SCF, GM-CSF, IL-3, and EPO (H4434; STEMCELL Technologies, Vancouver, Canada). Colonies were scored 14 days later.

4.8. Data Processing and Statistical Analysis

For DNA methylation analysis, low-quality CpGs were filtered out based on detection rate and confidence score. 754,931 out of 1,709,224 CpGs were conserved for further analysis. Linear regression and statistical modeling using the LIMMA R package [43] were used to identify differentially methylated CpGs (DMC) including maternal age, sex, ethnicity, batch, and library complexity in the linear model. We assessed enrichment for biological pathways performing GSEA using the ClusterProfiler package [44]. We performed transcription factor (TF) motif enrichment analysis using the HOMER tool [45] considering a 20 bp region around the DMCs.

For single-cell RNAseq (scRNA-seq) analysis, data were preprocessed using the CellRanger count pipeline (10x Genomics). Data filtering, normalization, and integration as well as cluster identifications were performed using Seurat (v4) pipeline. Pseudo-bulk differential expression analysis between LGA and CTRL cells within each hematopoietic lineage was performed using DESeq2 R package including batches and sex of samples in the negative binomial model[46]. Over representation test was performed on differentially expressed genes (DEGs) using enrichGO and enrichKEGG of the ClusterProfiler Package. The SCENIC workflow [47] was used to identify co-regulated genes module associated to a TF (regulons) and to generate cell-specific activity scores for each regulon. Differentiation trajectory analysis and pseudotime attribution were conducted with Monocle [24].

For single-cell ATAC-seq, data were preprocessed using the CellRanger ATAC pipeline (10x Genomics). Data filtering, normalization, and integration as well as clustering were performed using the Signac pipeline. Cell type identification was based on scRNA-seq annotation using a label transfer approach. Peaks calling at lineage level was performed using the MACS2 tool. Peaks specific to each lineage or differentially

accessible between LGA and Control were identified using the FindMarkers function with Logistic Regression (LR) models including cellular sequencing depth as a latent variable. TF motif enrichment on lineage or group-specific peaks was performed using the FindMotifs function. All peaks enrichment analysis was performed using hypergeometric tests. For final Gene Regulatory Network (GRN) construction, TF target interactions inferred with SCENIC were filtered out based on the presence of a corresponding TF motif in the peak associated with the target. Supplemental Table S9 contains information on the number of cells per sample.

4.9. Gene-Methylation Score

To compute the gene-methylation score, 2 steps were needed: (1) to generate a CpG score that reflects the association between CpG and gene, and (2) to concatenate CpG-scores at the gene level.

(1) CpG-score

CpGScore = (−log10(p_{cpg}) × meth.change) × LinkWeight × RegWeight

Where p_{cpg} is the nominal *p*-value of the differential methylation analysis, and meth.change is the difference between the percentage of methylation in LGA and the percentage of methylation in CTRL. LinkWeight represents the confidence in CpG-gene association and RegWeight represents the estimated regulatory influence of the considered CpG based on CD34+ specific genomic annotation defined using CD34+ specific histone marks as previously described [11] and EnsRegScore refers to regulatory regions defined based on the Ensembl Regulatory build hg19 genome annotation [48].

(2) To concatenate CpG-Scores at gene level: gene-methylation score

To summarize the CpG methylation change at the gene level, we aggregated the CpG-Scores into a methylation gene score by taking care to (i) alleviate the arbitrary number of CpGs per gene and (ii) interpret differently CpG influences located on the promoter of them in others genomic region.

The gene-methylation score is defined as:

Gene-methylation score = $\left(\sum CpG\ Score \times Weight_{n\_cpg}\right)_{\text{promoter}} + \left(\sum |CpG\ Score| \times Weight_{n\_cpg}\right)_{\text{other\_regions}}$

Where the $Weight_{nCpG}$ was optimized to alleviate the influence of the number of CpGs linked to a gene and defined as:

$$Weight_{nCpG} = \sqrt[3,8]{\frac{1}{\sum \frac{1}{|\cdots\cdots\cdots+1|}}}$$

The code to perform the analyses in this manuscript is available at https://github.com/umr1283/LGA_HSPC_PAPER.git (last accession date : 29th June 2022).

Supplementary Materials: The following supporting information cand be downloaded at: www.mdpi.com/xxx/s1 , Figure S1: Methylation data processing ; Figure S2: HSPC subpopulations analysis ; Figure S3: DEG analysis; Figure S4: ATAC-seq data processing ; Figure S5: TFs and pseudotime lineage specific characterization ; Supplementary Tables.

Author Contributions: A.P., A.C., Y.M.Z., E.D., L.B.-F. and F.D. were responsible for conducting research and analyzing data. M.D. and M.C. provided feedback on the data analysis. A.P., A.C., F.D., A.B. and P.F. contributed to writing the manuscript. F.D., F.H. and J.G. were responsible for designing the study. All authors have read and agreed to the published version of the manuscript.

Funding: Support for this project was provided by the Roadmap Epigenomics Program, R01 HD063791 (Einstein/Greally). Support was also provided by Einstein's Center for Epigenomics, including the Epigenomics Shared Facility and Computational Epigenomics Group.

Institutional Review Board Statement: : This study was approved by the Institutional Review Board of the Montefiore Medical Center and the Committee on Clinical Investigation at the Albert

Einstein College of Medicine and is in accordance with Health Insurance Portability and Accountability Act regulations.

Informed Consent Statement: This study was approved by the Institutional Review Board of the Montefiore Medical Center and the Committee on Clinical Investigation at the Albert Einstein College of Medicine and is in accordance with Health Insurance Portability and Accountability Act regulations. Written informed consent was obtained from all subjects before participation.

Data Availability Statement: The DNA methylation and gene expression data will be made available upon request to A.P., P.F. or F.D.

Acknowledgments: The authors thank the UMR 8199 LIGAN-PM Genomics platform (Lille, France) which belongs to the 'Federation de Recherche' 3508 Labex EGID (European Genomics Institute for Diabetes; ANR-10-LABX-46) and was supported by the ANR Equipex 2010 session (ANR-10-EQPX-07-01; 'LIGAN-PM'). The LIGAN-PM Genomics platform (Lille, France) is also supported by the FEDER and the Region Nord-Pas-de-Calais-Picardie. This project is cofunded in the frame of CPER CTRL program by the European Union—European Regional Development Fund (ERDF), Hauts de France Region (contract n°17003781), Métropole Européenne de Lille (contract n°2016_ESR_05), and French State (contract n°2017-R3-CTRL-Phase 1). The present work was also supported by the National Center for Precision Diabetic Medicine—PreciDIAB, which is jointly supported by the French National Agency for Research (ANR-18-IBHU-0001), by the European Union (FEDER), by the Hauts-de-France Regional Council and by the European Metropolis of Lille (MEL) and by the European Research Council (ERC Reg-Seq—715575). We thank "France Génomique" consortium (ANR-10-INBS-009).

Conflicts of Interest: The authors declare no competing financial interests in relation to the work described.

References

1. Doulatov, S.; Notta, F.; Laurenti, E.; Dick, J.E. Hematopoiesis: A human perspective. *Cell Stem Cell* **2012**, *10*, 120–136. https://doi.org/10.1016/j.stem.2012.01.006.
2. Eaves, C.J. Hematopoietic stem cells: Concepts, definitions, and the new reality. *Blood* **2015**, *125*, 2605–2613. https://doi.org/10.1182/blood-2014-12-570200.
3. Balistreri, C.R.; Garagnani, P.; Madonna, R.; Vaiserman, A.; Melino, G. Developmental programming of adult haematopoiesis system. *Ageing Res. Rev.* **2019**, *54*, 100918. https://doi.org/10.1016/j.arr.2019.100918.
4. Pardali, E.; Dimmeler, S.; Zeiher, A.M.; Rieger, M.A. Clonal hematopoiesis, aging, and cardiovascular diseases. *Exp. Hematol.* **2020**, *83*, 95–104. https://doi.org/10.1016/j.exphem.2019.12.006.
5. Lee, C.C.; Fletcher, M.D.; Tarantal, A.F. Effect of age on the frequency, cell cycle, and lineage maturation of rhesus monkey (Macaca mulatta) CD34+ and hematopoietic progenitor cells. *Pediatr. Res.* **2005**, *58*, 315–322. https://doi.org/10.1203/01.PDR.0000169975.30339.32.
6. Chandel, N.S.; Jasper, H.; Ho, T.T.; Passegue, E. Metabolic regulation of stem cell function in tissue homeostasis and organismal ageing. *Nat. Cell Biol.* **2016**, *18*, 823–832. https://doi.org/10.1038/ncb3385.
7. Guenechea, G.; Gan, O.I.; Dorrell, C.; Dick, J.E. Distinct classes of human stem cells that differ in proliferative and self-renewal potential. *Nat. Immunol.* **2001**, *2*, 75–82. https://doi.org/10.1038/83199.
8. Kamimae-Lanning, A.N.; Krasnow, S.M.; Goloviznina, N.A.; Zhu, X.; Roth-Carter, Q.R.; Levasseur, P.R.; Jeng, S.; McWeeney, S.K.; Kurre, P.; Marks, D.L. Maternal high-fat diet and obesity compromise fetal hematopoiesis. *Mol. Metab.* **2015**, *4*, 25–38. https://doi.org/10.1016/j.molmet.2014.11.001.
9. Al-Sweedan, S.A.; Musalam, L.; Obeidat, B. Factors predicting the hematopoietic stem cells content of the umbilical cord blood. *Transfus. Apher. Sci.* **2013**, *48*, 247–252. https://doi.org/10.1016/j.transci.2013.01.003.
10. Chandra, T.; Afreen, S.; Kumar, A.; Singh, U.; Gupta, A. Does umbilical cord blood-derived CD34+ cell concentration depend on the weight and sex of a full-term infant? *J. Pediatr. Hematol. Oncol.* **2012**, *34*, 184–187. https://doi.org/10.1097/MPH.0b013e318249adb6.
11. Delahaye, F.; Wijetunga, N.A.; Heo, H.J.; Tozour, J.N.; Zhao, Y.M.; Greally, J.M.; Einstein, F.H. Sexual dimorphism in epigenomic responses of stem cells to extreme fetal growth. *Nat. Commun.* **2014**, *5*, 5187. https://doi.org/10.1038/ncomms6187.
12. Wijetunga, N.A.; Delahaye, F.; Zhao, Y.M.; Golden, A.; Mar, J.C.; Einstein, F.H.; Greally, J.M. The meta-epigenomic structure of purified human stem cell populations is defined at cis-regulatory sequences. *Nat. Commun.* **2014**, *5*, 5195. https://doi.org/10.1038/ncomms6195.
13. Cabezas-Wallscheid, N.; Klimmeck, D.; Hansson, J.; Lipka, D.B.; Reyes, A.; Wang, Q.; Weichenhan, D.; Lier, A.; von Paleske, L.; Renders, S.; et al. Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. *Cell Stem Cell* **2014**, *15*, 507–522. https://doi.org/10.1016/j.stem.2014.07.005.

14. Karamitros, D.; Stoilova, B.; Aboukhalil, Z.; Hamey, F.; Reinisch, A.; Samitsch, M.; Quek, L.; Otto, G.; Repapi, E.; Doondeea, J.; et al. Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. *Nat. Immunol.* **2018**, *19*, 85–97. https://doi.org/10.1038/s41590-017-0001-2.
15. Zhang, Y.H.; Hu, Y.; Zhang, Y.; Hu, L.D.; Kong, X. Distinguishing three subtypes of hematopoietic cells based on gene expression profiles using a support vector machine. *Biochim. Biophys. Acta. Mol. Basis. Dis.* **2018**, *1864*, 2255–2265. https://doi.org/10.1016/j.bbadis.2017.12.003.
16. Zheng, S.; Papalexi, E.; Butler, A.; Stephenson, W.; Satija, R. Molecular transitions in early progenitors during human cord blood hematopoiesis. *Mol. Syst. Biol.* **2018**, *14*, e8041. https://doi.org/10.15252/msb.20178041.
17. Stoeckius, M.; Zheng, S.; Houck-Loomis, B.; Hao, S.; Yeung, B.Z.; Mauck, W.M., 3rd; Smibert, P.; Satija, R. Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome. Biol.* **2018**, *19*, 224. https://doi.org/10.1186/s13059-018-1603-1.
18. Zhu, J.; Emerson, S.G. Hematopoietic cytokines, transcription factors and lineage commitment. *Oncogene* **2002**, *21*, 3295–3313. https://doi.org/10.1038/sj.onc.1205318.
19. Venezia, T.A.; Merchant, A.A.; Ramos, C.A.; Whitehouse, N.L.; Young, A.S.; Shaw, C.A.; Goodell, M.A. Molecular signatures of proliferation and quiescence in hematopoietic stem cells. *PLoS Biol.* **2004**, *2*, e301. https://doi.org/10.1371/journal.pbio.0020301.
20. Tenen, D.G.; Hromas, R.; Licht, J.D.; Zhang, D.E. Transcription factors, normal myeloid development, and leukemia. *Blood* **1997**, *90*, 489–519.
21. Antonchuk, J.; Sauvageau, G.; Humphries, R.K. HOXB4-induced expansion of adult hematopoietic stem cells ex vivo. *Cell* **2002**, *109*, 39–45. https://doi.org/10.1016/s0092-8674(02)00697-9.
22. Sun, Y.; Zhou, B.; Mao, F.; Xu, J.; Miao, H.; Zou, Z.; Phuc Khoa, L.T.; Jang, Y.; Cai, S.; Witkin, M.; et al. HOXA9 Reprograms the Enhancer Landscape to Promote Leukemogenesis. *Cancer Cell* **2018**, *34*, 643–658 e645. https://doi.org/10.1016/j.ccell.2018.08.018.
23. Magnusson, M.; Brun, A.C.; Miyake, N.; Larsson, J.; Ehinger, M.; Bjornsson, J.M.; Wutz, A.; Sigvardsson, M.; Karlsson, S. HOXA10 is a critical regulator for hematopoietic stem cells and erythroid/megakaryocyte development. *Blood* **2007**, *109*, 3687–3696. https://doi.org/10.1182/blood-2006-10-054676.
24. Qiu, X.; Hill, A.; Packer, J.; Lin, D.; Ma, Y.A.; Trapnell, C. Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* **2017**, *14*, 309–315. https://doi.org/10.1038/nmeth.4150.
25. Min, I.M.; Pietramaggiori, G.; Kim, F.S.; Passegue, E.; Stevenson, K.E.; Wagers, A.J. The transcription factor EGR1 controls both the proliferation and localization of hematopoietic stem cells. *Cell Stem Cell* **2008**, *2*, 380–391. https://doi.org/10.1016/j.stem.2008.01.015.
26. Cartron, P.F.; Blanquart, C.; Hervouet, E.; Gregoire, M.; Vallette, F.M. HDAC1-mSin3a-NCOR1, Dnmt3b-HDAC1-Egr1 and Dnmt1-PCNA-UHRF1-G9a regulate the NY-ESO1 gene expression. *Mol. Oncol.* **2013**, *7*, 452–463. https://doi.org/10.1016/j.molonc.2012.11.004.
27. Takahashi, K.; Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **2006**, *126*, 663–676. https://doi.org/10.1016/j.cell.2006.07.024.
28. Feinberg, M.W.; Wara, A.K.; Cao, Z.; Lebedeva, M.A.; Rosenbauer, F.; Iwasaki, H.; Hirai, H.; Katz, J.P.; Haspel, R.L.; Gray, S.; et al. The Kruppel-like factor KLF4 is a critical regulator of monocyte differentiation. *EMBO. J.* **2007**, *26*, 4138–4148. https://doi.org/10.1038/sj.emboj.7601824.
29. Jiang, J.; Chan, Y.S.; Loh, Y.H.; Cai, J.; Tong, G.Q.; Lim, C.A.; Robson, P.; Zhong, S.; Ng, H.H. A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nat. Cell Biol.* **2008**, *10*, 353–360. https://doi.org/10.1038/ncb1698.
30. Marine, J.C.; McKay, C.; Wang, D.; Topham, D.J.; Parganas, E.; Nakajima, H.; Pendeville, H.; Yasukawa, H.; Sasaki, A.; Yoshimura, A.; et al. SOCS3 is essential in the regulation of fetal liver erythropoiesis. *Cell* **1999**, *98*, 617–627. https://doi.org/10.1016/s0092-8674(00)80049-5.
31. Matsui, K.; Ezoe, S.; Oritani, K.; Shibata, M.; Tokunaga, M.; Fujita, N.; Tanimura, A.; Sudo, T.; Tanaka, H.; McBurney, M.W.; et al. NAD-dependent histone deacetylase, SIRT1, plays essential roles in the maintenance of hematopoietic stem cells. *Biochem. Biophys. Res. Commun.* **2012**, *418*, 811–817. https://doi.org/10.1016/j.bbrc.2012.01.109.
32. Parmigiani, A.; Nourbakhsh, A.; Ding, B.; Wang, W.; Kim, Y.C.; Akopiants, K.; Guan, K.L.; Karin, M.; Budanov, A.V. Sestrins inhibit mTORC1 kinase activation through the GATOR complex. *Cell Rep.* **2014**, *9*, 1281–1291. https://doi.org/10.1016/j.celrep.2014.10.019.
33. Kaul, P.; Savu, A.; Yeung, R.O.; Ryan, E.A. Association between maternal glucose and large for gestational outcomes: Real-world evidence to support Hyperglycaemia and Adverse Pregnancy Outcomes (HAPO) study findings. *Diabet. Med.* **2022**, *39*, e14786. https://doi.org/10.1111/dme.14786.
34. Wang, J.; Moore, D.; Subramanian, A.; Cheng, K.K.; Toulis, K.A.; Qiu, X.; Saravanan, P.; Price, M.J.; Nirantharakumar, K. Gestational dyslipidaemia and adverse birthweight outcomes: A systematic review and meta-analysis. *Obes. Rev.* **2018**, *19*, 1256–1268. https://doi.org/10.1111/obr.12693.
35. Chen, K.Y.; Lin, S.Y.; Lee, C.N.; Wu, H.T.; Kuo, C.H.; Kuo, H.C.; Chuang, C.C.; Kuo, C.H.; Chen, S.C.; Fan, K.C.; et al. Maternal Plasma Lipids During Pregnancy, Insulin-like Growth Factor-1, and Excess Fetal Growth. *J. Clin. Endocrinol. Metab.* **2021**, *106*, e3461-e3472. https://doi.org/10.1210/clinem/dgab364.
36. Izzo, F.; Lee, S.C.; Poran, A.; Chaligne, R.; Gaiti, F.; Gross, B.; Murali, R.R.; Deochand, S.D.; Ang, C.; Jones, P.W.; et al. DNA methylation disruption reshapes the hematopoietic differentiation landscape. *Nat. Genet.* **2020**, *52*, 378–387. https://doi.org/10.1038/s41588-020-0595-4.

37. Wiench, M.; John, S.; Baek, S.; Johnson, T.A.; Sung, M.H.; Escobar, T.; Simmons, C.A.; Pearce, K.H.; Biddie, S.C.; Sabo, P.J.; et al. DNA methylation status predicts cell type-specific enhancer activity. *EMBO. J.* **2011**, *30*, 3028–3039. https://doi.org/10.1038/emboj.2011.210.
38. Barker, D.J. In utero programming of chronic disease. *Clin. Sci. (Lond.)* **1998**, *95*, 115–128.
39. Wilson, A.; Laurenti, E.; Trumpp, A. Balancing dormant and self-renewing hematopoietic stem cells. *Curr. Opin. Genet. Dev.* **2009**, *19*, 461–468. https://doi.org/10.1016/j.gde.2009.08.005.
40. Kotowski, M.; Safranow, K.; Kawa, M.P.; Lewandowska, J.; Klos, P.; Dziedziejko, V.; Paczkowska, E.; Czajka, R.; Celewicz, Z.; Rudnicki, J.; et al. Circulating hematopoietic stem cell count is a valuable predictor of prematurity complications in preterm newborns. *BMC. Pediatr.* **2012**, *12*, 148. https://doi.org/10.1186/1471-2431-12-148.
41. Yu, V.W.C.; Yusuf, R.Z.; Oki, T.; Wu, J.; Saez, B.; Wang, X.; Cook, C.; Baryawno, N.; Ziller, M.J.; Lee, E.; et al. Epigenetic Memory Underlies Cell-Autonomous Heterogeneous Behavior of Hematopoietic Stem Cells. *Cell* **2017**, *168*, 944–945. https://doi.org/10.1016/j.cell.2017.02.010.
42. Suzuki, M.; Jing, Q.; Lia, D.; Pascual, M.; McLellan, A.; Greally, J.M. Optimized design and data analysis of tag-based cytosine methylation assays. *Genome. Biol.* **2010**, *11*, R36. https://doi.org/10.1186/gb-2010-11-4-r36.
43. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic. Acids. Res.* **2015**, *43*, e47. https://doi.org/10.1093/nar/gkv007.
44. Yu, G.; Wang, L.G.; Han, Y.; He, Q.Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **2012**, *16*, 284–287. https://doi.org/10.1089/omi.2011.0118.
45. Heinz, S.; Benner, C.; Spann, N.; Bertolino, E.; Lin, Y.C.; Laslo, P.; Cheng, J.X.; Murre, C.; Singh, H.; Glass, C.K. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **2010**, *38*, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004.
46. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome. Biol.* **2014**, *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.
47. Aibar, S.; Gonzalez-Blas, C.B.; Moerman, T.; Huynh-Thu, V.A.; Imrichova, H.; Hulselmans, G.; Rambow, F.; Marine, J.C.; Geurts, P.; Aerts, J.; et al. SCENIC: Single-cell regulatory network inference and clustering. *Nat. Methods* **2017**, *14*, 1083–1086. https://doi.org/10.1038/nmeth.4463.
48. Zerbino, D.R.; Wilder, S.P.; Johnson, N.; Juettemann, T.; Flicek, P.R. The ensembl regulatory build. *Genome. Biol.* **2015**, *16*, 56. https://doi.org/10.1186/s13059-015-0621-5.

I.2. Complementary works

I.2.a. Materiels and methods

I.2.a.i. Reagents, primers and siRNAs

REAGENT	REFERENCE	SUPPLIER
Human CD34+ Cell Nucleofector™ Kit	VPA-1003	Lonza
LKLF/KLF2 siRNA (h)	sc-35818	santa-cruz
Egr-1 siRNA (h)	sc-29303	santa-cruz
SiRNA Control-A	sc-37007	santa-cruz
RNeasy Micro Kit (50)	74004	Qiagen
Human Methylcellulose Complete Media	HSC003	RnD systems
Invitrogen™ SuperScript™ III Reverse Transcriptase	18080093	fisher scientific
dNTP mix (10 mM each)	18427013	thermofisher
Random Primers	48190011	thermofisher
primers EGR1 and KLF2	see primers table	

PRIMER_NAME	SEQUENCE	TARGET	PRODUCT SIZE
pEGR1_forward	TTCAACCCTCAGGCGGACAC	EGR1	71
pEGR1_reverse	GAGATGTCAGGAAAAGACTCTGCG	EGR1	71
pKLF2_forward	AGAGGGTCTCCCTCGATGAC	KLF2	100
pKLF2_reverse	CTCGTCAAGGAGGATCGTGG	KLF2	100

I.2.a.ii. RNA velocity analysis

RNA velocity analysis was performed using Velocito to produce the spliced and unspliced gene expression matrices[556] and scVelo to estimate the RNA velocity across cells through dynamical modelling[557]. We then used both RNA velocity and the pseudotime analysis (published results) to estimate the cell fate of the cell, i.e. toward which differentiation state the cell is going to. To calculate this cell fate, we used the cell transition matrix produced by scVelo. The transition matrix gives the probability for each cell to differentiate in other cells based on this newly synthetized mRNA (pre-mRNA). Then, by multiplying these probabilities by the pseudotime of others cells we get a prediction of the future pseudotime of the cells, i.e toward with differentiation state the cell go. By subtracting this predicted pseudotime by this actual pseudotime, we obtained its expected pseudotime shift, representing the intensity of the cell differentiation, i.e. its differentiation bias.

I.2.a.iii. Rapidly processed samples

To produce rapidly processed samples, we removed the HTO multiplexing part from the scRNA-seq protocol. To pool samples together and allow cell multiplexing, we pool samples depending on their sex, because samples can then be demultiplex based on sex specific transcripts. Each cryopreserved CD34+ cells from each sample were thawed in a water bath at 37°C 1min before to be

resuspended in 10ml of pre-heated medium. Cell suspensions were filtered with a MACS pre-separation filter 30 µm and centrifuged 5min at 300g. Cell pellets were resuspended in Deionized Phosphate Buffer Saline 1X (DPBS, GIBCO™, Fisher Scientific 11590476) with 0.04% Bovine serum albumin (BSA) for counting on a Corning Cytosmart cell counter by Trypan blue (Trypan Blue solution, 11538886, Fisherscientific) counterstaining for viability check. Samples was pooled by 2 based on their sex (1 male and 1 female by pool) and cell suspension was loaded on a Chromium 10x Genomics controller following the manufacturer protocol using the chromium single-cell v3 chemistry with single indexing. After library preparation, the pool was sequenced using 100pb paired-end reads on NOVAseq 6000 system following the manufacturer recommendations (Illumina). Gene expression matrices were generated using the CellRanger count pipeline.

I.2.a.iv. Gene silencing

We used electroporation (nucleofection) to transfect siRNAs targeting KLF2 transcripts on cells. This experiment require a large number of cells because induce significant cell loss during electroporation step (>50%). To obtain a maximum of cells, CD34+ cells were isolated from fresh CTRL cord blood and put in culture overnight in StemPro™ CD34+ Cell Medium (GIBCO™, Fisher Scientific). Then, the Human CD34+ Cell Nucleofector™ Kit (Lonza) was used to perform the siRNAs electroporation following manufacturer recommendation. After centrifugate for 10min at 200g, the cell pellet was resuspended in 100 µL of nucleofector solution and 5 µL of 1 µM siRNAs solution was added (final concentration= 50 nM). The cells + siRNAs solution was transfered in a cuvette and put in the Nucleofector™ 2b Device (Lonza). The program U-008 was runned. After electroporation, 500 µL of pre-heated CD34+ Cell medium supplemented with stimulating cytokines (SCF,GM-CSF, Tpo and IL-6 ,STEMCELL Technologies) was added and cells was incubated in 24 well plaque overnight at 37°C in a humidified 5% CO2 incubator. Then, RNA isolation for RT-qPCR gene silencing validation, or scRNA-seq protocol was performed.

I.2.a.v. RT-qPCR

RNA was isolated using the RNeasy Micro Kit (Qiagen). RNA was retro transcribed into cDNA using the Invitrogen™ SuperScript™ III Reverse Transcriptase. 1ng of cDNA was mix with 1X SYBR Green master mix (Thermofisher), and 1µL of 10µM reverse / forward primers mix and H2O qsp 20 µL, and incubated in the QuantStudio 7 Pro qPCR system (Thermofischer). Relative normalized expression change was determined using the RPLP0 housekeeping gene expression as reference gene.

I.2.a.vi. Single-cell multimodal libraries preparation

CD34+ cells where incubated 2 hours in StemPro™ CD34+ Cell Medium with or without stimulating cytokines (SCF,GM-CSF, Tpo and IL-6) at 37°C in a humidified 5% CO2 incubator. Then nuclei

where isolated following 10X Genomics protocol of nuclei isolation for single-cell multiome ATAC + Gene expression sequencing[558]. Then, single-cell multimodal analysis was performed using Single Cell Multiome ATAC + Gene Expression Sequencing kit following manufacturer recommendations[559]. After library preparation, the Gene expression and ATAC libraries was sequenced separately following 10X Genomics recommandations. Feature-barcode matrices (of gene expression and peaks) were generated using the CellRanger ARC pipeline(10X Genomics).

I.2.a.vii. Data Processing and Statistical Analysis

Gene expression and peak counts matrices were filtered for low quality cells, and normalized following Seurat and Signac framework [560,561] as described in published method. Cells were annotated for hematopoietic lineage based on the hematomap by using the TransferLabel method of Seurat.

For rapidly processed samples, lineage specific pseudobulk differential expression analysis was performed using DESeq2[562].For siRNAs and single-cell multiome datasets, lineage specific differential expression analysis was performed using the wilcoxon test on the SCTransform normalized matrices. Functional enrichment analysis was performed using ClusterProfiler Package[563]. Regulons enrichment analysis in differentially expressed genes was performed using Fisher's exact test (phyper function) or the fgsea package[564] as described in the results. Regulon activity in cells was measured using the AUCell package[565].

I.2.b. Results

I.2.b.i. Validation of the LGA HSC differentiation bias

We observed an HSC shift toward more differentiated cells in LGA suggesting a differentiation bias of LGA HSC. However, it was unclear if this HSC proportion decrease in LGA was linked to a differentiation bias of LGA HSC during the time of cells preparation. Indeed, after thawing, cells preparation takes ~2 hours to prepare according to the sample multiplexing protocol (HTO), and some cues suggest that cells are responding to a stress or a stimulation regarding that the most active regulons in HSCs are related to the immediate early response (ARID5A, EGR1, KLF2, FOSB, and JUN; See Supplemental Table S7 of the article). To answer this question, I used two independent strategies: i) estimate the differentiation bias thanks to RNA velocity analysis, ii) compare the subpopulation shift with rapidly processed cells.

Estimate the differentiation bias thanks to RNA velocity analysis

To further validate the putative differentiation bias in LGA HSCs, we analyzed the dynamics of the transcriptional program in LGA and Control HSPCs thanks to the RNA velocity analysis (scvelo)[557]. We can then estimate if this program leads toward differentiation or stemness conservation, and thus predict the cell fate of each cell. Velocity calculation relies on the ratio of pre-mRNA over mature mRNA

(how a gene are actively transcribed) for each gene in each cell. Using this approach, we were able to estimate the RNAs velocity for 12 684 cells and to infer a transcriptional dynamic (velocity vectors) across all cells (Figure 16A). Interestingly, while the late branches (erythroid, myeloid and lymphoid progenitors) have a transcriptional dynamic toward differentiation, the main transcriptional dynamics of HSC and MPP cells are toward stemness conservation, so from MPP to HSCs, further supporting that the main transcriptional program in our cells are related to the regulation of activation. Indeed, we observed that the genes the most actively expressed (bigger RNA velocity and variance across HSPCs) are genes related to control of activation like FOS, DUSP1, FOSB, ZFP36, HES1, and NFKBIA (Figure 16B). Concordant with that, by looking at regulon's enrichment for the main contributors of this transcriptional dynamics model, we found the EGR1 regulon being the most enriched, while KLF2 and JUN are in the top5 (Figure 16C). We used this transcriptional dynamic to estimate toward which differentiation states the cells are transitioning and compared then LGA and CTRL cells dynamics (see method). We observed that LGA HSCs transition faster toward more differentiated cells compared to CTRL HSCs, while such differences doesn't appear significant in the others lineages (Figure 16D). These results confirms that LGA HSCs have a differentiation bias compared to CTRL HSC, supporting published transcriptional and functional alterations.

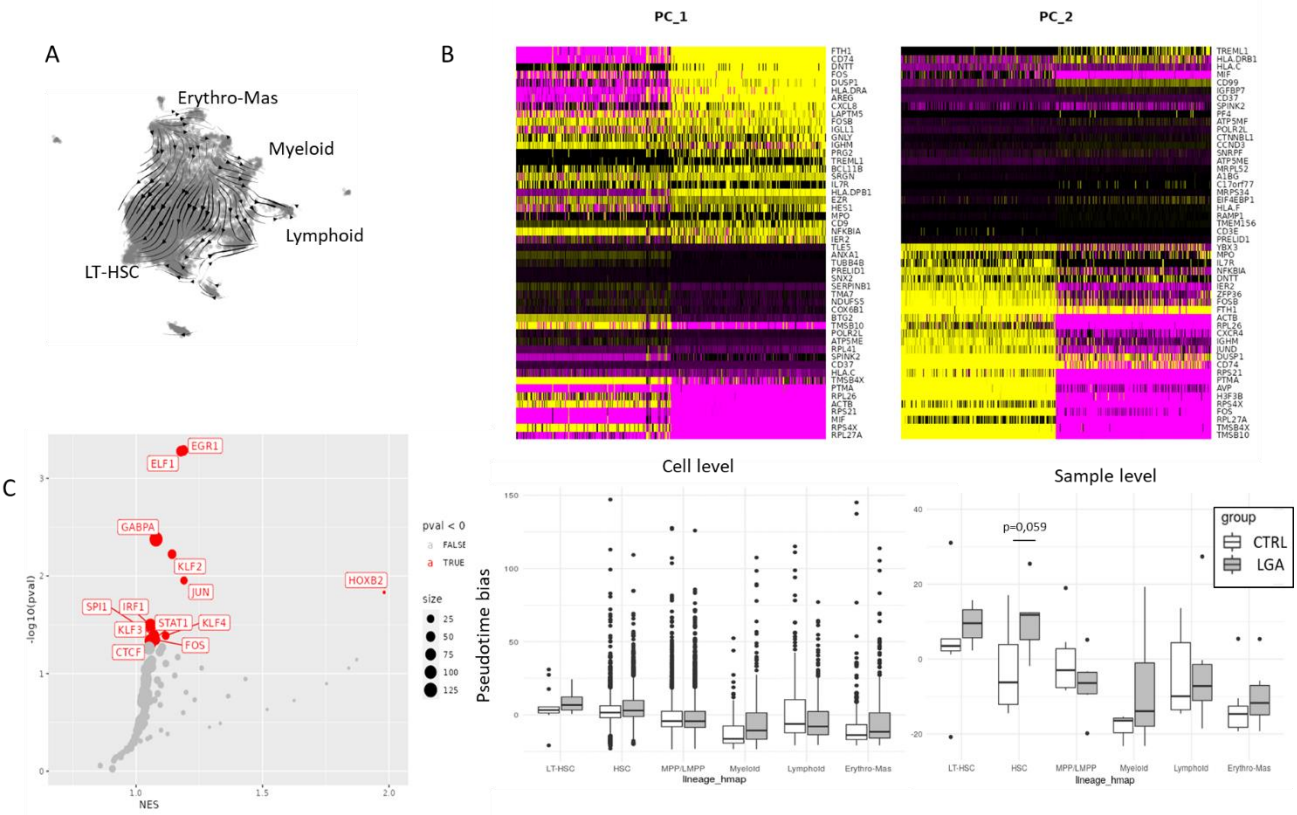


Figure 16: RNA velocity analysis. (A) Velocity stream representing the main transcriptional dynamics within HSPCs. (B) Top genes contributing to the first principal components (PC_1 and PC_2) of the gene velocity variability. (C) Regulons enrichment for main contributors of the transcriptional dynamics (higher change of RNA velocity across cells). (D) Predicted pseudotime shift for LGA and CTRL at cell level (left) and at sample level (right). Statistical difference between LGA and CTRL samples was tested using wilcoxon test, only HSC have a difference close to the statistical signifiance (p-value =0.05).

Compare the subpopulation shift with rapidly processed cells

To tests if cells respond to a stress/stimulation during time of cells preparation, we processed new LGA and CTRL samples without the Antibody based multiplexing (HTO) protocol reducing considerably cell preparation/incubation time (see methods). We compared first if there is a transcriptional difference between both protocols, i.e. if HTO protocol indeed induce a cell response, and compared then the LGA vs CTRL samples in both condition.

To observe if the HTO multiplexing protocol led to a stress/stimulation response, we processed three same CTRL samples with the two protocols (*n*=3 samples representing a total of 6776 rapidly process cells and 1749 HTO processed cells). We identified a strong sample wide gene expression difference between the two protocols at both cell and sample level (Figure 17A-B). We identified 1518

differentially expressed genes comparing HTO versus rapidly processed samples (adjusted *p*-value<0.05 and $|\log_2 FC|>0.5$), including 1075 upregulated in HTO samples enriched for pathways regulating stress response like Foxo, NF-kappa B, and MAPK signaling, but also biological process related to stress or extrinsic stimulation (Figure 17B). These results confirms that the HTO protocol trigger a stress on our cells. Then, we performed a DEG analysis at subpopulation level to know if this response was subpopulation specific. The results of the DEGs analysis at subpopulation level showed that the main cellular response to HTO protocol comes from the HSC subpopulation (Figure 18C). The upregulated genes on HSCs were similar to the one found regarding all HSPCs, confirming that the main transcriptional response to HTO protocol is in HSCs (Figure 17D). Futhermore, by using SCENIC to infer TF regulons activity across each cells and comparing regulon activity between HTO prepared and rapidly process cells, we observed that HTO prepared HSCs have a strong increased activity for FOSB, ARID5A, EGR1, JUN, KLF2, FOS, JUND and KLF4, further supporting the role of these regulons in the stress/stimulation response(Figure 17E). Based on G2/M gene expression signature, we observed also that the HSC and MPP cells prepared with HTO are more committed in the mitosis process (G2/M phase) than non HTO-prepared cells (Figure 17F). These results confirm that the HTO-based protocol trigger a cellular activation, especially in HSC, leading to entry in proliferation and differentiation process. The fact that HSC respond more to the environment compared to other HSPCs fit with the

fact that HSC activate quickly in the bone marrow niche in response to blood loss or other hematopoietics challenge to ensure blood homeostasis or respond to infection[371].

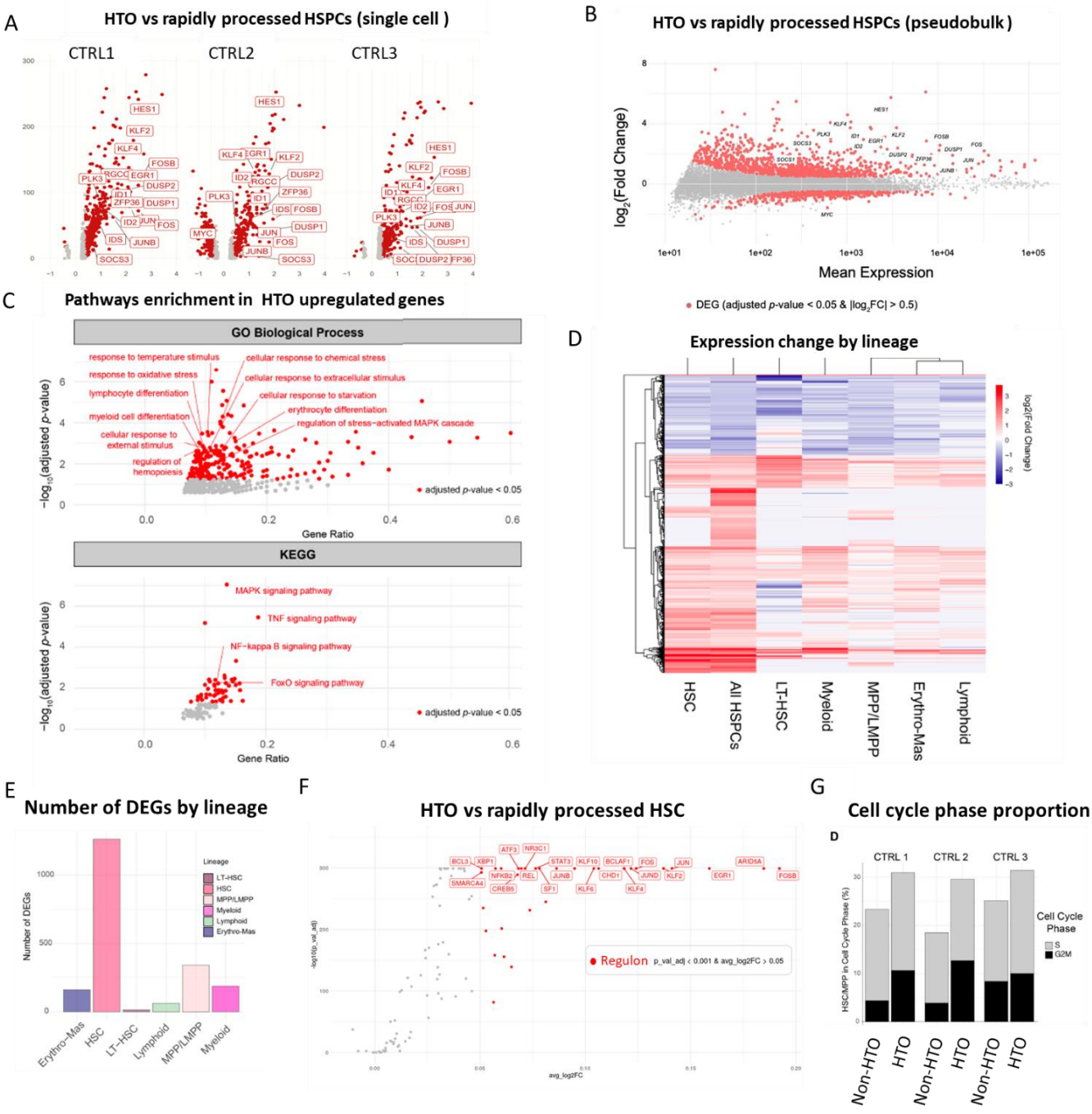


Figure 17 : HTO protocol effect. (A) volcano plot of differentially expressed genes (DEGs) in HTO versus non-HTO prepared cells for each individual. (B) MA plot of DEGs using pseudobulk analysis. (C) Functional enrichment of GO Biological Process and KEGG pathways enriched in upregulated genes in HTO prepared cells (D) Heatmap of the gene expression log2(FoldChange) comparing HTO vs non-HTO prepared samples for each lineage. (E) Number of DEGs by lineage. (F) Differential regulons activity between HTO

vs non-HTO samples (wilcoxon test). (G) Cell cycle phase proportion by sample within HSC/MPP cells

Thanks to these rapidly processed samples, we can study if the LGA specific transcriptional and functional alterations reflect a difference of cellular response to the stimulating/challenging environment during cells preparation. We processed 6 LGA and 7 CTRL samples with this quick cell manipulation condiditions. We annotated cells subpopulation based on the hematopoietics reference map (hematomap) established in the published results. We compared transcriptome within each subpopulation and subpopulation distribution similarly to the HTO processed samples. Based on these samples, we did not observe significant gene expression change across lineage between LGA and CTRL samples (Figure 18A). Concordantly, we did not observed significant subpopulation proportion change between LGA and CTRL (Figure 18B). These results suggest that rapidly processed samples LGA and CTRL samples are relatively similar in term of gene expression and HSPC composition when not challenged by stressful/stimulating environment and thus that both gene expression and subpopulation shift observed in HTO processed samples reflect a cellular intrinsic alteration of the response to stress/stimulation in LGA. These results are concordant with the fact that transcriptional alteration observed in LGA samples target the immediate early response genes, which are expressed specifically in HTO-prepared samples.

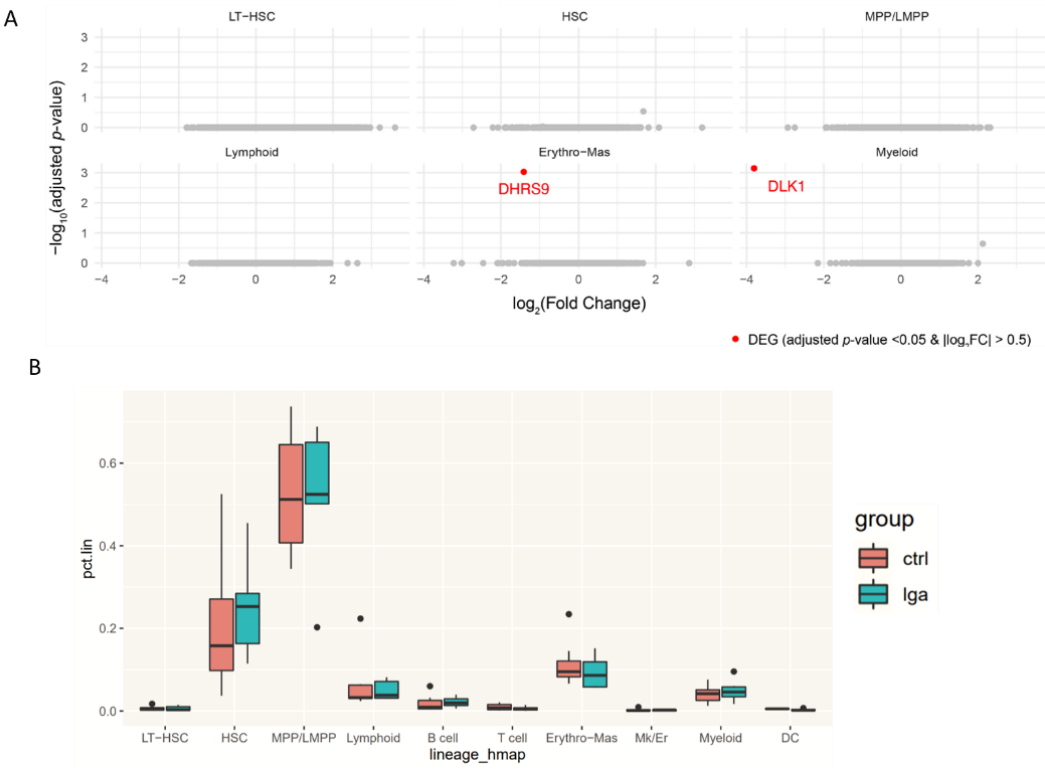


Figure 18 : LGA vs CTRL comparison from rapidly processed samples. (A) volcano plot of differentially expressed genes (DEGs) in LGA versus CTRL (pseudobulk DESeq2 analysis). (B) Subpopulation distribution in LGA and CTRL samples

I.2.b.ii. Validate the EGR1/KLF2/KLF4 regulatory network and its role in HSC activation

Our published results have highlighted a gene regulatory network, govern by EGR1, KLF2 and KLF4 TFs, being epigenetically and transcriptionally altered in LGA. These regulatory networks were inferred based on co-expression analysis of TFs and putative downstream target genes using our scRNA-seq data (SCENIC), and further filtered thanks to scATAC-seq data observing if accessible regions close to a gene have the specific TF motif. However, this method assume that the closest gene of an accessible region is the gene regulated by this accessible region. Regarding that DNA topology/ folding this assumption can be false in certain case because of DNA 3D conformation. Then, to validate the relevance of the inferred regulatory network of KLF2, EGR1 and KLF4, we used another strategy to associate accessible region with gene expression. This strategy leveraged the single cell multimodal assay assessing both chromatin accessibility and gene expression in the same cells. The main advantage of this analysis is that we can directly correlate the chromatin accessibility of an open chromatin region (peak) to the neighbor genes expression and therefore link a peak to a gene more precisely. We compared also stimulated versus unstimulated cells to highlight if this regulatory network is stimulation dependant. To do that we used IL6, CSF, Tpo and Flt3 cytokines known to

activate HSCs and allow their expansions. We recovered 4354 cells from unstimulated and stimulated condition (2001 for unstimulated, 2353 for stimulated), and annotated them thanks to the hematomap (Figure 19A).

Based on lineage subpopulation analysis, we observed a net decrease of HSC in the stimulated conditions compared to unstimulated conditions (Figure 19B). This change is accompagnied by a net increase of MPP/LMPP cells as well as DC progenitors. These results show that cytokines stimulation activate HSC toward differentiation process. By performing lineage specific differential expression analysis comparing stimulated to unstimulated HSC, we observed 581 DEGs, including 239 upregulated and 342 downregulated genes (Figure 19C). Importantly, we observed that the upregulated genes are strongly enriched for regulons of JUN, FOSB, ARID5A, FOS, and JUNB being the top5 most enriched regulons, while EGR1 and KLF2 regulons are also found significantly enriched (Figure 19D). These results support that the cell response observed in HTO prepared cells are similar to a response to physiological cytokines and confirms that the regulons altered in LGA HSC are regulons regulating stimulation response. These new data further support that the cellular response to stimulation is altered in LGA HSCs.

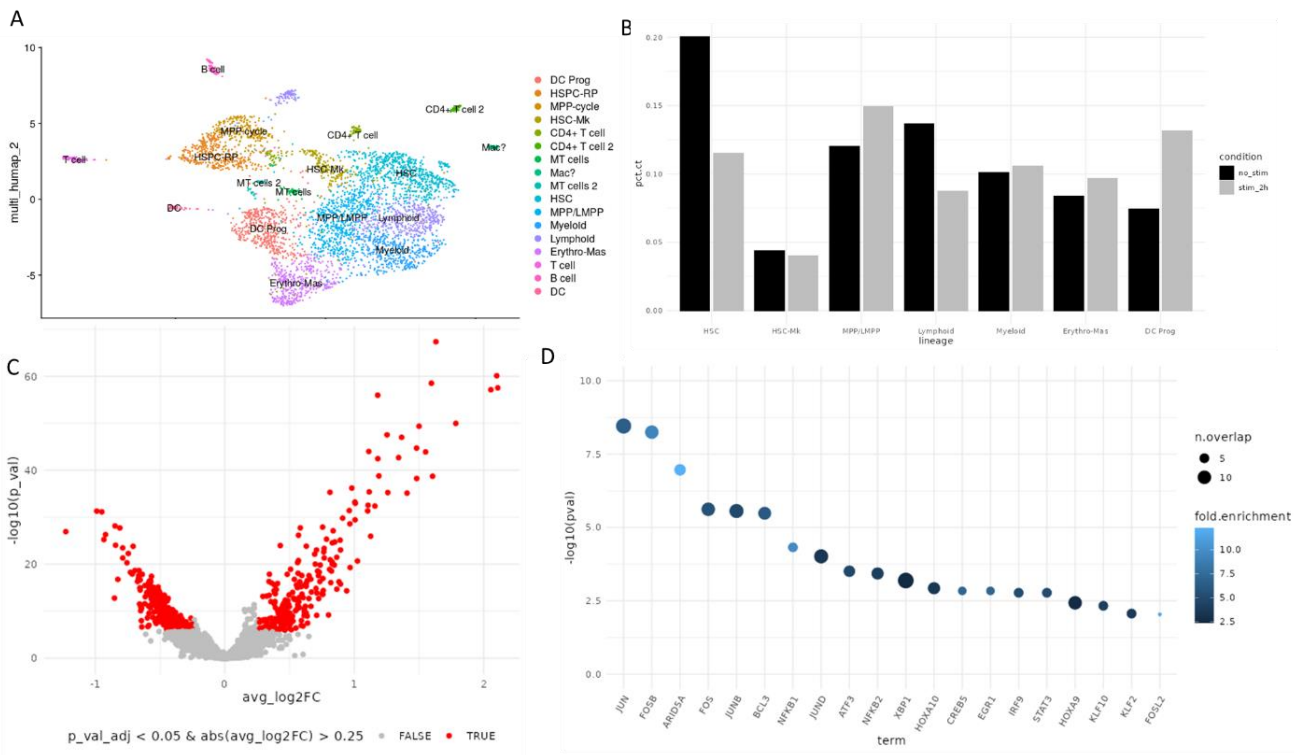


Figure 20: Single-cell multimodal analysis of cytokine stimulation. (A) UMAP representation of cells integrating both variability of gene expression and chromatin accessibility. (B) Subpopulation distribution in stimulated (grey bar) and unstimulated

(black bar) sample. (C) Differential expression analysis of stimulated versus unstimulated HSC (wilcoxon test). Red point represent DEGs (adjusted p-value<0.05 and $|log_2FC| > 0.25$). (D)Regulons enrichment in upregulated genes (Fisher's exact test)

To validate more precisely the gene regulatory network altered in LGA and the impact of methylation, we infer peak-genes links for all the genes taking part of the previously identified regulons (n=2476 genes). Using the correlation between peak accessibility and neighbors' gene expression, we identified 4281 peak-gene links across 1487 genes representing 60% of the interrogating genes (Figure 20A). These gene expression associated peaks are candidate cis-regulatory elements (CREs; e.g. enhancer) able to regulate associated gene expression. Considering genes associated to EGR1, KLF2 and KLF4 regulons (n=123 genes), we identified candidate CREs for 78 genes (63% of the genes). Interestingly, we found ZFP36L2 associated to a lot of candidat CREs (26) suggesting tight regulation (Figure 20B). This gene is an RNA-binding protein which is known to promote cell quiescence and to regulate erythroid differentiation[566]. Interestingly, it has also a role in promoting mRNA decay of immediate early genes (IEGs)[567], highlighting its role in regulating response to stimulation. Then, we focused on CREs containing EGR1, KLF2, or KLF4 TF motif. We observed that 46% of genes in EGR1 regulon have an EGR1 motif on a CREs (p-value < 0.05, over-representation test), while 45% for KLF2 (p-value < 0.01, over-representation test), and only 38% for KLF4 (non-significant, over representation test) (figure 20C). Even if this analysis was not able to validate every inferred TF-gene regulations, it shows that the EGR1 and KLF2 regulons inferred with previous methods are enriched for candidate CREs containing the corresponding TF motif, validating partially TF influence on these genes. However, these results were not able to validate the influence of KLF4 on inferred regulon. Together, these results highlight the interest of using single-cell multimodal data to find, or validate, TF-gene regulatory interaction based on correlation between chromatin accessibility and gene expression.

Then, to validate the putative influence of LGA associated methylation on gene expression, we integrate this newly identified regulatory information with methylation data. Overall, 2% of the ~750k queried CpGs fall in CREs, while 6% of the 4815 DMCs, showing 2.5 fold enrichment for DMCs in candidate regulatory elements (p-value<0.0001, over-representation test), further supporting putative DMCs impact on gene expression. Critically, these DMCs associated CREs are strongly enriched for EGR1/KLF2/KLF4 TF motifs with 91% of these CREs (198/216) having at least one of this TF motif (p-value < 0.0001, over-representation test; Figure 20D). These results further confirm that the DNA hyper-methylation in LGA target the EGR1/KLF2/KLF4 gene regulatory network.

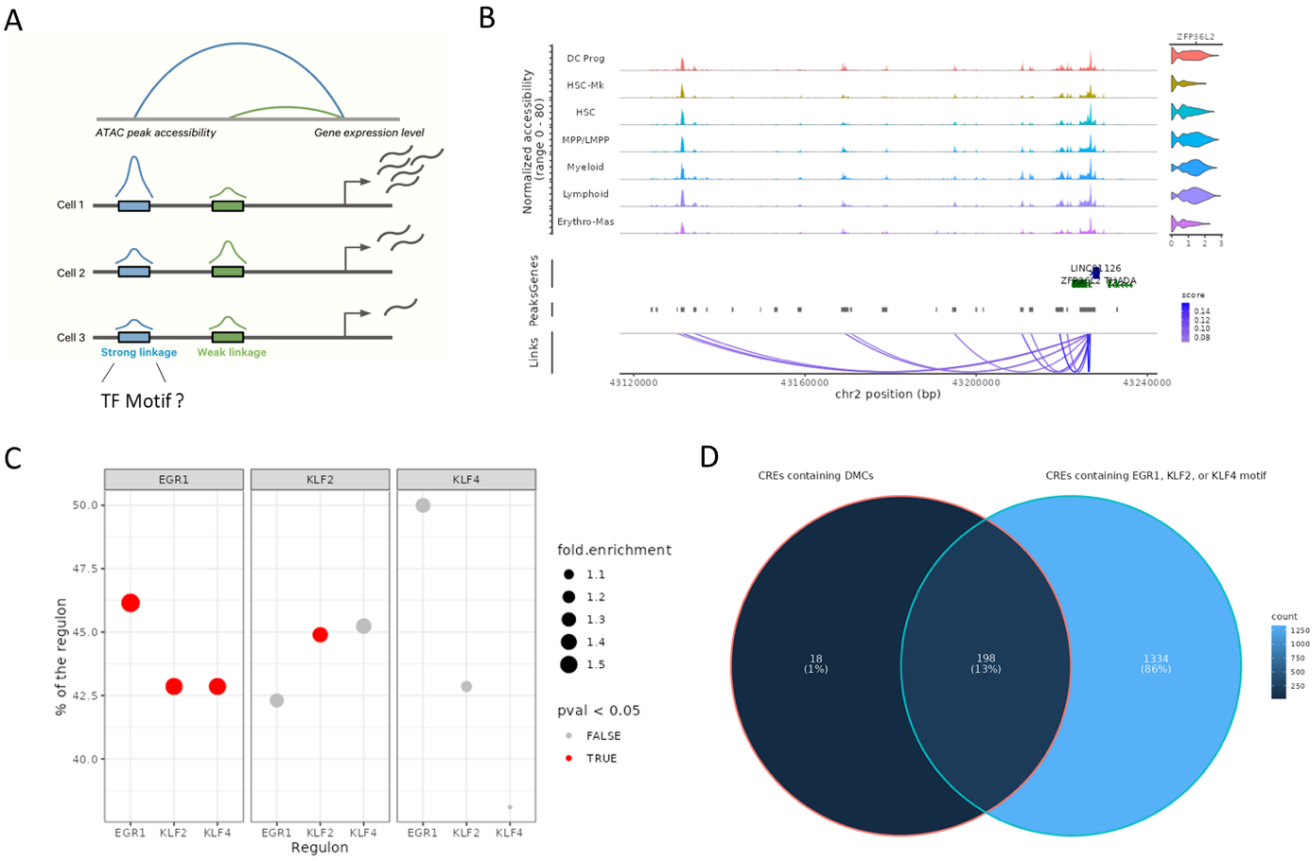


Figure 20: Validation of the EGR1-KLF2-KLF4 regulatory network epigenetics alteration. (A)Schema of peak gene linkage method. Reprinted from 10X genomics website support[R] . (B) Peak gene linkage in the ZFP36L2 locus. The score of the linkage correspond to the correlation between the peak accessibilty and the gene expression. (C) Over-representation of EGR1, KLF2 and KLF4 regulons (x axis) in CREs containing EGR1, KLF2 or KLF4 motif (boxes). P-value of the enrichment are calculated using Fisher's exact test. Red point represent p-value < 0.05. (D) Venn diagram of the intersection between CREs containing DMCs (left) and CREs containing EGR1, KLF2 or KLF4 motif.

Gene silencing-based validation

We found a correlation between hypermethylation, chromatin rearrangement, and decrease expression of EGR1/KLF2/KLF4 related transcriptional network associated with a reduce HSC ability to stay quiescent in response to challenging environment in LGA samples. To assess the causality between this network downregulation and the differentiation bias in LGA samples; we performed gene silencing experiments, using siRNAs targeting KLF2. We chose to target KLF2 because being the most upstream

[R] https://support.10xgenomics.com/single-cell-multiome-atac-gex/software/pipelines/latest/algorithms/feature-linkage

regulators based on co-regulatory network analysis results (SCENIC). This gene silencing experiment have also the potential to further validate the KLF2 influence on the downstream target genes (regulon) infer thanks to both scRNA-seq and scATAC-seq data. To perform this experiment, we used a fresh sample of CD34+ cells and transfect cells either with siRNAs targeting KLF2 transcripts (siKLF2 condition) or with negative control (siCTRL condition) and let the cells over-night in incubation on CD34+ cells optimized medium with stimulating cytokines (see Method). We first confirmed that siKLF2 indeed reduce expression of KLF2 through RT-qPCR (Figure 21A). We then performed scRNA-seq and recover 1533 and 2116 good quality cells from siCTRL and siKLF2 respectively. We annotated our cells based on the hematomap (Figure 21B-C) and found that the overnight incubation with media supplemented with cytokines lead to increase differentiated cells, because only 4% of cells were annotated as HSC compare to the 24% in our previous unincubated samples (Figure 21D). To observe if KLF2 silencing led to differentiation bias as observed in LGA samples, we looked at subpopulation distribution between siKLF2 and siCTRL samples. We observed a slight reduction of HSC and MPP cells in siKLF2 compared to siCTRL (Figure 21E). Even if the difference appears significiant using a chisquared test under the assumption that siCTRL distribution is the ground truth, our sample size is very limited. This experiment should be reproduce at least 3 times to further validate the statistical signifiance.

To validate the role of KLF2 regulating the downstream genes identified with previous analysis, we performed differential expression analysis on siKLF2 compared to siCTRL HSCs focusing on genes from the KLF2 regulon (n= 89). We decided to focus the test on HSC cells specifically because we have previously shown that KLF2 alterations mainly target HSCs. We further confirm that KLF2 is specifically active in HSC, by measuring the KLF2 regulons activity score using AUCell algorithm (Figure 21F). We observed only few DEGs between siKLF2 and siCTRL HSC passing adjusted p-value threshold 0.05 (n=4), but 15/89 (16%) genes at nominal p-value (Figure 21G). This relative weak result can be explained by the relative low number of HSCs and by the transfection efficiency (transfection efficiency was estimated being around 50% based on preliminary analysis).

To go beyond this relative finding, we performed an unsupervised differential expression analysis and gene set enrichment analysis (GSEA) to see if these transcriptional alterations were specific to the KLF2 regulon. We observed 4 regulons significantly enriched in downregulated genes (adjusted p-value <0.1), with KLF2 regulon being the second most enriched regulon (Figure 21H). Interestingly, the first most enriched regulon was STAT3, suggesting close link between KLF2 and STAT3 activity. These results confirm previous inferred data using both co-expression and chromatin accessibility and highlight putative regulatory role of KLF2 on STAT3 signaling.

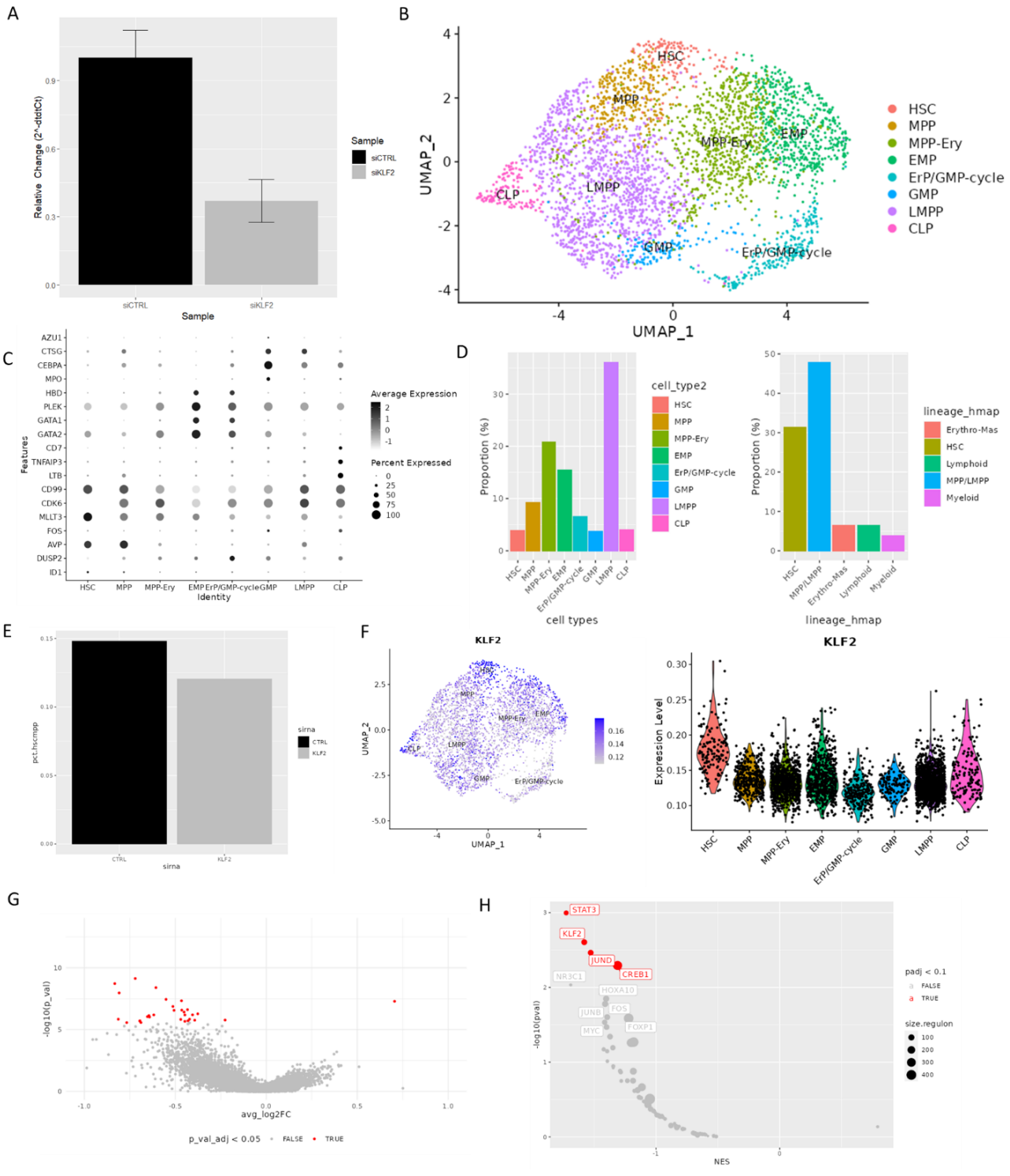


Figure 21: Gene silencing of KLF2. (A)Validation of the KLF2 gene expression silencing by RT-qPCR. (B) UMAP representation of the scRNA-seq dataset. Cell were annotated based on the hematomap reference and lineage specific markers expression. (C) Lineage specific markers expression. (D) Distribution of cell type in this experiment (left) compared to distribution from previous uncultured datasets(right). (E) proportion of

HSC and MPP cells in sample transfect with siRNAs targeting KLF2 (siKLF2) compared to sample exposed to siRNAs Control (siCTRL). (F) Activity score (AUCell) of KLF2 regulon in cells. (G) Differential expression analysis of HSCs exposed to siKLF2 compared to HSC exposed to siCTRL (wilcoxon test). Red point represent DEGs (adjusted p-value <0.05 and |log2FC|>0.25). (H) Regulons enrichment analysis using GSEA in differential expression results. NES: Normalized Enrichment Score. NES<0 indicate enrichment in downregulated genes in siKLF2 HSCs. Red point represent regulons significantly enriched (adjusted p-value<0.1)

I.3. Conclusion

I.3.a. Characterize the epigenetic memory of LGA HSPCs

Delahaye *et al* have previously shown that LGA HSPCs have a global DNA hypermethylation enriched for candidate cis regulatory elements and close to genes known to regulate metabolic function and stem cells properties[349]. Here we validated this DNA methylation alteration using an independent cohort and integrate all data to gain in discovery power.

One main challenge of DNA methylation analysis is to link with putative functional alteration. Indeed, functional enrichment analysis is performed at gene level and not at CpGs level. CpGs are heterogeneously distributed across the genome, which introduce a bias when linking to gene. Some tools have been developed trying to overcome this issue correcting for DNA methylation data specific bias[568–570]. Recently, Maksimovic *et al* have developed GOmeth and GOregion to better take into account methylation data specific bias allowing to perform functional enrichment analysis weighting each CpGs methylation differences according to the gene specific context[570].

Here, we weighted each CpG methylation change with tissue specific eQTL data, and histones profiles, as well as global regulatory annotation while correcting for CpG bias to obtain a score by gene reflecting the probability that the methylation change affect expression of them. We showed that this score better predict gene expression change than classical methods highlighting the predictive power of this method. Thanks to this approach, we predicted the DNA methylation change affecting genes regulating fetal growth and organ development, as well as stem cells pathway including notably Wnt signaling pathway and genes regulating cell fate commitment. Wnt signaling is a key pathway regulating stem cells differentiation for organogenesis as well as tissue regeneration[571–573]. DNA methylation of these pathways has been found to alter stem cells differentiation capacity in adipocytes

derived stem cells[574,575]. Then, these results further confirm putative impact of DNA methylation alteration in LGA HSPCs on regulation of their differentiation.

Even if different epigenetics layer (methylation, histone marks, chromatin conformation) co-exist in cells, the interactions between them are not always clear. Recently, a clear link between DNA methylation and chromatin accessibility have been found in Arabidopsis model, where Zhong *et al* showed that ablation of DNA methylation durably affect chromatin accessibility[576].

Here, integrating bulk DNA methylation with scATAC-seq data, we observed in our LGA model that the increased DNA methylation observed is associated with a decrease chromatin accessibility, showing in human cells this link between DNA methylation and chromatin accessibility. We found also that the most enriched TF motifs in DMCs regions are also those found in decrease accessibility regions further supporting the interplay between DNA methylation and chromatin accessibility. The TFs motifs enriched in epigenetically altered regions include TFs of the SP family like SP2 and SP3 encoding for zinc finger proteins having key role in Wnt signaling mediated embryogenesis[572]. They also include the zinc finger protein EGR1, which regulate hematopoetic stem cells proliferation and response to extrinsic stimuli[577,578] , as well as the Kruppel family TFs KLF2 and KLF4, which are key role in stem cells self renewal and regulation of differentiation[579–581]. Interestingly, within HSPC, LGA epigenetics alterations seems to affect specifically HSC. Indeed, we found that both DMCs and chromatin rearrangement are enriched in HSC specific open chromatin regions. These results highlight the interest to study epigenetics layers at single-cell level rather than bulk and suggest transcriptional consequences on HSC specifically. Together, we found in LGA a coordinated increase DNA methylation and decrease chromatin accessibility in candidat regulatory regions of TFs known to regulate proliferation and differentiation of stem cells suggesting impact on expression of these genes and regulation of differentiation in these cells.

I.3.b. Decipher the impact on gene expression

Single-cell transcriptomic data can both highlight transcriptomically distinct subpopulation and gene expression change between two conditions within the subpopulations. Several previous scRNA-seq studies have decipher the different subpopulations present in cord blood HSPC highlighting the transcriptomics and functional heterogeneity within phenotypically defined population[122,582–584]. Here we identified 7 distinct subpopulations in HSPCs recapitulating the different lineage found in the cord blood hematopoietic compartment, from LT-HSC to restricted erythroid, myeloid and lymphoid progenitors. Notably, we identified like a recent study on developmental hematopoiesis[585] the oligo-potential erythro-myeloid progenitors (EMP) connecting HSC/MPPs to erythroid, megakaryocytes, and

mast cells. This oligo-potential progenitor appears under the governance of GATA2 and TAL1 TF based on markers expression and regulons analysis.

To decipher LGA specific transcriptomic change in these subpopulations, we performed pseudobulk differential expression analysis allowing to find robust gene expression difference in LGA compared to CTRL samples. We observed that the main expression changes were in HSC, with a significant number of genes (285) being downregulated. These downregulated genes are enriched for genes negatively regulating cell growth signaling and proliferation, and response to stress or stimulation, suggesting decrease ability of LGA HSC to control differentiation and proliferation in response to stimulation. By integrating these results with the epigenomics change discuss in previous section, we observed that these downregulated genes were enriched for hyper-methylated genomics regions as well as regions with decrease accessibility highlighting the putative epigenetics influence on gene expression. Together, these results suggest that the epigenetics memory of fetal overgrowth impact a transcriptional program controlling differentiation and proliferation of HSC and response to stress/stimulation. Such mechanism of epigenetics memory was recently found on hair follicle bulge stem cells where previous exposure to wound damages was associated with durable chromatin accessibility change influencing stem cell's responses to future assaults[379].

Izzo *et al* recently demonstrated influences of DNA methylation on TF activity in the context of hematopoietic differentiation[393]. Researchers found TFs binding CpG rich DNA motifs, notably key erythroid TF Klf1 and Tal1, were affected by disruption of DNA methylation and driving cell fate change. Here, combining co-regulatory network analysis of scRNA-seq data (using SCENIC tool), and TF motif accessibility based on scATAC-seq data, we identified EGR1, KLF2 and KLF4 are putative TF upstream regulator affected by the epigenetics change. We confirmed that the regulons of these TFs were enriched for hypermethylated and decrease accessibility genomics regions, supporting the role of the epigenetics change on TF activity and targeted downstream genes. Furthermore, we found that these TFs have direct and indirects interactions. Indirect, because having co-downstream targets genes, and direct because regulating each other's. These interactions suggesting a common transcriptional program. Indeed, as exposed before, these TFs are all known to negatively regulate proliferation or differentiation, thus promoting HSC quiescence[577,586–589] . These evidence were confirmed in our data because their downstream genes are enriched for markers of HSC quiescence as well as for biological process regulating differentiation, proliferation, and response to stimulation.

Using single-cell multimodal (ATAC+ gene expression) data, we further confirm putative influence of EGR1 and KLF2, but not KLF4 on these downstream genes as well as their role in regulating HSC stimulation/activation. We further test the relevance of this regulatory network using gene silencing

experiments targeting KLF2. We found that the KLF2 knock-down was able to reduce expression of genes associated to the KLF2 regulon, but also to reduce HSC/MPP proportion further validating the KLF2 regulatory network and role in regulating differentiation of HSC. Taken together, we observed that HSCs exposed to fetal overgrowth have specific epigenetics alteration associated with a transcriptional downregulation of genes under the control of TFs, mostly EGR1 and KLF2, known to regulate differentiation and proliferation of HSCs. These results further suggest epigenetics programming of transcriptional activity in LGA, and consequences on HSC cell fates.

Interestingly, EGR1 was recently found as a key actor in shaping the brain DNA methylome by recruiting the DNA demethylase TET1 to regulate activation of downstream genes in response to life experience[590]. In our study, we observed that in LGA, EGR1 appears downregulated, with a decrease activity on downstream genes, while its putative binding regions are enriched for DNA hypermethylation. In light of its role in regulating DNA methylation, these results suggest then a direct link between EGR1 activity and the remodeling of DNA methylation observed in LGA.

Several evidence in our data show that the LGA response to stimulation/stresses is altered. The genes or regulons differentially expressed in LGA are strongly enriched for genes of the immediate response to external stimuli / stress. Notably, EGR1, SOCS3, JUNB, JUN, FOSB, DUSP2, IER2, and IER5 which are downregulated in LGA compared to control HSC, and known to regulate response to stress or stimulation taking part or the immediate early response[578]. The immediate early response genes (IEGs) are genes able to rapidly be transcribed, within minutes, following a stimulation. Our scRNA library preparation protocol require relatively long cells manipulation and incubation time on challenging environments (cold temperature, centrifugation...), due to sample multiplexing strategy (HTO) and are thus likely to trigger a cellular response through gene expression. These manipulation dependent gene expression was demonstrated in a previous study, aspecially for cryopreserved cells, as this is the case for our cells[591]. By processing samples with reduces cell preparation time, we confirmed that the HTO based cell preparation protocol triggers a cell response to stress / stimulation. To validate the physiologically relevance of the HTO stimulation, we characterized the HSPC response to physiological cytokines, notably IL-6, known to activate HSC through the JAK/STAT pathway[592–594], and found that cytokine stimulation activates similar regulons (JUN, FOSB, EGR1, STAT3...) that the the one with the HTO protocol supporting that the cellular response is physiological. Together, we showed that LGA functional changes are likely to represent an epigenetics programming of their response to stimulation/growth signaling. This observation fit with the concept that early exposure can have durable consequences on the ability of cells to respond to future environment as observed previously in others tissues like pancreas, adipose tissue, and muscle in the context of aging[314,316,317,595,596].

I.3.a. Assess the impact on HSPCs plasticity

HSC are mainly quiescent and located in the bone marrow niche. The exit of quiescence toward proliferation and differentiation is tightly regulated by HSC molecular response to cytokines and cell-cell interactions[597–600]. Several studies have found that this balance between HSC quiescence and differentiation is altered in aging[400]. Recently, Sureshchandra *et al* have shown in primate that HSC and progenitors ability can be altered as early as *in utero*. Indeed, they showed that maternal high fat western style diet was able to alter HSPC expansion and repopulation ability while reducing lymphoid potential[601].

Here we found an alteration of the control of the balance between quiescence and differentiation in HSCs of neonates exposed to excessive fetal growth. We found that LGA was associated with a decrease proportion of HSC while an increase of more differentiated cells (MPPs), as well as a decreased number of HSC/MPP derived colonies in LGA compared to CTRL HSPCs. These results support a bias toward differentiation in LGA HSPCs concordant with the epigenetics and transcriptomics alterations observed. RNA velocity analysis confirms this hypothesis showing that LGA HSCs have a bigger probability to differentiate compared to CTRL. Furthermore, rapidly processed LGA samples do not show significant differences with CTRL samples supporting that the LGA functional differences observed in HTO processed samples are an intrinsic alteration of the response to challenging/stimulating environment rather than a basal change.

Altogether, these results highlight that LGA specific epigenetics and transcriptional alterations of genes network promoting HSC quiescence, is associated with an HSC differentiation bias within HSPCs compartment, suggesting an epigenetics programming of the control of HSC differentiation in LGA. Bokeska *et al* have recently demonstrated such epigenetics programming on HSC. Indeed they showed that HSC early exposed to inflammatory challenges have durable epigenetics alterations and a decrease *in vivo* self renewal ability[602]. They also showed that these environmental challenges accelerate the cellular and molecular aging of HSCs with lifelong defect on tissue maintenance and regeneration. Our results highlight also a putative long-term effect of fetal overgrowth on the hematopoietic system and related chronic disorders. These results also corroborate with the findings of Sureshchandra *et al* in primates showing the role of early exposure to maternal high fat western style in programming hematopoiesis and related inflammatory status, with impair fetal bone marrow development and HSPC functions driving an hyperinflammatory phenotypes[601] . These evidence in light of our own results further support the key role of early detrimental exposure in future HSPCs dysfunctions with putative consequences on ACDs susceptibilities.

II. BIN1 AD GENETICS RISK STUDY

Alzeihmer's disease (AD), responsible of 70% of dementia and affecting more than 20% of elderly people, is the 7th leading cause of death worldwide. While several genetics, environmental and age related mechanisms have been identified contributing to AD, lot of unresolved questions regarding the cause of its development remain, so that no efficient treatments is yet available to prevent or cure this disease. AD is estimated having a 70% heritability suggesting strong genetics influences. Variant E4 of the APOE gene was found as the main genetics risk factor of AD, but not explain every diseases onset, as 1 third of AD patient do not have this variant. Then, it is critical to understand others genetics influences as well as the role of these gene on AD development. BIN1 locus is the 2nd most associated genetics risk but its role in AD development remain poorly understood.

Here we studied the role of BIN1 by deleting it in human iPSC derived cerebral organoid and neuronal culture models, leveraging scRNA-seq to identified in an unsupervised way the main influences on brain cell types. By conducting differential expression analysis between BIN1 knock out (KO) and wild type (WT) cells within each cell types identified we found that the transcriptional alterations in BIN1 KO mainly alter glutamatergic neurons and highlight pathways related to electrical activity, calcium related pathway, and synaptic transmission in both bi-dimensional and tridimensional cellular models. Comparing to scRNA-seq clinical data, we found similar transcriptional change in glutamatergic neurons of AD brains, suggesting similar functional alteration in AD. We then validated this functional alteration of neuronal electrical activity in 4-6 weeks bidimensional cultures using multi-electrode arrays, showing reduced frequency of spikes by burst while an increase amplitude, indicating neuronal hyperexcitability, but also a temporal disorganization of spikes in the neural networks, suggesting impaired capacity of BIN1 deleted cells to generate organized patterns of electrical activity. Leveraging single-cell transcriptomics signature in cerebral organoid, we found that proportion of neurons expressing genes signature of sustained electrical activity was significantly increased in BIN1 KO glutamatergic neurons from cerebral organoid, suggesting durable functional alterations of this neuronal hyperexcitability. Finally, we found that BIN1 was able to regulate the L-type voltage-gated calcium channel (LVGCC) Cav1.2, through direct interaction, and pharmacological inhibition of Cav1.2 partially rescued BIN1 KO mediated spikes desynchronization. LVGCC are known to regulate synchronous firing and its internalization prevent neurons hyperexcitability, bringing a mechanistic explanation to the role of BIN1 in regulating calcium signaling related neuronal hyperexcitability and neural network synchronization.

II.1. Published results

Orthis Saha*, Ana Raquel Melo de Farias*, Alexandre Pelletier*, Dolores Siedlecki-Wullich, Johanna Gadaut, Bruna Soares Landeira, Arnaud Carrier, Anaïs-Camille Vreulx, Karine Guyot, Amelie Bonnefond, Philippe Amouyel, Cláudio Marcos Queiroz, Devrim Kilinc, Fabien Delahaye, Jean-Charles Lambert, Marcos R. Costa. **Alzheimer's disease risk gene BIN1 modulates neural network activity through the regulation of L-type calcium channel expression in human induced neurons**. *BiorXiv, 2022.* https://doi.org/10.1101/2022.01.18.476601

Alzheimer's disease risk gene BIN1 modulates neural network activity through the regulation of L-type calcium channel expression in human induced neurons

Orthis Saha[1,#], Ana Raquel Melo de Farias[1,2,#], Alexandre Pelletier[3,#], Dolores Siedlecki-Wullich[1], Johanna Gadaut[1], Bruna Soares Landeira[1], Arnaud Carrier[3], Anaïs-Camille Vreulx[1], Karine Guyot[1], Amelie Bonnefond[3], Philippe Amouyel[1], Cláudio Marcos Queiroz[2], Devrim Kilinc[1], Fabien Delahaye[3], Jean-Charles Lambert[1] and Marcos R. Costa[1,2,*]

1. Univ. Lille, Inserm, CHU Lille, Institut Pasteur de Lille, U1167-RID-AGE facteurs de risque et déterminants moléculaires des maladies liées au vieillissement, DISTALZ, Lille, France
2. Brain Institute, Federal University of Rio Grande do Norte, Natal, Brazil
3. Univ. Lille, Inserm, CNRS, CHU Lille, Institut Pasteur de Lille, U1283-UMR 8199 EGID, Lille, France

# Equal contribution

* Correspondence should be addressed to:

Marcos Costa, MD, PhD
INSERM UMR1167
Institut Pasteur de Lille
1 rue du Pr. Calmette
59019 Lille CEDEX, France
Tel: 00 33 (0)3 20 87 77 10
marcos.costa@pasteur-lille.fr

Abstract

Bridging Integrator 1 (*BIN1*) is the second most important Alzheimer's disease (AD) risk gene after *APOE*, but its physiological roles and contribution to brain pathology are largely elusive. In this work, we tackled the short- and long-term effects of BIN1 deletion in human induced neurons (hiNs) grown in bi-dimensional cultures and in cerebral organoids. We show that *BIN1* loss-of-function leads to specific transcriptional alterations in glutamatergic neurons involving mainly genes associated with calcium homeostasis, ion transport and synapse function. We also show that BIN1 regulates calcium transients and neuronal electrical activity through interaction with the L-type voltage-gated calcium channel $Cav_{1.2}$ and regulation of activity-dependent internalization of this channel. Treatment with the $Cav_{1.2}$ antagonist nifedipine partly rescues neuronal electrical alterations in *BIN1* knockout hiNs. Together, our results indicate that BIN1 misexpression impairs calcium homeostasis in glutamatergic neurons, potentially contributing to the transcriptional changes and neural network dysfunctions observed in AD.

Introduction

The Bridging Integrator 1 (*BIN1*) is the second most associated genetic determinant with the risk of late-onset Alzheimer's disease (LOAD), after the Apolipoprotein E (*APOE*) gene[1–4], and it is only since the report of its association with AD more than ten years ago that its role in brain functions started to be investigated. In the adult human brain, *BIN1* is mainly expressed by oligodendrocytes, microglial cells, glutamatergic and GABAergic neurons[5–7] and its expression is reduced in the brains of AD patients compared to healthy individuals[7–9]. How this reduced expression of *BIN1* may affect AD pathogenesis remains poorly understood.

Changes in *BIN1* expression have been controversially associated with amyloid precursor protein (APP) processing towards the production of amyloid-beta (Aβ) peptides in cellular models[10,11]. However, we recently showed that BIN1 regulates endocytic trafficking in hiPSC-derived neurons (hiNs), without significantly affecting amyloidogenic APP processing[12] and BIN1 underexpression does not modify amyloid pathology in an AD-like mouse model[13]. A direct interaction between TAU and BIN1 has also been reported[14,15] potentially impacting learning and memory in a Tauopathy mouse model[16], Tau phosphorylation and propagation in vitro[16–19] or network hyperexcitability in rat hippocampal neurons[19].

Despite these advances, no consensus has been reached on the role of BIN1 in AD pathogenesis and even its physiological functions in human brain cells remain mostly unknown. Therefore, rather than developing an Ab/Tau-based hypothesis as in most previous reports, we decided to first develop an agnostic approach to capture a BIN1-dependent molecular landscape in cerebral organoids and neural cells derived from hiPSC underexpressing this gene.

Results

Transcriptional alterations in *BIN1* KO hiNs highlight pathways related to electrical activity and synaptic transmission

To unbiasedly study possible changes in gene expression in human neural cells in function of BIN1 expression, we generated *BIN1* wild-type (WT), and knockout (KO) cerebral organoids (COs)[20,21]. After 6.5 months of culture, COs were composed of all the major neural cell types identified by the expression of MAP2, GFAP and NESTIN, and we did not observe any gross differences in size or morphology of COs between genotypes (Fig. 1A). Western blot analyses confirmed the absence of BIN1 protein in *BIN1* KO COs (Fig. 1B). Using snRNA-seq, we recovered the transcriptional profile of 4398 nuclei that were grouped into 7 major cell clusters based on the expression of cell type markers (Fig. 1C-D). As observed in the human brain[7], BIN1 expression in COs was mainly detected in oligodendrocytes and glutamatergic neurons (Fig. 1D). Notably, we observed a significant reduction in the proportion of glutamatergic neurons in *BIN1* KO compared to WT COs (Fig. 1E), suggesting their selective loss or reduced differentiation. Using Wilcoxon test after sctransform normalization and variance stabilization of molecular count data[22], we detected 124, 75, 4 and 1 differently expressed genes (DEGs; |log2FC| >0.25 and FDR <0.05) respectively in glutamatergic neurons, astrocytes, NPCs and oligodendrocytes, when comparing gene expression in single cell populations of *BIN1* KO and WT COs (Fig. 1F; Sup. Table 1). Gene ontology (GO) term enrichment analysis for DEGs identified in *BIN1* KO glutamatergic neurons revealed a significant enrichment for several terms associated with synaptic transmission, calcium binding and ion channels (Fig. 1G; Sup. Table 2). In *BIN1* KO astrocytes, we found enrichment for GO terms associated with neuronal differentiation (Sup. Table 2). In addition, since BIN1 is expressed at very low level in WT COs astrocytes and we noticed several DEGs regulated by neuronal activity, such as *APBA1*, *GRIN2B*, *NPAS3* and *RORA* (Sup. Table 1)[23], changes in astrocytes are likely secondary to neuronal modifications/dysfunctions. Accordingly, we observed 65 DEGs in glutamatergic neurons of *BIN1* heterozygous (HET) compared to WT COs but only 6 DEGs in astrocytes (Sup. Fig. 1). Similar transcriptional alterations were observed in *BIN1* KO hiNs generated in bi-dimensional cultures (Sup. Fig. 2).

We next aimed at evaluating the cell-autonomous effect of *BIN1* deletion in glutamatergic neurons and, for this purpose, we generated *BIN1* WT or KO pure glutamatergic neuronal cultures by direct lineage-reprogramming of human NPCs (hNPCs) using doxycycline-inducible expression of ASCL1 (see online methods). After validation that ASCL1 expression efficiently reprogrammed hNPCs into highly pure neurons (hereafter ASCL1-hiNs; Fig. 1H), we added exogenous human cerebral cortex astrocytes to support functional neuronal maturation and synaptic connectivity[24]. After 4 weeks of differentiation and snRNA-seq analyses (n=3114 from 2 independent culture batches), we observed that ASCL1-hiNs (~70% of all the cells; see Online Methods for a full description of the cellular populations) were composed of glutamatergic neurons (~92%) with a small proportion of GABAergic neurons (~2%) or of cells co-expressing low levels of markers of both neuronal subtypes (~6%). We detected 675 DEGs (|log2FC| >0.25 and FDR <0.05) in *BIN1* KO compared to WT glutamatergic neurons, and only 1 DEG in GABAergic neurons (Fig. 1K; Sup. Table 3). As observed in COs (Fig. 1G) and spontaneously differentiated hiNs (Sup. Fig. 2), GO term enrichment analysis revealed a significant enrichment for terms associated with synaptic transmission, ion channel activity and calcium signaling pathways (Fig. 1L; Sup. Table 4). Noteworthy, exogenously added human astrocytes co-cultured with *BIN1* WT and KO hiNs showed a low number of DEGs (25 in Astro-I and 18 in Astro-II; Sup. Table 3), likely again reflecting an astrocyte reaction to primary changes in hiNs in response to *BIN1* deletion.

Altogether, results obtained from 2D and 3D models indicate that BIN1 loss-of-function leads to specific transcriptional changes associated with functional properties of glutamatergic neurons.

Molecular alterations in *BIN1* KO organoids and hiNs resemble those observed in the brains of AD patients

We then sought to evaluate whether molecular alterations in our neural models may recapitulate some of those observed in the brain of AD cases. For this purpose, we used a publicly available snRNA-seq dataset generated from the entorhinal cortex (EC) and superior frontal gyrus (SFG) of AD patients at different Braak stages[9]. We first observed a progressive and significant decrease in *BIN1* mRNA levels in glutamatergic neurons (Fig. 2A), suggesting that reduced BIN1 expression in this cell type may be a common feature occurring in the AD

pathology progression. We then compared DEGs identified in *BIN1* KO glutamatergic neurons (either from COs or ASCL1-hiNs) with those identified in the same cell subtype of AD brains (Sup. Table 5). Remarkably, DEGs identified in *BIN1* KO glutamatergic neurons (either from COs or ASCL1-hiNs) showed a statistically significant overlap with DEGs detected in this cell population in AD brains at different Braak stages (Fig. 2B). In astrocytes, however, a similar significant overlap could only be observed between COs and AD brains (Fig. 2B). GO analysis based on DEG overlap between *BIN1* KO ASCL1-hiNs and AD brain glutamatergic neurons indicated significant enrichment for pathways associated with glutamate receptor activity and gated channel activity (Fig. 2C). Similarly, DEG overlap between *BIN1* KO COs and AD brain glutamatergic neurons was significantly enriched for genes associated with glutamate receptor activity, gated channel activity and calcium ion binding (Fig. 2D; Sup. Table 6). No significant enrichment was observed for DEG overlap between *BIN1* KO COs and AD brain astrocytes (data not shown). Altogether, these observations suggest that *BIN1* loss-of-function is sufficient to elicit gene expression alterations in glutamatergic neurons in part similar of those observed in AD brains and associated with functional properties of glutamatergic neurons.

We finally investigated if AD-like biochemical modifications may occur in our different models by measuring the levels of phosphorylated TAU, APP, APP CTF-β and Aβ peptides. We detected an increase in the intracellular levels of phospho-TAU (Ser202, Thr205) in *BIN1* KO compared to WT cultures both in 2D and 3D cultures (Fig. 2E-H). In agreement with our previous observations in cerebral organoids (Lambert et al., 2022), we did not detect any significant differences neither in the concentrations of soluble Aβ(1-x) or Aβ(1-42), nor in the intracellular levels of full-length APP and APP CTF-β in *BIN1* KO compared to WT hiN cultures in 2D (Sup. Fig. 3). Altogether, these results indicate that BIN1 underexpression may be sufficient to induce AD-related Tau hyperphosphorylation in glutamatergic neurons.

Number of synaptic contacts is decreased in BIN KO organoids

Since synapse loss is also an early marker of AD development, we then assessed whether BIN1 deletion may affect synaptic connectivity in our different models. Using immunohistochemistry experiments, we did not find any significant differences in the number of putative synaptic contacts (% SYP assigned) in *BIN1* KO compared to WT ASCL1-hiNs, both at 4 and 6 weeks of differentiation (Fig. 3A-D). We also studied glutamatergic

163 synapses functionally using real-time imaging of ASCL1-hiNs expressing glutamate sensor
164 iGLUSnFr[25]. Like our observations based on immunocytochemistry, we did not detect
165 differences neither in the number of glutamatergic synapses (active spots) nor in the
166 frequency of events (change in fluorescence levels in active spots) in *BIN1* KO compared to
167 WT ASCL1-hiNs (Sup. Fig. 4; Sup. Movies 1 and 2). In contrast, *BIN1* KO COs showed a
168 significant reduction in the number of synaptic contacts (Fig. 3H), mainly due to a reduction
169 in the number of post-synaptic spots expressing HOMER1 (Fig. 3E-F). Thus, our data
170 indicate that long term BIN1 underexpression may affect synaptic connectivity, even if not
171 detectable at short term in 2D culture.
172
173 **_BIN1_ null deletion modifies electrical activity pattern in ASCL1-hiNs**
174 Although we cannot exclude that the latter observations may be linked to a difference
175 between 2D and 3D cultures per se, we postulated that the decrease in synaptic contacts
176 after long-term *BIN1* deletion may be a consequence of synapse down-scaling resulting from
177 chronically increased neuronal excitability due to deregulation of functional properties of
178 glutamatergic neurons[41,42]. To directly address this possibility, we used multi-electrode
179 arrays (MEA) to record and quantify multi-unit activity (MUA) in ASCL1-hiNs cultured in a
180 microfluidic device, which guides neurites into microchannels that are positioned over
181 recording electrodes (Sup. Fig. 5). As observed in dissociated cultures of cortical cells[26], 2D
182 cultures of ASCL1-hiNs cells exhibited a diverse range of spontaneous activity patterns,
183 including regular discharges, population bursts and period activity (Sup. Fig. 4). In this
184 respect, we found a conspicuous change in the temporal organization of MUA after *BIN1*
185 deletion, mainly characterized by an increased number of spike bursts at 4 weeks (Sup. Fig.
186 4). These alterations may result from compensatory adjustments in neuronal connectivity,
187 intrinsic membrane properties or both. To disentangle these possibilities, we used
188 waveform-based spike sorting to examine the functional consequences of *BIN1* deletion at
189 the single neuronal level (Fig. 4). We identified a similar number of single units per recording
190 electrode between genotypes (WT: 4.92±2.34; KO: 5.27±2.45), indicating that *BIN1* deletion
191 does not impair the expression neither the density of active units within culture
192 microchannels. However, we observed reduced single-unit activity (SUA) frequency (Fig. 4B)
193 and increased SUA amplitude (Fig. 4C) in *BIN1* KO compared to WT ASCL1-hiNs.
194 Interestingly, we could not detect significant changes in the number of bursts per neuron

(WT: 11.01±6.71; KO: 10.36±8.59), although the burst duration and the number of spikes within a burst were significantly decreased in *BIN1* KO compared to WT ASCL1-hiNs (Fig. 4D-E), demonstrating the pertinence of performing spike sorting in MEA data. With this approach, we demonstrate the temporal disorganization observed in *BIN1* KO hiNs networks (Fig. 4F) by computing the array-wide spike detection rate (ASDR), which reveals the strength of the synchronized population activity, and the autocorrelograms of SUAs, which allows the apprehension of synchronized periodicity. Both methods revealed striking modifications in the temporal organization of SUAs in *BIN1* KO compared to WT ASCL1-hiNs (Fig. 4G-I). While most spikes of *BIN1* WT neurons occurred at periodic intervals of about 8-10 s, the spikes of *BIN1* KO neurons were randomly distributed, suggesting that *BIN1* deletion in neurons impairs the capacity of these cells to generate organized patterns of electrical activity. Accordingly, the percentage of spikes occurring outside of bursts was significantly higher in *BIN1* KO than in WT ASCL1-hiNs (Fig. 4J).

Acute MEA recordings in 5-month-old COs also revealed a significant increase in spike frequency in *BIN1* KO compared to WT COs (Fig. 5A-B), but these experiments represent a very narrow time shot of COs differentiation. Therefore, to evaluate chronic alterations in neuronal electrical activity in this system, we developed an original approach based on the expression of activity-related genes (ARGs)[27]. While neurons stimulated with brief patterns of electrical activity transcribe rapid primary response genes (rPRGs) or early response genes (ERGs), those stimulated with sustained patterns of electrical activity express delayed primary response genes (dPRGs), secondary response genes (SRGs) and late response genes (LRGs) (Fig. 5C)[28,29]. Using Cell-ID[30], we analyzed the enrichment for these gene signatures (Sup. Table 7) in our COs at single-cell resolution. As expected, we observed that ARG signatures were predominantly enriched in neurons (Fig. 5D). Quantification of the proportion of neurons significantly enriched for specific signatures (p_{adj}<0.05) revealed a significantly higher proportion of glutamatergic neurons enriched for dPRGs, SRGs and LRGs in *BIN1* KO compared to WT COs (Fig. 5E). Enrichments for SRGs and LRGs were specific for this cell type and could not be observed either in GABAergic neurons (Fig. 5E) or in *BIN1* HET glutamatergic neurons (Sup. Fig. 6). Thus, *BIN1* deletion leads to alterations in neuronal electrical activity before observable changes in synaptic connectivity, suggesting that functional changes in *BIN1* KO ASCL1-hiNs are likely a consequence of altered cell-intrinsic properties.

227

228 **BIN1 regulates neuronal Ca^{2+} dynamics through LVGCCs**

229 Since we found significant enrichment for several terms associated with calcium
230 binding and ion channels, we postulated that actors of these pathways may be responsible
231 for such altered cell-intrinsic properties. To probe whether Ca^{2+} dynamics was altered in
232 *BIN1* KO ASCL1-hiNs, we performed calcium imaging in 4-week-old cultures (Sup. Movies 3
233 and 4). We observed a significant increase in the frequency of Ca^{2+} transients in *BIN1* KO
234 compared to WT ASCL1-hiNs, associated with changes in fluorescence dynamics indicative
235 of longer times to reach the maximum intracellular Ca^{2+} levels and to recover baseline levels
236 (Fig. 6A-F).

237 LVGCCs are key regulators of Ca^{2+} transients in neurons, which play a fundamental role
238 in neuronal firing and gene transcription regulation[31]. We thus sought to determine if BIN1
239 may interact and regulate LVGCC expression in hiNs, as previously described for
240 cardiomyocytes[32]. First, we performed proximity ligation assay (PLA) to probe a possible
241 interaction between BIN1 and $Cav1_{.2}$ or $Cav_{1.3}$, the two LVGCCs expressed in ASCL1-hiNs
242 (Sup. Fig. 7). We observed a widespread BIN1-$Cav_{1.2}$ PLA signal (Fig. 6G) and, to a lesser
243 extent, a BIN1-$Cav_{1.3}$ one in neurons (Sup. Fig. 7). Next, we quantified neuronal LVGCC
244 protein level and observed an increase in total $Cav_{1.2}$ levels in *BIN1* KO compared to WT
245 ASCL1-hiNs (Fig. 6H-I). Protein levels of neither $Cav_{1.3}$, nor the members of the Cav_2 family
246 ($Cav_{2.1}$, $Cav_{2.2}$ and $Cav_{2.3}$) were increased in the same cultures (Sup. Fig. 7), suggesting a
247 specific regulation of $Cav_{1.2}$ expression by BIN1.

248 Notably, LVGCCs are key regulators of the synchronous firing pattern in neurons[33] and
249 one of the homeostatic mechanisms protecting neurons from hyperexcitability involves
250 activity-dependent internalization of those channels[34]. Thus, to evaluate whether BIN1
251 deletion may impair this mechanism, we stimulated ASCL-hiNs with KCl 65nM for 30 min
252 and collected total and endosomal proteins for analysis. We confirmed an increase in the
253 global level of $Cav_{1.2}$ in BIN1 KO ASCL1-hiNs that was independent of KCl treatment (Fig. 6J).
254 However, $Cav_{1.2}$ expression in the endosomal fraction was increased by 50% after KCl
255 treatment in BIN1 WT, whereas this increase was only of 10% in BIN1 KO ASCL1-hiNs (Fig.
256 6K). This effect was specific for $Cav_{1.2}$ since both early endosome antigen 1 (EEA1) and $Cav_{1.3}$
257 expression increased in both BIN1 WT and KO ASCL1-hiNs at similar levels after KCl
258 treatment (Fig. 6K).

These last observations prompted us to investigate whether the network dysfunctions observed in *BIN1* KO ASCL1-hiNs may be related to the increased $Cav_{1.2}$ protein levels. For this purpose, we treated these cells with nifedipine, a specific antagonist of $Cav_{1.2}$ at a physiologically relevant concentration (50 nM) for 2 weeks. We observed a partial recovery of the oscillatory pattern of neuronal electrical activity observed in WT cells (Fig. 6L). Interestingly, the percentage of spikes outside bursts was not affected by nifedipine treatment in *BIN1* WT, but significantly decreased in *BIN1* KO ASCL-hiNs (Fig. 6M), indicating a partial recovery of burst organization. To note, no difference in firing rates was observed whatever the models and conditions (Fig. 6N). Altogether, these data support the view that BIN1 contributes to the regulation of electrical activity through the regulation of $Cav_{1.2}$ expression/localization in human neurons.

Discussion

In this work, we show that the AD genetic risk factor *BIN1*, plays a critical role in the regulation of neuronal firing homeostasis in glutamatergic neurons. Complete deletion of *BIN1* gene in these neurons is sufficient to alter the expression of the LVGCC $Cav_{1.2}$, leading to altered calcium homeostasis and neural network dysfunctions in human neurons *in vitro*. These functional changes are correlated with changes in the expression of genes involved in synaptic transmission and ion transport across the membrane, as well as increased Tau phosphorylation. In long-term neuronal cultures using COs, we confirm that *BIN1* loss-of-function affects electrical activity and leads to synapse loss, transcriptional and biochemical alterations resembling those observed in the AD brain. These results suggest that misexpression of BIN1 in glutamatergic neurons may contribute to early stages of AD pathophysiology by dysregulating neuronal firing homeostasis through LVGCCs.

Neuronal network dysfunctions are observed in AD patients at early stages of the disease and precede or coincide with cognitive decline [35–37]. Under physiological conditions, neuronal networks can maintain optimal output through regulation of synaptic and cell-intrinsic mechanisms[38]. Our results suggest that normal levels of BIN1 expression in glutamatergic neurons are fundamental to regulate neuronal firing rate homeostasis. Indeed, *BIN1* deletion in hiNs is sufficient to dysregulate network oscillations even without impacting the number of functional synaptic contacts, suggesting that the desynchronization observed in *BIN1* KO hiNs circuits are a consequence of miscarried homeostatic controls of neuronal activity.

One key mechanism controlling neuronal spiking activity is the regulation of Ca^{2+} homeostasis [31,33,39]. Increased neuronal electrical activity induces the turnover of LVGCCs from the plasma membrane through endocytosis[34] and regulates the transcription of genes encoding for calcium-binding proteins and calcium-mediated signaling[40], mechanisms aiming to restore local Ca^{2+} signaling cascades and protect cells against aberrant Ca^{2+} influx. We show that BIN1 interacts with $Cav_{1.2}$ in hiNs, similar to previous findings in cardiac T tubules[32] and in mouse hippocampal neurons[19] and provide evidence supporting a novel role for BIN1 in the regulation of activity-dependent internalization of $Cav_{1.2}$ in human neurons, thus linking BIN1 to firing homeostasis in human neurons through that LTVGCC.

303 Loss of Ca^{2+} homeostasis is an important feature of many neurological diseases and
304 has been extensively described in AD[41,42]. Interestingly, DEGs identified in glutamatergic
305 neurons in our different cell culture models are enriched for calcium-related biological
306 processes. This is also observed for DEGs detected both in glutamatergic neurons of *BIN1* KO
307 COs and in AD brains. Thus, reduced expression of *BIN1* in glutamatergic neurons may
308 contribute to the breakdown of Ca^{2+} homeostasis in the AD brain, potentially contributing to
309 neuronal circuit dysfunctions. Consistent with this hypothesis, we have previously shown a
310 significant reduction in the expression of the transcript encoding for the neuron-specific
311 BIN1 isoform 1 in bulk RNA-sequencing data from a large number of AD patients[7] and we
312 show in this work that *BIN1* expression is reduced in glutamatergic neurons of AD brains at
313 late Braak stages.
314 Altogether, our results suggest that *BIN1* misexpression in glutamatergic neurons may
315 primarily undermine Ca^{2+} homeostasis, leading to changes in neuronal electrical activity. In a
316 later stage, gene expression and circuit-level alterations such as synapse loss would occur,
317 likely because of altered neuronal electrical activity. A corollary to this model would be that
318 early treatments aiming to restore Ca^{2+} homeostasis and neuronal electrical activity may
319 have a beneficial impact in AD. Interestingly, a Mendelian randomization and a retrospective
320 population-based cohort study found evidence suggesting that Ca^{2+} channel blockers are
321 associated with a reduced risk of AD[43,44]. In the future, it would be interesting to study the
322 impact of these drugs for AD onset/progress as a function of genetic variants in the *BIN1*
323 locus.
324

Acknowledgements and financial support

The authors thank the BICeL platform of the Institut Biologie de Lille and the Vect'UB viral platform (INSERM US 005 – CNRS 3427 – TBMCore, Université de Bordeaux, France). The authors thank Karine Blary at the IEMN Lille for the microfabrication work. The Maestro Pro multiwell microelectrode array was acquired with the "Prix Claude Pompidou pour la Recherche sur l'Alzheimer (2021)" to MRC. This work was co-funded by the European Union under the European Regional Development Fund (ERDF) and by the Hauts de France Regional Council (contract no.18006176), the MEL (contract_2016_ESR_05), and the French State (contract no. 2018-3-CTRL_IPL_Phase2). This work was partly supported by the French RENATECH network (P-18-02737), Fondation pour la recherche médicale (ALZ201912009628, ALZ201906008477), PTR-MIAD, Fondation Recherche Alzheimer and by the Sanofi i-Awards Europe 2019. This work was also funded by the Lille Métropole Communauté Urbaine and the French government's LABEX DISTALZ program (Development of innovative strategies for a transdisciplinary approach to Alzheimer's disease). The UMR 8199 LIGAN-PM Genomics platform (Lille, France) belongs to the 'Federation de Recherche' 3508 Labex EGID (European Genomics Institute for Diabetes; ANR-10-LABX-46) and was supported by the ANR Equipex 2010 session (ANR-10-EQPX-07-01; 'LIGAN-PM'). The LIGAN-PM Genomics platform (Lille, France) is also supported by the FEDER and the Region Nord-Pas-de-Calais-Picardie and is a member of the "France Génomique" consortium (ANR-10-INBS-009).

Declaration of interests

The authors declare no competing interests.

Author contributions

Conceptualization, M.R.C.; Methodology, M.R.C., F.D., D.K., C.M.Q.; Investigation, M.R.C., O.S., B.S.L., A.R.M.F., A.C., A.P., D.S.W., F.D., K.G., D.K., C.M.Q.; Writing - Original Draft, M.R.C.; Writing - Reviews & Editing, M.R.C., F.D., J.C.L., C.M.Q., D.K.; Figures preparation: M.R.C., F.D., A.P., B.S.L., A.R.M.F., D.K. Supervision, M.R.C., F.D., J.C.L., D.K.; Funding Acquisition, M.R.C., J.C.L., F.D, D.K., P.A., A.B. All authors have read and approved the manuscript.

References

1. Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).
2. Kunkle, B. W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).
3. Bellenguez, C., Küçükali, F., Jansen, I., Andrade, V. & Moreno-grau, S. New insights on the genetic etiology of Alzheimer ' s and related dementia. 1–35 (2020).
4. Schwartzentruber, J. *et al.* Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. *Nat. Genet.* **53**, 392–402 (2021).
5. De Rossi, P. *et al.* Predominant expression of Alzheimer's disease-associated BIN1 in mature oligodendrocytes and localization to white matter tracts. *Mol. Neurodegener.* **11**, (2016).
6. De Rossi, P. *et al.* Neuronal BIN1 Regulates Presynaptic Neurotransmitter Release and Memory Consolidation. *Cell Rep.* **30**, 3520-3535.e7 (2020).
7. Marques-Coelho, D. *et al.* Differential transcript usage unravels gene expression alterations in Alzheimer's disease human brains. *npj Aging Mech. Dis.* **7**, 2 (2021).
8. Glennon, E. B. C. *et al.* BIN1 Is Decreased in Sporadic but Not Familial Alzheimer's Disease or in Aging. *PLoS One* **8**, 1–11 (2013).
9. Leng, K. *et al.* Molecular characterization of selectively vulnerable neurons in Alzheimer's disease. *Nat. Neurosci.* **24**, 276–287 (2021).
10. Ubelmann, F. *et al.* Bin1 and CD 2 AP polarise the endocytic generation of beta-amyloid . *EMBO Rep.* **18**, 102–122 (2017).
11. Miyagawa, T. *et al.* BIN1 regulates BACE1 intracellular trafficking and amyloid-β production. *Hum. Mol. Genet.* **25**, 2948–2958 (2016).
12. Lambert, E. *et al.* The Alzheimer susceptibility gene BIN1 induces isoform-dependent neurotoxicity through early endosome defects. *Acta Neuropathol. Commun.* **10**, 4 (2022).
13. Andrew, R. J. *et al.* Reduction of the expression of the late-onset Alzheimer's disease (AD) risk-factor BIN1 does not affect amyloid pathology in an AD mouse model. *J. Biol. Chem.* **294**, 4477–4487 (2019).

14. Chapuis, J. *et al.* Increased expression of BIN1 mediates Alzheimer genetic risk by modulating tau pathology. *Mol. Psychiatry* **18**, 1225–1234 (2013).
15. Sottejeau, Y. *et al.* Tau phosphorylation regulates the interaction between BIN1's SH3 domain and Tau's proline-rich domain. *Acta Neuropathol. Commun.* **3**, 58 (2015).
16. Sartori, M. *et al.* BIN1 recovers tauopathy-induced long-term memory deficits in mice and interacts with Tau through Thr348 phosphorylation. *Acta Neuropathol.* **138**, 631–652 (2019).
17. Calafate, S., Flavin, W., Verstreken, P. & Moechars, D. Loss of Bin1 Promotes the Propagation of Tau Pathology. *Cell Rep.* **17**, 931–940 (2016).
18. Lasorsa, A. *et al.* Structural basis of tau interaction with BIN1 and regulation by tau phosphorylation. *Front. Mol. Neurosci.* **11**, 1–12 (2018).
19. Voskobiynyk, Y. *et al.* Alzheimer's disease risk gene BIN1 induces Tau-dependent network hyperexcitability. *Elife* **9**, 1–25 (2020).
20. Lancaster, M. A. *et al.* Cerebral organoids model human brain development and microcephaly. *Nature* **501**, 373–379 (2013).
21. Trujillo, C. A. *et al.* Complex Oscillatory Waves Emerging from Cortical Organoids Model Early Human Brain Network Development. *Cell Stem Cell* **25**, 558-569.e7 (2019).
22. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *bioRxiv* 1–15 (2019) doi:10.1101/576827.
23. Farhy-Tselnicker, I. *et al.* Activity-dependent modulation of synapse-regulating genes in astrocytes. doi:10.7554/eLife.
24. Christopherson, K. S. *et al.* Thrombospondins are astrocyte-secreted proteins that promote CNS synaptogenesis. *Cell* **120**, 421–433 (2005).
25. Marvin, J. S. *et al.* Stability, affinity, and chromatic variants of the glutamate sensor iGluSnFR. *Nat. Methods* **15**, 936–939 (2018).
26. Wagenaar, D. A., Pine, J. & Potter, S. M. An extremely rich repertoire of bursting patterns during the development of cortical cultures. *BMC Neurosci.* **7**, 1–18 (2006).
27. Flavell, S. W. & Greenberg, M. E. Signaling Mechanisms Linking Neuronal Activity to Gene Expression and Plasticity of the Nervous System. *Annu. Rev. Neurosci.* **31**, 563–590 (2008).

421 28. Tyssowski, K. M. *et al.* Different Neuronal Activity Patterns Induce Different Gene
422 Expression Programs. *Neuron* **98**, 530-546.e11 (2018).
423 29. Hrvatin, S. *et al.* Single-cell analysis of experience-dependent transcriptomic states in
424 the mouse visual cortex. *Nat. Neurosci.* **21**, 120–129 (2018).
425 30. Cortal, A., Martignetti, L., Six, E. & Rausell, A. Gene signature extraction and cell
426 identity recognition at the single-cell level with Cell-ID. *Nat. Biotechnol.* (2021)
427 doi:10.1038/s41587-021-00896-6.
428 31. Simms, B. A. & Zamponi, G. W. Neuronal voltage-gated calcium channels: Structure,
429 function, and dysfunction. *Neuron* vol. 82 24–45 (2014).
430 32. Hong, T. T. *et al.* BIN1 localizes the L-type calcium channel to cardiac T-tubules. *PLoS*
431 *Biol.* **8**, (2010).
432 33. Plumbly, W., Brandon, N., Deeb, T. Z., Hall, J. & Harwood, A. J. L-type voltage-gated
433 calcium channel regulation of in vitro human cortical neuronal networks. *Sci. Rep.* **9**,
434 (2019).
435 34. Green, E. M., Barrett, C. F., Bultynck, G., Shamah, S. M. & Dolmetsch, R. E. The Tumor
436 Suppressor eIF3e Mediates Calcium-Dependent Internalization of the L-Type Calcium
437 Channel CaV1.2. *Neuron* **55**, 615–632 (2007).
438 35. Vossel, K. A. *et al.* Seizures and Epileptiform Activity in the Early Stages of Alzheimer
439 Disease. *JAMA Neurol.* **70**, 1158–1166 (2013).
440 36. Vossel, K. A., Tartaglia, M. C., Nygaard, H. B., Zeman, A. Z. & Miller, B. L. *Review*
441 *Epileptic activity in Alzheimer's disease: causes and clinical relevance*. vol. 16
442 www.thelancet.com/neurology (2017).
443 37. Lam, A. D. *et al.* Silent hippocampal seizures and spikes identified by foramen ovale
444 electrodes in Alzheimer's disease. *Nat. Med.* **23**, 678–680 (2017).
445 38. Turrigiano, G. G. & Nelson, S. B. Homeostatic plasticity in the developing nervous
446 system. *Nat. Rev. Neurosci.* **5**, 97–107 (2004).
447 39. Frere, S. & Slutsky, I. Alzheimer's Disease: From Firing Instability to Homeostasis
448 Network Collapse. *Neuron* vol. 97 32–58 (2018).
449 40. Dörrbaum, A. R., Alvarez-Castelao, B., Nassim-Assir, B., Langer, J. D. & Schuman, E. M.
450 Proteome dynamics during homeostatic scaling in cultured neurons. *Elife* **9**, (2020).
451 41. Bezprozvanny, I. & Mattson, M. P. Neuronal calcium mishandling and the
452 pathogenesis of Alzheimer's disease. *Trends Neurosci.* **31**, 454–463 (2008).

453 42. Carvalho, L. I., Lambert, J.-C. & Costa, M. R. Analysis of modular gene co-expression
454 networks reveals molecular pathways underlying Alzheimer's disease and progressive
455 supranuclear palsy. *medRxiv* 2021.09.21.21263793 (2021)
456 doi:10.1101/2021.09.21.21263793.
457 43. Ou, Y. N. *et al.* Genetically determined blood pressure, antihypertensive medications,
458 and risk of Alzheimer's disease: a Mendelian randomization study. *Alzheimer's Res.*
459 *Ther.* **13**, 1–9 (2021).
460 44. Wu, C. L. & Wen, S. H. A 10-year follow-up study of the association between calcium
461 channel blocker use and the risk of dementia in elderly hypertensive patients. *Med.*
462 *(United States)* **95**, (2016).
463
464

Figure legends

Figure 1: Transcriptional changes in BIN1 KO hiNs. (A) Immunohistochemistry for GFAP (red), MAP2 (green) and DAPI (blue) in 6.5-month-old *BIN1* WT and KO COs. (B) Western blots showing the isoforms of BIN1 detected in WT and the absence of BIN protein in KO COs. (C) UMAP representation of the different cell subtypes in COs identified using snRNA-seq. (D) Dot plot representing the expression for *BIN1* and key markers used to annotate cell subtypes. (E) Proportion of cell subpopulations in both genotypes (****p<0.0001; Chi-squared test). (F) Volcano plots representing DEGs comparing KO vs WT in astrocytes and glutamatergic neurons. DEGs with adjusted p-value <0.05 and |log2FC| >0.25 are shown in red. Gene labels are shown for top 10 genes in terms of log2FoldChange and p-value. (G) Functional enrichment analysis of DEGs identified in glutamatergic neurons. Bar plots representing the top 10 enriched gene ontology (GO) terms in biological processes (BP), cellular components (CC) and molecular function (MF) at p_{adj}<0.01. (H) Images showing *BIN1* WT and KO hiNs 7 days after the beginning of doxycycline treatment immunolabeled for neuronal markers MAP2 and TUBB3 and astrocyte marker GFAP and stained with DAPI. (I) UMAP representation of the different cell subtypes identified in ASCL1-hiNs cultures using snRNA-seq. (J) Dot plot representing expression of key markers used to annotate cell subtypes. (K) Volcano plot representing DEGs comparing *BIN1* KO vs WT glutamatergic neurons. DEGs with adjusted p-value <0.05 and |log2FC| >0.25 are shown in red. Gene labels are shown for calcium- and synapse-related genes. (L) Functional enrichment analysis of DEGs identified in glutamatergic neurons. Bar plots representing the top 10 enriched GO terms in each category at p_{adj}<0.01.

Figure 2: Similar molecular alterations in *BIN1* KO hiNs and glutamatergic neurons of the AD brain. (A) Box plot representing *BIN1* mRNA in expression through different Braak stages in entorhinal cortex (EC) and superior frontal gyrus (SFG) (***p_{adj}<0.001; Wilcoxon test). (B) Dot plot representing the overlap between DEGs identified in glutamatergic neurons of the AD brain and *BIN1* KO ASCL1-hiN cultures (left) or *BIN1* KO COs (right). (C-D) Network representation of enriched GO terms in overlapping DEGs between AD brains and glutamatergic neurons in culture. Enriched GO terms were identified using over-representation test. (E) Western blot for total TAU protein C-terminal (TAU-C), phosphorylated (p)-TAU at Ser202, Thr205 (AT8) and β-ACTIN in 4-week-old ASCL1-hiNs

cultures. (F) Quantification of TAU-C/β-ACTIN and p-TAU/TAU-C levels in BIN1 KO ASCL1-hiNs normalized to WT (*p=0.0262; Mann-Whitney test). (G) Western blot for total TAU protein C-terminal (TAU-C), phosphorylated (p)-TAU at Ser202, Thr205 (AT8) and β-ACTIN in 6.5month-old COs. (H) Quantification of TAU-C/β-ACTIN and p-TAU-TAU-C levels normalized to WT (*p=0.0357; #p=0.0714; Mann-Whitney test).

Figure 3: Similar synaptic density in *BIN1* WT and KO ASCL1-hiNs. (A-B) Immunocytochemistry using the astrocyte marker GFAP, neuronal marker MAP2, pre-synaptic marker SYP and post-synaptic marker HOMER1 in *BIN1* WT ASCL1-hiNs after 4 weeks of differentiation in a three-chamber microfluidic device. Scale bar = 200 µm. Rectangular box in A is magnified in B, allowing the identification of putative synaptic contacts (B'). (C-D) Fraction of SYP spots assigned by HOMER1 spots in MAP2 processes at 4 and 6 weeks ASCL1-hiNs cultures (n= 8 independent devices for each genotype). (E) Immunohistochemistry for HOMER1 (red), SYP (green) in 6.5-month-old *BIN1* WT and KO COs. (F) Quantifications of the number of SYP and HOMER1 spots, and the percentage of SYP assigned by HOMER1 spots in *BIN1* WT and KO COs (**p=0.0076; ***p=0.0002; Mann-Whitney test; n=3 COs per genotype).

Figure 4: Disorganization of neuronal activity in *BIN1* KO ASCL1-hiNs. (A) Raster plots showing the decomposition of multi-unity activity (MUA, black lines) into single-unit activity (SUA, colored lines) using spike waveform clustering. (B-E) Quantification of single-neuron firing rate (B; **p=0.0034), spike amplitude (C; *p=0.0106), burst duration (D; ****p<0.0001) and number of spikes per burst (E; ****p<0.0001) at 4 weeks. Mann-Whitney test; n= 5 independent experiments; WT: 376 neurons; KO: 416 neurons). (F) Raster plots showing SUA recorded from 5 different electrodes of *BIN1* WT (left) or KO (right) ASCL-hiNs cultures after 4 weeks of differentiation. (G) Array-wide spike detection rate (ASDR) plots based on SUA recorded in *BIN1* WT and KO ASCL1-hiNs cultures. Each line represents one independent culture batch. (H-I) Normalized autocorrelogram heatmap (H, each line refers to one SUA) and averaged correlation (I) for all SUAs recorded in 5 independent *BIN1* WT and KO ASCL1-hiNs cultures. (J) Percentage of spikes outside of bursts (*p=0.0417, Mann-Whitney test).

Figure 5: Altered electrical activity in *BIN1* KO COs. (A) Representative raster plots showing detected spikes in 5-month-old *BIN1* WT and KO COs recorded in a multi-well MEA device. (B) Spike frequency in Hz (**p=0.0068; Mann-Whitney test; n=4 WT and 3 KO COs). (C) Scheme indicating the different sets of ARGs regulated by brief and sustained patterns of electrical activity[28,29]. rPRGs: rapid primary response genes; dPRGs: delayed primary response genes; SRGs: secondary response genes; ERGs: early response genes; LRGs: late response genes; Exc – glutamatergic neurons; Inh – GABAergic neurons. (D) Feature plots showing the enrichment score of single cells for ARG signatures. Enrichment scores correspond to the $-\log10(p_{adj})$ of the Cell-ID-based enrichment test. (E) Proportions of glutamatergic (left) and GABAergic neurons (right) enriched for the different ARG signatures according to genotype (*p<0.05; ***p<0.001; Chi-squared test).

Figure 6: Altered frequency of calcium transients in *BIN1* KO ASCL1-hiNs. (A) Snapshot of a 4-week-old ASCL1-hiNs culture labeled with Oregon green BAPTA. (B) Representative plot of fluorescence change over time in 1000 frames. (C) Representative traces showing the fluorescence changes in *BIN1* WT and KO ASCL1-hiNs. Red dashed lines indicate the time to reach the fluorescence maximal intensity (raising time - t1) and to return to baseline (recovery time - t2). (D) Quantification of calcium transients in *BIN1* WT and KO ASCL1-hiNs (****p<0.0001; Mann-Whitney test; n= 3 independent cultures for each genotype; number of active cells per condition: 754 (WT), 1006 (KO)). (E-F) Quantification of rising time (t1) and recovery time (t2) for calcium transients (**p=0.0022; ****p<0.0001; Mann-Whitney test). (G) Images showing PLA spots using anti-BIN1 and anti-$Cav_{1.2}$ antibodies in 4-week-old *BIN1* WT and KO hiNs. Cells were also immunolabeled for the neuronal marker MAP2 (green), the astrocyte marker GFAP (white), and stained with DAPI (blue). (H) Western blot for Cav1.2 (without and with blocking peptide) and β-ACTIN in 4-week-old ASCL1-hiNs cultures. (I) Quantification of Cav1.2/β-ACTIN levels in *BIN1* WT and KO ASCL1-hiNs cultures ($^{\&}$p=0.0585; $^{\#}$p=0.0217; *p=0.0286; Unpaired t-test). (J) Western blot for $Cav_{1.2}$ and β-ACTIN in the total protein extracts from 4-week-old ASCL1-hiNs treated with KCl (+) or vehicle (-). Plot shows the quantification of $Cav_{1.2}$ normalized by β-ACTIN. (K) Western blot for $Cav_{1.2}$, $Cav_{1.3}$ and EEA1 in the endosomal protein extracts from 4-week-old ASCL1-hiNs treated with KCl (+) or vehicle (-). Plot shows the optical density of these proteins (****p<0.0001; Chi-square test). (L) Auto-correlograms of 4-week-old *BIN1* WT and KO hiNs treated or not with 50 nM

Nifedipine for 2 weeks. (M) Percentage of spikes outside of bursts (WT vs WT+NIF: **p_{adj}=0.0034; WT vs KO: *p_{adj}=0.0124; Dunn's multiple comparison test). (N) Average firing rates.

Supplementary data

Sup. Figure 1: Figure 1: Transcriptional changes in BIN1 HET COs. (A) Immunohistochemistry for GFAP (red), MAP2 (green) and DAPI (blue) in 6.5-month-old *BIN1* WT and HET COs. (B) Western blots showing the decrease in BIN1 expression in HET COs. (C) UMAP representation of the different cell subtypes in COs identified using snRNA-seq. (D) Cell proportions in each subpopulation in WT and HET COs. (E) Volcano plot representing DEG comparing HET vs WT in astrocytes and glutamatergic neurons. DEGs with adjusted p-value <0.05 and |log2FC| >0.25 are shown in red. Gene labels are shown for top 10 genes in terms of log2FoldChange and p-value. (F) Functional enrichment analysis of DEGs identified in *BIN1* HET glutamatergic neurons. Bar plots representing the top 10 enriched gene ontology (GO) terms in biological processes (BP), cellular components (CC) and molecular function (MF) at p_{adj}<0.01. (G) Venn diagram showing the overlap between DEGs identified in *BIN1* HET and KO glutamatergic neurons. (H) Bar plots representing the top 10 enriched GO:BP for common DEGs. (I) Immunohistochemistry for HOMER1 (red), synaptophysin (SYP, green) in 6.5-month-old *BIN1* WT and HET COs. (J) Quantification of the percentage of SYP assigned by HOMER1 spots in *BIN1* WT and HET COs (***p=0.0002; Mann-Whitney test; n=3 COs per genotype).

Sup. Figure 2: Transcriptional changes in spontaneously differentiated BIN1 KO hiNPCs. (A) UMAP representation of the different cell subtypes identified in 2D hiNPC cultures after 6 weeks of differentiation using snRNA-seq. (B) Proportion of cell subpopulations in both genotypes. (C) Dot plot representing expression of key markers used to annotate cell subtypes. (D) Volcano plot representing DEGs comparing *BIN1* KO vs WT glutamatergic neurons. DEGs with adjusted p-value <0.05 and |log2FC| >0.25 are shown in red. (E) Functional enrichment analysis of DEGs identified in glutamatergic neurons. Bar plots representing the top 10 enriched GO terms in each category at p_{adj}<0.01.

Sup. Figure 3: Normal APP processing in *BIN1* KO ASCL1-hiNs. (A) Western blots showing the expression of APP full-length, CTF-ß and ß-ACTIN at 4 weeks. (B) Quantification of the ratios APP/ß-ACTIN, CTF-ß/ß-ACTIN and CTF-ß/APP (n = 5 for each genotype). (C) Quantification of soluble $Aß_{1-x}$, $Aß_{1-42}$ and the ratio $Aß_{1-42/}$ $Aß_{1-x}$ in ASCL1-hiNs cultures at 3 and 4 weeks.

Sup. Figure 4: Normal glutamatergic transmission in *BIN1* KO ASCL1-hiNs. Box plots show the number of active spots per neuron and number of events detected by time-lapse video-microscopy in 4- or 6-week-old ASCL1-hiNs cultures transduced with the glutamate sensor iGLUSnFr (4 weeks: n= 378 BIN1 WT and 266 BIN1 KO ASCL1-hiNs; 6 weeks: n= 685 BIN1 WT and 629 BIN1 KO ASCL1-hiNs).

Sup. Figure 5: Increased spike burst frequency in *BIN1* KO ASCL-hiNs. (A) Bright-field image of ASCL1-hiNs cultures in microfluidic/MEA devices showing the cell chamber and micro channels. Neuron somata are mainly restricted to the cell chamber, whereas neuronal processes occupy microchannels. (B) Representative raster plots showing detected spikes in electrophysiological recordings of electrodes underneath the cell chamber and the micro channels, showing the higher sensitivity of the latter. (C) Raster plots showing MUA recorded for 1 minute in *BIN1* WT and KO ASCL1-hiNs after 4 weeks of differentiation. Each line represents one electrode localized side-by-side in our microfluidic/MEA array (as in panel A). (D-E) Quantification of the number of detected spikes at different time points (*p_{adj}=0.0141; ***p_{adj}=0.0006; Two-way ANOVA followed by Tukey's multiple-comparison test; n= 5 for each genotype). (F) Quantification of the number of spike bursts at different time points (**p=0.004; #p=0.0888; Mann-Whitney test).

Sup. Figure 6: Subtle increase in electrical activity in *BIN1* HET COs. (A) Representative raster plots showing detected spikes in 5-month-old *BIN1* WT and HET COs recorded in a multi-well MEA device. (B) Spike frequency in Hz (n=4 WT and 3 HET COs). (C) Proportions of glutamatergic neurons enriched for ARG signatures according to genotype (***p<0.001; Chi-squared test).

Sup. Figure 7: Expression of voltage-gated calcium channels in ASCL1-hiNs. (A) Violin plots showing the mRNA levels of Cav1 and Cav2 members of the voltage-gated calcium channel families L-type, P/Q-type, N-type and R-type detected in ASCL1 hiNs. (B) Images showing PLA spots using anti-BIN1 and anti-Cav1.3 antibodies in 4-week-old *BIN1* WT hiNs. Cells were also immunolabeled for the neuronal marker MAP2 (green), the astrocyte marker GFAP (white), and stained with DAPI (blue). (C) Western blots showing the expression of Cav1.3, Cav2.1, Cav2.2 and Cav2.3 in 4-week-old ASCL1h hiNs. (D) Quantification of protein expression.

Sup. Movies 1 and 2: Time-series of 1000 frames taken from *BIN1* WT and KO ASCL1-hiNs transduced with iGLUSnFr after 2 weeks of differentiation and imaged 2 weeks later. Videos are played at 100 fps.

Sup. Movies 3 and 4: Time-series of 1000 frames taken from *BIN1* WT and KO ASCL1-hiNs after 4 weeks of differentiation and labeled with Oregon Green BAPTA and imaged. Videos are played at 100 frames per second (fps).

Sup. Table 1: DEGs identified in different cell types/subtypes of COs.

Sup. Table 2: GO terms enriched for DEGs identified in different cell types/subtypes of COs.

Sup. Table 3: DEGs identified in different cell types/subtypes of ASCL1-hiNs cultures.

Sup. Table 4: GO terms enriched for DEGs identified in different cell types/subtypes of ASCL1-hiNs cultures.

Sup. Table 5: DEGs identified in different cell types/subtypes of the AD brain.

Sup. Table 6: GO terms enriched for DEGs commonly identified in *BIN1* HET or KO cells and the AD brain.

Sup. Table 7: List of ARGs used for Multiple Correspondence Analysis (MCA) in Cell-ID.

Online methods

Maintenance of cells and generation of hiNPCs and hiNs

hiPSCs modified for BIN1 in exon3 by CRISPR/Cas9 technology were sourced from Applied StemCell Inc. CA, USA. In addition to the BIN1 WT and KO hiPSCS, heterozygous (HET) iPSCS, harbouring a 1 bp insertion in one allele were also sourced Applied Stem Cells Inc. CA, USA.The parental cell line used for derivation of the cells was ASE 9109. The maintenance of these cells and the generation of hiNS, hiAs, and COs thereof, have been detailed in the publication by Lambert et al., 2022. All hiPSCs and their neuronal and glial cell derivatives including COs were maintained in media from Stemcell Technologies, Vancouver, Canada. Maintenance of cell cultures and COs were done following manufacturer's protocols which have been elucidated on the webpage of Stemcell Technologies. In addition, the embryoid body method detailed by Stemcell Technologies was used for the induction of BIN1 WT and KO hiPSCs. Cell numbers and viability were recorded using a LUNA™ Automated Cell Counter (Logos Biosystems, South Korea).

hiNs generated from ASCL1-transduced hiNPCs (protocol detailed in next section) were subjected to differentiation for 4 weeks. All differentiations were performed in tissue in 24-well cell imaging plates (0030741005, Eppendorf) culture dishes pre-coated with Poly-L-ornithine (P4957, Sigma-Aldrich) and Mouse Laminin (CC095, Sigma-Aldrich).

Differentiation protocol for induced hiNPCs

We differentiated neurons from virus-transduced hiNPCs according to an adapted protocol (Zhang et al., 2013; Yang et al., 2017). Briefly, hiNPCs are first transfected with the

TTA lentiviral construct and a passage later, the TetO-Ascl1-Puro lentiviral construct was transduced. These cells are maintained in NPM medium and expanded prior to differentiation. For differentiation of hiNs, hiNPCs are plated onto PLO/laminin-coated imaging plates at density 100,000 cells/well in NPM. After 24h, complete BrainPhys medium (BP) is added 1:1 together with 2 µg/mL doxycycline (Sigma-Aldrich) to induce TetO gene expression. The following day, 1 µg/mL puromycin (Sigma-Aldrich) was added to start cell selection. After 2-3 days (depending on the efficiency of antibiotic selection), 50,000 human cortical astrocytes were added in each well with BrainPhys containing doxycycline. After 24 hours, 2 µM of Ara-C (Cytosine β-D-arabinofuranoside) (Sigma-Aldrich) was added to arrest the proliferation of astrocytes. Half of the medium in each well was changed biweekly with fresh BrainPhys medium (StemCell Technologies) containing doxycycline until the 14th day. After that, the biweekly medium change was performed only with BrainPhys. Differentiation was allowed to continue for another 2 weeks prior to subjecting the cells to various experimental manipulations.

Human cortical astrocytes (Catalog # 1800) were sourced from ScienCell Research Laboratories, CA, USA. Maintenance and proliferation of astrocytes were done as per specifications mentioned on the datasheet from the provider which is available on their webpage.

This culture system was characterized using snRNA-seq showing that 70% of cells (n=3114 from 2 independent culture batches) expressed the pan-neuronal markers SOX11, SNAP25, DCX and RBFOX3, with 66% of cells co-expressing the glutamatergic neuron marker SLC17A6, less than 1.5% of cells co-expressing the GABAergic neuron markers DLX1, GAD1 and GAD2, and 5% of cells co-expressing low levels of markers of both neuronal subtypes. The remaining cells, immature astrocytes (Astro-I), mature astrocytes (Astro-II) and undifferentiated NPCs, represented about 15%, 8%, 4% of the cells, respectively. The first two cell populations likely represent two different states of astrocytes added to the cultures, whereas NPCs are likely cells that failed to reprogram into hiNs despite ASCL1 transduction.

Culture of Induced Neurons (hiNs) in Microfluidic Devices

Preparation of Microfluidic Devices: Three-compartment microfluidic neuron culture devices were used in which the presynaptic and postsynaptic chambers are connected to

719 the synaptic chamber by respectively long and short micro-channels. Details of the
720 microfluidic device design and fabrication have been previously described (Kilinc et al,
721 2020).

722 The homemade devices were placed individually in Petri dishes for easy handling and
723 UV sterilized for 30 min before coating for cell adhesion. The primary surface coating
724 consisted of poly-L-lysine (Sigma-Aldrich) at 20 µg/mL in borate buffer (0.31% boric acid,
725 0.475% sodium tetraborate, pH 8.5). All coated devices were incubated overnight at 37°C,
726 5% CO2. After a wash with DPBS, devices were then coated with 20 µg/mL laminin in DPBS
727 and incubated overnight at 37°C in 5% CO2. The following day, devices were carefully
728 washed once with DPBS before cell plating.

729 Cell Culture: In total, 30,000 NPCs resuspended in complete Neural Progenitor
730 Medium (NPM, Stemcell Technologies) containing 10 µM of Y-27632 ROCK inhibitor were
731 seeded per device, half at the entrance of the presynaptic somatic chamber and half at the
732 entrance of the postsynaptic somatic chamber. Microfluidic devices were microscopically
733 checked at the phase contrast to ensure the cells were correctly flowing into chambers.
734 After a minimum of 5 minutes to allow the cells to attach, devices were filled with NPM
735 (containing 10 µM of Y-27632 ROCK Inhibitor). Water was added to the Petri dishes to
736 prevent media evaporation, and these were then incubated at 37°C in a humidified 5% CO2
737 incubator. The spontaneous neuronal differentiation of NPCs started 24 hours later,
738 initiated by half medium change with complete BrainPhys Neuronal Medium. Induced
739 neuron cultures were maintained for 4 to 6 weeks with half of the medium replaced
740 biweekly with BrainPhys medium.

741 For induced neuron culture from NPCs transduced for Ascl1, doxycycline (2 µg/mL)
742 was added on the first day of half medium change to induce TetO gene expression. The
743 following day, puromycin (1 µg/mL) was added to start cell selection. Two days after the
744 puromycin selection, a total of 5,000 human cortical astrocytes (ScienCell Research
745 Laboratories, CA, USA) were added per device. After 24 hours, Ara-C (2 µM) was added to
746 stop their proliferation. Half of the medium was changed twice a week with complete
747 BrainPhys medium + 2 µg/mL doxycycline for 14 days. After that, half medium change was
748 performed only with BrainPhys medium.

749 Four microfluidic devices were employed for each experimental condition (*BIN1* KO vs
750 WT both for spontaneous neuronal differentiation and Ascl1 induction) and two

751 independent cultures were performed. To assess the time-course effect, neuron cultures
752 were stopped at 4 and 6 weeks.
753
754 *Generation of Cerebral Organoids*
755 Cerebral organoids (3D Cultures) were generated from wild-type, heterozygous and
756 knockout hiPSCs using a 4-stage protocol (Lancaster et al., 2013). The first step was the
757 Embryoid Body (EB) Formation Stage, where hiPCSc at 80%-90% confluency were detached
758 from the Vitronectin XF substrate using Accutase (#AT-104, Innovative Cell Technologies). To
759 form the EB, 9000 cells were plated per well in a 96-well round-bottom ultra-low
760 attachment plate containing EB seeding medium (Stem Cell Technologies). After two days,
761 the EBs were transferred to a 24-well ultra-low attachment plate containing Induction
762 Medium (Stem Cell Technologies), where each well receives 1-2 EBs. This was the Induction
763 Stage. Two days later, the EBs were ready for the Expansion Stage. The EBs were embedded
764 in Matrigel (Corning) and transferred to a 24-well ultra-low adherent plate with Expansion
765 Medium (Stem Cell Technologies). After three days, the medium was replaced by
766 Maturation Medium (Stem Cell Technologies) and the plate was placed in an orbital shaker
767 (100 rpm speed). During this final Maturation Phase, 75% medium change was done on a
768 biweekly basis. Organoids were allowed to mature for a period of 6.5 months.
769
770 *Viral Transductions*
771 Lentiviral constructs were produced by the Vect'UB platform within the TBM Core unit
772 at University of Bordeaux, Bordeaux, France (CNRS UMS 3427, INSERM US 005). The
773 lentiviral constructs used were TTA (ID # 571) and TetO-Ascl1-Puro (Addgene, Plasmid #
774 97329). Lentiviral infections were done in NPCs at P3 or P4. The viral constructs were
775 transduced at a multiplicity of infection (MOI) of 2.5. In brief, NPCs were plated at a
776 confluency of 1x106 cells per well of a 6-well plate. After 4 hours of plating the cells,
777 appropriate volumes of each lentiviral construct were mixed in complete Neural Progenitor
778 medium and 50 µl of the viral medium mix was then added to each well. We transduced the
779 TTA construct at first in the NPCs. Following one passage, the TTA-transduced cells were
780 transduced with the construct for Ascl1. Cells having both viral constructs were then further
781 expanded for 1 or 2 passages before being used for differentiation into hiNs.

782 The iGluSnFR construct was an adeno-associated viral vector (BS11-COG-AAV8)
783 sourced from Vigene Biosciences, MD, USA. The viral construct was transduced at a MOI of
784 5,000 at around 10 days of differentiation for the ASCL1-hiNs. Differentiation was allowed to
785 continue for a duration of 4 weeks prior to imaging.
786
787 *Immunocytochemistry and Immunohistochemistry*
788 Bidimensional (2D) cultures: All cells were fixed in 4% (w/v) paraformaldehyde
789 (Electron Microscopy Sciences, Catalog # 15712) for 10 minutes in the imaging plates.
790 Following, fixation, cells were washed thrice with PBS 0.1 M. Blocking solution (5% normal
791 donkey serum + 0.1% Triton X-100 in PBS 0.1 M) was added to fixed cells at room
792 temperature for 1 hour under shaking conditions. After the blocking step, primary
793 antibodies were added to cells in the blocking solution and incubated overnight at 4°C. The
794 following day, cells were washed with PBS 0.1 M thrice for 10 mins. Each. Alexa Fluor®--
795 conjugated secondary antibodies in blocking solution were then incubated with the cells for
796 2 hours at room temperature under shaking conditions ensuring protection from light.
797 Subsequently, 3 washes with 0.1 M PBS were done for 10 min each at room temperature
798 under shaking conditions with protection from light. Hoechst 33258 solution was added
799 during the second PBS wash. Cells were mounted with Aqua-Poly/Mount (Polysciences, Inc.)
800 and imaged directly in the cell imaging plates. All images were acquired using an LSM 880
801 Confocal Scanning Microscope housed at the Imaging Platform of the Pasteur Institute, Lille.
802 Duolink® Proximity Ligation Assays (PLA) was used to detect endogenous Protein-Protein
803 Interactions. The following pairs of antibodies were used: anti-BIN1 (rabbit, 182562, abcam)
804 and anti-Cav1.2 (mouse, 84814, abcam); or anti-BIN1 and anti-Cav1.3 (mouse, 85491,
805 mouse). Other antibodies used for immunocytochemistry were: MAP2 (188006 and
806 188004, Synaptic Systems), Beta III Tubulin (MAB1637, Sigma-Aldrich), GFAP (AB5804,
807 Millipore; and 173006, Synaptic Systems). All Alexa Fluor®-tagged secondary antibodies
808 were sourced from Jacskon ImmunoResearch Europe Ltd.
809 Microfluidic Devices: Cultured induced neurons were fixed in 4% paraformaldehyde in
810 PBS for 15 min at room temperature, washed three times with PBS, and permeabilized with
811 0.3% Triton X-100 in PBS for 5 min at room temperature. Cells were blocked in PBS
812 containing 5% normal donkey serum for 1 h at room temperature before overnight
813 incubation at 4°C with the following primary antibodies: MAP2 (188006, Synaptic Systems);

HOMER1 (160004, Synaptic Systems), Synaptophysin (101011, Synaptic Systems), and GFAP (AB5804, Millipore). Cells were washed twice with PBS and incubated with the following secondary antibodies for 2h at room temperature: DyLight™ 405 Donkey Anti-Chicken (703-475-155, Jackson ImmunoResearch), Alexa Fluor 594 Donkey Anti-Guinea Pig (706-585-148, Jackson ImmunoResearch), Alexa Fluor 488 Donkey Anti-Mouse (715-545-151, Jackson ImmunoResearch) and Alexa Fluor 647 Donkey Anti-Rabbit (711-605-152, Jackson ImmunoResearch). Cells were rinsed three times with PBS and microfluidic devices were mounted with 90% glycerol.

Samples were imaged with a LSM 880 confocal microscope with a 63X 1.4 NA objective. Images were acquired at zoom 2 in z-stacks of 0.5 µm interval. Typically, 6 images were acquired per device from the synapse chamber near the postsynaptic chamber such the image contains multiple dendrites. Images were deconvoluted using the Huygens software (Scientific Volume Imaging, Netherlands).

Cerebral Organoids: Cerebral organoids were fixed in 4% PFA (w/v) for 30 min at 4°C followed by three washes with PBS 0.1 M. Cerebral organoids were then placed in sucrose solution (30% w/v) overnight before being embedded in O.C.T (Tissue-Tek). Embedded tissue was sectioned at 20 µm using a Cryostar NX70 Cryostat (Thermo Scientific) and mounted slides were stored at −80°C until immunostaining was performed. For immunostaining, tissue sections were brought to room temperature and then rehydrated with 3 washes with 0.1 M PBS, each for 5 mins. Slides were then washed once with PBS with 0.2% Triton X-100 for 15 mins. Tissue was blocked using 10% of donkey serum in PBS 0.1 M for 1 h at room temperature. After blocking, primary antibodies were added to 0.2 % Triton X-100 and 10% of donkey serum in PBS 0.1 M at appropriate dilutions and incubated overnight at 4°C. The next day, slides were washed with PBS 0.1 M 3 times for 5 min each with gentle shaking. Subsequently, slides were incubated with Alexa Fluor®-conjugated secondary antibodies in 0.2 % Triton X-100 and 10% of donkey serum in PBS 0.1 M for 2 h at room temperature in the dark. After secondary antibody incubation, slides were washed 3 times with PBS for 5 min with gentle shanking. Nuclei were visualized by incubating the tissue for 5 min with Hoechst 33258 stain in PBS 0.1 M. Sections were mounted using aqueous mounting medium (Polysciences). Images were acquired using an LSM 880 Confocal Scanning Microscope in concert with the ZEISS ZEN imaging software housed at the Imaging Platform of the Pasteur Institute, Lille. Image acquisition was done at 40X for the

various cellular markers in Fig. 1. The antibodies used were MAP2 (188006, Synaptic Systems) and GFAP (AB5804, Sigma-Aldrich).

Quantification of Synaptic Connectivity

Synaptic connectivity was quantified as previously described (Kilinc et al, 2020). Briefly, images were analyzed with Imaris software (Bitplane, Zürich, Switzerland) by reconstructing Synaptophysin I and HOMER1 puncta in 3D. The volume and position information of all puncta were processed using a custom Matlab (MathWorks, Natick, MA) program. This program assigns each postsynaptic spot to the nearest presynaptic spot (within a distance threshold of 1 µm) and calculates the number of such assignments for all presynaptic puncta.

Immunoblotting

Samples from the 2D cultures or brain organoids were collected in RIPA buffer containing protease and phosphatase inhibitors (Complete mini, Roche Applied Science) and sonicated several times at 60%-70% for 10 seconds prior to use for the immunoblotting analyses. Protein quantification was performed using the BCA protein assay (ThermoFisher Scientific). 10 µg of protein from extracts were separated in NuPAGE 4-12% Bis-Tris Gel 1.0mm (NP0321BOX, Thermo Scientific) or 3-8% Tri-Acetate gel (EA03755BOX, Thermo Scientific) and transferred on to nitrocellulose membranes 0.2µm (#1704158, Bio-Rad). Next, membranes were incubated in milk (5% in Tris-buffered saline with 0.1% Tween-20 (TBST)) or SuperBlock (37536, ThermoFisher Scientific) to block non-specific binding sites for 1 hour at room temperature, followed by several washes with TBST 0.1% or TNT 1x as washing buffers. Immunoblottings were carried out with primary antibodies overnight at 4°C under shaking condition. The membranes were washed three times in the washing buffer, followed by incubation with HRP-conjugated secondary antibodies for 2 hours at room temperature under shaking condition. The membranes were washed three times in washing buffer, and the immune reactivity was revealed using the ECL chemiluminescence system (SuperSignal, ThermoScientific) and imaged using the Amersham Imager 600 (GE Life Sciences). Optical densities of bands were quantified using the Gel Analyzer plugin in Fiji–ImageJ. The primary antibodies used for the immunoblots were as follows: BIN1 (ab182562,Abcam), APP C-terminal (A8717, Sigma-Aldrich), Tau (A002401-2, Agilent)

878 Phospho-Tau(Clone: AT8) (MN1020,ThermoFisher Scientific), CaV1.3 (CACNA1D) (ACC-005,
879 Alomone), CaV2.1 (CACNA1A) (ACC-001, Alomone), CaV2.2 (CACNA1B) (ACC-002, Alomone),
880 CaV2.3 (CACNA1E) (ACC-006, Alomone), CaV1.2 (CACNA1C) (AGP-001 and ACC-003,
881 Alomone), blocking peptide for Anti-CaV1.2 (CACNA1C) (BLP-CC003, Alomone) and β-ACTIN
882 (A1978, Sigma-Aldrich). Secondary antibodies used for the immunoblots were Mouse-HRP
883 (115-035-003, Jackson ImmunoResearch), Rabbit-HRP (111-035-003, Jackson
884 ImmunoResearch), and Guinea pig-HRP (106-035-003, Jackson ImmunoResearch).
885
886 *Activity-dependent endosytosis assay*
887 ASCL1-hiNs (n=9 cultures from each genotype) were subjected to 30 min of
888 depolarization with 65 mM KCl or a mock treatment. Cells were then collected and pulled
889 for endosomal fraction purification using the Minute™ Endosome Isolation and Cell
890 Fractionation Kit (Invent Biotechnologies). Western blot was performed as described above.
891
892 *AlphaLISA measurements*
893 Cell culture media samples for AlphaLISA measurements were collected at the endof
894 the 3rd and 4th weeks of differentiation of the ASCL1-hiNs. Alpha-LISA kits specific for
895 human Aβ1–X (AL288C, PerkinElmer) and Aβ1–42 (AL276C, PerkinElmer) were used to
896 measure the amount of Aβ1–X and Aβ1–42 respectively in culture media. The human Aβ
897 analyte standard was diluted in the BrainPhys medium. For the assay, 2 µL of cell culture
898 medium or standard solution was added to an Optiplate-384 microplate (PerkinElmer). 2 µL
899 of 10X mixture including acceptor beads and biotinylated antibody was then added to the
900 wells with culture media or standard solution. Following incubation at room temperature
901 for an hour, 16 µL of 1.25X donor beads was added to respective wells and incubated at
902 room temperature for 1 hour. Luminescence was measured using an EnVision-Alpha Reader
903 (PerkinElmer) at 680-nm excitation and 615-nm emission wavelengths.
904
905 *Calcium and iGluSnFR Imaging*
906 Calcium imaging was performed in 2D cultures after 4-weeks (Ascl1-induced). Prior to
907 imaging, the cells were incubated with Oregon Green™ 488 BAPTA-1 (OGB-1) acetoxymethyl
908 (AM) (ThermoFisher Scientific) for 1 hour. A 2.5 mM stock solution of the calcium-indicator
909 dye was prepared in Pluronic™ F-127 (20% solution in DMSO) (ThermoFisher Scientific). 1 µL

of the dye solution was added to 400 µL of fresh BrainPhys medium in each well of a 24-well cell imaging plate. Existing BrainPhys media from the wells of the plate was removed and kept aside while the calcium-indicator dye was incubated in fresh BrainPhys medium. After the 1-hour incubation, the medium which was kept aside was replaced to each well. The 2D cultures were then ready to be filmed using a Spinning Disk Microscope housed at the Institut Pasteur de Lille, Lille, France using the MetaMorph imaging software.

For filming the calcium activity, 1000 images were taken using a 20X long-distance objective, 10 ms exposure time and 200ms intervals. For each well, 5 random fields were chosen, and the cellular activity was, thus, recorded.

For cells transduced with iGluSnFR, these cells were directly filmed after 4 weeks of differentiation and 500 images were taken using a 20X long-distance objective, 10 ms exposure time and 200ms intervals. Up to 8 fields per well were filmed, each field containing at least one fluorescent transduced cell along with its processes.

Analyses of Calcium Transients

All live recordings of neuronal calcium transients were first converted into .avi format after background subtraction using the FIJI software. Following these, the videos were subsequently opened using the free software for data analyses of calcium imaging, CALciumIMagingAnalyzer (CALIMA) made available online by Fer Radstake (Eindhoven University of Technology, The Netherlands). Each video recording of a field of cells was first downscaled to 2X in terms of size with a 10X zoom and was checked for the frame average mode. Moreover, in this first detection stage, pre-set filter parameters were adjusted and applied to enable the detection of the maximum number of fluorescent cells in each field. In the analysis tab, detection of the average activity was checked and for pre-processing, a median of 3 was applied. All cells within the pre-set filter parameters are detected as regions of interest (ROIs) in the detection stage. Cell activity from all detected ROIs is then recorded. However, in the subsequent analysis stage, only cells showing spiking frequencies with a standard deviation of at least 2 or more were taken into consideration. Data in the form of detection spikes and the correlation (peak) are extracted and exported as CSV files.

Electrophysiological recordings in 2D cultures and cerebral organoids

ASCL1-hiNs were cultured in the aforementioned microfluidic devices bound to multi-electrode arrays (256MEA100/30iR-ITO, Multi-Channel Systems, Germany). Extracellular action potentials were recorded in 5 different cultures for both genotypes at 2, 3, 4 and 6 weeks of differentiation using the MEA2100-256-System (Multi-Channel Systems). Before recordings, MEAs were let stabilize for 5 min on the headstage to reduce artifacts due to medium movement. Signals were recorded for 1 min, at 40 kHz sampling rate, using Multi Channel Experimenter 2.16.0 software (Multi-Channel Systems). Electrical activity in cerebral organoids was recorded using 256-6wellMEA200/30iR-ITO (Multi-Channel Systems, Germany). Briefly, 5-6-month-old cerebral organoids were mounted onto MEAs and kept for 2 h in complete Brainphys medium. Then, MEAs were placed on the headstage and let stabilize for 5 min before recordings. Signals were recorded for 5 min, at 10 kHz sampling rate using Multi-Channel Experimenter 2.16.0. For rescue experiments using a calcium channel blocker, ASCL1-hiNs were cultured MEA 96-well plates (CytoView MEA 96, Axion Biosystems, USA). Extracellular action potentials were recorded in 3 independent cultures for either genotype in the presence of 50nM nifedipine (Tocris Bioscience) or vehicle using the MaestroPro (Axion Biosystems, Inc, USA). Before recordings, MEAs were let stabilize for 5 min on the MaestroPro at 37°C and 5% CO_2. Signals were recorded for 3 min, at 12.5 kHz sampling rate, using AxIS Navigator software (Axion Biosystems).

Spikes were detected using a fixed amplitude threshold of 5.5 and 4.5 standard deviations (for the 2D and 3D cultures, respectively) of the high-pass filtered (>300 Hz) signal for positive- and negative-going signals. The detection included a dead time of 3 ms to account for the refractory period of action potentials. Quantification of the number of detected spikes (MUAs) and spike bursts (defined as at least 5 spikes within 50 ms) was performed using Multi-Channel Analyzer 2.16.0 software (Multi-Channel Systems).

Spike sorting and temporal structure of spontaneous activity

Channels containing detected waveforms were manually processed offline for spike waveform separation and classification using Offline Sorter v3 (Plexon, USA). Briefly, we applied principal component analysis (PCA) to cluster spike waveforms of similar morphologies. Using this approach, we identified from 2 to 10 well-isolated units per channel, and therefore, we considered this single-unit activity (SUA). For each SUA, we computed the average firing rate, the signal-to-noise ratio, the peak-to-trough amplitude

and duration, the average power (square amplitude of the average waveform), the mode of the interspike interval distribution, and their firing patterns. It has been demonstrated that dissociated neuronal cultures can develop complex discharge structures (Wagenaar, 2006). Here, we considered burst activity if the SUA presents periods of high-frequency discharges interspersed by regular or no discharges at all. Operationally, a burst must have at least 3 spikes within 100 ms and 200 ms intervals, for the interval between the first and the second, and the second and the third discharge, respectively. After the third spike, the maximal interval to consider a discharge part of the burst was 200 ms. Thus, we computed the SUA that presented bursts, the number of bursts (i.e., the burst frequency), the average burst duration, the number of spikes within each burst, the average burst frequency, and the inter-burst interval.

Two complementary approaches investigated the temporal structures of spike trains. In the first one, we computed the array-wide spike detection rate (ASDR), which is the number of spikes detected per unit of time, summed over all electrodes in the array. This method is commonly used in the literature to demonstrate synchronous activity (aka, bursts) in MUA data (Wagenaar 2006). The second approach uses the autocorrelation function (i.e., the probability of finding two spikes at a given time interval) to calculate the oscillation score and the oscillation period of every single unit (Muresan 2008:1333, J Neurophysiol). Briefly, the oscillation score was calculated as the averaged absolute magnitude difference between the positive and negative peaks of the smoothed autocorrelation function (bin size of 200 ms). The oscillation period was calculated as the averaged time interval of the positive peaks of the autocorrelation function.

snRNA-seq Library Preparation

Nuclei isolation and Hash-tagging with oligonucleotides steps were realized on ice with pre-cold buffers and centrifugations at 4°C. 6.5-month-old BIN1 WT, HET, and KO organoids were processed as previously (Lambert et al., 2022). 4-week-old cultured ASCL1-induced BIN1 WT and KO 2D cultures were washed in the imaging plate wells with 500 µL of Deionized Phosphate Buffer Saline 1X (DPBS, GIBCO™, Fisher Scientific 11590476). Cells were resuspended with wide bore tips in 500 µL Lysis Buffer (Tris-HCL 10mM, NaCl 10mM, MgCl2 3mM, Tween-20 0,1%, Nonidet P40 Substitute 0,1%, Digitonin 0,01%, BSA 1%, Invitrogen™ RNAseout™ recombinant ribonuclease inhibitor 0,04 U/µL). Multiple

mechanical resuspensions in this buffer were performed for a total lysis time of 15 mins., 500 µL of washing buffer was added (Tris-HCL 10mM, NaCl 10 mM, MgCl2 3 mM, Tween-20 0.1%, BSA 1%, Invitrogen™ RNAseout™ recombinant ribonuclease inhibitor 0,04 U/µL) and the lysis suspension was centrifuged 8 mins. at 500 g (used for all following centrifugation steps). Nuclei pellets were washed tree times with one filtration step by MACS pre-separation filter 20µm (Miltenyi Biotec). Nuclei pellets were resuspended in 100 µL of staining buffer (DPBS BSA 2%, Tween-20 0.01%), 10 µL of Fc blocking reagent HumanTruStainFc™ (422302, Biolegend) and incubated 5 min at 4°C. 1µl of antibody was added (Total-Seq™-A0453 anti-Vertebrate Nuclear Hashtag 3 MAb414 for the WT and Total-Seq™-A0454 anti-Vertebrate Nuclear Hashtag 4 MAb414 for the KO, 97286 and 97287 respectively, Biolegend) and incubated 15 mins. at 4°C. Nuclei pellets were washed three times in staining buffer with one filtration step by MACS pre-separation filter 20 µm (Miltenyi Biotec) to a final resuspension in 300 µL of staining buffer for Malassez cell counting with Trypan blue counterstaining (Trypan Blue solution, 11538886, Fisherscientific). Isolated nuclei were loaded on a Chromium 10X genomics controller following the manufacturer protocol using the chromium single-cell v3 chemistry and single indexing and the adapted protocol by Biolegend for the HTO library preparation. The resulting libraries were pooled at equimolar proportions with a 9 for 1 ratio for Gene expression library and HTO library respectively. Finally, the pool was sequenced using 100pb paired-end reads on NOVAseq 6000 system following the manufacturer recommendations (Illumina).

snRNA-seq Dataset Preprocessing

Unique Molecular Index (UMI) Count Matrices for gene expression and for Hash Tag Oligonucleotide (HTO) libraries were generated using the CellRanger count (Feature Barcode) pipeline. Reads were aligned on the GRCh38-3.0.0 transcriptome reference (10x Genomics). Filtering for low quality cells according to the number of RNA, genes detected, and percentage of mitochondrial RNA was performed. For HTO sample, the HTO matrix was normalized using centered log-ratio (CLR) transformation and cells were assigned back to their sample of origin using HTODemux function of the Seurat R Package (v4)[10]. Then, normalizations of the gene expression matrix for cellular sequencing depth, mitochondrial

1036 percentage and cell cycle phases using the variance stabilizing transformation (vst) based
1037 Seurat:SCTransform function were performed.
1038
1039 *snRNA-seq datasets integration and annotation*
1040 To integrate the datasets from independent experiments, the harmony R package
1041 (https://github.com/immunogenomics/harmony) was used. In order to integrate the
1042 datasets, the SCTransform normalized matrices was merged and PCA was performed using
1043 Seurat::RunPCA default parameter. The 50 principal components (dimensions) of the PCA
1044 were corrected for batch effect using harmony::RunHarmony function. Then, the 30 first
1045 batch corrected dimensions were used as input for graph-based cell clustering and
1046 visualization tool. Seurat::FindNeighbors using default parameters and Seurat::FindClusters
1047 function using the Louvain algorithm were used to cluster cells according to their batch
1048 corrected transcriptomes similarities. To visualize the cells similarities in a 2-dimension
1049 space, the Seurat::RunUMAP function using default parameter was used. Cell clusters were
1050 then annotated based on cell type specific gene expression markers.
1051
1052 *Differential gene expression and GO enrichment analyses*
1053 Gene expression within each main cell type was compared between conditions of
1054 interest using Wilcoxon test on the SCTransform normalized gene expression matrix. GO
1055 enrichment analysis on the differentially expressed genes was performed using the gost
1056 function of the gprofiler2 R package (CRAN).
1057
1058 *Activity-related genes (ARGs) signature enrichment analysis at single cell resolution*
1059 To study enrichment for activity-related genes (ARGs) signature across cerebral
1060 organoid cells, the CellID R package (https://github.com/RausellLab/CelliD) was used. ARGs
1061 obtained from Tyssowski et al. (2018) and Hravtin et al. (2018) (supplementary Table 7),
1062 were translated to the corresponding human gene name with the help of the biomaRt
1063 package using the respective Ensembl references. Then, the CellID::RunMCA was used to
1064 extract cell-specific gene signature and hypergeometric test was performed to test
1065 enrichment for ARGs in these cell signatures. To test the differential proportion of ARGs
1066 enriched cells in BIN1 deleted organoid compared to WT organoid, chi-squared test was
1067 performed.

Comparative analysis with specific DEGs in AD brains

To compare the transcriptomic change observed in BIN1 deleted cerebral organoid with those observed in AD brain, datasets from the work of Leng et al. (ref) and Morabito et al. (ref) were taken as 2 independent references. The raw gene expression matrix was normalized using Seurat::SCTransform and differential expression analysis was performed within each neuronal cell type using Wilcoxon test as used for our organoid dataset. AD related DEGs, thus, obtained were compared with our BIN1 related organoid DEGs in every cell type. To this end, the enrichment for AD-related DEGs in BIN1-related DEGs was tested using hypergeometric test. The background for this test was defined as all genes detected in both datasets. The p-value of this test was used as metrics to compare the significance of the gene overlap between neuronal cell types.

Statistical analysis

Statistical analysis was performed using GraphPad Prism version 8.0.0 (GraphPad Software, San Diego, California USA, www.graphpad.com) and R 4.2.0 (R Core Team, 2022, https://cran.r-project.org/bin/windows/base/old/4.2.0/). Bar plots show mean ± SD and individual values. Box plots show 1-99 percentile. Statistical tests and p values are indicated in figure legends.

Figure 2

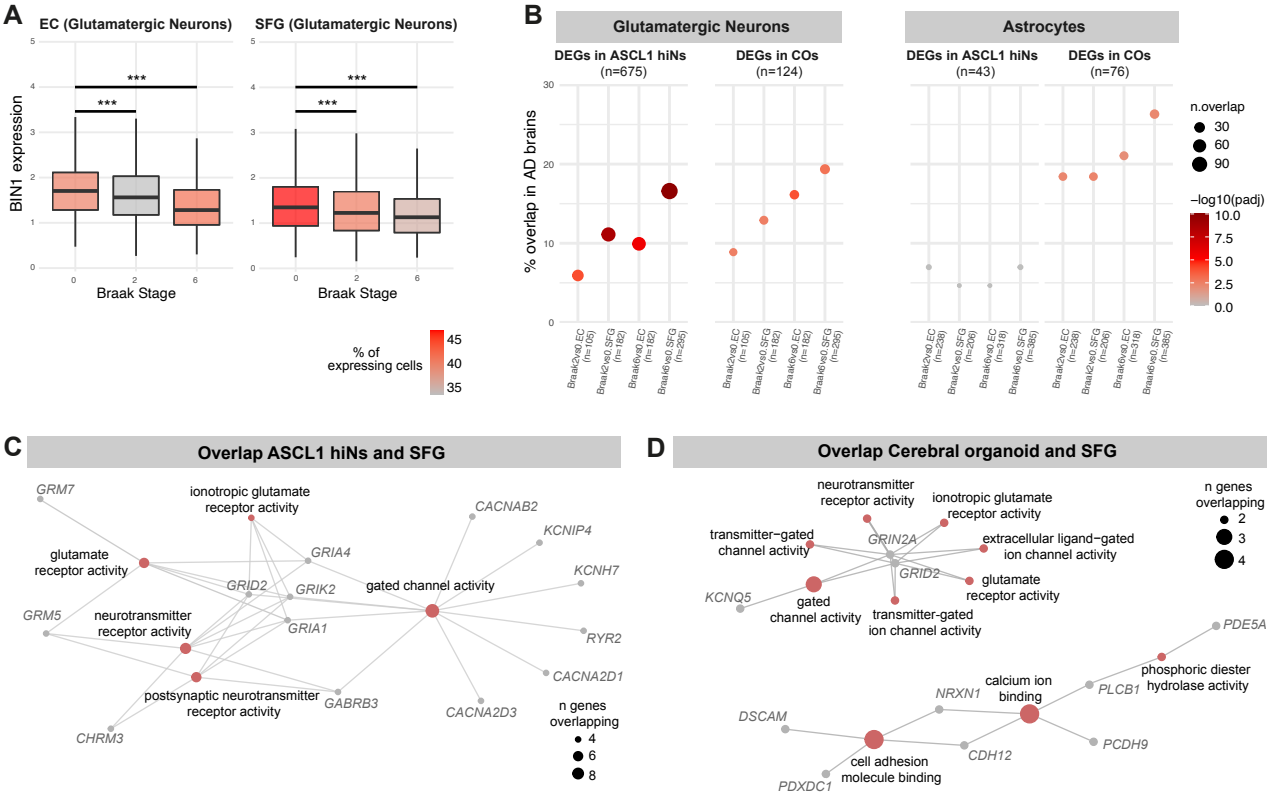

A
EC (Glutamatergic Neurons)
SFG (Glutamatergic Neurons)
BIN1 expression
Braak Stage
% of expressing cells
B
Glutamatergic Neurons
Astrocytes
DEGs in ASCL1 hiNs (n=675)
DEGs in COs (n=124)
DEGs in ASCL1 hiNs (n=43)
DEGs in COs (n=76)
% overlap in AD brains
n.overlap
−log10(padj)
C
Overlap ASCL1 hiNs and SFG
glutamate receptor activity
ionotropic glutamate receptor activity
gated channel activity
neurotransmitter receptor activity
postsynaptic neurotransmitter receptor activity
n genes overlapping
D
Overlap Cerebral organoid and SFG
neurotransmitter receptor activity
ionotropic glutamate receptor activity
transmitter−gated channel activity
extracellular ligand−gated ion channel activity
glutamate receptor activity
gated channel activity
transmitter-gated ion channel activity
calcium ion binding
phosphoric diester hydrolase activity
cell adhesion molecule binding
n genes overlapping
E
WT ASCL1 hiNs
KO ASCL1 hiNs
50 kDa
40 kDa
60 kDa
TAU-C
β-ACTIN
p-TAU (AT8)
F
Ratio TAU-C/β-ACTIN
Ratio p-TAU (AT8)/β-ACTIN
Ratio p-TAU (AT8)/TAU-C
G
WT COs
KO COs
P-TAU (AT8)
H
WT
KO

Figure 3

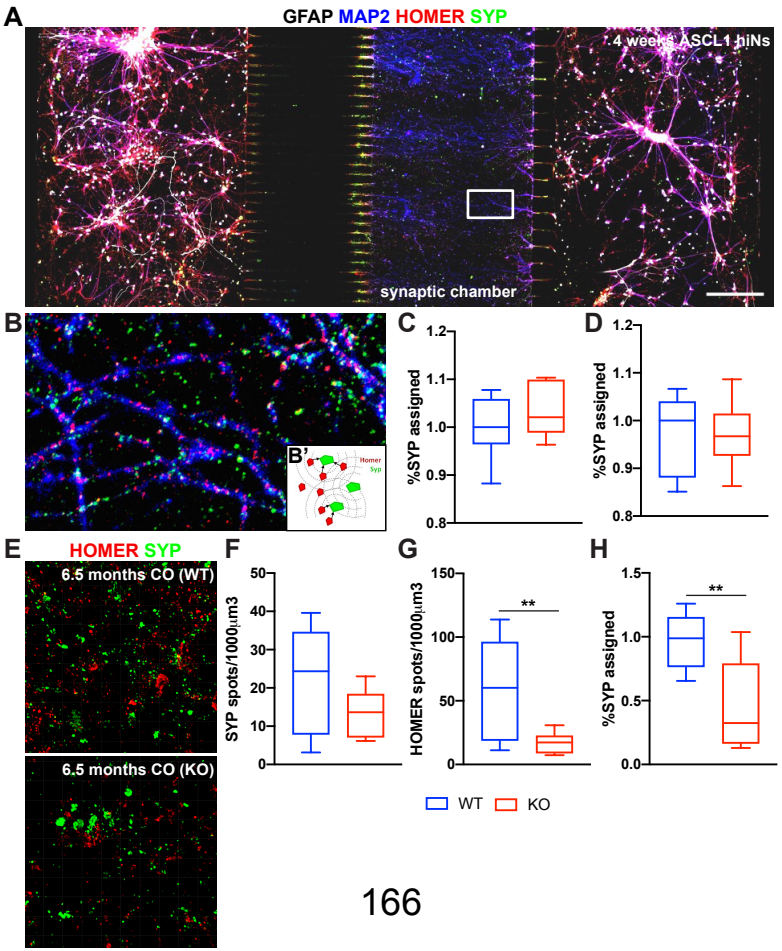

A
GFAP MAP2 HOMER SYP
4 weeks ASCL1 hiNs
synaptic chamber
B
B'
Homer
Syp
C
%SYP assigned
1.2
1.1
1.0
0.9
0.8
D
%SYP assigned
1.2
1.1
1.0
0.9
0.8
E
HOMER SYP
6.5 months CO (WT)
6.5 months CO (KO)
F
SYP spots/1000µm3
50
40
30
20
10
0
G
HOMER spots/1000µm3
150
100
50
0
**
H
%SYP assigned
1.5
1.0
0.5
0.0
**
WT
KO

Figure 4

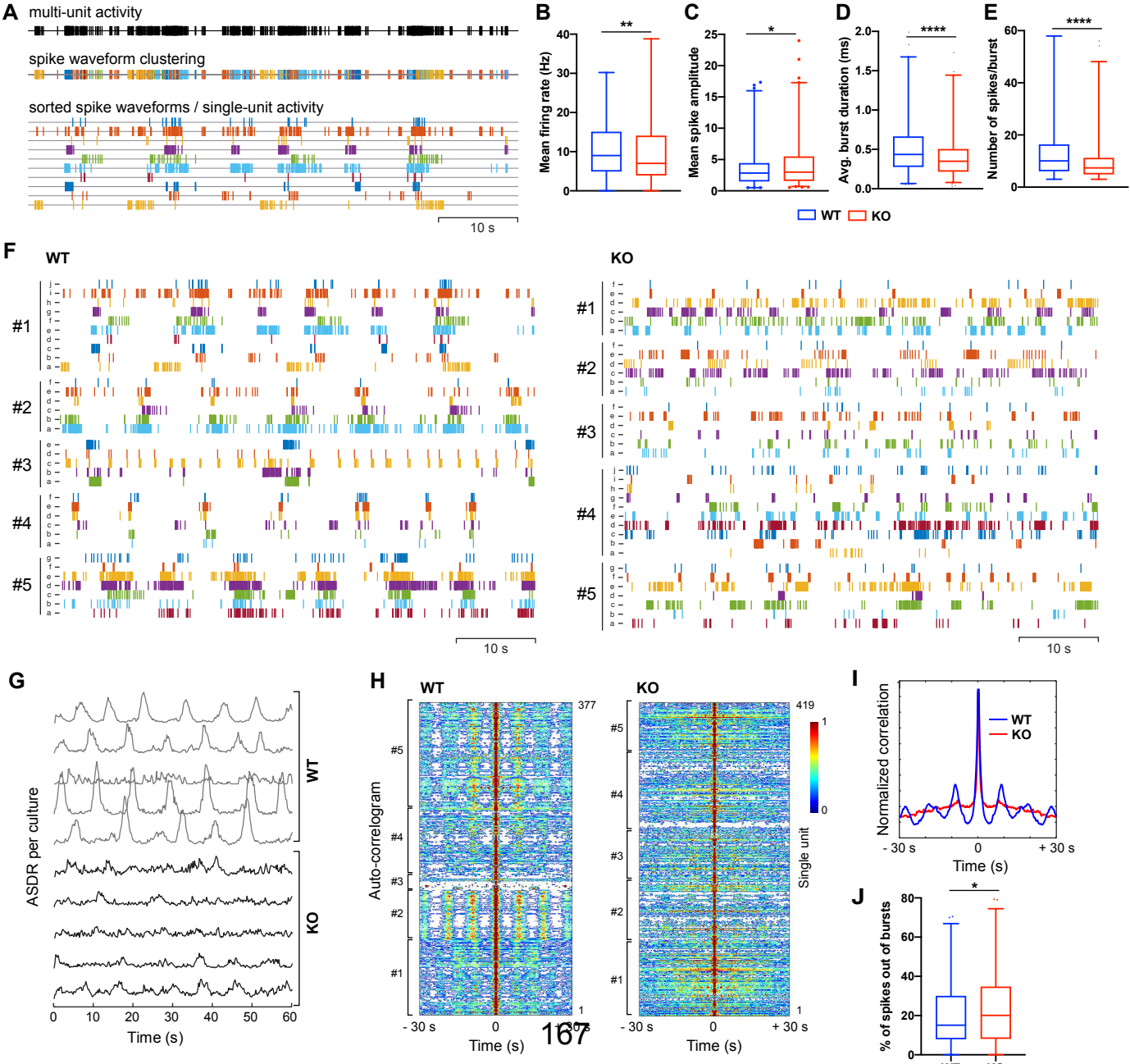

A
multi-unit activity
spike waveform clustering
sorted spike waveforms / single-unit activity
10 s
B
Mean firing rate (Hz)
**
C
Mean spike amplitude
*
D
Avg. burst duration (ms)
****
E
Number of spikes/burst
****
WT
KO
F
WT
KO
#1
#2
#3
#4
#5
10 s
G
ASDR per culture
WT
KO
0 10 20 30 40 50 60
Time (s)
H
WT
KO
Auto-correlogram
377
419
Single unit
- 30 s
0
+ 30 s
Time (s)
I
Normalized correlation
WT
KO
- 30 s
0
+ 30 s
Time (s)
J
% of spikes out of bursts
*
WT
KO

Figure 5

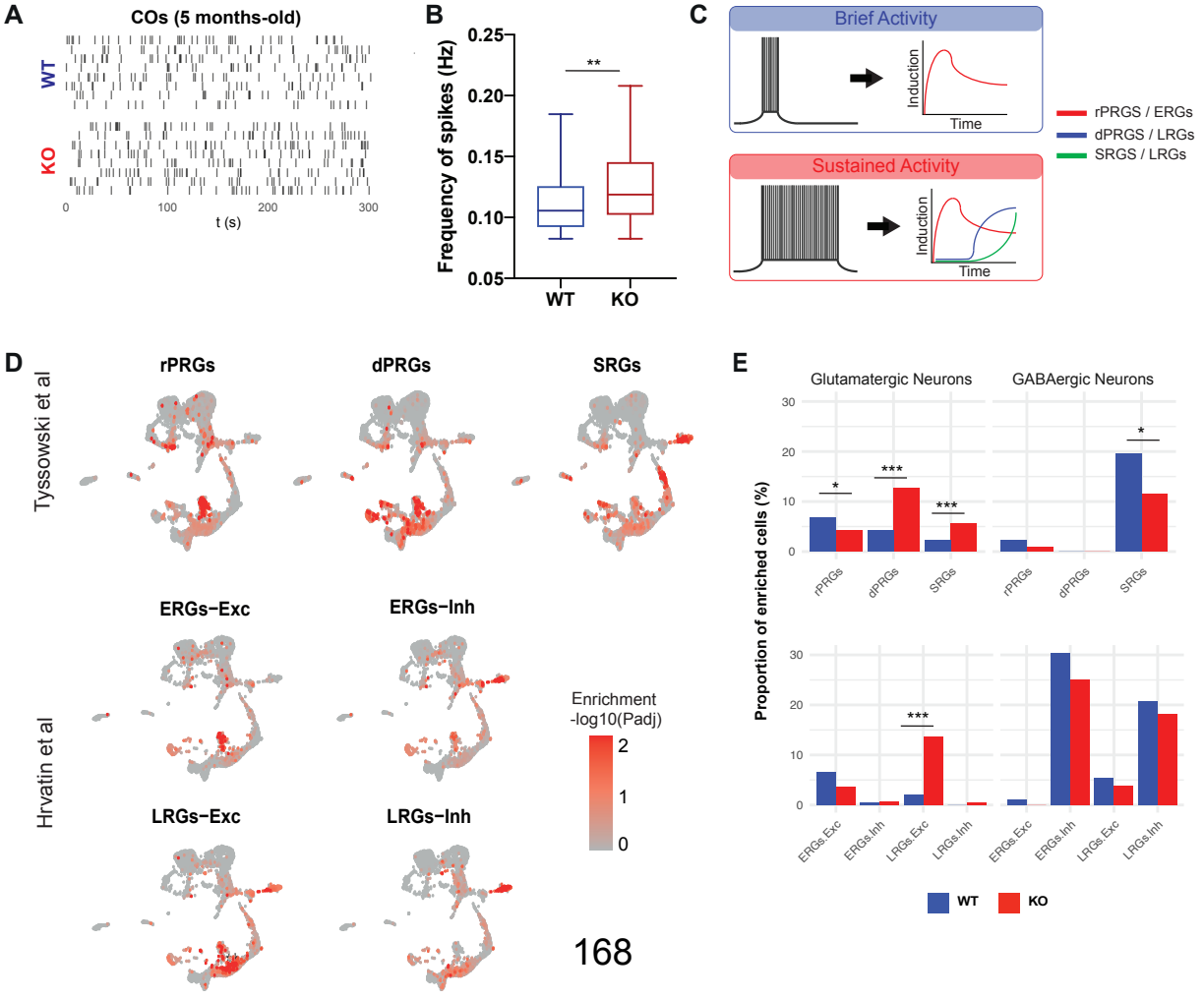

A
COs (5 months-old)
WT
KO
0
100
200
300
t (s)
B
Frequency of spikes (Hz)
0.25
0.20
0.15
0.10
0.05
**
WT
KO
C
Brief Activity
Induction
Time
Sustained Activity
Induction
Time
rPRGS / ERGs
dPRGS / LRGs
SRGS / LRGs
D
Tyssowski et al
Hrvatin et al
rPRGs
dPRGs
SRGs
ERGs−Exc
ERGs−Inh
LRGs−Exc
LRGs−Inh
Enrichment
-log10(Padj)
2
1
0
168
E
Glutamatergic Neurons
GABAergic Neurons
Proportion of enriched cells (%)
30
20
10
0
*
***
***
*
rPRGs
dPRGs
SRGs
ERGs.Exc
ERGs.Inh
LRGs.Exc
LRGs.Inh
WT
KO

Figure 1

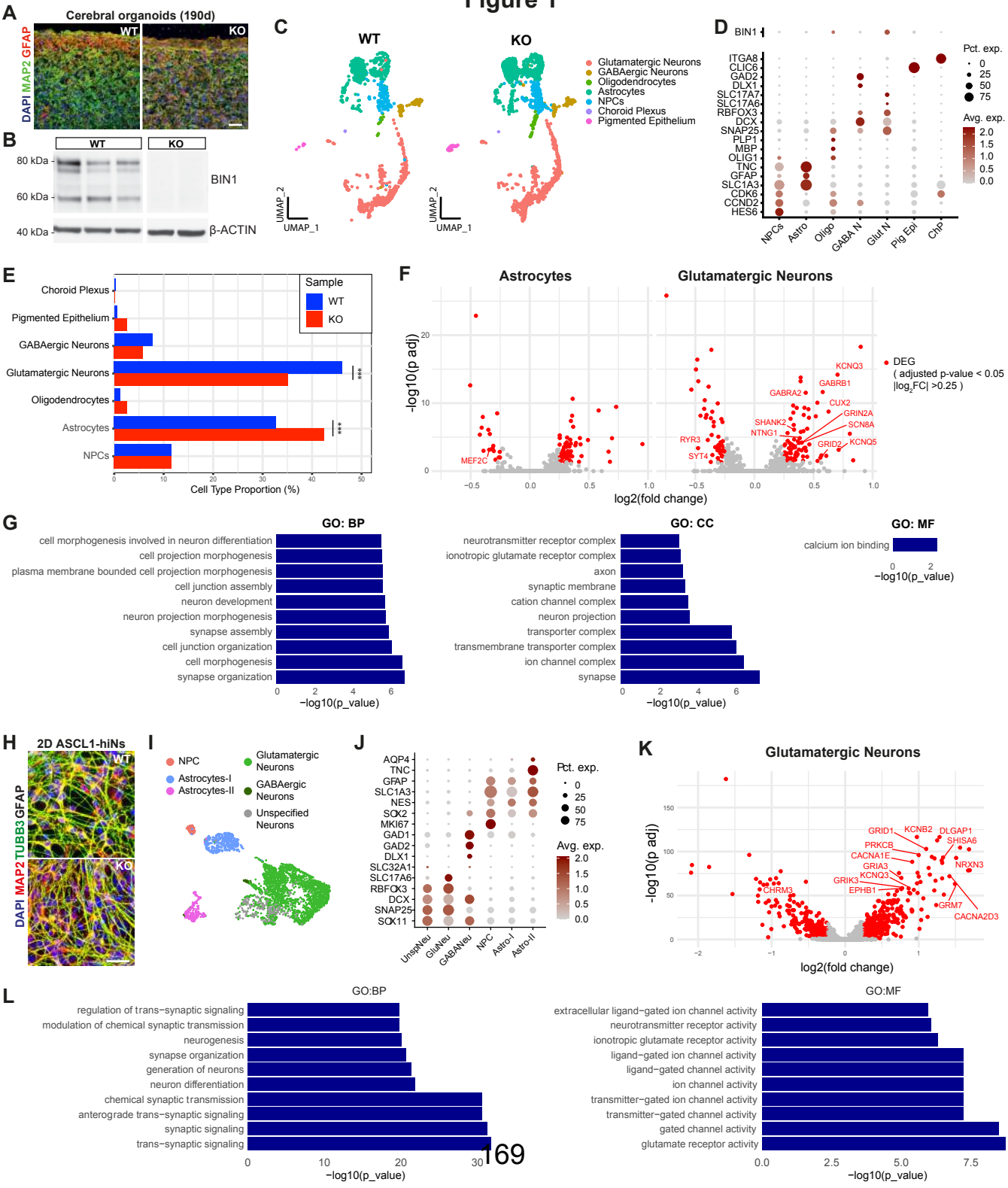
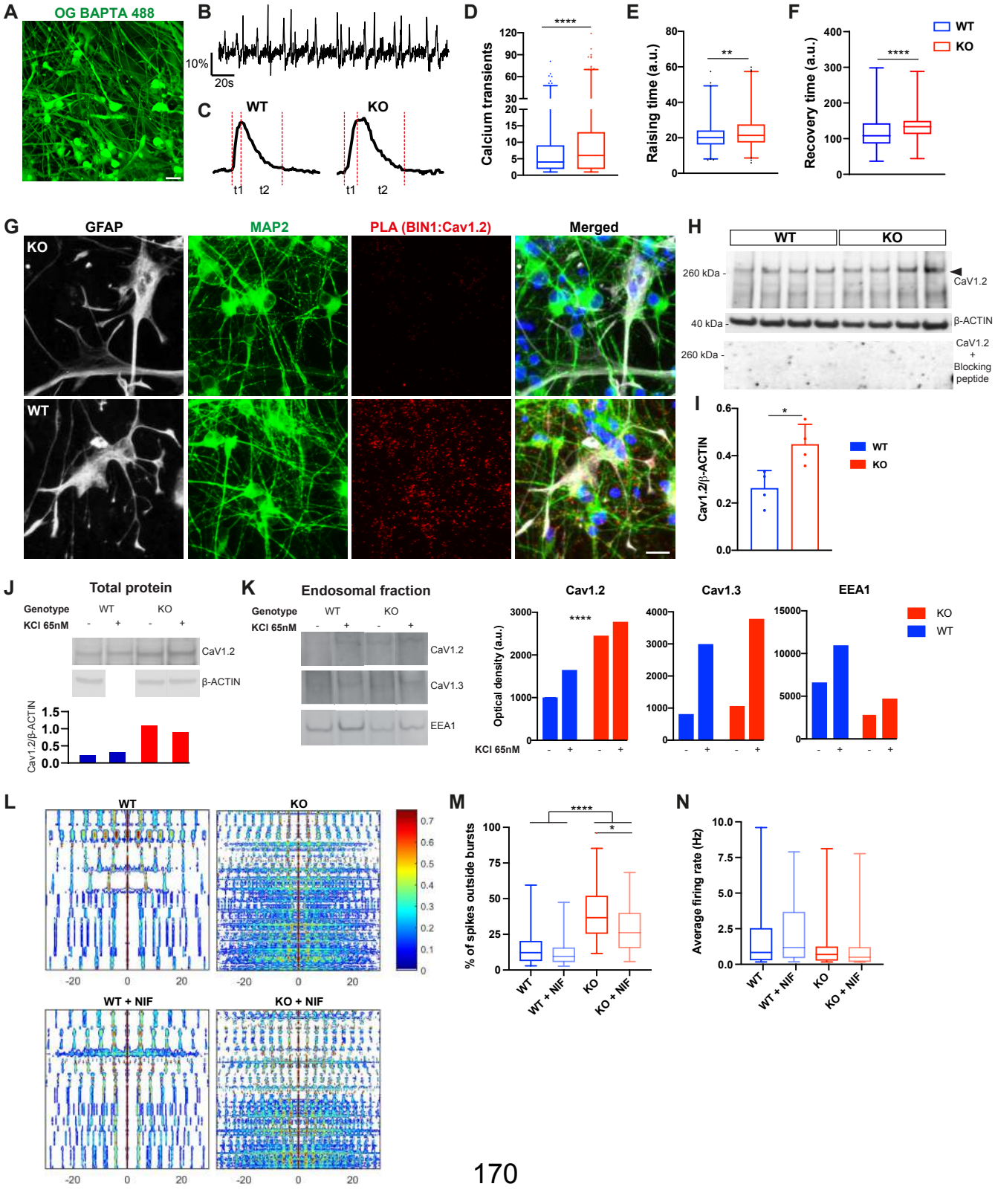

A
Cerebral organoids (190d)
WT
KO
DAPI MAP2 GFAP
B
WT
KO
80 kDa
60 kDa
40 kDa
BIN1
β-ACTIN
C
WT
KO
UMAP_2
UMAP_1
Glutamatergic Neurons
GABAergic Neurons
Oligodendrocytes
Astrocytes
NPCs
Choroid Plexus
Pigmented Epithelium
D
BIN1
ITGA8
CLIC6
GAD2
DLX1
SLC17A7
SLC17A6
RBFOX3
DCX
SNAP25
PLP1
MBP
OLIG1
TNC
GFAP
SLC1A3
CDK6
CCND2
HES6
NPCs
Astro
Oligo
GABA N
Glut N
Pig Epi
ChP
Pct. exp.
0
25
50
75
Avg. exp.
2.0
1.5
1.0
0.5
0.0
E
Choroid Plexus
Pigmented Epithelium
GABAergic Neurons
Glutamatergic Neurons
Oligodendrocytes
Astrocytes
NPCs
Sample
WT
KO
***
***
Cell Type Proportion (%)
F
Astrocytes
Glutamatergic Neurons
-log10(p adj)
log2(fold change)
DEG
(adjusted p-value < 0.05
|log2FC| >0.25)
MEF2C
RYR3
SYT4
SHANK2
NTNG1
KCNQ3
GABRB1
GABRA2
CUX2
GRIN2A
SCN8A
GRID2
KCNQ5
G
GO: BP
cell morphogenesis involved in neuron differentiation
cell projection morphogenesis
plasma membrane bounded cell projection morphogenesis
cell junction assembly
neuron development
neuron projection morphogenesis
synapse assembly
cell junction organization
cell morphogenesis
synapse organization
−log10(p_value)
GO: CC
neurotransmitter receptor complex
ionotropic glutamate receptor complex
axon
synaptic membrane
cation channel complex
neuron projection
transporter complex
transmembrane transporter complex
ion channel complex
synapse
−log10(p_value)
GO: MF
calcium ion binding
−log10(p_value)
H
2D ASCL1-hiNs
WT
KO
DAPI MAP2 TUBB3 GFAP
I
NPC
Astrocytes-I
Astrocytes-II
Glutamatergic Neurons
GABAergic Neurons
Unspecified Neurons
J
AQP4
TNC
GFAP
SLC1A3
NES
SOX2
MKI67
GAD1
GAD2
DLX1
SLC32A1
SLC17A6
RBFOX3
DCX
SNAP25
SOX11
UnspNeu
GluNeu
GABANeu
NPC
Astro-I
Astro-II
Pct. exp.
0
25
50
75
Avg. exp.
2.0
1.5
1.0
0.5
0.0
K
Glutamatergic Neurons
-log10(p adj)
log2(fold change)
GRID1
KCNB2
DLGAP1
SHISA6
PRKCB
CACNA1E
GRIA3
NRXN3
KCNQ3
GRIK3
EPHB1
CHRM3
GRM7
CACNA2D3
L
GO:BP
regulation of trans−synaptic signaling
modulation of chemical synaptic transmission
neurogenesis
synapse organization
generation of neurons
neuron differentiation
chemical synaptic transmission
anterograde trans−synaptic signaling
synaptic signaling
trans−synaptic signaling
−log10(p_value)
GO:MF
extracellular ligand−gated ion channel activity
neurotransmitter receptor activity
ionotropic glutamate receptor activity
ligand−gated ion channel activity
ligand−gated channel activity
ion channel activity
transmitter−gated ion channel activity
transmitter−gated channel activity
gated channel activity
glutamate receptor activity
−log10(p_value)

Figure 6


A
OG BAPTA 488
B
10%
20s
C
WT
KO
t1
t2
D
Calcium transients
****
E
Raising time (a.u.)
**
F
Recovery time (a.u.)
****
WT
KO
G
GFAP
MAP2
PLA (BIN1:Cav1.2)
Merged
KO
WT
H
WT
KO
260 kDa
CaV1.2
40 kDa
β-ACTIN
260 kDa
CaV1.2 + Blocking peptide
I
Cav1.2/β-ACTIN
*
WT
KO
J
Total protein
Genotype
WT
KO
KCl 65nM
CaV1.2
β-ACTIN
Cav1.2/β-ACTIN
K
Endosomal fraction
Genotype
WT
KO
KCl 65nM
CaV1.2
CaV1.3
EEA1
Cav1.2
Cav1.3
EEA1
Optical density (a.u.)
****
KCl 65nM
KO
WT
L
WT
KO
WT + NIF
KO + NIF
M
% of spikes outside bursts
****
*
WT
WT + NIF
KO
KO + NIF
N
Average firing rate (Hz)
WT
WT + NIF
KO
KO + NIF

II.2. Conclusion

BIN1 is mainly express in brain in oligodendrocytes, glutamatergic neurons, microglia and GABAergic neurons[543]. Past studies in rat hippocampal neurons, in mice, or drosophila have shown by over-expressing specific human isoforms, a potential role in endosomal trafficking, neurons excitability and long-term memory, interacting with Tau in both cellular and mouse models[544–546]. Here we showed in human cerebral organoid and bidimensional neuronal models leveraging single cell transcriptomics assay that the main transcriptional alteration of BIN1 deletion targets glutamatergic neurons and are enriched for genes involved in synapse organization and calcium channel related activity. These results support that the main activity of BIN1 occur in this cell type and highlight its putative role in regulating calcium channel related synaptic activity.

Neuronal calcium homeostasis is disturb in aging and is an important early cellular defects in AD, leading to neuronal hyperexcitability, defective long term memory, and ultimately neuronal cell death[603,470,477,478,604,605]. Here, we found that BIN1 deletion was sufficient to alter glutamatergic neurons transcriptional activity similarly than in AD brain and to drive neuronal hyperexcitability and neural network synchronization dysfunction. We also found that BIN1 deletion promotes in these neurons a gene expression signature of sustained electrical activity, suggesting long term impact on neurons function including long term memory. We then further validated BIN1 role in calcium homeostasis observing increase in calcium spike duration in BIN1 deleted neurons, as observed in aging[606,607]. We found then that the role of BIN1 in calcium homeostasis could pass through its ability to regulate membrane expression of the L-Type voltage gated calcium channel (LVGCC) Cav1.2, because found interacting direclty with this channel, and because BIN1 KO neurons have an increase Cav1.2 channel expression while reduced internalization capacity in response to stimulation. Cav1.2 play a key role in neurons excitability and long term memory [603], and its activity-dependant internalization allow calcium homeostasis and regulation of neuronal hyperexcitability[608]. Then, dysregulation of its expression at post-synaptic membrane could explained neuronal hyperexcitability observed in BIN1 KO neurons. Interestingly, expression of LVGCC increases in the aging brain and correlate with a neuronal hyperactivity[607,608]. Then, BIN1 role in AD pathogeneisis could pass through its ability to regulate neuronal excitability and synchronization through the LVGCC Cav1.2 internalization.

Recently, Canter et al have shown that pharmacogenetic inhibition of neuronal hyper-excitability reduced AD pathogenesis in mouse including Abeta deposition[609]. Here, we found that calcium channel blocker nifedipine, a specific antagonist of Cav1.2, was able to partially rescue the neuronal hyperactivity mediated by BIN1 deletion, reducing disorganized spikes, and permitting better neuronal activity synchronization. Together, these results highlight the putative cellular mechanisms associating the BIN1 genetics risk with AD pathogenesis and suggests that blocking Cav1.2 through

pharmacological inhibitor could be a good strategy to fight against the neuronal hyperexcitability mediated by BIN1 defects and/or found in aging.

172

GENERAL DISCUSSION

I. Single-cell genomics to identify putative early molecular and cellular mechanism of ACD development

The use of single-cell genomics in our two models allowed to decipher the influence of heterogeneity on early molecular and cellular disease mechanisms.

In both models, we were able to identify known cellular subpopulation using cell type specific markers, but also to redefine or highlight new heterogeneity within the studied tissues. In HSPCs, we found the main subpopulations, from LT-HSC to lineage-restricted progenitors but also confirm the continuous landscape of the hematopoietic differentiation. To note, even if the clustering algorithm generate discrete subpopulation within HSPCs, the limits between each subpopulation are in fact not clear and rather represent an arbitrary choice to facilitate further studies at subpopulation level. These results support the new model of continuous hematopoietic differentiation build through previous scRNA-seq studies but also raise the intrinsic subjectivity of defining subpopulation in the context of differentiation. This implies that comparing different scRNA-seq studies require first to integrate data and/or to have a same reference. Tools have been designed to facilitate this task, notably used a same scRNA-seq dataset reference to annotate cells from different studies or to integrate both in the same reducted dimensional space[182,189]. Same conclusions can be made with my second model using iPSC derived neural tissues. We identified known brain related cell types but also highlighted the continuous neural progenitor cells (NPCs) differentiation.

Furthermore, scRNA-seq has allowed us to identify new cellular heterogeneity within tissue, notably in HSCs, were transcriptomically distinct cell subpopulations were identified, including one, having STAT1 and IRF1 markers, both known to promote megakaryopoiesis[610] (See supplemental Figure S2 of the first article). These results support a priming of the HSC subpopulation toward megakaryocytes differentiation concordant with previous findings[416,611]. To validate the relevance of this gain in resolution, new sorting strategies could be implemented to isolate these different HSC clusters and study their putative activities or differentiation abilities. This intra-HSC heterogeneity can also rely on the transcriptional plasticity of HSC allowing different cell's fate[417,612,613].

Critically, the single-cell genomics analysis has allowed to decipher in our two models the cell type specific effect of disease epigenetics or genetics factor which would not have been discernable with classical bulk genomics assay. Regarding the first model, we highlighted the HSC specific epigenetics and transcriptional programming associated with excessive fetal growth and highlighted a decrease ability to regulate their activation or response to stimulation. In the second model, we found

that glutamatergic neurons were particularly affected by deletion of AD risk gene BIN1, with dysregulation of calcium related signaling and neuronal excitability. This is a new level of unsupervised discovery allowed by single-cell genomics. More than the unsupervised identification of molecular mechanisms involved in a disease similarly to bulk genomics, single cell genomics allow also the unsupervised identification of the cell types involved.

The cell resolution brings a new level of biological observations. It increases the number of individual biological observation enabling the study of gene co-expression across cells to identify associated biological module or molecular pathway. In the first model, we used SCENIC tools to infer the TFs and related gene regulatory network (GRN) affected by the epigenetics programming, highlighting the EGR1 and KLF2 related GRN being altered in HSC. This GRN inference is based on correlation, so interpretation of causality needs some caution. I used different strategy to ensure the relevance of this GRN, first integrating co-expression analysis with the chromatin information. Second with single-cell multiome ATAC+Gene expression allowing to correlate accessibility of an open-chromatin region containing a particular TF motif with neighbor gene expression. This multi strategy has allowed us to validate the relevance of the infered TF regulations but also to filter out non-robust regulatory link notably for KLF4 regulon identified with SCENIC, which was not validated based on the single-cell multiome analysis. New tools have emerged to leverage the single-cell multiome data, notably SCENIC+[195] which corrects putative bias of the previous version of SCENIC. It uses co-accessible cell type specific genomics region as candidate enhancer region to identify TF motif and subsequent GRN inference. Still, to validate the GRN inference, *in vitro* perturbation experiments are needed. Here, I used a gene silencing method to downregulate expression of TFs and see consequences on predicted downstream target genes. I was able to validate influence of KLF2 on downstream target genes. Nevertheless, such validation is still limited by the ability of cells to be transfected, the availability of efficient siRNAs and by the stress induced by the transfection method. To prevent putative siRNAs related limitations, others methods could be used notably CRISPR based knock out or silencing of candidate genes, associated with scRNA-seq (an approach called perturb-seq[134]). In the second model, rather than using a TF oriented approaches; we leveraged the cellular resolution to identify cell specific transcriptomic signature using CellID[614] and highlighted that BIN1 deletion leads to an increased proportion of glutamatergic neurons enriched for genes signature of sustained neuronal electrical activity, bringing functional insights on the BIN1 role in neuronal activity regulation.

II. Single-cell genomics to decipher influence of DNA methylation alterations on gene expression and cellular plasticity.

Single-cell genomics helped us to decipher cell specific influence of methylation on gene expression. Indeed, we demonstrated the cell type specific correlation between DNA methylation and gene expression changes confirming the importance to study epigenetics mechanism at cellular resolution. Role of DNA methylation in regulating transcription factors occupancy and subsequent gene expression has already been demonstrated in the context of NRF1 TF binding[615], and also recently in the context of hematopoiesis[393]. However, transcriptional response to DNA methylation still appears context specific[411]. Then, to really confirm the causal role of DNA methylation in regulating gene expression, it would be interesting to mimic the DNA methylation changes observed in cells exposed to excessive fetal growth in normally exposed cells. To do that, one interesting approach would be to exposed HSPCs to LGA related condition like high glucose, high IGF1 or high Insulin, with or without genetic deletion of DNMT3A, the *de novo* DNA methylation writer active in HSC[616]. This manipulation will allow us to test if LGA related condition induce stable epigenetics and transcriptomics changes as for HSPCs from LGA neonates (i) and if the deletion of DNMT3A prevent these changes(ii), which would confirm the epigenetics remodeling and impact on gene expression. Another approach could be to specifically change the DNA methylation of some key LGA affected regions using epigenome editing methods like dCas9-Dnmt3a allowing targeted *de novo* DNA methylation[617] and assess the impact on chromatin accessibility and gene expression.

We also linked the epigenetic programming with functional alteration of HSC, most specifically with differentiation ability/cellular plasticity. We found that DNA hypermethylation targets expression of genes (mostly of the EGR1/KLF2 regulatory network) regulating cell growth and/or HSC entry in proliferation/differentiation. The critical role of DNA methylation in regulating HSC differentiation has been recently confirmed in mice models[393]. To confirm the impact of LGA-related DNA methylation change on HSC differentiation, scRNA-seq and *in vitro* or *in vivo* differentiation/expansion assay can be performed on LGA-related cellular models +/-DNMT3A exposed in previous paragraph. Such alteration of HSC differentiation occurs naturally in aging notably through the accumulation of somatic mutation and epigenetics alterations which lead to clonal expansion of defective HSC[394]. These alterations are associated with inflammaging and ACD risk, notably CVD, and recent studies have confirmed direct role of such defective hematopoiesis on ACD development, further highlighting the potential role of LGA associated hematopoiesis alteration in ACD risk[54,56,58,398]. To really challenge this hypothesis, we could conduct an experiment with (immunodeficient) animal model transplanted for LGA-exposed or control HSC, and follow metabolic status in order to assess ACD risk.

An important finding is that the transcriptomics and functional changes do not seems detectable without added environmental challenges. It is still interesting to note that the chromatin accessibility changes observed in LGA samples are not dependent of future exposure. Indeed, chromatin accessibility was assessed thanks to single nucleus ATAC-seq that was independent of scRNA-seq assay and not require sustained exposure of cells in non-physiological environment. This consideration suggests that the decreased chromatin accessibility, as well as the increased DNA methylation are stably altered in LGA, but the transcriptional effect of such epigenetics change occurs only in challenging/stimulating condition. While this environmental challenge was not physiological, equivalent cellular response was observed using cytokines promoting HSC activation. The environment dependent functional alterations is rising interesting considerations. First, it highlights the importance of later environmental exposition in disease onset. Second, it reinforce the need of care when interpreting data as it can be directly impacted by the protocol used to measure them. The dependency of the measured biological element to the measurement/monitor echoes the fundamental problem of the measure in physics but is often neglected in biology and should probably be better considered.

An important remaining question is if the epigenetics programming could be reverse in life. To answer this question, experiments on animal model of excessive fetal growth or transplanted for HSC exposed to LGA-like conditions can be performed with assessment of the epigenetics landscape of HSC at different time points according to different diets or medications. Human longitudinal study could also be performed to assess if LGA HSC epigenetics alteration remain in time but the difficulties/invasiveness to isolate HSCs from bone marrow will certainly ask derivative strategy to follow these epigenetics biomarkers. Use of peripheral blood cells as a proxy of HSPC to assess such epigenetics alteration could be evaluated.

Finally, a parallel question is the relevance of such epigenetics programming in others stem cells compartment, notably mesenchymal stem cells (MSCs). MSCs are mainly located in bone marrow but can be found in nearly all tissues and give rise to numerous cells of the body notably adipocytes which are key players in the context of obesity and related ACD. To decipher if these stem cells are also epigenetically altered in LGA, MSCs can be isolated from umbilical cord[618] and have also the advantage to be easily isolable after birth from diverse tissue including blood, dermis and dental pulp[619], facilitating longitudinal studies.

III. Single-cell genomics to understand impact of genetics risk in heterogeneous and difficult to access tissue

One of the main contribution of single-cell genomics in the AD related study was the ability to decipher cellular and molecular mechanisms in heterogeneous and not accessible tissue like the brain. Indeed, before such assay, study of cellular mechanisms in brain models was restricted to cell imaging using immunophenotypic markers or electrochemical measurements, or cell subpopulation sorting using FACS but this required predefined hypothesis and cellular targets as well as consequent experimental procedures. Here, we were able to unsupervisely assess influence of an AD risk gene in several brain cell types and biological processes in one assay. Still, some challenges remain. First, our cellular models (bidimensional neuronal culture and tridimensional cerebral organoid) were composed only of neural progenitor cells-derived brain cell types, which excludes notably microglia, a cell type expressing BIN1 and known to play an important role in AD neuro-inflammation. Then, to decipher BIN1 role in microglia, it would be interesting to add microglia in our cellular models. To do so, it is possible to reprogrammed iPSC into hematopoetic progenitors and differentiate them into microglia[67]. Another possibility would be to use a conditional model of BIN1 deletion in mice allowing specific loss of function of BIN1 in microglia. Second, even if this study has highlighted BIN1 role in regulating calcium signaling, it cannot confirmed the influence of the AD risk variants on *BIN1*. Indeed, the BIN1 related risk variants do not lead to BIN1 deletion because they are not located in coding region. Some cues indicate that these variants could reduce expression of BIN1 in neurons[550], however little is known about the influence of these variants on gene expression regulation. To decipher the real effect of genetics variants on AD, several strategies could be used: first, assess the impact of AD risk haplotype on brain model using iPSC from careers of theses AD risk variants. It will enable to assess the impact of the whole haplotype associated with AD risk variants on these cells. However, it could not decipher the influence of specific variants. The second strategy could be to genetically modified control iPSC by adding the AD specific risk genetics variant to see if the variant alone can trigger AD related effect. Such studies have been led for APOE risk variants and in this case it appeared that the influence of such genetics risk was dependant of the whole haplotype[67], highlighting the importance of the risk variant interactions with individual-specific (epi)genetics background.

REFERENCES

1. Belikov, A.V. (2019). Age-related diseases as vicious cycles. Ageing Res Rev *49*, 11–26. 10.1016/j.arr.2018.11.002.

2. Afshin, A., Sur, P.J., Fay, K.A., Cornaby, L., Ferrara, G., Salama, J.S., Mullany, E.C., Abate, K.H., Abbafati, C., Abebe, Z., et al. (2019). Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet *393*, 1958–1972. 10.1016/S0140-6736(19)30041-8.

3. Kehler, D.S. (2019). Age-related disease burden as a measure of population ageing. The Lancet Public Health *4*, e123–e124. 10.1016/S2468-2667(19)30026-X.

4. Henning, R.J. (2018). Type-2 diabetes mellitus and cardiovascular disease. Future Cardiology *14*, 491–509. 10.2217/fca-2018-0045.

5. WHO | Diabetes (2013). https://web.archive.org/web/20130826174444/http://www.who.int/mediacentre/factsheets/fs312/en/.

6. Diabetes https://www.who.int/news-room/fact-sheets/detail/diabetes.

7. Bäckman, L., Jones, S., Berger, A.-K., Laukka, E.J., and Small, B.J. (2004). Multiple cognitive deficits during the transition to Alzheimer's disease. J Intern Med *256*, 195–204. 10.1111/j.1365-2796.2004.01386.x.

8. Dementia https://www.who.int/news-room/fact-sheets/detail/dementia.

9. Alzheimer's Disease Fact Sheet National Institute on Aging. https://www.nia.nih.gov/health/alzheimers-disease-fact-sheet.

10. Use of a cDNA microarray to analyse gene expression patterns in human cancer (1996). Nat Genet *14*, 457–460. 10.1038/ng1296-457.

11. Reuter, J.A., Spacek, D.V., and Snyder, M.P. (2015). High-Throughput Sequencing Technologies. Molecular Cell *58*, 586–597. 10.1016/j.molcel.2015.05.004.

12. Verheijen, J., and Sleegers, K. (2018). Understanding Alzheimer Disease at the Interface between Genetics and Transcriptomics. Trends in Genetics *34*, 434–447. 10.1016/j.tig.2018.02.007.

13. Marlow, G., Ellett, S., Ferguson, I.R., Zhu, S., Karunasinghe, N., Jesuthasan, A.C., Han, D.Y., Fraser, A.G., and Ferguson, L.R. (2013). Transcriptomics to study the effect of a Mediterranean-inspired diet on inflammation in Crohn's disease patients. Human Genomics *7*, 24. 10.1186/1479-7364-7-24.

14. Saini, J.S., Corneo, B., Miller, J.D., Kiehl, T.R., Wang, Q., Boles, N.C., Blenkinsop, T.A., Stern, J.H., and Temple, S. (2017). Nicotinamide Ameliorates Disease Phenotypes in a Human iPSC Model of Age-Related Macular Degeneration. Cell Stem Cell *20*, 635-647.e7. 10.1016/j.stem.2016.12.015.

15. Nicolae, D.L., Gamazon, E., Zhang, W., Duan, S., Dolan, M.E., and Cox, N.J. (2010). Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS. PLOS Genetics *6*, e1000888. 10.1371/journal.pgen.1000888.

16. Waddington, C.H. (1956). Genetic Assimilation of the Bithorax Phenotype. Evolution *10*, 1–13. 10.2307/2406091.

17. Tost, J. (2010). DNA methylation: an introduction to the biology and the disease-associated changes of a promising biomarker. Mol Biotechnol *44*, 71–81. 10.1007/s12033-009-9216-2.

18. Choy, M.-K., Movassagh, M., Goh, H.-G., Bennett, M.R., Down, T.A., and Foo, R.S.Y. (2010). Genome-wide conserved consensus transcription factor binding motifs are hyper-methylated. BMC Genomics *11*, 519. 10.1186/1471-2164-11-519.

19. Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes Dev. *25*, 1010–1022. 10.1101/gad.2037511.

20. Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science *328*, 916–919. 10.1126/science.1186366.

21. Lev Maor, G., Yearim, A., and Ast, G. (2015). The alternative role of DNA methylation in splicing regulation. Trends Genet *31*, 274–280. 10.1016/j.tig.2015.03.002.

22. Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., D'Souza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y., et al. (2010). Conserved Role of Intragenic DNA Methylation in Regulating Alternative Promoters. Nature *466*, 253–257. 10.1038/nature09165.

23. Medvedeva, Y.A., Khamis, A.M., Kulakovskiy, I.V., Ba-Alawi, W., Bhuyan, M.S.I., Kawaji, H., Lassmann, T., Harbers, M., Forrest, A.R., Bajic, V.B., et al. (2014). Effects of cytosine methylation on transcription factor binding sites. BMC Genomics *15*, 119. 10.1186/1471-2164-15-119.

24. Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res *21*, 381–395. 10.1038/cr.2011.22.

25. Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences *107*, 21931–21936. 10.1073/pnas.1016071107.

26. Audia, J.E., and Campbell, R.M. (2016). Histone Modifications and Cancer. Cold Spring Harb Perspect Biol *8*, a019521. 10.1101/cshperspect.a019521.

27. Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330. 10.1038/nature14248.

28. Fromm, B., Domanska, D., Høye, E., Ovchinnikov, V., Kang, W., Aparicio-Puerta, E., Johansen, M., Flatmark, K., Mathelier, A., Hovig, E., et al. (2020). MirGeneDB 2.0: the metazoan microRNA complement. Nucleic Acids Res *48*, D132–D141. 10.1093/nar/gkz885.

29. Alles, J., Fehlmann, T., Fischer, U., Backes, C., Galata, V., Minet, M., Hart, M., Abu-Halima, M., Grässer, F.A., Lenhof, H.-P., et al. (2019). An estimate of the total number of true human miRNAs. Nucleic Acids Res *47*, 3353–3364. 10.1093/nar/gkz097.

30. Friedman, R.C., Farh, K.K.-H., Burge, C.B., and Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. Genome Res *19*, 92–105. 10.1101/gr.082701.108.

31. Wutz, A., and Gribnau, J. (2007). X inactivation Xplained. Curr Opin Genet Dev *17*, 387–393. 10.1016/j.gde.2007.08.001.

32. Baker, M. (2010). MicroRNA profiling: separating signal from noise. Nat Methods *7*, 687–692. 10.1038/nmeth0910-687.

33. Bumgarner, R. (2013). Overview of DNA Microarrays: Types, Applications, and Their Future. Current Protocols in Molecular Biology *101*, 22.1.1-22.1.11. 10.1002/0471142727.mb2201s101.

34. Mahajan, A., Taliun, D., Thurner, M., Robertson, N.R., Torres, J.M., Rayner, N.W., Steinthorsdottir, V., Scott, R.A., Grarup, N., Cook, J.P., et al. (2018). Fine-mapping of an expanded set of type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet *50*, 1505–1513. 10.1038/s41588-018-0241-6.

35. Bellenguez, C., Küçükali, F., Jansen, I.E., Kleineidam, L., Moreno-Grau, S., Amin, N., Naj, A.C., Campos-Martin, R., Grenier-Boley, B., Andrade, V., et al. (2022). New insights into the genetic etiology of Alzheimer's disease and related dementias. Nat Genet *54*, 412–436. 10.1038/s41588-022-01024-z.

36. Ritchie, M.D., and Van Steen, K. (2018). The search for gene-gene interactions in genome-wide association studies: challenges in abundance of methods, practical considerations, and biological interpretation. Ann Transl Med *6*, 157. 10.21037/atm.2018.04.05.

37. Kunkle, B.W., Grenier-Boley, B., Sims, R., Bis, J.C., Damotte, V., Naj, A.C., Boland, A., Vronskaya, M., van der Lee, S.J., Amlie-Wolf, A., et al. (2019). Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nat Genet *51*, 414–430. 10.1038/s41588-019-0358-2.

38. Pan, G., King, A., Wu, F., Simpson-Yap, S., Woodhouse, A., Phipps, A., and Vickers, J.C. (2021). The potential roles of genetic factors in predicting ageing-related cognitive change and Alzheimer's disease. Ageing Research Reviews *70*, 101402. 10.1016/j.arr.2021.101402.

39. Jansen, I.E., Savage, J.E., Watanabe, K., Bryois, J., Williams, D.M., Steinberg, S., Sealock, J., Karlsson, I.K., Hägg, S., Athanasiu, L., et al. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nat Genet *51*, 404–413. 10.1038/s41588-018-0311-9.

40. Hidalgo, B., Irvin, M.R., Sha, J., Zhi, D., Aslibekyan, S., Absher, D., Tiwari, H.K., Kabagambe, E.K., Ordovas, J.M., and Arnett, D.K. (2014). Epigenome-Wide Association Study of Fasting Measures of Glucose, Insulin, and HOMA-IR in the Genetics of Lipid Lowering Drugs and Diet Network Study. Diabetes *63*, 801–807. 10.2337/db13-1100.

41. Arpón, A., Milagro, F.I., Ramos-Lopez, O., Mansego, M.L., Santos, J.L., Riezu-Boj, J.-I., and Martínez, J.A. (2019). Epigenome-wide association study in peripheral white blood cells involving insulin resistance. Sci Rep *9*, 2445. 10.1038/s41598-019-38980-2.

42. Xu, K., Zhang, X., Wang, Z., Hu, Y., and Sinha, R. (2018). Epigenome-wide association analysis revealed that SOCS3 methylation influences the effect of cumulative stress on obesity. Biological Psychology *131*, 63–71. 10.1016/j.biopsycho.2016.11.001.

43. Dick, K.J., Nelson, C.P., Tsaprouni, L., Sandling, J.K., Aïssi, D., Wahl, S., Meduri, E., Morange, P.-E., Gagnon, F., Grallert, H., et al. (2014). DNA methylation and body-mass index: a genome-wide analysis. The Lancet *383*, 1990–1998. 10.1016/S0140-6736(13)62674-4.

44. Smith, R.G., Pishva, E., Shireby, G., Smith, A.R., Roubroeks, J.A.Y., Hannon, E., Wheildon, G., Mastroeni, D., Gasparoni, G., Riemenschneider, M., et al. (2021). A meta-analysis of epigenome-wide association studies in Alzheimer's disease highlights novel differentially methylated loci across cortex. Nat Commun *12*, 3517. 10.1038/s41467-021-23243-4.

45. Li, Q.S., Sun, Y., and Wang, T. (2020). Epigenome-wide association study of Alzheimer's disease replicates 22 differentially methylated positions and 30 differentially methylated regions. Clinical Epigenetics *12*, 149. 10.1186/s13148-020-00944-z.

46. Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A *74*, 5463–5467.

47. Shendure, J., and Ji, H. (2008). Next-generation DNA sequencing. Nat Biotechnol *26*, 1135–1145. 10.1038/nbt1486.

48. The Human Genome Project Genome.gov. https://www.genome.gov/human-genome-project.

49. Behjati, S., and Tarpey, P.S. (2013). What is next generation sequencing? Archives of Disease in Childhood - Education and Practice *98*, 236–238. 10.1136/archdischild-2013-304340.

50. Next-Generation Sequencing (NGS) | Explore the technology https://emea.illumina.com/science/technology/next-generation-sequencing.html.

51. Fuchsberger, C., Flannick, J., Teslovich, T.M., Mahajan, A., Agarwala, V., Gaulton, K.J., Ma, C., Fontanillas, P., Moutsianas, L., McCarthy, D.J., et al. (2016). The genetic architecture of type 2 diabetes. Nature *536*, 41–47. 10.1038/nature18642.

52. Prokopenko, D., Lee, S., Hecker, J., Mullin, K., Morgan, S., Katsumata, Y., Weiner, M.W., Fardo, D.W., Laird, N., Bertram, L., et al. (2022). Region-based analysis of rare genomic variants in whole-genome sequencing datasets reveal two novel Alzheimer's disease-associated genes: DTNB and DLG2. Mol Psychiatry *27*, 1963–1969. 10.1038/s41380-022-01475-0.

53. Martincorena, I., and Campbell, P.J. (2015). Somatic mutation in cancer and normal cells. Science *349*, 1483–1489. 10.1126/science.aab4082.

54. Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P.V., Mar, B.G., Lindsley, R.C., Mermel, C.H., Burtt, N., Chavez, A., et al. (2014). Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. New England Journal of Medicine *371*, 2488–2498. 10.1056/NEJMoa1408617.

55. Lodato, M.A., Rodin, R.E., Bohrson, C.L., Coulter, M.E., Barton, A.R., Kwon, M., Sherman, M.A., Vitzthum, C.M., Luquette, L.J., Yandava, C.N., et al. (2018). Aging and neurodegeneration are associated with increased mutations in single human neurons. Science *359*, 555–559. 10.1126/science.aao4426.

56. Jaiswal, S., Natarajan, P., Silver, A.J., Gibson, C.J., Bick, A.G., Shvartz, E., McConkey, M., Gupta, N., Gabriel, S., Ardissino, D., et al. (2017). Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. New England Journal of Medicine *377*, 111–121. 10.1056/NEJMoa1701719.

57. Bonnefond, A., Skrobek, B., Lobbens, S., Eury, E., Thuillier, D., Cauchi, S., Lantieri, O., Balkau, B., Riboli, E., Marre, M., et al. (2013). Association between large detectable clonal mosaicism and type 2 diabetes with vascular complications. Nat Genet *45*, 1040–1043. 10.1038/ng.2700.

58. Fuster, J.J., Zuriaga, M.A., Zorita, V., MacLauchlan, S., Polackal, M.N., Viana-Huete, V., Ferrer-Pérez, A., Matesanz, N., Herrero-Cervera, A., Sano, S., et al. (2020). TET2-Loss-of-Function-Driven Clonal Hematopoiesis Exacerbates Experimental Insulin Resistance in Aging and Obesity. Cell Reports *33*, 108326. 10.1016/j.celrep.2020.108326.

59. Park, J.S., Lee, J., Jung, E.S., Kim, M.-H., Kim, I.B., Son, H., Kim, S., Kim, S., Park, Y.M., Mook-Jung, I., et al. (2019). Brain somatic mutations observed in Alzheimer's disease associated with aging and dysregulation of tau phosphorylation. Nat Commun *10*, 3090. 10.1038/s41467-019-11000-7.

60. Söllner, J.F., Leparc, G., Hildebrandt, T., Klein, H., Thomas, L., Stupka, E., and Simon, E. (2017). An RNA-Seq atlas of gene expression in mouse and rat normal tissues. Sci Data *4*, 170185. 10.1038/sdata.2017.185.

61. O'Rourke, J.A., Iniguez, L.P., Fu, F., Bucciarelli, B., Miller, S.S., Jackson, S.A., McClean, P.E., Li, J., Dai, X., Zhao, P.X., et al. (2014). An RNA-Seq based gene expression atlas of the common bean. BMC Genomics *15*, 866. 10.1186/1471-2164-15-866.

62. Leader, D.P., Krause, S.A., Pandit, A., Davies, S.A., and Dow, J.A.T. (2018). FlyAtlas 2: a new version of the Drosophila melanogaster expression atlas with RNA-Seq, miRNA-Seq and sex-specific data. Nucleic Acids Research *46*, D809–D815. 10.1093/nar/gkx976.

63. GTEx Portal https://gtexportal.org/home/.

64. The Human Protein Atlas https://www.proteinatlas.org/.

65. Karlsson, M., Zhang, C., Méar, L., Zhong, W., Digre, A., Katona, B., Sjöstedt, E., Butler, L., Odeberg, J., Dusart, P., et al. (2021). A single–cell type transcriptomics map of human tissues. Science Advances *7*, eabh2169. 10.1126/sciadv.abh2169.

66. Schwartzentruber, J., Cooper, S., Liu, J.Z., Barrio-Hernandez, I., Bello, E., Kumasaka, N., Young, A.M.H., Franklin, R.J.M., Johnson, T., Estrada, K., et al. (2021). Genome-wide meta-analysis, fine-mapping, and integrative prioritization implicate new Alzheimer's disease risk genes. Nat Genet *53*, 392–402. 10.1038/s41588-020-00776-w.

67. Tcw, J., Qian, L., Pipalia, N.H., Chao, M.J., Liang, S.A., Shi, Y., Jain, B.R., Bertelsen, S.E., Kapoor, M., Marcora, E., et al. (2022). Cholesterol and matrisome pathways dysregulated in astrocytes and microglia. Cell *185*, 2213-2233.e25. 10.1016/j.cell.2022.05.017.

68. Keenan, A.B., Jenkins, S.L., Jagodnik, K.M., Koplev, S., He, E., Torre, D., Wang, Z., Dohlman, A.B., Silverstein, M.C., Lachmann, A., et al. (2018). The Library of Integrated Network-Based Cellular Signatures NIH Program: System-Level Cataloging of Human Cells Response to Perturbations. Cell Systems *6*, 13–24. 10.1016/j.cels.2017.11.001.

69. Xia, F., Allen, J., Balaprakash, P., Brettin, T., Garcia-Cardona, C., Clyde, A., Cohn, J., Doroshow, J., Duan, X., Dubinkina, V., et al. (2022). A cross-study analysis of drug response prediction in cancer cell lines. Briefings in Bioinformatics *23*, bbab356. 10.1093/bib/bbab356.

70. Wang, M., Li, A., Sekiya, M., Beckmann, N.D., Quan, X., Schrode, N., Fernando, M.B., Yu, A., Zhu, L., Cao, J., et al. (2021). Transformative Network Modeling of Multi-omics Data Reveals Detailed Circuits, Key Regulators, and Potential Therapeutics for Alzheimer's Disease. Neuron *109*, 257-272.e14. 10.1016/j.neuron.2020.11.002.

71. Cogswell, J.P., Ward, J., Taylor, I.A., Waters, M., Shi, Y., Cannon, B., Kelnar, K., Kemppainen, J., Brown, D., Chen, C., et al. (2008). Identification of miRNA Changes in Alzheimer's Disease Brain and CSF Yields Putative Biomarkers and Insights into Disease Pathways. Journal of Alzheimer's Disease *14*, 27–41. 10.3233/JAD-2008-14103.

72. Calin, G.A., and Croce, C.M. (2006). MicroRNA signatures in human cancers. Nat Rev Cancer *6*, 857–866. 10.1038/nrc1997.

73. Cheng, G. (2015). Circulating miRNAs: Roles in cancer diagnosis, prognosis and therapy. Advanced Drug Delivery Reviews *81*, 75–93. 10.1016/j.addr.2014.09.001.

74. Ziller, M.J., Hansen, K.D., Meissner, A., and Aryee, M.J. (2015). Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. Nat Methods *12*, 230–232. 10.1038/nmeth.3152.

75. Takeshima, H., and Ushijima, T. (2019). Accumulation of genetic and epigenetic alterations in normal cells and cancer risk. npj Precis. Onc. *3*, 1–8. 10.1038/s41698-019-0079-0.

76. Hlady, R.A., Tiedemann, R.L., Puszyk, W., Zendejas, I., Roberts, L.R., Choi, J.-H., Liu, C., and Robertson, K.D. (2014). Epigenetic signatures of alcohol abuse and hepatitis infection during human hepatocarcinogenesis. Oncotarget *5*, 9425–9443. 10.18632/oncotarget.2444.

77. Lambert, M.-P., Paliwal, A., Vaissière, T., Chemin, I., Zoulim, F., Tommasino, M., Hainaut, P., Sylla, B., Scoazec, J.-Y., Tost, J., et al. (2011). Aberrant DNA methylation distinguishes hepatocellular carcinoma associated with HBV and HCV infection and alcohol intake. J Hepatol *54*, 705–715. 10.1016/j.jhep.2010.07.027.

78. Takeshima, H., Niwa, T., Toyoda, T., Wakabayashi, M., Yamashita, S., and Ushijima, T. (2017). Degree of methylation burden is determined by the exposure period to carcinogenic factors. Cancer Sci *108*, 316–321. 10.1111/cas.13136.

79. Yuan, W., Xia, Y., Bell, C.G., Yet, I., Ferreira, T., Ward, K.J., Gao, F., Loomis, A.K., Hyde, C.L., Wu, H., et al. (2014). An integrated epigenomic analysis for type 2 diabetes susceptibility loci in monozygotic twins. Nat Commun *5*, 5719. 10.1038/ncomms6719.

80. Lamoureux, F., Baud'huin, M., Rodriguez Calleja, L., Jacques, C., Berreur, M., Rédini, F., Lecanda, F., Bradner, J.E., Heymann, D., and Ory, B. (2014). Selective inhibition of BET bromodomain epigenetic signalling interferes with the bone-associated tumour vicious cycle. Nat Commun *5*, 3511. 10.1038/ncomms4511.

81. Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective Inhibition of Tumor Oncogenes by Disruption of Super-Enhancers. Cell *153*, 320–334. 10.1016/j.cell.2013.03.036.

82. Nativio, R., Lan, Y., Donahue, G., Sidoli, S., Berson, A., Srinivasan, A.R., Shcherbakova, O., Amlie-Wolf, A., Nie, J., Cui, X., et al. (2020). An integrated multi-omics approach identifies epigenetic alterations associated with Alzheimer's disease. Nat Genet *52*, 1024–1035. 10.1038/s41588-020-0696-0.

83. Lee, M.Y., Lee, J., Hyeon, S.J., Cho, H., Hwang, Y.J., Shin, J.-Y., McKee, A.C., Kowall, N.W., Kim, J.-I., Stein, T.D., et al. (2020). Epigenome signatures landscaped by histone H3K9me3 are associated with the synaptic dysfunction in Alzheimer's disease. Aging Cell *19*, e13153. 10.1111/acel.13153.

84. Pirola, L., Balcerczyk, A., Tothill, R.W., Haviv, I., Kaspi, A., Lunke, S., Ziemann, M., Karagiannis, T., Tonna, S., Kowalczyk, A., et al. (2011). Genome-wide analysis distinguishes hyperglycemia regulated epigenetic signatures of primary vascular cells. Genome Res. *21*, 1601–1615. 10.1101/gr.116095.110.

85. Brasacchio, D., Okabe, J., Tikellis, C., Balcerczyk, A., George, P., Baker, E.K., Calkin, A.C., Brownlee, M., Cooper, M.E., and El-Osta, A. (2009). Hyperglycemia Induces a Dynamic Cooperativity of Histone Methylase and Demethylase Enzymes Associated With Gene-Activating Epigenetic Marks That Coexist on the Lysine Tail. Diabetes *58*, 1229–1236. 10.2337/db08-1666.

86. Varshney, A., Scott, L.J., Welch, R.P., Erdos, M.R., Chines, P.S., Narisu, N., Albanus, R.D., Orchard, P., Wolford, B.N., Kursawe, R., et al. (2017). Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. Proceedings of the National Academy of Sciences *114*, 2301–2306. 10.1073/pnas.1621192114.

87. Gibbs, J.R., Brug, M.P. van der, Hernandez, D.G., Traynor, B.J., Nalls, M.A., Lai, S.-L., Arepalli, S., Dillman, A., Rafferty, I.P., Troncoso, J., et al. (2010). Abundant Quantitative Trait Loci Exist for DNA Methylation and Gene Expression in Human Brain. PLOS Genetics *6*, e1000952. 10.1371/journal.pgen.1000952.

88. Huang, S., Chaudhary, K., and Garmire, L.X. (2017). More Is Better: Recent Progress in Multi-Omics Data Integration Methods. Frontiers in Genetics *8*.

89. Yang, B.T., Dayeh, T.A., Volkov, P.A., Kirkpatrick, C.L., Malmgren, S., Jing, X., Renström, E., Wollheim, C.B., Nitert, M.D., and Ling, C. (2012). Increased DNA Methylation and Decreased Expression of PDX-1 in Pancreatic Islets from Patients with Type 2 Diabetes. Molecular Endocrinology *26*, 1203–1212. 10.1210/me.2012-1004.

90. Semick, S.A., Bharadwaj, R.A., Collado-Torres, L., Tao, R., Shin, J.H., Deep-Soboslay, A., Weiss, J.R., Weinberger, D.R., Hyde, T.M., Kleinman, J.E., et al. (2019). Integrated DNA methylation and gene expression profiling across multiple brain regions implicate novel genes in Alzheimer's disease. Acta Neuropathol *137*, 557–569. 10.1007/s00401-019-01966-5.

91. Wang, T.-H., Lee, C.-Y., Lee, T.-Y., Huang, H.-D., Hsu, J.B.-K., and Chang, T.-H. (2021). Biomarker Identification through Multiomics Data Analysis of Prostate Cancer Prognostication Using a Deep Learning Model and Similarity Network Fusion. Cancers *13*, 2528. 10.3390/cancers13112528.

92. Narayana, J.K., Aogáin, M.M., Ali, N.A.B.M., Tsaneva-Atanasova, K., and Chotirmall, S.H. (2021). Similarity network fusion for the integration of multi-omics and microbiomes in respiratory disease. European Respiratory Journal *58*. 10.1183/13993003.01016-2021.

93. Chierici, M., Bussola, N., Marcolini, A., Francescatto, M., Zandonà, A., Trastulla, L., Agostinelli, C., Jurman, G., and Furlanello, C. (2020). Integrative Network Fusion: A Multi-Omics Approach in Molecular Profiling. Frontiers in Oncology *10*.

94. Ma, T., and Zhang, A. (2017). Integrate multi-omic data using affinity network fusion (ANF) for cancer patient clustering. In 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 398–403. 10.1109/BIBM.2017.8217682.

95. Clark, C., Dayon, L., Masoodi, M., Bowman, G.L., and Popp, J. (2021). An integrative multi-omics approach reveals new central nervous system pathway alterations in Alzheimer's disease. Alzheimer's Research & Therapy *13*, 71. 10.1186/s13195-021-00814-7.

96. Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. Nat Commun *9*, 1911. 10.1038/s41467-018-04252-2.

97. Duan, C.-L., Liu, C.-W., Shen, S.-W., Yu, Z., Mo, J.-L., Chen, X.-H., and Sun, F.-Y. (2015). Striatal astrocytes transdifferentiate into functional mature neurons following ischemic brain injury. Glia *63*, 1660–1670. 10.1002/glia.22837.

98. Shen, S.-W., Duan, C.-L., Chen, X.-H., Wang, Y.-Q., Sun, X., Zhang, Q.-W., Cui, H.-R., and Sun, F.-Y. (2016). Neurogenic effect of VEGF is related to increase of astrocytes transdifferentiation into new mature neurons in rat brains after stroke. Neuropharmacology *108*, 451–461. 10.1016/j.neuropharm.2015.11.012.

99. Barbatelli, G., Murano, I., Madsen, L., Hao, Q., Jimenez, M., Kristiansen, K., Giacobino, J.P., De Matteis, R., and Cinti, S. (2010). The emergence of cold-induced brown adipocytes in mouse white fat depots is determined predominantly by white to brown adipocyte transdifferentiation. American Journal of Physiology-Endocrinology and Metabolism *298*, E1244–E1253. 10.1152/ajpendo.00600.2009.

100. Yoon, K.H., Ko, S.H., Cho, J.H., Lee, J.M., Ahn, Y.B., Song, K.H., Yoo, S.J., Kang, M.I., Cha, B.Y., Lee, K.W., et al. (2003). Selective β-Cell Loss and α-Cell Expansion in Patients with Type 2 Diabetes Mellitus in Korea. The Journal of Clinical Endocrinology & Metabolism *88*, 2300–2308. 10.1210/jc.2002-020735.

101. Guardado-Mendoza, R., Davalli, A.M., Chavez, A.O., Hubbard, G.B., Dick, E.J., Majluf-Cruz, A., Tene-Perez, C.E., Goldschmidt, L., Hart, J., Perego, C., et al. (2009). Pancreatic islet amyloidosis, β-cell apoptosis, and α-cell proliferation are determinants of islet remodeling in type-2 diabetic baboons. Proc. Natl. Acad. Sci. U.S.A. *106*, 13992–13997. 10.1073/pnas.0906471106.

102. Henquin, J.C., and Rahier, J. (2011). Pancreatic alpha cell mass in European subjects with type 2 diabetes. Diabetologia *54*, 1720–1725. 10.1007/s00125-011-2118-4.

103. Vázquez-Vela, M.E.F., Torres, N., and Tovar, A.R. (2008). White Adipose Tissue as Endocrine Organ and Its Role in Obesity. Archives of Medical Research *39*, 715–728. 10.1016/j.arcmed.2008.09.005.

104. Esteve Ràfols, M. (2014). Adipose tissue: Cell heterogeneity and functional diversity. Endocrinología y Nutrición (English Edition) *61*, 100–112. 10.1016/j.endoen.2014.02.001.

105. Keren-Shaul, H., Spinrad, A., Weiner, A., Matcovitch-Natan, O., Dvir-Szternfeld, R., Ulland, T.K., David, E., Baruch, K., Lara-Astaiso, D., Toth, B., et al. (2017). A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. Cell *169*, 1276-1290.e17. 10.1016/j.cell.2017.05.018.

106. Tabas, I. (2010). Macrophage death and defective inflammation resolution in atherosclerosis. Nat Rev Immunol *10*, 36–46. 10.1038/nri2675.

107. Nagele, R.G., Wegiel, J., Venkataraman, V., Imaki, H., Wang, K.-C., and Wegiel, J. (2004). Contribution of glial cells to the development of amyloid plaques in Alzheimer's disease. Neurobiology of Aging *25*, 663–674. 10.1016/j.neurobiolaging.2004.01.007.

108. Ozben, T., and Ozben, S. (2019). Neuro-inflammation and anti-inflammatory treatment options for Alzheimer's disease. Clinical Biochemistry *72*, 87–89. 10.1016/j.clinbiochem.2019.04.001.

109. Deczkowska, A., Keren-Shaul, H., Weiner, A., Colonna, M., Schwartz, M., and Amit, I. (2018). Disease-Associated Microglia: A Universal Immune Sensor of Neurodegeneration. Cell *173*, 1073–1081. 10.1016/j.cell.2018.05.003.

110. Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods *6*, 377–382. 10.1038/nmeth.1315.

111. Ramsköld, D., Luo, S., Wang, Y.-C., Li, R., Deng, Q., Faridani, O.R., Daniels, G.A., Khrebtukova, I., Loring, J.F., Laurent, L.C., et al. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol *30*, 777–782. 10.1038/nbt.2282.

112. Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat Methods *10*, 1096–1098. 10.1038/nmeth.2639.

113. Ludwig, L.S., Lareau, C.A., Ulirsch, J.C., Christian, E., Muus, C., Li, L.H., Pelka, K., Ge, W., Oren, Y., Brack, A., et al. (2019). Lineage Tracing in Humans Enabled by Mitochondrial Mutations and Single-Cell Genomics. Cell *176*, 1325-1339.e22. 10.1016/j.cell.2019.01.022.

114. Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202–1214. 10.1016/j.cell.2015.05.002.

115. Griffiths, J.A., Scialdone, A., and Marioni, J.C. (2018). Using single-cell genomics to understand developmental processes and cell fate decisions. Molecular Systems Biology *14*, e8046. 10.15252/msb.20178046.

116. Reyfman, P.A., Walter, J.M., Joshi, N., Anekalla, K.R., McQuattie-Pimentel, A.C., Chiu, S., Fernandez, R., Akbarpour, M., Chen, C.-I., Ren, Z., et al. (2019). Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis. Am J Respir Crit Care Med *199*, 1517–1536. 10.1164/rccm.201712-2410OC.

117. Velmeshev, D., Schirmer, L., Jung, D., Haeussler, M., Perez, Y., Mayer, S., Bhaduri, A., Goyal, N., Rowitch, D.H., and Kriegstein, A.R. (2019). Single-cell genomics identifies cell type–specific molecular changes in autism. Science *364*, 685–689. 10.1126/science.aav8130.

118. Rozenblatt-Rosen, O., Shin, J.W., Rood, J.E., Hupalowska, A., Regev, A., and Heyn, H. (2021). Building a high-quality Human Cell Atlas. Nat Biotechnol *39*, 149–153. 10.1038/s41587-020-00812-4.

119. Wang, Y.J., Schug, J., Won, K.-J., Liu, C., Naji, A., Avrahami, D., Golson, M.L., and Kaestner, K.H. (2016). Single-Cell Transcriptomics of the Human Endocrine Pancreas. Diabetes *65*, 3028–3038. 10.2337/db16-0405.

120. Chen, R., Wu, X., Jiang, L., and Zhang, Y. (2017). Single-Cell RNA-Seq Reveals Hypothalamic Cell Diversity. Cell Reports *18*, 3227–3241. 10.1016/j.celrep.2017.03.004.

121. Rosenberg, A.B., Roco, C.M., Muscat, R.A., Kuchina, A., Sample, P., Yao, Z., Graybuck, L.T., Peeler, D.J., Mukherjee, S., Chen, W., et al. (2018). Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science *360*, 176–182. 10.1126/science.aam8999.

122. Watcham, S., Kucinski, I., and Gottgens, B. (2019). New insights into hematopoietic differentiation landscapes from single-cell RNA sequencing. Blood *133*, 1415–1426. 10.1182/blood-2018-08-835355.

123. Velten, L., Haas, S.F., Raffel, S., Blaszkiewicz, S., Islam, S., Hennig, B.P., Hirche, C., Lutz, C., Buss, E.C., Nowak, D., et al. (2017). Human haematopoietic stem cell lineage commitment is a continuous process. Nat. Cell Biol. *19*, 271–281. 10.1038/ncb3493.

124. MacArthur, B.D., Please, C.P., and Oreffo, R.O.C. (2008). Stochasticity and the Molecular Mechanisms of Induced Pluripotency. PLOS ONE *3*, e3086. 10.1371/journal.pone.0003086.

125. Smith, Q., Stukalin, E., Kusuma, S., Gerecht, S., and Sun, S.X. (2015). Stochasticity and Spatial Interaction Govern Stem Cell Differentiation Dynamics. Sci Rep *5*, 12617. 10.1038/srep12617.

126. Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature *566*, 496–502. 10.1038/s41586-019-0969-x.

127. Wilk, A.J., Rustagi, A., Zhao, N.Q., Roque, J., Martínez-Colón, G.J., McKechnie, J.L., Ivison, G.T., Ranganath, T., Vergara, R., Hollis, T., et al. (2020). A single-cell atlas of the peripheral immune response in patients with severe COVID-19. Nat Med *26*, 1070–1076. 10.1038/s41591-020-0944-y.

128. Yang, A.C., Kern, F., Losada, P.M., Agam, M.R., Maat, C.A., Schmartz, G.P., Fehlmann, T., Stein, J.A., Schaum, N., Lee, D.P., et al. (2021). Dysregulation of brain and choroid plexus cell types in severe COVID-19. Nature *595*, 565–571. 10.1038/s41586-021-03710-0.

129. Ma, W.F., Hodonsky, C.J., Turner, A.W., Wong, D., Song, Y., Mosquera, J.V., Ligay, A.V., Slenders, L., Gancayco, C., Pan, H., et al. (2022). Enhanced single-cell RNA-seq workflow reveals coronary artery disease cellular cross-talk and candidate drug targets. Atherosclerosis *340*, 12–22. 10.1016/j.atherosclerosis.2021.11.025.

130. Xiao, Y., Hu, X., Fan, S., Zhong, J., Mo, X., Liu, X., and Hu, Y. (2021). Single-Cell Transcriptome Profiling Reveals the Suppressive Role of Retinal Neurons in Microglia Activation Under Diabetes Mellitus. Frontiers in Cell and Developmental Biology *9*.

131. Yang, A.C., Vest, R.T., Kern, F., Lee, D.P., Agam, M., Maat, C.A., Losada, P.M., Chen, M.B., Schaum, N., Khoury, N., et al. (2022). A human brain vascular atlas reveals diverse mediators of Alzheimer's risk. Nature *603*, 885–892. 10.1038/s41586-021-04369-3.

132. Son, J., Ding, H., Farb, T.B., Efanov, A.M., Sun, J., Gore, J.L., Syed, S.K., Lei, Z., Wang, Q., Accili, D., et al. (2021). BACH2 inhibition reverses **β** cell failure in type 2 diabetes models. J Clin Invest *131*. 10.1172/JCI153876.

133. Wang, C., Xiong, M., Gratuze, M., Bao, X., Shi, Y., Andhey, P.S., Manis, M., Schroeder, C., Yin, Z., Madore, C., et al. (2021). Selective removal of astrocytic APOE4 strongly protects against tau-mediated neurodegeneration and decreases synaptic phagocytosis by microglia. Neuron *109*, 1657-1674.e7. 10.1016/j.neuron.2021.03.024.

134. Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C.P., Jerby-Arnon, L., Marjanovic, N.D., Dionne, D., Burks, T., Raychowdhury, R., et al. (2016). Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell *167*, 1853-1866.e17. 10.1016/j.cell.2016.11.038.

135. Ursu, O., Neal, J.T., Shea, E., Thakore, P.I., Jerby-Arnon, L., Nguyen, L., Dionne, D., Diaz, C., Bauman, J., Mosaad, M.M., et al. (2022). Massively parallel phenotyping of coding variants in cancer with Perturb-seq. Nat Biotechnol *40*, 896–905. 10.1038/s41587-021-01160-7.

136. Marquina-Sanchez, B., Fortelny, N., Farlik, M., Vieira, A., Collombat, P., Bock, C., and Kubicek, S. (2020). Single-cell RNA-seq with spike-in cells enables accurate quantification of cell-specific drug effects in pancreatic islets. Genome Biol *21*, 106. 10.1186/s13059-020-02006-2.

137. Avey, D., Sankararaman, S., Yim, A.K.Y., Barve, R., Milbrandt, J., and Mitra, R.D. (2018). Single-Cell RNA-Seq Uncovers a Robust Transcriptional Response to Morphine by Glia. Cell Reports *24*, 3619-3629.e4. 10.1016/j.celrep.2018.08.080.

138. Philpott, M., Cribbs, A.P., Brown, T., Brown, T., and Oppermann, U. (2020). Advances and challenges in epigenomic single-cell sequencing applications. Current Opinion in Chemical Biology *57*, 17–26. 10.1016/j.cbpa.2020.01.013.

139. Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature *523*, 486–490. 10.1038/nature14590.

140. Klemm, S.L., Shipony, Z., and Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. Nat Rev Genet *20*, 207–220. 10.1038/s41576-018-0089-8.

141. Chiou, J., Zeng, C., Cheng, Z., Han, J.Y., Schlichting, M., Miller, M., Mendez, R., Huang, S., Wang, J., Sui, Y., et al. (2021). Single-cell chromatin accessibility identifies pancreatic islet cell type– and state-specific regulatory programs of diabetes risk. Nat Genet *53*, 455–466. 10.1038/s41588-021-00823-0.

142. Yuan, Q., and Duren, Z. (2022). Integration of single-cell multi-omics data by regression analysis on unpaired observations. Genome Biol *23*, 160. 10.1186/s13059-022-02726-7.

143. Novikova, G., Andrews, S.J., Renton, A.E., and Marcora, E. (2021). Beyond association: successes and challenges in linking non-coding genetic variation to functional consequences that modulate Alzheimer's disease risk. Mol Neurodegeneration *16*, 27. 10.1186/s13024-021-00449-0.

144. Pasquali, L., Gaulton, K.J., Rodríguez-Seguí, S.A., Mularoni, L., Miguel-Escalada, I., Akerman, İ., Tena, J.J., Morán, I., Gómez-Marín, C., van de Bunt, M., et al. (2014). Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. Nat Genet *46*, 136–143. 10.1038/ng.2870.

145. Corces, M.R., Shcherbina, A., Kundu, S., Gloudemans, M.J., Frésard, L., Granja, J.M., Louie, B.H., Eulalio, T., Shams, S., Bagdatli, S.T., et al. (2020). Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. Nature Genetics *52*, 1158–1168. 10.1038/s41588-020-00721-x.

146. Swarup, V., Morabito, S., Miyoshi, E., Michael, N., Shahin, S., Martini, A.C., Head, E., Silva, J., Leavy, K., and Perez-Rosendahl, M. (2021). Single-cell multi-omics analysis identifies dynamic

regulation of SREBF1 in Alzheimer's disease. Alzheimer's & Dementia *17*, e049956. 10.1002/alz.049956.

147. Armaka, M., Konstantopoulos, D., Tzaferis, C., Lavigne, M.D., Sakkou, M., Liakos, A., Sfikakis, P.P., Dimopoulos, M.A., Fousteri, M., and Kollias, G. (2022). Single-cell multimodal analysis identifies common regulatory programs in synovial fibroblasts of rheumatoid arthritis patients and modeled TNF-driven arthritis. Genome Med *14*, 78. 10.1186/s13073-022-01081-3.

148. Ma, S., Zhang, B., LaFave, L.M., Earl, A.S., Chiang, Z., Hu, Y., Ding, J., Brack, A., Kartha, V.K., Tay, T., et al. (2020). Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. Cell *183*, 1103-1116.e20. 10.1016/j.cell.2020.09.056.

149. Huang, L., Ma, F., Chapman, A., Lu, S., and Xie, X.S. (2015). Single-Cell Whole-Genome Amplification and Sequencing: Methodology and Applications. Annual Review of Genomics and Human Genetics *16*, 79–102. 10.1146/annurev-genom-090413-025352.

150. Single Cell Immune Profiling 10x Genomics. https://www.10xgenomics.com/products/single-cell-immune-profiling.

151. Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G., Ahmad, K., and Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun *10*, 1930. 10.1038/s41467-019-09982-5.

152. Bartosovic, M., Kabbe, M., and Castelo-Branco, G. (2021). Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. Nat Biotechnol *39*, 825–835. 10.1038/s41587-021-00869-9.

153. Luo, C., Rivkin, A., Zhou, J., Sandoval, J.P., Kurihara, L., Lucero, J., Castanon, R., Nery, J.R., Pinto-Duarte, A., Bui, B., et al. (2018). Robust single-cell DNA methylome profiling with snmC-seq2. Nat Commun *9*, 3824. 10.1038/s41467-018-06355-2.

154. Clark, S.J., Argelaguet, R., Kapourani, C.-A., Stubbs, T.M., Lee, H.J., Alda-Catalinas, C., Krueger, F., Sanguinetti, G., Kelsey, G., Marioni, J.C., et al. (2018). scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. Nat Commun *9*, 781. 10.1038/s41467-018-03149-4.

155. Ahn, J., Heo, S., Lee, J., and Bang, D. (2021). Introduction to Single-Cell DNA Methylation Profiling Methods. Biomolecules *11*, 1013. 10.3390/biom11071013.

156. Argelaguet, R., Clark, S.J., Mohammed, H., Stapel, L.C., Krueger, C., Kapourani, C.-A., Imaz-Rosshandler, I., Lohoff, T., Xiang, Y., Hanna, C.W., et al. (2019). Multi-omics profiling of mouse gastrulation at single-cell resolution. Nature, 1–5. 10.1038/s41586-019-1825-8.

157. Kelly, R.T. (2020). Single-cell Proteomics: Progress and Prospects. Molecular & Cellular Proteomics *19*, 1739–1748. 10.1074/mcp.R120.002234.

158. D. Duncan, K., Fyrestam, J., and Lanekoff, I. (2019). Advances in mass spectrometry based single-cell metabolomics. Analyst *144*, 782–793. 10.1039/C8AN01581C.

159. Burgess, D.J. (2019). Spatial transcriptomics coming of age. Nat Rev Genet *20*, 317–317. 10.1038/s41576-019-0129-z.

160. Crosetto, N., Bienko, M., and van Oudenaarden, A. (2015). Spatially resolved transcriptomics and beyond. Nat Rev Genet *16*, 57–66. 10.1038/nrg3832.

161. Lee, J.H., Daugharthy, E.R., Scheiman, J., Kalhor, R., Ferrante, T.C., Yang, J.L., Terry, R., Jeanty, S.S.F., Li, C., Amamoto, R., et al. (2014). Highly multiplexed subcellular RNA sequencing in situ. Science *343*, 1360–1363. 10.1126/science.1250212.

162. Xia, C., Fan, J., Emanuel, G., Hao, J., and Zhuang, X. (2019). Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. Proc Natl Acad Sci U S A *116*, 19490–19499. 10.1073/pnas.1912459116.

163. Rao, A., Barkley, D., França, G.S., and Yanai, I. (2021). Exploring tissue architecture using spatial transcriptomics. Nature *596*, 211–220. 10.1038/s41586-021-03634-9.

164. Cho, C.-S., Xi, J., Si, Y., Park, S.-R., Hsu, J.-E., Kim, M., Jun, G., Kang, H.M., and Lee, J.H. (2021). Microscopic examination of spatial transcriptome using Seq-Scope. Cell *184*, 3559-3572.e22. 10.1016/j.cell.2021.05.010.

165. Vickovic, S., Eraslan, G., Salmén, F., Klughammer, J., Stenbeck, L., Schapiro, D., Äijö, T., Bonneau, R., Bergenstråhle, L., Navarro, J.F., et al. (2019). High-definition spatial transcriptomics for in situ tissue profiling. Nat Methods *16*, 987–990. 10.1038/s41592-019-0548-y.

166. Spatial Transcriptomics - 10x Genomics https://www.10xgenomics.com/spatial-transcriptomics.

167. Wang, X., Allen, W.E., Wright, M.A., Sylwestrak, E.L., Samusik, N., Vesuna, S., Evans, K., Liu, C., Ramakrishnan, C., Liu, J., et al. (2018). Three-dimensional intact-tissue sequencing of single-cell transcriptional states. Science *361*, eaat5691. 10.1126/science.aat5691.

168. Callaway, E.M., Dong, H.-W., Ecker, J.R., Hawrylycz, M.J., Huang, Z.J., Lein, E.S., Ngai, J., Osten, P., Ren, B., Tolias, A.S., et al. (2021). A multimodal cell census and atlas of the mammalian primary motor cortex. Nature *598*, 86–102. 10.1038/s41586-021-03950-0.

169. Gyllborg, D., Langseth, C.M., Qian, X., Choi, E., Salas, S.M., Hilscher, M.M., Lein, E.S., and Nilsson, M. (2020). Hybridization-based in situ sequencing (HybISS) for spatially resolved transcriptomics in human and mouse brain tissue. Nucleic Acids Res *48*, e112. 10.1093/nar/gkaa792.

170. Moffitt, J.R., Bambah-Mukku, D., Eichhorn, S.W., Vaughn, E., Shekhar, K., Perez, J.D., Rubinstein, N.D., Hao, J., Regev, A., Dulac, C., et al. (2018). Molecular, Spatial and Functional Single-Cell Profiling of the Hypothalamic Preoptic Region. Science *362*, eaau5324. 10.1126/science.aau5324.

171. Alon, S., Goodwin, D.R., Sinha, A., Wassie, A.T., Chen, F., Daugharthy, E.R., Bando, Y., Kajita, A., Xue, A.G., Marrett, K., et al. (2021). Expansion Sequencing: Spatially Precise In Situ Transcriptomics in Intact Biological Systems. Science *371*, eaax2656. 10.1126/science.aax2656.

172. Maynard, K.R., Collado-Torres, L., Weber, L.M., Uytingco, C., Barry, B.K., Williams, S.R., Catallini, J.L., Tran, M.N., Besich, Z., Tippani, M., et al. (2021). Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. Nat Neurosci *24*, 425–436. 10.1038/s41593-020-00787-0.

173. Asp, M., Giacomello, S., Larsson, L., Wu, C., Fürth, D., Qian, X., Wärdell, E., Custodio, J., Reimegård, J., Salmén, F., et al. (2019). A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart. Cell *179*, 1647-1660.e19. 10.1016/j.cell.2019.11.025.

174. Fawkner-Corbett, D., Antanaviciute, A., Parikh, K., Jagielowicz, M., Gerós, A.S., Gupta, T., Ashley, N., Khamis, D., Fowler, D., Morrissey, E., et al. (2021). Spatiotemporal analysis of human intestinal development at single-cell resolution. Cell *184*, 810-826.e23. 10.1016/j.cell.2020.12.016.

175. Thrane, K., Eriksson, H., Maaskola, J., Hansson, J., and Lundeberg, J. (2018). Spatially Resolved Transcriptomics Enables Dissection of Genetic Heterogeneity in Stage III Cutaneous Malignant Melanoma. Cancer Res *78*, 5970–5979. 10.1158/0008-5472.CAN-18-0747.

176. Ji, A.L., Rubin, A.J., Thrane, K., Jiang, S., Reynolds, D.L., Meyers, R.M., Guo, M.G., George, B.M., Mollbrink, A., Bergenstråhle, J., et al. (2020). Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma. Cell *182*, 497-514.e22. 10.1016/j.cell.2020.05.039.

177. Navarro, J.F., Croteau, D.L., Jurek, A., Andrusivova, Z., Yang, B., Wang, Y., Ogedegbe, B., Riaz, T., Støen, M., Desler, C., et al. (2020). Spatial Transcriptomics Reveals Genes Associated with Dysregulated Mitochondrial Functions and Stress Signaling in Alzheimer Disease. iScience *23*, 101556. 10.1016/j.isci.2020.101556.

178. Chen, W.-T., Lu, A., Craessaerts, K., Pavie, B., Sala Frigerio, C., Corthout, N., Qian, X., Laláková, J., Kühnemund, M., Voytyuk, I., et al. (2020). Spatial Transcriptomics and In Situ Sequencing to Study Alzheimer's Disease. Cell *182*, 976-991.e19. 10.1016/j.cell.2020.06.038.

179. Cusanovich, D.A., Daza, R., Adey, A., Pliner, H.A., Christiansen, L., Gunderson, K.L., Steemers, F.J., Trapnell, C., and Shendure, J. (2015). Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. Science *348*, 910–914. 10.1126/science.aab1601.

180. Stuart, T., Srivastava, A., Madad, S., Lareau, C.A., and Satija, R. (2021). Single-cell chromatin state analysis with Signac. Nat Methods *18*, 1333–1341. 10.1038/s41592-021-01282-5.

181. Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. J. Stat. Mech. *2008*, P10008. 10.1088/1742-5468/2008/10/P10008.

182. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573-3587.e29. 10.1016/j.cell.2021.04.048.

183. Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression (Genomics) 10.1101/576827.

184. Dann, E., Henderson, N.C., Teichmann, S.A., Morgan, M.D., and Marioni, J.C. (2022). Differential abundance testing on single-cell data using k-nearest neighbor graphs. Nat Biotechnol *40*, 245–253. 10.1038/s41587-021-01033-z.

185. Zhao, J., Jaffe, A., Li, H., Lindenbaum, O., Sefik, E., Jackson, R., Cheng, X., Flavell, R.A., and Kluger, Y. (2021). Detection of differentially abundant cell subpopulations in scRNA-seq data. Proceedings of the National Academy of Sciences *118*, e2100293118. 10.1073/pnas.2100293118.

186. Luecken, M.D., Büttner, M., Chaichoompu, K., Danese, A., Interlandi, M., Mueller, M.F., Strobl, D.C., Zappia, L., Dugas, M., Colomé-Tatché, M., et al. (2022). Benchmarking atlas-level data integration in single-cell genomics. Nat Methods *19*, 41–50. 10.1038/s41592-021-01336-8.

187. Haghverdi, L., Lun, A.T.L., Morgan, M.D., and Marioni, J.C. (2018). Batch effects in single-cell RNA sequencing data are corrected by matching mutual nearest neighbours. Nat Biotechnol *36*, 421–427. 10.1038/nbt.4091.

188. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888-1902.e21. 10.1016/j.cell.2019.05.031.

189. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods *16*, 1289–1296. 10.1038/s41592-019-0619-0.

190. Hao, Y., Stuart, T., Kowalski, M., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., et al. (2022). Dictionary learning for integrative, multimodal, and scalable single-cell analysis. 2022.02.24.481684. 10.1101/2022.02.24.481684.

191. Open Problems - Multimodal Single-Cell Integration https://kaggle.com/competitions/open-problems-multimodal.

192. Granja, J.M., Corces, M.R., Pierce, S.E., Bagdatli, S.T., Choudhry, H., Chang, H.Y., and Greenleaf, W.J. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat Genet *53*, 403–411. 10.1038/s41588-021-00790-6.

193. Li, Z., Nagai, J.S., Kuppe, C., Kramann, R., and Costa, I.G. (2022). scMEGA: Single-cell Multiomic Enhancer-based Gene Regulatory Network Inference. 2022.08.10.503335. 10.1101/2022.08.10.503335.

194. Zhang, S., Pyne, S., Pietrzak, S., Siahpirani, A.F., Sridharan, R., and Roy, S. (2022). Inference of cell type-specific gene regulatory networks on cell lineages from single cell omic datasets. 2022.07.25.501350. 10.1101/2022.07.25.501350.

195. González-Blas, C.B., Winter, S.D., Hulselmans, G., Hecker, N., Matetovici, I., Christiaens, V., Poovathingal, S., Wouters, J., Aibar, S., and Aerts, S. (2022). SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. 2022.08.19.504505. 10.1101/2022.08.19.504505.

196. Wang, L., Trasanidis, N., Wu, T., Dong, G., Hu, M., Bauer, D.E., and Pinello, L. (2022). Dictys: dynamic gene regulatory network dissects developmental continuum with single-cell multi-omics. 2022.09.14.508036. 10.1101/2022.09.14.508036.

197. Wang, X., Park, J., Susztak, K., Zhang, N.R., and Li, M. (2019). Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. Nat Commun *10*, 380. 10.1038/s41467-018-08023-x.

198. Shree, A., Pavan, M.K., and Zafar, H. (2022). scDREAMER: atlas-level integration of single-cell datasets using deep generative model paired with adversarial classifier. 2022.07.12.499846. 10.1101/2022.07.12.499846.

199. Treppner, M., Haug, S., Köttgen, A., and Binder, H. (2022). Designing Single Cell RNA-Sequencing Experiments for Learning Latent Representations. 2022.07.08.499284. 10.1101/2022.07.08.499284.

200. Lopez, R., Regier, J., Cole, M.B., Jordan, M.I., and Yosef, N. (2018). Deep generative modeling for single-cell transcriptomics. Nat Methods *15*, 1053–1058. 10.1038/s41592-018-0229-2.

201. Bahrami, M., Maitra, M., Nagy, C., Turecki, G., Rabiee, H.R., and Li, Y. (2021). Deep feature extraction of single-cell transcriptomes by generative adversarial network. Bioinformatics *37*, 1345–1351. 10.1093/bioinformatics/btaa976.

202. Amodio, M., Youlten, S.E., Venkat, A., San Juan, B.P., Chaffer, C.L., and Krishnaswamy, S. (2022). Single-cell multi-modal GAN reveals spatial patterns in single-cell data from triple-negative breast cancer. Patterns *3*, 100577. 10.1016/j.patter.2022.100577.

203. Obesity and overweight https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight.

204. Glastras, S.J., Valvi, D., and Bansal, A. (2021). Editorial: Developmental Programming of Metabolic Diseases. Frontiers in Endocrinology *12*.

205. Flegal, K.M., Carroll, M.D., Kit, B.K., and Ogden, C.L. (2012). Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999-2010. JAMA *307*, 491–497. 10.1001/jama.2012.39.

206. Matta, J., Carette, C., Rives Lange, C., and Czernichow, S. (2018). Épidémiologie de l'obésité en France et dans le monde. La Presse Médicale *47*, 434–438. 10.1016/j.lpm.2018.03.023.

207. Shah, N.S., Wang, M.C., Freaney, P.M., Perak, A.M., Carnethon, M.R., Kandula, N.R., Gunderson, E.P., Bullard, K.M., Grobman, W.A., O'Brien, M.J., et al. (2021). Trends in Gestational Diabetes at First Live Birth by Race and Ethnicity in the US, 2011-2019. JAMA *326*, 660–669. 10.1001/jama.2021.7217.

208. Barker, D.J.P., and Osmond, C. (1986). INFANT MORTALITY, CHILDHOOD NUTRITION, AND ISCHAEMIC HEART DISEASE IN ENGLAND AND WALES. The Lancet *327*, 1077–1081. 10.1016/S0140-6736(86)91340-1.

209. Barker, D.J., Osmond, C., Simmonds, S.J., and Wield, G.A. (1993). The relation of small head circumference and thinness at birth to death from cardiovascular disease in adult life. BMJ *306*, 422–426. 10.1136/bmj.306.6875.422.

210. Barker, D.J.P., Hales, C.N., Fall, C.H.D., Osmond, C., Phipps, K., and Clark, P.M.S. (1993). Type 2 (non-insulin-dependent) diabetes mellitus, hypertension and hyperlipidaemia (syndrome X): relation to reduced fetal growth. Diabetologia *36*, 62–67. 10.1007/BF00399095.

211. Barker, D.J., Winter, P.D., Osmond, C., Margetts, B., and Simmonds, S.J. (1989). Weight in infancy and death from ischaemic heart disease. Lancet *2*, 577–580. 10.1016/s0140-6736(89)90710-1.

212. Rich-Edwards, J.W., Stampfer, M.J., Manson, J.E., Rosner, B., Hankinson, S.E., Colditz, G.A., Willett, W.C., and Hennekens, C.H. (1997). Birth weight and risk of cardiovascular disease in a cohort of women followed up since 1976. BMJ *315*, 396–400. 10.1136/bmj.315.7105.396.

213. Frankel, S., Elwood, P., Sweetnam, P., Yarnell, J., and Smith, G.D. (1996). Birthweight, body-mass index in middle age, and incident coronary heart disease. Lancet *348*, 1478–1480. 10.1016/S0140-6736(96)03482-4.

214. Huxley, R.R., Shiell, A.W., and Law, C.M. (2000). The role of size at birth and postnatal catch-up growth in determining systolic blood pressure: a systematic review of the literature. Journal of Hypertension *18*, 815–831.

215. Newsome, C.A., Shiell, A.W., Fall, C.H.D., Phillips, D.I.W., Shier, R., and Law, C.M. (2003). Is birth weight related to later glucose and insulin metabolism?—a systematic review. Diabetic Medicine *20*, 339–348. 10.1046/j.1464-5491.2003.00871.x.

216. Rich-Edwards, J.W., Kleinman, K., Michels, K.B., Stampfer, M.J., Manson, J.E., Rexrode, K.M., Hibert, E.N., and Willett, W.C. (2005). Longitudinal study of birth weight and adult body mass index in predicting risk of coronary heart disease and stroke in women. BMJ *330*, 1115. 10.1136/bmj.38434.629630.E0.

217. Hyppönen, E., Leon, D.A., Kenward, M.G., and Lithell, H. (2001). Prenatal growth and risk of occlusive and haemorrhagic stroke in Swedish men and women born 1915-29: historical cohort study. BMJ *323*, 1033–1034. 10.1136/bmj.323.7320.1033.

218. Roseboom, T.J., van der Meulen, J.H., Osmond, C., Barker, D.J., Ravelli, A.C., and Bleker, O.P. (2000). Plasma lipid profiles in adults after prenatal exposure to the Dutch famine. The American Journal of Clinical Nutrition *72*, 1101–1106. 10.1093/ajcn/72.5.1101.

219. McCarton, C.M., Wallace, I.F., Divon, M., and Vaughan, H.G., Jr (1996). Cognitive and Neurologic Development of the Premature, Small for Gestational Age Infant Through Age 6: Comparison by Birth Weight and Gestational Age. Pediatrics *98*, 1167–1178. 10.1542/peds.98.6.1167.

220. Forman, M.R., Cantwell, M.M., Ronckers, C., and Zhang, Y. (2005). Through the Looking Glass at Early-Life Exposures and Breast Cancer Risk. Cancer Investigation *23*, 609–624. 10.1080/07357900500283093.

221. Yu, Z.B., Han, S.P., Zhu, G.Z., Zhu, C., Wang, X.J., Cao, X.G., and Guo, X.R. (2011). Birth weight and subsequent risk of obesity: a systematic review and meta-analysis. Obesity Reviews *12*, 525–542. 10.1111/j.1467-789X.2011.00867.x.

222. Boney, C.M., Verma, A., Tucker, R., and Vohr, B.R. (2005). Metabolic syndrome in childhood: association with birth weight, maternal obesity, and gestational diabetes mellitus. Pediatrics *115*, e290-296. 10.1542/peds.2004-1808.

223. Harder, T., Rodekamp, E., Schellong, K., Dudenhausen, J.W., and Plagemann, A. (2007). Birth Weight and Subsequent Risk of Type 2 Diabetes: A Meta-Analysis. American Journal of Epidemiology *165*, 849–857. 10.1093/aje/kwk071.

224. Evagelidou, E.N., Giapros, V.I., Challa, A.S., Cholevas, V.K., Vartholomatos, G.A., Siomou, E.C., Kolaitis, N.I., Bairaktari, E.T., and Andronikou, S.K. (2010). Prothrombotic State, Cardiovascular, and Metabolic Syndrome Risk Factors in Prepubertal Children Born Large for Gestational Age. Diabetes Care *33*, 2468–2470. 10.2337/dc10-1190.

225. Moore, G.S., Kneitel, A.W., Walker, C.K., Gilbert, W.M., and Xing, G. (2012). Autism risk in small- and large-for-gestational-age infants. Am J Obstet Gynecol *206*, 314.e1-9. 10.1016/j.ajog.2012.01.044.

226. Colman, I., Ataullahjan, A., Naicker, K., and Van Lieshout, R.J. (2012). Birth weight, stress, and symptoms of depression in adolescence: evidence of fetal programming in a national Canadian cohort. Can J Psychiatry *57*, 422–428. 10.1177/070674371205700705.

227. Van Lieshout, R.J., and Boyle, M.H. (2011). Is bigger better? Macrosomia and psychopathology later in life. Obes Rev *12*, e405-411. 10.1111/j.1467-789X.2010.00816.x.

228. Herva, A., Pouta, A., Hakko, H., Läksy, K., Joukamaa, M., and Veijola, J. (2008). Birth measures and depression at age 31 years: the Northern Finland 1966 Birth Cohort Study. Psychiatry Res *160*, 263–270. 10.1016/j.psychres.2007.07.020.

229. Hales, C.N., and Barker, D.J. (1992). Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis. Diabetologia *35*, 595–601. 10.1007/BF00400248.

230. Phillips, D.I.W. (1996). Insulin resistance as a programmed response to fetal undernutrition. Diabetologia *39*, 1119–1122. 10.1007/BF00400663.

231. Eriksson, J.G., Forsén, T., Tuomilehto, J., Osmond, C., and Barker, D.J.P. (2003). Early adiposity rebound in childhood and risk of Type 2 diabetes in adult life. Diabetologia *46*, 190–194. 10.1007/s00125-002-1012-5.

232. Eriksson, J.G., Forsén, T., Tuomilehto, J., Jaddoe, V.W.V., Osmond, C., and Barker, D.J.P. (2002). Effects of size at birth and childhood growth on the insulin resistance syndrome in elderly individuals. Diabetologia *45*, 342–348. 10.1007/s00125-001-0757-6.

233. Hattersley, A.T., and Tooke, J.E. (1999). The fetal insulin hypothesis: an alternative explanation of the association of low bir thweight with diabetes and vascular disease. The Lancet *353*, 1789–1792. 10.1016/S0140-6736(98)07546-1.

234. Day, I.N.M., Chen, X., Gaunt, T.R., King, T.H.T., Voropanov, A., Ye, S., Rodriguez, S., Syddall, H.E., Sayer, A.A., Dennison, E.M., et al. (2004). Late Life Metabolic Syndrome, Early Growth, and Common Polymorphism in the Growth Hormone and Placental Lactogen Gene Cluster. The Journal of Clinical Endocrinology & Metabolism *89*, 5569–5576. 10.1210/jc.2004-0152.

235. Lindsay, R.S., Dabelea, D., Roumain, J., Hanson, R.L., Bennett, P.H., and Knowler, W.C. (2000). Type 2 diabetes and low birth weight: the role of paternal inheritance in the association of low birth weight and diabetes. Diabetes *49*, 445–449. 10.2337/diabetes.49.3.445.

236. Hyppönen, E., Smith, G.D., and Power, C. (2003). Parental diabetes and birth weight of offspring: intergenerational cohort study. BMJ *326*, 19–20. 10.1136/bmj.326.7379.19.

237. Beaumont, R.N., Kotecha, S.J., Wood, A.R., Knight, B.A., Sebert, S., McCarthy, M.I., Hattersley, A.T., Järvelin, M.-R., Timpson, N.J., Freathy, R.M., et al. (2020). Common maternal and fetal genetic variants show expected polygenic effects on risk of small- or large-for-gestational-age (SGA or LGA), except in the smallest 3% of babies. PLOS Genetics *16*, e1009191. 10.1371/journal.pgen.1009191.

238. Gluckman, P.D., Hanson, M.A., and Spencer, H.G. (2005). Predictive adaptive responses and human evolution. Trends in Ecology & Evolution *20*, 527–533. 10.1016/j.tree.2005.08.001.

239. Khan, I., Dekou, V., Hanson, M., Poston, L., and Taylor, P. (2004). Predictive Adaptive Responses to Maternal High-Fat Diet Prevent Endothelial Dysfunction but Not Hypertension in Adult Rat Offspring. Circulation *110*, 1097–1102. 10.1161/01.CIR.0000139843.05436.A0.

240. Yajnik, C.S., Fall, C.H.D., Coyaji, K.J., Hirve, S.S., Rao, S., Barker, D.J.P., Joglekar, C., and Kellingray, S. (2003). Neonatal anthropometry: the thin-fat Indian baby. The Pune Maternal Nutrition Study. Int J Obes Relat Metab Disord *27*, 173–180. 10.1038/sj.ijo.802219.

241. Ravelli, A., van der Meulen, J., Michels, R., Osmond, C., Barker, D., Hales, C., and Bleker, O. (1998). Glucose tolerance in adults after prenatal exposure to famine. The Lancet *351*, 173–177. 10.1016/S0140-6736(97)07244-9.

242. Ozanne, S.E., and Hales, C.N. (2004). Catch-up growth and obesity in male mice. Nature *427*, 411–412. 10.1038/427411b.

243. Guariguata, L., Linnenkamp, U., Beagley, J., Whiting, D.R., and Cho, N.H. (2014). Global estimates of the prevalence of hyperglycaemia in pregnancy. Diabetes Research and Clinical Practice *103*, 176–185. 10.1016/j.diabres.2013.11.003.

244. Li, Y., Ren, X., He, L., Li, J., Zhang, S., and Chen, W. (2020). Maternal age and the risk of gestational diabetes mellitus: A systematic review and meta-analysis of over 120 million participants. Diabetes Research and Clinical Practice *162*, 108044. 10.1016/j.diabres.2020.108044.

245. Kaul, P., Savu, A., Yeung, R.O., and Ryan, E.A. (2022). Association between maternal glucose and large for gestational outcomes: Real-world evidence to support Hyperglycaemia and Adverse Pregnancy Outcomes (HAPO) study findings. Diabet Med *39*, e14786. 10.1111/dme.14786.

246. Sacks, D.A., Liu, A.I., Wolde-Tsadik, G., Amini, S.B., Huston-Presley, L., and Catalano, P.M. (2006). What proportion of birth weight is attributable to maternal glucose among infants of diabetic women? Am J Obstet Gynecol *194*, 501–507. 10.1016/j.ajog.2005.07.042.

247. Lowe, L.P., Metzger, B.E., Dyer, A.R., Lowe, J., McCance, D.R., Lappin, T.R.J., Trimble, E.R., Coustan, D.R., Hadden, D.R., Hod, M., et al. (2012). Hyperglycemia and Adverse Pregnancy Outcome (HAPO) Study: Associations of maternal A1C and glucose with pregnancy outcomes. Diabetes Care *35*, 574–580. 10.2337/dc11-1687.

248. Barbry, F., Lemaitre, M., Ternynck, C., Wallet, H., Cazaubiel, M., Labreuche, J., Subtil, D., and Vambergue, A. (2022). HbA1c at the time of testing for gestational diabetes identifies women at risk for pregnancy complications. Diabetes & Metabolism *48*, 101313. 10.1016/j.diabet.2021.101313.

249. Lemaitre, M., Ternynck, C., Bourry, J., Baudoux, F., Subtil, D., and Vambergue, A. (2022). Association Between HbA1c Levels on Adverse Pregnancy Outcomes During Pregnancy in Patients With Type 1 Diabetes. The Journal of Clinical Endocrinology & Metabolism *107*, e1117–e1125. 10.1210/clinem/dgab769.

250. Tam, W.H., Ma, R.C.W., Ozaki, R., Li, A.M., Chan, M.H.M., Yuen, L.Y., Lao, T.T.H., Yang, X., Ho, C.S., Tutino, G.E., et al. (2017). In Utero Exposure to Maternal Hyperglycemia Increases Childhood Cardiometabolic Risk in Offspring. Diabetes Care *40*, 679–686. 10.2337/dc16-2397.

251. Pettitt, D.J., Aleck, K.A., Baird, H.R., Carraher, M.J., Bennett, P.H., and Knowler, W.C. (1988). Congenital susceptibility to NIDDM. Role of intrauterine environment. Diabetes *37*, 622–628. 10.2337/diab.37.5.622.

252. Dabelea, D., Hanson, R.L., Lindsay, R.S., Pettitt, D.J., Imperatore, G., Gabir, M.M., Roumain, J., Bennett, P.H., and Knowler, W.C. (2000). Intrauterine exposure to diabetes conveys risks for type 2 diabetes and obesity: a study of discordant sibships. Diabetes *49*, 2208–2211. 10.2337/diabetes.49.12.2208.

253. Huda, S.S., Brodie, L.E., and Sattar, N. (2010). Obesity in pregnancy: prevalence and metabolic consequences. Seminars in Fetal and Neonatal Medicine *15*, 70–76. 10.1016/j.siny.2009.09.006.

254. Rasmussen, K.M., and Yaktine, A.L. Weight Gain During Pregnancy: Reexamining the Guidelines. 250.

255. Gaudet, L., Ferraro, Z.M., Wen, S.W., and Walker, M. (2014). Maternal Obesity and Occurrence of Fetal Macrosomia: A Systematic Review and Meta-Analysis. Biomed Res Int *2014*, 640291. 10.1155/2014/640291.

256. Ludwig, D.S., and Currie, J. (2010). The Relationship Between Pregnancy Weight Gain and Birth Weight: A Within Family Comparison. Lancet *376*, 984–990. 10.1016/S0140-6736(10)60751-9.

257. Yu, Z., Han, S., Zhu, J., Sun, X., Ji, C., and Guo, X. (2013). Pre-Pregnancy Body Mass Index in Relation to Infant Birth Weight and Offspring Overweight/Obesity: A Systematic Review and Meta-Analysis. PLoS One *8*, e61627. 10.1371/journal.pone.0061627.

258. Oken, E., Rifas-Shiman, S.L., Field, A.E., Frazier, A.L., and Gillman, M.W. (2008). Maternal Gestational Weight Gain and Offspring Weight in Adolescence. Obstet Gynecol *112*, 999–1006. 10.1097/AOG.0b013e31818a5d50.

259. Crozier, S.R., Inskip, H.M., Godfrey, K.M., Cooper, C., Harvey, N.C., Cole, Z.A., and Robinson, S.M. (2010). Weight gain in pregnancy and childhood body composition: findings from the Southampton Women's Survey. Am J Clin Nutr *91*, 1745–1751. 10.3945/ajcn.2009.29128.

260. Fraser, A., Tilling, K., Macdonald-Wallis, C., Sattar, N., Brion, M.-J., Benfield, L., Ness, A., Deanfield, J., Hingorani, A., Nelson, S.M., et al. (2010). Association of maternal weight gain in pregnancy with offspring obesity and metabolic and vascular traits in childhood. Circulation *121*, 2557–2564. 10.1161/CIRCULATIONAHA.109.906081.

261. Gaillard, R., Steegers, E.A.P., Duijts, L., Felix, J.F., Hofman, A., Franco, O.H., and Jaddoe, V.W.V. (2014). Childhood cardiometabolic outcomes of maternal obesity during pregnancy: the Generation R Study. Hypertension *63*, 683–691. 10.1161/HYPERTENSIONAHA.113.02671.

262. Gaillard, R., Steegers, E. a. P., Franco, O.H., Hofman, A., and Jaddoe, V.W.V. (2015). Maternal weight gain in different periods of pregnancy and childhood cardio-metabolic outcomes. The Generation R Study. Int J Obes *39*, 677–685. 10.1038/ijo.2014.175.

263. Perng, W., Gillman, M.W., Mantzoros, C.S., and Oken, E. (2014). A prospective study of maternal prenatal weight and offspring cardiometabolic health in midchildhood. Annals of Epidemiology *24*, 793-800.e1. 10.1016/j.annepidem.2014.08.002.

264. Oostvogels, A.J.J.M., Stronks, K., Roseboom, T.J., van der Post, J.A.M., van Eijsden, M., and Vrijkotte, T.G.M. (2014). Maternal Prepregnancy BMI, Offspring's Early Postnatal Growth, and Metabolic Profile at Age 5–6 Years: the ABCD Study. The Journal of Clinical Endocrinology & Metabolism *99*, 3845–3854. 10.1210/jc.2014-1561.

265. Reynolds, R.M., Allan, K.M., Raja, E.A., Bhattacharya, S., McNeill, G., Hannaford, P.C., Sarwar, N., Lee, A.J., Bhattacharya, S., and Norman, J.E. (2013). Maternal obesity during pregnancy and premature mortality from cardiovascular event in adult offspring: follow-up of 1 323 275 person years. BMJ *347*, f4539. 10.1136/bmj.f4539.

266. Dalziel, S.R., Walker, N.K., Parag, V., Mantell, C., Rea, H.H., Rodgers, A., and Harding, J.E. (2005). Cardiovascular risk factors after antenatal exposure to betamethasone: 30-year follow-up of a randomised controlled trial. The Lancet *365*, 1856–1862. 10.1016/S0140-6736(05)66617-2.

267. Doyle, L.W., Ford, G.W., Davis, N.M., and Callanan, C. (2000). Antenatal corticosteroid therapy and blood pressure at 14 years of age in preterm children. Clin Sci (Lond) *98*, 137–142.

268. MacArthur, B.A., Howie, R.N., Dezoete, J.A., and Elkins, J. (1982). School progress and cognitive development of 6-year-old children whose mothers were treated antenatally with betamethasone. Pediatrics *70*, 99–105.

269. Phillips, D.I.W., Walker, B.R., Reynolds, R.M., Flanagan, D.E.H., Wood, P.J., Osmond, C., Barker, D.J.P., and Whorwood, C.B. (2000). Low Birth Weight Predicts Elevated Plasma Cortisol Concentrations in Adults From 3 Populations. Hypertension *35*, 1301–1306. 10.1161/01.HYP.35.6.1301.

270. Reynolds, R.M., Walker, B.R., Syddall, H.E., Andrew, R., Wood, P.J., Whorwood, C.B., and Phillips, D.I.W. (2001). Altered Control of Cortisol Secretion in Adult Men with Low Birth Weight and Cardiovascular Risk Factors1. The Journal of Clinical Endocrinology & Metabolism *86*, 245–250. 10.1210/jcem.86.1.7145.

271. Miller, G.E., Chen, E., and Parker, K.J. (2011). Psychological stress in childhood and susceptibility to the chronic diseases of aging: moving toward a model of behavioral and biological mechanisms. Psychol Bull *137*, 959–997. 10.1037/a0024768.

272. Fall, C.H.D. (2011). Evidence for the intra-uterine programming of adiposity in later life. Ann Hum Biol *38*, 410–428. 10.3109/03014460.2011.592513.

273. Eriksson, J.G., Forsén, T., Tuomilehto, J., Osmond, C., and Barker, D.J. (2001). Early growth and coronary heart disease in later life: longitudinal study. BMJ *322*, 949–953. 10.1136/bmj.322.7292.949.

274. Law, C.M., Shiell, A.W., Newsome, C.A., Syddall, H.E., Shinebourne, E.A., Fayers, P.M., Martyn, C.N., and de Swiet, M. (2002). Fetal, infant, and childhood growth and adult blood pressure: a longitudinal study from birth to 22 years of age. Circulation *105*, 1088–1092. 10.1161/hc0902.104677.

275. Fall, C.H.D., Sachdev, H.S., Osmond, C., Lakshmy, R., Biswas, S.D., Prabhakaran, D., Tandon, N., Ramji, S., Reddy, K.S., Barker, D.J.P., et al. (2008). Adult metabolic syndrome and impaired glucose tolerance are associated with different patterns of BMI gain during infancy: Data from the New Delhi Birth Cohort. Diabetes Care *31*, 2349–2356. 10.2337/dc08-0911.

276. Bhargava, S.K., Sachdev, H.S., Fall, C.H.D., Osmond, C., Lakshmy, R., Barker, D.J.P., Biswas, S.K.D., Ramji, S., Prabhakaran, D., and Reddy, K.S. (2004). Relation of serial changes in childhood body-mass index to impaired glucose tolerance in young adulthood. N Engl J Med *350*, 865–875. 10.1056/NEJMoa035698.

277. Adair, L.S., Fall, C.H.D., Osmond, C., Stein, A.D., Martorell, R., Ramirez-Zea, M., Sachdev, H.S., Dahly, D.L., Bas, I., Norris, S.A., et al. (2013). Associations of linear growth and relative weight gain during early life with adult health and human capital in countries of low and middle income: findings from five birth cohort studies. Lancet *382*, 525–534. 10.1016/S0140-6736(13)60103-8.

278. Owen, C.G., Whincup, P.H., and Cook, D.G. (2011). Breast-feeding and cardiovascular risk factors and outcomes in later life: evidence from epidemiological studies. Proc Nutr Soc *70*, 478–484. 10.1017/S0029665111000590.

279. Horta, B.L., Loret de Mola, C., and Victora, C.G. (2015). Long-term consequences of breastfeeding on cholesterol, obesity, systolic blood pressure and type 2 diabetes: a systematic review and meta-analysis. Acta Paediatr *104*, 30–37. 10.1111/apa.13133.

280. Fall, C.H., Borja, J.B., Osmond, C., Richter, L., Bhargava, S.K., Martorell, R., Stein, A.D., Barros, F.C., Victora, C.G., and COHORTS group (2011). Infant-feeding patterns and cardiovascular risk factors in young adulthood: data from five cohorts in low- and middle-income countries. Int J Epidemiol *40*, 47–62. 10.1093/ije/dyq155.

281. Cusick, S.E., and Georgieff, M.K. (2016). The Role of Nutrition in Brain Development: The Golden Opportunity of the "First 1000 Days." The Journal of Pediatrics *175*, 16–21. 10.1016/j.jpeds.2016.05.013.

282. Victora, C.G., Adair, L., Fall, C., Hallal, P.C., Martorell, R., Richter, L., Sachdev, H.S., and Maternal and Child Undernutrition Study Group (2008). Maternal and child undernutrition: consequences for adult health and human capital. Lancet *371*, 340–357. 10.1016/S0140-6736(07)61692-4.

283. Martorell, R. (2017). Improved nutrition in the first 1000 days and adult human capital and health. American Journal of Human Biology *29*, e22952. 10.1002/ajhb.22952.

284. Velasco, I., Bath, S.C., and Rayman, M.P. (2018). Iodine as Essential Nutrient during the First 1000 Days of Life. Nutrients *10*, 290. 10.3390/nu10030290.

285. Goulet, O. (2015). Potential role of the intestinal microbiota in programming health and disease. Nutrition Reviews *73*, 32–40. 10.1093/nutrit/nuv039.

286. Selma-Royo, M., Tarrazó, M., García-Mantrana, I., Gómez-Gallego, C., Salminen, S., and Collado, M.C. (2019). Shaping Microbiota During the First 1000 Days of Life. In Probiotics and Child Gastrointestinal Health: Advances in Microbiology, Infectious Diseases and Public Health Volume 10 Advances in Experimental Medicine and Biology., S. Guandalini and F. Indrio, eds. (Springer International Publishing), pp. 3–24. 10.1007/5584_2018_312.

287. Carding, S., Verbeke, K., Vipond, D.T., Corfe, B.M., and Owen, L.J. (2015). Dysbiosis of the gut microbiota in disease. Microbial Ecology in Health and Disease *26*, 26191. 10.3402/mehd.v26.26191.

288. Verdu, E.F., Hayes, C.L., and O' Mahony, S.M. (2016). Chapter 9 - Importance of the Microbiota in Early Life and Influence on Future Health. In The Gut-Brain Axis, N. Hyland and C. Stanton, eds. (Academic Press), pp. 159–184. 10.1016/B978-0-12-802304-4.00009-8.

289. Notre but 1000 jours qui comptent pour la santé. https://1000jourspourlasante.fr/futurs-parents/le-projet/notre-but/.

290. Widdowson, E.M., and McCance, R.A. (1975). A review: new thoughts on growth. Pediatr Res *9*, 154–156. 10.1203/00006450-197503000-00010.

291. Woodall, S.M., Johnston, B.M., Breier, B.H., and Gluckman, P.D. (1996). Chronic maternal undernutrition in the rat leads to delayed postnatal growth and elevated blood pressure of offspring. Pediatr Res *40*, 438–443. 10.1203/00006450-199609000-00012.

292. Ozanne, S.E., and Hales, C.N. (1999). The long-term consequences of intra-uterine protein malnutrition for glucose metabolism. Proc Nutr Soc *58*, 615–619. 10.1017/s0029665199000804.

293. Reusens, B., and Remacle, C. (2006). Programming of the endocrine pancreas by the early nutritional environment. Int J Biochem Cell Biol *38*, 913–922. 10.1016/j.biocel.2005.10.012.

294. Dahri, S., Snoeck, A., Reusens-Billen, B., Remacle, C., and Hote, J.J. (1991). Islet Function in Offspring of Mothers on Low-Protein Diet During Gestation. Diabetes *40*, 115–120. 10.2337/diab.40.2.S115.

295. Kind, K.L., Clifton, P.M., Grant, P.A., Owens, P.C., Sohlstrom, A., Roberts, C.T., Robinson, J.S., and Owens, J.A. (2003). Effect of maternal feed restriction during pregnancy on glucose tolerance in the adult guinea pig. Am J Physiol Regul Integr Comp Physiol *284*, R140-152. 10.1152/ajpregu.00587.2001.

296. Gardner, D.S., Tingey, K., Van Bon, B.W.M., Ozanne, S.E., Wilson, V., Dandrea, J., Keisler, D.H., Stephenson, T., and Symonds, M.E. (2005). Programming of glucose-insulin metabolism in adult sheep after maternal undernutrition. Am J Physiol Regul Integr Comp Physiol *289*, R947-954. 10.1152/ajpregu.00120.2005.

297. Petrik, J., Reusens, B., Arany, E., Remacle, C., Coelho, C., Hoet, J.J., and Hill, D.J. (1999). A Low Protein Diet Alters the Balance of Islet Cell Replication and Apoptosis in the Fetal and Neonatal Rat and Is Associated with a Reduced Pancreatic Expression of Insulin-Like Growth Factor-II1. Endocrinology *140*, 4861–4873. 10.1210/endo.140.10.7042.

298. Snoeck, A., Remacle, C., Reusens, B., and Hoet, J.J. (1990). Effect of a Low Protein Diet during Pregnancy on the Fetal Rat Endocrine Pancreas. NEO *57*, 107–118. 10.1159/000243170.

299. Cherif, H., Reusens, B., Dahri, S., and Remacle, C. (2001). A Protein-Restricted Diet during Pregnancy Alters in Vitro Insulin Secretion from Islets of Fetal Wistar Rats. The Journal of Nutrition *131*, 1555–1559. 10.1093/jn/131.5.1555.

300. Dahri, S., Reusens, B., Remacle, C., and Hoet, J.J. (1995). Nutritional influences on pancreatic development and potential links with non-insulin-dependent diabetes. Proceedings of the Nutrition Society *54*, 345–356. 10.1079/PNS19950003.

301. Reusens, B., and Remacle, C. (2001). Intergenerational Effect of an Adverse Intrauterine Environment on Perturbation of Glucose Metabolism. Twin Research and Human Genetics *4*, 406–411. 10.1375/twin.4.5.406.

302. Ozanne, S.E., Wang, C.L., Coleman, N., and Smith, G.D. (1996). Altered muscle insulin sensitivity in the male offspring of protein-malnourished rats. American Journal of Physiology-Endocrinology and Metabolism *271*, E1128–E1134. 10.1152/ajpendo.1996.271.6.E1128.

303. Ozanne, S.E., Smith, G.D., Tikerpae, J., and Hales, C.N. (1996). Altered regulation of hepatic glucose output in the male offspring of protein-malnourished rat dams. American Journal of Physiology-Endocrinology and Metabolism *270*, E559–E564. 10.1152/ajpendo.1996.270.4.E559.

304. Holness, M.J., Fryer, L.G.D., and Sugden, M.C. (1999). Protein restriction during early development enhances insulin responsiveness but selectively impairs sensitivity to insulin at low concentrations in white adipose tissue during a later pregnancy. British Journal of Nutrition *81*, 481–489. 10.1017/S0007114599000847.

305. Shepherd, P.R., Crowther, N.J., Desai, M., Hales, C.N., and Ozanne, S.E. (1997). Altered adipocyte properties in the offspring of protein malnourished rats. British Journal of Nutrition *78*, 121–129. 10.1079/BJN19970124.

306. Seckl (2004). Prenatal glucocorticoids and long-term programming. European Journal of Endocrinology *151*, U49–U62. 10.1530/eje.0.151u049.

307. Cottrell, E., and Seckl, J. (2009). Prenatal stress, glucocorticoids and the programming of adult disease. Frontiers in Behavioral Neuroscience *3*.

308. Jensen, E., Gallaher, B., Breier, B., and Harding, J. (2002). The effect of a chronic maternal cortisol infusion on the late-gestation fetal sheep. The Journal of endocrinology *174*, 27–36. 10.1677/joe.0.1740027.

309. Tangalakis, K., Lumbers, E., Moritz, K., Towstoless, M., and Wintour, E. (1992). Effect of cortisol on blood pressure and vascular reactivity in the ovine fetus. Experimental Physiology *77*, 709–717. 10.1113/expphysiol.1992.sp003637.

310. Levitt, N.S., Lindsay, R.S., Holmes, M.C., and Seckl, J.R. (1996). Dexamethasone in the Last Week of Pregnancy Attenuates Hippocampal Glucocorticoid Receptor Gene Expression and Elevates Blood Pressure in the Adult Offspring in the Rat. NEN *64*, 412–418. 10.1159/000127146.

311. Nyirenda, M.J., Lindsay, R.S., Kenyon, C.J., Burchell, A., and Seckl, J.R. (1998). Glucocorticoid exposure in late gestation permanently programs rat hepatic phosphoenolpyruvate carboxykinase and glucocorticoid receptor expression and causes glucose intolerance in adult offspring. J Clin Invest *101*, 2174–2181. 10.1172/JCI1567.

312. Lesage, J., Hahn, D., Léonhardt, M., Blondeau, B., Bréant, B., and Dupouy, J. (2002). Maternal undernutrition during late gestation-induced intrauterine growth restriction in the rat is associated with impaired placental GLUT3 expression, but does not correlate with endogenous corticosterone levels. The Journal of endocrinology *174*, 37–43. 10.1677/joe.0.1740037.

313. Gardner, D.S., Jackson, A.A., and Langley-Evans, S.C. (1997). Maintenance of Maternal Diet-Induced Hypertension in the Rat Is Dependent on Glucocorticoids. Hypertension *30*, 1525–1530. 10.1161/01.HYP.30.6.1525.

314. Blondeau, B., Garofano, A., Czernichow, P., and Bréant, B. (1999). Age-Dependent Inability of the Endocrine Pancreas to Adapt to Pregnancy: A Long-Term Consequence of Perinatal Malnutrition in the Rat1. Endocrinology *140*, 4208–4213. 10.1210/endo.140.9.6960.

315. Hales, C., Desai, M., Ozanne, S., and Crowther, N.J. (1996). Fishing in the Stream of Diabetes: From Measuring Insulin to the Control of Fetal Organogenesis. Biochemical Society transactions *24*, 341–350. 10.1042/bst0240341.

316. Petry, C.J., Dorling, M.W., Pawlak, D.B., Ozanne, S.E., and Hales, C.N. (2001). Diabetes in Old Male Offspring of Rat Dams Fed a Reduced Protein Diet. Int J Exp Diabetes Res *2*, 139–143. 10.1155/EDR.2001.139.

317. Ozanne, S.E., Dorling, M.W., Wang, C.L., and Nave, B.T. (2001). Impaired PI 3-kinase activation in adipocytes from early growth-restricted male rats. American Journal of Physiology-Endocrinology and Metabolism *280*, E534–E539. 10.1152/ajpendo.2001.280.3.E534.

318. Breton, C., Lukaszewski, M.-A., Risold, P.-Y., Enache, M., Guillemot, J., Rivière, G., Delahaye, F., Lesage, J., Dutriez-Casteloot, I., Laborie, C., et al. (2009). Maternal prenatal undernutrition alters the response of POMC neurons to energy status variation in adult male rat offspring. American Journal of Physiology-Endocrinology and Metabolism *296*, E462–E472. 10.1152/ajpendo.90740.2008.

319. Delahaye, F., Breton, C., Risold, P.-Y., Enache, M., Dutriez-Casteloot, I., Laborie, C., Lesage, J., and Vieau, D. (2008). Maternal Perinatal Undernutrition Drastically Reduces Postnatal Leptin Surge and Affects the Development of Arcuate Nucleus Proopiomelanocortin Neurons in Neonatal Male Rat Pups. Endocrinology *149*, 470–475. 10.1210/en.2007-1263.

320. Kirk, S.L., Samuelsson, A.-M., Argenton, M., Dhonye, H., Kalamatianos, T., Poston, L., Taylor, P.D., and Coen, C.W. (2009). Maternal Obesity Induced by Diet in Rats Permanently Influences Central Processes Regulating Food Intake in Offspring. PLOS ONE *4*, e5870. 10.1371/journal.pone.0005870.

321. Aerts, L., and Van Assche, F.A. (2006). Animal evidence for the transgenerational development of diabetes mellitus. The International Journal of Biochemistry & Cell Biology *38*, 894–903. 10.1016/j.biocel.2005.07.006.

322. Akyol, A., McMullen, S., and Langley-Evans, S.C. (2012). Glucose intolerance associated with early-life exposure to maternal cafeteria feeding is dependent upon post-weaning diet. British Journal of Nutrition *107*, 964–978. 10.1017/S0007114511003916.

323. Khan, I.Y., Taylor, P.D., Dekou, V., Seed, P.T., Lakasing, L., Graham, D., Dominiczak, A.F., Hanson, M.A., and Poston, L. (2003). Gender-linked hypertension in offspring of lard-fed pregnant rats. Hypertension *41*, 168–175. 10.1161/01.hyp.0000047511.97879.fc.

324. Taylor, P.D., McConnell, J., Khan, I.Y., Holemans, K., Lawrence, K.M., Asare-Anane, H., Persaud, S.J., Jones, P.M., Petrie, L., Hanson, M.A., et al. (2005). Impaired glucose homeostasis and mitochondrial abnormalities in offspring of rats fed a fat-rich diet in pregnancy. Am J Physiol Regul Integr Comp Physiol *288*, R134-139. 10.1152/ajpregu.00355.2004.

325. Cardenas-Perez, R.E., Fuentes-Mera, L., de la Garza, A.L., Torre-Villalvazo, I., Reyes-Castro, L.A., Rodriguez-Rocha, H., Garcia-Garcia, A., Corona-Castillo, J.C., Tovar, A.R., Zambrano, E., et al. (2018). Maternal overnutrition by hypercaloric diets programs hypothalamic mitochondrial fusion and metabolic dysfunction in rat male offspring. Nutr Metab (Lond) *15*, 38. 10.1186/s12986-018-0279-6.

326. Buckley, A.J., Keserü, B., Briody, J., Thompson, M., Ozanne, S.E., and Thompson, C.H. (2005). Altered body composition and metabolism in the male offspring of high fat–fed rats. Metabolism *54*, 500–507. 10.1016/j.metabol.2004.11.003.

327. Sertorio, M.N., César, H., de Souza, E.A., Mennitti, L.V., Santamarina, A.B., De Souza Mesquita, L.M., Jucá, A., Casagrande, B.P., Estadella, D., Aguiar, O., et al. (2022). Parental High-Fat High-Sugar Diet Intake Programming Inflammatory and Oxidative Parameters of Reproductive Health in Male Offspring. Front Cell Dev Biol *10*, 867127. 10.3389/fcell.2022.867127.

328. Aguilera, O., Fernández, A.F., Muñoz, A., and Fraga, M.F. (2010). Epigenetics and environment: a complex relationship. Journal of Applied Physiology *109*, 243–251. 10.1152/japplphysiol.00068.2010.

329. Hahn, O., Grönke, S., Stubbs, T.M., Ficz, G., Hendrich, O., Krueger, F., Andrews, S., Zhang, Q., Wakelam, M.J., Beyer, A., et al. (2017). Dietary restriction protects from age-associated DNA methylation and induces epigenetic reprogramming of lipid metabolism. Genome Biol *18*, 56. 10.1186/s13059-017-1187-1.

330. Denham, J., Marques, F.Z., O'Brien, B.J., and Charchar, F.J. (2014). Exercise: Putting Action into Our Epigenome. Sports Med *44*, 189–209. 10.1007/s40279-013-0114-1.

331. Zakhari, S. (2013). Alcohol Metabolism and Epigenetics Changes. Alcohol Res *35*, 6–16.

332. Kaur, G., Begum, R., Thota, S., and Batra, S. (2019). A systematic review of smoking-related epigenetic alterations. Arch Toxicol *93*, 2715–2740. 10.1007/s00204-019-02562-y.

333. Tiffon, C. (2018). The Impact of Nutrition and Environmental Epigenetics on Human Health and Disease. International Journal of Molecular Sciences *19*, 3425. 10.3390/ijms19113425.

334. Ford, D., Ions, L.J., Alatawi, F., and Wakeling, L.A. (2011). The potential role of epigenetic responses to diet in ageing. Proc Nutr Soc *70*, 374–384. 10.1017/S0029665111000851.

335. Burdge, G.C., Slater-Jefferies, J., Torrens, C., Phillips, E.S., Hanson, M.A., and Lillycrop, K.A. (2007). Dietary protein restriction of pregnant rats in the F0 generation induces altered methylation of hepatic gene promoters in the adult male offspring in the F1 and F2 generations. British Journal of Nutrition *97*, 435–439. 10.1017/S0007114507352392.

336. Gluckman, P.D., Hanson, M.A., Cooper, C., and Thornburg, K.L. (2008). Effect of In Utero and Early-Life Conditions on Adult Health and Disease. N Engl J Med *359*, 61–73. 10.1056/NEJMra0708473.

337. Tobi, E.W., Goeman, J.J., Monajemi, R., Gu, H., Putter, H., Zhang, Y., Slieker, R.C., Stok, A.P., Thijssen, P.E., Müller, F., et al. (2014). DNA methylation signatures link prenatal famine exposure to growth and metabolism. Nat Commun *5*, 5592. 10.1038/ncomms6592.

338. Heijmans, B.T., Tobi, E.W., Stein, A.D., Putter, H., Blauw, G.J., Susser, E.S., Slagboom, P.E., and Lumey, L.H. (2008). Persistent epigenetic differences associated with prenatal exposure to famine in humans. Proceedings of the National Academy of Sciences *105*, 17046–17049. 10.1073/pnas.0806560105.

339. Yoo, J.Y., Lee, S., Lee, H.A., Park, H., Park, Y.J., Ha, E.H., and Kim, Y.J. (2014). Can Proopiomelanocortin Methylation Be Used as an Early Predictor of Metabolic Syndrome? Diabetes Care *37*, 734–739. 10.2337/dc13-1012.

340. Díaz, M., García, C., Sebastiani, G., de Zegher, F., López-Bermejo, A., and Ibáñez, L. (2016). Placental and Cord Blood Methylation of Genes Involved in Energy Homeostasis: Association With Fetal Growth and Neonatal Body Composition. Diabetes *66*, 779–784. 10.2337/db16-0776.

341. Lillycrop, K.A., Phillips, E.S., Jackson, A.A., Hanson, M.A., and Burdge, G.C. (2005). Dietary protein restriction of pregnant rats induces and folic acid supplementation prevents epigenetic modification of hepatic gene expression in the offspring. J. Nutr. *135*, 1382–1386. 10.1093/jn/135.6.1382.

342. Pham, T.D., MacLennan, N.K., Chiu, C.T., Laksana, G.S., Hsu, J.L., and Lane, R.H. (2003). Uteroplacental insufficiency increases apoptosis and alters p53 gene methylation in the full-term IUGR rat kidney. Am J Physiol Regul Integr Comp Physiol *285*, R962-970. 10.1152/ajpregu.00201.2003.

343. Bogdarina, I., Welham, S., King, P.J., Burns, S.P., and Clark, A.J.L. (2007). Epigenetic modification of the renin-angiotensin system in the fetal programming of hypertension. Circ Res *100*, 520–526. 10.1161/01.RES.0000258855.60637.58.

344. Haworth, K.E., Farrell, W.E., Emes, R.D., Ismail, K.M.K., Carroll, W.D., Hubball, E., Rooney, A., Yates, A.M., Mein, C., and Fryer, A.A. (2014). Methylation of the FGFR2 gene is associated with high birth weight centile in humans. Epigenomics *6*, 477–491. 10.2217/epi.14.40.

345. Wilhelm, -Benartzi Charlotte S., Houseman, E.A., Maccani, M.A., Poage, G.M., Koestler, D.C., Langevin, S.M., Gagne, L.A., Banister, C.E., Padbury, J.F., and Marsit, C.J. (2012). In Utero Exposures, Infant Growth, and DNA Methylation of Repetitive Elements and Developmentally Related Genes in Human Placenta. Environmental Health Perspectives *120*, 296–302. 10.1289/ehp.1103927.

346. Su, R., Wang, C., Feng, H., Lin, L., Liu, X., Wei, Y., and Yang, H. (2016). Alteration in Expression and Methylation of IGF2/H19 in Placenta and Umbilical Cord Blood Are Associated with Macrosomia Exposed to Intrauterine Hyperglycemia. PLOS ONE *11*, e0148399. 10.1371/journal.pone.0148399.

347. Claycombe, K.J., Uthus, E.O., Roemmich, J.N., Johnson, L.K., and Johnson, W.T. (2013). Prenatal low-protein and postnatal high-fat diets induce rapid adipose tissue growth by inducing Igf2 expression in Sprague Dawley rat offspring. J Nutr *143*, 1533–1539. 10.3945/jn.113.178038.

348. Huang, R.-C., Galati, J.C., Burrows, S., Beilin, L.J., Li, X., Pennell, C.E., van Eekelen, J., Mori, T.A., Adams, L.A., and Craig, J.M. (2012). DNA methylation of the IGF2/H19 imprinting control region and adiposity distribution in young adults. Clin Epigenetics *4*, 21. 10.1186/1868-7083-4-21.

349. Delahaye, F., Wijetunga, N.A., Heo, H.J., Tozour, J.N., Zhao, Y.M., Greally, J.M., and Einstein, F.H. (2014). Sexual dimorphism in epigenomic responses of stem cells to extreme fetal growth. Nat Commun *5*, 5187. 10.1038/ncomms6187.

350. van Dijk, S.J., Peters, T.J., Buckley, M., Zhou, J., Jones, P.A., Gibson, R.A., Makrides, M., Muhlhausler, B.S., and Molloy, P.L. (2018). DNA methylation in blood from neonatal screening cards and the association with BMI and insulin sensitivity in early childhood. International Journal of Obesity *42*, 28–35. 10.1038/ijo.2017.228.

351. Gagné-Ouellet, V., Houde, A.-A., Guay, S.-P., Perron, P., Gaudet, D., Guérin, R., Jean-Patrice, B., Hivert, M.-F., Brisson, D., and Bouchard, L. (2017). Placental lipoprotein lipase DNA methylation alterations are associated with gestational diabetes and body composition at 5 years of age. Epigenetics *12*, 616–625. 10.1080/15592294.2017.1322254.

352. Chen, P., Piaggi, P., Traurig, M., Bogardus, C., Knowler, W.C., Baier, L.J., and Hanson, R.L. (2017). Differential methylation of genes in individuals exposed to maternal diabetes in utero. Diabetologia *60*, 645–655. 10.1007/s00125-016-4203-1.

353. Hjort, L., Martino, D., Grunnet, L.G., Naeem, H., Maksimovic, J., Olsson, A.H., Zhang, C., Ling, C., Olsen, S.F., Saffery, R., et al. (2018). Gestational diabetes and maternal obesity are associated with epigenome-wide methylation changes in children. JCI Insight *3*. 10.1172/jci.insight.122572.

354. del Rosario, M.C., Ossowski, V., Knowler, W.C., Bogardus, C., Baier, L.J., and Hanson, R.L. (2014). Potential epigenetic dysregulation of genes associated with MODY and type 2 diabetes in humans exposed to a diabetic intrauterine environment: An analysis of genome-wide DNA methylation. Metabolism *63*, 654–660. 10.1016/j.metabol.2014.01.007.

355. Obradovic, M., Sudar-Milovanovic, E., Soskic, S., Essack, M., Arya, S., Stewart, A.J., Gojobori, T., and Isenovic, E.R. (2021). Leptin and Obesity: Role and Clinical Implication. Frontiers in Endocrinology *12*.

356. Bouchard, L., Thibault, S., Guay, S.-P., Santure, M., Monpetit, A., St-Pierre, J., Perron, P., and Brisson, D. (2010). Leptin Gene Epigenetic Adaptation to Impaired Glucose Metabolism During Pregnancy. Diabetes Care *33*, 2436–2441. 10.2337/dc10-1024.

357. Lesseur, C., Armstrong, D.A., Paquette, A.G., Li, Z., Padbury, J.F., and Marsit, C.J. (2014). Maternal obesity and gestational diabetes are associated with placental leptin DNA methylation. American Journal of Obstetrics and Gynecology *211*, 654.e1-654.e9. 10.1016/j.ajog.2014.06.037.

358. Nogues, P., Dos Santos, E., Jammes, H., Berveiller, P., Arnould, L., Vialard, F., and Dieudonné, M.-N. (2019). Maternal obesity influences expression and DNA methylation of the adiponectin and leptin systems in human third-trimester placenta. Clinical Epigenetics *11*, 20. 10.1186/s13148-019-0612-6.

359. Attig, L., Vigé, A., Gabory, A., Karimi, M., Beauger, A., Gross, M.-S., Athias, A., Gallou-Kabani, C., Gambert, P., Ekstrom, T.J., et al. (2013). Dietary Alleviation of Maternal Obesity and Diabetes: Increased Resistance to Diet-Induced Obesity Transcriptional and Epigenetic Signatures. PLOS ONE *8*, e66816. 10.1371/journal.pone.0066816.

360. Aagaard-Tillery, K.M., Grove, K., Bishop, J., Ke, X., Fu, Q., McKnight, R., and Lane, R.H. (2008). Developmental origins of disease and determinants of chromatin structure: maternal diet modifies the primate fetal epigenome. J Mol Endocrinol *41*, 91–102. 10.1677/JME-08-0025.

361. Puppala, S., Li, C., Glenn, J.P., Saxena, R., Gawrieh, S., Quinn, A., Palarczyk, J., Dick Jr, E.J., Nathanielsz, P.W., and Cox, L.A. (2018). Primate fetal hepatic responses to maternal obesity: epigenetic signalling pathways and lipid accumulation. The Journal of Physiology *596*, 5823–5837. 10.1113/JP275422.

362. Tabachnik, T., Kisliouk, T., Marco, A., Meiri, N., and Weller, A. (2017). Thyroid Hormone-Dependent Epigenetic Regulation of Melanocortin 4 Receptor Levels in Female Offspring of Obese Rats. Endocrinology *158*, 842–851. 10.1210/en.2016-1854.

363. Matrisciano, F., Tueting, P., Dalal, I., Kadriu, B., Grayson, D.R., Davis, J.M., Nicoletti, F., and Guidotti, A. (2013). Epigenetic modifications of GABAergic interneurons are associated with the schizophrenia-like phenotype induced by prenatal stress in mice. Neuropharmacology *68*, 184–194. 10.1016/j.neuropharm.2012.04.013.

364. Babenko, O., Kovalchuk, I., and Metz, G.A.S. (2015). Stress-induced perinatal and transgenerational epigenetic programming of brain development and mental health. Neuroscience & Biobehavioral Reviews *48*, 70–91. 10.1016/j.neubiorev.2014.11.013.

365. Sierksma, A.S.R., Prickaerts, J., Chouliaras, L., Rostamian, S., Delbroek, L., Rutten, B.P.F., Steinbusch, H.W.M., and van den Hove, D.L.A. (2013). Behavioral and neurobiological effects of prenatal stress exposure in male and female APPswe/PS1dE9 mice. Neurobiol Aging *34*, 319–337. 10.1016/j.neurobiolaging.2012.05.012.

366. Hui, J., Feng, G., Zheng, C., Jin, H., and Jia, N. (2017). Maternal separation exacerbates Alzheimer's disease-like behavioral and pathological changes in adult APPswe/PS1dE9 mice. Behav Brain Res *318*, 18–23. 10.1016/j.bbr.2016.10.030.

367. Lesuis, S.L., Weggen, S., Baches, S., Lucassen, P.J., and Krugers, H.J. (2018). Targeting glucocorticoid receptors prevents the effects of early life stress on amyloid pathology and cognitive performance in APP/PS1 mice. Transl Psychiatry *8*, 1–11. 10.1038/s41398-018-0101-2.

368. Boulais, P.E., and Frenette, P.S. (2015). Making sense of hematopoietic stem cell niches. Blood *125*, 2621–2629. 10.1182/blood-2014-09-570192.

369. Mikkola, H.K.A., and Orkin, S.H. (2006). The journey of developing hematopoietic stem cells. Development *133*, 3733–3744. 10.1242/dev.02568.

370. Bruns, I., Lucas, D., Pinho, S., Ahmed, J., Lambert, M.P., Kunisaki, Y., Scheiermann, C., Schiff, L., Poncz, M., Bergman, A., et al. (2014). Megakaryocytes regulate hematopoietic stem cell quiescence through CXCL4 secretion. Nat Med *20*, 1315–1320. 10.1038/nm.3707.

371. Zhao, M., Perry, J.M., Marshall, H., Venkatraman, A., Qian, P., He, X.C., Ahamed, J., and Li, L. (2014). Megakaryocytes maintain homeostatic quiescence and promote post-injury regeneration of hematopoietic stem cells. Nat Med *20*, 1321–1326. 10.1038/nm.3706.

372. Chow, A., Lucas, D., Hidalgo, A., Méndez-Ferrer, S., Hashimoto, D., Scheiermann, C., Battista, M., Leboeuf, M., Prophete, C., van Rooijen, N., et al. (2011). Bone marrow CD169+ macrophages promote the retention of hematopoietic stem and progenitor cells in the mesenchymal stem cell niche. J Exp Med *208*, 261–271. 10.1084/jem.20101688.

373. Hao, Q., Shah, A., Thiemann, F., Smogorzewska, E., and Crooks, G. (1995). A functional comparison of CD34 + CD38- cells in cord blood and bone marrow. Blood *86*, 3745–3753. 10.1182/blood.V86.10.3745.bloodjournal86103745.

374. Wagner, J.E., Barker, J.N., DeFor, T.E., Baker, K.S., Blazar, B.R., Eide, C., Goldman, A., Kersey, J., Krivit, W., MacMillan, M.L., et al. (2002). Transplantation of unrelated donor umbilical cord blood in 102 patients with malignant and nonmalignant diseases: influence of CD34 cell dose and HLA disparity on treatment-related mortality and survival. Blood *100*, 1611–1618. 10.1182/blood-2002-01-0294.

375. Rocha, V., Labopin, M., Sanz, G., Arcese, W., Schwerdtfeger, R., Bosi, A., Jacobsen, N., Ruutu, T., de Lima, M., Finke, J., et al. (2004). Transplants of Umbilical-Cord Blood or Bone Marrow from Unrelated Donors in Adults with Acute Leukemia. New England Journal of Medicine *351*, 2276–2285. 10.1056/NEJMoa041469.

376. Bogunia-Kubik, K. (2001). [Cytokine production by adult and cord blood (CB) cells--comparison and explanation of differences. Postepy Hig Med Dosw *55*, 629–641.

377. Hordyjewska, A., Popiołek, Ł., and Horecka, A. (2015). Characteristics of hematopoietic stem cells of umbilical cord blood. Cytotechnology *67*, 387–396. 10.1007/s10616-014-9796-y.

378. Grskovic, B., Ruzicka, K., Karimi, A., Qujeq, D., and Müller, M.M. (2004). Cell cycle analysis of the CD133+ and CD133− cells isolated from umbilical cord blood. Clinica Chimica Acta *343*, 173–178. 10.1016/j.cccn.2004.01.023.

379. Gonzales, K.A.U., Polak, L., Matos, I., Tierney, M.T., Gola, A., Wong, E., Infarinato, N.R., Nikolova, M., Luo, S., Liu, S., et al. (2021). Stem cells expand potency and alter tissue fitness by accumulating diverse epigenetic memories. Science *374*, eabh2444. 10.1126/science.abh2444.

380. Louwen, F., Ritter, A., Kreis, N.N., and Yuan, J. (2018). Insight into the development of obesity: functional alterations of adipose-derived mesenchymal stem cells. Obesity Reviews *19*, 888–904. 10.1111/obr.12679.

381. Oñate, B., Vilahur, G., Camino-López, S., Díez-Caballero, A., Ballesta-López, C., Ybarra, J., Moscatiello, F., Herrero, J., and Badimon, L. (2013). Stem cells isolated from adipose tissue of obese patients show changes in their transcriptomic profile that indicate loss in stemcellness and increased commitment to an adipocyte-like phenotype. BMC Genomics *14*, 625. 10.1186/1471-2164-14-625.

382. López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M., and Kroemer, G. (2013). The Hallmarks of Aging. Cell *153*, 1194–1217. 10.1016/j.cell.2013.05.039.

383. Sen, G.L., Reuter, J.A., Webster, D.E., Zhu, L., and Khavari, P.A. (2010). DNMT1 maintains progenitor function in self-renewing somatic tissue. Nature *463*, 563–567. 10.1038/nature08683.

384. Bröske, A.-M., Vockentanz, L., Kharazi, S., Huska, M.R., Mancini, E., Scheller, M., Kuhl, C., Enns, A., Prinz, M., Jaenisch, R., et al. (2009). DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. Nat Genet *41*, 1207–1215. 10.1038/ng.463.

385. Beerman, I., and Rossi, D.J. (2015). Epigenetic Control of Stem Cell Potential during Homeostasis, Aging, and Disease. Cell Stem Cell *16*, 613–625. 10.1016/j.stem.2015.05.009.

386. Genovese, G., Kähler, A.K., Handsaker, R.E., Lindberg, J., Rose, S.A., Bakhoum, S.F., Chambert, K., Mick, E., Neale, B.M., Fromer, M., et al. (2014). Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N Engl J Med *371*, 2477–2487. 10.1056/NEJMoa1409405.

387. Challen, G.A., Sun, D., Jeong, M., Luo, M., Jelinek, J., Berg, J.S., Bock, C., Vasanthakumar, A., Gu, H., Xi, Y., et al. (2012). Dnmt3a is essential for hematopoietic stem cell differentiation. Nat Genet *44*, 23–31. 10.1038/ng.1009.

388. Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., Figueroa, M.E., Vasanthakumar, A., Patel, J., Zhao, X., et al. (2011). Tet2 Loss Leads to Increased Hematopoietic Stem Cell Self-Renewal and Myeloid Transformation. Cancer Cell *20*, 11–24. 10.1016/j.ccr.2011.06.001.

389. Shide, K., Kameda, T., Shimoda, H., Yamaji, T., Abe, H., Kamiunten, A., Sekine, M., Hidaka, T., Katayose, K., Kubuki, Y., et al. (2012). TET2 is essential for survival and hematopoietic stem cell homeostasis. Leukemia *26*, 2216–2223. 10.1038/leu.2012.94.

390. Ko, M., Bandukwala, H.S., An, J., Lamperti, E.D., Thompson, E.C., Hastie, R., Tsangaratou, A., Rajewsky, K., Koralov, S.B., and Rao, A. (2011). Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. Proceedings of the National Academy of Sciences *108*, 14566–14571. 10.1073/pnas.1112317108.

391. Cimmino, L., Dawlaty, M.M., Ndiaye-Lobry, D., Yap, Y.S., Bakogianni, S., Yu, Y., Bhattacharyya, S., Shaknovich, R., Geng, H., Lobry, C., et al. (2015). TET1 is a tumor suppressor of hematopoietic malignancy. Nat Immunol *16*, 653–662. 10.1038/ni.3148.

392. Buenrostro, J.D., Corces, M.R., Lareau, C.A., Wu, B., Schep, A.N., Aryee, M.J., Majeti, R., Chang, H.Y., and Greenleaf, W.J. (2018). Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. Cell *173*, 1535-1548.e16. 10.1016/j.cell.2018.03.074.

393. Izzo, F., Lee, S.C., Poran, A., Chaligne, R., Gaiti, F., Gross, B., Murali, R.R., Deochand, S.D., Ang, C., Jones, P.W., et al. (2020). DNA methylation disruption reshapes the hematopoietic differentiation landscape. Nat Genet *52*, 378–387. 10.1038/s41588-020-0595-4.

394. Asada, S., and Kitamura, T. (2021). Clonal hematopoiesis and associated diseases: A review of recent findings. Cancer Sci *112*, 3962–3971. 10.1111/cas.15094.

395. Pardali, E., Dimmeler, S., Zeiher, A.M., and Rieger, M.A. (2020). Clonal hematopoiesis, aging, and cardiovascular diseases. Experimental Hematology *83*, 95–104. 10.1016/j.exphem.2019.12.006.

396. Jaiswal, S., and Ebert, B.L. (2019). Clonal hematopoiesis in human aging and disease. Science *366*. 10.1126/science.aan4673.

397. Pang, W.W., Price, E.A., Sahoo, D., Beerman, I., Maloney, W.J., Rossi, D.J., Schrier, S.L., and Weissman, I.L. (2011). Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age. PNAS *108*, 20012–20017. 10.1073/pnas.1116110108.

398. Sano, S., Oshima, K., Wang, Y., MacLauchlan, S., Katanasaka, Y., Sano, M., Zuriaga, M.A., Yoshiyama, M., Goukassian, D., Cooper, M.A., et al. (2018). Tet2-mediated Clonal Hematopoiesis Accelerates Heart Failure through a Mechanism Involving the IL-1β/NLRP3 Inflammasome. J Am Coll Cardiol *71*, 875–886. 10.1016/j.jacc.2017.12.037.

399. Mann, M., Mehta, A., de Boer, C.G., Kowalczyk, M.S., Lee, K., Haldeman, P., Rogel, N., Knecht, A.R., Farouq, D., Regev, A., et al. (2018). Heterogeneous Responses of Hematopoietic Stem Cells to Inflammatory Stimuli are Altered with Age. Cell Rep *25*, 2992-3005.e5. 10.1016/j.celrep.2018.11.056.

400. Kovtonyuk, L.V., Fritsch, K., Feng, X., Manz, M.G., and Takizawa, H. (2016). Inflamm-Aging of Hematopoiesis, Hematopoietic Stem Cells, and the Bone Marrow Microenvironment. Frontiers in Immunology *7*.

401. van Greevenbroek, M.M.J., Schalkwijk, C.G., and Stehouwer, C.D.A. (2013). Obesity-associated low-grade inflammation in type 2 diabetes mellitus: causes and consequences. Neth J Med *71*, 174–187.

402. Asmat, U., Abad, K., and Ismail, K. (2016). Diabetes mellitus and oxidative stress—A concise review. Saudi Pharmaceutical Journal *24*, 547–553. 10.1016/j.jsps.2015.03.013.

403. Flynn, M.C., Kraakman, M.J., Tikellis, C., Lee, M.K.S., Hanssen, N.M.J., Kammoun, H.L., Pickering, R.J., Dragoljevic, D., Al-Sharea, A., Barrett, T.J., et al. (2020). Transient Intermittent Hyperglycemia Accelerates Atherosclerosis by Promoting Myelopoiesis. Circulation Research *127*, 877–892. 10.1161/CIRCRESAHA.120.316653.

404. Ferraro, F., Lymperi, S., Méndez-Ferrer, S., Saez, B., Spencer, J.A., Yeap, B.Y., Masselli, E., Graiani, G., Prezioso, L., Rizzini, E.L., et al. (2011). Diabetes impairs hematopoietic stem cell mobilization by altering niche function. Sci Transl Med *3*, 104ra101. 10.1126/scitranslmed.3002191.

405. Chandra, T., Afreen, S., Kumar, A., Singh, U., and Gupta, A. (2012). Does umbilical cord blood-derived CD34+ cell concentration depend on the weight and sex of a full-term infant? J Pediatr Hematol Oncol *34*, 184–187. 10.1097/MPH.0b013e318249adb6.

406. Al-Sweedan, S.A., Musalam, L., and Obeidat, B. (2013). Factors predicting the hematopoietic stem cells content of the umbilical cord blood. Transfusion and Apheresis Science *48*, 247–252. 10.1016/j.transci.2013.01.003.

407. Rowisha, M.A., El-Shanshory, M.R., El-Hawary, E.E., Ahmed, A.Y., and Altoraky, S.R.M. (2020). Impact of maternal and neonatal factors on umbilical cord CD34+ cells. Stem Cell Investig *7*, 5. 10.21037/sci.2020.03.01.

408. Kamimae-Lanning, A.N., Krasnow, S.M., Goloviznina, N.A., Zhu, X., Roth-Carter, Q.R., Levasseur, P.R., Jeng, S., McWeeney, S.K., Kurre, P., and Marks, D.L. (2015). Maternal high-fat diet and obesity compromise fetal hematopoiesis. Molecular Metabolism *4*, 25–38. 10.1016/j.molmet.2014.11.001.

409. Greally, J.M. (2018). The HELP-Based DNA Methylation Assays. Methods Mol Biol *1708*, 191–207. 10.1007/978-1-4939-7481-8_11.

410. Wijetunga, N.A., Delahaye, F., Zhao, Y.M., Golden, A., Mar, J.C., Einstein, F.H., and Greally, J.M. (2014). The meta-epigenomic structure of purified human stem cell populations is defined at cis-regulatory sequences. Nat Commun *5*, 5195. 10.1038/ncomms6195.

411. de Mendoza, A., Nguyen, T.V., Ford, E., Poppe, D., Buckberry, S., Pflueger, J., Grimmer, M.R., Stolzenburg, S., Bogdanovic, O., Oshlack, A., et al. (2022). Large-scale manipulation of promoter DNA methylation reveals context-specific transcriptional responses and stability. Genome Biology *23*, 163. 10.1186/s13059-022-02728-5.

412. Thompson, R.F., Fazzari, M.J., Niu, H., Barzilai, N., Simmons, R.A., and Greally, J.M. (2010). Experimental Intrauterine Growth Restriction Induces Alterations in DNA Methylation and Gene Expression in Pancreatic Islets of Rats *. Journal of Biological Chemistry *285*, 15111–15118. 10.1074/jbc.M109.095133.

413. Muller-Sieburg, C.E., Sieburg, H.B., Bernitz, J.M., and Cattarossi, G. (2012). Stem cell heterogeneity: implications for aging and regenerative medicine. Blood *119*, 3900–3907. 10.1182/blood-2011-12-376749.

414. Haas, S., Trumpp, A., and Milsom, M.D. (2018). Causes and Consequences of Hematopoietic Stem Cell Heterogeneity. Cell Stem Cell *22*, 627–638. 10.1016/j.stem.2018.04.003.

415. Issa, J.-P. (2014). Aging and epigenetic drift: a vicious cycle. J Clin Invest *124*, 24–29. 10.1172/JCI69735.

416. Sanjuan-Pla, A., Macaulay, I.C., Jensen, C.T., Woll, P.S., Luis, T.C., Mead, A., Moore, S., Carella, C., Matsuoka, S., Bouriez Jones, T., et al. (2013). Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. Nature *502*, 232–236. 10.1038/nature12495.

417. Pinho, S., Marchand, T., Yang, E., Wei, Q., Nerlov, C., and Frenette, P.S. (2018). Lineage-Biased Hematopoietic Stem Cells Are Regulated by Distinct Niches. Developmental Cell *44*, 634-641.e4. 10.1016/j.devcel.2018.01.016.

418. Pietras, E.M., Reynaud, D., Kang, Y.-A., Carlin, D., Calero-Nieto, F.J., Leavitt, A.D., Stuart, J.M., Göttgens, B., and Passegué, E. (2015). Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions. Cell Stem Cell *17*, 35–46. 10.1016/j.stem.2015.05.003.

419. Sookoian, S., Gianotti, T.F., Burgueño, A.L., and Pirola, C.J. (2013). Fetal metabolic programming and epigenetic modifications: a systems biology approach. Pediatr Res *73*, 531–542. 10.1038/pr.2013.2.

420. Silverman, B.L., Metzger, B.E., Cho, N.H., and Loeb, C.A. (1995). Impaired glucose tolerance in adolescent offspring of diabetic mothers. Relationship to fetal hyperinsulinism. Diabetes Care *18*, 611–617. 10.2337/diacare.18.5.611.

421. Hong, Y.H., and Lee, J.-E. (2021). Large for Gestational Age and Obesity-Related Comorbidities. J Obes Metab Syndr *30*, 124–131. 10.7570/jomes20130.

422. Chiavaroli, V., Giannini, C., D'Adamo, E., de Giorgis, T., Chiarelli, F., and Mohn, A. (2009). Insulin resistance and oxidative stress in children born small and large for gestational age. Pediatrics *124*, 695–702. 10.1542/peds.2008-3056.

423. Wei, J.-N., Sung, F.-C., Li, C.-Y., Chang, C.-H., Lin, R.-S., Lin, C.-C., Chiang, C.-C., and Chuang, L.-M. (2003). Low birth weight and high birth weight infants are both at an increased risk to have type 2 diabetes among schoolchildren in taiwan. Diabetes Care *26*, 343–348. 10.2337/diacare.26.2.343.

424. Henriksen, T. (2008). The macrosomic fetus: a challenge in current obstetrics. Acta Obstetricia et Gynecologica Scandinavica *87*, 134–145. 10.1080/00016340801899289.

425. Koyanagi, A., Zhang, J., Dagvadorj, A., Hirayama, F., Shibuya, K., Souza, J.P., and Gülmezoglu, A.M. (2013). Macrosomia in 23 developing countries: an analysis of a multicountry, facility-based, cross-sectional survey. The Lancet *381*, 476–483. 10.1016/S0140-6736(12)61605-5.

426. Schwartz, R., Gruppuso, P.A., Petzold, K., Brambilla, D., Hiilesmaa, V., and Teramo, K.A. (1994). Hyperinsulinemia and Macrosomia in the Fetus of the Diabetic Mother. Diabetes Care *17*, 640–648. 10.2337/diacare.17.7.640.

427. Susa, J.B., Neave, C., Sehgal, P., Singer, D.B., Zeller, W.P., and Schwartz, R. (1984). Chronic Hyperinsulinemia in the Fetal Rhesus Monkey: Effects of Physiologic Hyperinsulinemia on Fetal Growth and Composition. Diabetes *33*, 656–660. 10.2337/diab.33.7.656.

428. Wang, J., Moore, D., Subramanian, A., Cheng, K.K., Toulis, K.A., Qiu, X., Saravanan, P., Price, M.J., and Nirantharakumar, K. (2018). Gestational dyslipidaemia and adverse birthweight outcomes: a systematic review and meta-analysis. Obes Rev *19*, 1256–1268. 10.1111/obr.12693.

429. Chen, K.-Y., Lin, S.-Y., Lee, C.-N., Wu, H.-T., Kuo, C.-H., Kuo, H.-C., Chuang, C.-C., Kuo, C.-H., Chen, S.-C., Fan, K.-C., et al. (2021). Maternal Plasma Lipids During Pregnancy, Insulin-like Growth Factor-1, and Excess Fetal Growth. J Clin Endocrinol Metab *106*, e3461–e3472. 10.1210/clinem/dgab364.

430. Hebert, L.E., Weuve, J., Scherr, P.A., and Evans, D.A. (2013). Alzheimer disease in the United States (2010–2050) estimated using the 2010 census. Neurology *80*, 1778–1783. 10.1212/WNL.0b013e31828726f5.

431. Liu, C.-C., Kanekiyo, T., Xu, H., and Bu, G. (2013). Apolipoprotein E and Alzheimer disease: risk, mechanisms, and therapy. Nat Rev Neurol *9*, 106–118. 10.1038/nrneurol.2012.263.

432. Gasiorowska, A., Wydrych, M., Drapich, P., Zadrozny, M., Steczkowska, M., Niewiadomski, W., and Niewiadomska, G. (2021). The Biology and Pathobiology of Glutamatergic, Cholinergic, and Dopaminergic Signaling in the Aging Brain. Frontiers in Aging Neuroscience *13*.

433. Arias-Carrión, O., Stamelou, M., Murillo-Rodríguez, E., Menéndez-González, M., and Pöppel, E. (2010). Dopaminergic reward system: a short integrative review. Int Arch Med *3*, 24. 10.1186/1755-7682-3-24.

434. Bélanger, M., Allaman, I., and Magistretti, P.J. (2011). Brain energy metabolism: focus on astrocyte-neuron metabolic cooperation. Cell Metab *14*, 724–738. 10.1016/j.cmet.2011.08.016.

435. Prinz, M., and Priller, J. (2014). Microglia and brain macrophages in the molecular age: from origin to neuropsychiatric disease. Nat Rev Neurosci *15*, 300–312. 10.1038/nrn3722.

436. Bradl, M., and Lassmann, H. (2010). Oligodendrocytes: biology and pathology. Acta Neuropathol *119*, 37–53. 10.1007/s00401-009-0601-5.

437. Penney, J., Ralvenius, W.T., and Tsai, L.-H. (2020). Modeling Alzheimer's disease with iPSC-derived brain cells. Mol Psychiatry *25*, 148–167. 10.1038/s41380-019-0468-3.

438. Müller, U.C., Deller, T., and Korte, M. (2017). Not just amyloid: physiological functions of the amyloid precursor protein family. Nat Rev Neurosci *18*, 281–298. 10.1038/nrn.2017.29.

439. Shankar, G.M., Bloodgood, B.L., Townsend, M., Walsh, D.M., Selkoe, D.J., and Sabatini, B.L. (2007). Natural oligomers of the Alzheimer amyloid-beta protein induce reversible synapse loss by modulating an NMDA-type glutamate receptor-dependent signaling pathway. J Neurosci *27*, 2866–2875. 10.1523/JNEUROSCI.4970-06.2007.

440. Wei, W., Nguyen, L.N., Kessels, H.W., Hagiwara, H., Sisodia, S., and Malinow, R. (2010). Amyloid beta from axons and dendrites reduces local spine number and plasticity. Nat Neurosci *13*, 190–196. 10.1038/nn.2476.

441. Walsh, D.M., and Selkoe, D.J. (2007). A beta oligomers - a decade of discovery. J Neurochem *101*, 1172–1184. 10.1111/j.1471-4159.2006.04426.x.

442. Ma, T., and Klann, E. (2012). Amyloid β: linking synaptic plasticity failure to memory disruption in Alzheimer's disease. J Neurochem *120 Suppl 1*, 140–148. 10.1111/j.1471-4159.2011.07506.x.

443. John, A., and Reddy, P.H. (2021). Synaptic basis of Alzheimer's disease: Focus on synaptic amyloid beta, P-tau and mitochondria. Ageing Research Reviews *65*, 101208. 10.1016/j.arr.2020.101208.

444. Cacace, R., Sleegers, K., and Van Broeckhoven, C. (2016). Molecular genetics of early-onset Alzheimer's disease revisited. Alzheimer's & Dementia *12*, 733–748. 10.1016/j.jalz.2016.01.012.

445. Oakley, H., Cole, S.L., Logan, S., Maus, E., Shao, P., Craft, J., Guillozet-Bongaarts, A., Ohno, M., Disterhoft, J., Eldik, L.V., et al. (2006). Intraneuronal β-Amyloid Aggregates, Neurodegeneration, and Neuron Loss in Transgenic Mice with Five Familial Alzheimer's Disease Mutations: Potential Factors in Amyloid Plaque Formation. J. Neurosci. *26*, 10129–10140. 10.1523/JNEUROSCI.1202-06.2006.

446. Goedert, M., Clavaguera, F., and Tolnay, M. (2010). The propagation of prion-like protein inclusions in neurodegenerative diseases. Trends in Neurosciences *33*, 317–325. 10.1016/j.tins.2010.04.003.

447. Jucker, M., and Walker, L.C. (2011). Pathogenic protein seeding in alzheimer disease and other neurodegenerative disorders. Annals of Neurology *70*, 532–540. 10.1002/ana.22615.

448. Rodrigue, K.M., Kennedy, K.M., Devous, M.D., Rieck, J.R., Hebrank, A.C., Diaz-Arrastia, R., Mathews, D., and Park, D.C. (2012). β-Amyloid burden in healthy aging. Neurology *78*, 387–395. 10.1212/WNL.0b013e318245d295.

449. Kubo, A., Misonou, H., Matsuyama, M., Nomori, A., Wada-Kakuda, S., Takashima, A., Kawata, M., Murayama, S., Ihara, Y., and Miyasaka, T. (2019). Distribution of endogenous normal tau in the mouse brain. J Comp Neurol *527*, 985–998. 10.1002/cne.24577.

450. Yu, W., and Lu, B. (2012). Synapses and Dendritic Spines as Pathogenic Targets in Alzheimer's Disease. Neural Plast *2012*, 247150. 10.1155/2012/247150.

451. Naseri, N.N., Wang, H., Guo, J., Sharma, M., and Luo, W. (2019). The complexity of tau in Alzheimer's disease. Neuroscience Letters *705*, 183–194. 10.1016/j.neulet.2019.04.022.

452. Nagy, Z., Esiri, M.M., Jobst, K.A., Morris, J.H., King, E.M.-F., McDonald, B., Litchfield, S., Smith, A., Barnetson, L., and Smith, A.D. (1995). Relative Roles of Plaques and Tangles in the Dementia of Alzheimer's Disease: Correlations Using Three Sets of Neuropathological Criteria. DEM *6*, 21–31. 10.1159/000106918.

453. He, Z., Guo, J.L., McBride, J.D., Narasimhan, S., Kim, H., Changolkar, L., Zhang, B., Gathagan, R.J., Yue, C., Dengler, C., et al. (2018). Amyloid-β plaques enhance Alzheimer's brain tau-seeded pathologies by facilitating neuritic plaque tau aggregation. Nat Med *24*, 29–38. 10.1038/nm.4443.

454. Zempel, H., Thies, E., Mandelkow, E., and Mandelkow, E.-M. (2010). Aβ Oligomers Cause Localized Ca2+ Elevation, Missorting of Endogenous Tau into Dendrites, Tau Phosphorylation, and Destruction of Microtubules and Spines. J. Neurosci. *30*, 11938–11950. 10.1523/JNEUROSCI.2357-10.2010.

455. Mazanetz, M.P., and Fischer, P.M. (2007). Untangling tau hyperphosphorylation in drug design for neurodegenerative diseases. Nat Rev Drug Discov *6*, 464–479. 10.1038/nrd2111.

456. Martin, L., Latypova, X., Wilson, C.M., Magnaudeix, A., Perrin, M.-L., and Terro, F. (2013). Tau protein phosphatases in Alzheimer's disease: The leading role of PP2A. Ageing Research Reviews *12*, 39–49. 10.1016/j.arr.2012.06.008.

457. Ingelsson, M., Fukumoto, H., Newell, K.L., Growdon, J.H., Hedley-Whyte, E.T., Frosch, M.P., Albert, M.S., Hyman, B.T., and Irizarry, M.C. (2004). Early Abeta accumulation and progressive synaptic loss, gliosis, and tangle formation in AD brain. Neurology *62*, 925–931. 10.1212/01.wnl.0000115115.98960.37.

458. Barger, S.W., and Harmon, A.D. (1997). Microglial activation by Alzheimer amyloid precursor protein and modulation by apolipoprotein E. Nature *388*, 878–881. 10.1038/42257.

459. Sastre, M., Klockgether, T., and Heneka, M.T. (2006). Contribution of inflammatory processes to Alzheimer's disease: molecular mechanisms. International Journal of Developmental Neuroscience *24*, 167–176. 10.1016/j.ijdevneu.2005.11.014.

460. Heneka, M.T., Carson, M.J., Khoury, J.E., Landreth, G.E., Brosseron, F., Feinstein, D.L., Jacobs, A.H., Wyss-Coray, T., Vitorica, J., Ransohoff, R.M., et al. (2015). Neuroinflammation in Alzheimer's disease. The Lancet Neurology *14*, 388–405. 10.1016/S1474-4422(15)70016-5.

461. Yeh, F.L., Wang, Y., Tom, I., Gonzalez, L.C., and Sheng, M. (2016). TREM2 Binds to Apolipoproteins, Including APOE and CLU/APOJ, and Thereby Facilitates Uptake of Amyloid-Beta by Microglia. Neuron *91*, 328–340. 10.1016/j.neuron.2016.06.015.

462. Draber, P., Vonkova, I., Stepanek, O., Hrdinka, M., Kucova, M., Skopcova, T., Otahal, P., Angelisova, P., Horejsi, V., Yeung, M., et al. (2011). SCIMP, a transmembrane adaptor protein involved in major histocompatibility complex class II signaling. Mol Cell Biol *31*, 4550–4562. 10.1128/MCB.05817-11.

463. Liu, Z., Gu, Y., Chakarov, S., Bleriot, C., Kwok, I., Chen, X., Shin, A., Huang, W., Dress, R.J., Dutertre, C.-A., et al. (2019). Fate Mapping via Ms4a3-Expression History Traces Monocyte-Derived Cells. Cell *178*, 1509-1525.e19. 10.1016/j.cell.2019.08.009.

464. Spielman, L.J., Little, J.P., and Klegeris, A. (2014). Inflammation and insulin/IGF-1 resistance as the possible link between obesity and neurodegeneration. Journal of Neuroimmunology *273*, 8–21. 10.1016/j.jneuroim.2014.06.004.

465. Holmes, C., Cunningham, C., Zotova, E., Woolford, J., Dean, C., Kerr, S., Culliford, D., and Perry, V.H. (2009). Systemic inflammation and disease progression in Alzheimer disease. Neurology *73*, 768–774. 10.1212/WNL.0b013e3181b6bb95.

466. Mészáros, Á., Molnár, K., Nógrádi, B., Hernádi, Z., Nyúl-Tóth, Á., Wilhelm, I., and Krizbai, I.A. (2020). Neurovascular Inflammaging in Health and Disease. Cells *9*, 1614. 10.3390/cells9071614.

467. Giunta, B., Fernandez, F., Nikolic, W.V., Obregon, D., Rrapo, E., Town, T., and Tan, J. (2008). Inflammaging as a prodrome to Alzheimer's disease. Journal of Neuroinflammation *5*, 51. 10.1186/1742-2094-5-51.

468. Brini, M., Calì, T., Ottolini, D., and Carafoli, E. (2014). Neuronal calcium signaling: function and dysfunction. Cell Mol Life Sci *71*, 2787–2814. 10.1007/s00018-013-1550-7.

469. Hudmon, A., and Schulman, H. (2002). Neuronal CA2+/calmodulin-dependent protein kinase II: the role of structure and autoregulation in cellular function. Annu Rev Biochem *71*, 473–510. 10.1146/annurev.biochem.71.110601.135410.

470. LaFerla, F.M. (2002). Calcium dyshomeostasis and intracellular signalling in alzheimer's disease. Nat Rev Neurosci *3*, 862–872. 10.1038/nrn960.

471. Lopez, J.R., Lyckman, A., Oddo, S., LaFerla, F.M., Querfurth, H.W., and Shtifman, A. (2008). Increased intraneuronal resting [Ca2+] in adult Alzheimer's disease mice. Journal of Neurochemistry *105*, 262–271. 10.1111/j.1471-4159.2007.05135.x.

472. Kuchibhotla, K.V., Goldman, S.T., Lattarulo, C.R., Wu, H.-Y., Hyman, B.T., and Bacskai, B.J. (2008). Aβ Plaques Lead to Aberrant Regulation of Calcium Homeostasis In Vivo Resulting in Structural and Functional Disruption of Neuronal Networks. Neuron *59*, 214–225. 10.1016/j.neuron.2008.06.008.

473. Mattson, M.P., Cheng, B., Davis, D., Bryant, K., Lieberburg, I., and Rydel, R.E. (1992). beta-Amyloid peptides destabilize calcium homeostasis and render human cortical neurons vulnerable to excitotoxicity. J Neurosci *12*, 376–389.

474. Kawahara, M., and Kuroda, Y. (2000). Molecular mechanism of neurodegeneration induced by Alzheimer's beta-amyloid protein: channel formation and disruption of calcium homeostasis. Brain Res Bull *53*, 389–397. 10.1016/s0361-9230(00)00370-1.

475. Kagan, B.L., Hirakura, Y., Azimov, R., Azimova, R., and Lin, M.-C. (2002). The channel hypothesis of Alzheimer's disease: current status. Peptides *23*, 1311–1315. 10.1016/s0196-9781(02)00067-0.

476. Price, S.A., Held, B., and Pearson, H.A. (1998). Amyloid beta protein increases Ca2+ currents in rat cerebellar granule neurones. Neuroreport *9*, 539–545.

477. Dreses-Werringloer, U., Lambert, J.-C., Vingtdeux, V., Zhao, H., Vais, H., Siebert, A., Jain, A., Koppel, J., Rovelet-Lecrux, A., Hannequin, D., et al. (2008). A polymorphism in CALHM1 influences Ca2+ homeostasis, Aβlevels, and Alzheimer's disease risk. Cell *133*, 1149–1161. 10.1016/j.cell.2008.05.048.

478. Pierrot, N., Ghisdal, P., Caumont, A.-S., and Octave, J.-N. (2004). Intraneuronal amyloid-β1-42 production triggered by sustained increase of cytosolic calcium concentration induces neuronal death. Journal of Neurochemistry *88*, 1140–1150. 10.1046/j.1471-4159.2003.02227.x.

479. Pierrot, N., Santos, S.F., Feyt, C., Morel, M., Brion, J.-P., and Octave, J.-N. (2006). Calcium-mediated Transient Phosphorylation of Tau and Amyloid Precursor Protein Followed by Intraneuronal Amyloid-β Accumulation *. Journal of Biological Chemistry *281*, 39907–39914. 10.1074/jbc.M606015200.

480. Querfurth, H.W., and Selkoe, D.J. (1994). Calcium Ionophore Increases Amyloid .beta. Peptide Production by Cultured Cells. Biochemistry *33*, 4550–4561. 10.1021/bi00181a016.

481. Mattson, M.P. (1990). Antigenic changes similar to those seen in neurofibrillary tangles are elicited by glutamate and Ca2+ influx in cultured hippocampal neurons. Neuron *4*, 105–117. 10.1016/0896-6273(90)90447-N.

482. Mattson, M.P., Lovell, M.A., Ehmann, W.D., and Markesbery, W.R. (1993). Comparison of the effects of elevated intracellular aluminum and calcium levels on neuronal survival and tau immunoreactivity. Brain Research *602*, 21–31. 10.1016/0006-8993(93)90236-G.

483. Datta, D., Leslie, S.N., Wang, M., Morozov, Y.M., Yang, S., Mentone, S., Zeiss, C., Duque, A., Rakic, P., Horvath, T.L., et al. (2021). Age-related calcium dysregulation linked with tau pathology and impaired cognition in non-human primates. Alzheimer's & Dementia *17*, 920–932. 10.1002/alz.12325.

484. Berridge, M.J. (2010). Calcium hypothesis of Alzheimer's disease. Pflugers Arch - Eur J Physiol *459*, 441–449. 10.1007/s00424-009-0736-1.

485. Oddo, S., Caccamo, A., Shepherd, J.D., Murphy, M.P., Golde, T.E., Kayed, R., Metherate, R., Mattson, M.P., Akbari, Y., and LaFerla, F.M. (2003). Triple-Transgenic Model of Alzheimer's Disease with Plaques and Tangles: Intracellular Aβ and Synaptic Dysfunction. Neuron *39*, 409–421. 10.1016/S0896-6273(03)00434-3.

486. Townsend, M., Shankar, G.M., Mehta, T., Walsh, D.M., and Selkoe, D.J. (2006). Effects of secreted oligomers of amyloid β-protein on hippocampal synaptic plasticity: a potent role for trimers. The Journal of Physiology *572*, 477–492. 10.1113/jphysiol.2005.103754.

487. Walsh, D.M., Klyubin, I., Fadeeva, J.V., Cullen, W.K., Anwyl, R., Wolfe, M.S., Rowan, M.J., and Selkoe, D.J. (2002). Naturally secreted oligomers of amyloid β protein potently inhibit hippocampal long-term potentiation in vivo. Nature *416*, 535–539. 10.1038/416535a.

488. Klyubin, I., Walsh, D.M., Lemere, C.A., Cullen, W.K., Shankar, G.M., Betts, V., Spooner, E.T., Jiang, L., Anwyl, R., Selkoe, D.J., et al. (2005). Amyloid β protein immunotherapy neutralizes Aβ oligomers that disrupt synaptic plasticity in vivo. Nat Med *11*, 556–561. 10.1038/nm1234.

489. Rohn, T.T., Vyas, V., Hernandez-Estrada, T., Nichol, K.E., Christie, L.-A., and Head, E. (2008). Lack of Pathology in a Triple Transgenic Mouse Model of Alzheimer's Disease after Overexpression of the Anti-Apoptotic Protein Bcl-2. J. Neurosci. *28*, 3051–3059. 10.1523/JNEUROSCI.5620-07.2008.

490. Larson, J., Lynch, G., Games, D., and Seubert, P. (1999). Alterations in synaptic transmission and long-term potentiation in hippocampal slices from young and aged PDAPP mice. Brain Research *840*, 23–35. 10.1016/S0006-8993(99)01698-4.

491. Guo, Q., Fu, W., Sopher, B.L., Miller, M.W., Ware, C.B., Martin, G.M., and Mattson, M.P. (1999). Increased vulnerability of hippocampal neurons to excitotoxic necrosis in presenilin-1 mutant knock-in mice. Nat Med *5*, 101–106. 10.1038/4789.

492. Leissring, M.A., Akbari, Y., Fanger, C.M., Cahalan, M.D., Mattson, M.P., and LaFerla, F.M. (2000). Capacitative Calcium Entry Deficits and Elevated Luminal Calcium Content in Mutant Presenilin-1 Knockin Mice. J Cell Biol *149*, 793–798.

493. Rahman, M.A., Rahman, M.S., Uddin, M.J., Mamum-Or-Rashid, A.N.M., Pang, M.-G., and Rhim, H. (2020). Emerging risk of environmental factors: insight mechanisms of Alzheimer's diseases. Environ Sci Pollut Res *27*, 44659–44672. 10.1007/s11356-020-08243-z.

494. Xie, L., Kang, H., Xu, Q., Chen, M.J., Liao, Y., Thiyagarajan, M., O'Donnell, J., Christensen, D.J., Nicholson, C., Iliff, J.J., et al. (2013). Sleep Drives Metabolite Clearance from the Adult Brain. Science *342*, 373–377. 10.1126/science.1241224.

495. Kang, J.-E., Lim, M.M., Bateman, R.J., Lee, J.J., Smyth, L.P., Cirrito, J.R., Fujiki, N., Nishino, S., and Holtzman, D.M. (2009). Amyloid-β Dynamics are Regulated by Orexin and the Sleep-Wake Cycle. Science *326*, 1005–1007. 10.1126/science.1180962.

496. Zhou, J., Yu, J.-T., Wang, H.-F., Meng, X.-F., Tan, C.-C., Wang, J., Wang, C., and Tan, L. (2015). Association between Stroke and Alzheimer's Disease: Systematic Review and Meta-Analysis. Journal of Alzheimer's Disease *43*, 479–489. 10.3233/JAD-140666.

497. Arnold, S.E., Arvanitakis, Z., Macauley-Rambach, S.L., Koenig, A.M., Wang, H.-Y., Ahima, R.S., Craft, S., Gandy, S., Buettner, C., Stoeckel, L.E., et al. (2018). Brain insulin resistance in type 2 diabetes and Alzheimer disease: concepts and conundrums. Nat Rev Neurol *14*, 168–181. 10.1038/nrneurol.2017.185.

498. Cunningham, C., Wilcockson, D.C., Campion, S., Lunnon, K., and Perry, V.H. (2005). Central and Systemic Endotoxin Challenges Exacerbate the Local Inflammatory Response and Increase Neuronal Death during Chronic Neurodegeneration. J. Neurosci. *25*, 9275–9284. 10.1523/JNEUROSCI.2614-05.2005.

499. Cunningham, C., Campion, S., Lunnon, K., Murray, C.L., Woods, J.F.C., Deacon, R.M.J., Rawlins, J.N.P., and Perry, V.H. (2009). Systemic Inflammation Induces Acute Behavioral and Cognitive

Changes and Accelerates Neurodegenerative Disease. Biological Psychiatry *65*, 304–312. 10.1016/j.biopsych.2008.07.024.

500. Field, R., Campion, S., Warren, C., Murray, C., and Cunningham, C. (2010). Systemic challenge with the TLR3 agonist poly I:C induces amplified IFNα/β and IL-1β responses in the diseased brain and exacerbates chronic neurodegeneration. Brain, Behavior, and Immunity *24*, 996–1007. 10.1016/j.bbi.2010.04.004.

501. Holmes, C., Cunningham, C., Zotova, E., Culliford, D., and Perry, V.H. (2011). Proinflammatory cytokines, sickness behavior, and Alzheimer disease. Neurology *77*, 212–218. 10.1212/WNL.0b013e318225ae07.

502. Chou, R.C., Kane, M., Ghimire, S., Gautam, S., and Gui, J. (2016). Treatment for Rheumatoid Arthritis and Risk of Alzheimer's Disease: A Nested Case-Control Analysis. CNS Drugs *30*, 1111–1120. 10.1007/s40263-016-0374-z.

503. Llanos-González, E., Henares-Chavarino, Á.A., Pedrero-Prieto, C.M., García-Carpintero, S., Frontiñán-Rubio, J., Sancho-Bielsa, F.J., Alcain, F.J., Peinado, J.R., Rabanal-Ruíz, Y., and Durán-Prado, M. (2020). Interplay Between Mitochondrial Oxidative Disorders and Proteostasis in Alzheimer's Disease. Front. Neurosci. *0*. 10.3389/fnins.2019.01444.

504. Myung, N.-H., Zhu, X., Kruman, I.I., Castellani, R.J., Petersen, R.B., Siedlak, S.L., Perry, G., Smith, M.A., and Lee, H. (2008). Evidence of DNA damage in Alzheimer disease: phosphorylation of histone H2AX in astrocytes. AGE *30*, 209–215. 10.1007/s11357-008-9050-7.

505. Coppede, F., and Migliore, L. (2009). DNA Damage and Repair in Alzheimer's Disease. 10.2174/156720509787313970.

506. Zhang, P., Kishimoto, Y., Grammatikakis, L., Gottimukkala, K., Cutler, R.G., Zhang, S., Abdelmohsen, K., Bohr, V.A., Sen, J.M., Gorospe, M., et al. (2019). Senolytic therapy alleviates Aβ-associated oligodendrocyte progenitor cell senescence and cognitive deficits in an Alzheimer's disease model. Nature neuroscience *22*, 719. 10.1038/s41593-019-0372-9.

507. Edbauer, D., Neilson, J.R., Foster, K.A., Wang, C.-F., Seeburg, D.P., Batterton, M.N., Tada, T., Dolan, B.M., Sharp, P.A., and Sheng, M. (2010). Regulation of synaptic structure and function by FMRP-associated microRNAs miR-125b and miR-132. Neuron *65*, 373–384. 10.1016/j.neuron.2010.01.005.

508. Bekris, L.M., and Leverenz, J.B. (2015). The biomarker and therapeutic potential of miRNA in Alzheimer's disease. Neurodegener Dis Manag *5*, 61–74. 10.2217/nmt.14.52.

509. Banzhaf-Strathmann, J., Benito, E., May, S., Arzberger, T., Tahirovic, S., Kretzschmar, H., Fischer, A., and Edbauer, D. (2014). MicroRNA-125b induces tau hyperphosphorylation and cognitive deficits in Alzheimer's disease. EMBO J *33*, 1667–1680. 10.15252/embj.201387576.

510. Gatz, M., Reynolds, C.A., Fratiglioni, L., Johansson, B., Mortimer, J.A., Berg, S., Fiske, A., and Pedersen, N.L. (2006). Role of Genes and Environments for Explaining Alzheimer Disease. Archives of General Psychiatry *63*, 168–174. 10.1001/archpsyc.63.2.168.

511. Wightman, D.P., Jansen, I.E., Savage, J.E., Shadrin, A.A., Bahrami, S., Holland, D., Rongve, A., Børte, S., Winsvold, B.S., Drange, O.K., et al. (2021). A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. Nat Genet *53*, 1276–1282. 10.1038/s41588-021-00921-z.

512. de Rojas, I., Moreno-Grau, S., Tesi, N., Grenier-Boley, B., Andrade, V., Jansen, I.E., Pedersen, N.L., Stringa, N., Zettergren, A., Hernández, I., et al. (2021). Common variants in Alzheimer's disease and risk stratification by polygenic risk scores. Nat Commun *12*, 3417. 10.1038/s41467-021-22491-8.

513. Corder, E.H., Saunders, A.M., Strittmatter, W.J., Schmechel, D.E., Gaskell, P.C., Small, G.W., Roses, A.D., Haines, J.L., and Pericak-Vance, M.A. (1993). Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science *261*, 921–923. 10.1126/science.8346443.

514. Farrer, L.A., Cupples, L.A., Haines, J.L., Hyman, B., Kukull, W.A., Mayeux, R., Myers, R.H., Pericak-Vance, M.A., Risch, N., and van Duijn, C.M. (1997). Effects of Age, Sex, and Ethnicity on the Association Between Apolipoprotein E Genotype and Alzheimer Disease: A Meta-analysis. JAMA *278*, 1349–1356. 10.1001/jama.1997.03550160069041.

515. Jiang, Q., Lee, C.Y.D., Mandrekar, S., Wilkinson, B., Cramer, P., Zelcer, N., Mann, K., Lamb, B., Willson, T.M., Collins, J.L., et al. (2008). ApoE promotes the proteolytic degradation of Abeta. Neuron *58*, 681–693. 10.1016/j.neuron.2008.04.010.

516. Castellano, J.M., Kim, J., Stewart, F.R., Jiang, H., DeMattos, R.B., Patterson, B.W., Fagan, A.M., Morris, J.C., Mawuenyega, K.G., Cruchaga, C., et al. (2011). Human apoE isoforms differentially regulate brain amyloid-β peptide clearance. Sci Transl Med *3*, 89ra57. 10.1126/scitranslmed.3002156.

517. Calafate, S., Flavin, W., Verstreken, P., and Moechars, D. (2016). Loss of Bin1 Promotes the Propagation of Tau Pathology. Cell Reports *17*, 931–940. 10.1016/j.celrep.2016.09.063.

518. Miyagawa, T., Ebinuma, I., Morohashi, Y., Hori, Y., Young Chang, M., Hattori, H., Maehara, T., Yokoshima, S., Fukuyama, T., Tsuji, S., et al. (2016). BIN1 regulates BACE1 intracellular trafficking and amyloid-β production. Hum Mol Genet *25*, 2948–2958. 10.1093/hmg/ddw146.

519. Hong, T.-T., Smyth, J.W., Gao, D., Chu, K.Y., Vogan, J.M., Fong, T.S., Jensen, B.C., Colecraft, H.M., and Shaw, R.M. (2010). BIN1 Localizes the L-Type Calcium Channel to Cardiac T-Tubules. PLOS Biology *8*, e1000312. 10.1371/journal.pbio.1000312.

520. Yerbury, J.J., Poon, S., Meehan, S., Thompson, B., Kumita, J.R., Dobson, C.M., and Wilson, M.R. (2007). The extracellular chaperone clusterin influences amyloid formation and toxicity by interacting with prefibrillar structures. FASEB J *21*, 2312–2322. 10.1096/fj.06-7986com.

521. Narayan, P., Orte, A., Clarke, R.W., Bolognesi, B., Hook, S., Ganzinger, K.A., Meehan, S., Wilson, M.R., Dobson, C.M., and Klenerman, D. (2012). The extracellular chaperone clusterin sequesters oligomeric forms of the amyloid-β1−40 peptide. Nat Struct Mol Biol *19*, 79–83. 10.1038/nsmb.2191.

522. Hatters, D.M., Wilson, M.R., Easterbrook-Smith, S.B., and Howlett, G.J. (2002). Suppression of apolipoprotein C-II amyloid formation by the extracellular chaperone, clusterin. Eur J Biochem *269*, 2789–2794. 10.1046/j.1432-1033.2002.02957.x.

523. Kumita, J.R., Poon, S., Caddy, G.L., Hagan, C.L., Dumoulin, M., Yerbury, J.J., Stewart, E.M., Robinson, C.V., Wilson, M.R., and Dobson, C.M. (2007). The extracellular chaperone clusterin potently inhibits human lysozyme amyloid formation by interacting with prefibrillar species. J Mol Biol *369*, 157–167. 10.1016/j.jmb.2007.02.095.

524. Wang, H.-F., Wan, Y., Hao, X.-K., Cao, L., Zhu, X.-C., Jiang, T., Tan, M.-S., Tan, L., Zhang, D.-Q., Tan, L., et al. (2016). Bridging Integrator 1 (BIN1) Genotypes Mediate Alzheimer's Disease Risk by Altering Neuronal Degeneration. Journal of Alzheimer's Disease *52*, 179–190. 10.3233/JAD-150972.

525. Harold, D., Abraham, R., Hollingworth, P., Sims, R., Gerrish, A., Hamshere, M.L., Pahwa, J.S., Moskvina, V., Dowzell, K., Williams, A., et al. (2009). Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet *41*, 1088–1093. 10.1038/ng.440.

526. Lambert, J.C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., DeStafano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B., et al. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet *45*, 1452–1458. 10.1038/ng.2802.

527. Morgan, K. (2011). The three new pathways leading to Alzheimer's disease. Neuropathol Appl Neurobiol *37*, 353–357. 10.1111/j.1365-2990.2011.01181.x.

528. Seshadri, S., Fitzpatrick, A.L., Ikram, M.A., DeStefano, A.L., Gudnason, V., Boada, M., Bis, J.C., Smith, A.V., Carassquillo, M.M., Lambert, J.C., et al. (2010). Genome-wide analysis of genetic loci associated with Alzheimer disease. JAMA *303*, 1832–1840. 10.1001/jama.2010.574.

529. Fugier, C., Klein, A.F., Hammer, C., Vassilopoulos, S., Ivarsson, Y., Toussaint, A., Tosch, V., Vignaud, A., Ferry, A., Messaddeq, N., et al. (2011). Misregulated alternative splicing of BIN1 is associated with T tubule alterations and muscle weakness in myotonic dystrophy. Nat Med *17*, 720–725. 10.1038/nm.2374.

530. De Rossi, P., Buggia-Prévot, V., Clayton, B.L.L., Vasquez, J.B., van Sanford, C., Andrew, R.J., Lesnick, R., Botté, A., Deyts, C., Salem, S., et al. (2016). Predominant expression of Alzheimer's disease-associated BIN1 in mature oligodendrocytes and localization to white matter tracts. Molecular Neurodegeneration *11*, 59. 10.1186/s13024-016-0124-1.

531. Adams, S.L., Tilton, K., Kozubek, J.A., Seshadri, S., and Delalle, I. (2016). Subcellular Changes in Bridging Integrator 1 Protein Expression in the Cerebral Cortex During the Progression of Alzheimer Disease Pathology. Journal of Neuropathology & Experimental Neurology *75*, 779–790. 10.1093/jnen/nlw056.

532. Butler, M.H., David, C., Ochoa, G.C., Freyberg, Z., Daniell, L., Grabs, D., Cremona, O., and De Camilli, P. (1997). Amphiphysin II (SH3P9; BIN1), a member of the amphiphysin/Rvs family, is concentrated in the cortical cytomatrix of axon initial segments and nodes of ranvier in brain and around T tubules in skeletal muscle. J Cell Biol *137*, 1355–1367. 10.1083/jcb.137.6.1355.

533. Ellis, J.D., Barrios-Rodiles, M., Colak, R., Irimia, M., Kim, T., Calarco, J.A., Wang, X., Pan, Q., O'Hanlon, D., Kim, P.M., et al. (2012). Tissue-specific alternative splicing remodels protein-protein interaction networks. Mol Cell *46*, 884–892. 10.1016/j.molcel.2012.05.037.

534. Prokic, I., Cowling, B.S., and Laporte, J. (2014). Amphiphysin 2 (BIN1) in physiology and diseases. J Mol Med (Berl) *92*, 453–463. 10.1007/s00109-014-1138-1.

535. Ramjaun, A.R., and McPherson, P.S. (1998). Multiple amphiphysin II splice variants display differential clathrin binding: identification of two distinct clathrin-binding sites. J Neurochem *70*, 2369–2376. 10.1046/j.1471-4159.1998.70062369.x.

536. Shupliakov, O., Löw, P., Grabs, D., Gad, H., Chen, H., David, C., Takei, K., De Camilli, P., and Brodin, L. (1997). Synaptic vesicle endocytosis impaired by disruption of dynamin-SH3 domain interactions. Science *276*, 259–263. 10.1126/science.276.5310.259.

537. Wechsler-Reya, R., Sakamuro, D., Zhang, J., Duhadaway, J., and Prendergast, G.C. (1997). Structural analysis of the human BIN1 gene. Evidence for tissue-specific transcriptional regulation and alternate RNA splicing. J Biol Chem *272*, 31453–31458. 10.1074/jbc.272.50.31453.

538. Wigge, P., Köhler, K., Vallis, Y., Doyle, C.A., Owen, D., Hunt, S.P., and McMahon, H.T. (1997). Amphiphysin heterodimers: potential role in clathrin-mediated endocytosis. Mol Biol Cell *8*, 2003–2015. 10.1091/mbc.8.10.2003.

539. Wu, T., Shi, Z., and Baumgart, T. (2014). Mutations in BIN1 associated with centronuclear myopathy disrupt membrane remodeling by affecting protein density and oligomerization. PLoS One *9*, e93060. 10.1371/journal.pone.0093060.

540. Micheva, K.D., Ramjaun, A.R., Kay, B.K., and McPherson, P.S. (1997). SH3 domain-dependent interactions of endophilin with amphiphysin. FEBS Lett *414*, 308–312. 10.1016/s0014-5793(97)01016-8.

541. Slepnev, V.I., Ochoa, G.C., Butler, M.H., and De Camilli, P. (2000). Tandem arrangement of the clathrin and AP-2 binding domains in amphiphysin 1 and disruption of clathrin coat function by amphiphysin fragments comprising these sites. J Biol Chem *275*, 17583–17589. 10.1074/jbc.M910430199.

542. De Rossi, P., Nomura, T., Andrew, R.J., Masse, N.Y., Sampathkumar, V., Musial, T.F., Sudwarts, A., Recupero, A.J., Le Metayer, T., Hansen, M.T., et al. (2020). Neuronal BIN1 Regulates Presynaptic Neurotransmitter Release and Memory Consolidation. Cell Reports *30*, 3520-3535.e7. 10.1016/j.celrep.2020.02.026.

543. Marques-Coelho, D., Iohan, L. da C.C., Melo de Farias, A.R., Flaig, A., Lambert, J.-C., and Costa, M.R. (2021). Differential transcript usage unravels gene expression alterations in Alzheimer's disease human brains. npj Aging Mech Dis *7*, 1–15. 10.1038/s41514-020-00052-5.

544. Lambert, E., Saha, O., Soares Landeira, B., Melo de Farias, A.R., Hermant, X., Carrier, A., Pelletier, A., Gadaut, J., Davoine, L., Dupont, C., et al. (2022). The Alzheimer susceptibility gene BIN1 induces isoform-dependent neurotoxicity through early endosome defects. Acta Neuropathologica Communications *10*, 4. 10.1186/s40478-021-01285-5.

545. Voskobiynyk, Y., Roth, J.R., Cochran, J.N., Rush, T., Carullo, N.V., Mesina, J.S., Waqas, M., Vollmer, R.M., Day, J.J., McMahon, L.L., et al. (2020). Alzheimer's disease risk gene BIN1 induces Tau-dependent network hyperexcitability. eLife *9*, e57354. 10.7554/eLife.57354.

546. Sartori, M., Mendes, T., Desai, S., Lasorsa, A., Herledan, A., Malmanche, N., Mäkinen, P., Marttinen, M., Malki, I., Chapuis, J., et al. (2019). BIN1 recovers tauopathy-induced long-term memory deficits in mice and interacts with Tau through Thr348 phosphorylation. Acta Neuropathol *138*, 631–652. 10.1007/s00401-019-02017-9.

547. Crotti, A., Sait, H.R., McAvoy, K.M., Estrada, K., Ergun, A., Szak, S., Marsh, G., Jandreski, L., Peterson, M., Reynolds, T.L., et al. (2019). BIN1 favors the spreading of Tau via extracellular vesicles. Sci Rep *9*, 9477. 10.1038/s41598-019-45676-0.

548. Tan, M.-S., Yu, J.-T., and Tan, L. (2013). Bridging integrator 1 (BIN1): form, function, and Alzheimer's disease. Trends in Molecular Medicine *19*, 594–603. 10.1016/j.molmed.2013.06.004.

549. Chapuis, J., Hansmannel, F., Gistelinck, M., Mounier, A., Van Cauwenberghe, C., Kolen, K.V., Geller, F., Sottejeau, Y., Harold, D., Dourlen, P., et al. (2013). Increased expression of BIN1 mediates Alzheimer genetic risk by modulating tau pathology. Mol Psychiatry *18*, 1225–1234. 10.1038/mp.2013.1.

550. Glennon, E.B.C., Whitehouse, I.J., Miners, J.S., Kehoe, P.G., Love, S., Kellett, K.A.B., and Hooper, N.M. (2013). BIN1 Is Decreased in Sporadic but Not Familial Alzheimer's Disease or in Aging. PLOS ONE *8*, e78806. 10.1371/journal.pone.0078806.

551. Holler, C.J., Davis, P.R., Beckett, T.L., Platt, T.L., Webb, R.L., Head, E., and Murphy, M.P. (2014). Bridging integrator 1 (BIN1) protein expression increases in the Alzheimer's disease brain and correlates with neurofibrillary tangle pathology. J Alzheimers Dis *42*, 1221–1227. 10.3233/JAD-132450.

552. Ubelmann, F., Burrinha, T., Salavessa, L., Gomes, R., Ferreira, C., Moreno, N., and Guimas Almeida, C. (2017). Bin1 and CD2AP polarise the endocytic generation of beta-amyloid. EMBO reports *18*, 102–122. 10.15252/embr.201642738.

553. Andrew, R.J., Rossi, P.D., Nguyen, P., Kowalski, H.R., Recupero, A.J., Guerbette, T., Krause, S.V., Rice, R.C., Laury-Kleintop, L., Wagner, S.L., et al. (2019). Reduction of the expression of the late-onset Alzheimer's disease (AD) risk-factor BIN1 does not affect amyloid pathology in an AD mouse model. Journal of Biological Chemistry *294*, 4477–4487. 10.1074/jbc.RA118.006379.

554. Hong, T., Yang, H., Zhang, S.-S., Cho, H.C., Kalashnikova, M., Sun, B., Zhang, H., Bhargava, A., Grabe, M., Olgin, J., et al. (2014). Cardiac BIN1 folds T-tubule membrane, controlling ion flux and limiting arrhythmia. Nat Med *20*, 624–632. 10.1038/nm.3543.

555. Trujillo-Estrada, L., Sanchez-Mejias, E., Sanchez-Varo, R., Garcia-Leon, J.A., Nuñez-Diaz, C., Davila, J.C., Vitorica, J., LaFerla, F.M., Moreno-Gonzalez, I., Gutierrez, A., et al. (2021). Animal and Cellular Models of Alzheimer's Disease: Progress, Promise, and Future Approaches. Neuroscientist, 10738584211001752. 10.1177/10738584211001753.

556. La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498. 10.1038/s41586-018-0414-6.

557. Bergen, V., Lange, M., Peidli, S., Wolf, F.A., and Theis, F.J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. Nat Biotechnol *38*, 1408–1414. 10.1038/s41587-020-0591-3.

558. Sample Prep - Official 10x Genomics Support 10x Genomics. https://www.10xgenomics.com/support/single-cell-multiome-atac-plus-gene-expression/documentation/steps/sample-prep/nuclei-isolation-for-single-cell-multiome-atac-plus-gene-expression-sequencing.

559. Single Cell Multiome ATAC + Gene Expression 10x Genomics. https://www.10xgenomics.com/products/single-cell-multiome-atac-plus-gene-expression.

560. Analysis of Single-Cell Chromatin Data https://stuartlab.org/signac/.

561. Tools for Single Cell Genomics https://satijalab.org/seurat/.

562. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology *15*, 550. 10.1186/s13059-014-0550-8.

563. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation *2*. 10.1016/j.xinn.2021.100141.

564. Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M.N., and Sergushichev, A. (2021). Fast gene set enrichment analysis. 060012. 10.1101/060012.

565. AUCell (2022).

566. Zhang, L., Prak, L., Rayon-Estrada, V., Thiru, P., Flygare, J., Lim, B., and Lodish, H.F. (2013). ZFP36L2 is required for self-renewal of early burst-forming unit erythroid progenitors. Nature *499*, 92–96. 10.1038/nature12215.

567. Lin, N.-Y., Lin, C.-T., and Chang, C.-J. (2008). Modulation of immediate early gene expression by tristetraprolin in the differentiation of 3T3-L1 cells. Biochem Biophys Res Commun *365*, 69–74. 10.1016/j.bbrc.2007.10.119.

568. Dong, D., Tian, Y., Zheng, S.C., and Teschendorff, A.E. (2019). ebGSEA: an improved Gene Set Enrichment Analysis method for Epigenome-Wide-Association Studies. Bioinformatics *35*, 3514–3516. 10.1093/bioinformatics/btz073.

569. Ren, X., and Kuan, P.F. (2019). methylGSA: a Bioconductor package and Shiny app for DNA methylation data length bias adjustment in gene set testing. Bioinformatics *35*, 1958–1959. 10.1093/bioinformatics/bty892.

570. Maksimovic, J., Oshlack, A., and Phipson, B. (2021). Gene set enrichment analysis for genome-wide DNA methylation data. Genome Biology *22*, 173. 10.1186/s13059-021-02388-x.

571. Goessling, W., North, T.E., Lord, A.M., Ceol, C., Lee, S., Weidinger, G., Bourque, C., Strijbosch, R., Haramis, A.-P., Puder, M., et al. (2008). APC mutant zebrafish uncover a changing temporal requirement for wnt signaling in liver development. Dev Biol *320*, 161–174. 10.1016/j.ydbio.2008.05.526.

572. Weidinger, G., Thorpe, C.J., Wuennenberg-Stapleton, K., Ngai, J., and Moon, R.T. (2005). The Sp1-related transcription factors sp5 and sp5-like act downstream of Wnt/beta-catenin signaling in mesoderm and neuroectoderm patterning. Curr Biol *15*, 489–500. 10.1016/j.cub.2005.01.041.

573. Scheibner, K., Bakhti, M., Bastidas-Ponce, A., and Lickert, H. (2019). Wnt signaling: implications in endoderm development and pancreas organogenesis. Current Opinion in Cell Biology *61*, 48–55. 10.1016/j.ceb.2019.07.002.

574. Li, Y., Wang, L., Zhang, M., Huang, K., Yao, Z., Rao, P., Cai, X., and Xiao, J. (2020). Advanced glycation end products inhibit the osteogenic differentiation potential of adipose-derived stem cells by modulating Wnt/β-catenin signalling pathway via DNA methylation. Cell Proliferation *53*, e12834. 10.1111/cpr.12834.

575. Peng, S., Shi, S., Tao, G., Li, Y., Xiao, D., Wang, L., He, Q., Cai, X., and Xiao, J. (2021). JKAMP inhibits the osteogenic capacity of adipose-derived stem cells in diabetic osteoporosis by modulating the

Wnt signaling pathway through intragenic DNA methylation. Stem Cell Res Ther *12*, 120. 10.1186/s13287-021-02163-6.

576. Zhong, Z., Feng, S., Duttke, S.H., Potok, M.E., Zhang, Y., Gallego-Bartolomé, J., Liu, W., and Jacobsen, S.E. (2021). DNA methylation-linked chromatin accessibility affects genomic architecture in Arabidopsis. Proceedings of the National Academy of Sciences *118*, e2023347118. 10.1073/pnas.2023347118.

577. Min, I.M., Pietramaggiori, G., Kim, F.S., Passegué, E., Stevenson, K.E., and Wagers, A.J. (2008). The Transcription Factor EGR1 Controls Both the Proliferation and Localization of Hematopoietic Stem Cells. Cell Stem Cell *2*, 380–391. 10.1016/j.stem.2008.01.015.

578. Bahrami, S., and Drabløs, F. (2016). Gene regulation in the immediate-early response process. Advances in Biological Regulation *62*, 37–49. 10.1016/j.jbior.2016.05.001.

579. Jiang, J., Chan, Y.-S., Loh, Y.-H., Cai, J., Tong, G.-Q., Lim, C.-A., Robson, P., Zhong, S., and Ng, H.-H. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. Nat Cell Biol *10*, 353–360. 10.1038/ncb1698.

580. Takahashi, K., and Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell *126*, 663–676. 10.1016/j.cell.2006.07.024.

581. Feinberg, M.W., Wara, A.K., Cao, Z., Lebedeva, M.A., Rosenbauer, F., Iwasaki, H., Hirai, H., Katz, J.P., Haspel, R.L., Gray, S., et al. (2007). The Kruppel-like factor KLF4 is a critical regulator of monocyte differentiation. The EMBO Journal *26*, 4138–4148. 10.1038/sj.emboj.7601824.

582. Karamitros, D., Stoilova, B., Aboukhalil, Z., Hamey, F., Reinisch, A., Samitsch, M., Quek, L., Otto, G., Repapi, E., Doondeea, J., et al. (2018). Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. Nat. Immunol. *19*, 85–97. 10.1038/s41590-017-0001-2.

583. Zheng, S., Papalexi, E., Butler, A., Stephenson, W., and Satija, R. (2018). Molecular transitions in early progenitors during human cord blood hematopoiesis. Mol. Syst. Biol. *14*, e8041. 10.15252/msb.20178041.

584. Cheng, H., Zheng, Z., and Cheng, T. (2020). New paradigms on hematopoietic stem cell differentiation. Protein Cell *11*, 34–44. 10.1007/s13238-019-0633-0.

585. Ranzoni, A.M., Tangherloni, A., Berest, I., Riva, S.G., Myers, B., Strzelecka, P.M., Xu, J., Panada, E., Mohorianu, I., Zaugg, J.B., et al. (2021). Integrative Single-Cell RNA-Seq and ATAC-Seq Analysis of Human Developmental Hematopoiesis. Cell Stem Cell *28*, 472-487.e7. 10.1016/j.stem.2020.11.015.

586. Santaguida, M., Schepers, K., King, B., Sabnis, A.J., Forsberg, E.C., Attema, J.L., Braun, B.S., and Passegué, E. (2009). JunB Protects against Myeloid Malignancies by Limiting Hematopoietic Stem Cell Proliferation and Differentiation without Affecting Self-Renewal. Cancer Cell *15*, 341–352. 10.1016/j.ccr.2009.02.016.

587. Anwar, T., Sen, B., Aggarwal, S., Nath, R., Pathak, N., Katoch, A., Aiyaz, M., Trehanpati, N., Khosla, S., and Ramakrishna, G. (2018). Differentially regulated gene expression in quiescence versus senescence and identification of ARID5A as a quiescence associated marker. Journal of Cellular Physiology *233*, 3695–3712. 10.1002/jcp.26227.

588. Alder, J.K., Georgantas, R.W., III, Hildreth, R.L., Yu, X., and Civin, C.I. (2006). Kruppel-Like Factor 4 Upregulates p21 and Downregulates Proliferation of Human and Mouse HSPCs, but Is Not Essential for Mouse HSPC Repopulation. Blood *108*, 1317. 10.1182/blood.V108.11.1317.1317.

589. Dekker, R.J., Boon, R.A., Rondaij, M.G., Kragt, A., Volger, O.L., Elderkamp, Y.W., Meijers, J.C.M., Voorberg, J., Pannekoek, H., and Horrevoets, A.J.G. (2006). KLF2 provokes a gene expression pattern that establishes functional quiescent differentiation of the endothelium. Blood *107*, 4354–4363. 10.1182/blood-2005-08-3465.

590. Sun, Z., Xu, X., He, J., Murray, A., Sun, M., Wei, X., Wang, X., McCoig, E., Xie, E., Jiang, X., et al. (2019). EGR1 recruits TET1 to shape the brain methylome during development and upon neuronal activity. Nat Commun *10*, 3892. 10.1038/s41467-019-11905-3.

591. Denisenko, E., Guo, B.B., Jones, M., Hou, R., de Kock, L., Lassmann, T., Poppe, D., Clément, O., Simmons, R.K., Lister, R., et al. (2020). Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. Genome Biol *21*, 130. 10.1186/s13059-020-02048-6.

592. Bernad, A., Kopf, M., Kulbacki, R., Weich, N., Koehler, G., and Gutierrez-Ramos, J.C. (1994). Interleukin-6 is required in vivo for the regulation of stem cells and committed progenitors of the hematopoietic system. Immunity *1*, 725–731. 10.1016/S1074-7613(94)80014-6.

593. Mihara, M., Hashizume, M., Yoshida, H., Suzuki, M., and Shiina, M. (2011). IL-6/IL-6 receptor system and its role in physiological and pathological conditions. Clinical Science *122*, 143–159. 10.1042/CS20110340.

594. Tie, R., Li, H., Cai, S., Liang, Z., Shan, W., Wang, B., Tan, Y., Zheng, W., and Huang, H. (2019). Interleukin-6 signaling regulates hematopoietic stem cell emergence. Exp Mol Med *51*, 1–12. 10.1038/s12276-019-0320-5.

595. Fernandez-Twinn, D.S., Wayman, A., Ekizoglou, S., Martin, M.S., Hales, C.N., and Ozanne, S.E. (2005). Maternal protein restriction leads to hyperinsulinemia and reduced insulin-signaling protein expression in 21-mo-old female rat offspring. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology *288*, R368–R373. 10.1152/ajpregu.00206.2004.

596. Ozanne, S.E., Jensen, C.B., Tingey, K.J., Storgaard, H., Madsbad, S., and Vaag, A.A. (2005). Low birthweight is associated with specific changes in muscle insulin-signalling protein expression. Diabetologia *48*, 547–552. 10.1007/s00125-005-1669-7.

597. Scala, S., and Aiuti, A. (2019). In vivo dynamics of human hematopoietic stem cells: novel concepts and future directions. Blood Advances *3*, 1916–1924. 10.1182/bloodadvances.2019000039.

598. Seita, J., and Weissman, I.L. (2010). Hematopoietic stem cell: self-renewal versus differentiation. WIREs Systems Biology and Medicine *2*, 640–653. 10.1002/wsbm.86.

599. Kopp, H.-G., Avecilla, S.T., Hooper, A.T., and Rafii, S. (2005). The Bone Marrow Vascular Niche: Home of HSC Differentiation and Mobilization. Physiology *20*, 349–356. 10.1152/physiol.00025.2005.

600. Kumar, S., and Geiger, H. (2017). HSC Niche Biology and HSC Expansion Ex Vivo. Trends in Molecular Medicine *23*, 799–819. 10.1016/j.molmed.2017.07.003.

601. Sureshchandra, S., Chan, C.N., Robino, J.J., Parmelee, L.K., Nash, M.J., Wesolowski, S.R., Pietras, E.M., Friedman, J.E., Takahashi, D., Shen, W., et al. (2022). Maternal Western-Style Diet Impairs Bone Marrow Development and Drives a Hyperinflammatory Phenotype in Hematopoietic Stem and Progenitor Cells in Fetal Rhesus Macaques. 2021.04.26.441556. 10.1101/2021.04.26.441556.

602. Bogeska, R., Mikecin, A.-M., Kaschutnig, P., Fawaz, M., Büchler-Schäff, M., Le, D., Ganuza, M., Vollmer, A., Paffenholz, S.V., Asada, N., et al. (2022). Inflammatory exposure drives long-lived impairment of hematopoietic stem cell self-renewal activity and accelerated aging. Cell Stem Cell *29*, 1273-1284.e8. 10.1016/j.stem.2022.06.012.

603. Moore, S.J., and Murphy, G.G. (2020). The role of L-type calcium channels in neuronal excitability and aging. Neurobiology of Learning and Memory *173*, 107230. 10.1016/j.nlm.2020.107230.

604. Camandola, S., and Mattson, M.P. (2011). Aberrant subcellular neuronal calcium regulation in aging and Alzheimer's disease. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research *1813*, 965–973. 10.1016/j.bbamcr.2010.10.005.

605. Targa Dias Anastacio, H., Matosin, N., and Ooi, L. (2022). Neuronal hyperexcitability in Alzheimer's disease: what are the drivers behind this aberrant phenotype? Transl Psychiatry *12*, 1–14. 10.1038/s41398-022-02024-7.

606. Thibault, O., and Landfield, P.W. (1996). Increase in Single L-Type Calcium Channels in Hippocampal Neurons During Aging. Science *272*, 1017–1020. 10.1126/science.272.5264.1017.

607. Hemond, P., and Jaffe, D.B. (2005). Caloric restriction prevents aging-associated changes in spike-mediated Ca2+ accumulation and the slow afterhyperpolarization in hippocampal CA1 pyramidal neurons. Neuroscience *135*, 413–420. 10.1016/j.neuroscience.2005.05.044.

608. Green, E.M., Barrett, C.F., Bultynck, G., Shamah, S.M., and Dolmetsch, R.E. (2007). The Tumor Suppressor eIF3e Mediates Calcium-Dependent Internalization of the L-Type Calcium Channel CaV1.2. Neuron *55*, 615–632. 10.1016/j.neuron.2007.07.024.

609. Gail Canter, R., Huang, W.-C., Choi, H., Wang, J., Ashley Watson, L., Yao, C.G., Abdurrob, F., Bousleiman, S.M., Young, J.Z., Bennett, D.A., et al. (2019). 3D mapping reveals network-specific amyloid progression and subcortical susceptibility in mice. Commun Biol *2*, 1–12. 10.1038/s42003-019-0599-8.

610. Huang, Z., Richmond, T.D., Muntean, A.G., Barber, D.L., Weiss, M.J., and Crispino, J.D. (2007). STAT1 promotes megakaryopoiesis downstream of GATA-1 in mice. J Clin Invest *117*, 3890–3899. 10.1172/JCI33010.

611. Rodriguez-Fraticelli, A.E., Wolock, S.L., Weinreb, C.S., Panero, R., Patel, S.H., Jankovic, M., Sun, J., Calogero, R.A., Klein, A.M., and Camargo, F.D. (2018). Clonal analysis of lineage fate in native haematopoiesis. Nature *553*, 212–216. 10.1038/nature25168.

612. Popescu, D.-M., Botting, R.A., Stephenson, E., Green, K., Webb, S., Jardine, L., Calderbank, E.F., Polanski, K., Goh, I., Efremova, M., et al. (2019). Decoding human fetal liver haematopoiesis. Nature *574*, 365–371. 10.1038/s41586-019-1652-y.

613. Zhang, P., Li, X., Pan, C., Zheng, X., Hu, B., Xie, R., Hu, J., Shang, X., and Yang, H. (2022). Single-cell RNA sequencing to track novel perspectives in HSC heterogeneity. Stem Cell Research & Therapy *13*, 39. 10.1186/s13287-022-02718-1.

614. Cortal, A., Martignetti, L., Six, E., and Rausell, A. (2021). Gene signature extraction and cell identity recognition at the single-cell level with Cell-ID. Nat Biotechnol *39*, 1095–1102. 10.1038/s41587-021-00896-6.

615. Domcke, S., Bardet, A.F., Adrian Ginno, P., Hartl, D., Burger, L., and Schübeler, D. (2015). Competition between DNA methylation and transcription factors determines binding of NRF1. Nature *528*, 575–579. 10.1038/nature16462.

616. Ley, T.J., Ding, L., Walter, M.J., McLellan, M.D., Lamprecht, T., Larson, D.E., Kandoth, C., Payton, J.E., Baty, J., Welch, J., et al. (2010). DNMT3A Mutations in Acute Myeloid Leukemia. N Engl J Med *363*, 2424–2433. 10.1056/NEJMoa1005143.

617. Liu, X.S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R.A., and Jaenisch, R. (2016). Editing DNA Methylation in the Mammalian Genome. Cell *167*, 233-247.e17. 10.1016/j.cell.2016.08.056.

618. Liau, L.L., Ruszymah, B.H.I., Ng, M.H., and Law, J.X. (2020). Characteristics and clinical applications of Wharton's jelly-derived mesenchymal stromal cells. Curr Res Transl Med *68*, 5–16. 10.1016/j.retram.2019.09.001.

619. Hass, R., Kasper, C., Böhm, S., and Jacobs, R. (2011). Different populations and sources of human mesenchymal stem cells (MSC): A comparison of adult and neonatal tissue-derived MSC. Cell Commun Signal *9*, 12. 10.1186/1478-811X-9-12.