PhD degree in Systems Medicine (Curriculum in Molecular Oncology)

European School of Molecular Medicine (SEMM)

University of Milan “La Statale” and University of Naples “Federico II”

Settore disciplinare: MED/04

# Patient-specific Modelling of Syndromic Autism: Uncovering the Role of ADNP in Chromatin Dysregulation

*Ludovico Rizzuti*

European Institute of Oncology

University of Milan

Matricola n. R12114

Tutor*: Prof. Giuseppe Testa*

European Institute of Oncology, Milan

University of Milan, Milan

Human Technopole, Milan

PhD Coordinator*: Prof. Saverio Minucci*

Anno accademico 2020/2021

# TABLE OF CONTENTS

# LIST OF ABBREVIATION

**2D:** Two-dimensional
**3D**: Three-dimensional
**ADNP:** Activity-dependent neuroprotective protein
**ANR**: Anterior neural ridge
**aRG:** Apical radial glia
**ASD**: Autism spectrum disorder
**ATAC-seq:** Assay for transposase-accessible chromatin using sequencing
**BAF**: BRG1/BRM-associated factors
**BMP:** Bone morphogenetic protein
**bRG**: Basal radial glia
**CAS:** CRISPR associated genes
**CGI**: CpG island
**CHD**: Chromodomain-helicase-DNA-binding protein
**ChIP-seq:** ChIP-sequencing
**CNS:** Central nervous system
**CNV**: Copy-number variant
**CP:** Cortical plate
**CRISPR:** Clustered regularly interspaced short palindromic repeats
**DEGs:** Differentially expressed genes
**DNMTs:** DNA methyltransferases
**EGF:** Epidermal growth factor
**eQTL**: expression quantitative trait loci
**ESCs:** Embryonic stem cells
**FACS:** Fluorescence activated cell sorting
**FDR:** False discovery rate
**FGF:** Fibroblast growth factor
**GABA**: gamma-Aminobutyric acid
**GO:** Gene ontology
**GWAS**: Genome-wide association study
**H3K4me:** Mono-methylation of lysine 4 on histone H3
**H3K4me3:** Tri-methylation of lysine 4 on histone H3
**H3K9me3**: Tri-methylation of lysine 9 on histone H3
**H3K27ac:** Acetylation of lysine 27 on histone H3
**H3K27me1/2/3:** Mono-di-tri-methylation of the lysine 27 on histone H3
**H3K36me3:** Tri-methylation of lysine 36 on histone H3
**HDAC**: Histone deacetylase
**HDR:** Homologous directed recombination
**hESC**: human embryonic stem cell
**hiPSC**: human induced pluripotent stem cells
**hPSC**: human pluripotent stem cell
**HOX:** Homeobox
**ICM**: Inner cell mass
**ID:** Intellectual disability

**IP:** Intermediate progenitors
**iPSC:** Induced pluripotent stem cells
**iSVZ:** Inner subventricular zone
**IZ**: Intermediate zone
**KDM**: Lysine demethylase
**KO:** Knock out
**LGE**: Lateral ganglionic eminence
**LOF:** Loss of function
**MGE**: Medial ganglionic eminence
**MZ**: Marginal zone
**NCSC:** neural crest stem cells
**NHEJ**: Non homologous end joining
**NLS**: Nuclear localization signal
**NMD**: Nonsense mediated decay
**NPC**: Neural progenitor cell
**NuRD**: Nucleosome remodelling deacetylase
**oRG:** Outer radial glia
**OSKM:** Oct3/4, Sox2, Klf4 and c-Myc
**oSVZ:** Outer subventricular zone
**PAM:** protospacer adjacent motif
**PCA:** Principal component analysis
**PCW**: Post-conceptional week
**PcG:** Polycomb group of proteins
**PRC2:** Polycomb repressive complex 2
**PSC:** Pluripotent stem cells
**RG:** radial glia
**RNA-seq:** RNA sequencing
**RNP**: Ribonucleoprotein
**RPKM:** Reads per kilobase of transcript per million mapped reads
**SCNT**: Somatic cell nuclear transfer
**scRNAseq:** single-cell RNAseq
**sgRNA:** single-guide RNA
**Shh:** Sonic hedge-hog
**SNV**: Single-nucleotide variants
**SOX:** SRY-related HMG-box
**SP**: Sub-plate
**SVZ:** Subventricular zone
**SWI/SNF**: SWItch/Sucrose non-fermentable
**TAD**: Topologically associated domain
**TF:** Transcription factor
**TGF-β**: Transforming growth factor beta
**TrxG:** Tritorax group of proteins
**TSS**: Transcription start site
**vRG:** Ventral radial glia signature
**VZ:** Ventricular Zone
**ZFNs:** Zinc-finger nucleases (ZFNs)

# FIGURE INDEX

# TABLE INDEX

# 1. ABSTRACT

*ADNP* encodes Activity-Dependent Neuroprotective Protein, whose de novo heterozygous mutations cause Helsmoortel-Van der Aa Syndrome (HVDAS), a rare developmental syndrome affecting brain formation and neuronal functions, involving autism spectrum disorder and intellectual disability. Although *ADNP* is one of the single-gene most frequently mutated in ASD, its precise role in the syndrome onset has yet to be clarified. ADNP is the DNA-binding component of the newly identified chromatin remodeler complex ChAHP in mESC. It recognizes euchromatin regions to establish less accessible local chromatin domains and has also been recently identified as a new player in the regulation of genomic topology, competing with CTCF in the organization of chromatin architecture. Our aim is to understand the genetic and epigenetic implications of ADNP underlying this neurodevelopmental condition; we harnessed cell reprogramming to establish a highly informative cohort of patient-specific iPSCs and use it as a platform to develop meaningful model for the pathology, thus enabling the assessment of the ADNP pivotal relevance in both pluripotent and neuronally-patterned stages. We discovered an altered gene expression program associated with cell fate decision and neuronal lineage commitment, highlighting a neurodevelopmental disruption elicited by *ADNP* mutations already at the pluripotent stage. Employing CRISPR/Cas9-engineering, we FLAG-tagged the endogenous *ADNP* to assess its genomic occupancy and revealed a genome-wide distribution of ADNP at gene-regulatory elements and a predominant presence at transposable elements, Alu sequences in particular. We decoupled ADNP and CTCF interplay in our human iPSCs model, and found a global redistribution of active enhancer histone marks signature, which sustain upregulation with the intervention of EZH2-mediated derepression. Finally, HVDAS cortical organoid models show morpho-functional impairment in the early stages of neuronal differentiation, with decreased size and lower mitotic activity, coupled with accelerated maturation phenotype assessed through single-cell transcriptomic analysis. Altogether, with these results we delineate how ADNP deficiency affects pluripotent regulatory landscape and disease-relevant mechanisms that ultimately impact neuronal development and functionality.

# 2. INTRODUCTION

## 2.1 Genomic architecture of neurodevelopmental disorders

Neurodevelopmental disorders (NDDs) constitute a broad spectrum of diseases originated during the development of the central nervous system (CNS). They are characterized by an early childhood onset leading to varying degrees of neuropsychiatric impairment, often in combination with a plethora of accompanying manifestations, whose specific configurations represent both a diagnostic and therapeutic challenge. Under the collective nomenclature of NDDs falls a variety of conditions often found in comorbidity, thus resulting in a syndromic clinical picture. NDDs include attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorder (ASD), schizophrenia, bipolar disorder, communication or language disorders, intellectual disability (ID), global developmental delay and motor disorders. The incidence of these disorders constitutes a serious health problem in modern days; previous studies in distinct populations indicated a median global estimate of 62/10,000 for autism and 10.37/1000 for intellectual disability, with a steadily increasing trend for ASD[1,2,3,4]. Diagnosis and treatment of these disorders can be difficult; treatment often involves a combination of professional therapy, pharmaceuticals, and home- and school-based programs, while diagnosis is historically based on the identification of few candidates' mutant loci. The rapid and prominent improvements of sequencing technologies during the past decades have enabled a more meticulous and extensive genetic testing, which in turn has raised the ability to efficiently identify the genetic lesion underlying NDDs[5]. Whole-exome sequencing (WES) has been the most instrumental in the identification of the major coding variants, leading to the discovery of new pathogenetic mutations that are progressively being annotated in online repositories such as DDD Study and SFARI databases[6,7,8,9]. The most common configuration of WES analysis is the collection of parental samples along with the proband's one (trio analysis) in order to maximise the precision of variant interpretation through the "correction" for the genetic background. Then, variants detected by WES are filtered and categorized as pathogenic, likely pathogenic, uncertain significance, likely benign, or benign in accordance with updated variant interpretation guidelines[10]. However, since WES is limited to the detection of coding variants, a broader whole-genome sequencing (WGS) inquiry was deemed necessary in order to detect non-coding variants, which are increasingly studied for their pathogenetic potential and heavy influence on the phenotypical outcome[11]. For example, *de novo* mutations in highly conserved regulatory elements active during foetal brain development, are significantly and

specifically enriched in neurodevelopmental disorders, and have been found in 1-3% of patients with no exonic diagnostic variants[11].

What has emerged from large-scale WES and WGS analyses is convincing evidences for the massive genetic heterogeneity not only within but also between and across different NDDs, and considerable overlap of genes involved in more than one NDD. Exome- and genome-wide sequencing studies have identified approximately 5,500 single-gene disorders and traits caused by mutations in over 3800 genes, 1100 of which have been causally linked to neurodevelopmental disorders[12,8]. SFARI database lists 1089 genes and 2291 CNVs associated to ASD within a range of scores reflecting the strength of the causative evidence. These numbers outline the complexity of the genetic architecture of ID and ASD, which is composed of inherited and *de novo* mutations (mostly germline, with mosaicism detected in a small percentage of individuals)[7,13]. Moreover, each mutation contributes to the pathogenic mechanism through its specific burden: while mutations of genes in the SFARI list are fully penetrant mainly due to haploinsufficiency, mutations of many other genes are considered risk factors, impacting on the polygenic forms of such disorders[14,15,16]. ASD for example, is highly heritable, with both common and rare variants contributing to its etiology[17]. Common variants have been estimated to account for a major part of ASD liability as has been observed for other common neuropsychiatric disorders. By contrast, *de novo* mutations (mostly CNVs and gene-disrupting point mutations) have larger individual effects, but collectively explain < 5% of the overall liability. The net result of distinct combinations of variants can result in distinct severities of the disease, with a spectrum ranging from single-gene high risk rare variants to low effect size common variants[18,19] (**Fig. 1**).

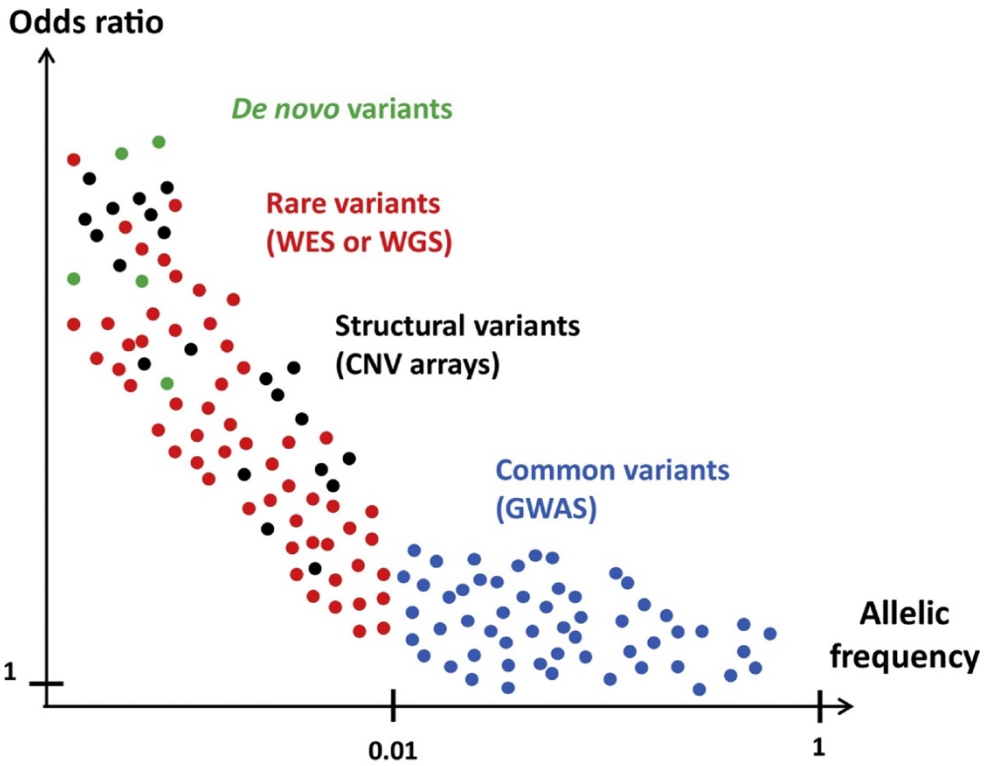



**Figure 1: ASD phenotype is the result of additive risk loci of both common and rare variants**. Adapted from Toma, Trends Genet (2020)

The majority of these CNV loci are also associated with developmental delay and other neuropsychiatric traits; for example, 16p11.2 duplications are associated with schizophrenia, while deletions of the same region are not. These observations are consistent with a model in which CNVs contribute risk to a number of neuropsychiatric disorders, however, the extent of this risk varies between phenotypes for each locus[20]. Concordantly, phenotype–genotype correlation studies highlighted both substantial variability in the clinical severity of patients with overlapping genetic etiology, and convergent phenotypes produced by distinct mutations (i.e., locus heterogeneity). This phenotypical variability is consistent with two main mechanisms of developmental dysfunction: 1) the pleiotropic effect of affected genes' product, whose mutations can influence brain- and non-brain-related activities, and 2) the functional redundancy of many of these genes due to the partial overlapping of pathways in which they are involved. In fact, genetic interaction (or epistasis) between such genes is a major determinant of genotype-phenotype correlations. Thanks to the study of familial NDDs, it has emerged that the phenotypical outcome essentially revolves around two main principles: gene vulnerability and mutational load[21]. As mentioned above, since most of the single genes accountable for NDDs are dosage-sensitive genes that produce haploinsufficient phenotypes, they fall within the category of highly vulnerable genes; their mutations (often *de novo*) are generally under negative selection and associated with

significant disease risk and high penetrance[22]. Disruption of one of these vulnerable genes has a high probability of inducing the onset of a disease phenotype also in the absence of other causative events, thus resulting in monogenic or oligogenic forms of NDDs. For example, copy number variations of the Williams-Beuren locus 7q11.23, show a combination of shared and symmetrically opposite phenotypes caused by the dosage imbalance of genes encompassed by a microdeletion or a microduplication (especially *GTF2I*), giving rise to Williams-Beuren syndrome or 7q-microduplication syndrome, respectively[23]. On the other hand, variants in genes that are not under negative selective pressure and are frequently transmitted in families for generations, are considered non-vulnerable genes (i.e., more tolerant to mutations)[24]; these mutations are not pathogenetic per se, thus being categorised as common variants with low disease risk and low penetrance, when considered alone. Nevertheless, studies have demonstrated that a significant portion of polygenic NDDs can be attributed to the individual burden of each common genetic variants, whose additive effect and reciprocal interactions can result in a diseased state or predisposing for comorbidity[16,25]. In line with the notion of mutational load, individuals with mutations in multiple genes are more likely to be affected than parental carriers, and the number of disrupting events positively correlates with the severity of the clinical signs observed[26]. To support this multifactorial nature of NDDs, the cumulative load of common genetic variants can also prime the genetic background to further genetic lesions, representing the first hit to subsequent possible pathological events. To complicate things further, the transcriptional characterisation of 200 mid-gestational human brains identified almost 8000 eQTLs, including several thousand prenatal regulatory regions that helped identifying novel candidate risk genes and account for significant genetic liability in neuropsychiatric diseases. Moreover, gene network analysis revealed that for both schizophrenia and ASD, common and rare genetic variation converges within disease-specific gene co-expression modules[27]. Overall, these genetic experimental evidences have led to the formulation of the "neurodevelopmental continuum model", in which NDDs are seen as representing the diverse range of outcomes that follow from disrupted or deviant brain development[28]. This model is also based on emerging evidence for shared genetic and environmental risk factors and predict a certain degree of overlapping pathogenetic mechanisms. In fact, despite the striking diversity of genetic causes identified to date and the extensive phenotypic variation, major molecular commonalities underlying NDDs are being increasingly found. Several shared, opposite or unique phenotypes, ranging from morphological features (e.g., craniofacial dysplasia or cardiovascular abnormalities) to more functional cognitive readouts (e.g., ID, speech and motor delay) have been associated with

specific subsets of  a genes, whose overlapping functional categories point of convergent mechanisms in neurodevelopmental disorders[25,20]

### 2.1.1 Chromatin basis of neurodevelopmental disorders

The overwhelming success of exome- and genome-wide association studies in discovering thousands of disease-associated genes is recently meeting the development of functional genomics approaches to elucidate the molecular mechanisms of these genes. Recently, mutations in genes encoding chromatin regulators and transcription factors have been recognized among the most common genetic causes of ID and ASD: 152 genes in the SFARI list fall into those definitions, of which 97 enrich the gene ontology category "positive regulation of gene expression" and 124 enrich "DNA-templated regulation of transcription"[29]. Mutations in genes that epigenetically regulate transcription either by modulating histone marks balance and inheritance at regulatory sequences or by reorganizing chromatin at multiple layers, are likely to have an immediate transcriptional impact, which can dramatically affect proper cell fate acquisition or cell physiology[14] (**Fig. 2**).

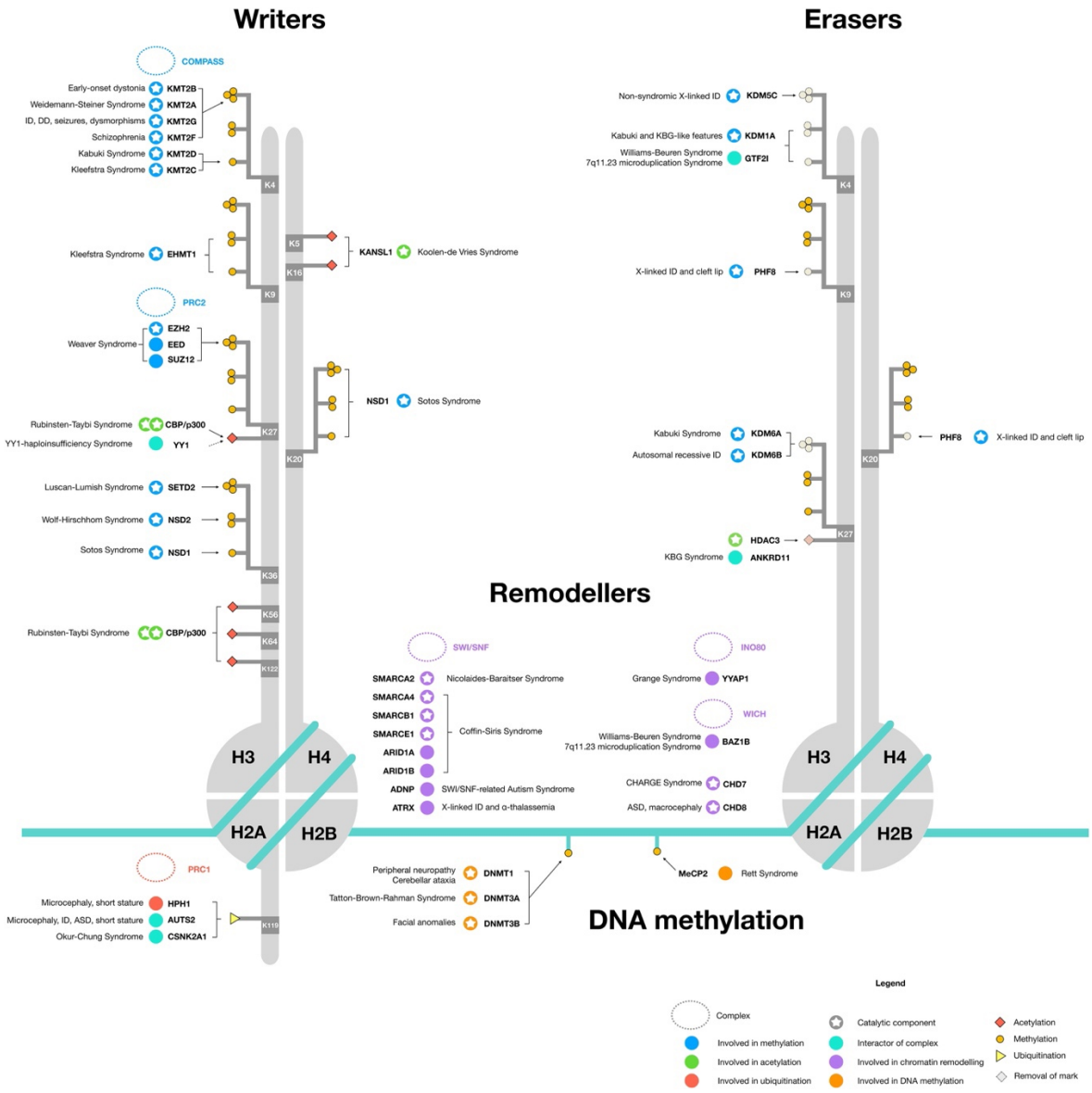



**Figure 2: Association of key epigenetic modulators to related neurodevelopmental disorders**. Adapted from Gabriele et al., Prog Neuropsychopharmacol Biol Psychiatry (2018)

Indeed, loss of function alleles of such genes found in ASD and ID have been mechanistically linked to *de novo* events such as CNV or SNV in heterozygosity[30]. Network analysis on exonic deletion events enriched in NDDs patients, highlighted recurrent and significantly enriched sets of genes that coalesce into protein-protein interaction network involving key chromatin regulators and synapse factors[31]. One paradigmatic example of how these two molecular functions intertwine and collectively contribute to ASD phenotype is *CHD8*, a chromatin remodeller whose targets are involved in axon and dendritic growth and cortical neurons migration[32,33]. Another study mapped ASD and ID risk genes onto co-expression networks depicting developmental trajectories and transcriptional profiles of foetal and adult cortical maturation, and showing that ASD genes are enriched in superficial cortical layers and glutamatergic projection neurons[34]. These results were further supported

by single-nucleus RNA sequencing of cortical tissue from patients with autism, which identified autism-associated transcriptomic changes in specific cell types. In particular, synaptic signalling of upper-layer cortico-cortical excitatory neurons and the molecular state of microglia are preferentially affected in ASD, with dysregulation of specific groups of genes correlating with clinical severity of autism[35].

The majority of epigenetic effectors causing NDDs are highly expressed during a window of vulnerability along the early phases of neuronal development. Their expression in the fully developed central nervous systems mostly localizes in the cortex, which contains neuronal circuits involved in higher cognitive functions such as language processing, social awareness and visuospatial construction[36]. The epigenetic dysregulation in human neural development can result both in NDDs caused by genetic lesions in chromatin regulators and in NDDs caused by environmentally-induced epigenetic dysfunction. The two possibilities are not mutually exclusive, as genetically encoded epigenetic vulnerabilities may enhance sensitivity to environmental disruptors, and viceversa[36]. Furthermore, the concomitant presence in NDDs of recurrent anomalies including peripheral nervous system, craniofacial, skeletal and cardiovascular defects, has been revealing an increasing spectrum of comorbidity between the cognitive/behavioural features that are hallmarks of NDDs and the involvement of neural crest-derived tissues[37]. Thus far, the most accountable pathological categories that explain a majority of the observed convergence are synapthopathies and enhanceropathies. While the former refers to NDDs sharing phenotypes that can be traced to synaptic dysfunction, the latter has gained increasing attention, encompassing disorders in which enhancers can be the directly mutated or, more usually, the genes involved in the control of enhancer elements are mutated[38]. The pathological relevance of mutations in the enhancers for developmental derangements is already well established, as reported for mutations at the sonic hedgehog enhancer locus, which are responsible for polydactyly, or in the long-range mutations that affect globin expression in hemoglobinopathies[39,40]. The pathogenic role of mutations in regulators of enhancer function is documented in the example of Gabriele-de Vries syndrome, which is caused by haploinsufficiency of the YY1 protein that mediates looping between enhancers and promoters[41,42]. Moreover, individual genetic background is strongly associated with enhancer and chromosomal organization, as recently demonstrated by a thorough characterization of genome-wide impact of human single nucleotide polymorphisms[43,44]. As a large fraction of causative genes in ASD are haploinsufficient transcriptional regulators, there is major interest in testing a selected subset of such genes to determine what insights into pathological mechanisms can be gathered by measuring the transcriptional consequences of their perturbation. One study harnessed

catalytically inactive Cas9-based transcriptional repression (dCas9-KRAB) to knockdown the expression of 13 ASD-related genes in a human cellular model of neuronal differentiation to capture the resulting transcriptional consequences using scRNA-seq. This effort identified individual transcriptional signatures after repression for each candidate gene, discovering shared transcriptional changes, and defining the impact of knockdown on the trajectory of neuronal differentiation[45]. Subsequent experiments confirmed a small hub of genes with the strongest statistical evidence of association with ASD that span historically relevant GO categories, including genes involved in expression regulation (*CHD8, ARID1B, CHD2, ADNP,* and *POGZ*), neuronal communication and synapse formation (*SCN2A, NRXN1, SYNGAP1,* and *ANK2*), and a kinase (*DYRK1A*), linking them with neurogenesis and neural progenitors maturation[46]. This is in line with other observations indicating proliferation and differentiation deficit as pivotal point of convergence for neurodevelopmental dysfunctions[47]. Moreover, the largest exome sequencing study of ASD to date (35,584 total samples) identified 102 high confidence ASD risk genes, of which 49 show higher frequencies of disruptive *de novo* variants in individuals ascertained to have severe neuro-developmental delay, whereas 53 show higher frequencies in individuals ascertained to have ASD and all of them are transcriptionally enriched in excitatory and inhibitory neuronal lineages[48]. Leveraging convergent evidences is critical for translating a list of genes into functional insights. The aim is to integrate many observations coming from "simple" analysis (e.g., risk-associated genetic variants) in order to generate hypotheses on "complex" functioning of the elements at stake. For instance, the pleiotropy of many NDDs can be minimized by focusing on points of strong convergence, which are expected to indicate the most relevant biological implications to the disorder in question[49]. To this end, syndromic forms of single-gene-accountable NDDs can be considered extremely informative as they provide an "all-in-one" picture of functional relevance that needs to be investigated and complemented to the other multi-factorial forms of NDDs.

## 2.2 Towards personalized medicine: molecular basis of pluripotency and stem cell experimental models

Human development is a complex multi-step process that from a zygote give rise to an embryo and eventually to the new-born organism. Understanding the main features of pluripotent stem cell is the biological prerequisite to comprehend the nature of the developing embryo and the foundational mechanisms that operate to drive cells and tissue differentiation. In the early phases of embryo formation, right after fertilization, different biological structures come in succession in a timely fashion during a short time window. The

zygote, the single diploid cell originated from the fusion of the spermatozoon and the ovum, undergoes the first cell divisions in a process called cleavage, which leads to the formation of a compact sixteen cell aggregate known as morula around 3-4 days after fertilization[50]. This is a mass of totipotent stem cells that further evolves into the blastocyst as mitotic events keep ongoing within the zona pellucida. At this stage, cells start organising and differentiating into distinct structures, resulting in a functional separation then enables the first lineage specification process in the mammalian development[51]. An outer layer called trophoblast provides nutrients to the embryo and will develop into extraembryonic structure such as the placenta, while the pluripotent stem cells that form the Inner Cell Mass (ICM) starts aggregating and polarising inside the blastocyst cavity[52,53]. Around 3 weeks after fertilisation, during gastrulation, they will arrange to form the primitive streak, which marks the beginning of the differentiation processes that will eventually generate the three main germ layers: endoderm, mesoderm and ectoderm. These represent the ancestor structures that will develop into the definitive tissues of the foetus and the adult organism[54]. Cells isolated from mammalian ICM prior to uterus implantation and grown in culture are known as Embryonic Stem Cell (ESC)[55,56]. Scientists have been classifying stem cells based on specific and fundamental features that they show *in vivo* and *in vitro*: reproduce themselves for long time maintaining an undifferentiated state, generate progeny able to differentiate into functional cell types, interact with the immediate environment to sustain self-renewal and differentiation[57,58].

In the past years, a specific set of criteria has been developed, especially in mouse, aimed at experimentally testing the requirements to be fulfilled in order for a cell to be considered a pluripotent stem cell (PSC)[59]. The principal and most stringent ones to assess functional pluripotency are: 1) Embryoid bodies formation and 2) teratoma formation, which assay the ability of PSC to generate cells representative of the three germ layers both *in vitro* and *in vivo*. 3) Generation of chimeric organism and 4) tetraploid complementation assay, which test the ability of PSC to generate *in vivo* all cell types needed to produce a fully functional and viable embryo, including germline-competent lineage[60,61,62,63,64]. Since the production of live organisms starting from human ESCs (either as chimeras or organisms entirely derived from these cells) is clearly not ethically feasible, researchers resorted to less stringent surrogate standards to assess pluripotency. Teratocarcinoma formation provided researchers with an *in vitro* model of development that resulted in the discovery of many of the factors recognized today as markers for pluripotent cells[59,65]. Starting with these experimental models, a valuable effort has been put in the identification of the molecular axes that sustain stem cell identity and modulate self-renewal or differentiation, both in mouse and human,

such as TGF-β, WNT/β-catenin and FGF pathways. Transforming growth factor beta (TGF-β) signalling has a prominent role in the earliest phase of cell fate decisions, including primitive streak formation, neural induction and mesoderm specification[66]. The TGF-β superfamily of ligands signals through two main branches: the SMAD1/5 branch, which transduces on behalf of bone morphogenetic protein (BMP) ligands and the TGFβ/Activin/Nodal branch, that involves the activation of SMAD2/3[67]. Upon activation by phosphorylation and association with a common SMAD4, the receptor-activated SMADs translocate to the nucleus and, in concert with other transcription factors, regulate gene expression[68,69]. In fact, hESC cultured in presence of recombinant Nodal exhibited prolonged expression of pluripotency markers[70]. There are also two inhibitory SMADs, SMAD6 and SMAD7, which inhibits both branches of TGFβ signalling, providing a repressive input on the pathway[68]. It has been shown that inhibition of SMAD2/3 activation results in significantly reduced expression of markers of pluripotency[71]. However, SMAD2/3 activation alone does not confer upon hESC the stem cell identity. Indeed, WNT pathway activation collaborates in the maintenance of pluripotency, directly or indirectly impinging on SMAD2/3 activation[67]. WNT pathway activation via pharmacological inhibitor of glycogen synthase kinase-3 (GSK-3), maintains the undifferentiated phenotype in hESC and sustains expression of the pluripotent state-specific transcription factors OCT-3/4, REX1 and NANOG[72,73]. Canonical WNT/β-catenin signalling role in pluripotency regulation has long been debated; according to the current most comprehensive interpretation, WNT cascade ultimately activate transcriptional programs with no intrinsic restriction in the type of biological event that may be controlled by these programs. Thus, WNT signals may promote cell proliferation at different developmental timepoints, and also control cell fate determination or terminal differentiation[74,75,76]. Pluripotency is critically dependent on physical interactions between pluripotency factors and transcriptional mediators of key morphogen signalling pathways. Thus, intermediaries of the WNT, BMP, and TGF-β pathways physically interact among each other and also co-occupy the genome with OCT4, SOX2, and NANOG to promote the pluripotent state and repress developmental stimuli[77,78,79]. FGFs represent another relevant protein family of potent regulators impacting on differentiation and proliferation in different types of stem cells including embryonic, hematopoietic, neural, spermatogonial, and bone marrow-derived mesenchymal stem cells[80,81]. Many members of the FGF family including FGF2, FGF4, FGF6, FGF7, FGF8, and FGF9 have been reported to control on the stemness of PSCs. Among them, it is clearly described that FGF2 and FGF4 are highly pertinent to maintain human stem cells in the undifferentiated state[82]. Among the downstream mediators of FGF2 signalling pathways, the MAPK pathway plays a pivotal role in maintaining the pluripotency of hPSCs. Expression

of FGF4 is restricted to undifferentiated embryonic stem cells and its transcription depends on a distally localized enhancer that contains consensus sites for SOX2/OCT4 complex, which controls embryonic stem cells maintenance[83]. Moreover, FGF4 expression is also induced by activation of TGFβ-related protein Nodal[84].

All these intertwined signalling cascades happen at a different scale in a diversified tissue-specific landscape defined as stem cell niche. From the embryo to the adult organs, the stem cell niche represents the microenvironment within the anatomic location where stem cells are found, and are committed to their fate[54]. In the niche, stem cells receive and integrate signals that maintain them in an undifferentiated state and induce their self-renewal as needed, controlling the proliferation and depletion balance. Therefore, the niche constitutes a basic unit of tissue physiology as critical as stem cell-autonomous functions in shaping our understanding of stem cell biology and its impact on health and disease[85,86]. To systematically investigate niches in a target tissue, stem cells first need to be identified by marking cells and following their lineages, then characterize cellular neighbours, expression patterns of signalling molecules and local environmental factors such as extracellular matrices. If commonalities are found, then the microenvironment can be perturbed to learn which aspects, if any, affect stem cell behaviour[87]. For example, in the intestinal crypt, fast-cycling stem cell and quiescent progenitor populations reside in distinct positions at the bottom of the crypt, whereas the transient amplified pool of precursors line the walls of the crypt, progressively differentiating as they reach the surface of the villi[88]. Many experimental approaches aimed at highlighting the structural properties of the niche and its influence in the adult tissue morphology employ *in vivo* genetic lineage tracing and live imaging lo localize and track the development of a stem cell pool during homeostatic conditions or processes, such as tissue regeneration, stem cell compartment growth, commitment or derangement[89,90].

### 2.2.1 Cell reprogramming and induced pluripotent stem cell (iPSC)

The prospect of personalized regenerative medicine promises to provide treatments for a wide range of degenerative diseases and medical conditions. An important first step in attaining this goal is the production of pluripotent stem cells directly from individual patients, thereby providing autologous material which can be further characterised or genetically manipulated[91]. Several methods were developed towards this achievement with the aim of generating actionable pluripotent stem cell *in vitro*. The first approach that was undertaken was the somatic cell nuclear transfer (SCNT), which in the first attempts ensured

the generation of viable tadpoles from terminally differentiated adult cells. This was the first demonstration that terminally differentiated adult nuclei maintain full developmental capacity, that cell differentiation can be reversed backwards to totipotency, and that it decreases with developmental age of the donor nucleus[92,93]. Obstacles associated with the requirement for a scarce cell type such as human oocytes as a recipient cytoplasm, were amplified by the inherent technical challenge of the method. Moreover, the experimental requirements for successful nuclear transfer in primates appear to be considerably different than other species. The observation that a pluripotent phenotype appeared to dominate following the fusion of murine somatic cells to ESC, seemed to promise that cell fusion might be an appealing alternative to the inefficient nuclear transfer. When this method was proven to successfully reprogram human cells as efficiently as mouse cells, this looked like a system that could be used for both the study of the mechanisms of nuclear reprogramming in human somatic cells, and eventually for the direct production of patient-specific pluripotent stem cells[94,95]. However, the method was limited by the need of overexpressing genes already known to be important for pluripotency, and by the tetraploid nature of the hybrid generated by the cell fusion, rising doubts about the physiological applicability of the process[96].

A new interpretation of cell reprogramming was provided when a first study in mouse and its subsequent application to human demonstrated that a straightforward, logical approach could lead to a direct reprogramming of fibroblasts into induced pluripotent-like stem cells (iPSC)[78,79]. Through a long and challenging process of gene selection, the authors found that iPSC could be generated from both embryonic and adult fibroblasts after four transcription factors – Oct4, Sox2, Klf4, and cMyc (OSKM) – were introduced using retrovirus, demonstrating the ability of a small set of defined transcription factors to directly push cells backwards along the epigenetic landscape[97]. Epigenetics is here operatively intended as the study of molecular factors and related mechanisms that can perpetuate alternative gene activity states in the context of the same DNA sequence. The four Yamanaka factors were since described in details and characterized for their ability to impact on potency and differentiation both *in vivo* and *in vitro*. Disruption of *Oct-3/4* results in the inappropriate differentiation of ICM and ESC to trophectoderm and extra-embryonic endoderm lineages in both mouse and human[98]. *Oct-3/4* null mouse embryos also die in utero during the peri-implantation stages of development and no ESC can be derived from them. The same holds true for *Sox2* null embryos, which die at the time of implantation due to a failure of primitive ectoderm development. *Sox2* is also associated with uncommitted dividing precursor cells of the developing central nervous system (CNS)[99,100]. *c-Myc* is one of the first proto-

oncogenes found in human cancers and its developmental role has been observed in mouse embryos homozygous for a *c-Myc* deletion that died between 9.5 and 10.5 days of gestation[101]. c-Myc acts by modifying the global chromatin structure surrounding the nearly 25,000 binding sites in the human genome target genes involved in development of heart, neural tube and erythropoiesis[102]. *Klf4* null embryos develop normally but new-born mice die within 15 hr and show an impaired differentiation in the skin and in the colon, thus indicating that it plays a crucial role as a switch from proliferation to differentiation[103].

First-generation iPSC were not equivalent to ESC in some aspects, leaving space to technical refinements that have progressively produced cells considerably more similar to physiological ESC in terms of transcriptional profiles and genome-wide epigenetic status[104,105]. Capturing naïve pluripotency is a major ambition in the field; human ESC derived from preimplantation embryos share features with primed mouse epiblast stem cells rather than naïve mESC[106,107]. A meta-analysis of the transcriptome of putative naïve hPSCs established with different protocols displayed a high degree of variation in gene expression programs when compared to human blastocysts[108]. Nonetheless, naïve hPSCs resembled human preimplantation embryos more than their primed counterparts or naïve mESC[109]. Although mouse pluripotent stem cells and mouse embryos have been widely used as references for human naïve conditions, it is unclear whether human *in vivo* naïve pluripotency is equivalent to that in the mouse. scRNAseq analysis of early and late human blastocysts discovered human-specific gene expression signatures of pluripotent cells in the ICM, expanding the coordinates of pluripotency spectrum and therefore the technical conditions required or culturing hESC[110,111].

Many studies had highlighted the importance of a core transcriptional regulatory circuitry sustaining pluripotency both in human ESC and then in iPSC. Transcription factors such as OCT4, NANOG and SOX2 are master regulators of early mammalian development modulating the expression of a common target group of developmentally important factors[112,113]. They collaborate to form regulatory hierarchies consisting not only of a highly integrated protein interaction network, but most importantly, they establish a solid autoregulatory and feedforward loops that enhance the maintenance of the pluripotent state[114,115]. OCT4, SOX2, and NANOG collectively target their own and each other promoters' regions to boost their expression, along with hundreds of other gene regulatory elements. Their function has been progressively characterized eventually describing a model in which they act co-ordinately occupying target regions in a sequential manner[116,110]. The master regulators occupy the promoters of active genes encoding transcription factors, signal transduction components, and chromatin-modifying enzymes that promote cell self-renewal.

However, these transcriptionally active genes account for only half of the targets of OCT4, SOX2, and NANOG in ESC. They also co-occupy the promoters of a large set of developmental transcription factors that are silent in ESC, but whose expression is associated with lineage commitment and cellular differentiation[113,117]. It is important to understand whether there are genuine differences in the global chromatin structure and the gene expression programs of human ESC and iPSC, given that such differences may impact the potential therapeutic use of iPSC[118]. A study mapping transcriptional and epigenetic profiles in 6 ESC versus 6 iPSC lines showed that their H3K4me3 and H3K27me3 occupancy enrichment as well as their protein-coding and noncoding transcriptome profiles were highly similar. As expected, nucleosomes marked with H3K4me3 occurred at the vast majority of actively-transcribed protein-coding genes in both ESC and iPSC while H3K27me3-modified nucleosomes occurred primarily in the promoters of repressed genes, many of which encode key regulators of development[118]. Another study showed proofs of faithful epigenetic reprogramming providing evidence of genome demethylation at key promoters of pluripotency genes and reactivation of somatically silenced X chromosome. iPSC also revealed other ES-like qualities, including growth factor responsiveness, the ability to act as reprogramming donors in cell fusion, as well as the ability to undergo ES-like differentiation both *in vitro* and *in vivo*, contributing to germline chimera[104].

Moreover, important functions have been attributed to the activity of epigenetic writers, readers and erasers in establishing ESC-specific chromatin patterns. Histone acetyltransferase complex P300/CBP promotes expression of pluripotency regulators critical for OSKM-mediated iPSC derivation as well for ESC[119]. On the other hand, H3K9me induced by the histone methyltransferase G9a, is associated with Oct4 inactivation, and its enzymatic inhibition can replace for Oct4 in transcription factor-induced pluripotency[120]. H3K9me and H4K20me are enriched in shortening telomere during cell aging, but their presence decrease at telomeres and pericentromeric repeats of iPSC, while telomere length increases to levels comparable to ESC, showing that histone methyltransferases inducing H3K9me3 and H4K20me3 might be key for telomere-associated pluripotency induction[121]. DNA methylation state is another important factor that plays a major role in the response of pluripotent cells to differentiation stimuli. By examining differentially methylated regions in ESC, iPSC and their parental fibroblasts, on a genome-wide scale it was observed that differentially hypomethylated cytosine-phospate-guanine (CpG) island were more enriched in human iPSC compared with their parental fibroblasts, which instead showed widespread CpG hypermthylation; those regions were also associated with bivalent chromatin marks, which identify developmental regulators. In addition, hypomethylated CpG island

overlapped with binding sites for OCT4, NANOG and SOX2[122]. Taken all together, these evidences show that iPSC and ESC share most of the fundamental epigenetic and transcriptional features, pointing to iPSC as a most valuable cell platform for developmentally relevant studies.

### 2.2.2 Application of pluripotent stem cell models in developmental biology and therapy

Generation of iPSC coming from different somatic tissues has represented a unique turning point for the *in vitro* disease modelling and pertaining therapeutic utility[123]. In particular, patient-specific iPSC allow researchers to investigate for the first time the onset and progression of degenerative and developmental diseases in cells that carry the entire set of an affected individual's genes, providing substrate for dissection of pathological mechanisms and clinically-relevant drug discovery (**Fig. 3**).

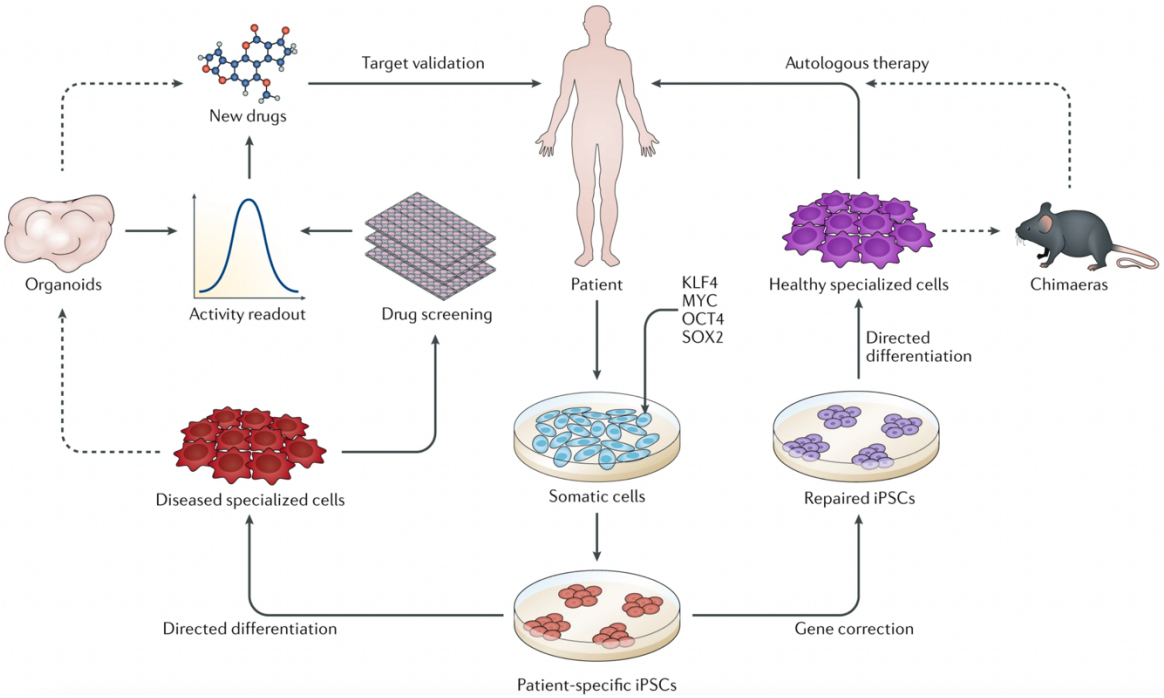



**Figure 3: Advances and complementarity of iPSCs applications.** Adapted from Rowe and Daley, Nat Rev Genet (2019)

Due to the limitations of integrating viral transfer system, scientists have been actively investigating other reprogramming methods such as the non-viral or non-integrative transfer systems, which are safer for therapeutic applications. Plasmid-based transfection of the Yamanaka factors demonstrated that even the transient overexpression of the pluripotency drivers was sufficient to achieve reprogramming[124]. Other approaches employ a Cre-loxP system, or piggyBac and Sleeping Beauty transposons to excise the integrated factors after

the ignition of pluripotency autonomous circuitry[125,126]. Introduction of synthetic mRNA encoding the reprogramming factors has also been proven to be an efficient method for creating integration-free pluripotent cells[127]. RNA-based iPSC have been generated using a single synthetic self-replicating Venezuelan equine encephalitis virus RNA replicon that expresses four reprogramming factors, *OCT4*, *KLF4*, *SOX2* with *c-MYC* or *GLIS1* at consistent high levels prior to regulated RNA degradation[128]. This is clinically significant when iPSCs are considered for transplant, as they represent a promising tool for regenerative medicine, in pathologies such as cardiomyopathies, stroke or spinal cord injuries. Technological advancements in iPSC engineering have also enabled a greater suitability for disease-modelling through a finer recapitulation of the features associated with their potency or differentiated state of interest. In addition, iPSCs are extremely more informative in the characterization of human-specific conditions compared to the already existing animal models, which are incapable of fully mimicking human complexity in terms of spatial and temporal regulation of gene expression during development. This in turn has facilitated the discovery of mechanisms behind those diseases for which we still do not have a molecular model, and therefore a treatment. In order to develop such models, a functional characterization is performed based on the knowledge about the physiology of cell types under study. Indeed, many disease-relevant cell type such as neurons are not easily accessible in patients, and the manipulation of iPSC has unlocked the way to the direct handling of challenging tissues and their associated disorders through specific differentiation protocols.

While the utility of iPSCs have been demonstrated mainly for Mendelian disorders, iPSCs have been also derived from patients affected by complex and multigenic diseases such as Down Syndrome. An extensive characterization of this disorders in patient-derived iPSCs highlighted a globally deregulated transcriptome and a neuronal signature of syndromic early brain development[129]. Schizophrenia is another complex disorder that has been successfully modelled using iPSC-derived neurons which display reduced neuronal connectivity, reduced neurite extensions, and reduced glutamate receptor expression, metrics consistent with phenotypes observed in post-mortem schizophrenia brains[130]. The properties of iPSC make them feasible for the identification of genes and pathways involved in differentiation and proliferation, and their perturbation via chemical compounds that can broadly or specifically activate or interrupt signal cascades during patterned development[131]. Autism Spectrum Disorder (ASD) is one more example of disease-oriented investigation where iPSC offers a powerful approach to identify novel regulators of development. The identification of pathological genetic background both in coding and non-coding genome, and epigenetic

players that operate in neuronal development can shed light on the onset of common ASD phenotypical traits such as developmental delay, intellectual disability or can help to clinically define the idiopathic ASD cases[132,23]. The development of new drugs is one of the biggest investments with steadily increasing new number of chemical compounds being experimented at multiple stages of development in multiple pathological conditions[23,133,134,]. An important challenge of iPSC technology is the individual variability between iPSC; the variable outcome of differentiation process is unpredictable and mostly caused by genetic background differences as well as the reprogramming methods used in a specific cell line. Thus, in efforts to model a disease, care must be taken in discriminating small phenotypic differences of technical origin from a disease-relevant phenotype when comparing patient and control iPSCs[135]. The generation of isogenic pairs of patient versus control iPSCs that differ exclusively at the disease-causing mutation has been used to control for the variation and have led to defining subtle relevant differences in monogenic diseases[136,137,138,139]. The problem is exacerbated when studying polygenic conditions like ASD, where low effect size loci contribute to the phenotype[25]. Since in this case the phenotypic differences can be elusive, the use of isogenic pairs of disease-derived and control cells is even more important. One of the most exciting approaches enabled by genome editing technologies such as CRISPR/Cas9, is to genetically and functionally test the empirical data generated by genome-wide association studies (GWAS), engineering variant alleles observed in these studies. This can be done by correcting the mutation in the patient-specific genetic background, or introducing a mutation that would phenocopy the disorder of interest in a wild-type context.

## 2.3 Genetic and epigenetic interplay in chromatin dynamics and expression control

Development in mammals requires that each cellular unit composing the embryo acts co-ordinately to ensure the proper orchestration of essential activities such as chromosome segregation, cell division, metabolism, and intercellular communication. Lineage specification is an ensemble of multi-step processes that eventually allows different cells or groups of cells to develop correctly by acquiring distinctive fates in a timely fashion. This is accomplished through nuclear reorganization and epigenetic remodelling that synergistically drive specific transcriptional programs so that each cell type will eventually manifest distinct phenotypic and functional characteristics[140]. The developmental potential entailed in these activities is encoded in each cell's genome and actualized through proteins that act together

to shape DNA into a specific chromatin structure, or recognize and bind specific sequences in the DNA to regulate gene expression. Superimposed epigenetic changes provide a flexible and dynamic way to modify gene expression patterns by enhancing or inhibiting gene activity. These modifications play an important role not only in differentiation processes but also in the maintenance of cell identity[141]. The extensive level of genome annotation reached to this day has charted a vast library of regulatory elements, which are being increasingly well defined in their developmental role[142,143,144]. Along with genetic landscape investigation, chromatin profiling also provides a systematic mean for detecting cis- and trans-regulatory elements, and identify the specific roles of chromatin remodellers, writers, readers and erasers in controlling DNA access and mediating context-dependent signals. Recurrent combinations of histone modifications and transcription factors binding sites can correspond with active, poised or repressed state of promoters, strong or weak enhancers, insulators, actively transcribed regions or inactive domains[145,146,147]. Moreover, local and distal regulatory elements also have implications in the interpretation of disease-associated variants that can be detected through GWAS. Pathogenetic mutations in expression quantitative trait loci (eQTL) have been identified, demonstrating that alterations in cis-regulatory sequences can impact on gene products or its dosage, in turn leading to developmental disorders[148,149,150].

In complex organisms such as mammals, the organization of DNA into structured nucleoprotein complexes not only represents extraordinary packaging system for the large size of the genome, but also provides mechanisms for selective contacts or access of transcription factors (TF) to binding sites in a highly cell-specific manner[151,152,153]. As cells replicate during differentiation, chromatin undergoes general changes in accessibility, with repercussion on transcriptional programs required for cell fate determination. In addition, investigating the open chromatin landscape of the whole human genome has revealed novel relationships among chromatin state, transcriptional activity, DNA methylation and regulatory factor occupancy patterns[154,155]; around 3 million chromatin accessible sites have been identified, encompassing all known experimentally validated cis-regulatory sequences and uncovering cell-selective control regions, some of which have also been associated with disease conditions[156,157]. Although DNA and chromatin transitions can occur at many levels of biological organization, the impact of epigenetic regulation can be summarised in four main areas (**Fig. 4**).

1) DNA methylation is the primary level of epigenetic modification that contribute in shaping cell development and cell identity.

2) A plethora of distinctive histone modifications are associated with multiple activity or altered states for promoters, enhancers, large and small genomic domains.

3) The highly dynamic structure of chromatin is functionally represented by both closed state, which inhibits the access to the underlying DNA sequence, and open state, which allows localized binding of TFs to cognate binding sites.

4) Recent methodological approaches aimed at the study of long-range interactions in genomic architecture are opening new perspectives on the investigation of nuclear structural properties, while posing questions on how spatial regulation is finely implemented into the chromatin context[151].

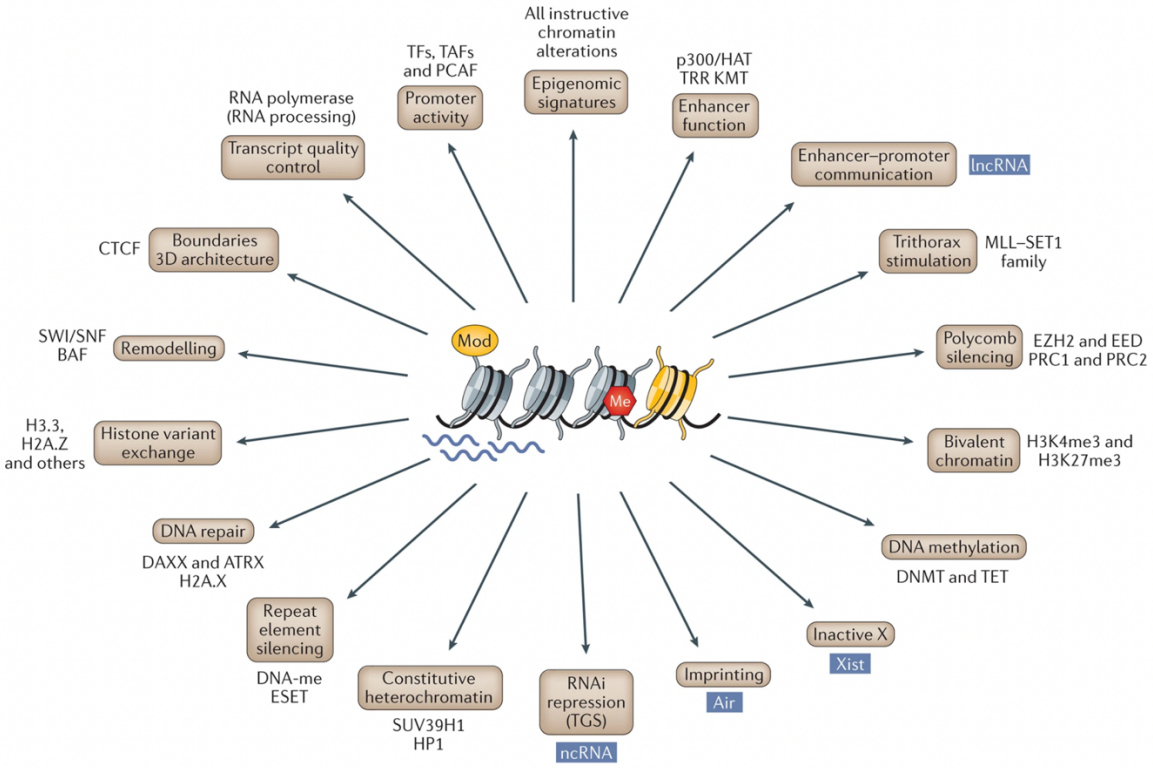



**Figure 4: Chromatin impact on epigenome functions**. Adapted from David Allis and Jenuwein, Nat Rev Genet (2016)

### 2.3.1 DNA methylation

Among the hallmarks of mammalian genome epigenetic modifications is DNA methylation. 75% of CpG dinucleotides in somatic cells carry a methyl-cytosine, while unmethylated DNA is typically found in embryonic stem cell genome or in neural and hematopoietic

precursors[158]. Methylation is deposited by different DNA methyltransferases enzymes (DNMT) according to the function or the timing of the modification. The two general classes of enzymatic activities are maintenance methylation and *de novo* methylation. Maintenance activity is necessary to preserve DNA methylation after every replication cycle. DNMT1 is the methyltransferase responsible for the renewal of parent DNA methylation patterns into the daughter cells during replication. This maintenance is executed through UHRF1, which binds hemi-methylated DNA and stimulates DNMT1 transferase activity. Therefore, a single complex contains both the writer and the reader of the methyl CpG mark, and both are essential for DNA methylation maintenance[159]. In mouse, disruption of Dnmt1 results in lethality before E 10.5[160]. DNMT3a and DNMT3b are the *de novo* methyltransferases that catalyse the establishment of DNA methylation patterns early during development. Deletion of Dnmt3a is compatible with embryonic development, but mice die within the first weeks after birth[161]. DNA methylation undergoes an extensive reconfiguration during development, that starts with the erasure of the paternal blueprint during embryo pre-implantation stages, and the set-up of a new methylation profile during post-implantation stages. From this developmental timepoint onwards, methylation patterns become stage- and tissue-specific, with changes that define each individual cell type lasting stably over a long period[162]. As demonstrated by the phenomenon of genomic imprinting (the parent-specific monoallelic silencing of selected genes) or X chromosome inactivation, DNA methylation is a highly dynamic and reversible process. However, the methylation profile of a differentiated cell represents a developmental milestone and an epigenetic barrier ensuring stable and heritable cell identity both in mitotic and post-mitotic cell types.

Of great interest for gene regulatory mechanisms is a specific category of CpG-rich sequences termed CpG islands (CGI), which are mostly devoid of methylation and enriched for permissive chromatin marks. CGI are few hundreds base pair long regions with a CG content higher than expected, usually associated with promoter sequences. Around 60-70% of human genes have a CpG island in their promoter regions, with biologically meaningful separation of genes based on their CGI content and properties[163]. Several experiments have outlined high plasticity and genome-wide divergent methylation patterns during lineage commitment when differentiating ESCs into distinct cell lineages *in vitro*. A comprehensive analysis of DNA methylation at single-base resolution in the mammalian frontal cortex has shown striking postnatal alterations in neuronal methylation profiles that occur as synapses develop from foetal to adult stage, suggesting that DNA methylation is important in the maturation of neurons in the developing brain[158]. Moreover, the majority of genes driving early lineage-specific expression in ESC-derived precursor cells are CpG-rich and are

located in large genomic domains devoid of DNA methylation, referred to as methylations valleys[164,165]. Silencing of CGI promoters is achieved through dense CpG methylation or Polycomb proteins recruitment. CGIs are therefore generically equipped to influence local chromatin structure and simplify regulation of gene activity. Many studies performing comparison of methylation distribution across tissues showed that CGIs are susceptible to be differentially methylated in a tissue-specific manner, correlating with changes in gene expression[166,167,168]. A global gain of methylation during cellular differentiation ensures the expression of the appropriate gene subset that define a certain lineage, thus mediating the silencing of pluripotency-associated genes and other lineage-specific markers.

### 2.3.2 Functional characteristics of histone post-translational modifications

Chemical modifications to DNA and histone proteins form a complex regulatory network that modulates chromatin structure and genome function. In the past decades, and more recently with the advent of new omic technologies, we have seen a remarkable progress in our ability to characterize histone modifications at a global scale, thus enlarging our knowledge of chromatin patterns and their functional activity. Operationally, histone modifications can either disrupt chromatin contacts or affect the recruitment of transcriptional factors to chromatin. Their presence on histones can dictate the higher-order chromatin structure and can orchestrate the recruitment of enzyme complexes to manipulate DNA[169]. The core histones that compose the nucleosome are H2A, H2B, H3, H4 subunits (and their variants), and are subject to more than 100 different post-translational modifications (PTM). International and collective efforts in the past decades such as ENCODE and Roadmap Epigenomics projects have contributed to profile all these histone marks, storing an enormous amount of data into free repositories of ChIP-seq experiments performed in hundreds of human cell types[170,171]. The most studied are the covalent modifications such as acetylation, methylation, phosphorylation, and ubiquitination, which occur primarily at specific positions within the amino-terminal histone tails. Whereas lysine acetylation almost always correlates with chromatin accessibility and transcriptional activity, lysine methylation can have different effects depending on which residue is modified[172]. Methylation of histone H3 lysine 4 (H3K4) and H3 lysine 36 (H3K36) is associated with transcribed chromatin. In contrast, methylation of H3 lysine 9 (H3K9), H3 lysine 27 (H3K27), and H4 lysine 20 (H4K20) generally correlate with repression. Lysine can be mono-, di-, or tri-methylated, providing further functional diversity to each site of methylation. For example, both mono- and tri-methylation on K4 of histone H3 (H3K4me1

and H3K4me3) are activation markers, but with different nuances: H3K4me1 typically marks transcriptional enhancers, while H3K4me3 marks gene promoters, while tri-methylation of K36 (H3K36me3) is an activation marker associated with gene bodies of transcribed regions. In contrast, tri-methylation on K9 and K27 of histone H3 (H3K9me3 and H3K27me3) are repressive signals with unique functions: H3K27me3 is a temporary signal at promoter and enhancer regions that controls development regulators in embryonic stem cells. H3K9me3 is instead a permanent signal for heterochromatin formation in gene-poor chromosomal regions enriched with transposable elements, satellite repeats, telomeres, and peri-centromeres, referred to as constitutive heterochromatin[157]. Part of the large methylated chromatin domains is the result of methyltransferase activity of Polycomb group (PcG) of proteins, in particular Polycomb repressive complex 2 (PRC2), which catalyse the successive trimethylation of H3K27 that ultimately yields to H3K27me3. PcG proteins are epigenetic modifiers organized within multiprotein complexes originally described for their role in Hox gene clusters regulation in *Drosophila melanogaster*[173]. They coordinate expression state of key developmental genes in multiple cell types and tissue contexts, including embryonic and adult stem cells, and are essential for cell fate transitions and proper establishment of cell identity in vertebrates. Indeed, PRC2 occupies a special set of developmental genes in ESC that must be repressed to maintain pluripotency and that are poised for activation upon differentiation stimuli[174,175]. What makes chromatin context relevant for gene regulation is often the combination of different histone modifications composing narrow or broad chromatin domains such as the so-called bivalent domains (although this model is being challenged[176]), and the additive occupancy of the specific epigenetic readers[177]. Bivalent promoters associated to Polycomb present large H3K27me3 domains deposited by PRC2 methyltransferase subunit EZH2, and smaller H3K4me3 deposited by Trithorax group (TrxG) of proteins. This chromatin signature typical of transcriptionally-poised genes is later resolved into H3K4me3-only or H3K27me3-only lineage-committed promoters upon differentiation. Deletion of EZH2 in ESCs results in loss of neurogenic capacity, pointing at a crucial role during neurodevelopment. Indeed, it is known that different sets of genes are differentially regulated by Polycomb during neurodevelopment, resulting in the fine-tuned control of the neural maturation[178]. Ezh2 knock-down model of chicken embryos show defects in the apico-basal polarity of neuroblasts leading to impaired neural tube organisation[179]. PRC2 also acts in concert with DNA methylation for regulating gene expression during cortical development. In particular H3K27me3 promoters are more prone to acquire DNA methylation during neuronal differentiation, defining restriction and potential of neural progenitor cells (NPCs)[180]. Furthermore, steady-state H3K27me3 levels along development reflect the equilibrium

between methylation, mediated by EZH enzymes, and demethylation, catalysed by KDM6A or KDM6B. Combinations of H3K4me1 with either H3K27me3 or H3K27ac at enhancers is another perfect example of the synergistic outcome that different modifications have at specific loci. Developmental enhancers in pluripotent stem cells exist in two distinct classes that share enrichment for H3K4me1, P300 binding, and nucleosomal depletion. However, they are also characterised by distinctive chromatin signatures based on their activation state. Enhancers with H3K27me3 mark are in a poised state that bookmarks them for a response to developmental cues, while enhancers enriched for H3K27ac are actively functioning in the transcription of lineage-specific genes[181,182]. The deposition of H3K4me1 at transcriptional enhancers is catalysed by the SET domain of KMT2D (also known as MLL4), a histone-lysine *N*-methyltransferase belonging to Trithorax group of proteins that functionally counterbalance Polycomb group of proteins for maintenance of gene activity. KMT2D is also required for the binding of histone H3K27 acetyltransferases CBP and P300 on enhancers, and cell type-specific gene expression during lineages differentiation[183,184]. It is particularly interesting that mutations in KMT2D are highly recurrent in tumorigenesis[185], and its haploinsufficiency has been described as the cause of a neurodevelopmental disorder called Kabuki syndrome, an autosomal dominant disease characterised by craniofacial abnormalities, intellectual disabilities and organs malformations[38,186,187].

### 2.3.3 Chromatin remodelling

Chromatin exists in a dynamic balance between genome packaging and genome access, and its higher-order structure can influence and be influenced by local gene activity and epigenetic modifications. Local organization of chromatin elements is a prerogative of many specialized protein complexes; among them, chromatin remodellers play an instructive role in the configuration of DNA domains and regulatory sequences, and in turn of gene expression. Nucleosomal rearrangement through specialized remodellers is necessary in order to slide or eject nucleosomes and expose regulatory regions required for transcription factors to activate gene expression[188]. On the contrary, certain remodellers promote dense nucleosome packaging directly upstream of transcription start sites. There are four different families of chromatin remodelling complexes, and all four utilize DNA-dependent ATPase domain which enables the break of histone-DNA contacts. The four families of ATP-dependent chromatin remodellers are the SWI/SNF (known as the BAF complex in human), the INO80, the ISWI and the CHD families. All of them also share high affinity for nucleosome and can recognize covalent histone modifications; moreover, they all have domains for interaction with other chromatin factors[189]. Together, these shared properties

allow nucleosome engagement, selection, and remodelling. However, all four are also specialized for particular purposes and biological contexts, imparted by unique structural properties like catalytic domains or subunits. Most remodellers are specialized to preferentially conduct one main function: nucleosome assembly and organization (including installing or removing histone variants) and chromatin accessibility[190]. Nucleosome assembly and organization. Following replication, histone chaperones deliver histone complexes (H3–H4 tetramers and H2A–H2B dimers) to nascent DNA where assembly remodellers such as the ISWI and CHD subfamily complex help in the maturation of the canonical octameric nucleosomes. Next, they form nucleosome arrays by spacing nucleosomes at relatively fixed distance. This assembly and spacing process also takes place during transcription at locations where nucleosomes have been dynamically ejected. Assembly remodellers such as Nucleosome Remodelling Deacetylase (NuRD) complex establishes densely packed chromatin profile through deacetylases activity of HDAC1 and HDAC2 and chromodomain helicase DNA-binding protein CHD4. Remodellers of the INO80 subfamily conduct the replication-independent removal of a particular histones within a nucleosome and its replacement with either a canonical or a variant histone. Common examples of editing include the replacement of canonical H2A or H3 histones with related variants, which is assisted by editing remodellers[190,191]. On the other hand, rendering the chromatin more accessible to proteins and RNA involves sliding nucleosomes along the DNA or ejecting full nucleosomes. SWI/SNF subfamily remodellers are largely involved in this function, favouring the exposure of binding sites for transcription activators and transcription repressors at gene promoters or enhancers[189].

During evolution, and in particular in the transition to vertebrates, there has been a large increase in the number of possible complexes as a result of gene expansion and divergence, and thus the ability to combinatorially assemble several subunits encoded by gene families. For example, different SWI/SNF complexes are found in different tissues according to their activity and unique developmental roles; mouse ESCs produce a peculiar complex called esBAF. This complex regulates the core pluripotency transcriptional network of mouse ESCs but is also crucial for the exit from the ESC state: a study employing RNAi-mediated depletion of components of esBAF prevents silencing of Nanog, and also hinders chromatin compaction and heterochromatin formation during differentiation[192,193]. As ESCs differentiate into neuronal progenitors, the esBAF complex undergoes several subunit exchanges which lead to the formation of a new cell type-specific complex called npBAF. BRG1 in this complex is required for the self-renewal of neuronal progenitors and for the normal differentiation of neurons[194]. As neuronal progenitors leave their stem-cell niche in

the subventricular zone of the brain and exit from mitosis, they switch again the composition of BAF complex, replacing npBAF with the post-mitotic neuron-specific nBAF[195]. Importantly, this switch in subunit expression is not simply correlated with differentiation status, but rather helps determine fate and proper synaptic plasticity[196]. Chromatin remodellers exert their role in gene regulation also through modulation of promoter activity; SWI/SNF complexes are indeed preferentially targeted to distal lineage specific enhancers and interact with p300 to modulate histone H327 acetylation[197]. Concordantly with these evidences, mutations in SWI/SNF components such as BRG1 impact on early development and cell cycle control, as homozygous mutations in mice are embryonic lethal or lead to increased cell proliferation, respectively[198]. Loss of Brg1 correlates with an increase in proliferating neural progenitors and an expansion of Sox2-positive cells in embryos at a later stage of neurogenesis[199]. Moreover, heterozygous mutations identified in several components of the BAF complex (SMARCA2, SMARCA4, SMARCB1, SMARCE1, ARID1A and ARID1B) lead to distinct but largely overlapping intellectual disability syndromes referred to as "SWI/SNF-related ID syndromes"[200].

### 2.3.4 3D genome: functions and dysfunctions of spatial organization

All the approximately 3.2 billion base pairs that make up the human diploid genome (equivalent to a length of circa 2 meters of DNA) have to be compacted in a nucleus whose diameter measures around 10um. Although three-dimensional (3D) architecture must be robust to ensure proper nuclear activity, it also needs to be flexible enough to allow changes to occur, such as those regularly happening during cell cycle. Recent results suggest that the global structural landscape is resistant to perturbations during development, but individual genes often switch between active and inactive chromosome compartments, and specific interactions frequently change both within and between chromatin domains[201,202]. A hierarchical folding of DNA has emerged from many studies employing different approaches to investigate the architecture of the genome, including sequencing-based methods, super-resolution microscopy and computational and synthetic modelling approaches[203]. Chromosomes contain hundreds of millions of base pairs that fold in nucleosomes, chromatin fibres, chromosome domains, compartments and finally in chromosome territories[204]. Therefore, chromatin is a multi-scale environment, and regulatory information resides at all levels (**Fig. 5**).

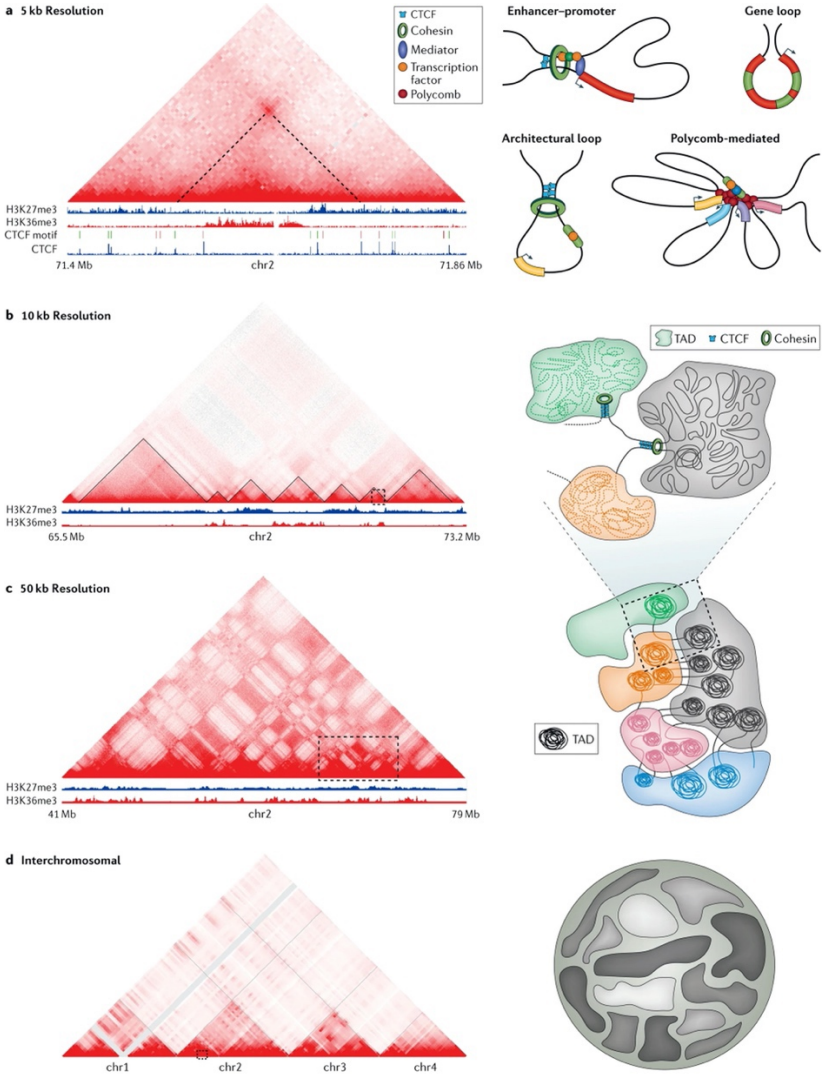



**Figure 5: Hierarchical organization of chromatin structure**. Adapted from Cavalli and Bonev, Nat Rev Genet (2016)

For example, large patterns spanning entire chromosomes reflect the spatial separation of euchromatin and heterochromatin and distal interaction between chromatin regions bearing shared chromatin modifications. The active and inactive regions are referred to as A and B compartments, respectively. The B compartments are predominantly localized in the nuclear periphery and surrounding the nucleoli, while A compartments are in the interior of the nucleus[205]. Given the close link between compartmentalization and the transcriptional state of certain chromatin regions, the causal relationships between genome organization and transcription seem to reside mostly in the epigenetic differences between compartments. This means that is not the act of transcription per se but the epigenetic state (active versus inactive) of chromatin regions that determines its affinity with other regions, indicating that epigenetic activity instructs the pattern of compartmentalization in the nucleus[206,207].

A key feature of vertebrate genomes is the relatively long distances linearly separating cis-regulatory elements from their target genes. Precise patterns of gene expression are controlled by mainly three classes of regulatory elements: promoters, enhancers and boundary elements. During differentiation and lineage commitment, these elements form specific interactions in dynamic higher-order chromatin structures[208]. Promoters are typically located near the transcription start sites (TSS) of genes while enhancers and boundary elements are commonly found within genes, and more frequently distributed at various distances (1–1,000 kb) flanking the genes within non-coding DNA. In order to elicit its effect, an enhancer is thought to be brought into close spatial proximity with its target promoter. These interactions have been proposed to form an active chromatin hub, in which high local concentrations of transcription factors and RNA polymerase II lead to transcription[39,209,210]. Many studies in the last decade have tried to address open questions about how the topological contacts between promoters and enhancers are defined in a spatio-temporal fashion. Techniques such as Hi-C and single-molecule microscopy have remarkably increased our power in answering those questions and extended our knowledge of the chromatin organization at high resolution, describing a spatial organization that is now believed to be pervasive throughout the genome, consistent across different cell types and highly conserved between mice and humans[211,212,213].

One of the most relevant breakthroughs is the finding that chromosomes are spatially segregated into sub-megabase scale domains, called topologically associating domains (TADs) which can substantially vary in size (range from tens to hundreds of kilobases), and even include different chromatin states within the same domain[214,215]. One of the defining attribute of such architectural feature is that genomic elements contained within the confines of a TAD interact with each other much more frequently than with regions located in adjacent TADs, and the different chromatin states often assemble into sub-TADs of the same epigenetic content[216,217]. The genome therefore appears as a succession of these structures demarcated by precise boundaries. TADs are made of heterogeneous interactions at different range, representing dynamic and preferential contacts that are summarised in a model called “loop extrusion” theory; this model, which is nowadays the most supported, poses that chromatin fibre is continuously and extensively looped and that nested looping processes give rise to the higher-order structures, as observed in the interphase nuclei by experimental approaches such as Hi-C[218]. In mammals, strong chromatin loops are observed at the borders of the majority of TADs, and they anchor to specific sites that establish the confines of the loops. Indeed, what really distinguishes background contacts, such as those among random points within a TAD, from regulatory or structural chromatin loops is the stability of the

loop itself, which is increased by the presence of specific boundary elements. CCCTC-binding factor (CTCF) oriented positioning together with cohesin juxtaposition are the two major players responsible for directional loop extrusion; they function as loop anchor and loop hinge, respectively[212,219]. Evidences supporting cohesin and CTCF prominent role in genome partitioning come from experiments using conditional KO of CTCF through auxin-mediated degradation or manipulation of its binding sites, which show radical perturbations in chromatin structure, and relatively little but relevant changes in gene expression[220,221,222]. Disruption of TADs can indeed rewire long-range regulatory elements and result in pathogenic phenotypes as exemplified by human limb malformations caused by deletions, inversions, or duplications altering the structure of the TAD spanning enhancers that drive limb formation[223]. The global misfolding effect is even more dramatic upon depletion of cohesin, as chromatin loops and TADs are drastically weakened. Furthermore, the amount and the residence time of cohesin directly determine the chromatin loop size: cells deficient for the cohesin release factor WAPL display chromatin loops with increased size, whereas inactivation of the cohesin loading complex NIPBL results in shorter chromatin loops[224,225,226]. Overall, these observations demonstrate the functional importance of TADs and chromatin loops in genome partitioning and in turn in gene expression, and indicate criteria for predicting the pathogenicity of human structural variants, particularly in non-coding regions of the human genome[223]. This last aspect presents unique challenges for deciphering disease etiology, particularly in attempting to distinguish causative mechanisms from secondary phenotypes when multiple cell types are involved. For example, several studies have characterized genome-wide chromatin wiring in fetal brain tissues and cultured neural cells. The identification of tissue-relevant chromatin contacts can be used for the annotation of novel enhancer-promoter interactions during human brain development or for the biological interpretation of risk variants in a complex neuropsychiatric disorders such as schizophrenia[227,228,229,230].

## 2.4 Development of the human brain

Over the past decades, considerable advances have been made in the understanding of the basic stages and mechanisms of mammalian brain development. Studies elucidating the biology of neural organization from the macro-anatomic to the cellular and molecular level have provided a picture of brain development as consequence of a complex series of adaptive processes operating within a highly constrained, genetically organized and constantly changing context[231]. The human brain expresses numerous genes; approximately 80%–95% of protein-coding genes are expressed in at least one brain region during at least one period

of development or adulthood. Co-expressed genes suggest anatomical structures, cell types, and molecular pathways that are potentially critical to brain development and function. Studies of model organisms have provided fundamental insights into human neurodevelopmental processes, however, despite commonalities shared with other mammals, there are striking interspecies differences that result in divergent features in cognition and behaviour[232,233]. Moreover, comparison of brain RNA-seq from different species reveals evolutionarily conserved and divergent gene expression patterns[232]. For example, the human brain as a whole, and most importantly the areas of the cerebral neocortex, develops more slowly than the brains of other primates, particularly, through a longer gestational time as well as longer childhood and adolescence period. Moreover, the developing human CNS possesses certain exclusive features, such as expanded proliferative zones and diverse subtypes of neural stem and progenitor cells with enhanced proliferative capacities that facilitate brain expansion, especially of the neocortex.

Human brain development begins in the early embryonic maturation at third post-conceptional week (PCW) with the differentiation of the neural progenitor cells and extends throughout foetal development, and post-natally until late adolescence and adulthood. After gastrulation, the embryo is constituted by cells organised in endoderm, mesoderm and ectoderm, which will give rise to all tissues of the foetus. During the embryonic period of brain development, the primitive nervous system starts emerging from neuroectoderm, a peculiar portion of the ectoderm that undergoes neurulation when neural progenitor cells (NPCs) begin to appear. Neurulation processes lead first to the formation of the neural plate which in turn develops into the neural tube. Here NPCs self-organise into a single layer of cell with apico-basal orientation, leaving the apical portion facing towards the cavity of the neural tube, which will give rise to the ventricular system[234,235]. At this stage of neurodevelopment, another cell population start emerging; neural crest stem cells (NCSC) are a vertebrate-specific transient multipotent cell population that originates at the neural tube border and continues until neural tube closure. After delamination from the roof plate, neural crest specifiers promote epithelial to mesenchymal transition and migration to different regions of the embryo depending on their position along the anterior-posterior axis. After migration to the target tissue, NCSCs can acquire various terminally differentiated cell identity such as those of peripheral neurons, chondrocyte, adipocytes, osteocytes and craniofacial structures[236]. Between PCW3 and PCW8, three primary vesicles along the antero-posterior axis appear: prosencephalon (embryonic precursor of forebrain), mesencephalon (precursor of midbrain) and rhombencephalon (precursor of hindbrain). The neural progenitor cells in the most rostral region of the neural tube will give rise to the brain

(prosencephalon) while the more caudally positioned cells will give rise to the hindbrain and spinal column (rhombencephalon). These primary vesicles will further subdivide into secondary structures constituting the primordium of the brain: the prosencephalon differentiates into the telencephalon and diencephalon, the rhombencephalon into the metencephalon and myelencephalon, while the mesencephalon does not subdivide[237]. The pool of neural progenitor cells that is formed at the end of gastrulation is too small to achieve a massive neuron production that will eventually generate billions of neurons that make up the mature brain. Thus, the first step of neuron production is aimed at increasing the size of the neural progenitor cell population. From the end of gastrulation through approximately PCW6 in humans, the population of neural progenitors divides by a symmetrical mode, producing two identical neural progenitor cells. Over multiple rounds of cell division between PCW3 and PCW6, symmetrical cell division amplify the size of the neural progenitor pool, which at this stage is referred to as neuroepithelium. Beginning around PCW6, the mode of cell division steadily shift from symmetrical to asymmetrical, thus giving birth to two different types of cells: one neural progenitor and one neuron[238]. The new progenitor cells remain in the proliferative zone and continue to divide, while the postmitotic neurons leaves the proliferative zone to take their place in the developing neocortex. The shift to asymmetrical cell division among the progenitor population is gradual; initially includes only a small proportion of progenitors, and finally those numbers increase dramatically by the end of cortical neurogenesis, which is completed around PCW15[239].

During the foetal period of brain development, which lasts from PCW9 to the end of gestation, the development of cortical and subcortical areas takes place. The gross morphology of the developing brain undergoes striking changes during this time. The brain gradually folds from a lissencephalic structure into a mature shape with characteristic sulci and gyri in a process called gyrification[240]. The changes that occur in the gross anatomy of the foetal brain reflect dramatic changes occurring at the cellular level. Most cortical neurons are generated during this time and completion of neurogenesis is largely achieved by mid-gestation[231]. New-born cortical neurons migrate in a sequential fashion to form the six layers of the neocortex and, once reached their final position, they differentiate by extending intra- and inter-layer dendritic and axonal processes[241]. By the end of the foetal period, most of the major pathways including patterns of sensorimotor contacts between neocortical areas are established[242].

### 2.4.1 Molecular principles of neurogenesis in the developing human cortex

The cerebral cortex is a bilateral structure located in the superficial portion of the cerebral hemispheres (telencephalon). It is the centre of higher cognitive functions, emotional processing and intellectual ability and consists of neurons, astrocytes, oligodendrocytes, blood vessels and ependyma[243]. The largest region of the cerebral cortex is the neocortex, whose neurons are organized in six layers (L1-6) aligned into a columnar (or radial) units that receive similar inputs and serve similar functions[244]. During evolution of the mammalian brain, the most salient morphological change has been the increased surface area of the neocortex, with a concomitant increase in its laminar complexity[245,246]. Observations of the developing neocortex in humans and non-human primates compared to the most studied model organisms such as mouse, rats and ferrets, have revealed how differences in neural progenitor cell populations can result in variable size and shape of the neocortex. Indeed, the human brain is roughly three times larger than the chimpanzee brain, and structures like the prefrontal cortex (the centre of cognition and decision-making) shows the most pronounced increase in size and complexity along phylogeny[243]. This shows that species-specific interplay between intrinsic genetic programmes and extrinsic morphogenetic patterning results in the expansion, layering, folding and functional organisation of cortical structures.

Seminal works have focused on identifying transcriptional signatures and molecular markers that specify regional identity of neural progenitors and thereby contribute to cortical formation[247,248,249,250]. Early forebrain patterning is regulated by interactions and gradients of molecules along the dorso-ventral and antero-posterior axes. Soon after neural induction is initiated, a region of the rostral neural plate known as the anterior neural ridge (ANR), the border between neural and non-neural ectoderm, begins to express Fgf proteins. Fgf family proteins are secreted from the rostral patterning centre and signal together with proteins secreted from two other regions, a dorsal centre which expresses bone morphogenic proteins (Bmp) and Wnt proteins, and a ventral centre which expresses sonic hedgehog (Shh), located where the ganglionic eminences are specified[251]. Fgf signalling centre confers positional identity along the rostral-caudal axis in a dose-dependent manner, with high doses defining the rostral-most cortical identity. The rostral patterning centre expresses at least five Fgf genes: Fgf3, Fgf8, Fgf15, Fgf17 and Fgf18, all of which are implicated in a region-specific control of proliferation in the cortical neuroepithelium. Fgf8 has the most pronounced dosage-dependent function in early forebrain patterning as this protein promotes progenitor proliferation, regulates neural specification, and inhibits cell death. Fgf8 hypomorph or conditional KO mice show severe patterning deficits and holoprosencephaly[252]. At the earliest stages of telencephalic morphogenesis, Fgf8 promotes the expression of Foxg1, which has a fundamental role in fostering the progenitor cell state and repressing

differentiation, by inhibiting Bmp signalling. The dorsal patterning centre, located in the cortical hem, has a central role in telencephalic development and secrets molecules of the Bmp and Wnt families that control the morphogenesis of structures belonging to the medial and dorsal pallium, the choroid plexus, hippocampus and neocortex[253]. Bmp signalling represses Fgf8 expression, and Fgf signalling in turn represses Bmp4 and Wnt8b expression, thus linking the functions of the dorsal and rostral patterning centres through reciprocal antagonistic control. The Bmp antagonists chordin and noggin are expressed in the ANR and notochord, and are required for Shh expression in the ventral centre plate, which is in turn required form maintenance of the normal levels of Fgf8 expression in the ANR. Ectopic expression and conditional inactivation of Bmp signalling alter development of the dorsal midline affecting the patterning of the dorsal pallium[252]. During early neurogenesis, Wnt proteins play an active role in symmetric divisions, whereas later, during neurogenesis, they promote neuronal differentiation through expression of N-myc, Ngn1 and NeuroD1. Overall, these opposing effect of oriented gradients impact on the expression of several master transcription factors, including COUP-TFI, Emx2, Pax6, and Sp8, whose controlled expression across the embryonic forebrain axes determines size and cellular position within cortical areas by specifying or repressing regional identities[242]. For example, the balance between dorsal excitatory and ventral inhibitory neurons is mainly regulated by a transcriptional axis involving the expression of Foxg1 and Nkx2.1, respectively, and by the cascade of TFs signalling deriving from their activity[254]. Within the excitatory glutamatergic neurons forming the six layers of the neocortex, laminar organisation is controlled by layer-specific transcription factors. Sox5 and Tbr1 are genes specifically expressed by early-born neurons that will occupy L6 and L5. On the other hand, late-born neurons differentiation is characterised by the expression of Satb2 or Pou3f2 and Pou3f3, highly enriched in neurons of L2-5[255]. Finally, the positioning of neurons and the establishment of specific connections largely depend on cytoskeleton regulation by modulators such as doublecortin (Dcx) and tubulin subunits (Tuba1a, Tubb2b and Tubb3), among others[256].

During patterning of the forebrain, the primordial structure is subdivided into distinct dorso-ventral and antero-posterior domains. The cerebral cortex arises from dorsal telencephalic proliferative zones called ventricular and sub-ventricular zone (VZ and SVZ), while the ventral telencephalon contains two proliferating cell masses, the lateral and medial ganglionic eminences (LGE and MGE)[257,258,259,260]. The neuroepithelium organises in a stratified layer forming the VZ, where they amplify through proliferative divisions prior to the onset of neurogenesis. Neuroepithelial cells then differentiate into apical radial glia (aRG), which can be distinguished by their parental cells for a different morphology and the

expression of astroglial markers (including glial fibrillary acidic protein, GFAP)[261]. aRG cells organise radially and coordinate neurogenesis and neuronal migration through the intermediate zone (IZ) towards the basal surface, thus forming the nascent cortical plate (CP), which subsequently develops into L2-6 of the postnatal neocortex[255]. Incoming CP neurons split into the superficial marginal zone (MZ), which develops into L1 of the mature cortex, and the deeper subplate (SP), which is situated below L6. The neurons of the MZ and the SP, the first to achieve morphological maturity and form synapses, are thought to play crucial roles in the migration and synaptogenesis of CP neurons, as well as in the formation of proper cortical efferent and afferent projections. The formation of the neocortex proceeds with an inside-out pattern, as earlier born neurons form the deep layers and later born neurons form the upper layers (**Fig. 6**).

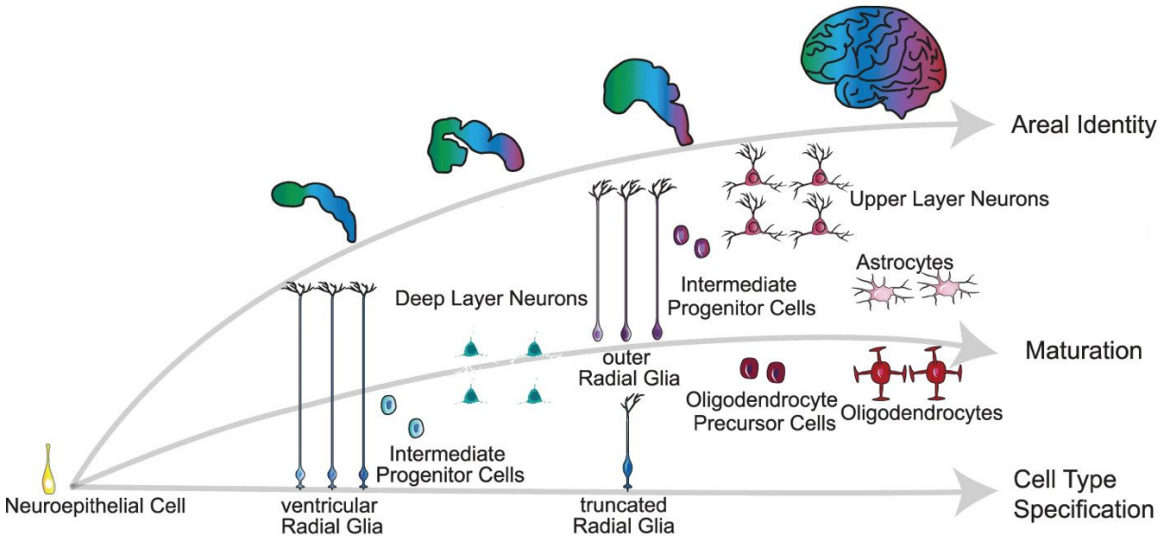



**Figure 6: Schematic representation of mammalian cortical development**. Adapted from Badhuri et al., Nature (2020)

Each layer has distinct connections within the cortex and with subcortical structures, but they are all mainly constituted by two major cell types. Roughly 80% are excitatory projection neurons that use glutamate as their neurotransmitter, and whose axons extends long distances to their synaptic targets. The remainder are interneurons that tangentially migrate into the cortical layers from the LGE and MGE, use GABA as their inhibitory neurotransmitter, and have axons that typically synapse on nearby neurons.

Another neurogenic proliferative compartment, the subventricular zone, appears above the VZ and enlarges dramatically over the course of early and mid-fetal development[262]. The VZ and the SVZ will give rise to all the excitatory glutamatergic projection neurons (also

known as pyramidal neurons) within the telencephalon, and subsequently glial cells[235]. The "radial unit hypothesis" of neocortical development proposes that functional radial units are established by the proliferation of precursor cells in the embryonic ventricular zone, giving rise to related neurons that migrate into the cortex along shared radial glial fibre guides[245]. Time-lapse imaging of GFP-labelled clones demonstrated that, *in vivo*, aRG cells generate neurons by multiple rounds of self-renewing asymmetric division, and that new-born neurons migrate along clonally related radial glial cell's fibre to the cortical plate[263,238]. aRG cells indeed have cell bodies situated in the ventricular zone, and a long fibre that extend to the pial surface that serve as scaffold for radial neuron migration[264]. Further studies have showed that the asymmetrical pattern of aRG cells division give rise to a daughter aRG cell, and either an intermediate progenitor cell (IP cells, also known as a transit amplifying cell) or nascent neuron[265,232]. IP cells migrate superficially from the VZ, where their parent cells remain, to the SVZ. Imaging studies have demonstrated that while aRG divisions are asymmetric and associated with self-renewal, IP cells of the SVZ (or basal radial glia, bRG) usually undergo one terminal symmetric division that produces two neurons thus depleting the progenitor cell pool[266]. Therefore, neurons can be generated both directly from aRG, or indirectly through the generation of IP, but only indirect neurogenesis leads to the amplification of the final number of neurons of the cerebral neocortex in mammals[266]. In fact, analysis of aRG versus IP cell divisions showed that, although aRG cells outnumber IP cells at early stages, IP cells account for the majority of neuron-producing cell divisions during all stages, implying that a major role of aRG cells in neurogenesis is to make neuronally committed IP cells[267]. bRG show virtually identical molecular expression profiles and neurogenic properties as aRG, except lacking apical processes, while transcription factor expression was found to be a criterium for differential classification of progenitor cells, with aRG expressing high levels of PAX6 and IPs expressing TBR2 (or EOMES)[268]. A human genetic study showed that individuals with homozygous silencing of TBR2 exhibited microcephaly and polymicrogyria, further supporting the importance of SVZ and IP cells in the control of neocortical size[269]. Primate corticogenesis is distinguished by the appearance of a large SVZ that subdivide into two subregions, an inner (iSVZ) and outer subventricular zone (oSVZ), the latter of which is not shared with lissencephalic organisms. Thymidine-labelling studies in primates indicate that proliferation of cells within the oSVZ coincides with the major wave of cortical neurogenesis, suggesting that the oSVZ contributes to neuron production[270]. In addition, cells in the oSVZ express aRG markers such as nestin (NES), vimentin (VIM), PAX6, and GFAP, as well as the IP cell marker TBR2 (EOMES)[271,272]. Cellular heterogeneity of the oSVZ, which includes both aRG and IP cell types, have highlighted the importance of this area for neuron production during human fetal

development. In particular, ~40% of oSVZ progenitor cells express nuclear and cytoplasmic markers typical of RG and also possess radial fibers. However, unlike aRG, which are bipolar, the oSVZ radial glia-like cells (oRG cells) are unipolar, with a basal fiber that ascends toward the pia but without an apical fiber that descends toward the ventricle[273]. Dynamic imaging of cultured slices of fetal neocortex showed that oRG cells division follows a distinctive behaviour whereby the cell soma rapidly jumps along the radial fiber prior to cell division, a process termed mitotic somal translocation[267]. oRG cells, upon completing mitotic somal translocation, typically divide with a horizontal cleavage plane, with the more basal daughter cell always inheriting the basal fiber. Cell fate analysis of oRG cell clones showed that they undergo multiple rounds of asymmetric division, functioning as founder cells for an extended population of IP cells that lose SOX2 and gain TBR2 expression, and another oRG, thus expanding the boundary of the oSVZ and enlarging the overall neuronal pool[274]. Neurogenesis in the oSVZ begins around PCW11, circa one month after the VZ-directed neurogenesis, and continues up until mid-gestation, when the oSVZ expands dramatically to become the predominant germinal region in the neocortex. Importantly, oSVZ expansion does not occur at the expense of progenitor cells in the VZ, implying that the oSVZ is generated by proliferation of the residing cells rather than by delamination and migration of VZ progenitor cells. The coexistence of oRG cells with ventricular RG cells demonstrates that human RG consist of two major subclasses, each functioning as neural stem cells in their respective locations[233]. Single-cell characterisation of oRG cells highlighted preferentially expressed genes sets related to extracellular matrix formation, migration, and stemness (including TNC, PTPRZ1, FAM107A, HOPX, and LIFR) and that these molecular features underlay distinctive behaviours for oRG cell cycle progression, and extensive proliferative potential[275]. Subsequently, oRG cells were also found in the developing mouse cortex, but at lower abundance than in gyrified brain such as the ones of ferrets or humans[276]. The appearance of a distinguished oSVZ during mid-gestation of primate neurodevelopment and the fact that it coincides with a major wave of neurogenesis has suggested that oRG cells might be the key cell type linked to the expansion and differentiation of the cortical sub-regions in primates, and that is primarily involved in initiation and development of cortical convolutions in diverse mammalian species[267,270,274,277] (**Fig. 7**). Although most of our current knowledge about developmental features of the mammalian neurogenesis comes from traditional animal models, they cannot fully recapitulate all the properties of the developing human brain. This raises the issue of generating an appropriate human experimental model, which could also represent a platform for disease modelling and translational applications.

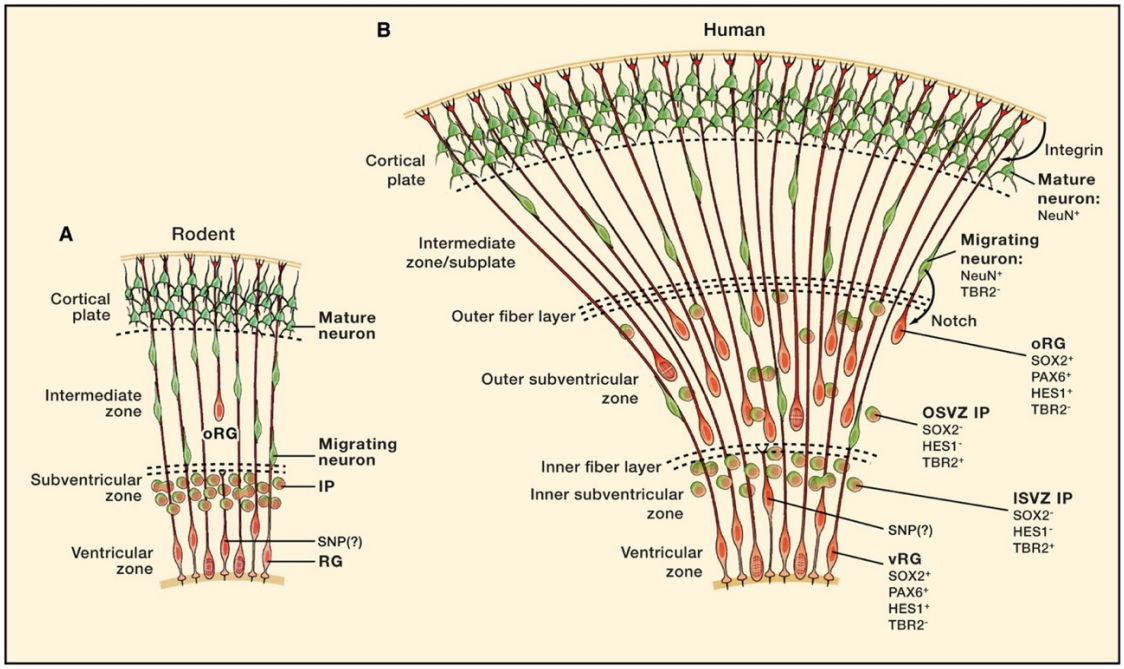



**Figure 7: Outer subventricular zone. (A)** Lissencephalic corticogenesis. **(B)** Human corticogenesis. Adapted from Lui et al., Cell (2011).

### 2.4.2 Experimental models to investigate human brain development

For decades, the study of cellular and molecular architectures of the developing human CNS has relied on the availability of foetal primary tissue, with several limitations hampering the progress of the field: 1) limited access to this kind of specimen, particularly at early developmental stages, 2) difficulty in *in vitro* maintenance, with complete maturation impossible to achieve or sustain for long periods, 3) ethical constrains. A relatively small number of research groups has reported the use of primary human foetal neuron cultures either as 2D layers or neurospheres. Initial attempts to culture primary dissociated foetal neurons showed that they contained abundant morphologically well differentiated neurons with complex dendritic arbours as well as numerous cell-cell contacts and synapses[278]. However, such degree of *in vitro* complexity has not been reliably reproduced, and classification of neuronal and gial cell types has relied mostly on a molecular description rather than morphological and functional profiling. The advent of single-cell RNA-seq (scRNA-seq) along with other omics technologies has enabled a much more fine-grained picture of brain cell populations and their molecular signature. One study applying scRNA-seq on surgically resected neurons from adult individuals cultured in 2D for almost 90 days, revealed that neuronal-subtype-specific genes identified from animal models are often lowly expressed or even absent in adult human neurons, while more complex signatures including non-coding RNAs are more informative of subtype specification[279]. The use of primary foetal tissue in combination with genome-wide technologies has also allowed in-depth

characterizations of the developing epigenome. Analysis of distribution of H3K27ac, H3K4me1, H3K4me3, and H3K27me3 in different human foetal samples has revealed a significant incidence of brain-specific features such as increased presence of super enhancer sites, with an average distance from target genes higher than in mice, as well as increased frequency of bivalent domains[280]. Moreover, high resolution mapping of the H3K4me3 across primates, including human children and adults, identified 471 sequences with human-specific enrichment or depletion, among which there are neuronal selectively methylated loci, including the TSS of several neuropsychiatric susceptibility genes[281]. Interestingly, genes with these human-specific enrichments showed only minimal overlap with the Simons Foundation Autism Research Initiative database (SFARI), nor they conformed to a single gene ontology (GO) category, therefore DNA sequences subject to differential histone methylation in human or chimpanzee prefrontal cortex are not part of a specific cellular signalling pathway or function, but rather influence widespread functional domains[282].

Primary and post-mortem tissues have allowed a detailed overview of the epigenetic landscape of human brain development, however, both approaches entail considerable limitations such as the lack of resolution given by the intrinsic cellular heterogeneity of the brain or glia-to-neuron ratios that could lead to considerable fluctuations across normal development or in certain pathological states. Furthermore, most studies rely on the analysis of bulk homogenates resulting in measurements of the average effects rather than cell type specific information[283]. To amend this drawbacks, one alternative has been the characterisation of dissected sub-areas or the use of fluorescence activated cell sorting (FACS) for the isolation of more homogeneous cell populations. For instance, analysis of germinal layers dissected from human foetal tissue has been instrumental in identifying the key TFs and pathways that characterize their constituent cells, as well as a key role of extra cellular matrix in their development[284]. Nevertheless, while a significant portion of the epigenomic organization in the neuronal nucleus is preserved, these approaches cannot avoid the loss of some types of histone modification, chromosomal loopings and other higher order chromatin structures[285]. In order to study functional properties in a 3D microenvironment, one of the most recent approaches has been the grafting of primary human tissue into mouse CNS through reprogramming technologies. Human iPSC-derived NPCs, due to their preserved plastic properties, are implanted in rodent brains resuming differentiation programs *in situ* once embedded and resulting for example in “humanized” rodent cortex, or alternatively, the implantation of *in vitro* differentiated neurons, which have been shown to integrate into mouse cortical circuits[286,287].

A significant breakthrough came with the development of methods that allowed the isolation and propagation of progenitors from the CNS in defined culturing conditions. This was initially achieved from dissection of the lateral wall of the striatum to obtain cells of the SVZ and then the expansion of the proliferating population which aggregates in suspension into what is known as neurospheres[288]. Neural stem cells can assemble and clonally grow into neurospheres when cultured in ultra-low attachment conditions with a medium containing necessary growth factors such as FGF and EGF. With the increasing experimentation on pluripotent stem cells, neurospheres originally derived from primary samples have been later substituted by iPSC-based production of neurospheres starting from 2D rosettes or NPCs in a neuronally committed environment and suspension cultures[288]. These iPSC-derived neurospheres are routinely used as a mean to maintain and study somatic stem cells and under the right exposure to precise morphogen cocktails, constitute and intermediate stage in multiple neuronal differentiation protocols[289]. While being a useful platform to investigate general properties of neural stem cells such as proliferation, self-renewal and multipotency, or a convenient substrate for drug testing, neurospheres are derived from an already advanced progenitor state potentially primed to specific lineages, thus failing to reproduce the sequential appearance of intermediate progenitor populations typical of brain development. Moreover, due to the frequent requirement for re-plating in 2D to achieve high maturation levels, they are devoid of intrinsic 3D organization found in the cortex[290].

iPSC-based neuronal differentiation has been a major goal for years, leading to a variety of approaches that would rapidly and reproducibly yield a homogeneous population of CNS-specific cells. First attempts discovered the presence of NPCs arising from embryoid bodies by modulating their culturing conditions, in particular, exposing them to retinoic acid[291]. Embryoid bodies can be subsequently dissociated for the isolation of NPCs that are cultivated in defined media complemented with neuronal patterning factors. However, as this model could not provide enough homogenous neural cell population, a new culturing conditions using stronger patterning factors was required to favour neuronal differentiation. This strategy allowed the derivation of neural precursors in the shape of rosettes: radial arrangements of columnar cells resembling a cross section of the developing neural tube, that express many of the characteristic markers of neuroepithelial cells and retain the ability to differentiate into the main classes of progeny of neuroepithelial cells *in vivo* such as neurons, oligodendrocytes, and astrocytes[292,293,290]. In order to optimize this approach, in the last two decades protocols started to focus on interfering with the SMAD pathway, achieving a drastic improvement in ectodermal specification and neuronal differentiation, with full neuroectoderm conversion achieved by simultaneous inhibition of the SMAD pathway

(TGF-β and BMP type I receptors), strategy known as dual SMAD inhibition. This approach has represented the foundation of most of the current methods for neural induction as well as a platform for further differentiation into region-specific neuronal type such as dopaminergic or spinal motoneurons[294,295]. Despite these remarkable achievements, differentiating cultures take usually months to reach neuronal maturity, limiting large-scale studies. Therefore, parallel efforts have been made for developing faster alternatives relying on the over-expression of master regulators of neurogenesis in combination with small molecules to obtain synaptically active neurons in less than one month[296,297]. Studies analysing the single cell content of neuronal population obtained with NGN2 over-expression show substantial molecular heterogeneity in the neuron types generated across multiple iPSC clones or individuals, with populations that express genes associated with neurons of the peripheral nervous system. This highlights the notion that neuron fate acquisition is sensitive to dosage and duration of master regulators expression and that heterogeneity can confound results that are sensitive to neuron type, if analysed as a whole[298].

#### *2.4.2.1 A focus on 3D models: brain organoids*

The advent of brain organoids constitutes the most recent and promising alternative to interrogate human brain development. *In vitro* organogenesis takes advantage of the properties of pluripotent cells that undergo a differentiation process through the manipulation of culture conditions, triggering intrinsic developmental programs that spontaneously recapitulate global aspects of specific tissue development. *In vitro* organoidogenesis can achieve the highest degree of tissue homology in terms of cell composition, organization and gene expression thanks to the increasing knowledge of the molecular and physical processes orchestrating tissue differentiation. Direct control of pluripotent stem cells growth can result in spatially organised functional units similar to the ones of the physiological reference organ, and most importantly, with the same developmental and temporal patterns found *in vivo*. These experimental systems offer an unprecedented level of resolution in the study of human organ development and have shown particular relevance for the brain, for both physiological and pathological research[299,300] (**Fig. 8**).

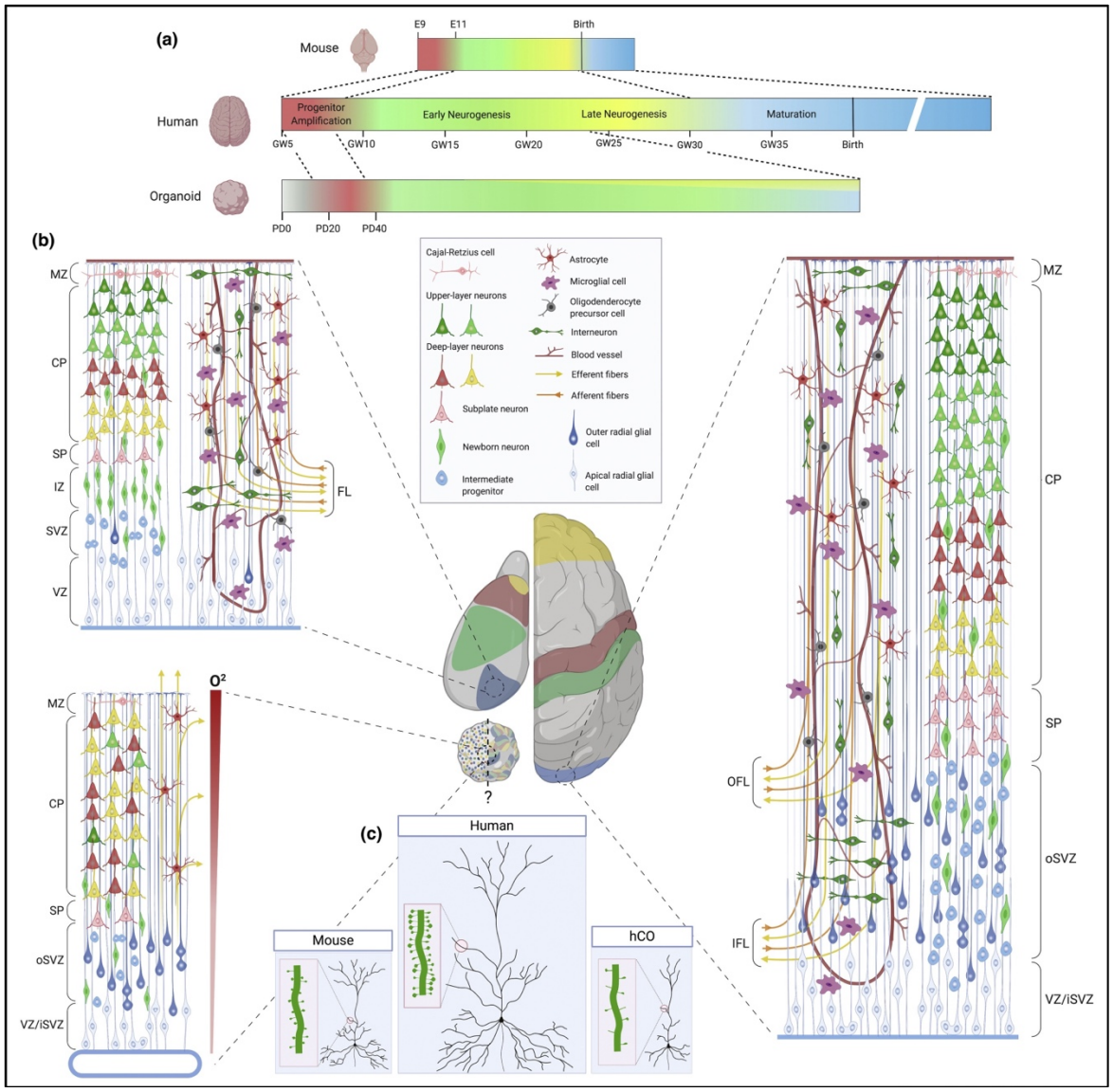



**Figure 8: Cortical development in mouse and human brain and human organoid models**. **(A)** Timeline of mouse, human and organoid for cortical development. **(B)** Comparison of cortical area spatial distribution in mouse and human cortex and in current organoid models. **(C)** Schematic representation of layer 2/3 pyramidal neurons and dendritic spine morphology in mouse, human cortex and human organoids. Adapted from Le Bail et al., Curr Opin Neurobiol (2021).

Given the complexity of an organ like the brain, the resemblance is only limited; nevertheless, many studies performed in different laboratories have generated brain organoids through protocols selectively adapted to address specific questions, with a certain extent of reproducibility among them[301]. Indeed, several key features of *in vivo* brain organogenesis are recapitulated by *in vitro* brain organoids, making them attractive models for studies of certain aspects of brain development. In particular, common aspects to all brain organoid protocols include: 1) formation of self-organizing structures, in particular rosettes which display features of the embryonic neural tube, including a neuroepithelium with apico-basal polarity and subdivided proliferative and differentiating zones that appear in a timely

manner, 2) generation of different subpopulations of progenitors usually absent in their 2D counterparts (e.g. apical and basal radial glial cells, deep and superficial layer neurons in an ordered temporal fashion), and 3) a considerable degree of compartmentalization at the extracellular level that includes the production of their own extracellular matrix (ECM)[302]. However, the absence of a proper ECM scaffold, vascularization, as well basement and apical membranes, has led to some adjustments in protocol parameters in order to compensate such as embedding in synthetic matrices, shaking, endothelial-like support, different combinations of incubation gases and adaptation to air-liquid interface culturing[303,304]. Depending on the brain area of interest, different region-specific organoids can be generated such as forebrain, midbrain, cortex, and hippocampus, through the use of patterning factors and morphogens. The first main published protocols could achieve the differentiation of pluripotent stem cells into virtually all the principal cerebral structures (whole-brain organoid) with no use of specific patterning molecules, but rather relying on basic neuronal induction medium, thus allowing cells to unleash their cell-autonomous differentiation potential[305,306]. Subsequently, other studies optimized protocols for the generation of patterned organoids such as the cortical organoids. The peculiarity of this method is the absence of extracellular scaffolding and the use of heavy patterning towards dorsal telencephalon through sequential exposure to dual SMAD inhibitors for neuroepithelial specification, EGF-FGF2 for neural progenitor expansion and the neurotrophins BDNF and NT3 for cortical glutamatergic neurons differentiation. Such mature cortical brain organoids display the main cellular components of the developing neocortex: neurons from deep and superficial cortical layers, astrocytes and all the temporal sequence of progenitor cell, including oRG cells[307]. Finally, another striking experimental milestone showing the reliability of brain organoids as *in vitro* model for the human developing brain properties, is the recording of functional network-synchronized firing activity coming from the spontaneous electrical waves produced by cortical organoids[308]. Despite the ongoing debate on cell-type fidelity and gene signature in brain organoids possibly impaired by enhanced cellular stress pathways[309], it is now well established that overall organoid gene expression program and epigenome are highly similar to the ones of human foetal tissue and can reproducibly form its cell diversity[301,310]. Lineage relationship between foetal cortex and cortical organoids at single-cell level highlighted that 80% of genes implicated in neocortex disease or evolution that change expression along the foetal cortex lineage, have similar expression profiles in organoids, meaning that organoid cells use similar sets of genes as their foetal counterparts to perform processes such as NPCs proliferation and self-renewal, production of ECM, migration, adherence, delamination, and differentiation[311]. Indeed, cerebral organoids follow the *in vivo* timeline and can recapitulate

early (8-10 PCW) to mid-fetal (22-26 PCW) human brain development much better than their 2D counterparts, maintaining the overall cell type composition of the foetal telencephalon. Moreover, the relationship between epigenetic profile and transcriptional readouts parallelly assessed in foetal cortex and cortically patterned organoids showed that major changes in enhancer activity occur at the transition between neural stem cells and progenitors, and that human-specific enhancers are mostly active in the earliest stages of cortical development, when they target genes that regulate the growth of radial glial cells. In addition, active enhancers described in organoids were enriched for ASD *de novo* variants that disrupt TF binding sites, providing informative mutations in cis-regulatory elements relevant for genetic program driving telencephalon development[312].

The ability to reprogramme human somatic cells into iPSC and then direct those cells towards a specific cell fate has revolutionized the study of human embryo and organ development[313]. This enables the generation of brain organoids that can be applied to diverse research and translational purposes: evolutionary inquiries elucidating the human distinctive developmental characteristics such as the presence of oRG during the proliferative timeframe of the organoid maturation, or disease-modelling of individual-specific features that are preserved during the differentiation process and that are extremely relevant for the study of pathological mechanisms underlying neurodevelopmental disorders[314,315,316,317]. Brain organoids have already been successful in reproducing disease phenotypes in multiple contexts. Cellular phenotypes such as increased rates of proliferation of neural progenitor cells and neuron numbers in ASD individuals, have been shown using different protocols of cortical and cerebral organoids[318]. The complexity of brain organoids and their protracted process of differentiation has shown to be an advantage over its 2D counterparts by uncovering an heterocronicity of the ASD phenotype, a feature masked when pushing the differentiation of iPSCs-derived neurons in a short timeframe that skip progenitor phases[319]. Interestingly, brain organoids have been proven to reproduce different degrees of morphological and dynamic features of neurodevelopment as they have been used not only to model a pathological condition linked to a clear genetic basis such as the microcephaly caused by CDK5RAP2 mutation, but also to recapitulate the cortical enlargement in macrocephaly of idiopathic origin, highlighting the preservation of the genetic background contribution to disease phenotype. Indeed, organoids from idiopathic cases showed differences in NPCs proliferation, neuronal differentiation and synaptic assembly as well as an overproduction of GABAergic inhibitory neurons, chiefly mediated by dysregulation of FOXG1, corroborating the hypothesis of glutamatergic/GABAergic imbalance in ASD[306,317]. In opposition to the study of macroscopic morphological readouts, brain

organoids have also proven to be greatly insightful when scoring mild and heterogeneous manifestations, as in schizophrenia patients where the apparent absence of perceivable alterations in the cortical organization hide a subtle defective apical progenitor proliferation, as well as alteration of neuronal maturation marked by TBR1 and Reelin disproportions[320].

## 2.5 Helsmoortel-Van der Aa Syndrome: the role of ADNP in chromatin regulation

Helsmoortel-Van der Aa syndrome (HVDAS) is an autosomal-dominant neurodevelopmental disorder caused by *de novo* mutations in *ADNP* gene displaying intellectual disability/developmental delay at early onset, autism spectrum disorder, craniofacial dysmorphisms, hypotonia, and congenital heart disease. With a 0.17% estimated prevalence of pathogenic variants in individuals diagnosed with ASD, *ADNP* can be considered as one of the most commonly mutated single genes causing autism[321,322]. The human *ADNP* spans about 40Kb of genomic DNA and consists of five exons encoding for a multi-domain protein harboring nine zinc fingers, a C-terminal homeobox domain, a binding site for the heterochromatin protein 1 (HP1), and a nuclear localization sequence. ADNP has been first characterized as a neurotropic protective factor preventing cell death in suffering neurons and later on described as an embryonic lethal gene in a knock-out mouse model, pointing to its vital role during central nervous system development[323,324,325]. The link with a human pathology has been established in a screening study aimed at uncovering novel candidate genes for ASD susceptibility in a large cohort of autism trios, when recurrent *de novo* protein-truncating variants (i.e., frameshift and nonsense mutations) in *ADNP* were found[18]. Mutations in *ADNP* fall predominantly in the exon 5 of the coding frame of the gene (only the last three exons out of five are translated), where most of the protein functional domains are located. In fact, the detection of mutant mRNAs in the blood of patients and the presence of truncated ADNP upon ectopic expression of the mutant gene in HEK293T cells argues against an haploinsufficient model, as Nonsense Mediated Decay (NMD) is not preventing the degradation of the mRNA carrying the premature termination codon[326,322]. ADNP has been mainly described as the DNA-binding subunit of a recently defined complex called ChAHP, including also CHD4 (already known as chromatin remodelling helicase subunit of repressive complex NuRD) and HP1$\gamma$[327]. In fact, ADNP is essential for ChAHP formation, as the whole complex is unable to assemble in its absence. The molecular characterization performed in mESC has highlighted a pool of around 15,000 binding sites for ADNP, whose targets involves regulatory regions like gene promoters[328]. In mESC,

ChAHP recognizes euchromatic regions establishing local inaccessible domains that, unlike the classic HP1-mediated genomic silencing, do not require the presence of H3K9me3-modified nucleosomes (**Fig. 9**).

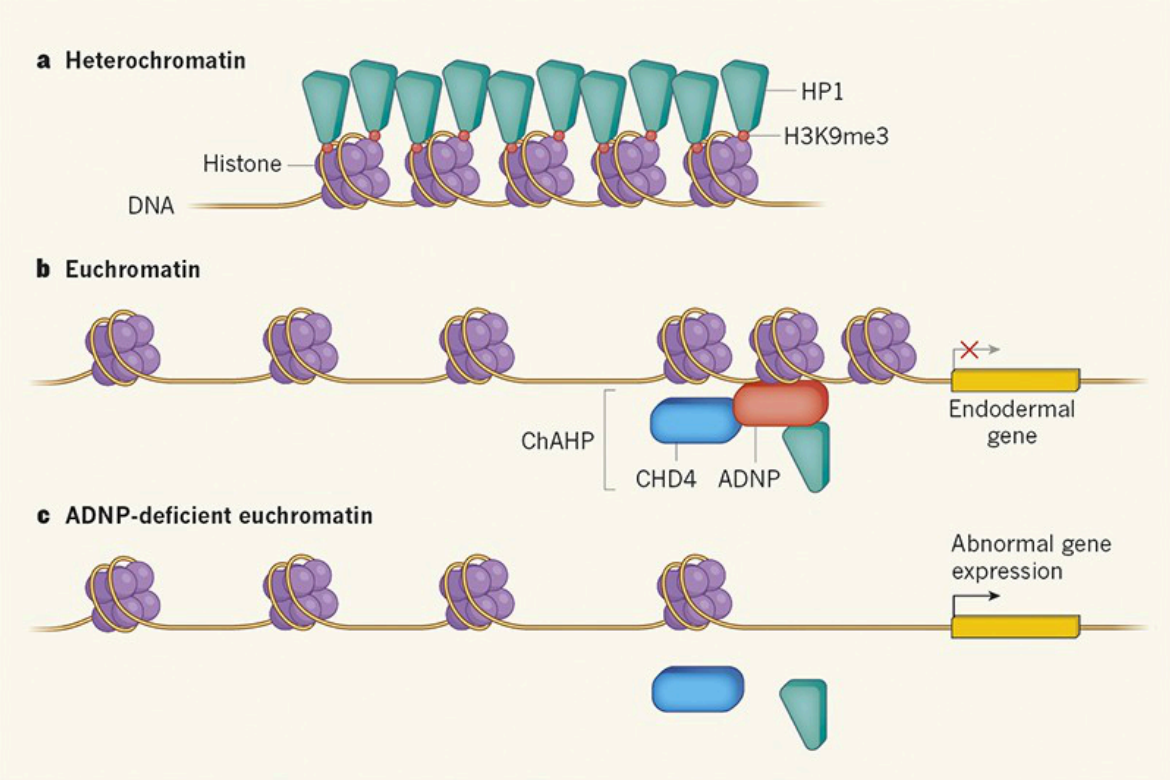



**Figure 9: ChAHP complex controls gene expression through chromatin compaction**. Adapted from Jensen and Lorincz - Nature News and Views (2018)

Experiments using mESCs depleted of ADNP have demonstrated that these cells fail to properly differentiate towards neuroectodermal state, concurrently causing an abnormal activation of endo/mesodermal-specifying genes under the control of ChAHP complex[327]. While this suggests an evident involvement of ADNP in restricting aberrant gene expression during lineage commitment, ChAHP has also been proven to directly modulate spatial chromatin organization; it competes with a subset of binding sites of CTCF that are mostly localized at evolutionarily young transposable element SINE B2, counteracting local chromatin looping[328]. Moreover, serum starved mESC display the recruitment of general transcription factor TFIIIC at a subset of Alu elements marked by ADNP presence and located near cell-cycle genes, altering their chromatin accessibility by direct acetylation of histone H3 lysine18 (H3K18). In turn, Alu elements come in contact with distant CTCF sites near promoter of other cell-cycle genes, which also become hyperacetylated at H3K18, ensuring basal transcription of cell-cycle genes and their re-activation upon serum re-exposure[329]. Altogether, these observations shape a model in which ADNP exerts its fundamental roles in hierarchical degrees of genomic regulatory function. Further

observations have been collected by subsequent studies which validated *ADNP* among a subset of the most significantly enriched genes in NDDs and assessed its related phenotypical features when mutated. In particular, one large exome-seq screening study, places *ADNP* in the category of genes that are associated with severe NDDs, and show an expression pattern that progressively increase from early stages of neuronal maturation towards the mature state of both excitatory or inhibitory lineages[48]. Moreover, distinctive features of brain maturation were measured upon *ADNP* homolog CRISPR-mediated mutagenesis in *X.tropicalis*, such as reduced telencephalon size. In addition, iPSC-derived NPCs mutants showed altered relative abundance of proliferative cells with increased presence of Ki67+ cells[46]. In line with these results, *ADNP* was also found to impact on proper timing of expression patterns driving neurodevelopment: CRISPR-mediated repression of *ADNP* and few other high-confidence ASD-related genes impact on a shared regulatory network causing decreased expression of neuronal maturation markers, resulting in morphological and transcriptional differentiation delay.

# 3. RESULTS

## 3.1 Establishment of patient-derived HVDAS iPSCs cohort

We selected a highly informative cohort of individuals with Helsmoortel-Van der Aa syndrome, who were ascertained for full clinical and molecular diagnosis[321]. This cohort included six individuals homogeneously divided between males and females, carrying *de novo* pathological frameshift or nonsense mutations spanning throughout the functional domains located in the last exon of ADNP (see **Table 1** in Materials and Methods section). Five samples coming from established biobanks or biopsies of healthy individuals were used as control. The settlement of such experimental design is aligned with recommendations from a benchmarking work describing number of lines and clones to be used in iPSC-based studies[330]. Our design (5 controls' lines vs 6 patients' lines) approaches the false discovery rate (FDR) and sensitivity in transcriptomic data observed in experimental designs employing isogenic controls. Moreover, the molecular characterisation of recurring mutations located in different domains of ADNP protein was considered a helpful prerequisite to start building a genotype-phenotype correlation that would allow to expose pathological commonalities among patients, and in parallel, pinpoint mutation-specific features. Our iPSCs cohort is meant to be a meaningful platform to dissect the ground truth of patient-specific pluripotent state through transcriptional and chromatin profiling, and to investigate pathologically relevant lineages upon patterned differentiation. A representative scheme of ADNP structure and patient-specific mutations is depicted in **Fig. 10**.

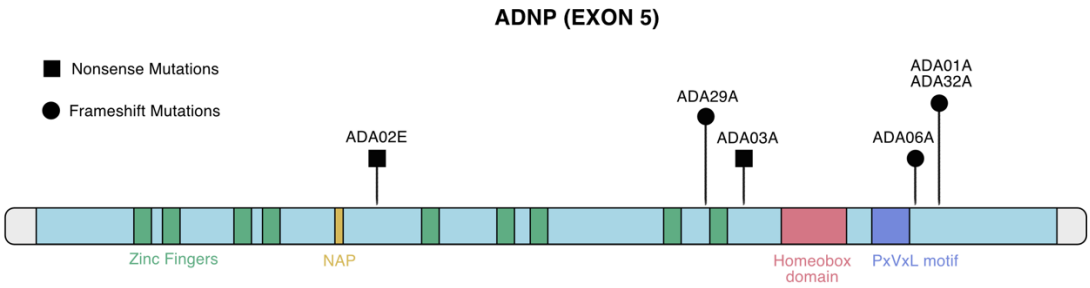



**Figure 10**: ***ADNP* mutational landscape of iPSCs cohort.** Schematic representation of *ADNP* mutations carried by iPSCs cohort relative to protein functional domains located in the exon 5.

We reprogrammed skin fibroblasts into iPSCs by transfection with synthetic mRNAs encoding the five pluripotency factors *POU5F1 (OCT4), SOX2, KLF4, LIN28A and MYC*. In addition to a higher efficiency in terms of the number of reprogrammed colonies, this

integration-free approach circumvents the residual permanence of reprogramming transgenes and its detrimental impact in terms of clones heterogeneity, variability in differentiation proficiency, insertional mutagenesis and reprogramming factor-induced DNA damage[128,329,332]. As expected, transcriptional assessment of iPSC state showed the expression of the main master regulators of pluripotency. Expression of reprogramming factors and other pluripotency markers, along with the downregulation of mesenchymal-specific markers demonstrates the successful reprogramming of the original fibroblasts (**Fig. 11**).

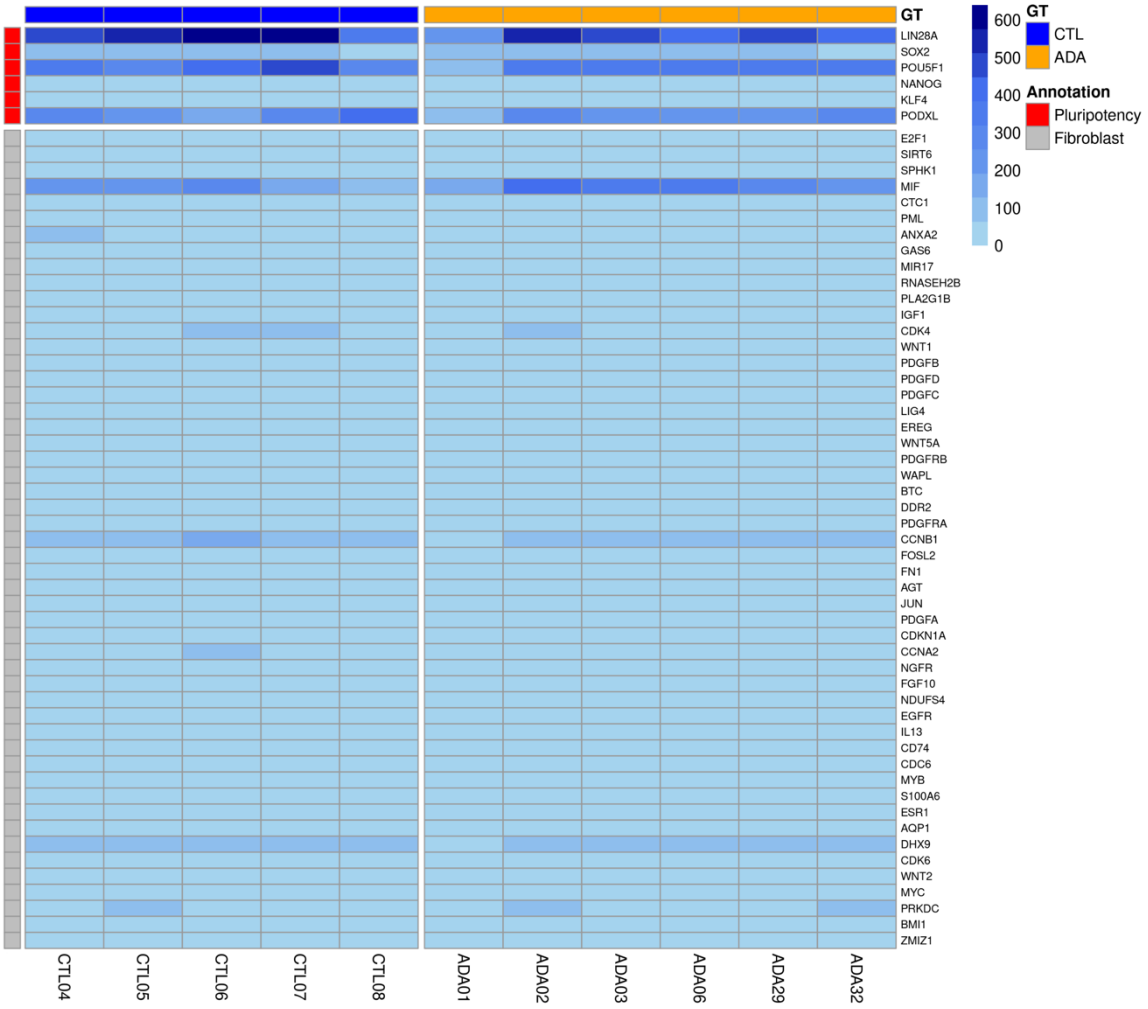



**Figure 11: Transcriptional validation of iPSC state**. Heatmap listing the expression values (RPKM) of each gene associated either with pluripotency (red) or with fibroblast (grey) cell identity, for the iPSC cohort

## 3.2 HVDAS iPSCs are transcriptionally primed to disease-relevant dysregulation

We sought to understand whether HVDAS pluripotent state already harboured any transcriptional imbalance; We therefore used at least 2 clones per each individual in the

cohort to perform bulk RNA-seq. We observed a significant transcriptional dysregulation involving 440 differentially expressed genes (DEGs) (FDR < 0.05, FC > |1.5|), with a prevalent upregulation in HVDAS iPSCs compared to controls (307 upregulated and 133 downregulated), consistently with concomitant upregulation of POLR2A, a gene encoding a subunit composing the DNA-binding domain of RNA polymerase II. (**Fig. 12A,B,C**).

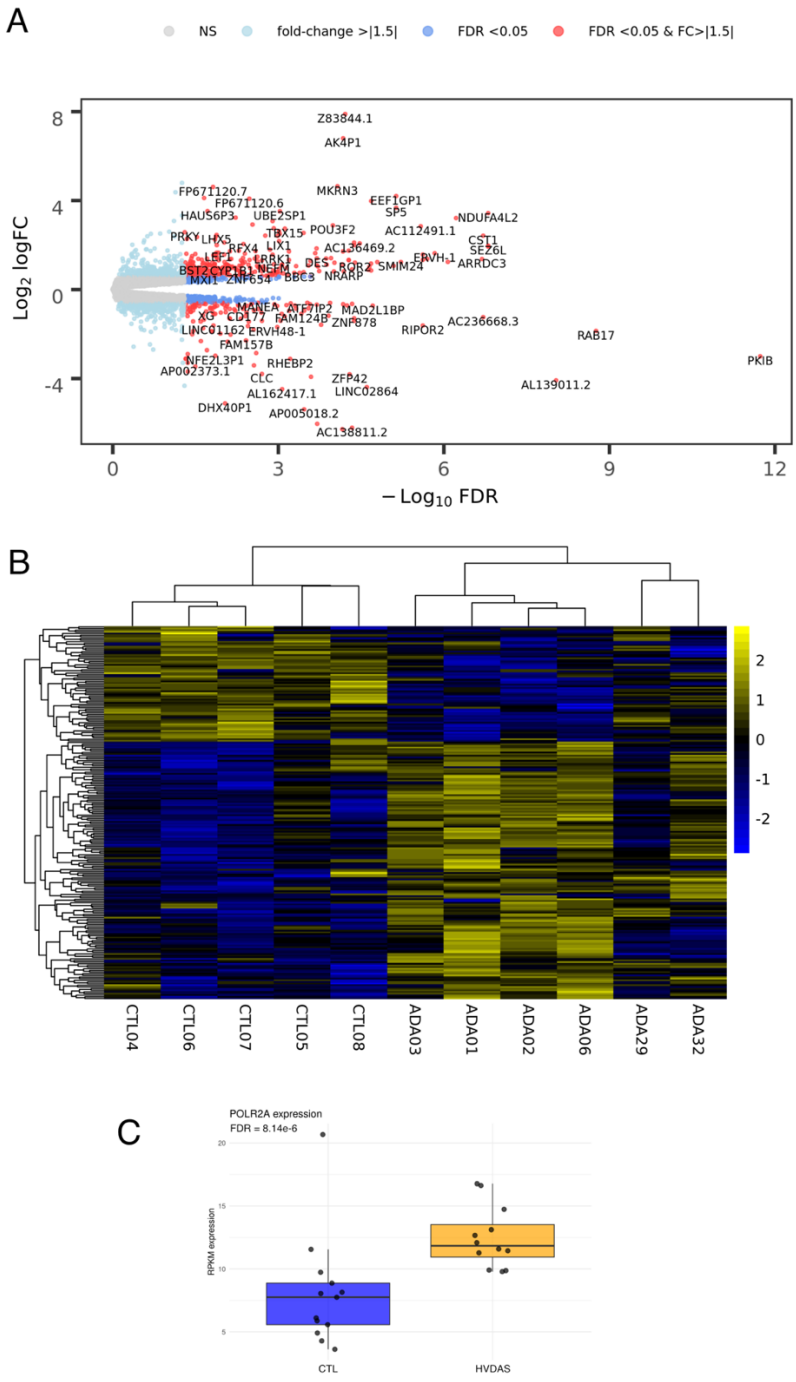



**Figure 12: Differential expression analysis. (A)** Volcano plot of differential expressed genes annotated according to the colour legend. **(B)** Heatmap of differentially expressed genes with corresponding z-score for each gene (rows) and the clustering of each iPSC line (columns). Parameters used in this analysis are: FDR < 0.05 and FC > |1.5|. **(C)** Boxplot of the distribution of POLR2A expression values (RPKM) and statistically significant difference calculated as adjusted p-value (FDR) between control and HVDAS samples.

This is in line with the notion that ADNP (and ChAHP complex) acts as a transcriptional repressor that maintains a compacted chromatin state in the promoter of certain lineage-specifying genes, allowing for proper cell fate acquisition. We first checked the expression levels of *ADNP* in order to understand if it was present in the form of both wild-type and mutant transcripts. We detected a homogeneous level of reads derived from mutant *ADNP* alleles among all the patients, although with an expression pattern different from the expected biallelic ratio, with mutant reads accounting for less than 50% in each patient (**Fig. 13A**). Wild-type *ADNP* mRNA levels were homogenously comparable between controls and patients (**Fig. 13B**).

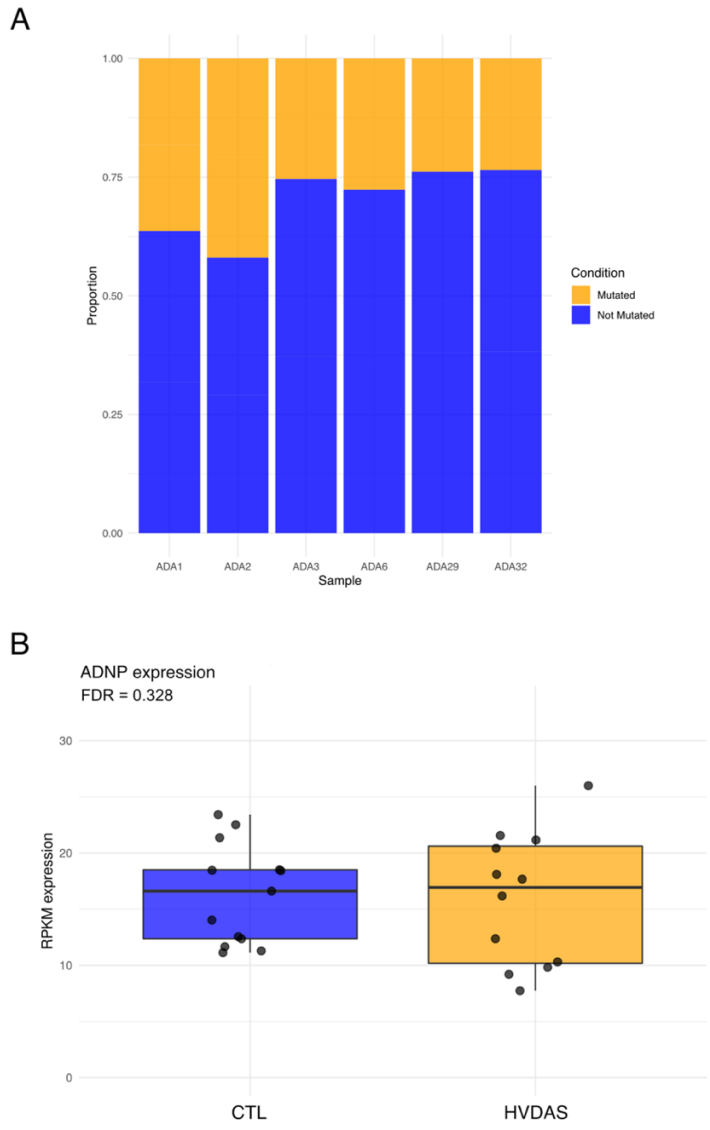



**Figure 13: *ADNP* wild-type and mutant alleles expression levels. (A)** Barplot showing the proportion of reads coming from *ADNP* mutated allele versus reads coming from *ADNP* wild-type allele in each HVDAS lines. **(B)** Boxplot of the distribution of *ADNP* expression values (RPKM) and significance calculated as adjusted p-value (FDR) between control and HVDAS samples.

Principal component analysis (PCA) did not show a clear separation between controls and patients neither in the direction of PC1 nor PC2, suggesting that genotype might be a major contributor to the variation of the dataset but is only accounting for part of that variation, while there might be other variables to explain transcriptional differences (**Fig. 14**).

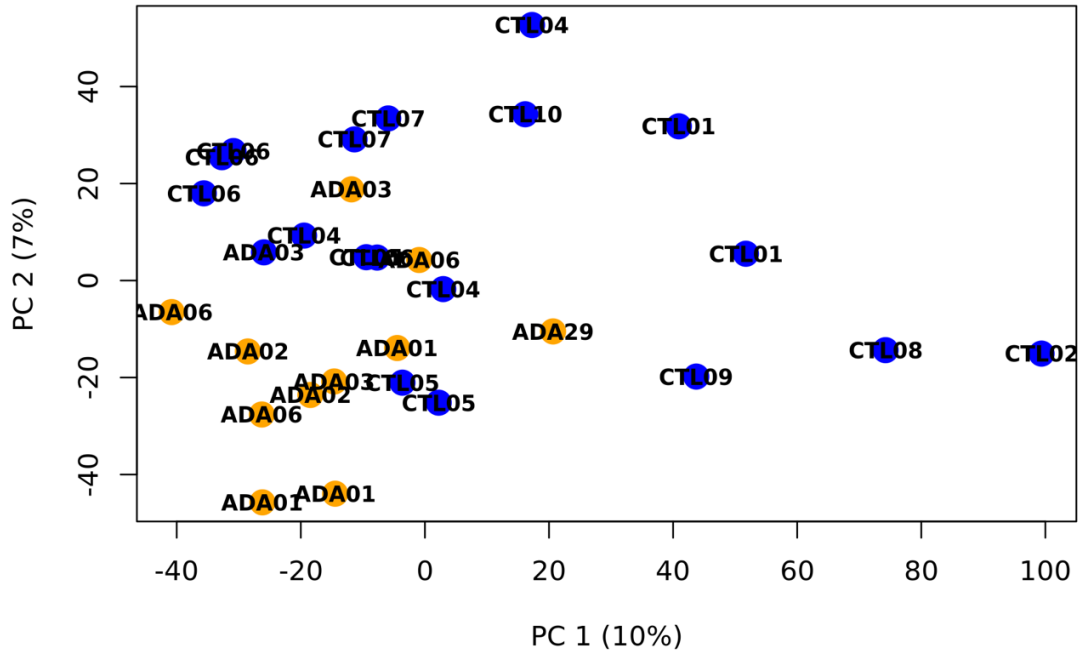



**Figure 14: Dimensionality reduction of iPSC cohort**. Principal component analysis of data points representative of different clones for all the iPSC lines, projected over the first two principal components.

To have a better understanding of the genes interested by the misregulation we looked into the Gene Ontology (GO) categories enriched by differentially expressed genes. GO analysis showed significant enrichments for categories encompassing biological processes related to the developmental function of ADNP. Lineage specification was represented by many different GO terms, in particular the ectodermal and neuroectodermal ones such as "epithelium morphogenesis", "forebrain development" "regulation of nervous system" and the broader "regulation of cell development". On the other hand, GO molecular function showed enrichment for "DNA-binding transcription activity". Moreover, upregulated and downregulated genes enriched for different gene categories, indicating a well-focused dysregulation upon *ADNP* mutations (**Fig. 15 A,B,C**).

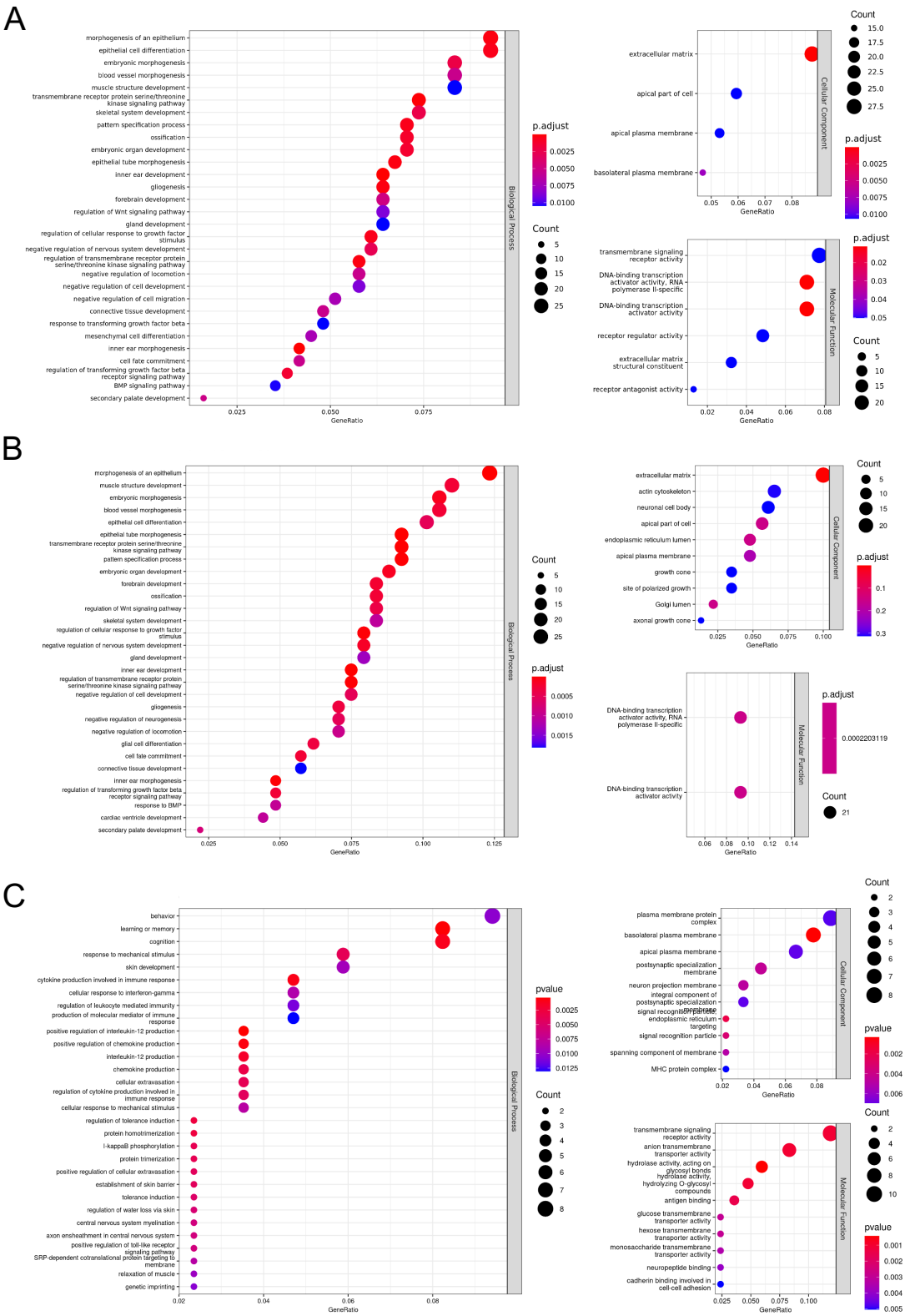



**Figure 15: Gene ontology analysis iPSCs: (A)** Top Biological Processes, Cellular Components and Molecular Functions GO terms enriched by differentially expressed genes, annotated according to the legend indicating gene count and adjusted p-value (FDR). **(B)** Top Biological Process, Cellular Component and Molecular Function GO terms enriched by upregulated genes, annotated according to the legend indicating gene count and adjusted p-value (FDR). **(C)** Top Biological Process, Cellular Component and Molecular Function GO terms enriched by downregulated genes, annotated according to the legend indicating gene count and p-value.

When we looked into details of some significant categories, we found that all the included genes were upregulated in patients compared to controls, and many of such genes were transcription factors involved in neurodevelopment (**Fig. 16 A,B,C**), pointing to ADNP as an upstream player modulating a cascade of other transcription factors involved in neuroectodermal cell fate decision.

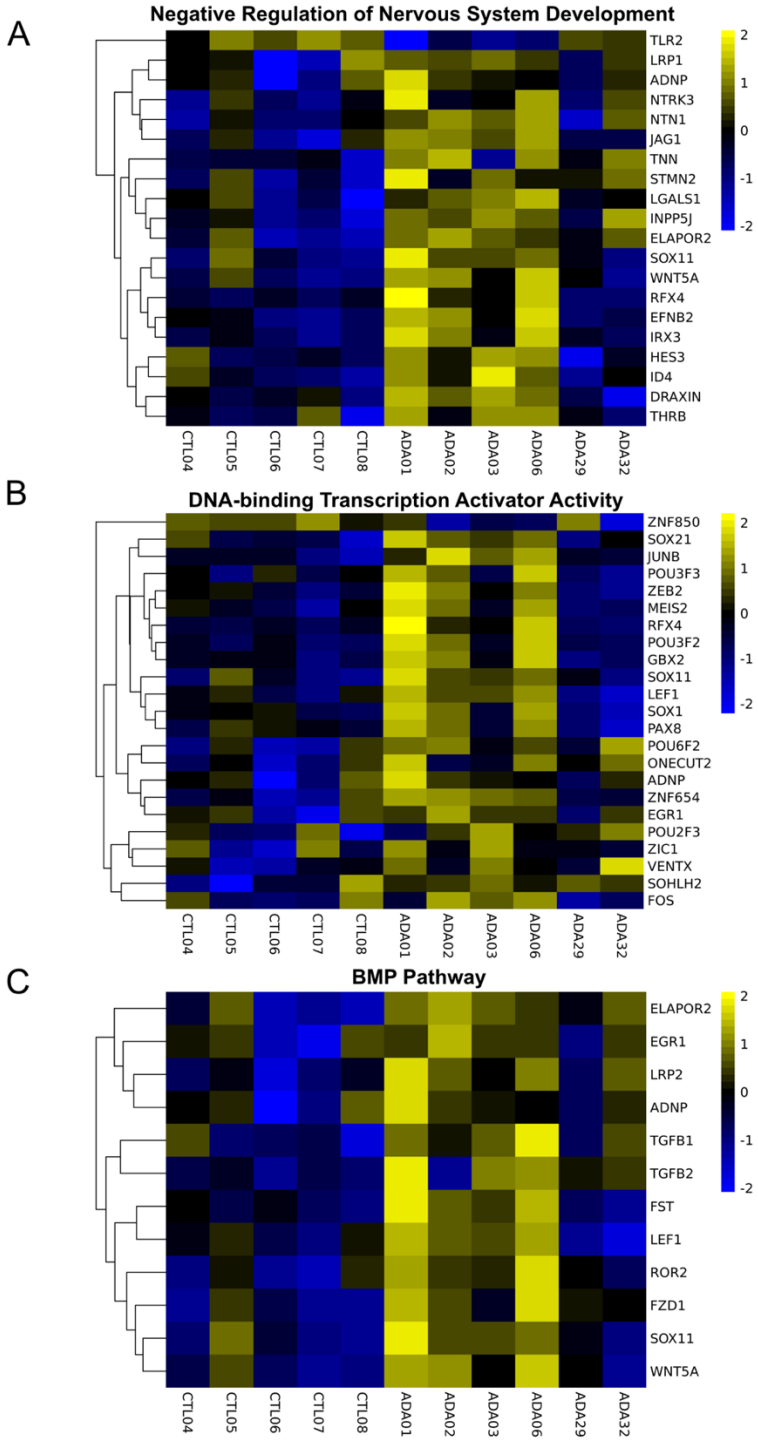



**Figure 16: GO categories-related upregulation. (A-C)** Heatmaps of the differentially expressed genes belonging to the indicated GO categories, with corresponding z-score. The expression levels of ADNP are also indicated for each heatmap.

Given the example of previous works indicating a strong impact of ADNP knock-out on endo/mesoderm-specifying genes, we checked if that was the case also in our model. Only few genes were observed to be differentially expressed among those involved in endoderm and mesoderm lineage commitment, therefore suggesting either a different pool of targets for hADNP compared to the mouse ortholog, or simply an unmatchable phenotype due to different genotype conditions (**Fig. 17A,B**).

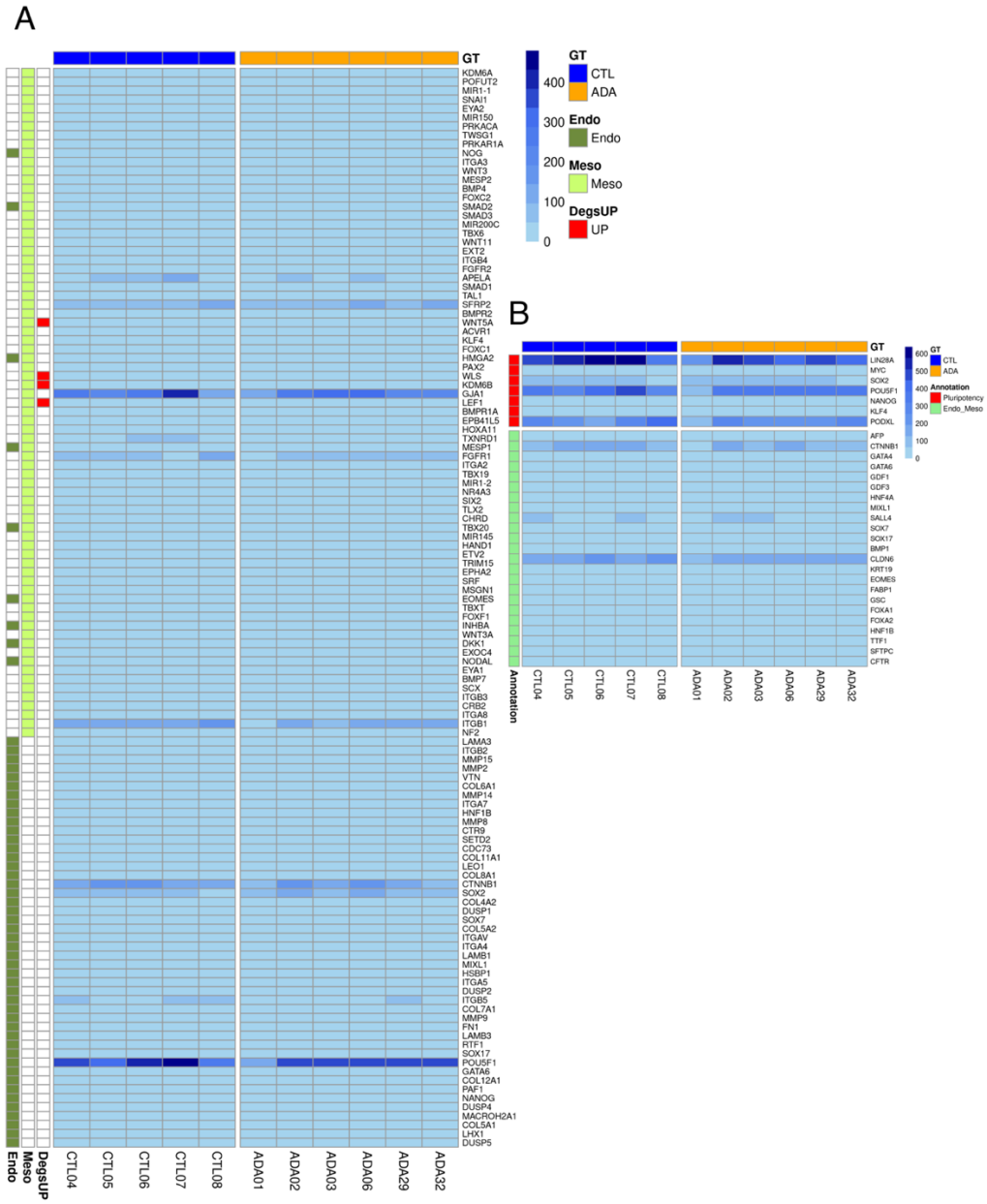



**Figure 17: Transcriptional signature of endoderm and mesoderm-specifying genes. (A)** Heatmap listing the genes relevant for endodermal and mesodermal cell fate specification, with reported RPKM expression values for each iPSC line and annotation of differential expression coming from RNA-seq analysis. **(B)** Heatmap listing the genes relevant for primitive endoderm commitment and the pluripotency genes, with reported RPKM expression values for each iPSC line. No DEGs were found in this list of genes.

We then performed a master regulator analysis to pinpoint key transcription factors whose targets are present in our DEGs list. Interestingly, we found that PRC2 catalytic components EZH2 and SUZ12 were the most significantly enriched hits (**Fig. 18**).

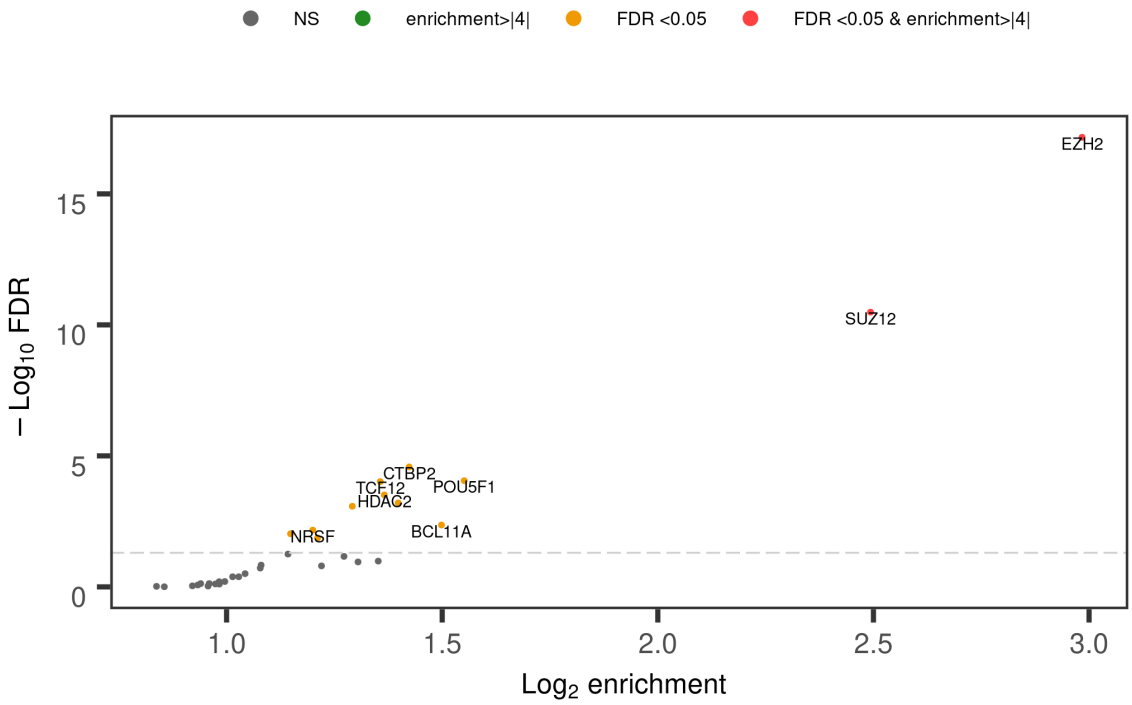



**Figure 18: iPSC Transcription Factor enrichment analysis**. Master regulators identified upon differential expression analysis are ranked based on the FDR of the overlap between their target and the DEGs pool, and the enrichment of such targets among the total number of DEGs.

Given the renowned function of PRC2 in establishing and maintaining a repressed state of developmental genes, its presence among master regulators suggests a coordinated activity with ADNP in the aberrant derepression of lineage-specifying genes upon ADNP impairment. In fact, the overlap between DEGs and known EZH2 targets was heavily disproportionate towards upregulated genes (ten times more than downregulated ones) (**Fig. 19A,B**) and, most of them, as expected, belonged to developmental process or transcription factor activity GO terms (**Fig. 19C**).

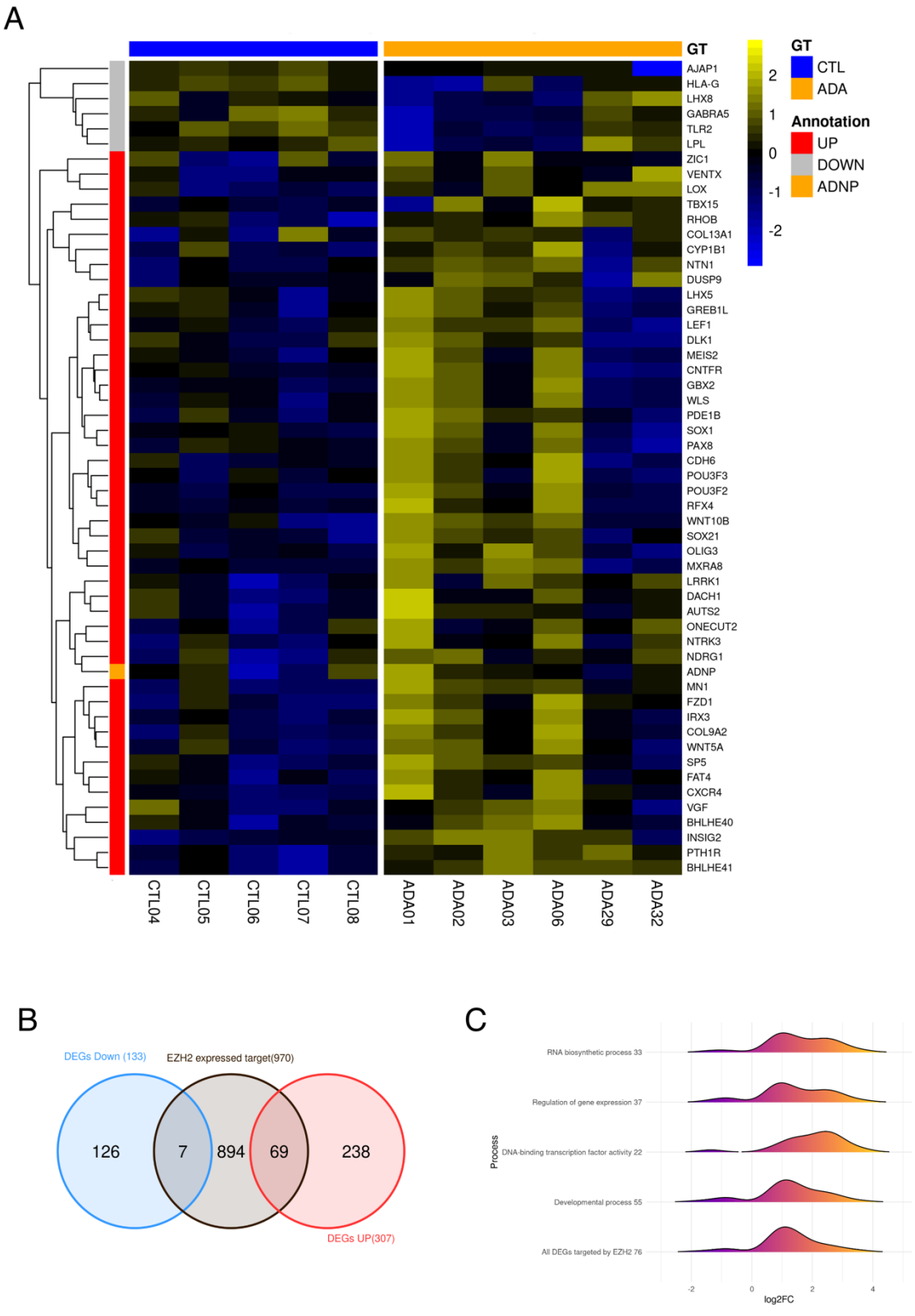



**Figure 19: Prevalent upregulation of EZH2 targets. (A)** Heatmap of the differentially expressed EZH2 targets with the related annotation of differential expression direction (upregulated or downregulated) according to the colour legend. **(B)** Venn diagram of the intersection between the expressed EZH2 targets and downregulated or upregulated DEGs. Hypergeometric test was applied to evaluate the enrichment of EZH2 targets among the upregulated DEGs and downregulated DEGs. DEGs Up: FDR = $7.11e^{-23}$. DEGs Down: FDR = 0.8 **(C)** Ridgeplot showing the fold-change distribution of individual categories enriched by genes targeted by EZH2 together with all the differentially expressed genes targeted by EZH2.

This strongly points to an intervention of EZH2 in the transcriptional impairment triggered by *ADNP* mutations, which leads to the release of many of its direct targets from repressive chromatin state. We then built a map of interactions based on linkages between genes and their respective GO terms enriched from the differential expression analysis, as shown in the Gene-concept network plot (**Fig. 20**).

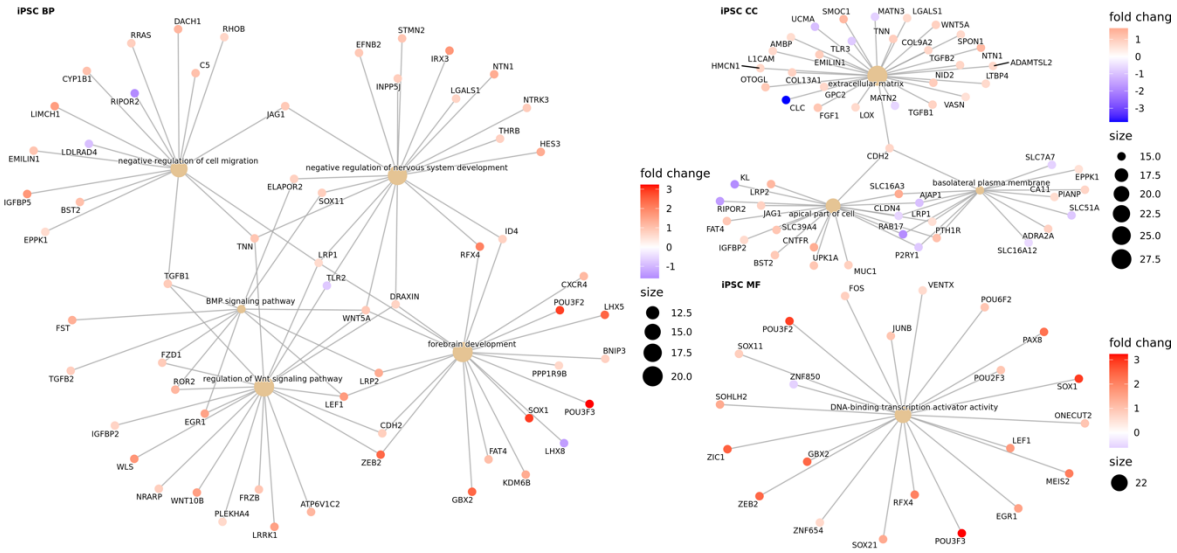



**Figure 20: GO categories networks**. Gene-concept network plot showing the interactions across genes belonging to significant GO categories retrieved from differential expression analysis. The size of each category and the fold-change of each gene are reported for Biological Processes, Cellular Components and Molecular Functions.

Once again, almost all of the genes involved in neurodevelopmental categories were upregulated, and some of them provided an overlapping regulatory hub among WNT signalling, forebrain development, BMP signalling and regulation of migration. In addition, genes included in forebrain development and functionally associated with DNA-binding transcription activation were the most upregulated.

## 3.3 Neural crest stem cells recapitulate iPSCs transcriptional dysregulation

Since the syndromic picture of HVDAS patients include craniofacial abnormalities, we followed up on the transcriptional dysfunction observed already at iPSC stage and differentiated iPSCs into neural crest stem cells (NCSCs). This cell population arises during gastrulation and continues its development upon delamination and migration from the neural tube to different target tissues, acquiring a remarkable differentiation potential which give

rise to peripheral neurons, chondrocytes, melanocytes and cells composing the craniofacial scaffold. We harnessed a published protocol to quickly and efficiently generate homogenous population of NCSCs[333]. We patterned differentiation of 4 HVDAS lines and 3 control lines by cultivating these cells with media enriched in GSK3 inhibitor which enhances the canonical WNT signalling pathway, and SB431542, a small molecule that inhibits BMP/Activin/TGFβ pathways. We first checked the acquisition of neural crest morphology at the end of the differentiation protocol through fluorescence-activated cell sorting (FACS), using established NCSC markers. We observed $p75^{+}$ (NGFR) and $HNK1^{+}$ cells after 21 days of differentiation with minimal iPSCs contamination in the double positive NCSC population (**Fig. 21**).

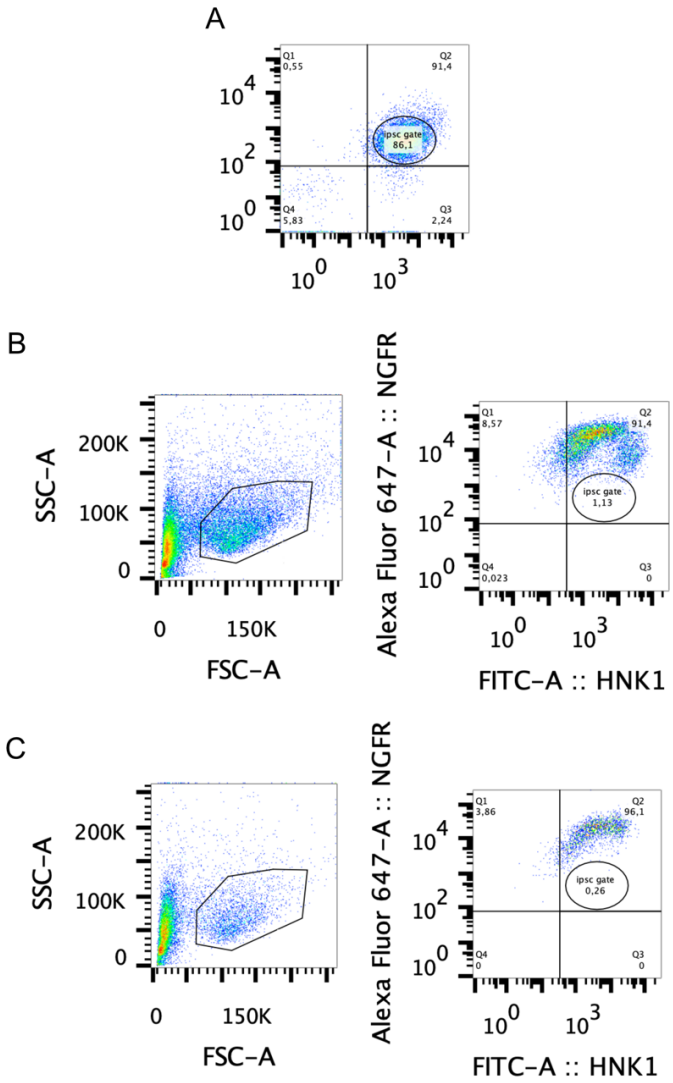



**Figure 21: FACS validation of NCSC differentiation. (A)** FACS quadrants showing iPSC population that served as negative control to set the gate for further identification of NCSC. **(B-C)** Validation of differentiation of one control and one patient iPSC lines via detection of NCSC-specific markers HNK1 and p75 (also known as NGFR).

Subsequently, we performed bulk RNA-seq on the differentiated lines. PCA analysis showed clear distinction between original iPSC and newly derived NCSC populations (**Fig. 22**).

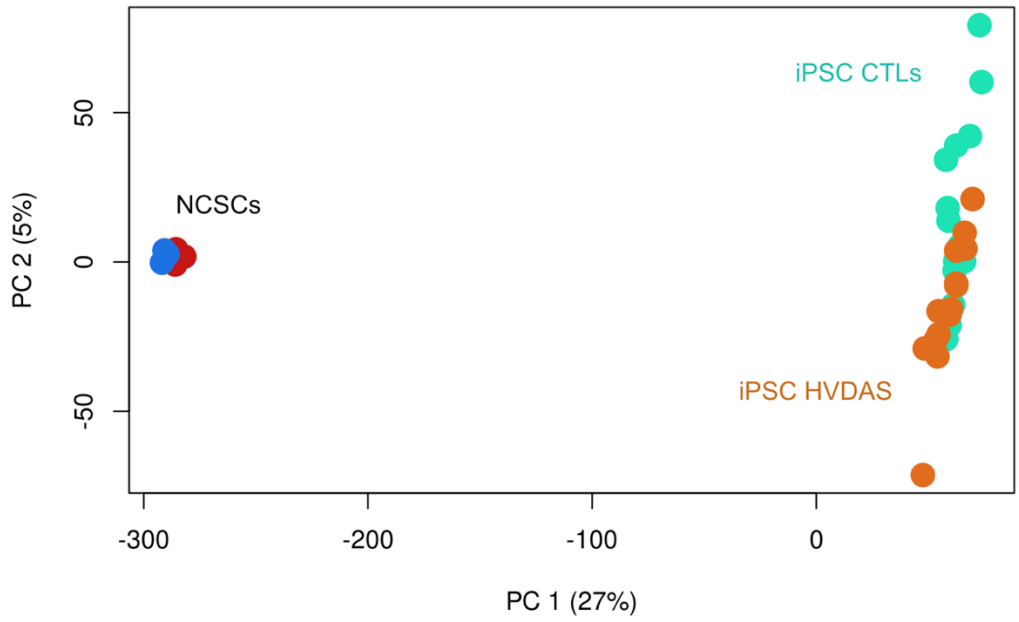



**Figure 22: Dimensionality reduction of NCSC cohort compared to iPSC cohort**. Principal component analysis showing data points representative of each NCSC line and iPSC line, projected over the first two principal components.

As in the case of iPSCs, we observed transcriptional upregulation also for NCSCs, with 149 genes upregulated and 105 downregulated (FDR < 0.05, FC > |1.5|) in the HVDAS compared to controls (**Fig. 23A,B**).

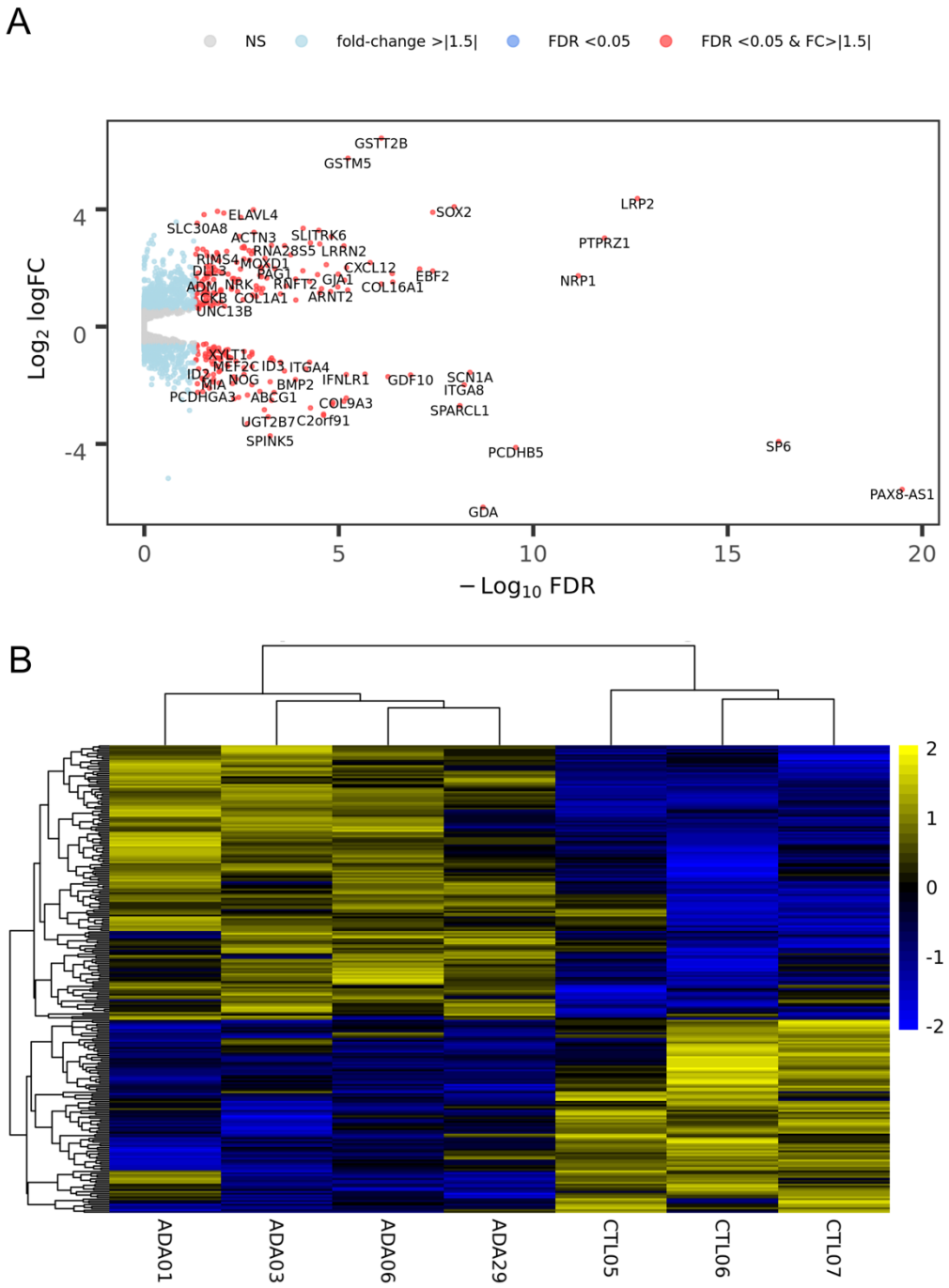



**Figure 23: NCSC differential expression analysis. (A)** Volcano plot of differential expressed genes annotated according to the colour legend. **(B)** Heatmap of differentially expressed genes with corresponding z-score for each gene (rows) and the clustering of each iPSC line (columns). Parameters used in this analysis are: FDR < 0.05 and FC > |1.5|.

Consistently with iPSCs, GO analysis showed an enrichment for categories involved in general development, in particular for nervous system development, with the top enriched category ("cell morphogenesis involved in neuron differentiation") showing mostly upregulated genes in the patients (**Fig. 24A,B**), concordantly with previous observation on general derepression caused by *ADNP* mutations.

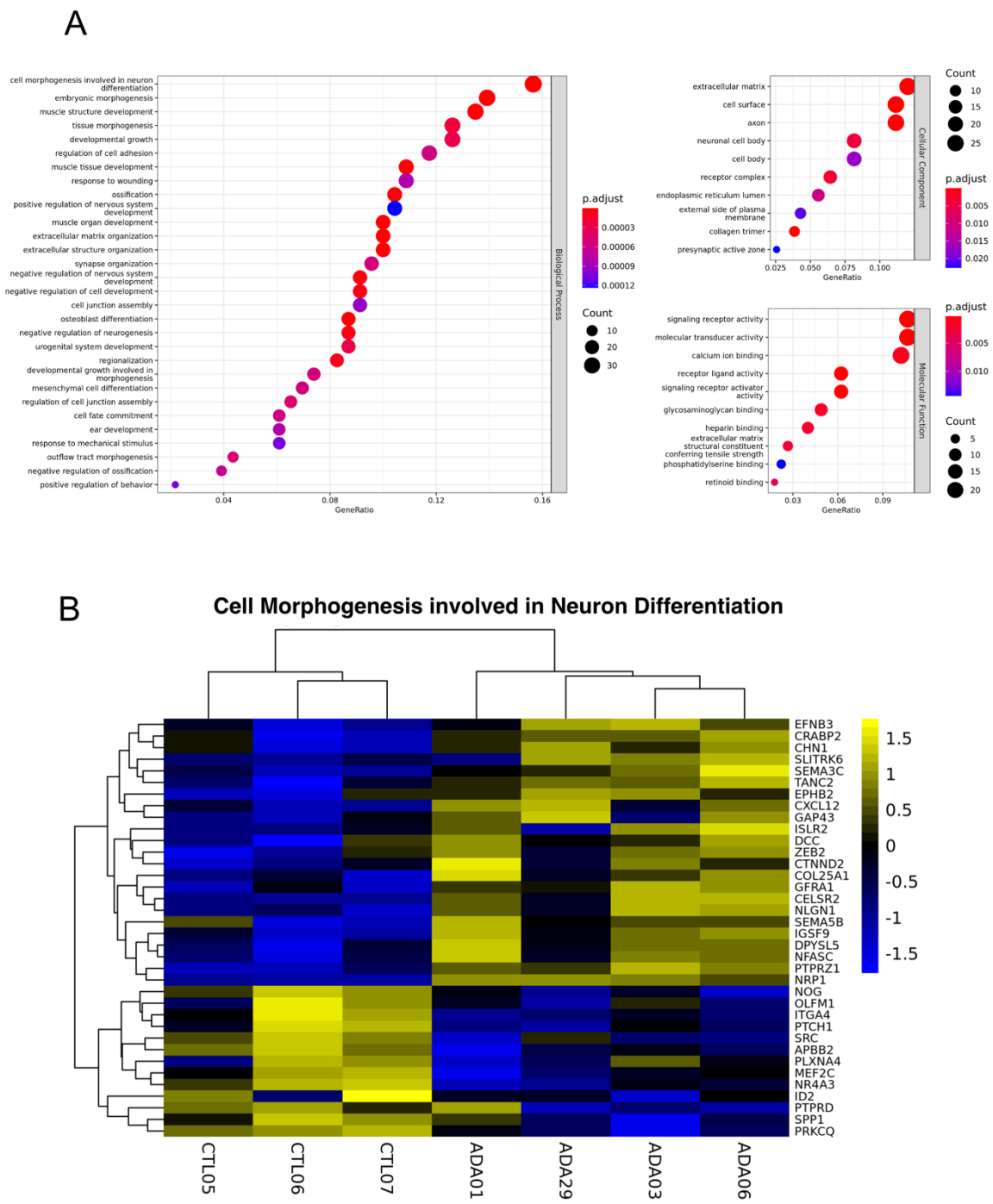



**Figure 24: Gene ontology analysis NCSCs (A)** Top Biological Processes, Cellular Components and Molecular Functions GO enriched terms retrieved from differentially expressed genes, annotated according to the legend indicating gene count and adjusted p-value (FDR). **(B)** Heatmap showing the differentially expressed genes belonging to the indicated GO category, with corresponding z-score.

As in iPSCs, GO cellular components enriched for extracellular matrix; suggesting a common ground of both cell types linking connective tissue alterations with certain HVDAS morphological aspects, Moreover, like in iPSCs, master regulator analysis exposed SUZ12 and especially EZH2 as the most enriched transcription factors targeting the majority of the DEGs (**Fig. 25**).

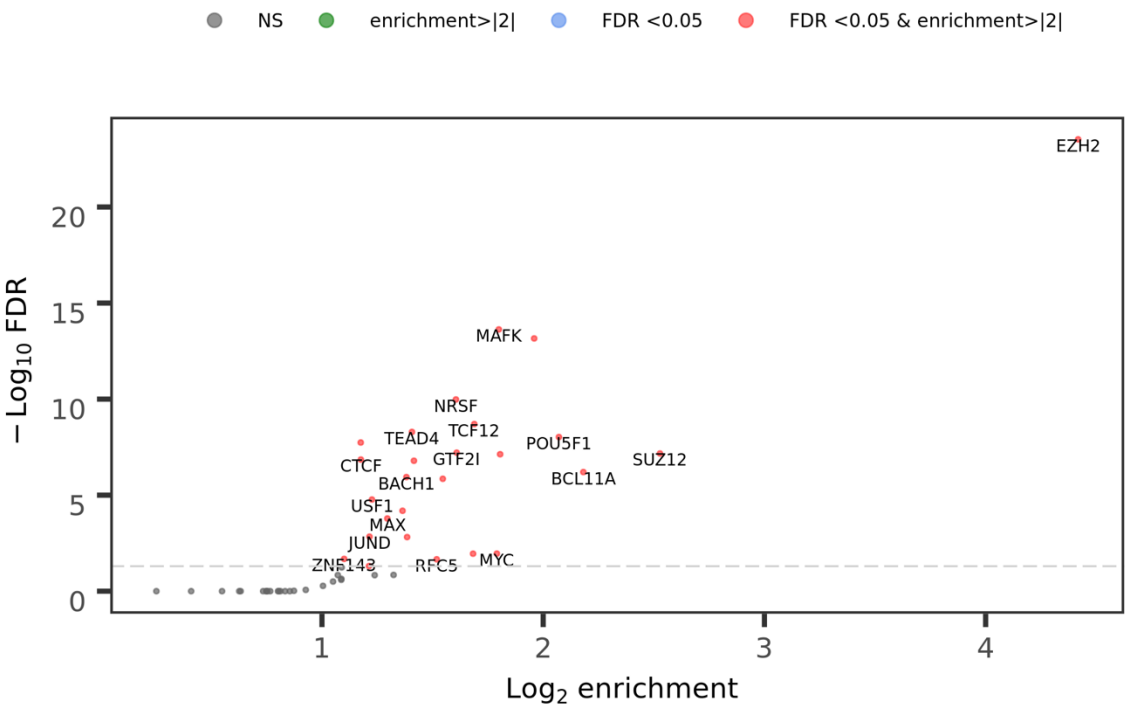



**Figure 25: NCSC Transcription Factor enrichment analysis**. Master regulators identified upon differential expression analysis are ranked based on the FDR of the overlap between their target and the DEGs pool, and the enrichment of such targets among the total number of DEGs.

This piece of data provides one more hint to a putative interplay between ADNP and PRC2 in establishing cell identity through coordinated gene expression programs possibly initiated at the pluripotent state and lasting towards more differentiated cell-types.

## 3.4 CRISPR/Cas9 engineering for endogenous tagging of ADNP locus

Given earlier evidences in mouse embryonic stem cells and knowing the structural properties of human ADNP, we sought to understand where ADNP exerts its DNA-binding activity and whether that binding could be informative on ADNP regulatory targeting of dysregulated genes. To that end, we performed ChIP-seq experiments in our cohort of patient and control lines. However, due to the lack of a commercial ChIP-grade antibody, we decided to endogenously tag the *ADNP* gene by designing a CRISPR-based engineering that would result in a FLAG-tagged protein. This strategy not only allows the detection of ADNP binding sites on chromatin using an ordinary anti-FLAG antibody, but importantly, preserves the cell-line-specific dosage of the protein, avoiding episomal transient overexpression or viral-associated insertional mutagenesis. Therefore, we designed a construct harbouring: 1) Kozak sequence to boost initiation of translation starting from the ATG of the hygromycin, 2) hygromycin resistance gene for the selection of edited colonies, 3) P2A self-cleaving peptide allowing the translational separation of the hygromycin resistance protein and the downstream tagged ADNP, 4) 3XFLAG sequence to tag the 5' of the translated region,

which begins at the exons 3 of *ADNP*, resulting in a N-terminus FLAG-tagged endogenous protein, 5) 5' and 3' 1000 base pairs long homology arms (HA) flanking the elements mentioned above for the homologous recombination at the cut site. This construct was then electroporated along with the Cas9 and gRNA assembled to form a ribonucleoprotein (RNP) complex, using a gRNA with high on-target score and minimal exonic off-targets (calculated according to Benchling online software). The protospacer adjacent motif (PAM), the three nucleotides sequence that triggers the Cas9 to cut, was located in the vicinity of the ADNP endogenous ATG in order to enhance the homologous recombination precisely at the starting of exon 3, where the protein starts being translated. Moreover, because the 3'HA falls in a coding sequence (exon 3), we changed one base of the arm that encompass the PAM in order to impede the Cas9 to cut the construct or the newly edited region. Finally, we designed the construct to target the 5' of the translated portion of the gene instead of the 3' so to enable the endogenous tag of all the patient lines with only one construct, suitable for all the downstream patient-specific C-terminal mutations. A schematic representation of the editing strategy is depicted in **Fig. 26**.

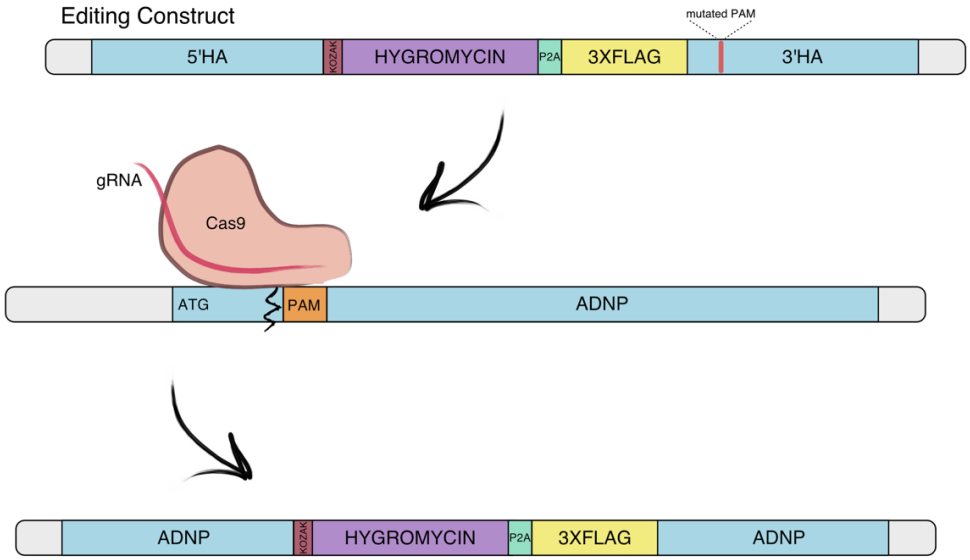



**Figure 26: Schematic illustration of CRISPR/Cas9 strategy for endogenous tagging of *ADNP*.** CRISPR ribonucleoprotein complex mediates the insertion of the editing construct at the PAM situated close to the ATG located in the exon 3 of ADNP locus through homologous recombination, thus resulting in the N-terminus FLAG-tag of ADNP protein. Hygromycin resistance gene is knocked-in to allow the selection of the edited clones. P2A self-cleaving peptide allows the translational separation of hygromycin resistance protein and FLAG-ADNP.

Of note, the employment of Cas9 protein instead of a plasmid-based Cas9 transfection allows for a shorter editing time-window, increasing the ratio of on-target over off-target events. Indeed, unlike plasmids, RNPs cleave genomic DNA rapidly after transfection and persist in cells for only about a day before being degraded, thus minimising unintended cuts in the

genome and maximising accurate editing[334,335,336,337]. We electroporated the tagging construct and the RNP complex in one iPSC control line (CTL08A) and one iPSC patient line carrying the mutation p.Tyr719* (ADA03A), and selected colonies for 10 days culturing cells in medium complemented with hygromycin. We then screened the selected clones with a panel of different PCRs to check for the presence of the whole construct and the resulting genotype. All of the screened clones coming from the control iPSCs showed recombination in heterozygosity, while for the patient's line we managed to isolate two homozygous clones (**Fig. 27A,B,C**). We finally validated with Sanger sequencing the correct in-frame knock-in of the construct in the heterozygous control clone (hereafter referred to as CTL08-FLAG) and the homozygous HVDAS clone (hereafter referred to as ADA03-FLAG) (**Fig. 27D**).

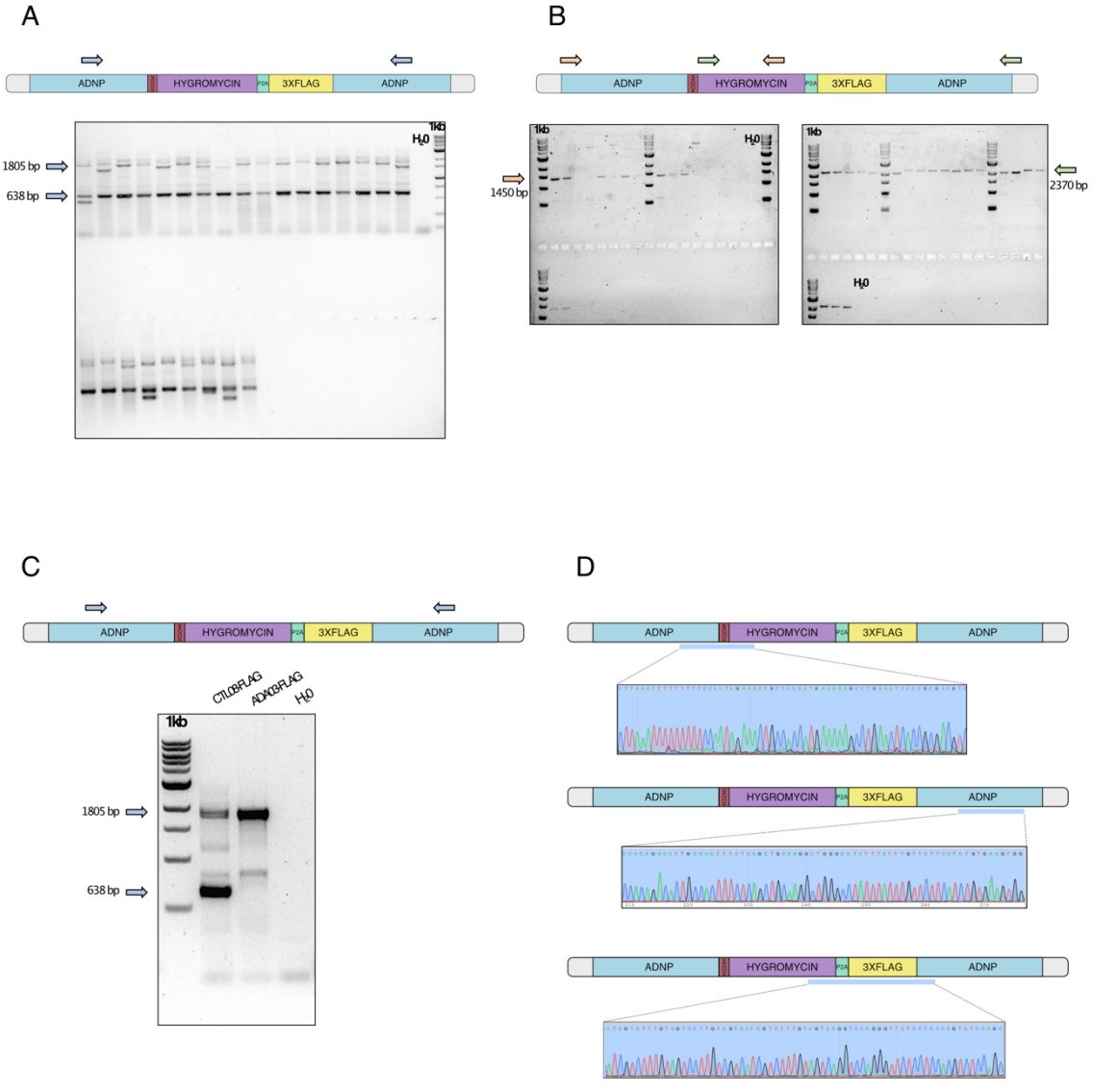



**Figure 27: PCR screening of CRISPR/Cas9-edited clones. (A-C)** Three different PCR reactions with related gel electrophoresis showing the amplification of different portions of the edited locus. In **C**, the PCR allows the genotypic discrimination between iPSC edited lines: CTL08-FLAG is heterozygous while ADA03-FLAG is homozygous. **(D)** Representative Sanger sequencing electropherograms showing the in-frame integrity of portions of the edited locus.

To assess the presence of the FLAG-tagged protein in both lines we performed Western blot and effectively observed the correct size band for both control and patient wild-type ADNP (approximately 150 KDa) (**Fig. 28**). Notably, we could not see any trace of the truncated protein in ADA03A-FLAG, which argues in favour of nonsense-mediated decay occurring upon transcription of the ADNP mutated allele.

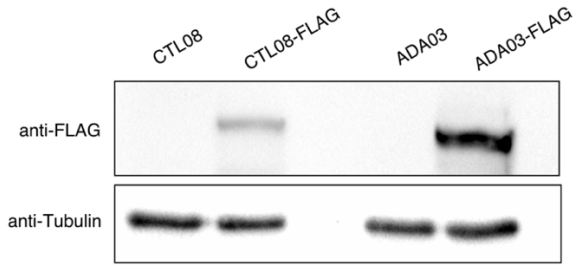



**Figure 28: ADNP-FLAG detection in CTL08-FLAG and ADA03-FLAG iPSCs**. Western Blot showing the presence of ADNP-FLAG protein in both CTL08-FLAG and ADA03-FLAG iPSC lines. Parental untagged lines served as negative controls. Tubulin served as loading positive control.

## 3.5 ADNP genome-wide mapping highlights preferential localization at transposable elements

After successful validation of FLAG-tagged lines we performed ChIP-seq experiments in three replicates for each line. PCA analysis did not show a clear distinction between CTL08-FLAG and ADA03-FLAG replicates, while a neat separation was observed compared to input (**Fig. 29A**). CTL08-FLAG ChIP-seq revealed ~33,000 regions significantly enriched for ADNP, while we observed a severe reduction in the number of binding sites of ADA03-FLAG ChIP-seq, amounting to ~9,000 regions (**Fig. 29B**), almost completely overlapping with CTL08-FLAG peaks (**Fig. 29C**).

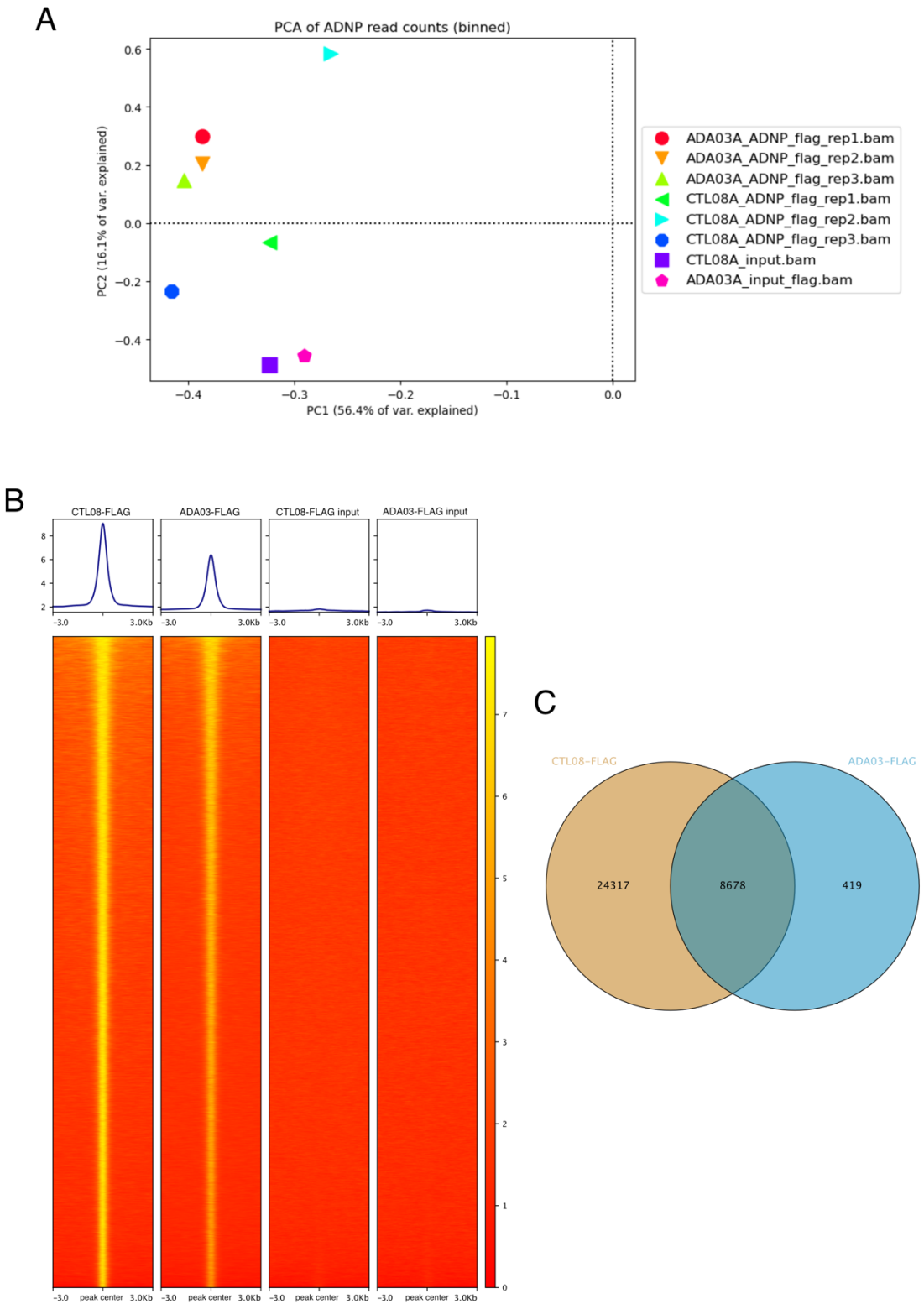



**Figure 29: Dimensionality reduction and genome-wide mapping of ADNP ChIP-seq. (A)** PCA representing the distribution of variation of each ChIP-seq replicate performed in CTL08-FLAG and ADA03-FLAG. **(B)** Heatmap and corresponding enrichment curve of ADNP ChIP–seq across all significant top peaks in the FLAG-tagged iPSCs. Each row represents a 6-kb window centred on peak midpoints, sorted by the average ADNP ChIP-seq signal intensity of 3 technical replicates. Input signals for the same windows are shown on the right. **(C)** Venn diagram of the overlap of top significant peaks in CTL08-FLAG and ADA03-FLAG.

GO analysis on ADNP-bound genes in both lines showed significant enrichments for cell morphogenesis categories, in particular axon and synapse organization (**Fig. 30**).

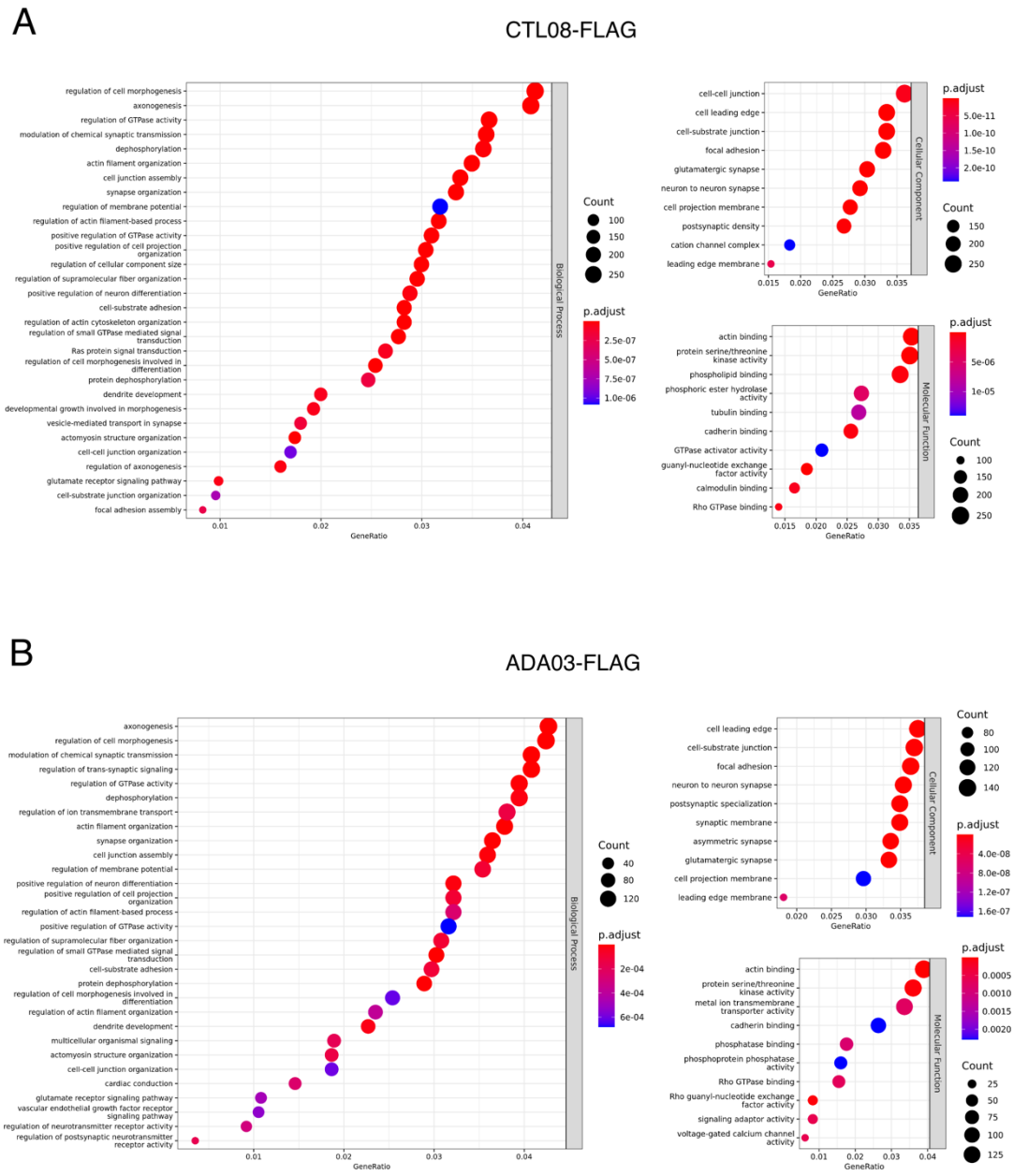



**Figure 30: ChIP-seq Gene Ontology analysis. (A-B)** Top Biological Processes, Cellular Components and Molecular Functions GO terms enriched by ChIP-seq ADNP targets, annotated according to the legend indicating gene count and adjusted p-value (FDR).

We analysed the genome-wide distribution of ADNP signal and found its presence at many different genomic features, including promoters, in line with previous results showing that ChAHP complex targets regions associated with protein-coding genes (**Fig. 31A**)[327]. Interestingly, the majority of the peaks (around 70%) localized outside of the coding sequences, mainly at intronic and distal intergenic regions. Moreover, when we looked at

the abundance of ADNP relative to Transcription Start Site (TSS), circa 65% of its binding sites were located farther than 10 kb from the TSS (**Fig. 31B**).

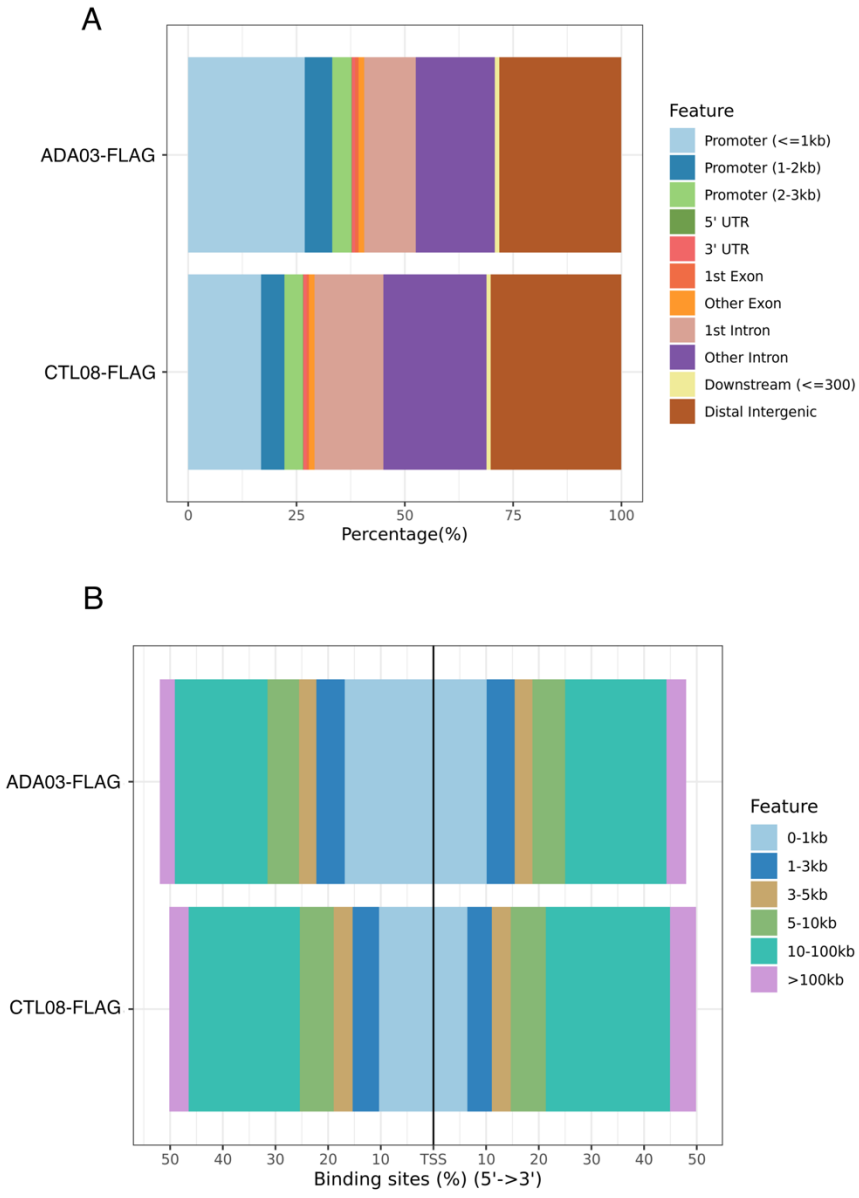



**Figure 31: ADNP genome-wide distribution and relative-to-TSS. (A)** Percentage distribution of ADNP peaks at genomic features in CTL08-FLAG and ADA03-FLAG. **(B)** Percentage distribution of ADNP peaks relative to TSS in CTL08-FLAG and ADA03-FLAG.

Indeed, we found that ADNP was mostly enriched at transposable elements (Alu in particular), where its occupancy was much higher than in any other genomic features, and proportionally reduced in ADA03-FLAG (**Fig. 32A,B**). Moreover, given the repressive role of ADNP and the chromatin release upon its knock-out[327], we performed ATAC-seq experiments in order to assess whether the same mechanism held true in our iPSCs. We then

probed the chromatin state in the top enriched ADNP-bound regions, namely the transposable elements. We found no evident differences in the chromatin compaction around ADNP binding sites, meaning that, unlike mouse ADNP KO, the chromatin state is not plainly altered by its loss and could be influenced by a different mechanism in our genotypic context (**Fig. 32C**).

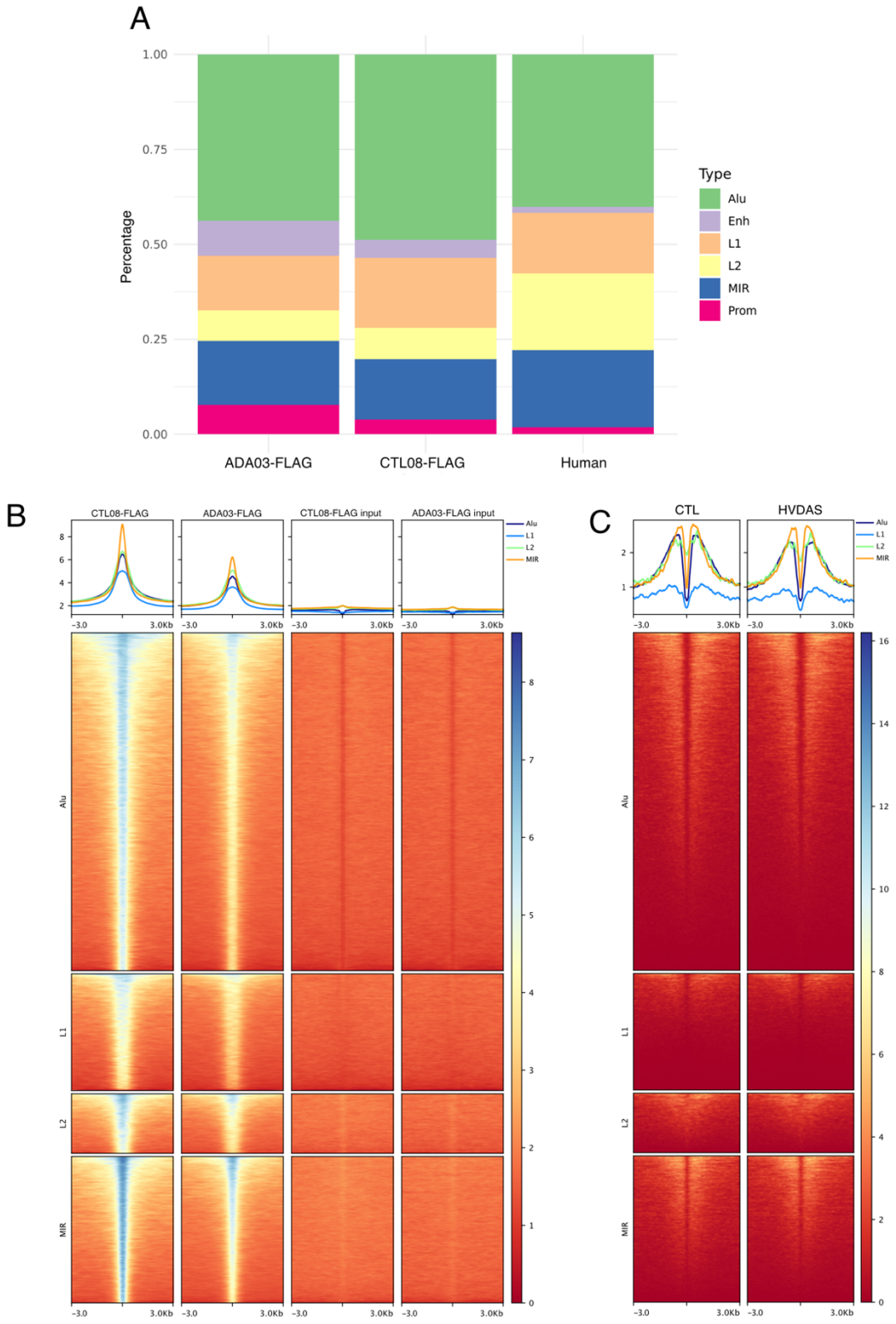



**Figure 32: ADNP distribution at the most enriched genomic features. (A)** Percentage abundance of ADNP-bound genomic features in CTL08-FLAG and ADA03-FLAG compared to the abundance of the same features in the human genome. **(B)** Heatmap and corresponding enrichment curve of ADNP ChIP–seq across the transposable element (Alu, L1, L2, MIR) significant top peaks in the FLAG-tagged iPSCs. Each row represents a 6-kb window centred on peak midpoints, sorted by the average ADNP ChIP-seq signal intensity of 3 technical replicates. Input signals for the same windows are shown on the right. **(C)** Heatmap and corresponding enrichment curve of ADNP ATAC–seq across the transposable element (Alu, L1, L2, MIR) significant top peaks in the all the iPSCs cohort. Each row represents a 6-kb window centred on peak midpoints, sorted by the average ATAC-seq signal intensity.

Given the widespread presence of ADNP at the Alu elements, we investigated in further details this specific TE class. Alu elements are a subclass of primate-specific repetitive elements that belong to the class of retroelements called SINE (short interspersed nuclear elements) and that collectively account for 11% of the human genome, greatly contributing to the plasticity of primates evolution[338,339]. We found that the most ADNP-enriched Alu elements were the evolutionarily young ones (Y subfamily), in line with a recent work demonstrating that ChAHP complex in mESC is principally positioned at least diverged SINE B2, the mouse homologs of the human Alu elements (**Fig. 33A,B**)[328].

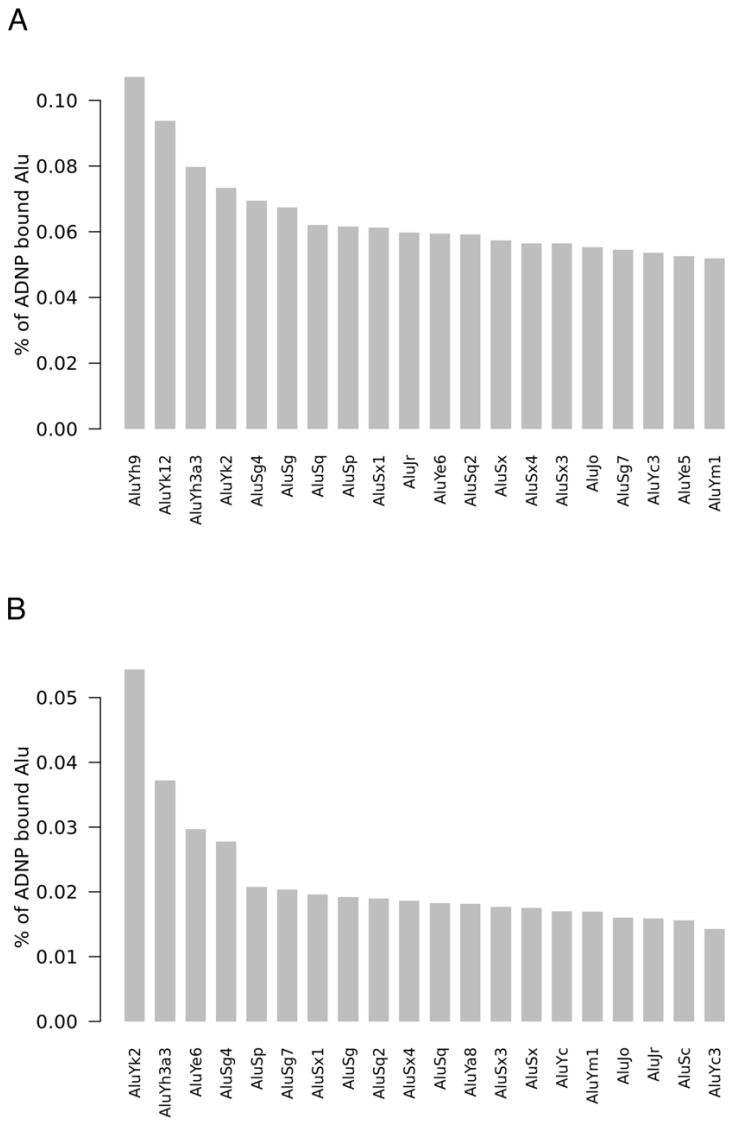



**Figure 33: ADNP binding sites enriched in Alu subtypes.** List of Alu elements subfamilies ranked according to the number of ADNP binding sites.

We then asked what could be the functional effect of ADNP occupancy at Alu on nearby or distal genes, and discovered that while ADNP is only rarely found to directly bind promoter

sequences, a higher enrichment was observed in different subclasses of Alu elements that were in close proximity to promoters **(Fig. 34A-F)**.

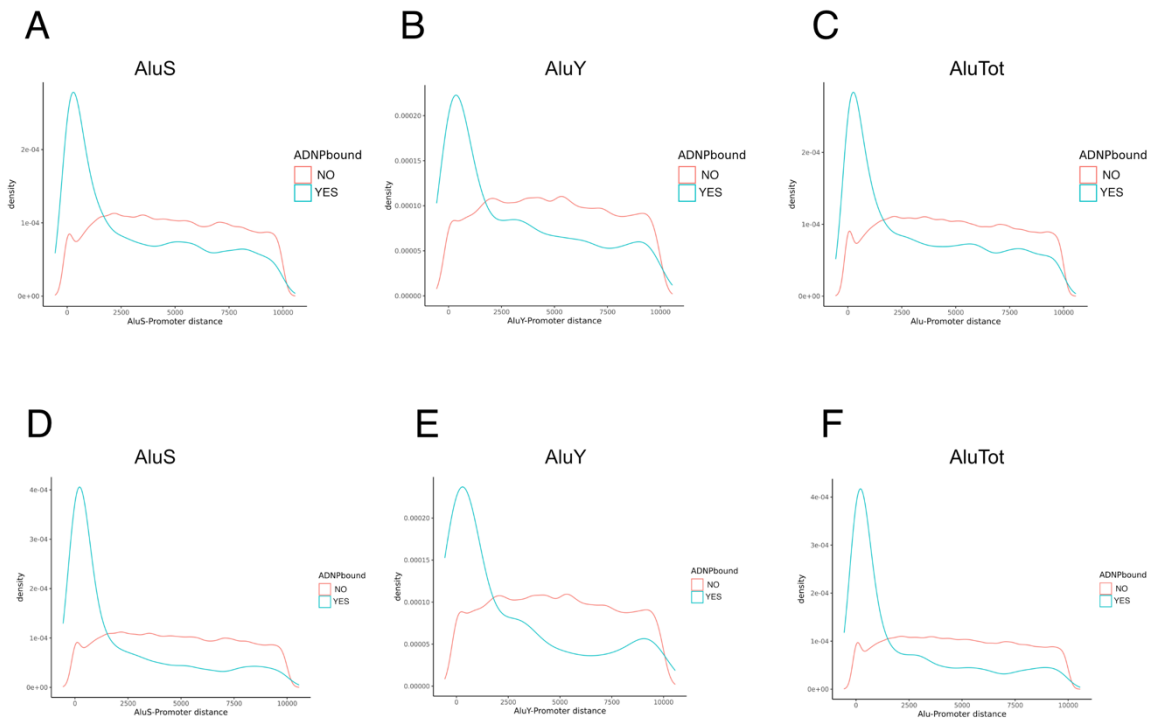



**Figure 34: ADNP occupancy at Alu elements close to promoters. (A-F)** Frequency distribution of ADNP-bound Alu elements within 10,000 bp from the TSS. **A-C** refer to CTL08-FLAG ChIP-seq dataset while **D-F** refer to ADA03-FLAG ChIP-seq dataset.

Given the widespread dissemination of Alu in the genome, these data suggest that ADNP might operate a pervasive control on gene expression. This prompted us to investigate to what extent ADNP binding sites could explain the transcriptional dysregulation observed in iPSCs. We then isolated ADNP peaks falling into gene regulatory regions (i.e., promoters and enhancers) in our RNA-seq dataset of 16,756 expressed genes, and found 5,379 genes targeted by ADNP in the pool of 7,443 total gene-associated ADNP peaks produced by CTL08-FLAG ChIP-seq. Among the 440 DEGs (FDR < 0.05, FC > 1.5), 107 were associated with ADNP peaks, of which 88 upregulated and 19 downregulated, meaning that ~20% of the total DEGs number has a gene regulatory element targeted by ADNP, resulting in significant enrichment among the total number of DEGs, especially in the upregulated ones. (**Fig. 35A**). When comparing the peaks of ADNP in CTL08-FLAG with the ones in ADA03-FLAG, we found 2,264 genes with a loss of ADNP binding sites. (**Fig. 35B**). Again, this ensemble (i.e., gene-associated promoters/enhancers losing ADNP) was significantly enriched among the DEGs, and mostly overlapping with the upregulated genes. Overall, these results reveal that ADNP directly targets a pool of differentially expressed genes, and that these genes are preferentially derepressed and upregulated upon *ADNP* mutations.

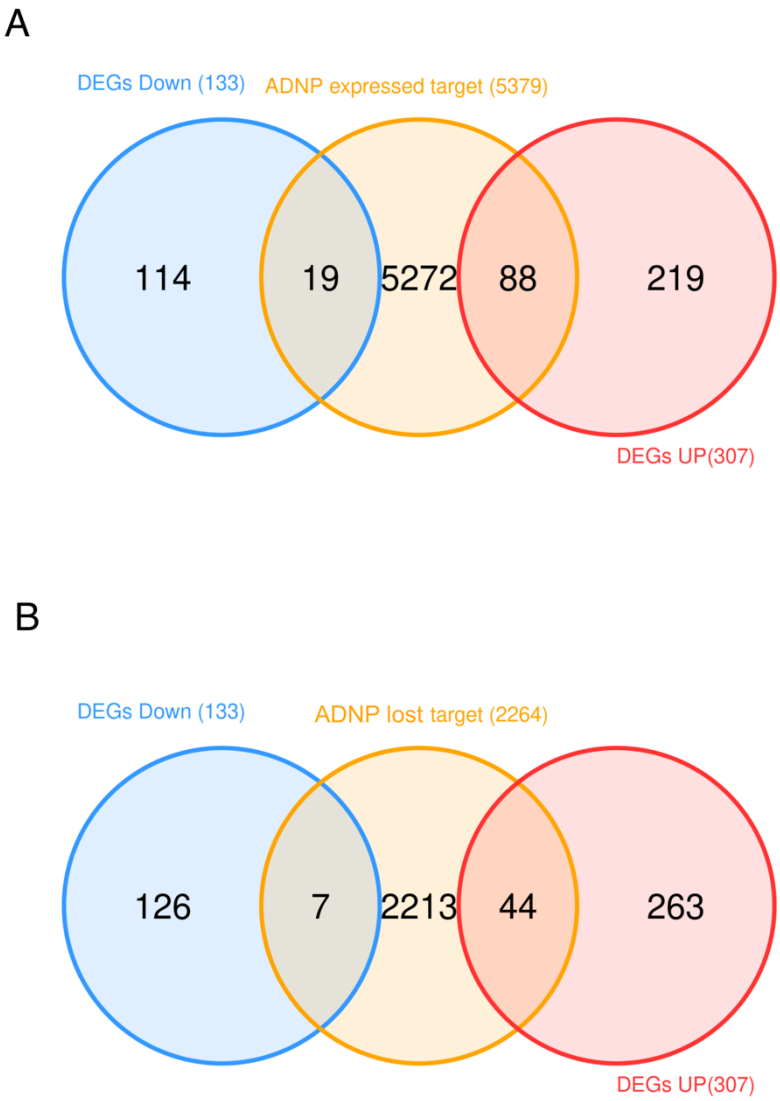



**Figure 35: Overlap of ADNP targets or lost targets with DEGs. (A)** Venn diagram showing the intersection of the ADNP regulatory targets (i.e., promoters and enhancers) with the differentially expressed genes, subdivided into upregulated and downregulated. Hypergeometric test was used to assess the enrichment of ADNP targets among DEGs: p-val = 0.0006. Pearson's chi-squared test was used to assess the enrichment of ADNP targets among the upregulated DEGs versus downregulated DEGs: p-val = 0.004. **(B)** Venn diagram showing the intersection of the genes that lose ADNP at regulatory targets with the differentially expressed genes, subdivided into upregulated and downregulated. Hypergeometric test was used to assess the enrichment of ADNP targets among DEGs: p-val = 0.0034. Pearson's chi-squared test was used to assess the enrichment of ADNP lost targets among the upregulated DEGs versus downregulated DEGs: p-val = 0.01.

To gain a better perception of ADNP targets, we extracted the DNA-binding motifs from CTL08-FLAG ChIP-seq and found that there was no overwhelming consensus motif, but modest enrichment of motifs resembling different transcription factors such as HBP1 and TCF4, both acting as transcriptional modulators and involved in neuronal development, especially via WNT signalling pathway (**Fig. 36A**). Motifs retrieved from ADA03-FLAG did not drastically differ from the ones obtained in the control line. (**Fig. 36B**). To check the ADNP occupancy profile in the light of the motif enrichment, we checked its genomic

features distribution relative to the ones predicted by the DNA-binding motif of HBP1. In percentage, ADNP seems to occupy the promoters more frequently than how much one would expect to observe for HBP1, which could be consequence of the ADNP preferential binding for the Alu elements close to promoters. On the other hand, distal intergenic regions were the most concordant between CTL08-FLAG ChIP-seq and HBP1 motif presence.

A

| Rank | Motif | P-value | log P-pvalue | % of Targets | % of Background | STD(Bg STD) | Best Match/Details |
|---|---|---|---|---|---|---|---|
| 1 |  | 1e-339 | -7.815e+02 | 29.38% | 0.95% | 29.2bp (54.1bp) | PB0132.1_Hbp1_2/Jaspar(0.747) |
| 2 |  | 1e-194 | -4.488e+02 | 25.22% | 1.97% | 28.4bp (56.4bp) | Zac1(Zf)/Neuro2A-Plagl1-ChIP-Seq(GSE75942)/Homer(0.618) |
| 3 |  | 1e-61 | -1.416e+02 | 23.05% | 6.69% | 44.9bp (57.7bp) | VDR/MA0693.2/Jaspar(0.668) |
| 4 |  | 1e-60 | -1.395e+02 | 4.15% | 0.06% | 32.8bp (46.1bp) | ZNF16(Zf)/HEK293-ZNF16.GFP-ChIP-Seq(GSE58341)/Homer(0.775) |
| 5 |  | 1e-51 | -1.195e+02 | 18.99% | 5.33% | 39.3bp (56.1bp) | TCF4(bHLH)/SHSY5Y-TCF4-ChIP-Seq(GSE96915)/Homer(0.806) |
| 6 |  | 1e-51 | -1.176e+02 | 20.18% | 6.04% | 44.1bp (56.9bp) | PB0090.1_Zbtb12_1/Jaspar(0.706) |

B

| Rank | Motif | P-value | log P-pvalue | % of Targets | % of Background | STD(Bg STD) | Best Match/Details |
|---|---|---|---|---|---|---|---|
| 1 |  | 1e-1297 | -2.988e+03 | 13.92% | 0.56% | 36.3bp (53.8bp) | PB0132.1_Hbp1_2/Jaspar(0.752) |
| 2 |  | 1e-625 | -1.441e+03 | 6.34% | 0.21% | 44.5bp (73.3bp) | TEAD3(TEA)/HepG2-TEAD3-ChIP-Seq(Encode)/Homer(0.716) |
| 3 |  | 1e-605 | -1.395e+03 | 11.29% | 1.28% | 36.0bp (54.1bp) | Zac1(Zf)/Neuro2A-Plagl1-ChIP-Seq(GSE75942)/Homer(0.627) |
| 4 |  | 1e-447 | -1.031e+03 | 4.93% | 0.20% | 46.0bp (67.3bp) | Hoxd10(Homeobox)/ChickenMSG-Hoxd10.Flag-ChIP-Seq(GSE86088)/Homer(0.789) |
| 5 |  | 1e-412 | -9.504e+02 | 21.32% | 7.14% | 55.0bp (59.3bp) | VDR/MA0693.2/Jaspar(0.705) |

C

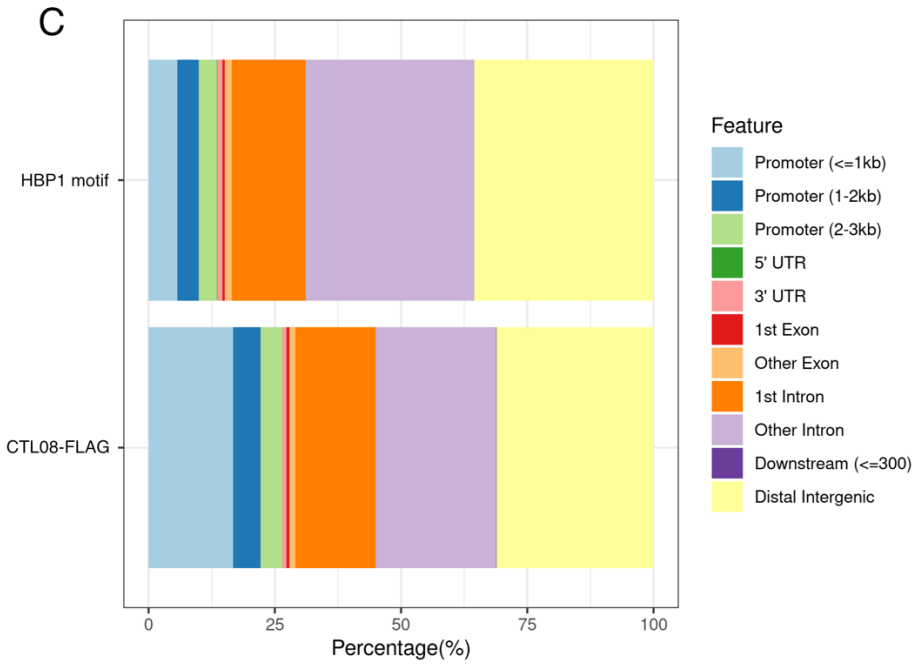



**Figure 36: ADNP DNA-binding consensus motif enrichment. (A)** ADNP binding motifs retrieved from CTL08-FLAG ChIP-seq predicted by HOMER and ranked by p-values (motif enrichment over genomic background). Frequency of occurrence and consensus motifs of the best matched genes are indicated. **(B)** The same analysis has been performed for ADNP binding motifs retrieved from ADA03-FLAG. **(C)** Distribution of CTL08-FLAG peaks relative to the HBP1 DNA-binding motif, split into different genomic features.

### 3.5.1 ADNP targets transposable elements in a CTCF-independent manner

Notably, we could not retrieve the CTCF consensus motif, as reported in the case of ChAHP complex described in mESC[328], probably indicating a different DNA-binding pattern for hADNP. Nevertheless, given the ADNP preferential occupancy at Alu elements and MIR, and taken into account that CTCF DNA-binding motif pool in human encompasses various type of transposable elements, we investigated a putative interplay between ADNP and CTCF. We performed a ChIP-seq of CTCF in three replicates for both CTL08-FLAG and ADA03-FLAG to understand whether *ADNP* mutations could impact on its occupancy, inspired by the effect of *ADNP* knock-out on CTCF in mESC. PCA analysis did not show any distinctions between CTL08-FLAG and ADA03-FLAG replicates, while a neat separation was observed compared to input (**Fig. 37A**). We found ~23,000 significant top CTCF peaks in CTL08-FLAG and ~20,000 significant top CTCF peaks in ADA03-FLAG, thus already excluding drastic differences between the two genotypes (**Fig. 37B**).

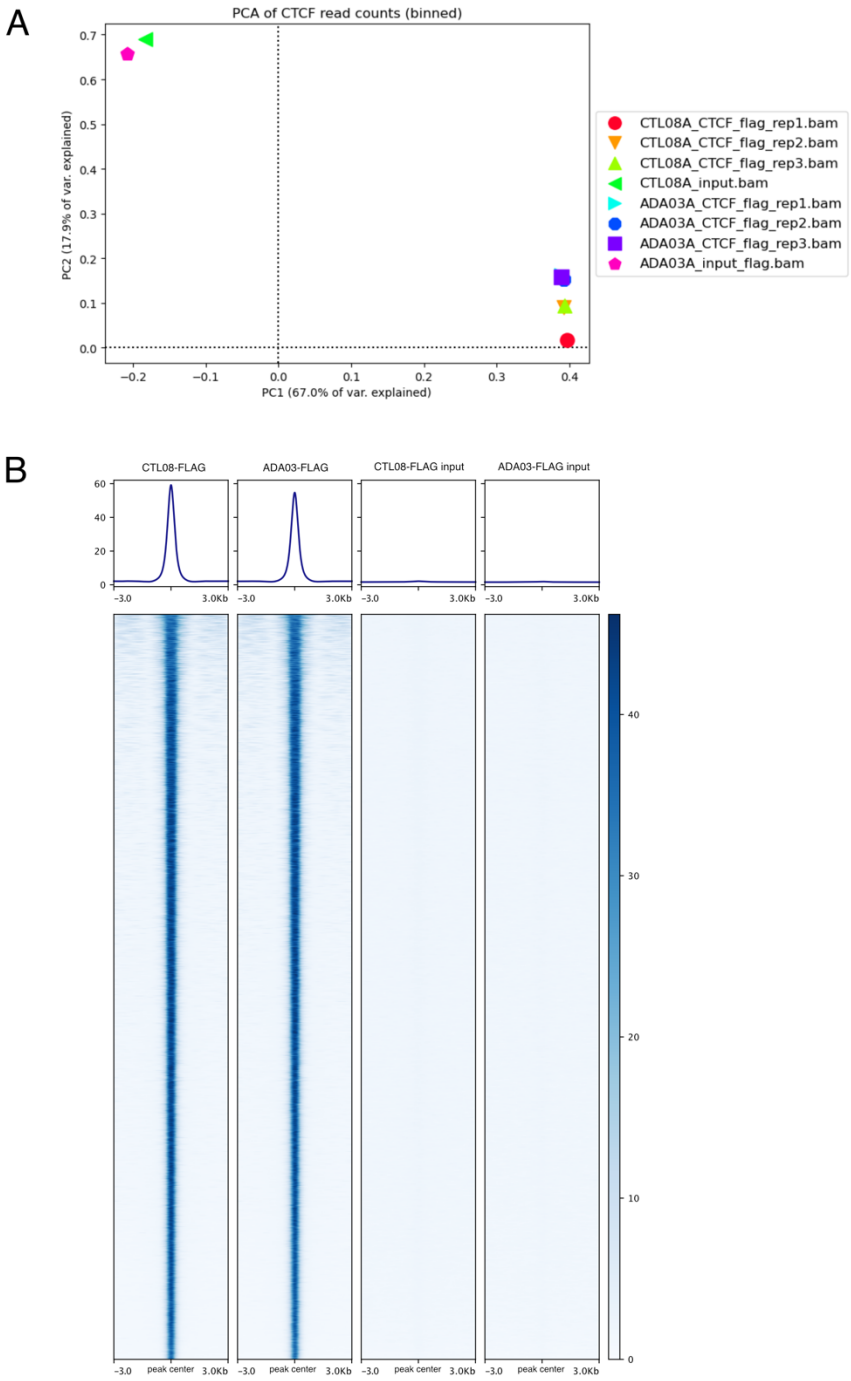



**Figure 37**: **Dimensionality reduction and genome-wide mapping of CTCF ChIP-seq. (A)** Principal component analysis of the CTCF ChIP-seq performed in three replicates for both CTL08-FLAG and ADA03-FLAG. **(B)** Heatmap and corresponding enrichment curve of CTCF ChIP–seq across all significant top peaks in the FLAG-tagged iPSCs. Each row represents a 6-kb window centred on peak midpoints, sorted by the average ADNP ChIP-seq signal intensity of 3 technical replicates. Input signals for the same windows are shown on the right.

To gain insights into the putative interplay between ADNP and CTCF we compared their positioning in both control and patient lines. Although we could not observe any significant difference in the amount of CTCF binding sites between the two genotypes, we did see a redistribution of CTCF targets in the patient line (**Fig. 38A**). Indeed, while the majority of the CTCF peaks was preserved among the control and patient line, we found 2,813 CTL08-

FLAG-exclusive and 1,027 ADA03-FLAG-exclusive peaks. This gain or loss of CTCF peaks in HVDAS iPSCs might indicate a selective rewiring of loop anchors or TAD boundaries that could spatially affect insulation and in turn regulation of gene expression. To further evaluate the interaction between ADNP and CTCF we looked at the shared portion of binding sites between the top peaks of CTCF and ADNP in the FLAG-tagged lines ChIP-seq (**Fig. 38B,C**).

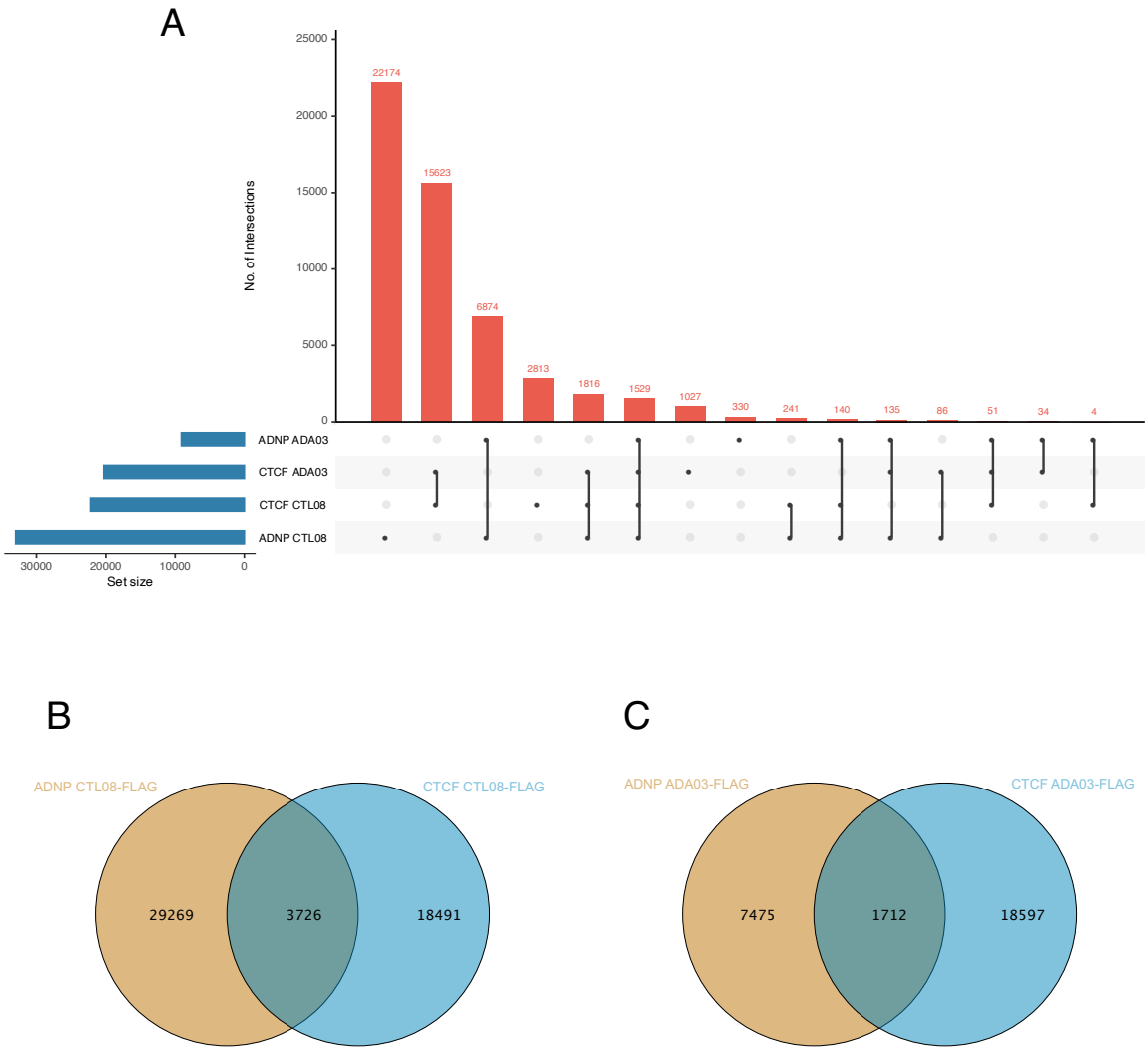



**Figure 38: Overlap of ADNP and CTCF binding sites. (A)** Upset plot representing the number of intersections between ADNP and CTCF peaks obtained from the FLAG-tagged lines. Horizontal bars represent peaks annotated to individual set, vertical bars represent peaks annotated jointly to specified combinations of sets, and dots represent the intersection point among the sets. **(B)** Venn diagram of the overlap of top significant ADNP and CTCF peaks in CTL08-FLAG. **(C)** Venn diagram of the overlap of top significant ADNP and CTCF peaks in ADA03-FLAG.

The extent of this overlap reached 11% in the case of CTL08-FLAG peaks and 18% for ADA03-FLAG peaks. These numbers, consistently with the DNA-binding motif enrichment, show that hADNP binding sites are not as extensively shared with CTCF as in

mESC. We then mapped the ADNP most targeted regions, asking if the competitive model proven in mESC held true in iPSCs. When we looked at the ADNP and CTCF positioning we observed differential binding patterns at the centre of the ADNP-enriched transposable elements, which displayed almost complete absence of CTCF, except for the MIR elements, where a certain extent of shared signal was still retained. Moreover, this positioning pattern does not seem to be modified upon *ADNP* mutations, as we found the same anti-correlation in the presence of the two signals at the ADNP-enriched regions in the HVDAS peaks (**Fig. 39A-D**). Interestingly, LINE1 transposable element category is the only one with the greater variability, exhibiting a higher loss of occupancy in the HVDAS. These results shows that ADNP and CTCF are unlikely to occupy the same pool of binding sites, confirming the non-competitive chromatin binding mechanism of the two factors in human iPSCs.

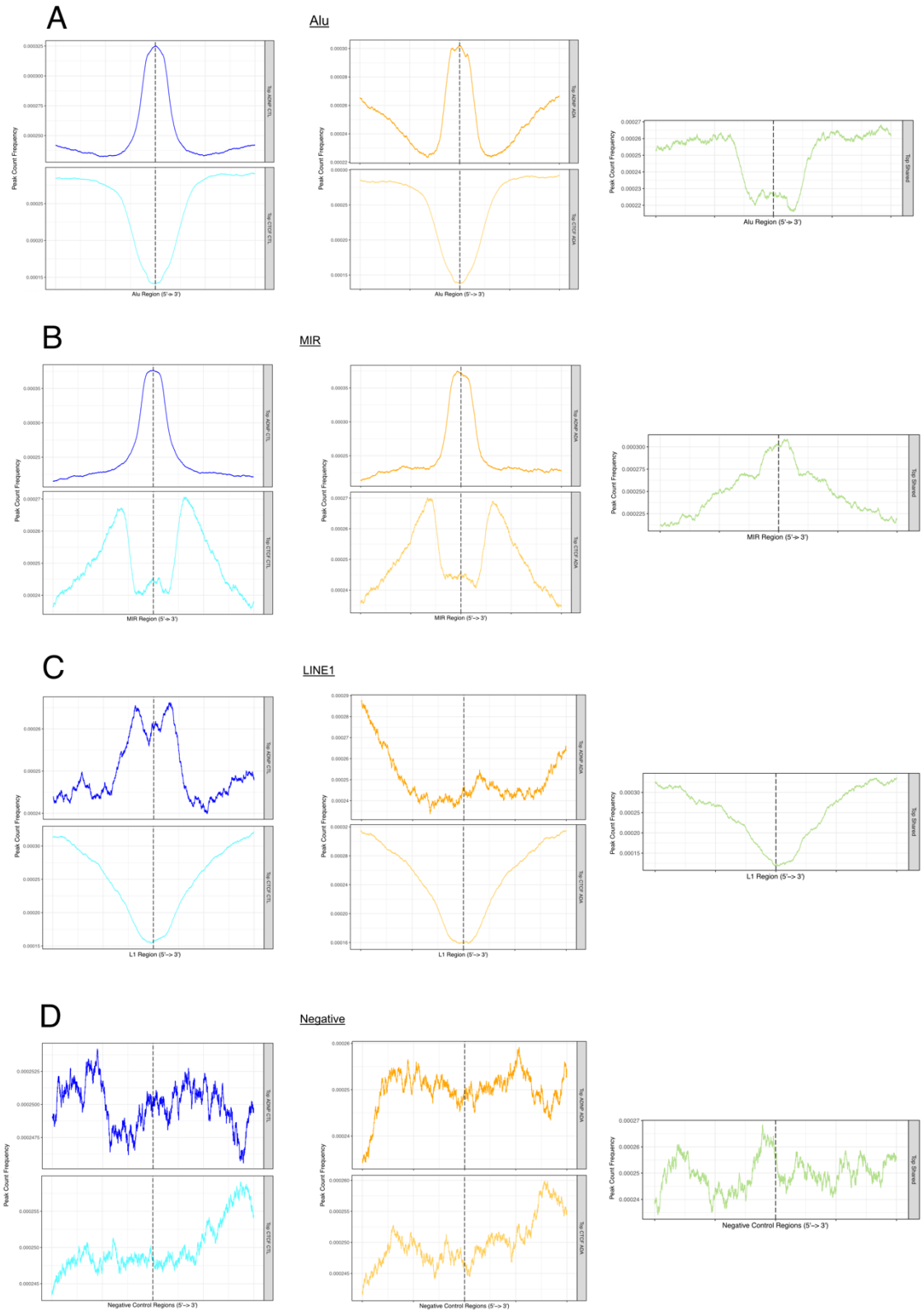



**Figure 39: ADNP and CTCF positioning at ADNP-enriched transposable elements**. **(A-C)** Windows of 4 kb (+2 -2 kb relative to region centre) displaying the ADNP, CTCF and shared top ChIP-seq signals for each ADNP-enriched indicated transposable element. CTL08-FLAG ChIP-seq signal is coloured in shades of blue, ADA03-FLAG ChIP-seq signal is coloured in shades of orange, and shared ChIP-seq signal is coloured in green. **(D)** ChIP-seq signal window as depicted in **A-C** but randomly shifted of 4 kb to serve as negative control.

## 3.6 ADNP mutations elicit a chromatin rearrangement of active enhancer histone marks that sustain transcriptional upregulation

The implications of genetic lesions in the epigenetic players that are affected in neurodevelopmental disorders often result in chromatin rearrangement caused by a disruption in writing, reading, erasing and remodelling functions (**Fig. 8**). To gain insights into the impact that *ADNP* mutations have on chromatin landscape we interrogated active enhancers markers and heterochromatin markers in our iPSCs to understand if a causative link existed with the transcriptional imbalance. We performed ChIP-seq experiments targeting histone post-translational modifications (HPTM) such as H3K27 acetylation (H3K27ac), H3K4 mono-methylation (H3K4me1) and H3K9 tri-methylation (H3K9me3) in all the cohort iPSC lines. First, we checked the quality of our data by performing PCA analysis for each ChIP-seq, visualizing the correct clustering of the samples and their separation from the input (**Fig. 40A**).

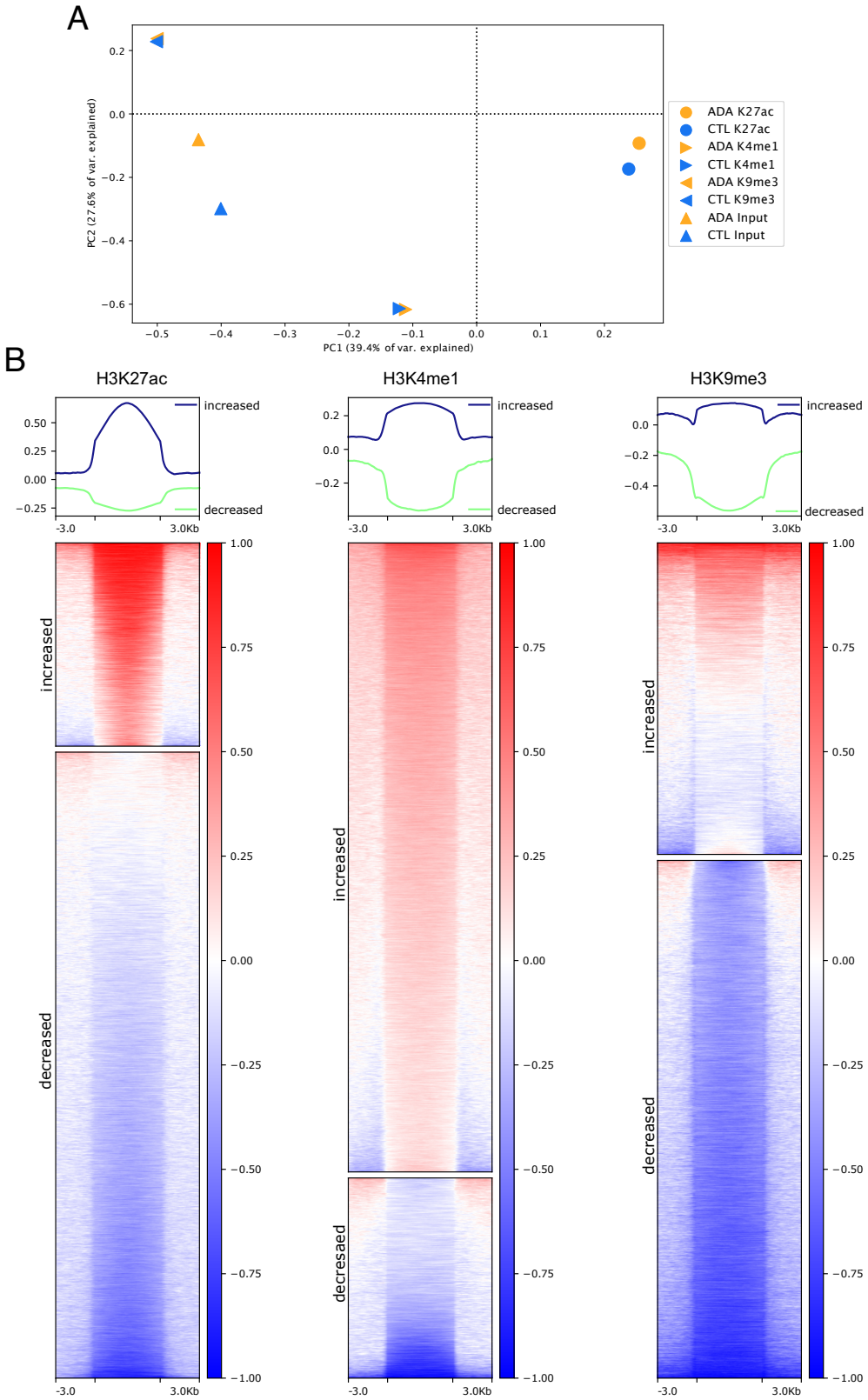



**Figure 40: Dimensionality reduction and genome-wide quantification of H3K27ac, H3K4me1 and H3K9me3 ChIP-seq. (A)** Principal component analysis of the HPTMs ChIP-seq performed in all the cohort lines. **(B)** Heatmap and corresponding enrichment curve quantifying the increased (blue) and decreased (green) HPTMs ChIP–seq signal across all significant top peaks found in all the cohort lines. Each row represents a 6-kb window centred on peak midpoints, sorted by the average ChIP-seq signal intensity of all the biological replicates (n = 13), and subdivided into increased and decreased signal chunks.

We observed a strong increase in all the three HPTMs genome-wide distribution (**Fig. 40B**) and subsequently examined, for each histone modification, the genomic features where the gain was more pronounced and quantitatively compared it between controls and patients (**Fig. 41**).

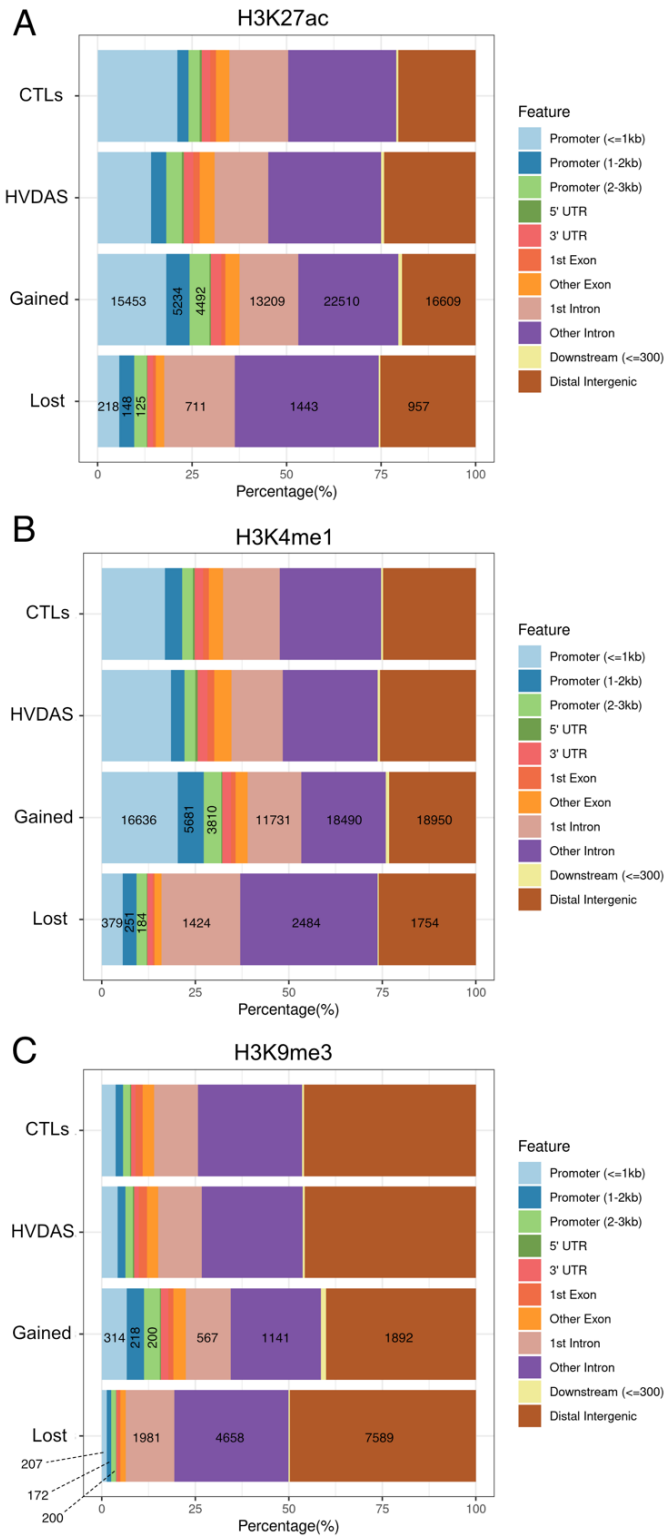



**Figure 41: HPTMs distribution genome-wide. (A-C)** Percentage distribution of the indicated HPTM peaks at different genomic features, divided per genotype (CTL and HVDAS). Number of regions gained and lost in the HVDAS genotype are reported.

We found a prominent gain of H3K27ac and H3K4me1 deposition in HVDAS iPSCs at many different genomic features compared to controls, while for H3K9me3 the gain was less noticeable, while loss was higher. Features falling in the vicinity of promoters (within 3 kb from TSS) were extremely more enriched in both H3K27ac and H3K4me, and so were intronic and distal intergenic regions. When we looked at the abundance of the histone modification peaks relative to the TSS, we found a strong increase especially of H3K27ac and H3K4me1 in both proximal and distal regions (**Fig. 42**).

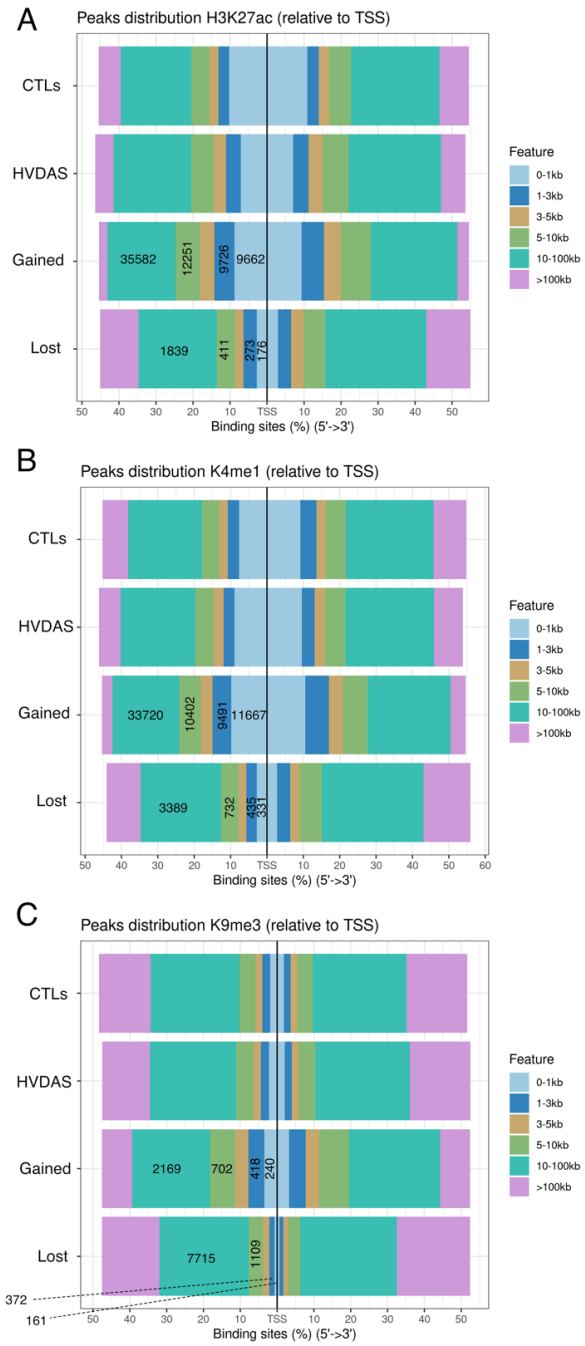



**Figure 42: HPTMs distribution relative-to-TSS. (A-C)** Percentage distribution relative to TSS of the indicated HPTM peaks, divided per genotype (CTL and HVDAS). Number of regions gained and lost in the HVDAS genotype are reported (the number takes into account both 5' and 3' of the feature).

The presence of such markers at distal sites is strongly indicative of cis-regulatory regions, thus suggesting a widespread enhancers activation. In order to accurately identify those regions as enhancers, we confronted our H3K27ac and H3K4me1 gained regions with a list of experimentally validated enhancers coming from a free available repository called 4DNucleome[340]. For each histone modifications we reported the number of genes associated to an enhancer with gained or lost peaks, defining gained as a region with a peak exclusively called in HVDAS lines. We focused on the regions gaining active enhancer marks, and we found 7,691 genes associated with enhancers gaining H3K27ac, 8,535 genes associated with enhancers gaining H3K4me1, and 5,888 genes associated with enhancers gaining both histone modifications. As a negative control, we checked the trend of H3K9me3 ad found that its gain among the enhancer-associated genes was negligible, and in fact similar to the number of regions losing H3K27ac and H3K4me1 at the enhancers (**Fig. 43**).

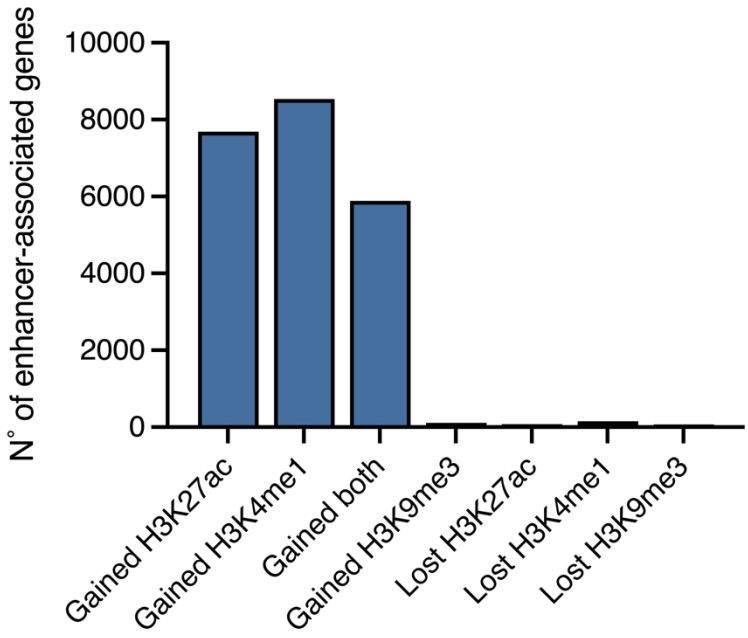



**Figure 43: Number of enhancer-associated genes gaining or losing HPTMs**. Barplot showing the number of enhancer-associated genes with gained or lost H3K27ac, H3K4me1, H3K9me3 or gained both (H3K27 and H3K4me1).

We then asked if such newly activated enhancers could be linked with our iPSC transcriptional phenotype. We therefore looked at the expressed genes in our RNA-seq dataset that had at least one HPTM peak at the enhancers (8,592 for H3K27ac and 8,394 for H3K4me1), so to probe their putative activation in the overlap with DEGs, filtering out those gene whose enhancers could not be associated with any histone modifications. This integrated analysis showed that most of the DEGs with gained H3K27ac and H3K4me1 were associated with upregulated genes, which greatly outnumbered the DEGs losing the same histone modifications (**Fig. 44A,B,C,D**).

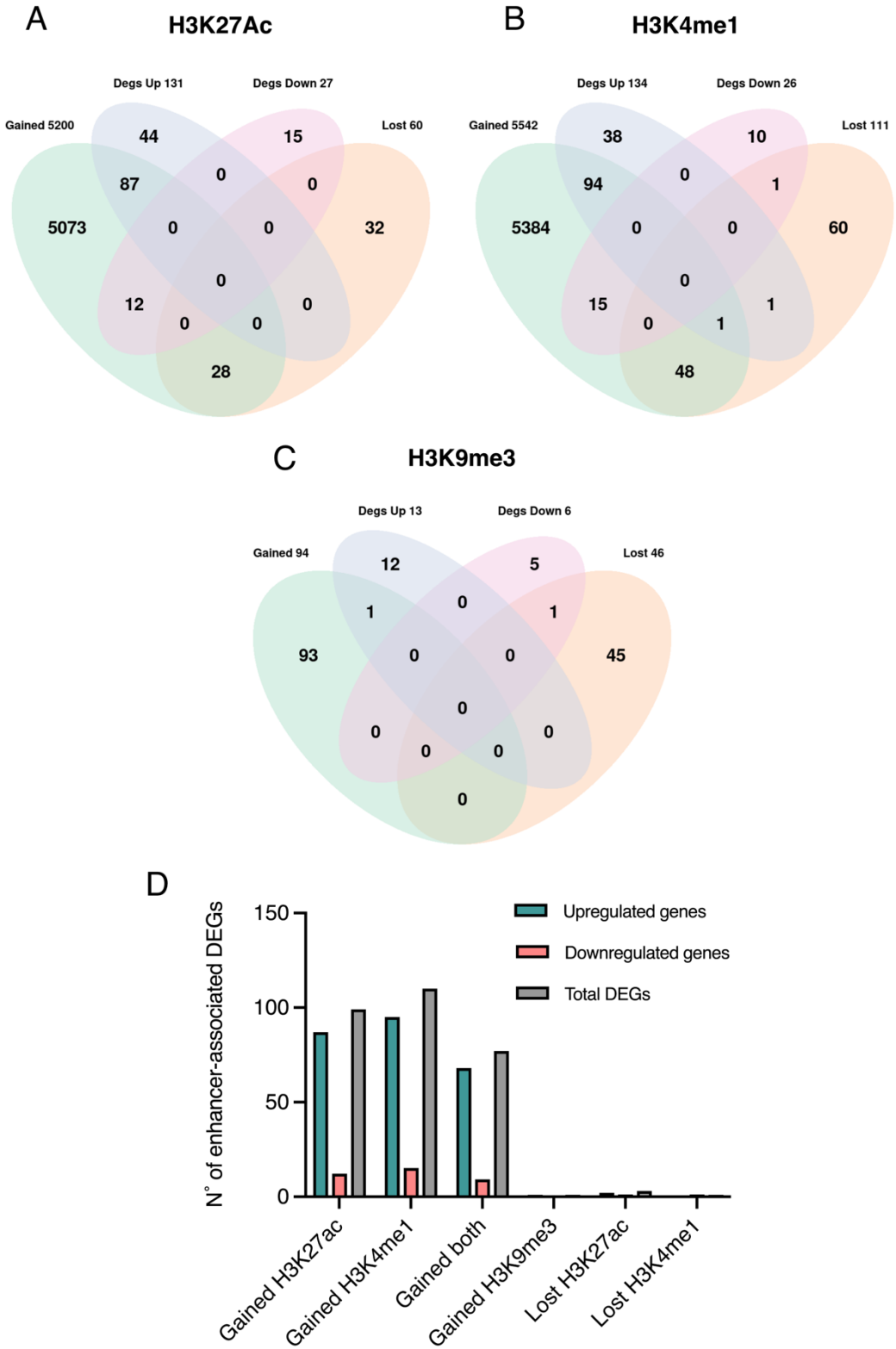



**Figure 44: Integration of differential expression analysis with differential HPTMs abundance at the enhancers. (A-C)** Venn diagrams of the intersection between differentially expressed genes (DEGs Up and DEGs Down) and differential distribution of the indicated HPTM. None of the intersection resulted significant at the hypergeometric tests. **(D)** Barplot showing the number of enhancer-associated DEGs with gained or lost H3K27ac, H3K4me1, H3K9me3 or gained both (H3K27 and H3K4me1). Upregulated, downregulated, and the total number of DEGs are reported according to the colour legend.

However, the overlap of DEGs uniquely enriched for either H3K27ac or H3K4me1 at the enhancers, was not significant. On the other hand, the overlap between genes gaining H3K27ac at enhancers and genes gaining H3K4me1 at enhancers was substantial, indicating that there is a consistent common pool of genes acquiring both markers (**Fig. 45A**). Similarly, we found a solid enrichment of upregulated DEGs whose enhancers gained one or both histone marks (**Fig. 45B,C)**, demonstrating not only that *ADNP* mutations result in an extensive enhancer engagement and modulation, but also providing epigenetic mechanisms supporting the notion that the transcriptional alteration is oriented towards gene activation. Moreover, only the upregulated DEGs gained both active histone marks resulting in significant enrichment over the downregulated ones, whose gain was negligible (**Fig. 45D**).

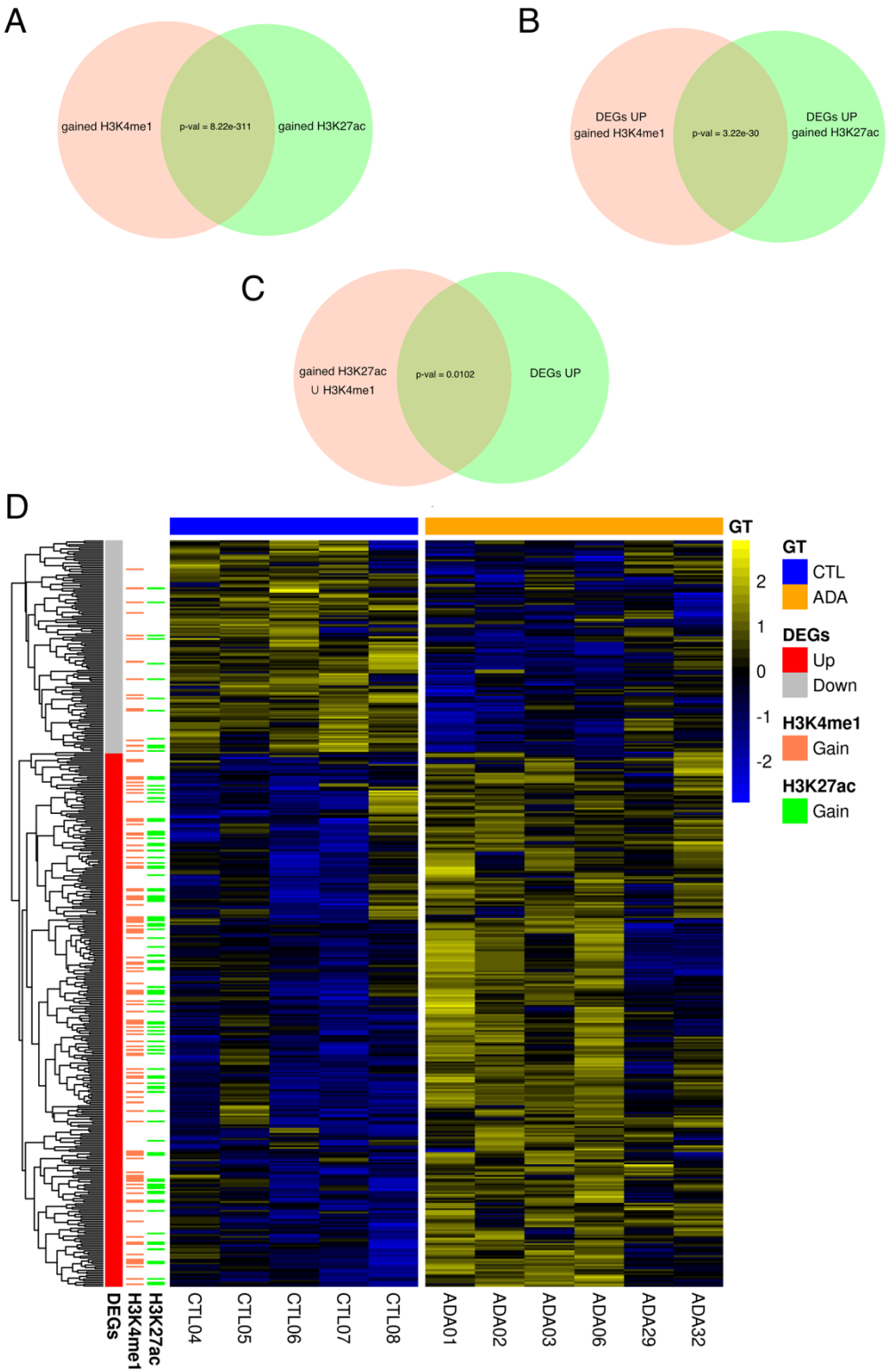



**Figure 45: Concomitant enhancer gain of H3K4me1 and H3K27ac with DEGs upregulation. (A)** Venn diagrams showing the intersection of genes gaining H3K4me1 or H3K27ac. **(B)** Venn diagram showing the intersection of upregulated genes gaining H3K4me1 or H3K27ac. **(C)** Venn diagram showing the intersection of upregulated genes gaining either H3K4me1 or H3K27ac or both HPTMs with upregulated genes. For each intersection the significance of the hypergeometric test is reported. **(D)** Heatmap of the DEGs with the related annotation of differential expression direction (upregulated or downregulated) according to the colour legend. Orange and green side bars indicate a gain of H3K4me1 or H3K27ac, respectively. Statistical analysis of the enrichment of HPTM in upregulated DEGs versus downregulated DEGs was performed with Pearson's chi-squared test. Gained H3K27ac in DEGs UP: p-val = $8.75e^{-05}$. Gained H3K4me1 in DEGs UP: p-val = $1.5e^{-4}$. Gained both in DEGs UP: p-val = $3.31e^{-05}$.

The distribution of gene expression for both HVDAS and control samples integrated with the analysis of the histone marks performed above is further represented by gene sets multiple comparison in **Fig. 46A**. Despite the pervasive gain of HPTMs (which affects gene expression levels also in the control lines), the gain of the active enhancer marks sustains the upregulation especially in the DEGs, which indeed are mostly upregulated (**Fig. 46B**).

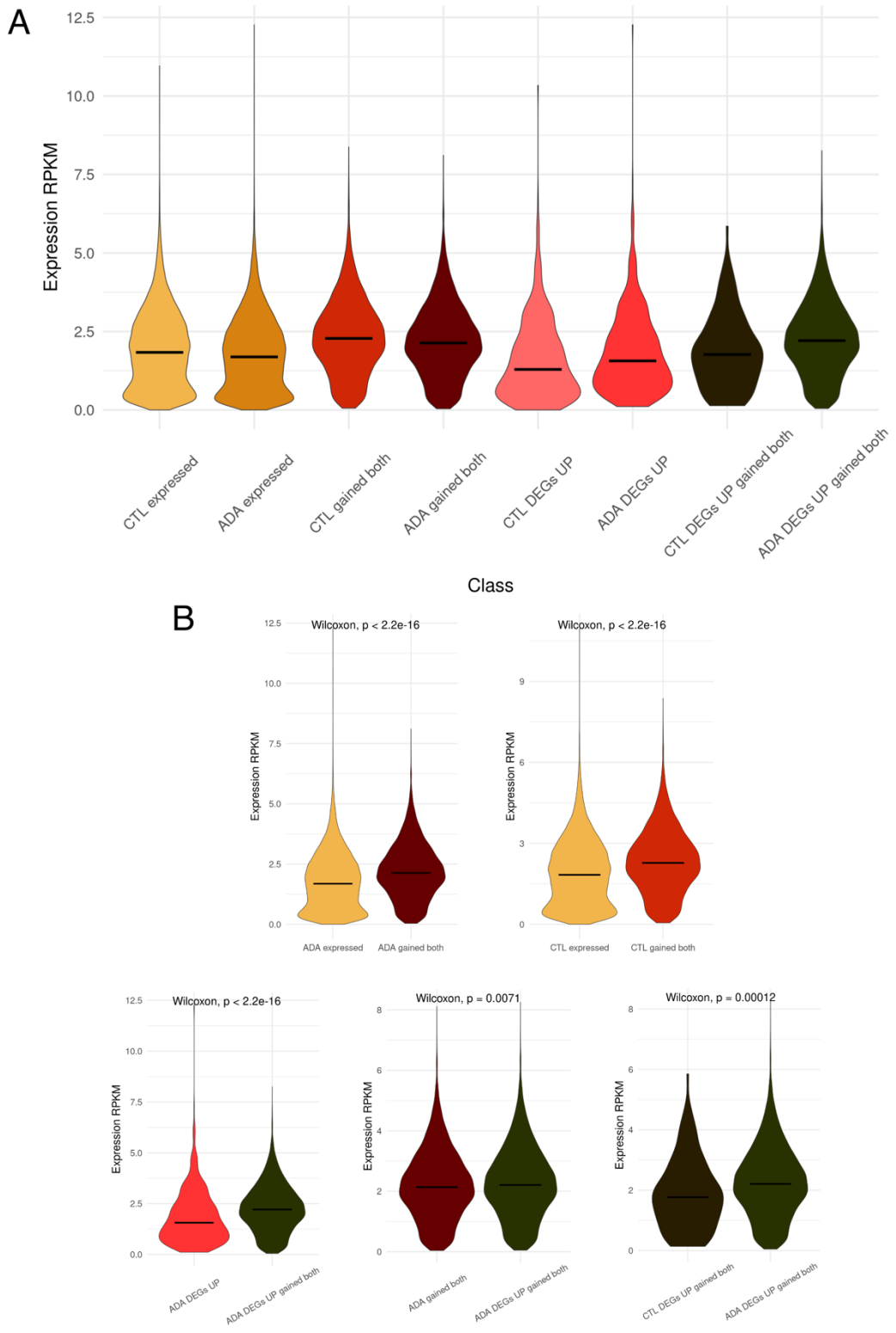



**Figure 46: Expression distribution of genes and DEGs gaining H3K4me1 and H3K27ac and multiple gene sets comparison. (A)** Violin plots showing the distribution of gene sets expression values (RPKM) integrated with active enhancers marks quantification. **(B)** Multiple comparisons between the indicated gene sets were performed using Wilcoxon rank-sum test. Statistical significance is reported above each comparison.

Finally, in the light of the newly found enhancer activation and to further explore the involvement of PRC2 in the transcriptional upregulation, we coupled the analyses performed using the list of DEGs and the enhancers gaining HPTMs, with the EZH2 targets coming

from ENCODE, a publicly available ChIP-seq data repository, focusing on its occupancy at gene regulatory elements (i.e., promoters and enhancers). We found a significant juxtaposition of active enhancer marks and EZH2 targets among upregulated DEGs but not in the downregulated ones, hence reinforcing the idea that chromatin rearrangements engage PRC2 functioning and converge to establish a more active transcriptional program upon *ADNP* mutations (**Fig. 47**).

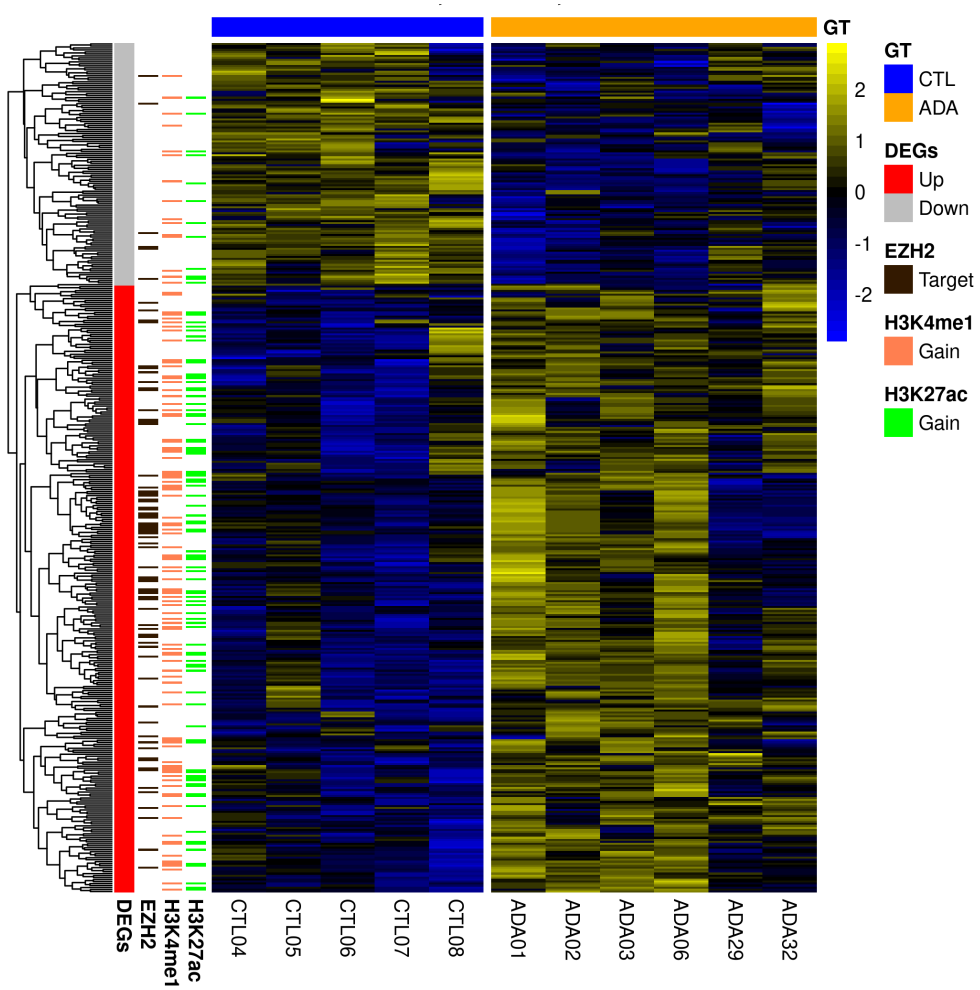



**Figure 47: Integrated analysis of DEGs targeted by EZH2 and HPTMs enhancer gain**. Heatmap of the DEGs with the related annotation of differential expression direction (upregulated or downregulated) according to the colour legend. Orange and green side bars indicate a gain of H3K4me1 or H3K27ac, respectively. Black side bars indicate the genes targeted by EZH2 according to ENCODE ChIP-seq data. Hypergeometric test was applied to evaluate the enrichment of EZH2-targeted DEGs gaining both active enhancer marks and the upregulated genes: p-val = $2.14e^{-6}$. Statistical analysis of the enrichment of EZH2 targets among the upregulated DEGs versus downregulated DEGs was performed with Pearson's chi-squared test: p-val = $6.61e^{-5}$.

Considering the enrichment of EZH2 in the DEGs regulatory regions, we sought to verify how many of these targets were shared with ADNP, testing if it as well might intervene in the active enhancer marks redistribution. We did not see any significant evidence of the overlap between HPTMs gain and ADNP peaks (hypergeometric test: p-val = 0.61); the same resulted from the intersection with the ADNP-lost binding sites (hypergeometric test:

p-val = 0.22). Non-significant result was observed also in the overlap of ADNP and EZH2 targets. Altogether, these evidences decouple ADNP and EZH2 in establishing chromatin active enhancer marks redistribution, highlighting a putative role for PRC2 but not for ADNP.

## 3.7 Functional characterisation of HVDAS 3D models

### 3.7.1 HVDAS cortical spheroids show morphological abnormalities

To have a better understanding of the lineage-specific alterations occurring during neuronal differentiation we generated cortical organoids from the whole cohort of control and patient lines. Considering the results coming from the analysis of iPSCs, we were mostly interested in the early phases of neurodevelopment. To this end, we initially employed the protocol developed in the Pasca laboratory, described in Sloan et al. in 2015, which allows the timely recapitulation of foetal cortex maturation through the generation of laminated cerebral cortex-like structures referred to as cortical spheroids[307]. We first sought to assess general morphological properties of cortical spheroids; therefore, we differentiated controls and HVDAS lines in two independent batches including different iPSC lines and recorded morphological features of individual spheroid at multiple stages of differentiation. We only managed to grow them until day 25, at which point HVDAS spheroids where already looking unhealthy, not homogenously developed and with loose ring structure anticipating partial or full disintegration. However, we managed to collect pictures with ScanR high-content microscopy system in the very early days of differentiation (within the first 2 weeks) before spheroids broke apart. As illustrated in **Fig. 48A and B**, HVDAS cortical spheroids showed reduced global size and more irregular shape compared to controls, suggesting an impaired differentiation occurring in the early stages of neuronal maturation or even during embryoid body formation.

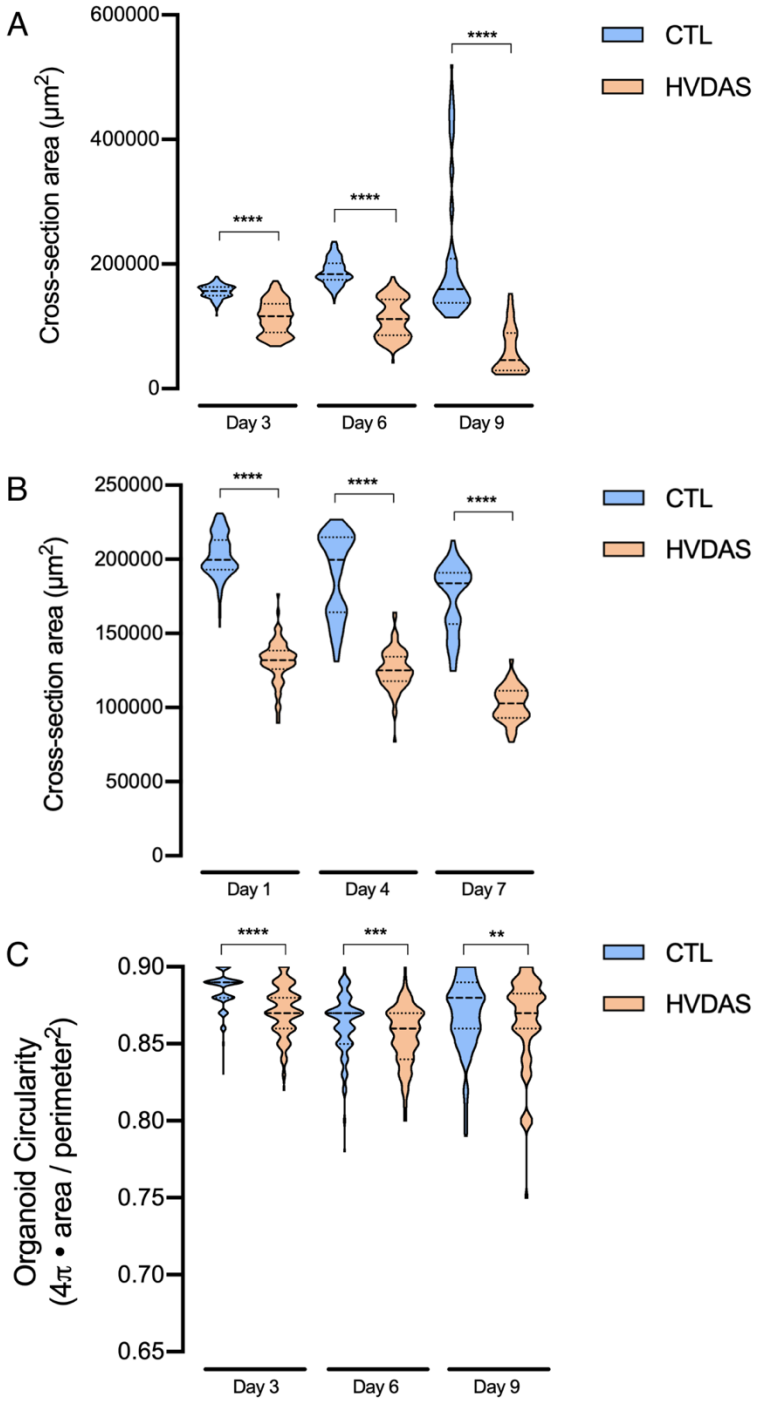



**Figure 48: Morphological analysis of cortical spheroids. (A)** Measurements of the spheroid cross-section area in control (blue) versus HVDAS (orange) lines measured at day 3, 6 and 9 of differentiation. Day 3: CTL n = 85, HVDAS n = 92; Day 6: CTL n = 82, HVDAS n = 79; Day 9: CTL n = 47, HVDAS n = 45; Two-tailed Mann-Whitney **** p-val = 0.0001. **(B)** Measurements of the spheroid cross-section area in control (blue) versus HVDAS (orange) lines measured at day 1, 4 and 7 of differentiation. Day 1: CTL n = 95, HVDAS n = 80; Day 4: CTL n = 73, HVDAS = 75, Day 7: CTL n = 39, HVDAS n = 43; Two-tailed Mann-Whitney **** p-val = 0.0001. **(C)** Cortical spheroid circularity measured at day 3, 6 and 9 in control and HVDAS lines. Number of spheroids (n) is equal to the one reported in **A**. Two-tailed Mann-Whitney; ** p-val = 0.01; *** p-val = 0.001; **** p-val = 0.0001.

### 3.7.2 HVDAS cortical organoids show morphological abnormalities and proliferation impairment

In order to keep investigating 3D models of neurodevelopment in HVDAS conditions, we resorted to a different organoid generation protocol, developed in the Muotri laboratory, described in Trujillo et al. in 2019[308]. This method, similarly to the one developed in Pasca laboratory, produces temporally organized neural populations that are dynamically patterned towards dorsal telencephalon, resulting in 3D structures that recapitulate cerebral cortex development, which are referred to as cortical organoids. We generated two independent batches of such organoids and observed consistent and reproducible morphological abnormalities in HVDAS samples at two different time-point of differentiation. As in the case of the spheroids, HVDAS cortical organoids surviving rate was lower, moreover, they showed reduced global size and increased irregular shape when imaged with transmitted light via bright-field contrast microscopy (**Fig. 48A,B**). Because a thorough morphometrical analysis like the one performed in the spheroids could not be duplicated in this case, we performed immunostaining on organoids approximately after 30 days of differentiation in order to confirm the altered morphology phenotype and to understand where this could derive from. In particular, we were interested in dissecting the proliferation and mitotic activity that could impact on the different morphological features of HVDAS cortical organoids. We therefore collected and clarified HVDAS and control organoids in order to perform immunostaining using DAPI and an antibody directed against the phosphorylation of Ser28 in histone H3 (pH3), hallmark of cells actively undergoing mitosis. With this setup, we obtained 3D pictures comprehensive of the superficial and inner architecture of each organoid enabling the high-resolution quantification of the number of cells actively dividing in a single field-of-view. By measuring the signal in the DAPI channel, this analysis revealed that HVDAS organoids indeed displayed a significant reduced volume compared to controls (**Fig. 48C**). We then asked if reduced size was a consequence of altered proliferation by visualising the signal of pH3. The number of ‘blobs’ were normalized to each organoid spatial dimension, resulting in a density measurement. Following quantification, our results revealed a significant reduction in cells undergoing mitosis in HVDAS organoids when compared to controls (**Fig. 48D**).

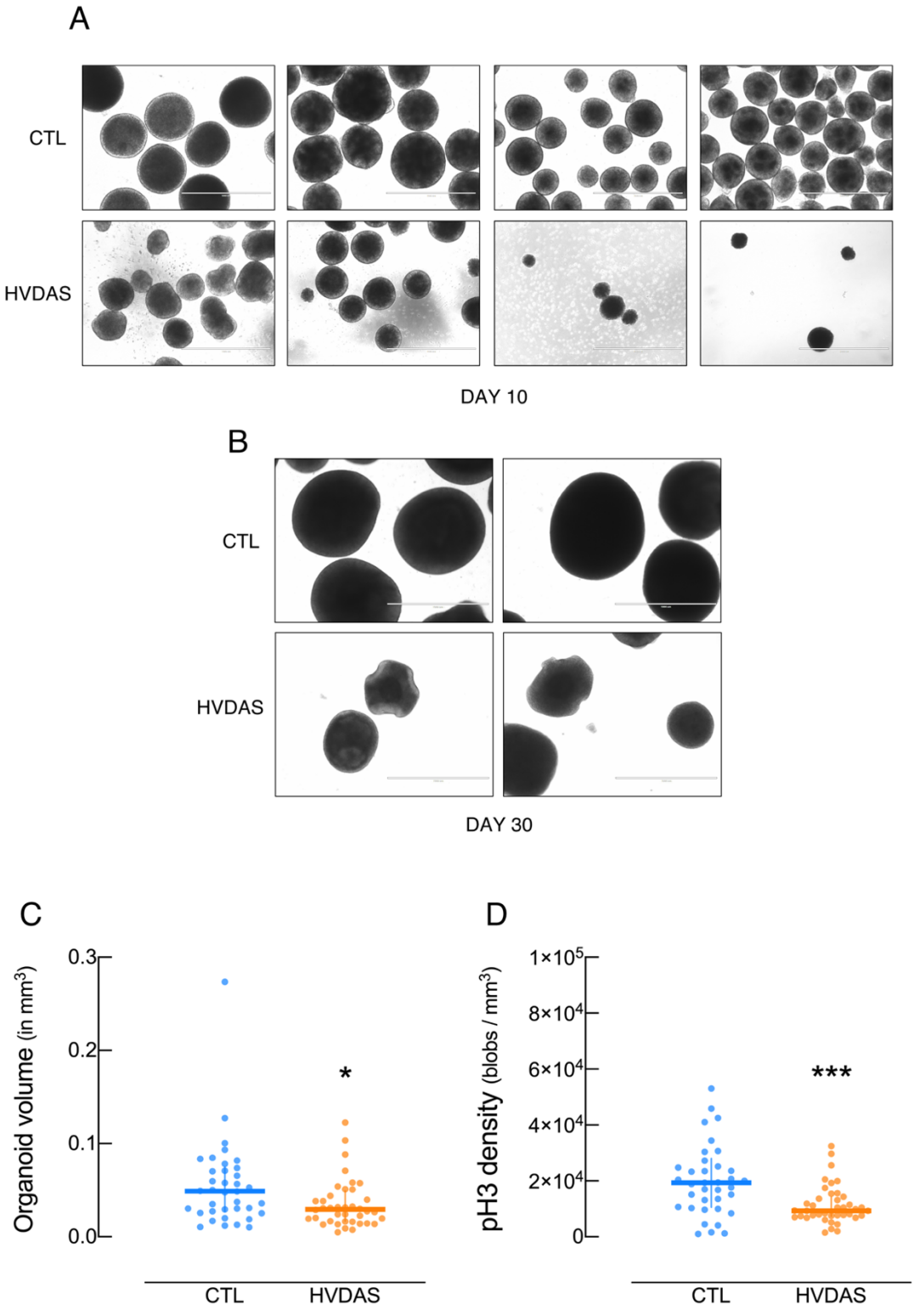



**Figure 49: Morphological analysis of cortical organoids. (A)** Representative pictures taken with EVOS microscope of controls and HVDAS cortical organoids at day 10 of differentiation. Pictures represent different control and HVDAS lines of the same differentiation batch. **(B)** Representative pictures of controls and HVDAS cortical organoids at day 30 of differentiation taken with EVOS-FL microscope. Pictures represent different control and HVDAS lines of a same differentiation batch. **(C)** Volume measurement in control and HVDAS organoids by quantifying the signal in the DAPI channel. HVDAS n = 40, 0.02943 $mm^3$; CTL n = 38, 0.04900 $mm^3$; Two-tailed Mann-Whitney test: * p-val = 0.0201). **(D)** phospho-H3 density measurements normalized for each organoid's size. HVDAS n = 40, 9348 blobs/$mm^3$; CTL n = 38, 19392 blobs/$mm^3$; Two-tailed Mann-Whitney test: *** p-val = 0.0002.

Because the analysis is performed in 3D, both DAPI volume calculation through channel segmentation, and the count of pH3 'blobs', were highly accurate, avoiding repetitive counts of the same cells in different stacks (**Fig. 50**). Altogether, results coming from our 3D models point to a consistent morphological derangement characterised by restrained growth likely caused by reduced mitotic and proliferative activity occurring in early stages of cortical development.

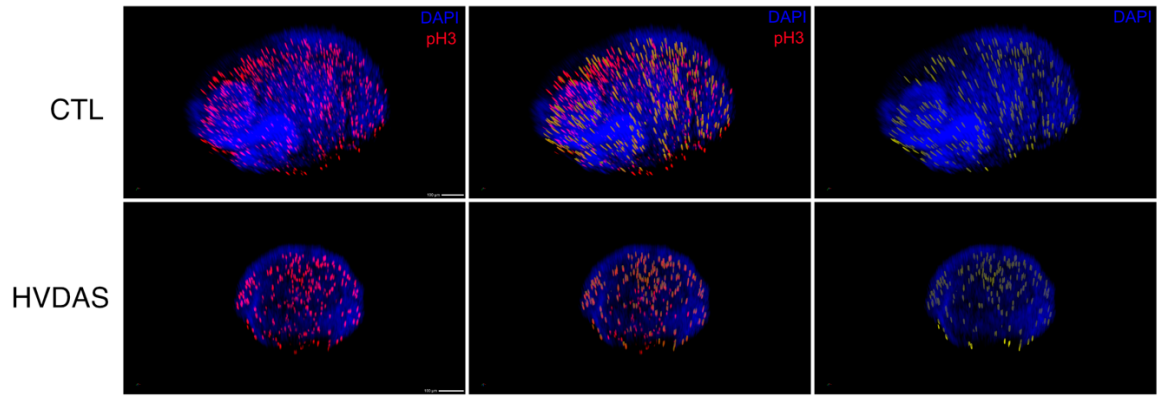



**Figure 50: 3D imaging of cortical organoids.** Representative picture of cleared and immunostained control and HVDAS cortical organoids stained with DAPI and anti-pH3. The cross-section blobs identified by segmenting pH3-positive cells that fall into DAPI signal are shown in yellow for both conditions.

### 3.7.3 HVDAS cortical organoids exhibit increased number of PAX6-positive progenitor cells

To follow-up on the morphological and proliferative aberrations we then investigated the cellular composition of HVDAS and control cortical organoids, by immunostaining the organoids using a panel of antibodies directed against neural progenitor markers. We therefore cleared and imaged organoids looking at the distribution and density of SOX2 and PAX6, both widely used markers of neural stem and progenitor cells, and Nestin, a cytoskeletal protein described as neuroepithelial marker[341,342] (**Fig. 51A**). We quantified the area occupied by each marker normalizing for individual organoid's volume, and discovered a significant increase in the percentage of PAX6 occupied area in HVDAS cortical organoids compared to controls (**Fig. 51B,C**), revealing a higher abundance of the progenitor pool in patients, indicative of maturation imbalance caused by ADNP mutations. However, we did not observe a similar change in the SOX2 measurements, suggesting that despite the prevalent co-localization in the ventricular zone, they might act on different molecular axis involved the proliferation and maturation of precursor cells.

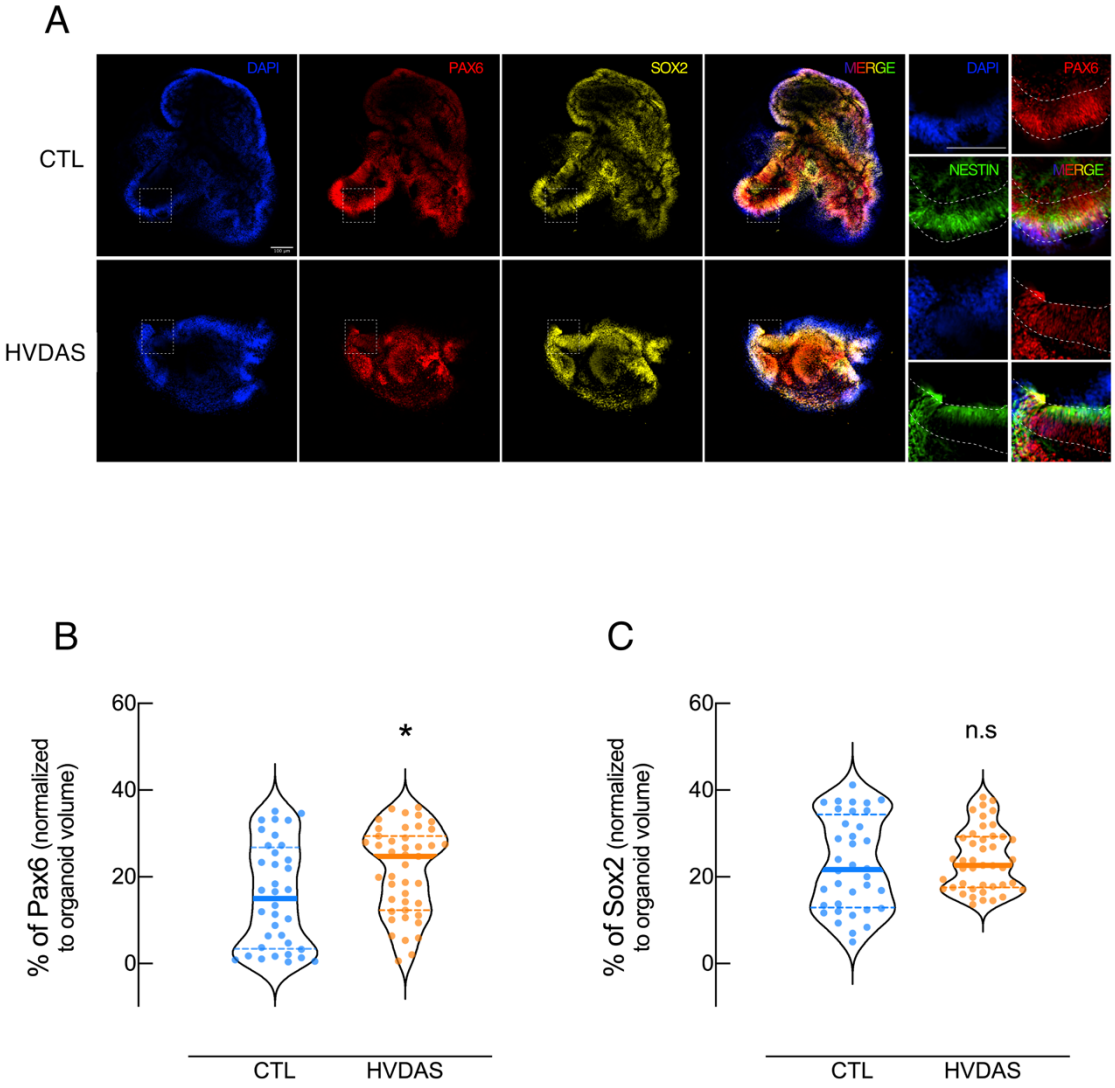



**Figure 51: 3D imaging of cortical spheroids for the characterization and relative abundance of NPCs.** **(A)** Representative picture of cleared and immunostained control and HVDAS cortical organoids stained with DAPI, anti-PAX6, anti-SOX2 and anti-Nestin. The portion of the organoids enclosed in the dashed square is magnified in the quadrants located in the right-most part of the figure. **(B)** Percentage of PAX6-positive cells in control and HVDAS organoids measured by the area of PAX6 signal and normalized to the organoid volume (sum of the area occupied by DAPI in each individual stack, multiplied by the z-step interval used during whole-organoid acquisition, 5 µm). CTL n = 36, 15.03% PAX6-occupied area; HVDAS n = 43, 24.76% PAX6-occupied area; Two-tailed Mann-Whitney test: * p-val = 0.0245. **(C)** Same analysis of **B** was performed. CTL n = 36, 21.68% SOX2-occupied area: HVDAS n = 43, 22.62% SOX2-occupied area; Two-tailed Mann-Whitney test: ns.

## 3.8 Single-cell transcriptomic of HVDAS organoids reveals global accelerated neuronal maturation

To further characterise the pathological relevance of *ADNP* mutations in neuronal differentiation and the cellular diversity of HVDAS organoids, we performed single-cell RNA-seq on 1-month-old cortical organoids. We processed and analysed 7 samples (5

HVDAS and 2 control) for a grand total of 27,885 cells, and demultiplexed each line contribution by clustering the whole dataset based on the genotype. We observed a neat separation between the control and HVDAS cells, indicating a basic discrepancy in transcriptomic composition (**Fig. 52A,B**).

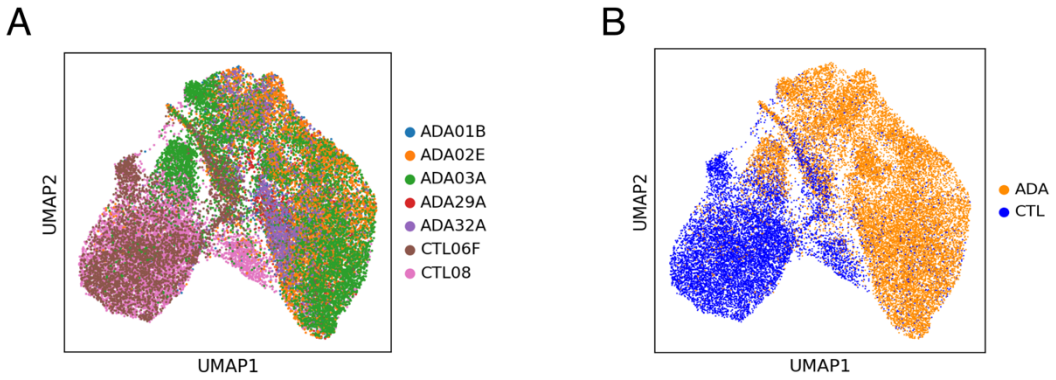



**Figure 52: Dimensionality reduction of scRNA-seq dataset based on genotype. (A)** UMAP of the scRNA-seq dataset highlighting each multiplexed line used in the experiment, as reported in the colour legend. **(B)** UMAP of the scRNA-seq dataset labelled based on the genotype, as reported in the colour legend.

We next used Leiden clustering algorithm to identify group of cells based on their transcriptional affinity, and assign a gene expression signature to each of them (**Fig. 53 A,B)**. We further superimposed an additional clustering based on the expression of marker genes to check for the correct differentiating cell types and their relative abundance; we found populations of cells principally clustering into four major groups: progenitors, intermediate progenitors, glia, and glutamatergic neurons, concordantly with the Muotri Lab reference dataset[308] (**Fig. 53 C,D**) (**Fig. 54A-D**).

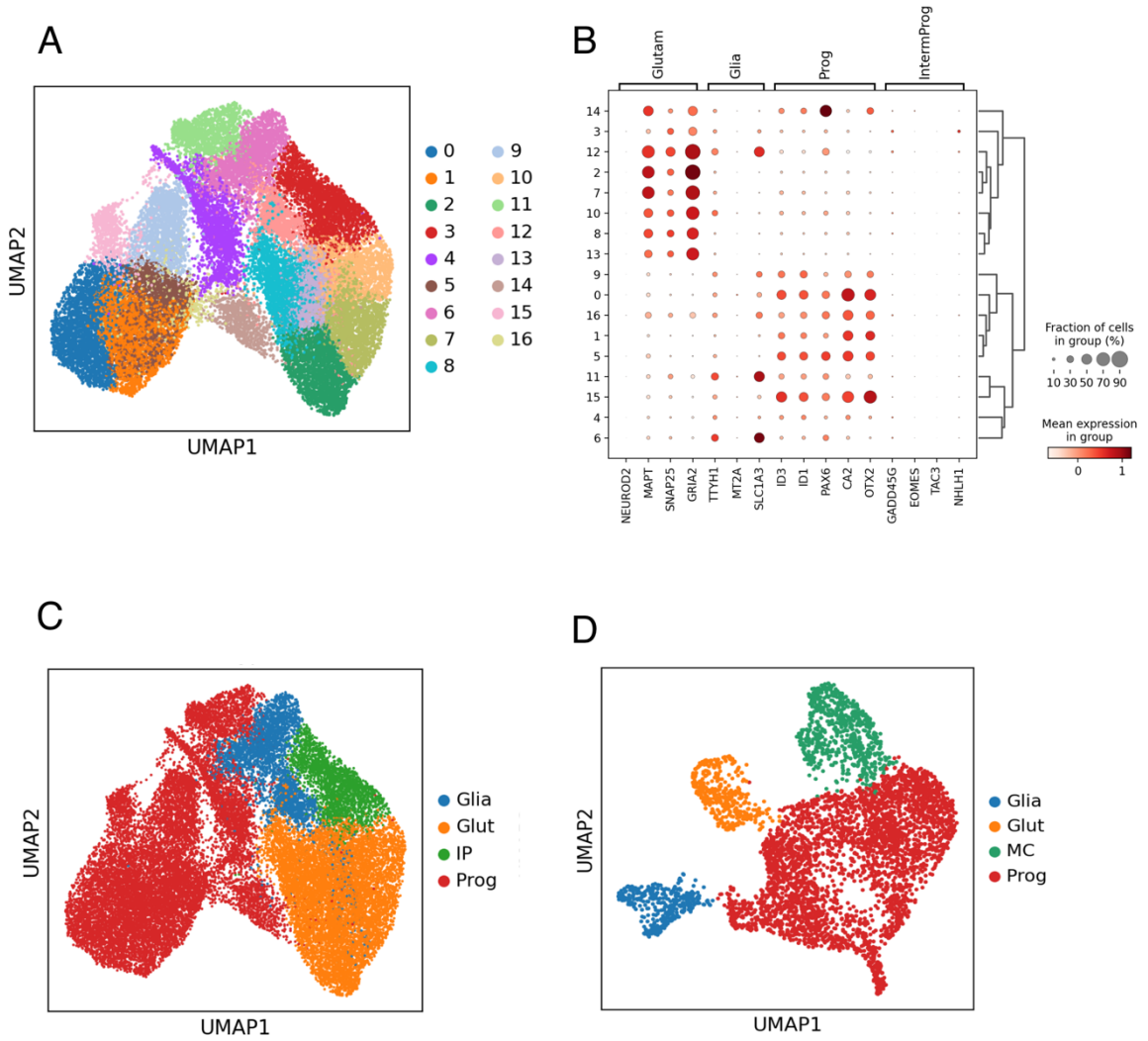



**Figure 53: Leiden clustering algorithm and neuronal clusters identification (A)** UMAP clusters as produced using the Leiden clustering algorithm. **(B)** Dotplot showing the fraction of cells expressing marker genes for each cluster identified with the Leiden algorithm. The mean expression values are reported. **(C)** UMAP of the scRNA-seq dataset representing the four major cell types identified. **(D)** UMAP of the cell-type composition in the 1-month-old organoids grown in the Muotri Lab (see reference in the text).

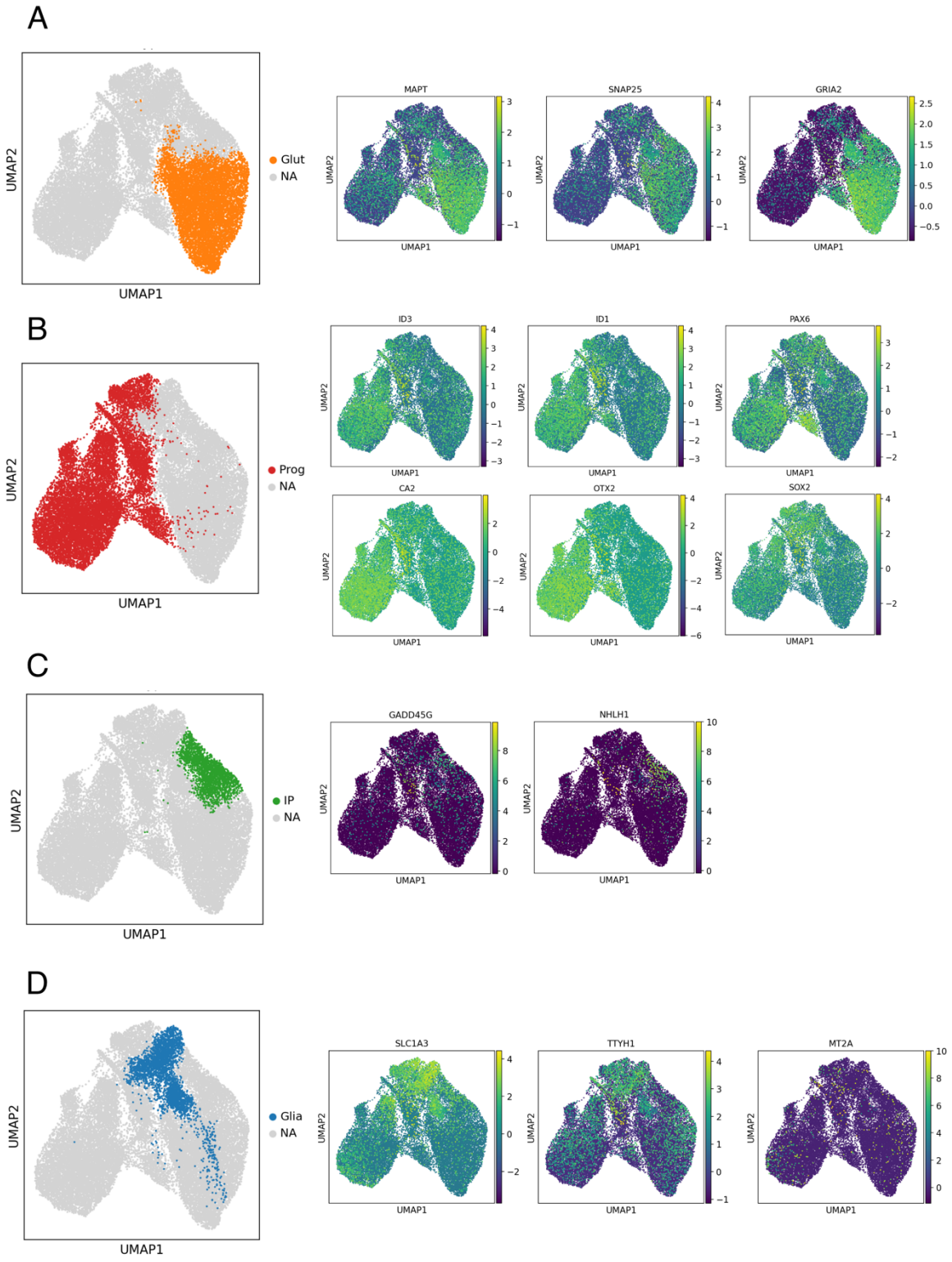



**Figure 54: Distribution of marker genes in the four major cell types identified. (A-D)** UMAPs representing the expression of marker genes distributed in the whole dataset. Group of marker genes which identify individual clusters are reported for glutamatergic neurons **(A),** progenitors **(B),** intermediate progenitors **(C)** and glia **(D)**.

These data demonstrate a marked separation in the cell composition of the HVDAS organoids compared to controls. Indeed, the vast majority of control cells consisted of

progenitors, while a minor portion consisted of glutamatergic neurons, concordantly with the reference stage-related data. Conversely, the HVDAS cells displayed a much higher abundance of cells in a more mature differentiation state, with glutamatergic neurons accounting for the majority of the cell population (**Fig. 55A,B**). In fact, when compared to the Muotri Lab reference dataset, HVDAS-specific cell composition looked more similar to the 3-months-old organoids rather than the 1-month-old ones, which instead remained the closest reference for the control-specific cell composition.

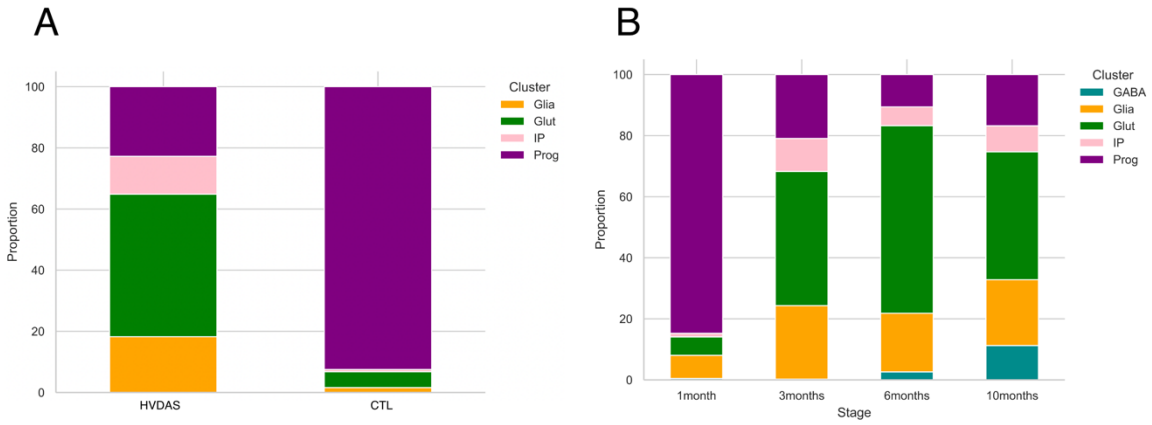



**Figure 55: Cell composition of cortical organoids compared to the Muotri Lab reference. (A)** Barplot of the cell composition in our cortical organoids, with relative abundance of HVDAS-specific and control-specific cell populations. **(B)** Barplot of the cell composition using the reference data from Muotri Lab cortical organoids, divided per stage of maturation.

Moreover, we pinpointed a minor cluster of GABAergic neurons, which was appearing unexpectedly early in the HVDAS organoids, confirming the accelerated maturation phenotype (**Fig. 56A,B**). Interestingly, at the iPSCs stage, HVDAS cells expressed a higher level of *GAD1* compared to controls (**Fig. 56C**).

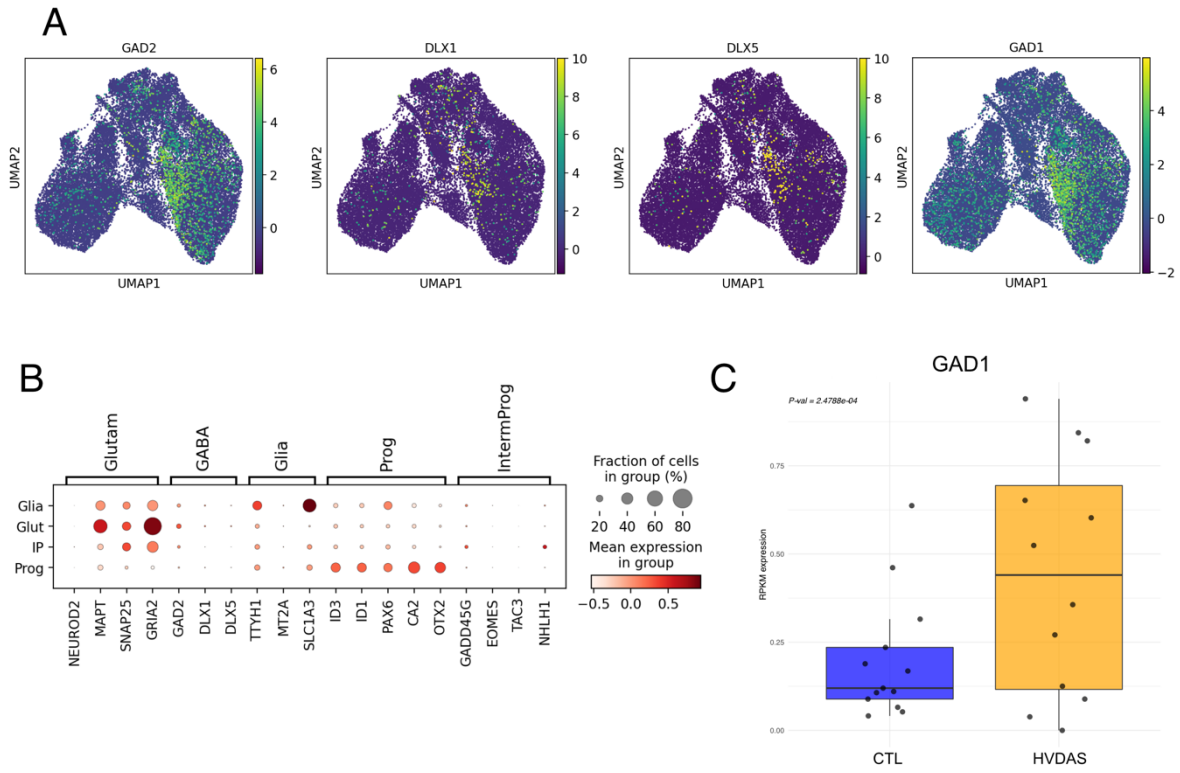



**Figure 56: GABAergic neurons gene signature in HVDAS cortical organoids. (A)** UMAP of the distribution of marker gene expression used to identify GABAerigc neurons. **(B)** Dotplot showing the fraction of cells expressing marker genes for each major neuronal cluster. The mean expression values are reported. **(C)** Expression levels of GAD1 in iPSCs. P-value is reported to indicate differential expression between HVDAS and CTL.

# 4. DISCUSSION

Our goal with this work is the dissection of the molecular pleiotropic effects elicited by *ADNP* mutations on transcriptional and chromatin profile, linking the two aspects under the unifying concept of ADNP as an upstream (co)regulator of neuronal differentiation. We obtained meaningful insights into the molecular function of ADNP by investigating different developmental stages of the largest cohort of HVDAS patients to date. We showed a transcriptional alteration heavily skewed towards upregulation, in line with described molecular function of ADNP. However, A genetic model that accounts for the mutational effect underlying the complex patients' phenotype is currently lacking, as mutations in *ADNP* are not mechanistically framed in a blunt loss-of-function or gain-of-function, nor to haploinsufficiency. In fact, despite the latter being the most recurrent disease mechanism in NDDs[29,49,48], is to be excluded in the case of HVDAS as the subcellular tracking of the truncated ADNP in HEK293T cells argues against an haploinsufficient model[326], but rather advocates for the need of a closer look at the truncated species of the protein, and consequently, at the patient-specific outcomes of ADNP function according to the relative mutation. For example, ADNP mutations falling over or nearby the nuclear localization signal (NLS) can impair ADNP nuclear localization. The most common *de novo* mutation found in HVDAS patients, p.Tyr719* was indeed described to restrict the correct nuclear localization of ADNP[326]. In addition, the very same mutation has been reported to disrupt the PxVxL motif of ADNP and, as a consequence, the interaction with the chromoshadow domain of HP1γ, resulting in the failure of mESC ChAHP complex assembly and its ability to operate chromatin compaction[327]. Interestingly, while we detected mutant allele transcripts in our RNA-seq, we could not see any trace of the truncated protein, at least not when looking at the expected molecular weight in the Western blot, indicating that a more suitable explorative methodology such as mass spectrometry would be required to further address this question. Since in our ChIP-seq experiment the dosage-dependent signal is equalized (only one allele is tagged in CTL08-FLAG, and only the wild-type allele seems to be properly functioning in ADA03-FLAG), the drastic reduction in ChIP-seq signal suggests that the mutant ADNP might disturb the correct localization of the wild-type protein or its interaction with the complex partners in a dominant-negative fashion. Unexpectedly, we observed a rather strong increase in the wild-type protein level of ADA03-FLAG. This might indicate a compensational mechanism that possibly act on the translational level rather than the transcriptional one, as we did not observe a significant increase in the mRNA levels of the wild-type *ADNP* nor the paralog *ADNP2* in the HVDAS lines[343]. The compensation

would then act to raise the quantity of the wild-type protein to counterbalance the loss of the mutant allele. Further experiments in this direction such as ribosome footprint profiling could reveal exciting new mechanisms of translational control upon *ADNP* genetic perturbation. The dominant-negative mutational effect coupled with ADNP genome-wide deficiency could also explain why the transcriptional dysregulation observed in HVDAS iPSCs is prevalently causing an upregulation of the differentially expressed genes. Indeed, almost one-fifth of the DEGs were associated with a binding site of ADNP in their regulatory sequence, meaning that a substantial part of its targets could directly be affected by its malfunctioning. We found that ADNP-bound genes were significantly enriched among the upregulated genes compared to the downregulated ones, and so were the genes losing ADNP binding in the HVDAS iPSCs. These results indicate that the direct targeting of ADNP is likely underlying the derepression and increased expression of at least a pool of its target genes. Of note, among the differentially expressed genes, we found an upregulation of many WNT pathway-related genes, suggesting their involvement in neuronal-specific developmental alteration produced by *ADNP* mutations. Interestingly, a recent work investigating the role of ADNP in neuronal differentiation found that ADNP is required for neural induction by enhancing WNT signalling and that loss of ADNP promotes β-Catenin degradation via ubiquitin-proteasome pathway, resulting in down-regulation of key neuroectoderm developmental genes[344]. These opposite findings might underlie differential functioning of ADNP between mouse and human.

We found EZH2 and SUZ12 as master regulators targeting the majority of the iPSC DEGs, recurring in the transcriptomic analysis of the NCSCs as well. Both factors compose the catalytic subunit of PRC2, one of the most researched multiprotein complexes contributing to basic developmental processes in vertebrates. PRC2 coordinates the expression state of key developmental genes by repressing their regulatory elements through the deposition of H3K27me3 thus maintaining ESC identity; such heterochromatin domains can be further modulated upon differentiation stimuli, triggering mechanisms essential for cell fate transitions and proper establishment of cell identity[174,175]. Recent evidences demonstrated that genes under the control of PRC2 are occupied by a poised form of polymerase II[345,346] and that poised enhancers establish physical interactions with their target genes already in ESCs, in a PRC2-dependent manner[347]. Since most of our DEGs were PRC2 targets, and a major portion of them simultaneously gained active histone marks at their enhancers, this helped us speculating that the genome-wide redistribution of priming/activating enhancer marks observed in HVDAS might impact on the derepression of genes under the control of PRC2-dependent enhancers. Moreover, the parallel increased expression of other epigenetic

modifier such as KDM6 (FDR = 0.01, logFC = 1.38), which demethylates the lysine 27 of histone H3, KDM4B (FDR = 0.0005, logFC = 0.75), which demethylates the lysine 9 of histone H3, and KMT2D (FDR = 0.04, logFC = 0.6), which methylates the lysine 4 of histone H3, are in line with the global upregulated transcriptional profile of HVDAS iPSCs. Furthermore, given the concomitant upregulation of *POLR2A*, it will be important to understand whether poised RNAPII also occupies poised enhancers and its role in the mechanisms that lead to the induction of genes involved in neuroectoderm differentiation. Further experiments to support this model will be required, aimed especially at mapping PRC2 and H3K27me3 in order to better define poised enhancers and promoters, and evaluate the putative switch of repressive chromatin marks with active ones at the gene regulatory regions.

ChIP-seq experiments discovered a widespread localization of ADNP at the transposable elements, especially the Alu repeats, part of the short interspersed nuclear elements (SINE) together with mammalian-wide interspersed repeats (MIR). These classes of retro-transposable elements, and Alu in particular, have been successfully moving along the genome making use of the molecular machinery encoded by LINE, contributing to the genetic diversity of the human genome and expanding the complexity of its regulatory functions[338]. Indeed, Alu elements have been proven to harbour CpG residues and most importantly to host a number of TF binding sites[348], while MIR have been reported to provide insulator- and enhancer-like functions in the human genome[349]. Some of the TF binding sites are specific to certain Alu subfamilies (AluS, AluJ, AluY) indicating the tendency of these sequences to evolve into regulatory elements such as enhancers[350]; this is further demonstrated by the enrichment of active chromatin marks at the Alu and their ability to establish long range interactions with promoters, detected in Hi-C experiments[351]. We highlighted a preferential ADNP binding at evolutionarily young Alu, and a more pronounced enrichment in the vicinity of promoters. We also found a global loss of ADNP chromatin binding in HVDAS iPSCs with a perturbed localization in all the transposable elements. While this is expected in the perspective of a dominant negative mutational mechanism, we unexpectedly found that the absence of ADNP did not result in the predicted chromatin release at the lost binding sites. ATAC-seq experiments showed no marked difference in the chromatin accessibility state at the level of the ADNP targets in HVDAS iPSCs. This could indicate a different mechanism of chromatin remodelling in human compared to mouse, or the dosage-dependent retention of ADNP functions unless completely removed. Curiously, the genome-wide reduction of ADNP occupancy spared those peaks mainly falling in the promoter sequences (1 kb away from TSS), which argues

in favour of a conservation of ADNP binding sites that are probably needed to exert its task. In fact, while human ADNP is still lacking a precise functional characterisation, with this work we contributed to shed light on one of the most intriguing functions described in mESC: its role in the topological architecture of the 3D genome. We know that mESC 3D genome architecture is globally unperturbed in the absence of ADNP, however, local modulations of long- and short-range interactions were observed. The malfunctioning of ChAHP could in fact directly or indirectly change promoter-enhancer interaction frequencies and impact gene expression, especially by locally counteracting CTCF chromatin looping[328]. We highlighted a divergent genome distribution of ADNP and CTCF in human iPSCs compared to mESC, with a markedly reduced pool of common targets. However, we also showed an overlap, albeit limited, between ADNP and CTCF binding sites, and most importantly, a certain extent of modulation in CTCF occupancy upon *ADNP* mutations. Deeper studies will be needed to elucidate whether the gain and loss of CTCF and ADNP peaks in HVDAS iPSCs can affect the gene expression program by rearranging the chromatin conformation. On the other hand, the higher number of hADNP binding sites compared to mADNP already suggests an evolutionary shift along mammal development and possibly an even more relevant function of ADNP in the increased complexity of human genome topology. Indeed, since our ChIP-seq experiments did not retrieve the CTCF DNA-binding consensus motif, ADNP could acts as an anchor point for TADs or enhancer-promoter loops in a CTCF-independent manner, as further corroborated by ADNP positioning at Alu close to promoters and other transposable elements such as MIR[352], where CTCF was mostly absent. Interestingly, LINE-1 are the most variable TE in our dataset, showing a robust decrease of ADNP signal in the HVDAS iPSCs. The increase in somatic retrotransposition that occurs during mammalian neural development involves the activation of transcription of LINE-1 elements that may be caused by a derepression of mobile elements. Moreover, misregulation of retrotransposition correlates with many neurological disorders, including neurodegeneration, Rett syndrome and schizophrenia. Elevated LINE-1 expression together with lower binding affinity of repressive MeCP2 protein and histone H3K9me3 was observed in the cerebellum, suggesting a lessening of epigenetic repression and consequently an increase in chromatin accessibility[353]. Although chromatin compaction state did not change in HVDAS at the level of LINE-1, these pieces of data highlights yet one more difference between human and mouse models, demanding a thorough functional characterization of the enriched transcription factors DNA-binding motif (e.g., HBP1 and TCF4) and the identification of other putative physical interactors of hADNP in order to unveil alternative human-specific ChAHP complexes or functions. Our work then paves the way for additional experiments such as Hi-C, intended at the investigation of the impact that

ADNP exercise on the spatial chromatin organization, and the consequences of a putative genome rewiring occurring upon *ADNP* mutations, that could ultimately affect neuronal differentiation and function.

The increased reliability achieved in the last years in the generation of 3D cerebral and cortical organoids enabled the recapitulation of the dynamic features of neurodevelopment which in turn helped us defining a suitable model to elucidate the cellular characterization of ADNP impairment. We harnessed two different protocols for organoid generation in order to model neurodevelopmental properties of HVDAS patients. Cortical spheroids (produced with Pasca Lab protocol) showed reduced global size and irregular shape, indicating that ADNP is necessary for establishing proper development of dorsal telencephalon in the early stages of neuronal differentiation. These results were further confirmed in cortical organoids generated with Muotri Lab protocol. Not only the morphological alteration was replicated in 1-month-old HVDAS organoids, but we also observed a reduction in mitotic activity (scored as decreased number of pH3-positive cells) and an increase in PAX6-positive progenitor cells at the same differentiation stage. These results show that the reduced organoid size in our 3D models could be directly linked with the impaired proliferation rate of NPCs; moreover, the higher abundance of NPCs is suggestive of a delayed differentiation phenotype occurring in HVDAS organoids, probably established in the early phases of neurodevelopment. Is worth mentioning that SOX2-positive cells did not differ from the control to HVDAS organoids, suggesting that despite the abundant co-localization with PAX6 in the ventricular zone, they might act on different molecular axes impacting the maturation of precursor cells. Altogether, these evidences are in line with a recently published work identifying convergent mechanisms of neuronal differentiation upon downregulation of ASD high-confidence risk genes[45]. In particular, they found that downregulation of *ADNP* in a 2D human cellular model of neuronal differentiation produces a phenotype of delayed maturation that is more evident in the early stages of neurodevelopment. Moreover, this cellular phenotype correlated with higher incidence of microcephaly[354,355], which is mechanistically consistent with the NPCs proliferation defects and also sustained by another recent work that reported the reduced telencephalon size in a *X.tropicalis* model of CRISPR-induced *ADNP* loss-of-function[46], accompanied by an increased ratio of NPCs to neurons. Our morpho-functional analysis of 3D cortical organoids, reminiscent of the above-mentioned phenotypes, suggests a common outcome deriving either from haploinsufficiency or putative dominant-negative mutational models involving *ADNP*.

On the other hand, the results emerging from single-cell RNA-seq provide an apparently opposite scenario; we observed a more mature state of cell differentiation in HVDAS samples compared to controls. These results are consistent with recently published work performing a systematic reconstruction of transcriptional and morphological neuronal growth trajectories which revealed a significant developmental acceleration in differentiating neurons in ASD subjects compared to control neurons[319]. These data highlight the presence of aberrant maturational dynamics displayed as an higher ratio of DCX- over SOX2-positive cells in 5-week-old cerebral organoids in ASD cells, indicative of more abundant early maturing neurons when compared to the trajectories of controls; this is further confirmed through the morphological assessment of the aberrantly complex branching patterns and increased thickness of cortical plate in ASD neurons. As additional evidence of such accelerated maturational phenotype, we pinpointed a minor cluster of cells expressing GABAergic-specific markers, which appeared noticeably soon (1 month) compared to the reference transcriptomic atlas provided in Muotri Lab (6 months). Interestingly, one of the main GABAergic-specific gene was particularly upregulated already at iPSCs stage, indicating once again a possible priming of HVDAS transcriptional program that is further exacerbated upon differentiation cues. While this seems to be discordant with the higher presence of PAX6-positive cells from the immunostaining analysis, it still remains a mechanistically valid cause for the decreased organoids' size. The lower proliferation rate observed in terms of pH3-positive cells could indeed be connected with the accelerated differentiation phenotype, which concordantly manifest as decreased mitotic activity in the diminished pool of NPCs, and a simultaneous enrichment of early post-mitotic cells. Further immunostainings aimed at the detection of early-maturing neuronal markers will be needed to assess the prevalence of such cell type in the HVDAS samples compared to controls. Similarly, giving the sparseness of transcriptomic results in single-cell, a deeper coverage scRNA-seq could help to emphasize the differential expression of precursors- vs mature neurons-specific genes.

Because we managed to grow cortical organoids using the Muotri Lab protocol more successfully than we did with the Pasca Lab cortical spheroids, we can speculate on the increased vulnerability shown by the latter. The growth failure exhibited especially by the HVDAS spheroids revealed blunt phenotypic manifestations once neuronal induction began, resulting in aberrant size and integrity of the 3D spheres especially after day 6, when FGF and EGF start inducing a more extensive neuronal precursors proliferation and maturation. The morphological catastrophe occurring in the first 25 days of both batches of differentiation could be a consequence of the higher sensitivity of HVDAS spheroids to the

small molecules inducing telencephalon differentiation. Among these, the concentration of dorsomorphin supplied in the media during the first days of differentiation might play a role, as it is the only discriminating factor between Pasca and Muotri protocol; however, there are no evidences of altered neuronal induction based on different concentration of dorsomorphin for these protocols.

# 5. MATERIALS AND METHODS

## 5.1 Experimental Procedures

### 5.1.1 Cell reprogramming

Skin fibroblasts from HVDAS patients were received from Prof. Frank Kooy (Antwerpen Universiteit, Belgium) and Prof. Even Eichler (University of Washington). Reprogramming was performed using non integrating self-replicating mRNAs as previously described[128] (Stemgent, 00-0071). All control lines were reprogrammed starting from healthy donor biopsies; CTL08A was the only line purchased from the Wellcome Trust Sanger Institute and was the only reprogrammed using the Sendai virus (CytoTune-iPS 2.0 Sendai Reprogramming Kit; Thermo Fisher Scientific, A16517). The details regarding mutant hiPSC used in this work are summarised in **Table 1**.

| Sample Names | Sex | Biopsy | N° iPSC clones | Protein Mutation | Genomic Mutation | cDNA Mutation |
|---|---|---|---|---|---|---|
| ADA1 | M | Fibroblast | 3 | p.Asn832Lysfs*80 | g.49508752_49508755delTTTA | c.2496_2499delTAAA |
| ADA2 | F | Fibroblast | 3 | p.Ser404* | g.49510040G>T | c.1211C>A |
| ADA3 | F | Fibroblast | 3 | p.Tyr719* | g.49509094_49509095insT | c.2156_2157insA |
| ADA6 | M | Fibroblast | 3 | p.Lys831Ilefs*81 | g.49508757_49508760delTTAA | c.2491_2494delTTAA |
| ADA29 | F | Fibroblast | 1 | p.Ser711Lysfs*24 | g.49509122dupG | c.2129dupC |
| ADA32 | M | Fibroblast | 1 | p.Asn832Lysfs*80 | g.49508752_49508755delTTTA | c.2496_2499delTAAA |

**Table 1:** List of HVDAS patient-derived iPSCs cohort with related information

### 5.1.2 iPSC maintenance

hiPSC were cultured in TeSR-E8 medium (Stemcell technologies, 05990) supplemented with penicillin-streptomycin (P/S, 100 U/mL; Thermo Fisher Scientific, 15140-122), with daily media change, at 37 °C, 5 % CO2 and 3 % O2 in standard incubators. hiPSC were grow on matrigel-coated dishes prepared as follows: matrigel stock solution (Corning, 354248) was diluted 1:40 in DMEM /F- 12 1:1 medium (Lonza, BE12-614F and Thermo Fisher Scientific, 11765054, respectively) supplemented with P/S 100 U/mL and used to coat dishes for 15-30 minutes at 37 °C. Passages 1:6 and 1:8 were performed using ReLeSR (Stemcell technologies, 05872) or Accutase solution (Sigma- Aldrich, A6964). ReLeSR was used to detach hiPSC in clumps for expansion and standard maintenance, while Accutase solution was used for single cell passaging; in this case, ROCK inhibitor 5µM (Sigma, Y0503) was added to the culture overnight to enhance single hiPSC survival. Cryopreservation of hiPSC was performed by single cell dissociation and storage in complete TeSR- E8 medium plus 10% DMSO supplemented with ROCK inhibitor 5µM.

### 5.1.3 NCSC differentiation

hiPSC were differentiated into NCSC as previously described[333]. NCSC differentiation required 15- 20 days and was carried out as follows: 90% confluent hiPSC were detached with Accutase solution and plated on matrigel coated dishes in TeSR-E8 medium supplemented with 5μM ROCK inhibitor at a density of ~9.2 × 104 cells per $cm^2$. The day after, NCSC differentiation medium was added and changed every day for 15- 20 days. NCSC medium was composed of DMEM-F-12 1:1, 10% probumin Life Science Grade from the 20% stock solution (20% m/v in DMEM F-12 1:1, stock solution; Millipore, 821001), P/S 100 U/mL, 2 mM L-Glutamine, 1% NEAA, 0.1% 1000X trace elements complex (CA055-010, Gentaur Italy Srl), 0.2% 50 mM b-mercaptoethanol, 10 μg/ml Transferrin, bovine (Holo form) (Life Technologies, 11107-018), 50 μg/ml (+)-Sodium L-ascorbate (Sigma, A4034), 10 ng/ml Heregulin-1 (Peprotech, 100-03), 200 ng/ml LONGÒR3 IGF-I (Sigma, 85580C), 8 ng/ml FGF2, 3 μM GSK3 inhibitor IX (BIO) (MedChem express, HY-10580) and 20 μm SB431542 (MedChem express, HY-10431). Cells were passaged every 4-5 days and plated at high concentration (1:1 the first time and 1:2 the following ones) on matrigel coated dishes for the entire duration of the differentiation. Upon differentiation, NCSC were stocked as stable lines and cultured in the medium and splitting ratios. FACS following staining for HNK1 (Sigma, c6680) and NGFR (Advanced Targeting System, AB-N07) was performed to assess NCSC identity.

### 5.1.4 RNA extraction and library preparation for RNA-seq

Total RNA was extracted from fresh pellets of iPSC or NCS using the RNeasy Mini Kit (Qiagen, 74104). Purified RNA was quantified using a NanoDrop spectrophotometer and RNA quality was checked with an Agilent 2100 Bioanalyzer using the RNA nano kit (Agilent, 5067-1512). Library preparation for RNA sequencing was performed according to TruSeq Total RNA sample preparation protocol (Illumina, RS-122-2202), starting from 250 ng - 1 μg of total RNA. cDNA library quality was assessed on Agilent 2100 Bioanalyzer, using the high sensitivity DNA kit (Agilent 5067-4626). Libraries were sequenced with the Illumina Novaseq 6000 machine at a read length of 50 bp paired-end and a coverage of 35 million reads per sample.

### 5.1.5 Western Blot

Cells were grown to confluency on 10-well plates, collected in PBS, pelleted by 2 min centrifugation at 400 g, and pellets were then resuspended in 100 μl RIPA buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1% Triton X-100, 0.5 mM EDTA, and 5% glycerol)

supplemented with protease inhibitor cocktail (PIC), 1 mM PMSF, and 1 mM dithiothreitol (DTT). Proteins were extracted for 30 min on ice, the lysates were centrifuged at 16,000 g for 20 min at 4 °C, and the protein concentration in the supernatant was determined using the Biorad protein assay (Biorad, #5000006). For western blotting, 35 µg of protein were resolved on polyacrylamide gels, which were transferred on nitrocellulose membrane (Cytiva, 10600124), blocked for 30 min in 2.5% non-fat dry milk in TBS plus 0.05% Tween 20 (TBST), and stained with primary antibodies at 4 °C O/N. The primary antibodies used for western blotting were mouse anti-Flag (1:500, Sigma-Aldrich clone M2, F1804) and rabbit anti-tubulin (1:5,000, Abcam). Signal was detected with corresponding horseradish peroxidase (HRP)-conjugated secondary antibodies and imaged with ChemiDoc XRS+ System (Biorad).

### 5.1.6 ChIP-seq

ChIP-seq for histone marks were performed in all iPSC lines of the cohort. Approximately $10^7$ cells were harvested detaching them with Accutase and then crosslinked with formaldehyde 1% in PBS 1X (Sigma, F8775) for 8 minutes at RT in rotation. Crosslinking was stopped by adding glycine at a final concentration of 0.125 mM and incubation on ice for 5 minutes. Cells were pelleted by centrifugation at 500 g for 3 min at 4°C and the pellet was washed once in ice cold PBS. Cell were lysed in 10 mL buffer A (50 mM HEPES pH 8.0, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP40, 0.25% Triton X-100) for 10 minutes on ice. After centrifugation the pellet was resuspended in 5 mL buffer B (10 mM Tris pH 8, 1 mM EDTA, 0.5 mM EGTA and 200 mM NaCl) and incubated for 5 min on ice. Then nuclei were pelleted by centrifugation at 500 g for 3 min at 4°C, resuspended in 150 ul buffer C (50 mM Tris pH 8, 5 mM EDTA, 1% SDS, 100 mM NaCl, 1x Roche complete mini protease inhibitors) and incubated on ice for 10 min. Then 3.5 mL ice cold TE buffer was added and chromatin was sheared in 13 mL tubes in a Branson Sonifier 450 (Marshall Scientific) device for 4 pulses (30 s ON / 30 s OFF) at 30% amplitude. Then 350 ul 10x ChIP buffer (0.1% SDS, 10% Triton X-100, 12mM EDTA, 167mM Tris-HCl pH 8, 1.67M NaCl) was added, chromatin was transferred into 2 mL Eppendorf tubes and spun for 10 min at 13000 g, 4°C. 1% sheared chromatin was saved as input control, the rest was transferred into fresh tubes. 5 ug of antibody were used for H3K27ac (abcam, ab4729), H3K4me1 (abcam, ab8895), and H3K9me3 (abcam, ab8898) IPs while 10 ug were used for ADNP (Sigma-Aldrich, F1804 monoclonal anti-FLAG M2) and CTCF (Cell Signaling, #2899) IPs. Samples were incubated overnight on a rotating wheel at 4°C. The next morning, 40 ul Protein G Dynabeads (mixed 1:1 and washed with 1x ChIP buffer) were added and incubation was continued for another 4 h. ChIPs were washed for 1 minute each with 4x LSB (10mM Tris-

HCl pH 8.0, 1mM EDTA pH 8.0, 140mM NaCl, 1% Triton X-100, 0.1% SDS, 0.1% Na-deoxycholate), 1x HSB (10mM Tris-HCl pH 8.0, 1mM EDTA pH 8.0, 360mM NaCl, 1% Triton X-100, 0.1% SDS, 0.1% Na-deoxycholate), 2x LiSB (10mM Tris-HCl pH 8.0, 1mM EDTA, pH 8.0, 250mM LiCl, 0.5% NP-40, 0.5% Na-deoxycholate), 1x TEplus (10mM Tris-HCl pH 8.0, 1mM EDTA, 50mM NaCl). Beads were transferred to a fresh tube during the last wash and wash buffer was completely removed before adding 100 ul elution buffer (10mM Tris-HCl pH 8.0, 1mM EDTA pH 8.0, 150mM NaCl, 1% SDS) and incubating 30 minutes at 65°C with constant shaking. Elution was repeated once more with 50ul elution buffer for 10 minutes and eluates were pooled, 2 ul RNaseA (20ug/ul) were added and samples were incubated for 1 h at 37°C. Input samples were adjusted to 150ul total volume with elution buffer and processed similar to ChIP samples. Then 2ul Proteinase K (20mg/ml) was added and samples were incubated 2 h at 55°C and then O/N h at 65°C. 2X volumes of AMPure XP beads and 1 volume Isopropanol was added and samples were vigorously mixed and incubated for 10 min at RT. Then beads were collected on a magnetic rack, washed twice with 80% EtOH and DNA was eluted in 30ul 10mM Tris pH8.0 for 5 min at 37°C. DNA libraries were prepared by Genomic Unit at the IFOM/IEO/IIT campus according to established protocol[356] and sequenced on the Illumina Novaseq 6000 instrument at 50bp paired-end read length and a coverage of 35 million reads per sample.

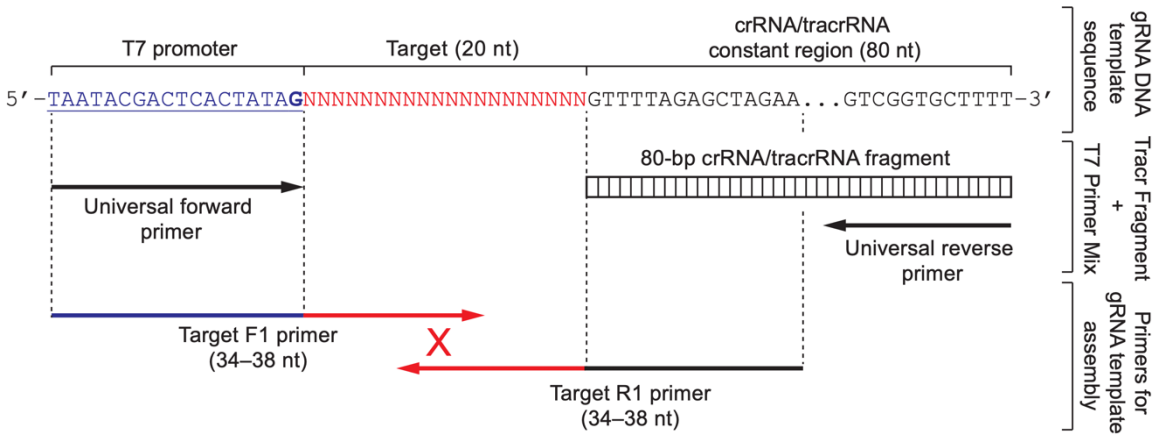
### 5.1.7 CRISPR/Cas9 engineering

The donor plasmid was designed to knock-in 3XFLAG tag into control and patient lines. HAs were designed to flank the recombination cassette according to the updated assembly of the human genome (GRCh38.p13). The construct was synthesised by GeneArtTM (Thermo Fisher Scientific). sgRNAs design was performed using the CRISPR design tool of Benchling available at the following website: https://www.benchling.com/. sgRNA was chosen according to the quality score and the position of the cut. Two scores were evaluated, the on-target and the off-target score, both ranging from 0 to 100 , defined based on algorithm by Doench et al and Hsu et al[357,358]. The off-target score indicates the inverse probability of the Cas9/gRNA complex to bind sequences in the rest of the genome. The on-target score represented the cleavage efficiency of the Cas9. sgRNA was synthesised using the GeneArtTM Precision gRNA Synthesis Kit (Thermo Fisher Scientific, A29377) by in vitro transcription on a DNA template containing the T7 promoter, the target gRNA sequence and the constant part of the crRNA/tracrRNA. The DNA template was generated by PCR amplification using two oligonucleotides. The forward primer, harbouring the target gRNA sequence and the T7 promoter; and the reverse primer, that overlaps with the forward primer

at the gRNA target sequence, and with the constant part of the crRNA/tracrRNA. The sequence of the constant portion of the crRNA/tracrRNA to include in the oligonucleotides was provided by the manufacturer. A schematic summary of the gRNA specifics and the oligonucleotides used for the PCR assembly is represented in **Table 2**. Oligonucleotides were synthesised by Sigma-Aldrich. After assembling the PCR reaction according to the manufacturer instructions, the reaction proceeded as illustrated below.

| | | | |
|---|---|---|---|
| T7 promoter | | | |
| Target gRNA | | | |
| crRNA/tracrRNA | | | |
| | | | |
| **gRNA Squence** | **On-target score** | **Off-target score** | |
| AATACTTAGTGACATTGGGT | 54.4 | 72 | |
| | | | |
| **Oligonucleotides for PCR assembly** | | | |
| TAATACGACTCACTATAGAATACTTAGTGACATTGGGT | | | |
| TTCTAGCTCTAAAACACCCAATGTCACTAAGTATT | | | |
| | | | |
| **gRNA DNA template sequence** | | | |
| TAATACGACTCACTATAGAATACTTAGTGACATTGGGTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT | | | |

**Table 2.** gRNA sequence used to engineer endogenous locus of ADNP reported with on-target and off-target score. Colour legend marks the portions of the oligonucleotides used to PCR-assemble the DNA template necessary for the in-vitro transcription of the gRNA.




In vitro transcription was performed for 4 hours at 37°C using the T7 RNA polymerase and template DNA was removed using DNAseI (both provided by the kit). In vitro transcribed RNA was purified using the gRNA Clean Up reagents of the kit. sgRNAs were loaded on a denaturing gel as a control. The concentration was determined using a NanoDrop spectrophotometer.

CRISPR/Cas9 conditions were standardised using a Cas9 purified by Cogentech biochemistry facility (Cogentech SRL). The latter was available solubilised in transduction buffer (5x Transduction buffer: 500 mM NaCl, 25 mM NaH2PO4, 250 mM NDSB-201, 150

mM glycerol, 75 mM glycine, 1.25 mM MgCl2, 1 mM 2-mercaptoethanol at pH 8.0 in milliQ water[359].

The Neon Transfection System (Thermo Fisher Scientific) with the 100 µL kit (Thermo Fisher Scientific, MPK10096) was used to electroporate hiPSC. 70-80 % confluent cells were pre-treated for 2-4 hours with 10 µM ROCK inhibitor to enhance cell-survival after electroporation. At the moment of the electroporation, hiPSC were detached using Accutase solution (1 mL x 10 cm plate), centrifuged for 3 minutes at 160g with a standard centrifuge. Cells were then resuspended in PBS and counted with an automated cell counter. In the meantime, the ribonucleic complex Cas9/sgRNA (molar ratio: 1:2,58) was assembled and incubated 10 minutes at 37˚C to enhance the complex formation. Once formed, the complex is stable for two hours. For each electroporation reaction, $4x10^5$ cells, 1.5 µg donor plasmid and corresponding amounts of Cas9/gRNA complex were used (10 µg of Cas9 each $4x10^5$ cells). Cells were resuspended in electroporation buffer T (provided by the kit) in order to have $4x10^5$ cells in 50 µL. The donor plasmid was added to the Cas9/gRNA complex for the HDR-mediated integration of the construct in the control and patient lines. The mixture was added to 50 µL of cells and 120 µL volume was reached adding buffer T. The Neon pipet was used to slowly introduce the electroporation mix in the 100 µL tip avoiding bubbles formation. The 20 µL excess was necessary to avoid bubble formation. The tip was introduced into the electroporation station containing 3 mL of buffer E2 (provided by the kit). Electroporation was performed with the following conditions: amplitude of the pulse: 900 V, pulse width: 20 ms, n° of pulses: 3. Electroporated cells were directly plated on matrigel coated dishes in TeSR-E8 supplemented with 10 µM ROCK inhibitor. A total 3,2 $x10^6$ cells were electroporated in 8 rounds of $4x10^5$, and plated in 15 cm in a dish to allow clonal growth.

Hygromycin B (Merck, 31282-04-9) selection was carried out after 48h from the electroporation using a final concentration of 50 µg/mL for 10 days. This concentration was previously standardised in the lab as the optimal for hiPSC. After 10-15 days individual colonies were big enough for manual picking. Clones were transferred in Eppendorf tubes, centrifuged 160g for 5’ and then the pellet was resuspended in TeSR-E8 supplemented with 5 µM ROCK inhibitor and plated in 48 well plates for expansion, screening and freezing.

Cells were incubated O/N at 60°C with Bradley Lysis Buffer (10 mM Tris-HCl pH 7.5, 10 mM EDTA, 0,5% SDS, 10 mM NaCl in ddH2O) containing Proteinase K (1 mg/ml). The day after, ice-cold EtOH/NaCl mix (EtOH 100%, NaCl 5M) was added to precipitate the DNA and the cells were incubated 30’ at -80°C and then centrifuged 20 minutes at 3000 rpm. The pellet was then washed twice in cold 70% EtOH and centrifuged 10 minutes at

3000 rpm. Once EtOH was completely removed, DNA was eluted in 30 µl of warm TE buffer (Tris-HCl pH8, EDTA 1 mM pH 8.0). DNA was then incubated 10 minutes at 56˚C. Nanodrop spectrophotometer was used for quantification. 25 ng of genomic DNA were used for PCR amplification. The extracted DNA was used to screen via PCR for correctly edited clones and further validation was performed using Sanger sequencing. The screening strategy was performed by sequential PCRs on multiple clones to verify the integration of the cassette in the correct locus. Clones were subjected to amplification with two primers annealing on the resistance cassette. Positive clones were then screened for the correct integration at the 5 minutes and at the 3 minutes using couples of primers annealing outside and inside the donor construct. Lastly, clones with the correct integration site were amplified with two primers designed in the HAs and spanning the whole construct as test of heterozygosity. Amplicons obtained from the PCRs were subjected to Sanger sequencing to confirm the presence of the cassette and the absence of unwanted rearrangements around the recombination site. All oligos were ordered by Sigma-Aldrich. All Sanger sequencing experiments were performed by Cogentech SRL. Primers used in the screening and validation are listed in **Table 3**.

| | |
|---|---|
| Fw | CTCGTGCTTTCAGCTTCGATGTAGG |
| Rv | GTCTTAGGGTAAAATCCCACTTGTTCC |
| | |
| Fw | ACAAACTTCTGTTTTAAGGAACTTGGG |
| Rv | ACGGCGGGAGATGCAATAGG |
| | |
| Fw | GCCCAGAAGGTTGAGGCTATGG |
| Rv | CTCAATGGTAGCCTTCTGCTTCAGG |
| | |
| Fw | CTCGTGCTTTCAGCTTCGATGTAGG |
| Rv | TGTAGTCCTTGTCGTCATCGTCTTTG |

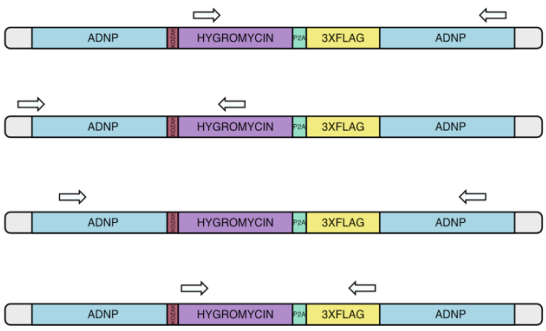



**Table 3**. Primers used for PCR screening of the selected clones.

### 5.1.8 ATAC-seq (Omni-ATAC)

ATAC-seq was performed as reported in Corces et al[360]; an improved version of the original protocol with the main advantages of higher accuracy in calling genome-wide accessible regions, and reduced mitochondrial DNA contamination. Omni-ATAC was first thoroughly standardised to achieve the best working conditions using a Tn5 prepared and purified in IEO crystallography unit. 50,000 iPSCs were collected and centrifuged at 500g for 5 min in a pre-chilled centrifuge and then briefly resuspended in ice-cold ATAC Resuspension Buffer (ATAC-RSB buffer: Tris-HCl pH 7.4 10mM, NaCl 10mM, MgCl2 3mM) supplemented with NP-40 0.1%, Tween-20 0.1% and Digitonin 0.01%. Samples were incubated on ice for

3 minutes. Lysis was washed out with ATAC-RSB supplemented with Tween-20 0.1% without NP-40 or Digitonin. Nuclei were then centrifuged for 10 minutes at 500g in a pre-chilled centrifuge. To allow transposase reaction, samples were resuspended in 50ul of ice-cold transposition mixture (TD buffer 2X, MEDS-loaded Tn5 100nM, PBS 33%, digitonin 0.01%, Tween-20 0.1%) and then incubated for 30 minutes at 37C on agitation (1000 rpm). Tn5 was pre-loaded with pre-annealed Mosaic End double-stranded (MEDS) oligonucleotides as described in Picelli et al protocol[361]. To clean up transposase reaction, samples were purified with Zymo DNA Clean and Concentrator-5 kit (Zymo Research), according to manufacture instructions. Eluted tagmented DNA was PCR amplified for 5 cycles using NEBNext Master Mix (NEB) and barcoded with Unique Dual Indexes (UDIs) which mitigate sample misassignment due to index hopping during de-multiplexing. 5 ul of the total 50 ul PCR reaction were collected for qPCR quantification using Viia7 Real-Time PCR system in order to assess the right number of additional cycles required to obtain optimal complexity during library amplification[155]. 8 final PCR cycles (5 pre-amplification + 3 extra cycles) was established as the gold standard for our iPSCs samples with conditions as reported in **Table 4**.

| **Cycling Conditions** | |
|---|---|
| 72 C | 5 min |
| 98 C | 30 sec |
| **5 cycles of** | |
| 98 C | 10 sec |
| 63 C | 30 sec |
| 72 C | 1 min |
| **Hold at 4 C** | |

**Table 4**: Omni-ATAC PCR conditions for library preparation

DNA fragments obtained at the end of library preparation underwent a doubled-sided size selection to remove primer dimers and fragments larger than 1000bp. To remove DNA fragments > 1000bp, 0.5X volumes of Agencourt AMPure beads XP (Beckman Coulter) were added to the samples, then incubated for 10 minutes at room temperature. The supernatant, containing DNA fragments < 1000bp, was transferred in a new tube and incubated for 10 minutes at room temperature with 1.3X original volume AMPure beads. Supernatant, containing primer dimers, has been discarded and the DNA-beads complex was

washed 3X with Ethanol 80% and eluted in water. Libraries was quantified by Qubit DNA High sensitivity (Thermo Fisher), checked with Bioanalyzer high-sensitivity and sequenced on an Illumina NovaSeq 6000 at 50bp paired-end read length and a coverage of 60 million reads per sample.

### 5.1.9 Cortical spheroids generation (Pasca protocol)

Cortical organoids were generated using an adaptation of the previously described protocol published by Pasca et al[262]. hiPSC were expanded on matrigel-coated 10 cm plates and dissociated at 60% confluency with Accutase solution for 3 minutes. Cells were centrifuged to remove the enzymatic suspension (160g for 3 minutes). After resuspension in TeSR-E8 medium supplemented with 5 μM ROCK inhibitor cells were counted with a TC20 automatic cell counter (Biorad) and seeded into 96 ultra-low attachment well plates (S-bio Duotech, MS- 9096UZ) at a final concentration of 2 x $10^4$ cells in each well. EB are generated by single-cell aggregation, therefore plates were centrifuged at 160g for 3 minutes to enhance EB formation. The following day medium was not changed, leaving EB undisturbed. At this point, Dorsomorphin (5 μM, MedChem express, HY-13418A) and TGF-β inhibitor SB-431542 (10 μM, MedChem express, HY-10431) were used to perform dual-SMAD inhibition, pushing neuroectoderm specification. Dual-SMAD inhibition was kept for a total of 5 days, with daily media change. On day 6 the differentiation medium 2 was added until day 25 with daily media change for the first 12 days, and then every other day. The differentiation medium 2 was composed of neurobasal medium (Thermo Fisher Scientific, 12348017) supplemented with 1X B-27 supplement without vitamin A (Thermo Fisher Scientific 12587001), 2 mM L-Glutamine, P/S, 100 U/mL, 20 ng/mL FGF2 and 20 ng/mL EGF (Thermo Fisher Scientific, PHG0313). Human FGF2 and EGF were used to amplify the pool of neural progenitors. On day 12, organoids were moved to ultra-low attachment 10 cm dishes and grown on shakers to enhance oxygen and nutrient supply. On day 26, FGF2 and EGF were replaced with 20 ng/mL brain-derived neurotrophic factor (BDNF) (Peprotech 450-02) and 20 ng/mL neurotrophin-3 (NT3, Peprotech 450-03) to promote differentiation of neural progenitors towards the glutamatergic fate.

### 5.1.10 Cortical organoids generation (Muotri protocol)

Cortical organoids were generated using an adaptation of the previously described protocol published by Trujillo et al[308]. Feeder-free iPSCs were fed daily with mTeSR1 for at least 7

days and colonies are then split with Accutase, collected and centrifuged at 160g for 3 min. Cell pellet were then resuspended in mTeSR1supplemented with 10 μM TGF-β inhibitor SB-431542 (MedChem express, HY-10431), 1 uM Dosromorphin (MedChem express, HY-13418A) and 5 μM ROCK inhibitor and counted with TC20 automatic cell counter. Approximately 4 x $10^6$ cells were transferred to one well of a 6-well plate and kept in suspension under rotation (95 rpm) for 24 hours to form free-floating spheres. The media was daily changed for the following 2 days. After 3 days, mTeSR1 was replaced by Media 1 [Neurobasal A (Life Technologies) supplemented with GlutaMAX, 1% N2 NeuroPlex (Gemini Bio-Products), 1% B27 + Vitamin A (Gemini Bio-Products), 1% NEAA (Life Technologies), 1% PS (Life Technologies), 10 μM SB and 1 μM Dorsomorphin] for 7 days. Then, the cells were maintained in Media 2 [Neurobasal A with GlutaMAX, 1% B27 + vitamin A, 1% NEAA and 1% PS] supplemented with 20 ng/mL FGF2 (Life Technologies) for 7 days, followed by 7 additional days in Media 2 supplemented with 20 ng/mL of FGF2 and 20 ng/mL EGF (PeproTech) to favour neural progenitors' proliferation. Next, cells were transferred to Media 3 [Media 2 supplemented with 10 ng/mL of BDNF, 10 ng/mL of GDNF, 10 ng/mL of NT-3 (all from PeproTech), 200 μM L-ascorbic acid and 1 mM dibutyryl-cAMP (Sigma-Aldrich)] to promote maturation, gliogenesis and activity. After 7 days, cortical organoids were maintained in Media 2 for as long as needed, with media changes every 3-4 days.

### 5.1.11 Cortical spheroids morphometrical characterization (ScanR microscopy)

Organoid morphometric properties were assessed for each individual cell line and bright-field images were taken using the OLYMPUS IX81-ZDC inverted microscope, equipped with a Hamamatsu ORCA-ER B/W CCD camera at a magnification of 4X/0.16 (dry). Acquisition was fully automated with the ScanR Acquisition software (version 2.4.0.13), in which a single field of view was acquired for each individual well in 96-well format. Images were analyzed using the open access Fiji software (FIJI-ImageJ v2.1.0, USA) and data was plotted using Prism (version 8.4.0). Images were processed using a custom-tailored semi-automated workflow. Briefly, after contrast was enhanced equally in all field of view, images were segmented and binarized, and their specific area, perimeter, circularity, and solidity were analyzed using Fiji's Analyze Particle plugin and saved on the region of interest manager. The measured parameters correspond to the largest cross-section of each organoid.

### 5.1.12 Cortical organoid clarification and immunofluorescence for 3D imaging analysis

#### *5.1.12.1 Clearing protocol and immunostaining*

The method described in this paragraph is fully performed using reagents and manufacturer's instructions provided in the MACS® Clearing Kit (Miltenyi Biotec, 130-126-719).

Cortical organoids of each sample were collected from shaking 6-well plates and gently transferred in 48-well plates for fixation using cut pipette tips. Organoids were washed 3 times in PBS to remove residual medium and then fixed for 20 minutes at room temperature with 2 ml of para-formaldehyde (4%). Organoids were washed again three times in PBS to remove fixative solution. Up to three organoids were permeabilized together in 0.5 ml of Permeabilization Solution in 1.5 ml Eppendorf tubes for 6 hours at room temperature under slow continuous rotation. After incubation, Permeabilization Solution was discarded and substituted with freshly prepared 1X Antibody Staining Solution (10X Antibody Staining Solution is diluted 1:10 with sterile water beforehand) complemented with primary antibodies in a final volume of 0.4 ml of 1X Antibody Staining Solution. Up to three organoids were incubated together with gentle shaking for 40 hours at 37˚C. To remove unbound antibodies, Antibody Staining Solution was discarded and replaced with fresh ones, and incubated for 30 minutes at room temperature with slow continuous rotation. These steps were repeated 5 times to ensure full removal of unbound antibodies. Secondary antibodies were diluted according to manufacturer's recommendations in a final volume of 0.4 ml of 1X Antibody Staining Solution and added to the 1.5 Eppendorf tubes containing cortical organoids and then incubated for 1 hour at room temperature. To remove unbound antibodies, Antibody Staining Solution was discarded and replaced with fresh ones, and incubated for 30 minutes at room temperature with slow continuous rotation. These steps were repeated 5 times to ensure full removal of unbound antibodies. A summary of antibody details used with this protocol is reported below in **Table 5**. For the embedding, agarose gel was prepared by dissolving 1.5% agarose in double-distilled water in a microwave. Organoids were transferred to the bottom of a 15 ml Falcon tubes and agarose solution was poured on top of them once it was slightly cooled down. After gel solidification (15-20 minutes), agarose block was cut into approximately 5x5 mm pieces, each containing up to three cortical organoids, making sure that they were located in the corner of the gel block to facilitate imaging conditions. Dehydration solutions were prepared by diluting absolute ethanol in sterile water to obtain 50% and 70% ethanol solutions containing 2% Tween-20. Up to three embedded organoids were dehydrated with a series of ethanol dilutions in 15 ml Falcon tubes at room temperature under slow continuous rotation: 50% ethanol was

incubated for 2 hours, 70% ethanol for 2 hours, and finally 100% ethanol O/N. Newly thawed 2.5 ml of Clearing Solution was added into new 15 ml Falcon tubes containing dehydrated organoids, and incubated at room temperature under slow continuous rotation for 3 hours. Clearing Solution was replaced with fresh one after 3 hours, and incubation was continued for additional 3 hours. Clearing Solution was discarded and substituted with Imaging Solution to proceed with imaging acquisition.

| Antibodies | Dilution | Company | Catalog N° |
|---|---|---|---|
| Goat anti-Sox2 polyclonal | 1:250 | R&D systems | AF2018 |
| Rabbit anti-Pax6 polyclonal | 1:250 | BioLegend | 901301 |
| Mouse anti-Nestin monoclonal | 1:250 | Millipore | MAB5326 |
| Rabbit anti-BLBP polyclonal | 1:250 | Millipore | ABN14 |
| Mouse anti-MAP2 monoclonal | 1:250 | BD Transduction Laboratories | 610460 |
| Rat anti-Histone H3 (phospho S28) monoclonal | 1:500 | Abcam | Ab10543 |
| Anti-goat Alexa 647 | 1:400 | Jackson ImmunoResearch | 705-605-147 |
| Anti-rabbit Alexa 594 | 1:400 | Jackson ImmunoResearch | 705-165-147 |
| Anti-mouse Alexa 488 | 1:400 | Jackson ImmunoResearch | 715-545-150 |

**Table 5:** List of antibodies used for the immunostaining of cleared cortical organoids.

### *5.1.12.2 Acquisition and data analysis*

For measuring NPCs abundance in control and mutated lines, organoids were immunostained and cleared as described above, and then visualised in a Yokogawa Spinning Disk Field Scanning Confocal System (Yokogawa CSU-W1 25µm-50µm pinhole dual disk, Nikon, Japan), equipped with motorised stage x-y-z, and a Prime BSI camera (Teledyne Photometrics, Arizona, USA), in confocal mode. Four channel (DAPI, Nestin, Pax6, and Sox2) z-stack images of the whole organoid (Z-step intervals of 5 µm) were acquired using the Nikon NIS Elements AR software (version 5.02.03) at a 10x/0.3 magnification (dry, no binning). Images were later analyzed with an open-source software (FIJI-ImageJ v2.1.0,

USA), in which all organoids were measured using a semi-automated macro with the criteria of analysis fully established and applicable to all images. The DAPI channel was used in all samples to normalize all area and/or count measurements (Pax6 and Sox2; PH3). All channels analyzed were transformed into binary images using available methods ("Otsu" for DAPI and Sox2; "Yen" for Pax6) and converted to masks, from which the area was later measured, and calculated individually accounting for each organoid volume. For counting cells actively undergoing mitosis in the two conditions, images were processed in the Arivis Vision 4D software (version 3.5.0, Arivis, Germany), using a fully automated custom pipeline. Briefly, to facilitate the downstream 3D workflow, DAPI and PH3 channels were processed separately, and advanced image enhancement filters were applied. For volume measurements, the DAPI channel in each organoid was later segmented using the "Li" thresholder, and objects smaller than 10000 $\mu m^3$ were filtered out. For counting the number of PH3-positive cells, the "Blob Finder" analysis operator was used to segment cells and the following parameters were applied: 1) averaged diameter of 6 µm; 2) 5% probability threshold; 3) 60% split sensitivity. Finally, raw data was processed, and the number of particles was normalized to DAPI volume. All data is presented as median of *n* different organoids for all analysis.

### 5.1.13 Single-cell Multiome (scATAC-seq + scRNA-seq)

Cortical organoids were dissociated using Dissociation Buffer composed of 0.5% BSA, 2 mM EDTA, 0.1% Accutase, 0.4 mg/ml Collagenase/Dispase (Sigma-Aldrich, 10269638001), 0.05 mg/ml DNAse I (Sigma-Aldrich, 10104159001) diluted into final volume of PBS, and filtered with 0.22 µm filters. Few homogeneous-sized organoids were incubated for approximately 40 minutes on a rotating wheel at 37˚C with 1 ml of Dissociation Buffer, manually pipetting every 10 minutes. Organoids were transferred to a new Eppendorf leaving behind undissociated pieces, and then centrifuged at 300g for 5 minutes at 4˚C. Cells were resuspended in 1 ml of PBS-BSA 0.04% and filtered with 40 µm Flowmi cell strainer (Bel-Art, H13680-0040) to remove the majority of debris. Cell were counted manually and then multiplexed together with other samples to obtain a final number of 1 million cells, with equal representation from each sample (e.g., 250K cells in case of 4 samples multiplexed). This step allows to pool together different genotypes and then de-multiplexed them based on their transcriptome already profiled at iPSC stage. Multiplexed cell solution was washed twice with PBS-BSA 0.04% and pellet was resuspended in 100 µl of Lysis Buffer [10mM Tris-HCl (pH 7.4), 10mM NaCl, 3mM $MgCl_2$, 0.1% Tween-20 (Bio-Rad, 1662404), 0.1% Nonidet P40 Substitute (Sigma-Aldrich, 74385), 0.01% Digitonin

(Thermo Fisher, BN2006), 1% BSA, 1 mM DTT (Sigma-Aldrich, 646563), 1U/µl RNase inhibitor (Sigma-Aldrich, 3335399001), nuclease-free water], and incubated 5 minutes on ice. Lysis Buffer was washed away three times with 1 ml of Wash Buffer [10mM Tris-HCl (pH 7.4), 10mM NaCl, 3mM $MgCl_2$, 1% BSA, 0.1% Tween-20, 1 mM DTT, 1U/µl RNase inhibitor, nuclease-free water]. Assuming 50% nuclei loss during lysis, nuclei were resuspended in Diluted Nuclei Buffer [1X 10x Genomics Nuclei Buffer, 1 mM DTT, U/µl RNase inhibitor, nuclease-free water] according to the reference table provided by 10X protocol appendix. The volume of Nuclei Diluted Buffer is critical to fit the right range of concentration based on the number of targeted nuclei recovery, therefore avoiding overcrowding of the Chromium machine during tagmentation and GEM preparation steps. Resuspended nuclei were passed again through a Flowmi cell strainer and counted to check for the right concentration feasible for the number of targeted nuclei. 5,000 was the number of targeted nuclei recovery for each multiplexed sample in all the experiments (e.g., 20,000 nuclei were targeted in the reactions with 4 multiplexed samples). DNA libraries were prepared by Genomic Unit at the IFOM/IEO/IIT campus according to manufacturer protocol and sequenced on the Illumina Novaseq 6000 instrument at a coverage of 50,000 reads per nucleus.

## 5.2 Bioinformatic Procedures

### 5.2.1 RNA-seq analysis

Bulk RNA-seq was quantified using Salmon v 1.4 on Gencode GRCh38.p13 human genome assembly. Differential Expression Analysis was performed using generalized linear models (GLM) using trimmed mean of M-Values (TMM) normalization and robust estimate of the negative binomial dispersion parameter for each gene, with expression levels specified by a log-linear model, using observation weights. Master regulatory analysis was performed via recursive hypergeometric tests: for each transcription factors a gene-set derived from ChIP-seq experiments and motif-aware databases (e.g., Jaspar and TFBS) is compared to the target list of differentially expressed genes. Correction by background was performed against expressed genes. Gene Onthology (GO) and pathway enrichments were performed using topGO and clusterProfiler packages, eventual alternative background gene lists are specified in the text. Annotated heatmaps were generated using *pheatmap* library.

### 5.2.2 ChIP-seq analysis

Chromatin Immunoprecipitation sequencing (ChIP-seq) reads were trimmed for library-specific adaptor contamination before being aligned to the GRCh38.p13 human genome assembly with Bowtie 2 (removing multi-mapping reads with samtools). We performed peak calling via MACS 2.1 using narrow settings for ADNP and CTCF (-q 0.05) and broad settings for histone marks (H3K27ac, H3K4me1 and H3K4me3). Peaks overlap were perfomed by means of bedtools v2.28, and bound genes were defined as intersecting promoters (from 500bp upstream to 250bp downstream TSS), or intersecting enhancers (cell-type specific 4DGenome consortium peak sets). Presence of peaks in exclusion lists from ENCODE was verified. Reference peak sets for ADNP and CTCF were chosen by filtering MACS2 Score and qValue: peaks with score equal or higher to median, and qValue higher than first quartile were retained. Reference peak sets for histone marks were defined as present in at least half controls (3 out of 6 samples). We identified "lost peaks" as regions preferentially found in CTLs (i.e., regions in at least 5 out of 6 controls, and none of HVDAS, plus regions found in all controls and at most 1 out of 3 HVDAS lines), and "gained peaks" as regions preferentially found in HVDAS lines (following the same logic). Motif enrichments were performed with Homer v.4.11 using default parameters. Principal Component Analysis (PCA) was performed and represented using DeepTools v3.5 package on multiBamSummary output using 100bp binning. Heatmaps were generated using deepTools plotHeatmap on bamCoverage (RPGC normalization) or bamCompare (normalization on Inputs) ouputs, as stated in the text. Reference peak files were chosen depending on histone marks or TF, as stated in the previous sections, or as specified in the text. Unless otherwise specified in the text, the number of mapped reads was used as the library size, mean difference genewise exact test was used for assessing the statistical significance of the intensive changes in ChIP-seq signals in each differential analysis, using edgeR v.3.12.1. Master regulatory analysis was performed as it follows: for each transcription factors a gene-set derived from ChIP-seq experiments and motif-aware databases (e.g., Jaspar) is compared via hyper-geometric test to the target list of genes, correcting by background (e.g., expressed genes for differentially expressed genes in RNA-seq, or bound genes against differentially marked regions in ChIP-seq). GO and pathway enrichments were performed using ChIPseeker and clusterProfiler packages, background genes are specified in the text.

### 5.2.3 ATAC-seq analysis

Reads were aligned using bowtie 2 on the same human hg38 reference used for ChIPseq. Fragment length distributions were plotted to verify the quality of each sequencing run. Reads were aggregated by genotype. Nucleosome free, mono- and di-nucleosome reads were extracted depending on their size. Reads shifting was performed using DeepTools alignmentSieve. Peak calling was performed using Genrich ATAC mode. Motif enrichment, differential peak calling between patients and controls, and quantitative analyses were performed as for ChIP-seq.

### 5.2.3 Single-cell Multiomic analysis

Reads were aligned using cellranger-arc v.2.11. *In silico* demultiplexing has been performed on each batch of sequencing using single nucleotide variants (SNP) called from bulk RNA-seq using internal pipelines (alignment with STAR on cellranger reference, followed by GATK). Multiomic data was split into Gene Expression (RNA) and Chromatin Accessibility (ATAC) components before quality control. They are separately filtered according to the type of observable (counts per cell, mitochondrial genes expression/percentage of mitochondrial reads, percentage of reads in peaks, TSS enrichments, etc). The two modes were normalized independently following best practices (scaling and single-cell library size normalization for RNA-seq, TF-IDF for ATAC-seq). Label transferring has been used to verify that cells cluster coherently in ATAC and RNA. We performed two independent dimensionality reductions (UMAP) on ATAC and RNA. Clustering was performed via leiden algorithm (resolution 1.0). Integration of multiple batches was performed using Harmony. Ingestion on published reference data was performed for cell-type association and benchmarking.

# BIBLIOGRAPHY


1. Sherr, E. H. *et al.* Neurodevelopmental disorders and genetic testing: Current approaches and future advances: Genetic Testing for Neurodevelopmental Disorders. *Ann Neurol.* n/a-n/a (2013) doi:10.1002/ana.23950.
2. Elsabbagh, M. *et al.* Global prevalence of autism and other pervasive developmental disorders. *Autism Res* **5**, 160–179 (2012).
3. Maulik, P. K., Mascarenhas, M. N., Mathers, C. D., Dua, T. & Saxena, S. Prevalence of intellectual disability: a meta-analysis of population-based studies. *Res Dev Disabil* **32**, 419–436 (2011).
4. Rice, C. E. *et al.* Evaluating Changes in the Prevalence of the Autism Spectrum Disorders (ASDs). *Public Health Rev* **34**, 1–22 (2012).
5. Gilissen, C. *et al.* Genome sequencing identifies major causes of severe intellectual disability. *Nature* **511**, 344–347 (2014).
6. Neale, B. M. *et al.* Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **485**, 242–245 (2012).
7. Iossifov, I. *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).
8. Wright, C. F. *et al.* Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *The Lancet* **385**, 1305–1314 (2015).
9. SFARI | Simons Foundation Autism Research Initiative. *SFARI* https://www.sfari.org/.
10. Savatt, J. M. & Myers, S. M. Genetic Testing in Neurodevelopmental Disorders. *Frontiers in Pediatrics* **9**, 52 (2021).
11. Short, P. J. *et al.* De novo mutations in regulatory elements in neurodevelopmental disorders. *Nature* **555**, 611–616 (2018).

12. Amberger, J. S., Bocchini, C. A., Scott, A. F. & Hamosh, A. OMIM.org: leveraging knowledge across phenotype–gene relationships. *Nucleic Acids Research* **47**, D1038–D1043 (2019).
13. Sanders, S. J. *et al.* Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* **70**, 863–885 (2011).
14. Vitriolo, A., Gabriele, M. & Testa, G. From enhanceropathies to the epigenetic manifold underlying human cognition. *Human Molecular Genetics* **28**, R226–R234 (2019).
15. McCarroll, S. A. & Hyman, S. E. Progress in the Genetics of Polygenic Brain Disorders: Significant New Challenges for Neurobiology. *Neuron* **80**, 578–587 (2013).
16. Niemi, M. E. K. *et al.* Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. *Nature* **562**, 268–271 (2018).
17. Robinson, E. B. *et al.* Autism spectrum disorder severity reflects the average contribution of de novo and familial influences. *Proc Natl Acad Sci U S A* **111**, 15161–15165 (2014).
18. O'Roak, B. J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* **485**, 246–250 (2012).
19. Cardoso, A. R. *et al.* Essential genetic findings in neurodevelopmental disorders. *Human Genomics* **13**, 31 (2019).
20. Sanders, S. J. *et al.* Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87**, 1215–1233 (2015).
21. Parenti, I., Rabaneda, L. G., Schoen, H. & Novarino, G. Neurodevelopmental Disorders: From Genetics to Functional Pathways. *Trends in Neurosciences* **43**, 608–621 (2020).

22. Krumm, N. *et al.* Excess of rare, inherited truncating mutations in autism. *Nat Genet* **47**, 582–588 (2015).
23. Adamo, A. *et al.* 7q11.23 dosage-dependent dysregulation in human pluripotent stem cells affects transcriptional programs in disease-relevant lineages. *Nat Genet* **47**, 132–141 (2015).
24. Iossifov, I. *et al.* Low load for disruptive mutations in autism genes and their biased transmission. *Proc Natl Acad Sci USA* **112**, E5600–E5607 (2015).
25. Autism Spectrum Disorder Working Group of the Psychiatric Genomics Consortium *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet* **51**, 431–444 (2019).
26. Guo, H. *et al.* Inherited and multiple de novo mutations in autism/developmental delay risk genes suggest a multifactorial model. *Molecular Autism* **9**, 64 (2018).
27. Walker, R. L. *et al.* Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms. *Cell* **179**, 750-771.e22 (2019).
28. Owen, M. J. & O'Donovan, M. C. Schizophrenia and the neurodevelopmental continuum:evidence from genomics. *World Psychiatry* **16**, 227–235 (2017).
29. The DDD Study *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).
30. Berkel, S. *et al.* Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. *Nat Genet* **42**, 489–491 (2010).
31. Pinto, D. *et al.* Convergence of Genes and Cellular Pathways Dysregulated in Autism Spectrum Disorders. *The American Journal of Human Genetics* **94**, 677–694 (2014).
32. Xu, Q. *et al.* Autism-associated CHD8 deficiency impairs axon development and migration of cortical neurons. *Molecular Autism* **9**, 65 (2018).

33. Quesnel-Vallières, M., Weatheritt, R. J., Cordes, S. P. & Blencowe, B. J. Autism spectrum disorder: insights into convergent mechanisms from transcriptomics. *Nat Rev Genet* **20**, 51–63 (2019).

34. Parikshak, N. N. *et al.* Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008–1021 (2013).

35. Velmeshev, D. *et al.* Single-cell genomics identifies cell type–specific molecular changes in autism. *Science* **364**, 685–689 (2019).

36. Gabriele, M., Lopez Tobon, A., D'Agostino, G. & Testa, G. The chromatin basis of neurodevelopmental disorders: Rethinking dysfunction along the molecular and temporal axes. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* **84**, 306–327 (2018).

37. Vega-Lopez, G. A., Cerrizuela, S., Tribulo, C. & Aybar, M. J. Neurocristopathies: New insights 150 years after the neural crest discovery. *Dev Biol* **444 Suppl 1**, S110–S143 (2018).

38. Smith, E. & Shilatifard, A. Enhancer biology and enhanceropathies. *Nat Struct Mol Biol* **21**, 210–219 (2014).

39. Lettice, L. A. *et al.* A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* **12**, 1725–1735 (2003).

40. Higgs, D. R. & Wood, W. G. Long-range regulation of alpha globin gene expression during erythropoiesis. *Curr Opin Hematol* **15**, 176–183 (2008).

41. Gabriele, M. *et al.* YY1 Haploinsufficiency Causes an Intellectual Disability Syndrome Featuring Transcriptional and Chromatin Dysfunction. *The American Journal of Human Genetics* **100**, 907–925 (2017).

42. Weintraub, A. S. *et al.* YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* **171**, 1573-1588.e28 (2017).
43. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
44. van Arensbergen, J. *et al.* High-throughput identification of human SNPs affecting regulatory element activity. *Nat Genet* **51**, 1160–1169 (2019).
45. Lalli, M. A., Avey, D., Dougherty, J. D., Milbrandt, J. & Mitra, R. D. High-throughput single-cell functional elucidation of neurodevelopmental disease–associated genes reveals convergent mechanisms altering neuronal differentiation. *Genome Res.* **30**, 1317–1331 (2020).
46. Willsey, H. R. *et al.* Parallel in vivo analysis of large-effect autism genes implicates cortical neurogenesis and estrogen in risk and resilience. *Neuron* **109**, 788-804.e8 (2021).
47. Ernst, C. Proliferation and Differentiation Deficits are a Major Convergence Point for Neurodevelopmental Disorders. *Trends in Neurosciences* **39**, 290–299 (2016).
48. Satterstrom, F. K. *et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568-584.e23 (2020).
49. Willsey, A. J. *et al.* The Psychiatric Cell Map Initiative: A Convergent Systems Biological Approach to Illuminating Key Molecular Pathways in Neuropsychiatric Disorders. *Cell* **174**, 505–520 (2018).
50. Forgacs, G. & Newman, S. A. *Biological Physics of the Developing Embryo*. (Cambridge University Press, 2005). doi:10.1017/CBO9780511755576.
51. *Seeley's Anatomy & Physiology*. (2019).

52. Schoenwolf, G., Steven Bleyl, Brauer, P. & Francis-West, P. *Larsen's Human Embryology*. (Churchill Livingstone Elsevier, 2009).
53. Rossant, J. Stem Cells and Early Lineage Development. *Cell* **132**, 527–531 (2008).
54. Gilbert, S. F. & Gilbert, S. F. *Developmental Biology*. (Sinauer Associates, 2000).
55. Thomson, J. A. *et al.* Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science* **282**, 1145–1147 (1998).
56. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
57. What is a Stem Cell? in *The Science of Stem Cells* 1–11 (John Wiley & Sons, Inc., 2017). doi:10.1002/9781119235293.ch1.
58. Knoblich, J. A. Mechanisms of Asymmetric Stem Cell Division. *Cell* **132**, 583–597 (2008).
59. Smith, K. P., Luong, M. X. & Stein, G. S. Pluripotency: Toward a gold standard for human ES and iPS cells. *Journal of Cellular Physiology* **220**, 21–29 (2009).
60. Boland, M. J. *et al.* Adult mice generated from induced pluripotent stem cells. *Nature* **461**, 91–94 (2009).
61. Zhao, X. *et al.* iPS cells produce viable mice through tetraploid complementation. *Nature* **461**, 86–90 (2009).
62. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J. C. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *PNAS* **90**, 8424–8428 (1993).
63. Gropp, M. *et al.* Standardization of the Teratoma Assay for Analysis of Pluripotency of Human ES Cells and Biosafety of Their Differentiated Progeny. *PLOS ONE* **7**, e45532 (2012).

64. Wen, D., Saiz, N., Rosenwaks, Z., Hadjantonakis, A.-K. & Rafii, S. Completely ES Cell-Derived Mice Produced by Tetraploid Complementation Using Inner Cell Mass (ICM) Deficient Blastocysts. *PLOS ONE* **9**, e94730 (2014).
65. Andrews, P. W. *et al.* Embryonic stem (ES) cells and embryonal carcinoma (EC) cells: opposite sides of the same coin. *Biochem Soc Trans* **33**, 1526–1530 (2005).
66. Muñoz-Sanjuán, I. & Brivanlou, A. H. Neural induction, the default model and embryonic stem cells. *Nat Rev Neurosci* **3**, 271–280 (2002).
67. James, D., Levine, A. J., Besser, D. & Hemmati-Brivanlou, A. TGFβ/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development* **132**, 1273–1282 (2005).
68. Shi, Y. & Massagué, J. Mechanisms of TGF-β Signaling from Cell Membrane to the Nucleus. *Cell* **113**, 685–700 (2003).
69. Massagué, J., Seoane, J. & Wotton, D. Smad transcription factors. *Genes Dev.* **19**, 2783–2810 (2005).
70. Vallier, L., Reynolds, D. & Pedersen, R. A. Nodal inhibits differentiation of human embryonic stem cells along the neuroectodermal default pathway. *Dev Biol* **275**, 403–421 (2004).
71. Vallier, L. *et al.* Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. *Development* **136**, 1339–1349 (2009).
72. Sato, N., Meijer, L., Skaltsounis, L., Greengard, P. & Brivanlou, A. H. Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nat Med* **10**, 55–63 (2004).
73. Nusse, R. *et al.* Wnt Signaling and Stem Cell Control. *Cold Spring Harb Symp Quant Biol* **73**, 59–66 (2008).

74. Clevers, H. Wnt/β-Catenin Signaling in Development and Disease. *Cell* **127**, 469–480 (2006).

75. ten Berge, D. *et al.* Embryonic stem cells require Wnt proteins to prevent differentiation to epiblast stem cells. *Nat Cell Biol* **13**, 1070–1075 (2011).

76. Sokol, S. Y. Maintaining embryonic stem cell pluripotency with Wnt signaling. *Development* **138**, 4341–4350 (2011).

77. Mullen, A. C. & Wrana, J. L. TGF-β Family Signaling in Embryonic and Somatic Stem-Cell Renewal and Differentiation. *Cold Spring Harb Perspect Biol* **9**, a022186 (2017).

78. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126**, 663–676 (2006).

79. Takahashi, K. *et al.* Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell* **131**, 861–872 (2007).

80. Huang, Y. *et al.* Type 2 Fibroblast Growth Factor Receptor Signaling Preserves Stemness and Prevents Differentiation of Prostate Stem Cells from the Basal Compartment. *J Biol Chem* **290**, 17753–17761 (2015).

81. Tian, R. *et al.* Fibroblast growth factor-5 promotes spermatogonial stem cell proliferation via ERK and AKT activation. *Stem Cell Res Ther* **10**, 40 (2019).

82. Mossahebi-Mohammadi, M., Quan, M., Zhang, J.-S. & Li, X. FGF Signaling Pathway: A Key Regulator of Stem Cell Pluripotency. *Frontiers in Cell and Developmental Biology* **8**, 79 (2020).

83. Yuan, H., Corbi, N., Basilico, C. & Dailey, L. Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes Dev* **9**, 2635–2645 (1995).

84. Guzman-Ayala, M., Ben-Haim, N., Beck, S. & Constam, D. B. Nodal protein processing and fibroblast growth factor 4 synergize to maintain a trophoblast stem cell microenvironment. *PNAS* **101**, 15656–15660 (2004).

85. Fuchs, E. & Horsley, V. Ferreting out stem cells from their niches. *Nat Cell Biol* **13**, 513–518 (2011).

86. Scadden, D. T. The stem-cell niche as an entity of action. *Nature* **441**, 1075–1079 (2006).

87. Spradling, A., Drummond-Barbosa, D. & Kai, T. Stem cells find their niche. *Nature* **414**, 98–104 (2001).

88. Ritsma, L. *et al.* Intestinal crypt homeostasis revealed at single-stem-cell level by in vivo live imaging. *Nature* **507**, 362–365 (2014).

89. Rompolas, P., Mesa, K. R. & Greco, V. Spatial organization within a niche as a determinant of stem-cell fate. *Nature* **502**, 513–518 (2013).

90. Li, L. & Clevers, H. Coexistence of Quiescent and Active Adult Stem Cells in Mammals. *Science* **327**, 542–545 (2010).

91. Rodolfa, K. T. Inducing pluripotency. in *StemBook* (Harvard Stem Cell Institute, 2008).

92. Gurdon, J. B., Elsdale, T. R. & Fischberg, M. Sexually Mature Individuals of Xenopus laevis from the Transplantation of Single Somatic Nuclei. *Nature* **182**, 64–65 (1958).

93. Hochedlinger, K. & Jaenisch, R. Monoclonal mice generated by nuclear transfer from mature B and T donor cells. *Nature* **415**, 1035–1038 (2002).

94. Tada, M., Takahama, Y., Abe, K., Nakatsuji, N. & Tada, T. Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. *Curr Biol* **11**, 1553–1558 (2001).

95. Tada, M. *et al.* Pluripotency of reprogrammed somatic genomes in embryonic stem hybrid cells. *Dev Dyn* **227**, 504–510 (2003).

96. Cowan, C. A., Atienza, J., Melton, D. A. & Eggan, K. Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. *Science* **309**, 1369–1373 (2005).

97. Waddington, C. H. The strategy of the genes. A discussion of some aspects of theoretical biology. With an appendix by H. Kacser. *The strategy of the genes. A discussion of some aspects of theoretical biology. With an appendix by H. Kacser.* (1957).

98. Chambers, I. *et al.* Functional Expression Cloning of Nanog, a Pluripotency Sustaining Factor in Embryonic Stem Cells. *Cell* **113**, 643–655 (2003).

99. Avilion, A. A. *et al.* Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* **17**, 126–140 (2003).

100. Zappone, M. V. *et al.* Sox2 regulatory sequences direct expression of a (beta)-geo transgene to telencephalic neural stem cells and precursors of the mouse embryo, revealing regionalization of gene expression in CNS stem cells. *Development* **127**, 2367–2382 (2000).

101. Davis, A. C., Wims, M., Spotts, G. D., Hann, S. R. & Bradley, A. A null c-myc mutation causes lethality before 10.5 days of gestation in homozygotes and reduced fertility in heterozygous female mice. *Genes Dev* **7**, 671–682 (1993).

102. Myc influences global chromatin structure. *The EMBO Journal* **25**, 2723–2734 (2006).

103. Chen, X. *et al.* Krüppel-like Factor 4 (Gut-enriched Krüppel-like Factor) Inhibits Cell Proliferation by Blocking G1/S Progression of the Cell Cycle*. *Journal of Biological Chemistry* **276**, 30423–30428 (2001).

104. Maherali, N. *et al.* Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1**, 55–70 (2007).

105. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313–317 (2007).

106. Gafni, O. *et al.* Derivation of novel human ground state naive pluripotent stem cells. *Nature* **504**, 282–286 (2013).

107. Theunissen, T. W. *et al.* Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell* **15**, 471–487 (2014).

108. Huang, K., Maruyama, T. & Fan, G. The naive state of human pluripotent stem cells: a synthesis of stem cell and preimplantation embryo transcriptome analyses. *Cell Stem Cell* **15**, 410–415 (2014).

109. Takashima, Y. *et al.* Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell* **158**, 1254–1269 (2014).

110. Li, M. & Belmonte, J. C. I. Deconstructing the pluripotency gene regulatory network. *Nat Cell Biol* **20**, 382–392 (2018).

111. Durruthy-Durruthy, J. *et al.* Spatiotemporal Reconstruction of the Human Blastocyst by Single-Cell Gene-Expression Analysis Informs Induction of Naive Pluripotency. *Dev Cell* **38**, 100–115 (2016).

112. Boyer, L. A. *et al.* Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. *Cell* **122**, 947–956 (2005).

113. Loh, Y.-H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* **38**, 431–440 (2006).

114. Rizzino, A. The Sox2-Oct4 Connection: Critical players in a much larger interdependent network integrated at multiple levels. *Stem Cells* **31**, 1033–1039 (2013).

115. Jaenisch, R. & Young, R. Stem Cells, the Molecular Circuitry of Pluripotency and Nuclear Reprogramming. *Cell* **132**, 567–582 (2008).

116. Wang, J. *et al.* A protein interaction network for pluripotency of embryonic stem cells. *Nature* **444**, 364–368 (2006).

117. Thomson, M. *et al.* Pluripotency Factors in Embryonic Stem Cells Regulate Differentiation into Germ Layers. *Cell* **145**, 875–889 (2011).

118. Guenther, M. G. *et al.* Chromatin Structure and Gene Expression Programs of Human Embryonic and Induced Pluripotent Stem Cells. *Cell Stem Cell* **7**, 249–257 (2010).

119. Mali, P. *et al.* Butyrate Greatly Enhances Derivation of Human Induced Pluripotent Stem Cells by Promoting Epigenetic Remodeling and the Expression of Pluripotency-Associated Genes. *STEM CELLS* **28**, 713–720 (2010).

120. Feldman, N. *et al.* G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nat Cell Biol* **8**, 188–194 (2006).

121. Marión, R. M. *et al.* A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. *Nature* **460**, 1149–1153 (2009).

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

123. Dimos, J. T. *et al.* Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. *Science* **321**, 1218–1221 (2008).

124. Okita, K., Nakagawa, M., Hyenjong, H., Ichisaka, T. & Yamanaka, S. Generation of Mouse Induced Pluripotent Stem Cells Without Viral Vectors. *Science* **322**, 949–953 (2008).

125. Woltjen, K. *et al.* piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature* **458**, 766–770 (2009).

126. Grabundzija, I. *et al.* Sleeping Beauty transposon-based system for cellular reprogramming and targeted gene insertion in induced pluripotent stem cells. *Nucleic Acids Research* **41**, 1829–1847 (2013).

127. Mandal, P. K. & Rossi, D. J. Reprogramming human fibroblasts to pluripotency using modified mRNA. *Nat Protoc* **8**, 568–582 (2013).

128. Yoshioka, N. *et al.* Efficient Generation of Human iPSCs by a Synthetic Self-Replicative RNA. *Cell Stem Cell* **13**, 246–254 (2013).

129. Briggs, J. A. *et al.* Integration-Free Induced Pluripotent Stem Cells Model Genetic and Neural Developmental Features of Down Syndrome Etiology. *STEM CELLS* **31**, 467–478 (2013).

130. Brennand, K. J. *et al.* Modelling schizophrenia using human induced pluripotent stem cells. *Nature* **473**, 221–225 (2011).

131. Cavallo, F. *et al.* High-throughput screening identifies histone deacetylase inhibitors that modulate GTF2I expression in 7q11.23 microduplication autism spectrum disorder patient-derived cortical neurons. *Mol Autism* **11**, 88 (2020).

132. Liu, X. *et al.* Idiopathic autism: Cellular and molecular phenotypes in pluripotent stem cell derived-neurons. *Mol Neurobiol* **54**, 4507–4523 (2017).

133. Marchetto, M. C. N. *et al.* A Model for Neural Development and Treatment of Rett Syndrome Using Human Induced Pluripotent Stem Cells. *Cell* **143**, 527–539 (2010).

134. Bar-Nur, O., Caspi, I. & Benvenisty, N. Molecular analysis of FMR1 reactivation in fragile-X induced pluripotent stem cells and their neuronal derivatives. *J Mol Cell Biol* **4**, 180–183 (2012).

135. Hockemeyer, D. & Jaenisch, R. Induced pluripotent stem cells meet genome editing. *Cell Stem Cell* **18**, 573–586 (2016).

136. Soldner, F. *et al.* Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. *Cell* **146**, 318–331 (2011).

137. Steyer, B. *et al.* Scarless Genome Editing of Human Pluripotent Stem Cells via Transient Puromycin Selection. *Stem Cell Reports* **10**, 642–654 (2018).

138. Paquet, D. *et al.* Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* **533**, 125–129 (2016).

139. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281–2308 (2013).

140. Nord, A. S. *et al.* Rapid and Pervasive Changes in Genome-wide Enhancer Usage during Mammalian Development. *Cell* **155**, 1521–1531 (2013).

141. Ostuni, R. *et al.* Latent Enhancers Activated by Stimulation in Differentiated Cells. *Cell* **152**, 157–171 (2013).

142. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet* **15**, 272–286 (2014).

143. Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**, 557-572.e24 (2017).

144. Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nat Rev Genet* **17**, 661–678 (2016).

145. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).

146. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**, 311–318 (2007).

147. Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol* **28**, 817–825 (2010).

148. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580–585 (2013).

149. Jasinska, A. J. *et al.* Genetic variation and gene expression across multiple tissues and developmental stages in a nonhuman primate. *Nat Genet* **49**, 1714–1721 (2017).

150. Nicolae, D. L. *et al.* Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS. *PLOS Genetics* **6**, e1000888 (2010).

151. Voss, T. C. & Hager, G. L. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nat Rev Genet* **15**, 69–81 (2014).

152. Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13**, 613–626 (2012).

153. Hardison, R. C. & Taylor, J. Genomic approaches towards finding cis-regulatory modules in animals. *Nat Rev Genet* **13**, 469–483 (2012).

154. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).

155. Buenrostro, J., Wu, B., Chang, H. & Greenleaf, W. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* **109**, 21.29.1-21.29.9 (2015).

156. Zhang, S. *et al.* Allele-specific open chromatin in human iPSC neurons elucidates functional disease variants. *Science* **369**, 561–565 (2020).

157. Rivera, C. M. & Ren, B. Mapping human epigenomes. *Cell* **155**, 39–55 (2013).

158. Lister, R. *et al.* Global Epigenomic Reconfiguration During Mammalian Brain Development. *Science* **341**, 1237905 (2013).

159. Cavalli, G. & Heard, E. Advances in epigenetics link genetics to the environment and disease. *Nature* **571**, 489–499 (2019).

160. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).

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

162. Cedar, H. & Bergman, Y. Programming of DNA Methylation Patterns. *Annual Review of Biochemistry* **81**, 97–117 (2012).

163. Suelves, M., Carrió, E., Núñez-Álvarez, Y. & Peinado, M. A. DNA methylation dynamics in cellular commitment and differentiation. *Briefings in Functional Genomics* **15**, 443–453 (2016).

164. Gifford, C. A. *et al.* Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell* **153**, 1149–1163 (2013).

165. Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134–1148 (2013).

166. Ji, H. *et al.* Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature* **467**, 338–342 (2010).

167. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev* **25**, 1010–1022 (2011).

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

169. Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **128**, 693–705 (2007).

170. Moore, J. E. *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).

171. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).

172. Bernstein, B. E., Meissner, A. & Lander, E. S. The Mammalian Epigenome. *Cell* **128**, 669–681 (2007).

173. Schuettengruber, B. & Cavalli, G. Recruitment of Polycomb group complexes and their role in the dynamic regulation of cell fate choice. *Development* **136**, 3531–3542 (2009).

174. Aloia, L., Di Stefano, B. & Di Croce, L. Polycomb complexes in stem cells and embryonic development. *Development* **140**, 2525–2534 (2013).

175. Simon, J. A. & Kingston, R. E. Mechanisms of Polycomb gene silencing: knowns and unknowns. *Nat Rev Mol Cell Biol* **10**, 697–708 (2009).

176. Shah, R. N. *et al. Re-evaluating the role of nucleosomal bivalency in early development*. 2021.09.09.458948 https://www.biorxiv.org/content/10.1101/2021.09.09.458948v1 (2021) doi:10.1101/2021.09.09.458948.

177. Bernstein, B. E. *et al.* A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* **125**, 315–326 (2006).

178. Pereira, J. D. *et al.* Ezh2, the histone methyltransferase of PRC2, regulates the balance between self-renewal and differentiation in the cerebral cortex. *Proc Natl Acad Sci U S A* **107**, 15957–15962 (2010).

179. Akizu, N. *et al.* EZH2 regulates neuroepithelium structure and neuroblast proliferation by repressing p21. *Open Biol* **6**, 150227 (2016).

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

181. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–283 (2011).

182. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences* **107**, 21931–21936 (2010).

183. Wang, C. *et al.* Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition. *Proc Natl Acad Sci USA* **113**, 11871–11876 (2016).

184. Froimchuk, E., Jang, Y. & Ge, K. Histone H3 lysine 4 methyltransferase KMT2D. *Gene* **627**, 337–342 (2017).

185. Zhang, J. *et al.* Disruption of KMT2D perturbs germinal center B cell development and promotes lymphomagenesis. *Nat Med* **21**, 1190–1198 (2015).

186. Ng, S. B. *et al.* Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* **42**, 790–793 (2010).

187. Fasciani, A. MLL4-associated condensates counterbalance Polycomb-mediated nuclear mechanical stress in Kabuki syndrome. *Nature Genetics* **52**, 40 (2020).

188. Clapier, C. R. & Cairns, B. R. The biology of chromatin remodeling complexes. *Annu Rev Biochem* **78**, 273–304 (2009).

189. de Dieuleveult, M. *et al.* Genome-wide nucleosome specificity and function of chromatin remodellers in ES cells. *Nature* **530**, 113–116 (2016).

190. Clapier, C. R., Iwasa, J., Cairns, B. R. & Peterson, C. L. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat Rev Mol Cell Biol* **18**, 407–422 (2017).

191. Basta, J. & Rauchman, M. The Nucleosome Remodeling and Deacetylase (NuRD) Complex in Development and Disease. *Transl Res* **165**, 36–47 (2015).

192. Ho, L. *et al.* An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proc Natl Acad Sci U S A* **106**, 5187–5191 (2009).

193. Schaniel, C. *et al.* Smarcc1/Baf155 couples self-renewal gene repression with changes in chromatin structure in mouse embryonic stem cells. *Stem Cells* **27**, 2979–2991 (2009).

194. Sokpor, G., Xie, Y., Rosenbusch, J. & Tuoc, T. Chromatin Remodeling BAF (SWI/SNF) Complexes in Neural Development and Disorders. *Front Mol Neurosci* **10**, 243 (2017).

195. Lessard, J. *et al.* An Essential Switch in Subunit Composition of a Chromatin Remodeling Complex during Neural Development. *Neuron* **55**, 201–215 (2007).

196. Vogel-Ciernia, A. *et al.* The neuron-specific chromatin regulatory subunit BAF53b is necessary for synaptic plasticity and memory. *Nat Neurosci* **16**, 552–561 (2013).

197. Alver, B. H. *et al.* The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. *Nat Commun* **8**, 14648 (2017).

198. de la Serna, I. L., Ohkawa, Y. & Imbalzano, A. N. Chromatin remodelling in mammalian differentiation: lessons from ATP-dependent remodellers. *Nat Rev Genet* **7**, 461–473 (2006).

199. Seo, S., Richardson, G. A. & Kroll, K. L. The SWI/SNF chromatin remodeling protein Brg1 is required for vertebrate neurogenesis and mediates transactivation of Ngn and NeuroD. *Development* **132**, 105–115 (2005).

200. Kosho, T. *et al.* Clinical correlations of mutations affecting six components of the SWI/SNF complex: Detailed description of 21 patients and a review of the literature. *American Journal of Medical Genetics Part A* **161**, 1221–1237 (2013).

201. Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).

202. Ghavi-Helm, Y. *et al.* Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**, 96–100 (2014).

203. Jerkovic´, I. & Cavalli, G. Understanding 3D genome organization by multidisciplinary methods. *Nat Rev Mol Cell Biol* **22**, 511–528 (2021).

204. Gibcus, J. H. & Dekker, J. The Hierarchy of the 3D Genome. *Molecular Cell* **49**, 773–782 (2013).

205. Rowley, M. J. *et al.* Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol Cell* **67**, 837-852.e7 (2017).

206. van Steensel, B. & Furlong, E. E. M. The role of transcription in shaping the spatial organization of the genome. *Nat Rev Mol Cell Biol* **20**, 327–337 (2019).

207. Rada-Iglesias, A., Grosveld, F. G. & Papantonis, A. Forces driving the three-dimensional folding of eukaryotic genomes. *Mol Syst Biol* **14**, e8214 (2018).

208. Oudelaar, A. M. & Higgs, D. R. The relationship between genome structure and function. *Nat Rev Genet* **22**, 154–168 (2021).

209. Palstra, R.-J. *et al.* The β-globin nuclear compartment in development and erythroid differentiation. *Nat Genet* **35**, 190–194 (2003).

210. Denholtz, M. *et al.* Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. *Cell Stem Cell* **13**, 602–616 (2013).

211. Stevens, T. J. *et al.* 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59–64 (2017).

212. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).

213. Zheng, M. *et al.* Multiplex chromatin interactions with single-molecule precision. *Nature* **566**, 558–562 (2019).

214. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**, 458–472 (2012).

215. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).

216. Boettiger, A. N. *et al.* Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **529**, 418–422 (2016).

217. Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**, 1665–1680 (2014).

218. Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion. *Cell Reports* **15**, 2038–2049 (2016).

219. Bauer, B. W. *et al.* Cohesin mediates DNA loop extrusion by a “swing and clamp” mechanism. *Cell* **184**, 5448-5464.e22 (2021).

220. Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944.e22 (2017).

221. Guo, Y. *et al.* CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* **162**, 900–910 (2015).

222. de Wit, E. *et al.* CTCF Binding Polarity Determines Chromatin Looping. *Molecular Cell* **60**, 676–684 (2015).

223. Lupiáñez, D. G. *et al.* Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* **161**, 1012–1025 (2015).

224. Schwarzer, W. *et al.* Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017).

225. Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320.e24 (2017).

226. Haarhuis, J. H. I. *et al.* The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell* **169**, 693-707.e14 (2017).

227. Won, H. *et al.* Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523–527 (2016).

228. Rajarajan, P. *et al.* Neuron-specific signatures in the chromosomal connectome associated with schizophrenia risk. *Science* **362**, eaat4311 (2018).

229. Song, M. *et al.* Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. *Nat Genet* **51**, 1252–1262 (2019).

230. Lu, L. *et al.* Robust Hi-C Maps of Enhancer-Promoter Interactions Reveal the Function of Non-coding Genome in Neural Development and Diseases. *Molecular Cell* **79**, 521-534.e15 (2020).

231. Stiles, J. & Jernigan, T. L. The Basics of Brain Development. *Neuropsychol Rev* **20**, 327–348 (2010).

232. Bae, B.-I., Jayaraman, D. & Walsh, C. A. Genetic Changes Shaping the Human Brain. *Developmental Cell* **32**, 423–434 (2015).

233. Taverna, E., Götz, M. & Huttner, W. B. The Cell Biology of Neurogenesis: Toward an Understanding of the Development and Evolution of the Neocortex. *Annu. Rev. Cell Dev. Biol.* **30**, 465–502 (2014).

234. Kostović, I. & Jovanov-Milosević, N. The development of cerebral connections during the first 20-45 weeks' gestation. *Semin Fetal Neonatal Med* **11**, 415–422 (2006).

235. Silbereis, J. C., Pochareddy, S., Zhu, Y., Li, M. & Sestan, N. The Cellular and Molecular Landscapes of the Developing Human Central Nervous System. *Neuron* **89**, 248–268 (2016).

236. Muñoz, W. A. & Trainor, P. A. Neural crest cell evolution: how and when did a neural crest cell become a neural crest cell. *Curr Top Dev Biol* **111**, 3–26 (2015).

237. Stiles, J. & Stiles, E. P. of C. S. J. *The Fundamentals of Brain Development: Integrating Nature and Nurture*. (Harvard University Press, 2008).

238. Weissman, T., Noctor, S. C., Clinton, B. K., Honig, L. S. & Kriegstein, A. R. Neurogenic radial glial cells in reptile, rodent and human: from mitosis to migration. *Cereb Cortex* **13**, 550–559 (2003).

239. Clancy, B., Darlington, R. B. & Finlay, B. L. Translating developmental time across mammalian species. *Neuroscience* **105**, 7–17 (2001).

240. Gazzaniga, M. S. *The Cognitive Neurosciences*. (MIT Press, 2009).

241. Bayer, S. A., Altman, J., Russo, R. J. & Zhang, X. Timetables of neurogenesis in the human brain based on experimentally determined patterns in the rat. *Neurotoxicology* **14**, 83–144 (1993).

242. Bishop, K. M., Rubenstein, J. L. R. & O'Leary, D. D. M. Distinct actions of Emx1, Emx2, and Pax6 in regulating the specification of areas in the developing neocortex. *J Neurosci* **22**, 7627–7638 (2002).

243. Rubenstein, J. L. R. Development of the Cerebral Cortex: Implications for Neurodevelopmental Disorders. *J Child Psychol Psychiatry* **52**, 339–355 (2011).

244. Mountcastle, V. B. The columnar organization of the neocortex. *Brain* **120**, 701–722 (1997).

245. Rakic, P. Specification of cerebral cortical areas. *Science* **241**, 170–176 (1988).

246. Hill, R. S. & Walsh, C. A. Molecular insights into human brain evolution. *Nature* **437**, 64–67 (2005).

247. Shimogori, T., Banuchi, V., Ng, H. Y., Strauss, J. B. & Grove, E. A. Embryonic signaling centers expressing BMP, WNT and FGF proteins interact to pattern the cerebral cortex. *Development* **131**, 5639–5647 (2004).

248. Sur, M. & Rubenstein, J. L. R. Patterning and Plasticity of the Cerebral Cortex. *Science* **310**, 805–810 (2005).

249. Grove, E. A. & Fukuchi-Shimogori, T. Generating the cerebral cortical area map. *Annu Rev Neurosci* **26**, 355–380 (2003).

250. Hoch, R. V., Rubenstein, J. L. R. & Pleasure, S. Genes and signaling events that establish regional patterning of the mammalian forebrain. *Semin Cell Dev Biol* **20**, 378–386 (2009).

251. Ohkubo, Y., Chiang, C. & Rubenstein, J. L. R. Coordinate regulation and synergistic actions of BMP4, SHH and FGF8 in the rostral prosencephalon regulate morphogenesis of the telencephalic and optic vesicles. *Neuroscience* **111**, 1–17 (2002).

252. Storm, E. E. *et al.* Dose-dependent functions of Fgf8 in regulating telencephalic patterning centers. *Development* **133**, 1831–1844 (2006).

253. Caronia-Brown, G., Yoshida, M., Gulden, F., Assimacopoulos, S. & Grove, E. A. The cortical hem regulates the size and patterning of neocortex. *Development* **141**, 2855–2865 (2014).

254. Goulburn, A. L., Stanley, E. G., Elefanty, A. G. & Anderson, S. A. Generating GABAergic cerebral cortical interneurons from mouse and human embryonic stem cells. *Stem Cell Res* **8**, 416–426 (2012).

255. Kwan, K. Y., Šestan, N. & Anton, E. S. Transcriptional co-regulation of neuronal migration and laminar identity in the neocortex. *Development* **139**, 1535–1546 (2012).

256. Manzini, M. C. & Walsh, C. A. What disorders of cortical development tell us about the cortex: one plus one doesn't always make two. *Curr Opin Genet Dev* **21**, 333–339 (2011).

257. Rakic, P. Evolution of the neocortex: a perspective from developmental biology. *Nat Rev Neurosci* **10**, 724–735 (2009).

258. Olson, E. C. & Walsh, C. A. Smooth, rough and upside-down neocortical development. *Curr Opin Genet Dev* **12**, 320–327 (2002).

259. Nadarajah, B. & Parnavelas, J. G. Modes of neuronal migration in the developing cerebral cortex. *Nat Rev Neurosci* **3**, 423–432 (2002).

260. Anderson, S. A., Marín, O., Horn, C., Jennings, K. & Rubenstein, J. L. Distinct cortical migrations from the medial and lateral ganglionic eminences. *Development* **128**, 353–363 (2001).

261. Malatesta, P., Appolloni, I. & Calzolari, F. Radial glia and neural stem cells. *Cell Tissue Res* **331**, 165–178 (2008).

262. Dehay, C., Kennedy, H. & Kosik, K. S. The outer subventricular zone and primate-specific cortical complexification. *Neuron* **85**, 683–694 (2015).

263. Noctor, S. C., Flint, A. C., Weissman, T. A., Dammerman, R. S. & Kriegstein, A. R. Neurons derived from radial glial cells establish radial units in neocortex. *Nature* **409**, 714–720 (2001).

264. Rakic, P. Mode of cell migration to the superficial layers of fetal monkey neocortex. *J Comp Neurol* **145**, 61–83 (1972).

265. Malatesta, P., Hartfuss, E. & Götz, M. Isolation of radial glial cells by fluorescent-activated cell sorting reveals a neuronal lineage. *Development* **127**, 5253–5263 (2000).

266. Noctor, S. C., Martínez-Cerdeño, V., Ivic, L. & Kriegstein, A. R. Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases. *Nat Neurosci* **7**, 136–144 (2004).

267. Hansen, D. V., Lui, J. H., Parker, P. R. L. & Kriegstein, A. R. Neurogenic radial glia in the outer subventricular zone of human neocortex. *Nature* **464**, 554–561 (2010).

268. Englund, C. *et al.* Pax6, Tbr2, and Tbr1 are expressed sequentially by radial glia, intermediate progenitor cells, and postmitotic neurons in developing neocortex. *J Neurosci* **25**, 247–251 (2005).

269. Baala, L. *et al.* Homozygous silencing of T-box transcription factor EOMES leads to microcephaly with polymicrogyria and corpus callosum agenesis. *Nat Genet* **39**, 454–456 (2007).

270. Lukaszewicz, A. *et al.* G1 phase regulation, area-specific cell cycle control, and cytoarchitectonics in the primate cortex. *Neuron* **47**, 353–364 (2005).

271. Bayatti, N. *et al.* A molecular neuroanatomical study of the developing human neocortex from 8 to 17 postconceptional weeks revealing the early differentiation of the subplate and subventricular zone. *Cereb Cortex* **18**, 1536–1548 (2008).

272. Mo, Z. & Zecevic, N. Is Pax6 critical for neurogenesis in the human fetal brain? *Cereb Cortex* **18**, 1455–1465 (2008).

273. Fietz, S. A. *et al.* OSVZ progenitors of human and ferret neocortex are epithelial-like and expand by integrin signaling. *Nat Neurosci* **13**, 690–699 (2010).

274. Lui, J. H., Hansen, D. V. & Kriegstein, A. R. Development and Evolution of the Human Neocortex. *Cell* **146**, 18–36 (2011).

275. Pollen, A. A. *et al.* Molecular Identity of Human Outer Radial Glia during Cortical Development. *Cell* **163**, 55–67 (2015).

276. Hevner, R. F. & Haydar, T. F. The (Not Necessarily) Convoluted Role of Basal Radial Glia in Cortical Neurogenesis. *Cereb Cortex* **22**, 465–468 (2012).

277. Smart, I. H. M., Dehay, C., Giroud, P., Berland, M. & Kennedy, H. Unique Morphological Features of the Proliferative Zones and Postmitotic Compartments of the Neural Epithelium Giving Rise to Striate and Extrastriate Cortex in the Monkey. *Cerebral Cortex* **12**, 37–53 (2002).

278. Lopez-Tobon, A., Caporale, N., Trattaro, S. & Testa, G. Three-dimensional models of human brain development. in *Stem Cell Epigenetics* 257–278 (Elsevier, 2020). doi:10.1016/B978-0-12-814085-7.00011-8.

279. Spaethling, J. M. *et al.* Primary Cell Culture of Live Neurosurgically Resected Aged Adult Human Brain Cells and Single Cell Transcriptomics. *Cell Rep* **18**, 791–803 (2017).

280. Yan, L. *et al.* Epigenomic Landscape of Human Fetal Brain, Heart, and Liver *. *Journal of Biological Chemistry* **291**, 4386–4398 (2016).

281. Human-Specific Histone Methylation Signatures at Transcription Start Sites in Prefrontal Neurons. https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001427.

282. Fischbach, G. D. & Lord, C. The Simons Simplex Collection: a resource for identification of autism genetic risk factors. *Neuron* **68**, 192–195 (2010).

283. Ninkovic, J. *et al.* The BAF Complex Interacts with Pax6 in Adult Neural Progenitors to Establish a Neurogenic Cross-Regulatory Transcriptional Network. *Cell Stem Cell* **13**, 403–418 (2013).

284. Fietz, S. A. *et al.* Transcriptomes of germinal zones of human and mouse fetal neocortex suggest a role of extracellular matrix in progenitor self-renewal. *PNAS* **109**, 11836–11841 (2012).

285. Mitchell, A. C. *et al.* The Genome in Three Dimensions: A New Frontier in Human Brain Research. *Biological Psychiatry* **75**, 961–969 (2014).

286. Chen, C., Kim, W.-Y. & Jiang, P. Humanized neuronal chimeric mouse brain generated by neonatally engrafted human iPSC-derived primitive neural progenitor cells. *JCI Insight* **1**, (2016).

287. Linaro, D. *et al.* Xenotransplanted Human Cortical Neurons Reveal Species-Specific Development and Functional Integration into Mouse Visual Circuits. *Neuron* **104**, 972-986.e6 (2019).

288. Hofrichter, M. *et al.* Comparative performance analysis of human iPSC-derived and primary neural progenitor cells (NPC) grown as neurospheres in vitro. *Stem Cell Research* **25**, 72–82 (2017).

289. Zeng, H. *et al.* Specification of Region-Specific Neurons Including Forebrain Glutamatergic Neurons from Human Induced Pluripotent Stem Cells. *PLOS ONE* **5**, e11853 (2010).

290. Fedorova, V. *et al.* Differentiation of neural rosettes from human pluripotent stem cells in vitro is sequentially regulated on a molecular level and accomplished by the mechanism reminiscent of secondary neurulation. *Stem Cell Research* **40**, 101563 (2019).

291. Li, M., Pevny, L., Lovell-Badge, R. & Smith, A. Generation of purified neural precursors from embryonic stem cells by lineage selection. *Current Biology* **8**, 971–974 (1998).

292. Elkabetz, Y. *et al.* Human ES cell-derived neural rosettes reveal a functionally distinct early neural stem cell stage. *Genes Dev.* **22**, 152–165 (2008).

293. Wilson, P. G. & Stice, S. S. Development and differentiation of neural rosettes derived from human embryonic stem cells. *Stem Cell Rev* **2**, 67–77 (2006).

294. Watanabe, K. *et al.* Directed differentiation of telencephalic precursors from embryonic stem cells. *Nat Neurosci* **8**, 288–296 (2005).

295. Chambers, S. M. *et al.* Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nat Biotechnol* **27**, 275–280 (2009).

296. Zhang, Y. *et al.* Rapid Single-Step Induction of Functional Neurons from Human Pluripotent Stem Cells. *Neuron* **78**, 785–798 (2013).

297. Qi, Y. *et al.* Combined small-molecule inhibition accelerates the derivation of functional cortical neurons from human pluripotent stem cells. *Nat Biotechnol* **35**, 154–163 (2017).

298. Lin, H.-C. *et al.* NGN2 induces diverse neuron types from human pluripotency. *Stem Cell Reports* **16**, 2118–2127 (2021).

299. Sato, T. *et al.* Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature* **459**, 262–265 (2009).

300. Di Lullo, E. & Kriegstein, A. R. The use of brain organoids to investigate neural development and disease. *Nat Rev Neurosci* **18**, 573–584 (2017).

301. Velasco, S. *et al.* Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. *Nature* **570**, 523–527 (2019).

302. Kelava, I. & Lancaster, M. A. Stem Cell Models of Human Brain Development. *Cell Stem Cell* **18**, 736–748 (2016).

303. Lancaster, M. A. *et al.* Guided self-organization and cortical plate formation in human brain organoids. *Nat Biotechnol* **35**, 659–666 (2017).

304. Cakir, B. *et al.* Engineering of human brain organoids with a functional vascular-like system. *Nat Methods* **16**, 1169–1175 (2019).

305. Eiraku, M. *et al.* Self-Organized Formation of Polarized Cortical Tissues from ESCs and Its Active Manipulation by Extrinsic Signals. *Cell Stem Cell* **3**, 519–532 (2008).

306. Lancaster, M. A. *et al.* Cerebral organoids model human brain development and microcephaly. *Nature* **501**, 373–379 (2013).

307. Paşca, A. M. *et al.* Functional cortical neurons and astrocytes from human pluripotent stem cells in 3D culture. *Nat Methods* **12**, 671–678 (2015).

308. 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).

309. Bhaduri, A. *et al.* Cell stress in cortical organoids impairs molecular subtype specification. *Nature* **578**, 142–148 (2020).

310. Luo, C. *et al.* Cerebral Organoids Recapitulate Epigenomic Signatures of the Human Fetal Brain. *Cell Reports* **17**, 3369–3384 (2016).

311. Camp, J. G. *et al.* Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. *PNAS* **112**, 15672–15677 (2015).

312. Amiri, A. *et al.* Transcriptome and epigenome landscape of human cortical development modeled in organoids. *Science* **362**, eaat6720 (2018).

313. Le Bail, R., Bonafina, A., Espuny-Camacho, I. & Nguyen, L. Learning about cell lineage, cellular diversity and evolution of the human brain through stem cell models. *Current Opinion in Neurobiology* **66**, 166–177 (2021).

314. Pollen, A. A. *et al.* Establishing Cerebral Organoids as Models of Human-Specific Brain Evolution. *Cell* **176**, 743-756.e17 (2019).

315. Mariani, J. *et al.* Modeling human cortical development in vitro using induced pluripotent stem cells. *PNAS* **109**, 12770–12775 (2012).

316. Lancaster, M. A. & Knoblich, J. A. Organogenesis in a dish: Modeling development and disease using organoid technologies. *Science* **345**, 1247125–1247125 (2014).

317. Mariani, J. *et al.* FOXG1-Dependent Dysregulation of GABA/Glutamate Neuron Differentiation in Autism Spectrum Disorders. *Cell* **162**, 375–390 (2015).

318. Marchetto, M. C. *et al.* Altered proliferation and networks in neural cells derived from idiopathic autistic individuals. *Mol Psychiatry* **22**, 820–835 (2017).

319. Schafer, S. T. *et al.* Pathological priming causes developmental gene network heterochronicity in autistic subject-derived neurons. *Nat Neurosci* **22**, 243–255 (2019).

320. Stachowiak, E. K. *et al.* Cerebral organoids reveal early cortical maldevelopment in schizophrenia—computational anatomy and genomics, role of FGFR1. *Transl Psychiatry* **7**, 1–24 (2017).

321. Helsmoortel, C. *et al.* A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP. *Nat Genet* **46**, 380–384 (2014).

322. Breen, M. S. *et al.* Episignatures Stratifying Helsmoortel-Van Der Aa Syndrome Show Modest Correlation with Phenotype. *The American Journal of Human Genetics* **107**, 555–563 (2020).

323. Zamostiano, R. *et al.* Cloning and Characterization of the Human Activity-dependent Neuroprotective Protein. *Journal of Biological Chemistry* **276**, 708–714 (2001).

324. Pinhasov, A. *et al.* Activity-dependent neuroprotective protein: a novel gene essential for brain formation. *Developmental Brain Research* **144**, 83–90 (2003).

325. Mandel, S., Rechavi, G. & Gozes, I. Activity-dependent neuroprotective protein (ADNP) differentially interacts with chromatin to regulate genes essential for embryogenesis. *Developmental Biology* **303**, 814–824 (2007).

326. Cappuyns, E., Huyghebaert, J., Vandeweyer, G. & Kooy, R. F. Mutations in ADNP affect expression and subcellular localization of the protein. *Cell Cycle* **17**, 1068–1075 (2018).

327. Ostapcuk, V. *et al.* Activity-dependent neuroprotective protein recruits HP1 and CHD4 to control lineage-specifying genes. *Nature* **557**, 739–743 (2018).

328. Kaaij, L. J. T., Mohn, F., van der Weide, R. H., de Wit, E. & Bühler, M. The ChAHP Complex Counteracts Chromatin Looping at CTCF Sites that Emerged from SINE Expansions in Mouse. *Cell* **178**, 1437-1451.e14 (2019).

329. Ferrari, R. *et al.* TFIIIC Binding to Alu Elements Controls Gene Expression via Chromatin Looping and Histone Acetylation. 13.

330. Germain, P.-L. & Testa, G. Taming Human Genetic Variability: Transcriptomic Meta-Analysis Guides the Experimental Design and Interpretation of iPSC-Based Disease Modeling. *Stem Cell Reports* **8**, 1784–1796 (2017).

331. Warren, L. *et al.* Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell* **7**, 618–630 (2010).

332. Pasi, C. E. *et al.* Genomic instability in induced stem cells. *Cell Death Differ* **18**, 745–753 (2011).

333. Menendez, L. *et al.* Directed differentiation of human pluripotent cells to neural crest stem cells. *Nat Protoc* **8**, 203–212 (2013).

334. Gaj, T. *et al.* Targeted gene knock-in by homology-directed genome editing using Cas9 ribonucleoprotein and AAV donor delivery. *Nucleic Acids Research* **45**, e98 (2017).

335. Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J.-S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Research* **24**, 1012–1019 (2014).

336. Liang, X. *et al.* Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *Journal of Biotechnology* **208**, 44–53 (2015).

337. Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* **3**, e04766 (2014).

338. Deininger, P. Alu elements: know the SINEs. *Genome Biol* **12**, 236 (2011).

339. Cordaux, R. & Batzer, M. A. The impact of retrotransposons on human genome evolution. *Nat Rev Genet* **10**, 691–703 (2009).

340. Dekker, J. The 4D nucleome project. 9.

341. Hutton, S. R. & Pevny, L. H. SOX2 expression levels distinguish between neural progenitor populations of the developing dorsal telencephalon. *Developmental Biology* **352**, 40–47 (2011).

342. Bernal, A. & Arranz, L. Nestin-expressing progenitor cells: function, identity and therapeutic implications. *Cell Mol Life Sci* **75**, 2177–2195 (2018).

343. El-Brolosy, M. A. *et al.* Genetic compensation triggered by mutant mRNA degradation. *Nature* **568**, 193–197 (2019).

344. Sun, X., Peng, X., Cao, Y., Zhou, Y. & Sun, Y. ADNP promotes neural differentiation by modulating Wnt/β-catenin signaling. *Nat Commun* **11**, 2984 (2020).

345. RNA polymerase II primes Polycomb-repressed developmental genes throughout terminal neuronal differentiation. *Molecular Systems Biology* **13**, 946 (2017).

346. Brookes, E. *et al.* Polycomb Associates Genome-wide with a Specific RNA Polymerase II Variant, and Regulates Metabolic Genes in ESCs. *Cell Stem Cell* **10**, 157–170 (2012).

347. Cruz-Molina, S. *et al.* PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. *Cell Stem Cell* **20**, 689-705.e9 (2017).

348. Bourque, G. *et al.* Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Research* **18**, 1752–1762 (2008).

349. Wang, J. *et al.* MIR retrotransposon sequences provide insulators to the human genome. *Proc Natl Acad Sci USA* **112**, E4428–E4437 (2015).

350. Cao, Y. *et al.* Widespread roles of enhancer-like transposable elements in cell identity and long-range genomic interactions. *Genome Res.* **29**, 40–52 (2019).

351. Su, M., Han, D., Boyd-Kirkup, J., Yu, X. & Han, J.-D. J. Evolution of Alu Elements toward Enhancers. *Cell Reports* **7**, 376–385 (2014).

352. Jjingo, D. *et al.* Mammalian-wide interspersed repeat (MIR)-derived enhancers and the regulation of human gene expression. *Mobile DNA* **5**, 14 (2014).

353. Misiak, B., Ricceri, L. & Sąsiadek, M. M. Transposable Elements and Their Epigenetic Regulation in Mental Disorders: Current Evidence in the Field. *Frontiers in Genetics* **10**, 580 (2019).

354. Huynh, M.-T. *et al.* A heterozygous microdeletion of 20q13.13 encompassing ADNP gene in a child with Helsmoortel–van der Aa syndrome. *Eur J Hum Genet* **26**, 1497–1501 (2018).

355. Petruzzi, M., Stella, A., Capra, V., Contaldo, M. & della Vella, F. Oro-Dental Manifestations in a Pediatric Patient Affected by Helsmoortel-Van der Aa Syndrome. *International Journal of Environmental Research and Public Health* **18**, 8957 (2021).

356. Blecher-Gonen, R. *et al.* High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. *Nat Protoc* **8**, 539–554 (2013).

357. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184–191 (2016).

358. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827–832 (2013).

359. D'Astolfo, D. S. *et al.* Efficient intracellular delivery of native proteins. *Cell* **161**, 674–690 (2015).

360. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**, 959–962 (2017).

361. Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).