PhD degree in Systems Medicine (curriculum in Molecular Oncology)

European School of Molecular Medicine (SEMM),

University of Milan and University of Naples "Federico II"

Settore disciplinare: MED/04

The developmental logic of enhanceropathies caused by KMT2D and YY1 haploinsufficiency through patient-derived disease modeling platforms

Michele Gabriele

IEO, Milan

Matricola n. R10727

Supervisor: Prof. Giuseppe Testa

IEO, Milan

Anno accademico 2017-2018

Contents

List of Abbreviations

abbreviation	definition
5caC	5-carboxylcytosine
5fC	5-formilcytosine
5hmC	5-hydroxymethylcytosine
5mC	in 5-methyl-Cytosine
AAV	adeno-associated virus
ChIP-seq	chromatin immunoprecipitation coupled with sequencing
COMPASS	COMplex of Proteins ASsociated with Set1
CTD	Carboxyterminal domain
DEA	Differential expression analyses
DEG	Differentially expressed gene
DMR	Differentially methylated region
DNA	Deoxyribonucleic Acid
EBV	Epstein Barr virus
ePB	ePiggyBac
eRNA	enhancer RNA
GADEVS	Gabriele-De Vries syndrome
GO	Gene Ontology
H3K27Ac	Histone 3 lysine 27 Acetylation
H3K27me3	Histone 3 lysine 27 tri-methylation
H3K4me1	Histone 3 lysine 4 mono-methylation
H3K4me1	Histone 3 lysine 4 di-methylation
H3K4me1	Histone 3 lysine 4 tri-methylation
iNeuron	neurons obtained by Ngn2 overexpression
IPA	ingenuity pathway analysis
KDa	Kilo Dalton
KO	Knock-out
KS	Kabuki Syndrome
mRNA	messenger RNA
NCSC	neural crest stem cells
ND	Neurodevelopmental disoderd
NMD	Nonsense mediated decay
NTD	Aminoterminal domain
PCA	Principal component analysis
PcG	Polycomb Group
PRC2	Polycomb Repressive Complex 2
PRE	Polycomb Responsive Elements
PTC	premature termination codon
PTM	post-translational modification
RNA	Ribonucleic Acid
RPKM	Reads Per Kilobase Million
TF	Transcription factor
Trl	Trithorax-related

Trx   Trithorax
VEE   Venezuelan Equine Encephalopathy virus

Index of figures and tables

ABSTRACT

The human brain is generated by developmental events orchestrated by fine-tuned transcriptional events. The vulnerability of this process is highlighted by the fact that environmental insults in early stages of development, or germline mutations in chromatin remodelers, often lead to neurodevelopmental disorders. An historical challenge for studying neurodevelopmental disorders has been the lack of reliable human platforms. Nowadays, technological advancements allowed us to use human cellular models to study the neurons and the molecular implications of specific mutations. In this work is described a human modeling platform of Kabuki Syndrome, caused by germline mutations in the enzyme KMT2D, which is involved in enhancer activation, and of Gabriele-de Vries syndrome, caused by mutations in YY1, which mediates DNA looping between gene promoters and enhancers. These two diseases thus belong to the class of enhanceropathies. Using induced pluripotent stem cells and their differentiation in disease-relevant cell types such as neural crest stem cells and cortical excitatory neurons, it was possible to identify molecular and functional phenotypes unique to Kabuki syndrome that contribute to the understanding of its molecular pathogenesis. Moreover, starting from genetic observations of individuals with intellectual disabilities who lacked a molecular diagnosis, YY1 was identified as the mutated gene responsible for a novel neurodevelopmental disorder, which is caused by an improper regulation of transcriptional enhancers. The synergic study of enhanceropathies may lead to the identification of altered gene regulatory networks, convergent to common targets, responsible for shared clinical features, thus paving the way for the identification of shared actionable pathways.

INTRODUCTION

Chromatin Structure and Epigenetics

The genetic information is transmitted by the deoxyribonucleic acid (DNA) (Avery et al., 1944; Hershey and Chase, 1952), which in living beings is composed by the deoxyribonucleosides of the nitrogenous bases guanine (G) , cytosine (C), adenine (A), and thymine (T). DNA has a negative double helix structure, in which bases of the opposite strand interacts with hydrogen bonds. Guanine interacts with cytosine, and adenine with thymine by three and two hydrogen bonds, respectively (Watson and Crick, 1953).

Genetic information can also be propagated from mother cell to daughter cell and from parents to offspring without changing the sequence of DNA, by the set of phenomena included in the term epigenetics. The term epigenetics (from the greek επὶ, epì = "above" and γεννετικός, gennetikòs = "genetics") was originally used by Conrad Waddington in 1942, to describe the *"causal interactions between genes and their products which bring the phenotype into being"* (Waddington, 1942). Nowadays, a variety of meanings are included in this term:

1) The set of phenomena responsible for the inheritance of a phenotype across generations (of a cell or individuals) without changes in nucleic acid sequence (Boniolo and Testa, 2012; Meloni and Testa, 2014);

2) The mechanism by which a phenotype is stably maintained across the lifespan of an organism, including non-dividing cells (Beck et al., 2010).

3) The use as synonymous of gene regulation and chromatin modifications ("On the use of the word 'epigenetic,'" 2007)

In all the three definitions, the mechanism by which epigenetics is implemented is through DNA covalent modifications and/or by chemical modifications of proteins strictly associated with the DNA itself, such as histones. These molecular modifications are able to influence a phenotype, without varying the genotype, by influencing gene expression.

There is a variety of DNA covalent modification, in animals the most relevant and studied one is the cytosine methylation in 5-methyl-Cytosine (5mC) (Aravind et al., 2014). This modification has the peculiarity of not changing the pairing with G, consequently it does not introduce mutations when propagated through cell divisions, and influences gene expression. Cytosine is methylated by two *de novo* enzymes (DNMT3A and DNMT3B) and a maintenance methylase (DNMT1). 5mC cannot be directly removed. Indeed it can be diluted passively following cell division or it can be sequentially enzymatically oxidized to 5-hydroxymethylcytosine (5hmC) and to 5-formilcytosine (5fC) and 5-carboxylcytosine (5caC), which can be removed by the base excision repair (Jin and Liu, 2018). In eukaryotes, DNA resides in nuclei in chromatin: a molecular complex between DNA and proteins. The core unit of chromatin is the nucleosome, constituted by DNA wrapped in basic proteins called histones. The nucleosome includes ~147 bp, folded in 1.7 turns, and an octamer of histones composed by a pair of histones 3 (H3), histone 4 (H4), histone 2A (H2A), and histone 2B (H2B) (Luger et al., 1997). Humans bear numerous histones isoforms, of them H2A is the most variable, with 19 genes; H2B has 16 variants; H3 has 8 variants; H4 is encoded in multiple clusters but does not possess variants (Kamakaka and Biggins, 2005; Talbert and Henikoff, 2010). Post-translational modification of histones, as well as their isoforms, are involved in any cellular process such as gene expression regulation, DNA repair, cell cycle, recombination, apoptosis, and cell division (Lawrence et al., 2016). A linker histone, the histone 1 (H1), is responsible for stabilizing the DNA around the nucleosome, promotes the

folding in a higher order of chromatin structure, and regulates cell cycle and gene expression, as well (Cutter and Hayes, 2015).

Transcriptional Enhancers

Since the discovery of the Lac operon in *Escherichia coli*, it has become common knowledge that non-coding DNA sequences can directly regulate gene expression (Jacob and Monod, 1961). In prokaryotes, these sequences are located near the transcription start site (TSS) and are a binding site for activators or repressors, which quickly change gene expression as a consequence of the replacement of a nutrient, or to respond to environmental stimuli in each organism. In metazoans, transcription regulation ensures gene expression in a specific tissue at a specific developmental window, with the final result of the organization of a complex organism derived from a single cell. This time-space regulation is mainly mediated by transcriptional enhancers, which are short DNA sequences (~100–1,000 bp) located in the same chromosome (*in cis*) of the target genes and are positioned at any distance from them, from close proximity up to megabases of distance (Long et al., 2016). Enhancers were identified and named in 1981, when the transient expression of rabbit β-globin gene in HeLa was massively increased if linked with a 72bp repeated region of the early genes of SV40, independently from its distance or position with respect to the gene (Banerji et al., 1981). Transcriptional enhancers have a pivotal role in the multi-step highly regulated gene expression regulation, which is accompanied by chromatin reorganization, DNA modifications, and histone post-translational modifications (HPTMs). Given the heterogeneous nature and sparse distribution of enhancers, their direct identification is difficult. Using chromatin immunoprecipitation coupled with sequencing (ChIP-seq), histone modifications can be used to identify enhancers, their activation status and to distinguish them from promoters (ENCODE Project Consortium, 2012; Heintzman

et al., 2007; Visel et al., 2009). The presence of histone 3 lysine 4 monomethylation (H3K4me1) identifies the presence of enhancers, which can be activated, when marked by histone 3 lysine 27 acetylation (H3K27ac), or poised if marked by histone 3 lysine 27 trimethylation (H3K27me3) (Rada-Iglesias et al., 2011). Enhancers physically contact target promoters in a loop structure, that once is formed can be inherited during cell division and differentiation, to ensure that the target promoter will start and maintain the gene transcription only in the correct cell type and at right time(Ghavi-Helm et al., 2014; Jost et al., 2014). The loop structure is the result of the interaction between multiple structural proteins, such as cohesin, the mediator complex, YY1, histone modifiers, and chromatin remodelers (Hu and Tee, 2017; Weintraub et al., 2017), as depicted in

Figure 1. The important similarity between enhancer and promoters is even underscored by the capability of the RNA pol2 associated with enhancers to bidirectionally transcribed them, producing what are called enhancer RNA (eRNA), which are associated with enhancer activity (Tippens et al., 2018). These transcripts bind YY1 and have been shown to increase its recruitment at the enhancer site (Sigova et al., 2015). Nevertheless, they are not needed to sustain the transcription of target genes and are not accumulated (Rahman et al., 2017). While histone modifications are a reliable signal for the presence of enhancers, and their activation status, they do not reveal about the position of the target genes. For this reason, the identification of target genes is more difficult, primarily because enhancers can interact with multiple genes and are located anywhere in the same chromosome. Nowadays, next-generation technologies (NGS) have become helpful to develop sequencing techniques to identify interactions between enhancers and target promoters by chromosome conformation capture techniques and their modifications, such as HiC or HiChIP, to identify the contacts between the genomic target of a protein of interest and its distal targets (Davies et al., 2017; Mumbach et

al., 2016). Once identified, enhancers need functional validation. Classically, episomal reporters used were, but they fail to reproduce the appropriate chromatin context and they activate the INF-I response, for this reason, several techniques have been developed during years to increase the accuracy and reliability of such assays and to study such interactions in the native chromatin context (Fulco et al., 2016; Inoue et al., 2017; Muerdter et al., 2018).
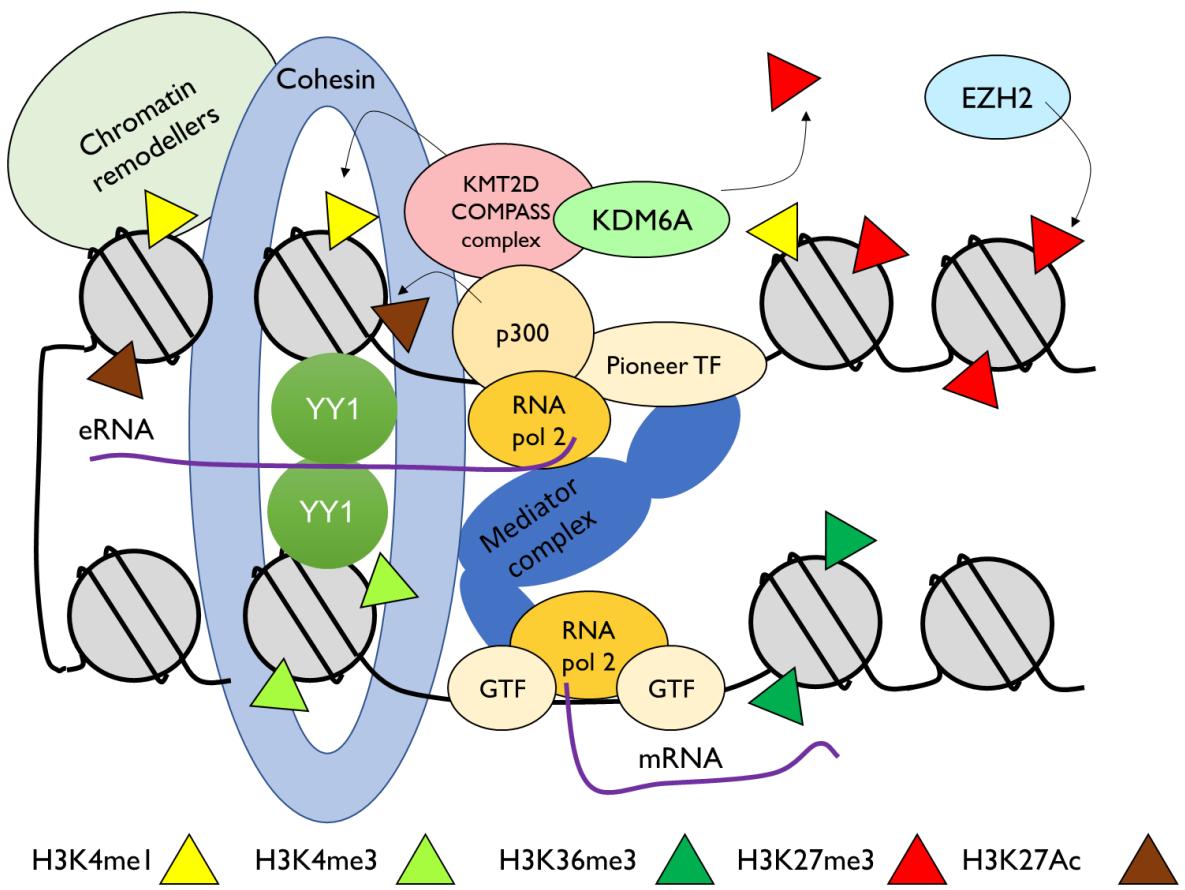
Figure 1 GTF: general transcription factor; TF: transcription factors; eRNA: enhancer RNA; mRNA: messenger RNA

The Early Development

The human development starts with the fusion of the two maternal and paternal gametes to form the zygote, the only totipotent cell, that it undergoes asynchronous sequential mitotic divisions. These cells, the blastomeres, divide for four days (E4) during which at 8-cell stage undergo compaction, a process by which they connect each other's by tight gap junctions and they separate from the zona pellucida, further cell division will form the 16-cells stage called morula. The outer cells give rise to the trophoblast cells that at E5, during the cavitation process, secrete liquid inside forming the cavity called blastocele, characteristic of the new embryonic structure called blastocyst. Here, the trophoblast cells are clearly separated to the inner cell mass (ICM) cells, which are pluripotent. The trophoblast cells will give rise to the chorion and the embryonic portion of the placenta. Differently, the ICM cells will give rise to the embryo, yolk sac, allantois, and amnion. This first commitment step is regulated by the mutually exclusive expression of NANOG and OCT4 in the ICM, and CDX2 in the trophectoderm (Ralston and Rossant, 2005). After the implantation, at about E9 there is a second commitment event during which the ICM divides into the epiblast and the hypoblast. The epiblast cells are the last pluripotent cells: they will give rise to the three germ layers (endoderm, mesoderm, ectoderm), and to the amniotic ectoderm; the hypoblast will give rise to the yolk sac. During this commitment phase, the hypoblast express GATA6 that repress the self-sustaining pluripotency core network composed by NANOG, OCT4, and SOX2 (Boyer et al., 2005) that are still expressed in the epiblast (Li and Belmonte, 2017). Between day E13 and E20 the embryo undergoes gastrulation, the process that the developmental biologist Lewis Wolpert defined as the most critical moment of life "*It is not birth, marriage or death, but gastrulation which is truly the most*

important time in your life"[1]. During this process, the three embryonic layers are formed by cell migration across the primitive streak and towards the rostral part of the embryo.

The Neural Crest Stem Cells

After the end of the third gestational week (3 GW) the embryo finishes the gastrulation and starts the process by which develops the central nervous system (CNS). A portion the dorsal mesoderm sends signals to the dorsal ectoderm which elongates its cells in columnar cells, forming the neural plate, which is negative for the pluripotency factors NANOG and OCT4 but still express SOX2, which is necessary to start neuronal differentiation and for neuronal progenitors maintenance (Zhang and Cui, 2014), and start to express PAX6 (Osumi et al., 2008). At E21 it starts the neurulation, the process during which the neural ridges (the regions that encompass the neural plate laterally) rise, fold and eventually fuse forming the neural tube. The closure starts in the middle region of the embryo and continues towards both the rostral and caudal region. The neural tube that is located above the notochord and below the epidermal ectoderm, to which exchanges morphogenetic signals. From this region, the neural crest stem cells (NCSC) are formed and start to migrate ventrally in multiple regions of the embryo according to their rostral-caudal positioning. The NCSC will give rise to a multiplicity of tissues and organs, for this reason, they are sometimes referred to as the "fourth embryonic layer" (Dupin et al., 2018). NCSC give rise to every component of the peripheric nervous system, the enteric nervous system, cranial nerves, and melanocytes. The cranial neural crests give rise to the anterior facial cartilage and bones, adrenal

---

[1] From Egg to Embryo: Determinative Events in Early Development (1999), Ch. 1 : Regional specification in animal development

gland. The cardiac NC give rise to the muscle-connective tissue wall of the large arteries and the septum, responsible for the separations of the pulmonary circulation from the aorta. Also, NCSCs contribute to teeth formation, thymus, parathyroid, and thyroid glands (Vega-Lopez et al., 2018) as summarized in Figure 2. The class of diseases in which neural crest are affected is called neurocristopathies and, given the anatomical contribution of the NC, it is not surprising to find recurrent clinical manifestations such as craniofacial abnormalities, neurological, cardiac, and immunological dysfunction in such disease category.
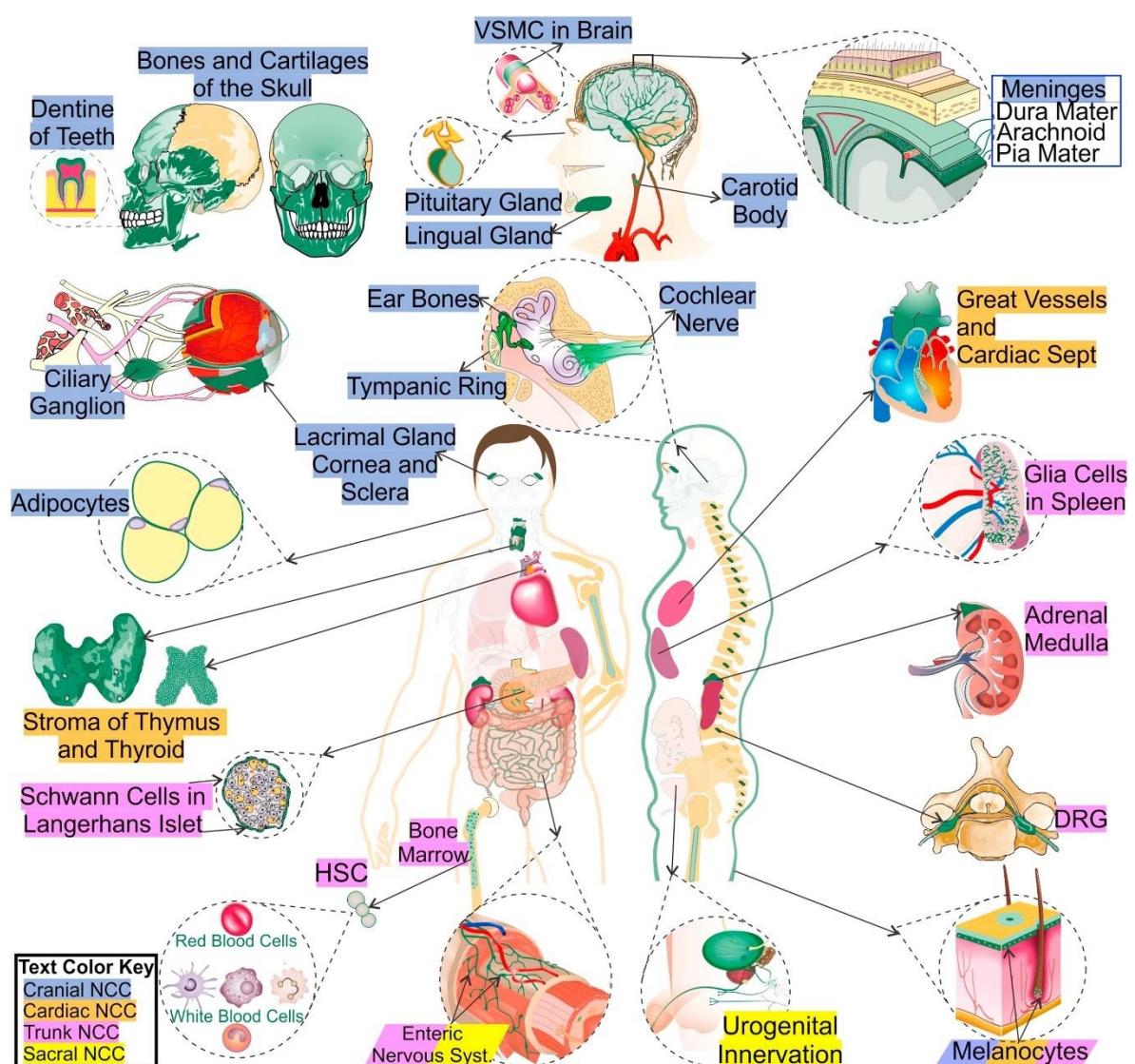
Figure 2 summary rapresentation of neural crest multipotency, adapted from Vega-Lopez et al., 2018

Central nervous system development

After the completion of the neural tube, the neuronal progenitors form a single layer surrounding the hollow center. The rostral progenitors will give rise to the brain, while the neuronal progenitors in the caudal neural tube will give rise to the hindbrain and the spinal cord. This cavity will become the brain ventricular system, for this reason, this region is now called ventricular zone (VZ). Before the neural tube closure, the anterior end starts to develop the three brain vesicles prosencephalon, the mesencephalon, and rhombencephalon, which will form the forebrain, the midbrain, and the hindbrain, respectively, as depicted in Figure 3. At this stage, PAX6 is expressed only in forebrain, hindbrain, and spinal cord (Osumi et al., 2008), while OTX2, another important transcription factor that starts to be expressed at morula stage, is now expressed only in the forebrain and midbrain (Beby and Lamonerie, 2013).
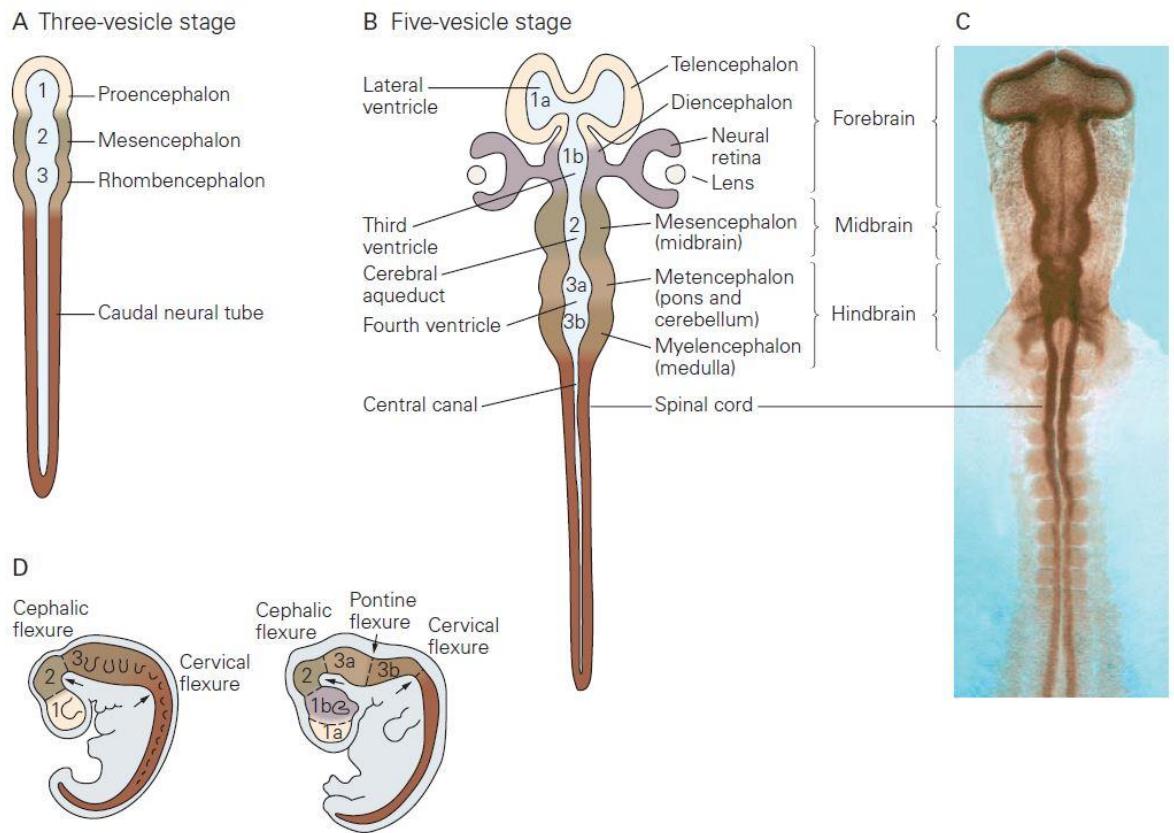
Figure 3 Stages of central nervous system development from the neural tube. Adapted from Kandel – Principles of Neural Science 5th edition.

The main changes in the brain morphology happen within the fetal period (e.g period between GW9 to the end of gestation) during which gyri and sulci are formed. During this period the neural progenitors expand their pool and differentiate in neurons. In particular, most of the neurons of the adult brain is formed by mid-gestation (the time span that includes GW18-24). Neural progenitors divide symmetrically starting from the end of gastrulation (~ GW3) to GW6. Afterward, there is a switch to asymmetrical divisions, by which each progenitor gives rise to a neuronal progenitor and a post-mitotic neuron.

The neocortex development

The neocortex is the part of the CNS responsible for behavior and cognition; it is composed of six stratified layers of organized neurons, generated from proliferative waves of migrating neurons in contact to the VZ. The neocortex is indeed one of the most complex areas of the human brain, and it is remarkably different from other specie ones. This structure was first identified in reptiles during the Carboniferous Period, and it has increased in size and connectivity complexity during evolution. There are important anatomical and functional differences between the human neocortex and the most studied mammalian model *Mus musculus* one, especially in the prefrontal cortex structure, which constitutes almost one-third of the neocortex and it is responsible for complex cognitive tasks (Carlén, 2017). The neocortex is composed of neurons that migrate from the proliferative zone in the VZ towards the pial surface, generated in sequential migratory waves. The first migratory neurons contact the pial surface and generate the preplate (PP). The second migratory wave splits the PP into the subplate (SP) and the marginal zone (MZ), which contain the Cajal-Retzius cells (CR). The CR cells produce Reelin, which controls the positioning of migrating neurons stopping their migration. Consequently, every following neuronal migratory wave will get through the most upper layer taking its place in the developing neocortex. For this reason, apart from the very first migratory wave, the neocortex possesses an inside-out structure in which the oldest neurons are the closest to the VZ. The cell called "radial glial guides" was initially thought to support neuronal migration from the VZ to the pial surface thanks to their long process extended from the VZ to the pial surface (Rakic, 1972), was later revealed to be the asymmetrical neuronal progenitor, which possesses the ability of replicating both in the VZ and in the sub-ventricular zone (SVZ) (Noctor et al., 2004, 2001). The neocortex structure with developmental stages is represented in Figure 4.
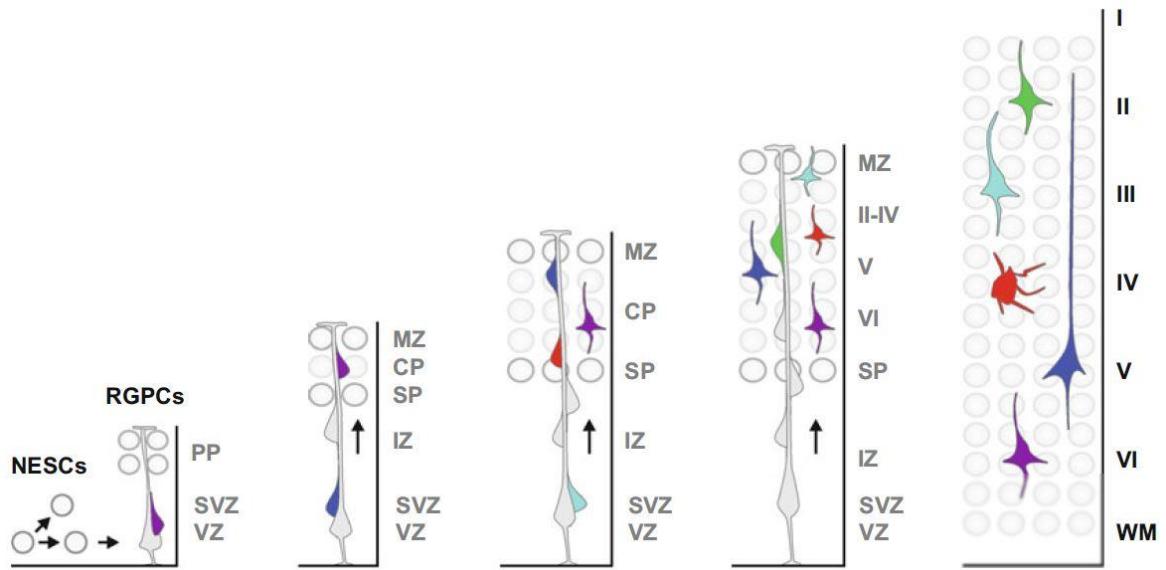
Figure 4 Representation of the inside-out cortical layer development. NESC: neuroepithelial stem cells, divide symmetrically; RGPC: radial glia progenitor cells, divide asymmetrically. PP: preplate; VZ: ventricular zone; SVZ sub-ventricular zone; MZ: marginal zone; CP: cortical plate; SP: subplate; IZ:intermediate zone; WM: white matter; I-VI: cortical layers 1-6. Color code represent proliferative and migratory waves. Adapted from Nguyen and Hippenmeyer – Cellular and Molecular Control of Neuronal Migration, Springer, 2014.

The final neocortex is not only formed by neurons originated from the described proliferative zone, indeed, it is also composed by the interneurons that tangentially migrate to the neocortex from a proliferative region in the lateral and medial ganglionic eminences, comprised between the VZ and the striatum (Stiles and Jernigan, 2010). The transcription factors Ngn1 and Ngn2 (*NEUROG2*, Neurogenin-2) have a critical role in the differentiation of the glutamatergic cortical neurons, and they represses the GABAergic phenotype (Schuurmans et al., 2004). Figure 5 summarizes the developmental origin of cortical excitatory neurons and GABAergic neurons.

.
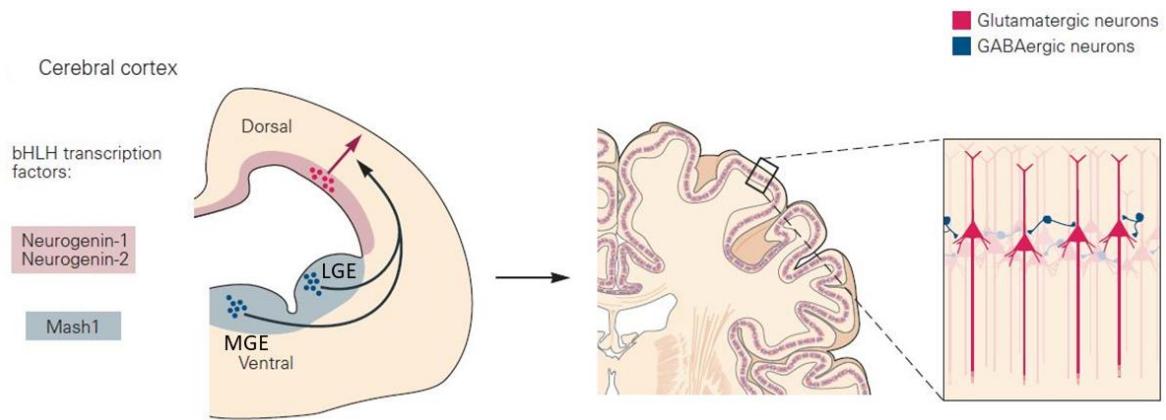
Figure 5 Transcription factors involved in the generation of glutamatergic or GABAergic interneurons. MGE: medial ganglionic eminence; LGE: lateral ganglionic eminence. Adapted from Kandel – Principles of Neural Science 5th edition.

Neurodevelopmental disorders and chromatin biology dysregulation

Neurodevelopmental disorders (NDDs) are a vast class of early onset neurological diseases that encompass intellectual disabilities (ID), autism spectrum disorders (ASD), attention-deficit/hyperactivity disorder (ADHD), schizophrenia, bipolar disorder, learning disabilities, and major depressive disorder. NDDs are caused both by environmental factors (such us alcohol consumption during pregnancy, causing the fetal alcohol syndrome) and genetic factors. It is difficult to estimate a precise incidence and prevalence of NDDs, especially concerning IDs, since it is identified by IQ tests, which are often heterogeneous and not accurate (Ropers, 2010). It has been estimated that *de novo* mutations in coding sequence explain 42% of cases of severe NDDs with a prevalence mean of 1 in 295 birth caused by dominant monogenic *de novo* mutations in the coding sequence of genes, of which 59% operate by loss of function and 41% by altered function (Deciphering Developmental Disorders Study, 2017). Concerning *de novo* single nucleotide variants (SNVs) in non-coding elements such as fetal brain-active elements and highly conserved

elements, it has been estimated that they account for 1-3% and 0,5% of NDDs, respectively (Short et al., 2018). It then emerges the importance of regulatory elements for the molecular pathogenesis of these diseases. Indeed, taking in account the list of 910 genes responsible for ASDs (January 2018), provided by the Simons Foundation Autism Research Initiative (SFARI)[2], it emerges the gene ontology (GO) enrichment "chromatin binding proteins", preceded by three categories of molecular functions involved in neuronal biology, as shown in Figure 6 (Gabriele et al., 2018). At the current state of knowledge, it is known that germline mutations in genes coding for chromatin remodelers, histone post-translational modifiers, and DNA methylases causes NDDs (Figure 7). To highlight the importance of these genes and of transcriptional fine regulation in the onset of NDDs it is possible to observe their spatial and temporal expression pattern in the brain. The region with the highest expression pattern of these genes is the brain cortex, in particular in the Brodmann area 9, located in the prefrontal cortex, known to be involved high cognitive functions and frequently disrupted in several NDDs. The temporal expression pattern further underscores their importance for neurodevelopment. Most of them show an increased expression pattern during the gestational period (e.g. the timespan between conception and GW8) and the fetal period (e.g period between GW9 to the end of gestation, and comprises the midgestation, which include GW18-24), which are thought to be the most susceptible developmental stage for neocortex development Figure 8 (Gabriele et al., 2018; Stiles and Jernigan, 2010).

---

[2] https://www.sfari.org/resource/sfari-gene/
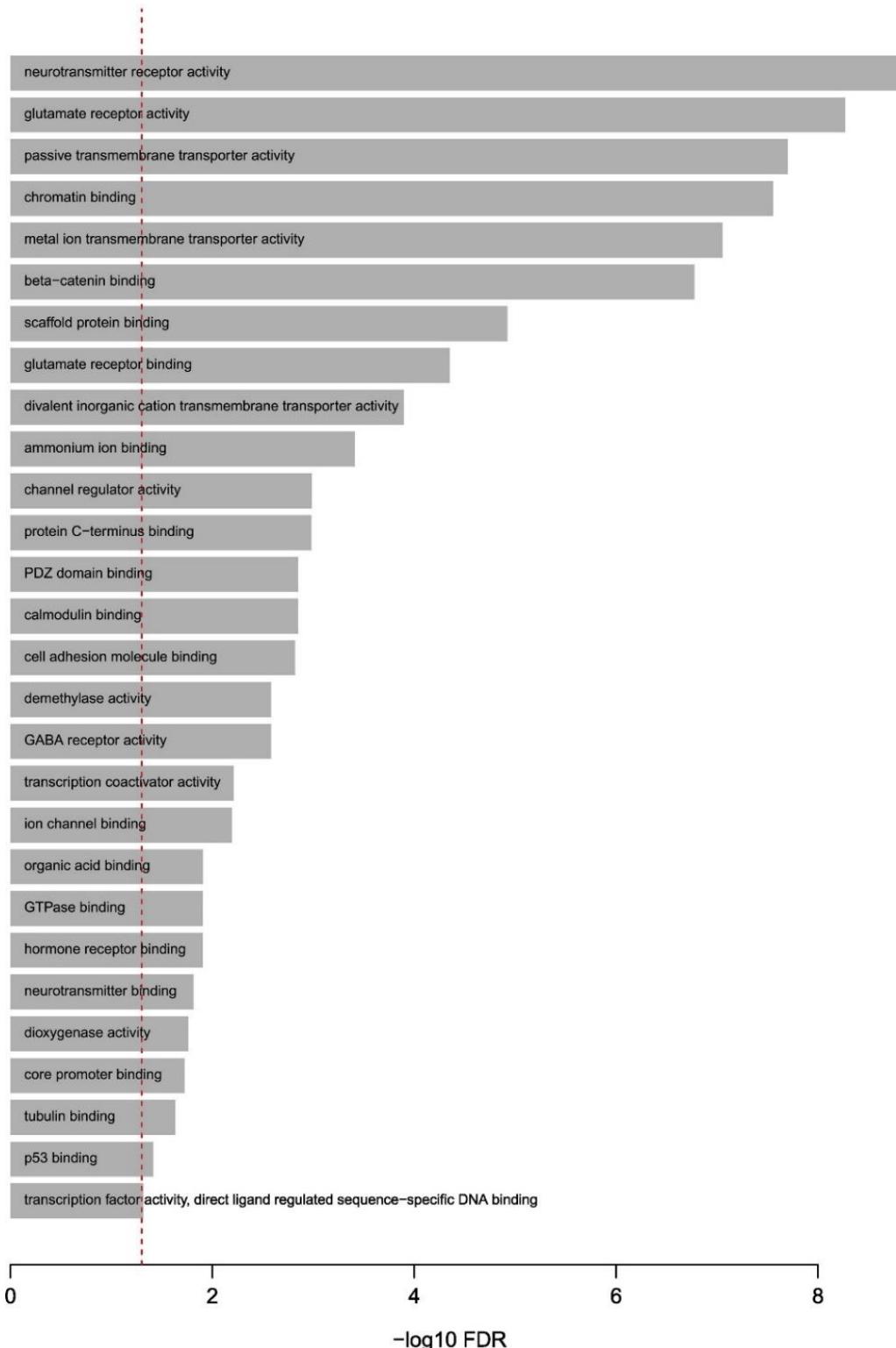
Figure 6 Gene Ontology analysis of SFARI genes responsible for ASDs, adapted from Gabriele et al., 2018
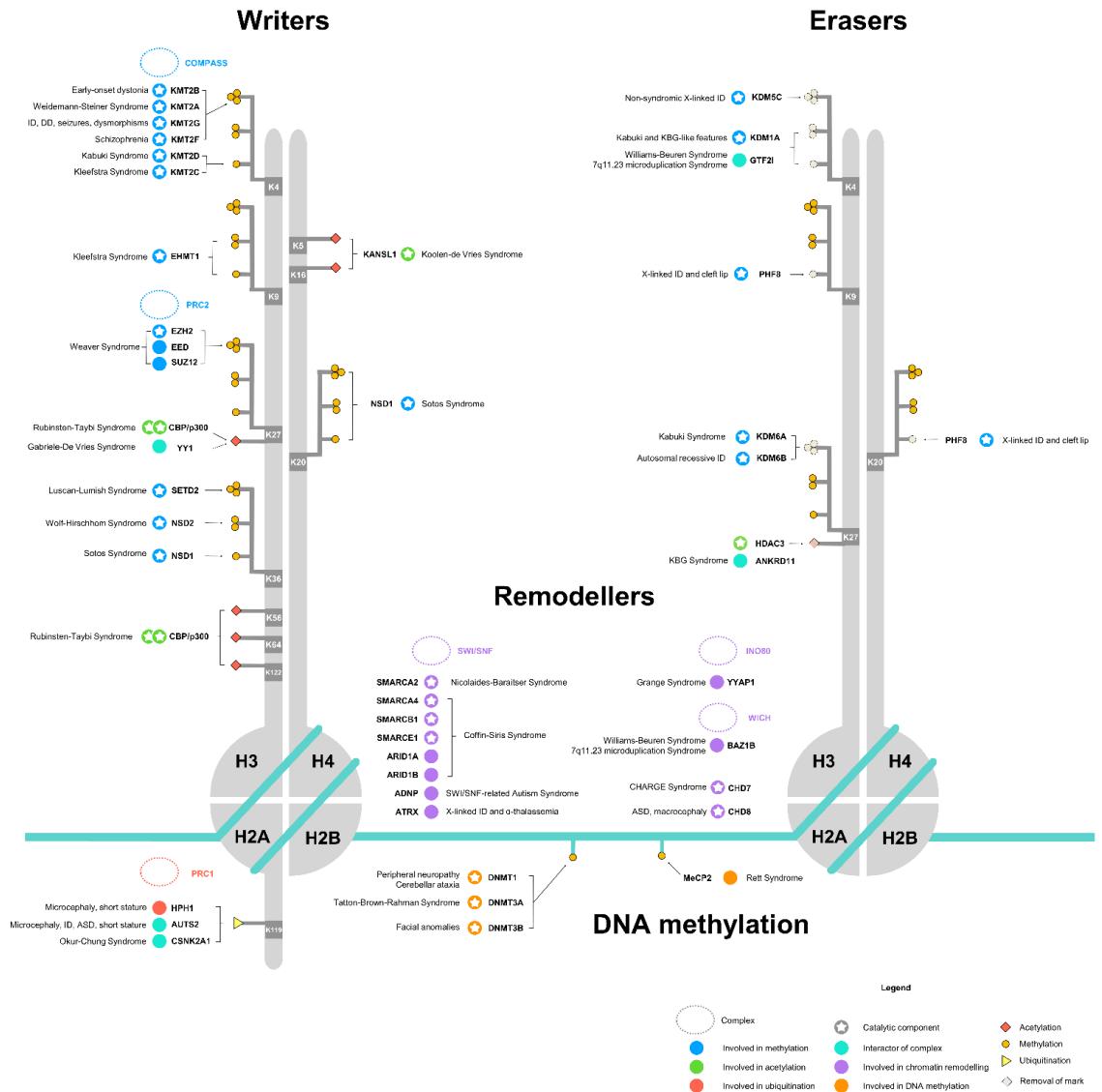
Figure 7 Representation of several genes coding for protein involved in chromatin remodelling, DNA methylation, and histone post-translation modifications and relative syndromes caused by germline mutations. Adapted from Gabriele et al., 2018
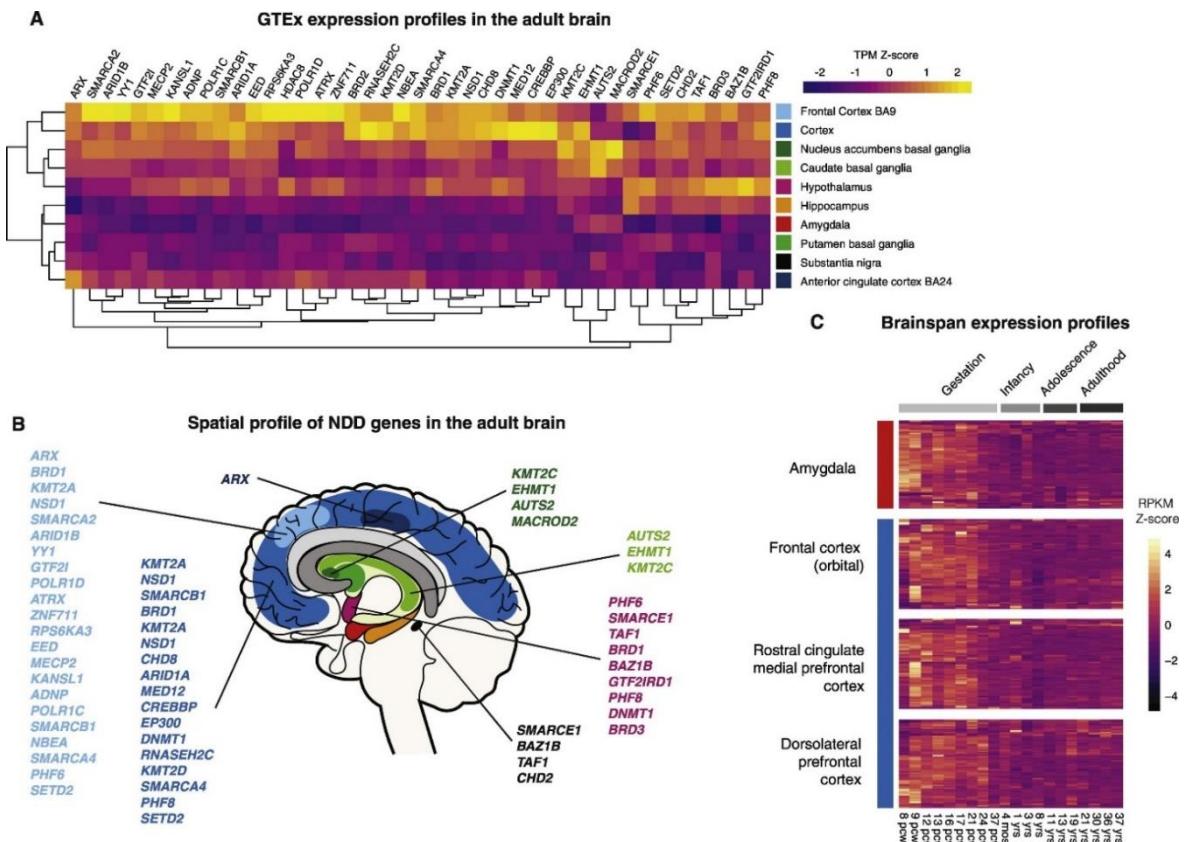
Figure 8 A) Spatial expression pattern according to the GTEx brain tissues; B) representation of spatial expression pattern; C) Temporal expression pattern according to the BrainSpan atlas. Adapted from Gabriele et al., 2018

Disease Modelling and Induced Pluripotent Stem Cells

The study of NDDs has the significant limit of lacking a proper model to study their molecular pathogenesis. As described above, the brain is a sophisticated and complex organ, several aspects of its functionality such as the processes of decision making, the molecular basis of learning and memory, or how the neuronal circuits elaborate information are still needed to be addressed (Adolphs, 2015). A reason for this limit is that some peculiarities are uniquely present in humans, but they are studied in model organisms, as surrogates, for the lack of availability of a human model, for obvious ethical reasons. Direct access to the human brain and its use as specimen is possible only using post-mortem samples. Even to test the existence of human adult neural stem cells, researchers had to used post-mortem tissues, taken from cancer patients treated with BrdU for diagnostic purposes (Eriksson et al.,

17

1998). Especially for the study of developmental diseases, it is obvious the impossibility of experimenting in humans. However, it is important to address certain questions in the right cell type of the most appropriate organism.

Nowadays, thanks to the isolation of embryonic stem cells (ESC) from the ICM of the embryonic blastocyst, both from mouse and human (Evans and Kaufman, 1981; Thomson et al., 1998), it is virtually possible, using the knowledge coming from decades of developmental biology, to differentiate ESC in any cell type present in an organism (Vazin and Freed, 2010). For this reason, stem cell biology permitted to reproduce *in vitro* human developmental processes that were not possible to study before. It is now possible to study the onset and molecular pathology of diseases for which a model was not present (Craft and Johnson, 2017). Nevertheless, the use of human ESC raises the ethical issue of the sacrifice of a human embryo, a possible future individual. Moreover, given the limited availability of hESC clones, strictly connected to the ethical concerns of deriving them, while they are a good model for basic molecular studies, the lack of biological replicates becomes a problem for drug screening studies, in which individual genotypes can interact with drug metabolism. In addition, the clinical potential of hESC is impaired by the immunological response of the receiver, given the differences in the immunohistocompatibility complexes. Another issue of hESC is that they represent a human being that was never formed. Consequently, any information regarding phenotypes is not possible to be inferred (Ilic and Ogilvie, 2017). The solution to all these problems was obtained by Shinya Yamanka, when in 2006 identified the cocktail of transcription factors Oct3/4, Sox2, Klf4, and c-Myc (Takahashi and Yamanaka, 2006) able to reprogram adult somatic cells back to induced pluripotent stem cells (iPSCs) with the property of contributing to all the three embryonic layers and germline of mice, if positive for Nanog (Okita et al., 2007). A short time was needed to replicate the result also in human (Takahashi et al., 2007). The first

generations of iPSC were not devoid of problems, such as the use of integrative vectors for the delivery of the expression cassette and the use of the oncogene c-Myc, which confers oncogenic potential (Hyun et al., 2007). Research has overcome these problems by using non-integrative methods, as the use of the self-replicating RNA viruses Sendai virus (Fusaki et al., 2009) and the backbone of the Venezuelan Equine Encephalitis (VEE) (Yoshioka et al., 2013) to solve the integration problems, and by substituting the oncogene c-Myc by the maternal factor GLIS1 (Nakagawa et al., 2008).

Kabuki Syndrome

Kabuki syndrome (KS) is a rare multisystemic neurodevelopmental disorder first described in 1981 (Kuroki et al., 1981; Niikawa et al., 1981). KS prevalence is estimated to be 1/33000 (Adam and Hudgins, 2005). It is a dominant, autosomal genetic disease, and a heterogeneous clinical phenotype characterizes it.

KS clinical phenotype features characteristic facial dimorphisms, which encompass long palpebral fissures with eversion of the eyelid, arched eyebrows, a short columella with a depressed nasal tip and prominent ears. Moreover, KS is characterized by mild to moderate mental retardation. Also, KS includes postnatal growth retardation and short stature, skeletal abnormalities, fetal fingerpads, cleft lip and/or palate. Additionally, KS encompasses functional abnormalities such as seizures, hearing loss, feeding problems, endocrinologic alterations such as premature thelarche in females, increased susceptibility to infections such as Otis and also autoimmune disorders (Kuroki et al., 1981; Niikawa et al., 1981). Furthermore, KS patients have been described to have congenital heart defects (Ohdo et al., 1985); the most common alterations are the juxtaductal coarctation of the aorta, the ventriculoseptal defect, and atrial septal defects (Digilio et al., 2001;

Hughes and Davies, 1994). More than 25% of patients have been reported to have renal anomalies such as malposition of the kidneys, renal hypoplasia or dysplasia, hydronephrosis, and fusion defects of the kidneys (Ilyina et al., 1995; Kawame et al., 1999; Niikawa et al., 1988). Interestingly, most of the systemic KS clinical anomalies can be traced back to a neural crest embryological origin (Vega-Lopez et al., 2018). Even the cardiac embryological defects can be traced back to alterations in cardiac neural crest (Kappetein et al., 1991; Keyte and Hutson, 2012).

The genetic cause of KS was unknown until 2010 when *KMT2D*[3] was identified as the gene primarily mutated (Ng et al., 2010) to cause the syndrome. Moreover, mutations of *KDM6A*[4] were later identified as responsible for being causative of KS in a minor percentage of cases (Lederer et al., 2012).

Mutations in *KMT2D* and *KDM6A* explain 60-70% and 10% of KS cases, respectively; the remaining cases are caused by a mutation in unidentified genes (Miyake et al., 2013). Knockdown experiments, with morpholino, in *D. rerio*, underlined the involvement of *KDM6A* and *KMT2D* in the development of brain and heart and recapitulated the craniofacial anomalies typical of KS (Van Laarhoven et al., 2015). According to the cardiological clinical phenotype, *KMT2D* has been found to regulate heart development in a KO mouse model. Mice with only one *Kmt2d* allele were viable and with mild defects in the ascending aorta; KO mice embryos lacked somites and headfolds at E8.0. Interestingly this study suggests that KMT2D is mainly involved in cardiac development through H3K4me2 deposition (S.-Y. Ang et al., 2016).

---

[3] In current literature, the gene coding for KMT2D can be found with many synonyms': *ALR*, *CAFL114*, *MLL4* and *MLL2*. To avoid any confusion, I will refer to histone modifiers only according to the nomenclature proposed by Allis et al. (Allis et al., 2007).
[4] Already known as UTX.

KMT2D gene

KMT2D resides on chromosome 12q13.12 and codes for the lysine 4 histone 3 (H3K4) methyltransferase KMT2D, a 5537 amino acids 593 KDa protein containing at the Amino-Terminal Domain (NTD) 2 plant homeotic domains (PHDs) clusters, containing three PHDs each cluster (Ruthenburg et al., 2007). Also, at the Carboxyterminal domain (CTD), KMT2D contains a SET catalytical domain, another PHD, and an FY-rich N/C-terminal (FYRN and FYRC) domain. Moreover, it includes nine nuclear receptor interacting motifs (LXXLLs) and a high mobility group (HMG-I) (Froimchuk et al., 2017).

KDM6A gene

KDM6A is located on the X chromosome but escapes its inactivation (Greenfield et al., 1998). It codes for a protein composed of 1401 aminoacidic residues, which weights around 154KDa. The NTD contains tetratricopeptide repeats, which are known to mediate protein-protein interactions (Smith et al., 1995) while at the CTD contains a treble-clef zinc finger, that may be involved both for DNA binding or protein-protein interactions (Ginalski et al., 2004). KDM6A is the only demethylase, together with KDM6B[5], able to counteract the post-translational modification performed by the Polycomb Repressive Complex 2 (PRC2) by performing di- and trimethyl H3K27 demethylation with the JmjC-domain (Agger et al., 2007; Lan et al., 2007). The Y chromosome hosts a KDM6A homolog named *UTY,* which is known to be catalytically dead (Hong et al., 2007). Its presence is necessary, since female *Kdm6a* KO mice die during embryonic development E11.5-12-5 (Shpargel et al., 2012; Wang et al., 2012).

---

[5] Already known as JMJD3.

KMT2D and UTX function in KMT2D/COMPASS complex

KMT2D fulfills its role in the KMT2D/COMPASS (COMplex of Proteins ASsociated with Set1) complex, an evolutionarily conserved multi-subunit complex able to perform H3 post-translational modifications. The methylase catalytic core of the COMPASS complexes are the enzymes KMT2. In mammals, there are six different H3K4 methylases (KMT2A-G, excluding the catalytically inactive KMT2E), which arise from duplication events occurred during evolution. In comparison to *Drosophila melanogaster*, KMT2A/B proteins are homologous of Trithorax (Trx); KMT2C/D of Trithorax-related (Trl) and KMT2F/G of Set1. In relation to yeast, Trx, Trl, and Set1 of *D. melanogaster* derive from the yeast Set1. While Set1 in yeast is able of catalyzing all the three methylation status of H3K4, together with the duplication, the metazoan COMPASS complexes acquired the ability to catalyze specific methylation status, with different regulatory function (Mohan et al., 2011; Shilatifard, 2012). In metazoan, the COMPASS complexes share a set of core subunits and contain a set of complex specific ones. The subunits common to all the COMPASS complexes are ASH2L and RBBP5, which interact and activate all the KMT2 family proteins (Li et al., 2016); WDR5, which binds the FYRN/FYRC domains of KMT2C and KMT2D (Cho et al., 2007); and DPY30. The specific subunits of the KMT2D/COMPASS complex are PTIP, PA1, NCOA6, and KDM6A, the histone demethylase able to remove the lysine 27 histone 3 tri-methylation (H3K27me3) (Kruidenier et al., 2012). Initially, KMT2D was first purified from HeLa nuclear extracts and was shown to belong to the "ASCOM" complex, together with ASH2L, NCOA6[6], RbBP5, and KMT2C (Goo et al., 2003). A later study on PTIP-associated proteins shown that endogenous PTIP interacts with ASH2L, RbBP5, WDR5, DPY30, NCOA6, KDM6A, PA1, KMT2C, and KMT2D. Furthermore, it was

---

[6] Called ASC-2 in the reference.

shown that PTIP and PA1 associate In the same protein complexes, which contain KMT2D or KMT2C and they exhibit robust H3K4 methylation activity (Cho et al., 2007). Additionally, during the same period, independent research corroborated the same discoveries (Issaeva et al., 2007; Patel et al., 2007).

Following ChIP-chip experiments of KDM6A in lung and foot fibroblasts, 92% of binding events were found to occupy HOX loci selectively, both in transcriptionally active and inactive HOX genes. Conversely, in mouse ESC ChIP-chip KDM6A was excluded from HOX genes, which are inactive and marked by H3K27me3 (Lan et al., 2007). In parallel, independent research showed the KDM6A role in HOXA1/3 and HOXB1/3 activation, following retinoic acid treatment in pluripotent testicular embryonal carcinoma cell line NT2/D1 (Agger et al., 2007; Lee et al., 2007). Moreover, it has shown that after retinoic acid administration, H3K27 demethylation and of PRC2 occupancy decrease in the promoter regions of HOXA1/3 and HOXB1/3, and it is subsequent to ASH2L recruitment and increase in H3K4me3 (Lee et al., 2007). Since both KDM6A and KMT2D are in the same molecular complex, it is not surprising that also KMT2D was found to be required to mediate retinoic acid response on its target genes (Guo et al., 2012). Afterward, the interaction with the retinoic acid receptor (RAR) was found to be mediated by the KMT2D/C COMPASS complex subunit NCOA6 (Rocha-Viegas et al., 2014). Moreover, estrogen receptor alpha (ERα), in the presence of its ligand, interacts directly with KMT2D thanks to its LXXLL motifs and triggers the localization of KMT2D to promoter regions of estrogen target genes, which are activated (Mo et al., 2006).

Nowadays, the research identified KMT2D as responsible for the deposition of the mono-methyl group on H3K4 (H3K4me1), a hallmark of enhancer regions, while not being directly involved in H3K4me3 (Lee et al., 2013a).

It is now known that the demethylation conducted by KDM6A makes H3K27 a substrate for p300/CBP acetylation (H3K27Ac), which is recruited by KMT2D (Wang et al., 2016). Consequently, the combined action of KMT2D and KDM6A is required to control cell fate transition, as shown in macrophages and ESC (Kaikkonen et al., 2013; Wang et al., 2016), by changing the enhancer status to an active state, which is defined by the presence of H3K4me1 and H3K27Ac whilst lacking H3K4me3 and H3K27me3 (Creyghton et al., 2010; Heintzman et al., 2007; Rada-Iglesias et al., 2011). Additionally, KMT2D/C are required for priming of enhancers and super-enhancers during cell differentiation of mouse brown adipose tissue, followed by their activation by CBP/p300 binding (Lai et al., 2017).

As with KMT2A and B, KMT2C and KMT2D have a partially redundant role: the deletion of either one of these two genes affects global H3K4me1 deposition only marginally. Mice model for Kmt2c and Kmt2d have phenotype onset remarkably different: *Kmt2c* depleted mice die after birth without morphological alterations, whereas *Kmt2d* KO mice die during embryonic development, at embryonic stage E9.5, due to its essential role in heart development, adipogenesis, and myogenesis. Accordingly, conditional KO in cardiac precursor and myocardium leads to embryonic lethality for severe cardiac defects (Siang-Yun Ang et al., 2016). Moreover, the conditional depletion of *Kmt2d* in the somatic precursor showed decreased brown adipose tissues formation and muscle mass levels, leading the mice to die by breathing dysfunction after birth (Lee et al., 2013a). Similar observations were reproduced after knocking down, by morpholino experiments, *KMT2D* and *KDM6A* orthologs in *Zebrafish. T*he authors observed their involvement in the heart, brain, and craniofacial development (Van Laarhoven et al., 2015).

KMT2D and KDM6A enzymatic-independent role

There is convergent evidence pointing to the fact that the role of KMT2D and KDM6A is not solely restricted to their catalytical activity. Indeed, research has shown an enzymatic-independent role for both. The presence of the SET domain in KMT2D it is not only essential for the catalytic activity but also for the stability of the protein. The catalytical activity is dispensable for mediate long-range interactions and promoting enhancer RNA (eRNA) transcription. Indeed, the loss of KMT2D catalytic activity only marginally impacts target gene expression and H3K27Ac deposition (Dorighi et al., 2017; Jang et al., 2017). However, the presence of KMT2D itself is essential to prevent the KMT2D COMPASS Complex to collapse and for KDM6B stability (Jang et al., 2017; Lee et al., 2013a).

In support to the hypothesis that these enzymes have alternative role than their catalytical activity, it is necessary to underscore that KS individuals have been found mutated in *KDM6A* both in females and males, which have only one copy (Banka et al., 2015; Lederer et al., 2012). As with KMT2D, also KDM6B catalytical activity has been found to be dispensable for enhancer activation and cell-type specific gene transcription for mammary luminal differentiation and *C. elegans* development (Vandamme et al., 2012; Yoo et al., 2016). These evidence permits to speculate that if KDM6A have an independent role from its catalytical activity, also the paralogue UTY[7] might have a similar role and might compensate for KDM6A loss in male KS individuals. Indeed, mouse UTY can compensate for *Kdm6a* deletion and can rescue development. The UTY function is independent of catalytical activity since both *in vitro* assay with the human UTY (Agger et al., 2007; Lan et al., 2007) and *in vivo* experiments with the mouse one showed its lack of enzymatic activity (Shpargel et al., 2012). KDM6A and UTY have redundant function in embryonic development,

---

[7] While *KDM6A* resides on the X chromosome, *UTY* is located in the Y chromosome.

and they both influence the deposition of H3K4 methylation since also UTY, interacting with RBBP5, is included in the KMT2D compass complex (Shpargel et al., 2012). Also, UTY partially recover NC conditional KO of *Kdm6A* phenotype, which appears to be majorly involved in NC post-migratory developmental stage (Shpargel et al., 2017)

A convergent line of evidence suggests that the enzymatic-independent function of KMT2D, KDM6A, and UTX is mediated by their ability to interact with chromatin remodelers of the SWI/SNF family. Indeed, both KDM6A and KDM6B have been found to mediate the interaction between the catalytical subunit of the SWI/SNF complex Brg1[8] and the transcription factor T-bet, of the T-box family (Miller et al., 2010). Moreover, also UTY has been found to interact with BRG1 (Shpargel et al., 2012). In addition, both wild-type and enzymatically inactive KDM6A have been found to mediate the interaction between Brg1 and Tbx5, and to facilitate Brg1 recruitment to the enhancers of the cardiac genes *ANF* and *Baf60c* (Lee et al., 2012). In addition to the interaction with Brg1, KDM6A has been found to directly bind, together with SMARCA2, to the first conserved zinc fingers of CBP by its TRP domains (Tie et al., 2012, p. 27). Moreover, both the SET domains of KMT2D and KMT2C have been found to directly interact with SMARCB1 (Lee et al., 2009). Finally, elegant research has shown that neither KMT2D nor KDM6A catalytical activities are necessary for gene activation, but it is their cooperativity with p300 that stimulates mostly target gene expression (Wang et al., 2017). The most updated model for the mechanism of the KMT2D COMPASS complex is summarized in Figure 9.

---

[8] In human SMARCA4
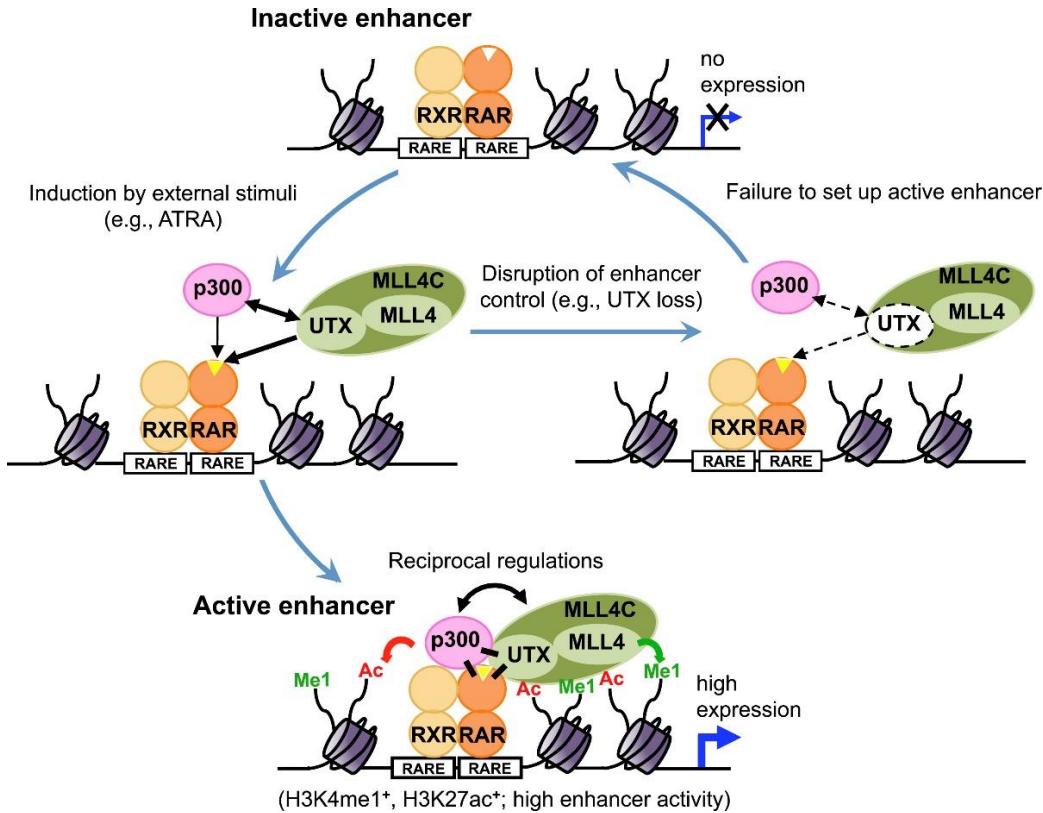
Figure 9 model for KMT2D COMPASS complex function. MLL4C: MLL4 complex; UTX: KDM6A; RARE: retinoic acid response elements. Adapted from Wang et al., 2017

Altogether, these piece of research, support the involvement of *KMT2D* and *KDM6A* in the etiology of KS but they are not explaining the clinical phenotype with an identification of the downstream targets altered as consequence of the heterozygous deletion of one of the two genes. Heterozygous mice for *Kmt2d* display increased serum bile acid, glucose tolerance, and insulin sensitivity and in this animal model *Kmt2d* was shown to regulate the hepatic circadian clock (Kim et al., 2015). Moreover, following studies highlighted the role of *Kmt2d* as an epigenetic mediator of overnutrition-induce murine steatosis (Kim et al., 2016).

Nowadays, research is defining a new class of diseases demarcated by alteration of enhancer regulation, and it is referring to this class as enhanceropathies (Smith and Shilatifard, 2014). Consequently, due to the players involved in KS, this syndrome can be included in this category. An altered stoichiometry of KMT2D/COMPASS

complex is relevant for the pathogenesis of KS, and it has already been shown that fine regulation of PRC2 axis is a key feature in neurodevelopment (Burgold et al., 2012, 2008; Fragola et al., 2013; Park et al., 2014; Testa, 2011). For these reasons, and as mentioned before, due to the importance of the integer presence of KMT2D, to better study a cellular model that can recapitulate KS pathogenesis, it is interesting to analyze the impact of KMT2D haploinsufficiency.

KMT2D mutations in cancer

Over the last years, *KMT2D* has emerged as one of the most frequently mutated genes across a wide spectrum of cancers, with a driver role in the majority of them (Kandoth et al., 2013; Pasqualucci et al., 2011; Rao and Dou, 2015). In normal cells, the *KMT2D* loss is associated with malignant transformation, and it is thought to be a tumor suppressor (Ford and Dingwall, 2015). Furthermore, high levels of KMT2D are related to poor prognosis in breast cancer patients (Kim et al., 2014). Conversely, *KMT2D* deletion in proliferating transformed cells have antiproliferative effect as it has been shown after its knockdown cancer cell lines (Guo et al., 2013).

Moreover, *KMT2D* is the most mutated chromatin modifier in medulloblastoma and it has a role in its biology (Northcott et al., 2012). Also, KMT2D has shown as a necessary regulator of metastatic melanoma cell migration (Bossi et al., 2016). More recently, acute deletion of *KMT2D* has been associated with genomic instability selectively affecting actively transcribed genes, thus providing a suggestive link between its enhancer-marking function and the preservation of genome integrity (Kantidakis et al., 2016). Surprisingly, although nearly three-quarters of KMT2D mutations in KS have the same pattern of domain loss to of the somatic cancer mutations (Rao and Dou, 2015), nowadays there is no association between KS and cancer. It is worthy of mention that also KDM6A is found mutated in several tumors and thought to work as a tumor suppressor, in part by counteracting PRC2 (Ezponda

et al., 2017, 2017; van Haaften et al., 2009). Unexpectedly, the incidence of cancer in KS has not been found to be increased in comparison to the general population. Nevertheless, some case reports show cancer occurrence in few KS individuals, and the authors invite the clinical community to carefully examine KS patients for malignancy (Bernier et al., 2017; Roma et al., 2015; Tumino et al., 2010).

YY1 haploinsufficiency syndrome

After the first identification of an individual with ID carrying a *YY1 de novo* mutation (Lisenka E L M Vissers et al., 2010), an international clinical effort discovered additional 23 ID individuals with *de novo YY1* mutations (Gabriele et al., 2017). YY1 haploinsufficiency syndrome is a rare autosomal dominant ND, following defined by the OMIM committee as Gabriele-De Vries syndrome[9] (GADEVS). People with YY1 mutations share ID with variable cognitive impairment, dysmorphic facial features, delayed psychomotor development, feeding problems, and movement abnormalities. The facial dysmorphism shared among individuals included facial asymmetry with a broad forehead, an upper-lip indentation shaped like a Gingko leaf, and fullness of the upper eyelids. In the represented cohort, half individuals displayed behavioral anomalies like anxiety and autistic features. Few patients featured dystonia and a variety of congenital abnormalities, such as esophageal atresia, cleft palate, hydronephrosis, craniosynostosis, and Epstein anomaly (Gabriele et al., 2017).

YY1, gene description and function

YY1 is located on chromosome 14q32.2 and it codes for an ubiquitous (Patten et al., 2018) zinc finger protein of 414 amino acids (Shi et al., 1991). It is homologous of the *D. melanogaster Pleiohomeotic* gene, which codes for a transcription factor able to mediate the recruitment of Polycomb Group (PcG) proteins to the Polycomb Responsive Elements (PRE) on DNA (Brown et al., 1998). *YY1* was named after the *Yin-Yang*, the Lao Tsu idea of equilibrium, given the fact that, in mammals, YY1 was initially found to function both as activator and repressor. It was first described that

---

[9] OMIM: #617557

in presence or absence of the adenovirus E1A oncoprotein it represses or activates, respectively, the adeno-associated virus (AAV) P5 promoter (Shi et al., 1991). Following the line of research on *Drosophila*, initially YY1 was studied as PRC2 interactor and was thought to mediate its repression. Several pieces of research have showed YY1 interaction with PRC2 and reported a possible role as repressor. However, most pieces of research show YY1 as an activator, especially in nervous system development (He and Casaccia-Bonnefil, 2008). The first piece of research in which unbiased genome-wide analysis were performed suggested YY1 to have a major involvement in gene activation of highly transcribed genes, while having a negative role in nuclear and nucleolar small non-coding RNAs biogenesis, in mouse ESCs (Vella et al., 2012). Moreover, by mass-spectrometry analysis, no protein of the Polycomb-group was found to immunoprecipitated with YY1, in mESC (Vella et al., 2012). Nowadays, the functional role of YY1 seems to be clarified and can be compatible both with the activating and repressive reports. Indeed, YY1 mediates the formation of the structural loop between enhancer and promoters (Weintraub et al., 2017). Literature showing YY1 as repressor can be compatible with the later description of poised enhancer (Calo and Wysocka, 2013). For example, in myoblasts, together with PRC2, YY1 occupies MyoD promoter, which is then expressed in myotubes. YY1 removal, even if it is associated with H3K27me3 decrease, was not triggering gene activation (Caretti et al., 2004). Indeed, it is not YY1 *per se* function to activate or repress a gene, but the interactors associated with a chromatin loop. YY1 main interactors are p300/CBP (Lee et al., 1995), the INO80 chromatin remodeling complex, as well as the two RNA helicases Ddx5 and Ddx3x (Cai et al., 2007; Vella et al., 2012; Wu et al., 2007). A recent study, in which eQTL were analyzed in 25 tissues, showed YY1 to have predominantly an activating role (Reshef et al., 2018).

Moreover, as well as other transcription factors, YY1 binds RNA and it has been shown to mediated X-chromosome inactivation by binding *Xist* RNA (Jeon and Lee, 2011) and by promoting *Xist* expression (Makhlouf et al., 2014). The N-terminal region is responsible for the RNA-binding property of YY1, which make it capable of binding enhancer RNAs (eRNA) of active enhancers and DNA of promoters. For this reason, it has been proposed that RNA contributes to stabilize YY1 occupancy in regulatory elements (Sigova et al., 2015). A further study has shown that also the Zinc Finger domain is capable of binding RNA, with low specificity (Wai et al., 2016).

While YY1 homozygous deletion results in embryonic lethality, the pleiotropy of YY1 haploinsufficiency has been already shown in a mouse model. Its heterozygosity leads to serious growth retardation, proliferative and neurological defects such as exencephaly, pseudoventricles, and asymmetry of the developing brain, even though at incomplete penetrance (Donohoe et al., 1999). Other lines of research have underscored the direct relevance of YY1 for neuronal development (He and Casaccia-Bonnefil, 2008). Indeed, YY1 activity is necessary for oligodendrocytes for global nerves myelination (He et al., 2007) and in Schwann cells, where it mediates the myelination gene expression program neuregulin-dependent (He et al., 2010). Moreover, recent research has showed that *YY1* knockdown impairs enhancer-promoter interactions of neuronal progenitor cells regulators (Beagan et al., 2017).

Aim of the Thesis

An historical challenge for studying neurodevelopmental disorders has been the lack of reliable human platforms. Especially the human brain has anatomical and functional properties that are not possible to reconstruct in animal models. This has always constituted a major challenge for the study of neurodevelopmental disorders. Only recently it was possible to study this class of diseases in human reliable and informative cellular models (Linda et al., 2018). Moreover, the molecular diagnosis of neurodevelopmental disorders caused by rare sporadic germline mutations, and the understanding of their molecular impact, it has become possible only recently thanks to the development of next-generation sequencing techniques. In this work I pursued two connected line of research:

I) the generation of a reliable human model for KS, and the study of the molecular and functional impact in primary cell types, at the pluripotent stage, and in disease-relevant cell types such as NCSC, and cortical neurons;

II) II) defining the molecular mechanism underlying the intellectual disability caused by mutations in *YY1*, and the generation of a disease modeling platform based on iPSCs, which will permit us to study the impact of *YY1* mutations in specific disease-relevant cell types in the future.

MATERIAL AND METHODS

Cell Culture

Fibroblasts were cultured in RPMI 1640, FBS 15%, L-Glutamine 1%, Penicillin-Streptomycin 1%. Trypsin was used to passage fibroblasts. iPSCs were cultured with mTeSR-1 (StemCell Technologies) or Essential 8 (E8, Thermo fisher) in feeder-free conditions on hES-qualified Matrigel (BD Biosciences) diluted 1:40. iPSCs were passaged with ReLSR (StemCell Technologies). Accutase (Sigma) was used when single cell for counting were needed, supplementing the medium with rock inhibitor 5 μM Y-27632 (Sigma. NCSC were cultured with a defined medium as previously described (Menendez et al., 2013). Cortical Neurons were induced as described elsewhere (Y. Zhang et al., 2013) and maintained with neurobasal medium fully complemented and conditioned overnight on mouse astrocytes, or with neurobasal plus (Thermoscientific). LCL were grown in RPMI 1640, FBS 15%, HEPES 1%, L-Glutamine 1%, Penicillin-Streptomycin 1%. All samples were tested for mycoplasma.

Reprogramming into iPSCs

The six fibroblasts samples: KAB_1, KAB_4, KAB_5, CTL_1, CTL_4, and CTL_5 were reprogrammed into iPSC using a non-integrative method (Yoshioka et al., 2013) based on the self-replicating Venezuelan Equine Encephalopathy (VEE) virus. The reprogramming was performed according to the manufacturer's protocol (Simplicon™ RNA Reprogramming Technology, Millipore). VEE replicon presence removal was evaluated by RT-qPCR using already primers already published (Yoshioka et al., 2013). The fibroblast samples: KAB_3, KAB_7, GADEVS samples and control LCL were reprogrammed using Sendai virus according to following the

manufacturer's protocol (CytoTune™-iPS 2.0 Sendai Reprogramming Kit, ThermoFisher Scientific).

iPSCs Differentiation into NCSC and Cortical Neurons

iPSCs were differentiated into NCSC using a published protocol (Menendez et al., 2013). NCSC purity was assessed by FACS analysis using antibodies against NK-1 and NGFR as previously described (Adamo et al., 2015).

iPSCs Differentiation into Cortical Neurons

iPSCs were engineered using a ePiggyBac (ePB) transposon. iPSCs were electroporated with the Neon system (ThermoFisher) with the following parameters: 1200V, 30 ms, 1 pulse using 0,5 μg of an helper plasmid expressing a transposase and 4,5 μg of donor plasmid with a transposable element with the following genetic configuration: hUbC promoter - rtTA - T2A – BsdR -TRE – *Ngn2* – P2A – EGFP – T2A – PuroR as represented in Figure 10 to allow the selection of by administration of blasticidin 5μg/mL. Neuronal differentiation was driven by doxycycline administration followed by Ngn2 expression, differentiation and maturation were performed as already described (Zhang et al., 2013). Stainings on neurons were performed using nitric acid treated coverslips coated overnight with poly-L-Lysine.
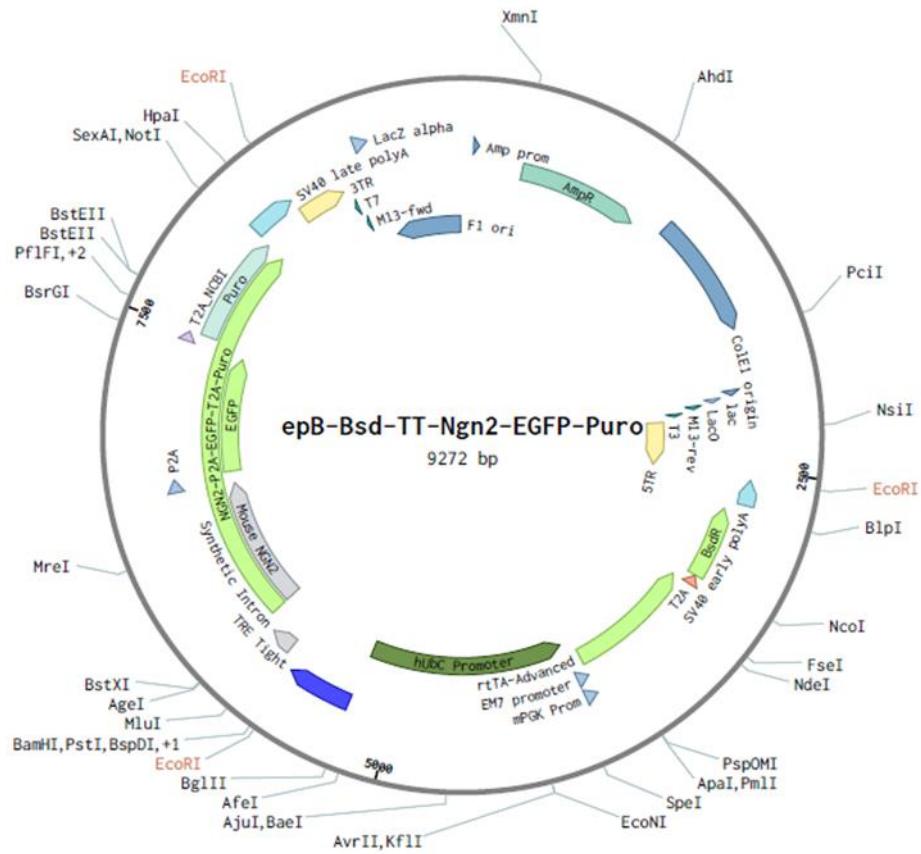
Figure 10 Schematic represenation of Ngn2 ePB donor plasmid

Immuno-Fluorescence Stainings

Immunofluorescence stainings were performed on living cells using the TRA-1-60 live cell staining, DyLight 488 conjugated(Stemgent, 09-0068). Stainings on fixed cells were performed washing samples with cold PBS. Cells were treated for 10 minutes with 4% Paraformaldehyde/4% sucrose, washed with PBS, and permeabilized with Triton X-100 diluted in PBS. A blocking of 1 hour was performed in serum matched with the specie of the secondary antibody. After washes with PBS, samples were incubated 2 hours at room temperature or overnight at 4°C with the primary antibody diluted in blocking buffer. Samples washed with PBS to remove primary antibodies and incubated with secondary antibodies for 1 hour at room temperature. Secondary antibodies were removed by washing with PBS. Afterward, DAPI was given for 5 minutes, removed with PBS, and samples were finally mounted on coverslides with Moviol mounting medium. Stainings were performed using

36

Nanog (Everest Biotech, EB06860), Oct3/4 (Santacruz, sc-9081), TUJ1 (Covance, MMS-435P), MAP2B (BD Biosciences, 610460), VGLUT1 (Synaptic System, 135 303).

Protein extraction and immunoblotting

Protein extraction was performed using RIPA buffer (150 mM NaCl, 1.0% NP-40, 0.5% sodium deoxycholate, Protease inhibitor cocktail (Sigma), 0.1% SDS and 50 mM Tris, pH 8.0) to lysate cells. Lysis was followed by sonication using the Bioruptor Sonication System (UCD200) for three cycles of 30s at high power with 30s pauses and centrifuged at 13,000g for 15 min. Protein quantification was performed with Bradford protein assay (Bio-Rad). For protein electrophoresis 20–40 μg of protein extracts were loaded on home-made 10% SDS-PAGE gels. Transfers were performed with a tension of 120V for 1 hour onto nitrocellulose membranes and blocked in TBS-T 0.2% (50 mM Tris, pH 7.5, 150 mM NaCl and 0.2% Tween-20) and 5% milk. Primary and secondary antibodies were diluted in TBS-T 0.2%, 5% milk. Odyssey Infrared Imaging System (LI-COR Biosciences) was used to acquire images. Densitometry was performed using ImageJ software. The secondary antibodies were a α-rabbit IRdye680LT or a α-mouse IRdye800LT (LI-COR Biosciences).

RNA-seq transcriptome analysis

RNA samples were extracted using the RNeasy Micro Plus kit (Qiagen) according to the manufacturer's protocol. RLT buffer containing β-mercaptoethanol was directly added to samples. RNA Concentration and purity was evaluated by the NanoDrop (NanoDrop Technologies). Before RNA-seq library construction, RNA integrity was assessed by 2100 Bioanalyzer (Agilent Technologies). Only samples with RIN > 9

were used for library preparation. Prior library preparation ERCC spike-ins (ThermoFisher Scientific) were used to facilitate data normalization. RNA-seq libraries were prepared following manufacter protocols for Truseq-stranded RiboZero depletion (Illumina). Libraries were sequenced using a HiSeq2000 (Illumina) at a depth of 30 million reads, paired-end, 50bp sequencing. RNA-seq analysis was performed directly from the reads using Salmon 6.1(Patro et al., 2015), which has been showed to be excellent in tracking relative gene expression differences (Germain et al., 2016), using hg38 and the Refseq annotation complemented with the sequences of the ERCC spike-ins.

ChIP-seq epigenomic analysis

Chromatin immunoprecipitations were performed as previously described (Frank et al., 2001) with some modification. Cross-linking was performed using formaldehyde 1% in PBS. Glycine was added to the final concentration 125 mM to quench formaldehyde. Cells were re-suspended using ChIP SDS buffer (0.5% SDS, 5 mM EDTA, NaCl 100mM and 50 mM Tris-HCl at pH 8.1). Pellets were collected at 400g for 30 min and resuspended in IP buffer (0.5% SDS, 5 mM EDTA, NaCl 100mM and 50 mM Tris-HCl at pH 8.6, Triton X-100 1,5%). Cells were sonicated to obtain chromatin size of 300 bp DNA fragments, using the Branson digital sonifier (Emerson Industrial Automation). Chromatin for ChIP was quantified using Bradford protein assay (Bio-Rad) or normalized on cell number.

For each chromatin modification, IPs were performed using 100 µg of chromatin. Antibody used were: Ab8580 (H3K4me3), Ab8895 (H3K4me31), Ab4729 (H3K27Ac), 9733B (H3K27me3), sc-1703 and sc-281 (YY1). Chromatin was incubated overnight with 2 µg of antibody at 4°C. Protein G - Sepharose 4B (Thermo Fisher) were incubated with chromatin and antibodies mix for 4 hours at 4°C.

Washes were conducted by using low and high salt buffers (1% Triton X-100, 0.1% SDS, 150mM or 500mM NaCl , 2mM EDTA, 20 mM Tris-HCl ph8). Decross-linking was performed at 65°C for 3 hours. DNA was collected with QIAquick PCR Purification Kit (Qiagen). Libraries were prepared as already described (Blecher-Gonen et al., 2013) with adaptations for the automated system Biomek FX.

ChIP-seq reads were trimmed using scythe 0.981 (min 4 nucleotides) before being aligned to the hg38 genome using Bowtie 1.0 (-v 2 -m 1), peaks were called using MACS 2.0.9 with default settings (except for H3K27ac and H3K27me3 for which the 'broad' option was enabled).

Differential expression analyses

Differential expression analyses (DEA) were performed with EdgeR (Robinson et al., 2010) with the parameter "estimateGLMRobustDisp", taking in account genotypes, sex, family (when possible), and batches when present. Gene ontology analysis were performed by in-house tool, Ingenuity pathway analysis (IPA, Qiagen) and the online tool webgestalt[10], which permits to load a custom universe of genes as background (taking in account the expressed genes for each specific cell type) selecting Over-Representation Analysis.

Micro electrode arrays recordings and Neuronal network analysis

Eight iPSCs (4 lines from Kabuki patients, 3 lines reprogrammed from half-matched and one unmatched healthy subjects) were differentiated into upper-layer excitatory cortical neurons as described above. The derived neurons were plated onto MEAs or

---

[10] http://www.webgestalt.org

glass coverslips pre-coated with adhesion promoting factors (Poly-L-Ornithine 50 μg/mL and Laminin 20 μg/mL) at a final density of 100 and 600 cells/mm2 respectively. After two days, rat astrocytes were added in the culture at the same density. Recordings of the spontaneous activity of hiPSCs-derived neuronal networks were performed during the 5th week in vitro. Recordings were performed using the 24-well MEA system (Multichannel Systems, MCS GmbH, Reutlingen, Germany). Spontaneous electrophysiological activity of hiPSC-derived neuronal network grown on MEAs was recorded for 20 min. The signal was sampled at 10 KHz, filtered with a high-pass filter (i.e. butterworth, 100 Hz cutoff frequency) and the noise threshold was set at ±4.5 standard deviations. During the recording, the temperature was maintained constant at 37 °C, and the evaporation and pH changes of the medium was prevented by inflating a constant, slow flow of humidified gas (5% CO_2, 20% O_2, 75% N_2) onto the MEA. Data analysis was performed off-line by using a custom software package named SPYCODE (Bologna et al., 2010) developed in MATLAB (The Mathworks, Natick, MA, USA). Cells plated on coverslips were transfected with a pCAG-dsRED plasmids 7 days after plating. After 23 days in vitro, cells were fixed and mounted for imaging. Neurons were imaged using an Axio Imager Z1with 568nm laser light and Axiocam 506 mono. Neurons were digitally reconstructed using Neurolucida 360 software (MBF–Bioscience, Williston, ND, USA).

RESULTS

Kabuki syndrome modelling

Cohort of patients and fibroblasts characterization

Thanks to the collaboration with the Telethon biobank we obtained Fibroblasts samples of seven KS patients and five matched un-affected relatives. As depicted in Figure 11 all individuals carry indel or nonsense mutations that eventually lead to a premature termination codon (PTC), which is predicted to activate nonsense mediated decay (NMD).

Name	family	sex	mutation
KAB_1	a	m	p.C197AfsX11
KAB_2	b	f	p.Y223X
KAB_3	c	f	p.P235PfsX26
KAB_4	d	f	p.R2635X
KAB_5	e	f	p.R3321X
KAB_6	-	m	p.K1885QfsX18
KAB_7	-	f	p.R2635X
CTL_1	a	m	
CTL_2	b	f	
CTL_3	c	f	
CTL_4	d	f	
CTL_5	e	f	
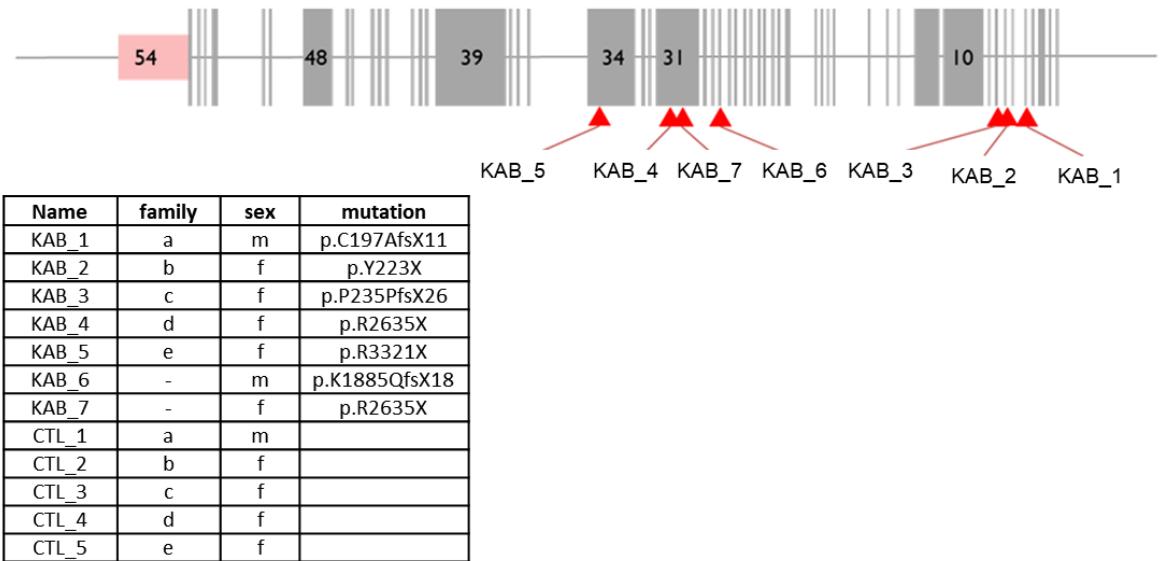
Figure 11 Schematic represenation of KTM2D exons with mutations position for availavle KS individuals. RThe number in the sample name identifies the family. Controls individuals are half-matched, sex-matched unaffected relatives.

Heterozygous mutations of KMT2D do not affect bluk H3K4 methylation deposition

Even though the KS individuals carry a heterozygous mutation on *KMT2D*, in fibroblasts samples the bulk level of H3K4 post-translational modification was not affected, as shown in Figure 12.
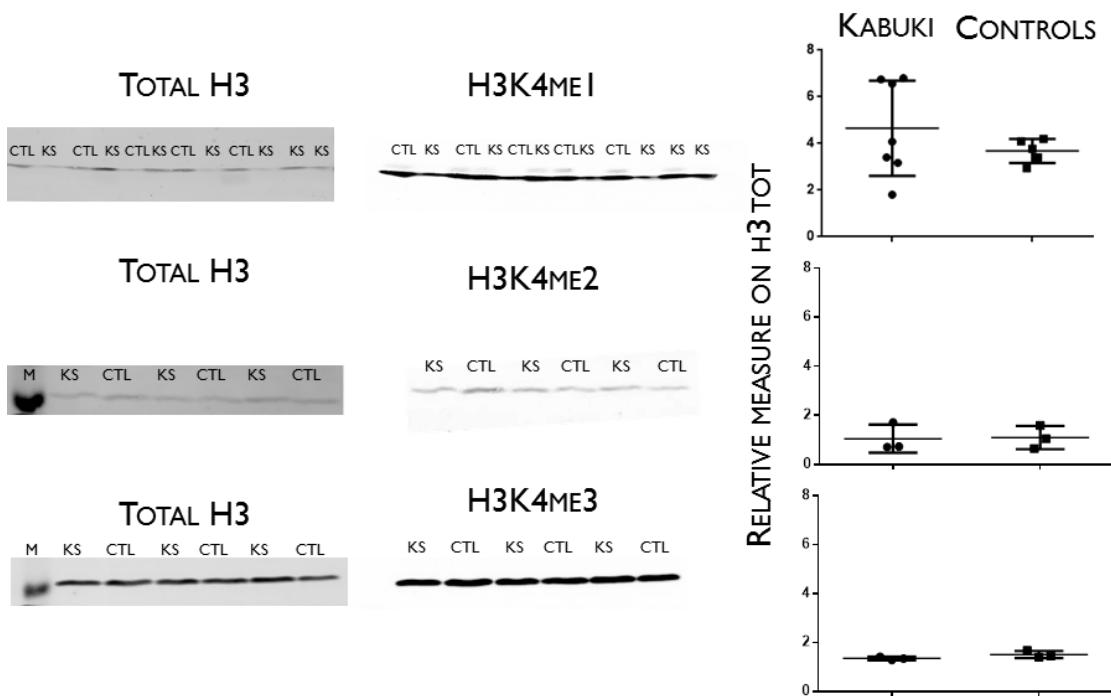
Figure 12 heterozygous KMT2D mutations do not affect bulk H3K4 post-translational methylations. Western Blotting for H3K4 modification in fibroblasts lysates.

Molecular characterization of KS primary fibroblast samples

Transcriptome analysis was performed on the whole cohort of primary fibroblast samples. DEA identified 189 Differentially Expressed Genes (DEGs) as shown in Figure 13. Although the number of DEGs is not high, it is possible to identify statistically significant enrichment that can be related with the clinical phenotype, as depicted in Figure 14.
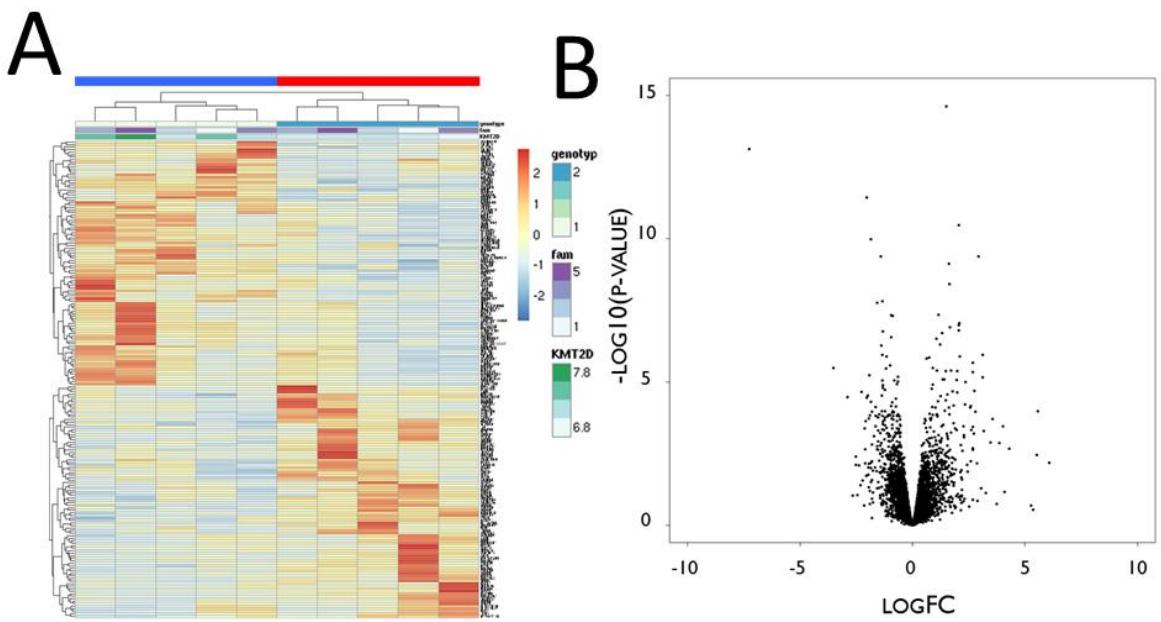
Figure 13 Transcriptional characterization of KS primary fibroblasts and half matched samples. A: Differentially Expressed Genes (DEGs) detected by EdgeR paired analysis. Conditions: Red: Kabuki; Blue: B: Volcano Plot: gene expression distribution

Figure 14 Goseq enrichment analysis in fibroblasts RNA-seq data

Given the enzymatic activity of KMT2D the genomic distribution of H3K4me1 was analysed. ChIP-seq result for this histone modification identified 7455 (Figure 15)

differentially methylated regions (DMRs) between patients and control. Moreover, Principal Component Analysis (PCA) highlights a clear clustering separation between patients and controls (Figure 15B). Although the bulk H3K4me1 does not change, H3K4me1 distribution is affected at specific loci in KS.
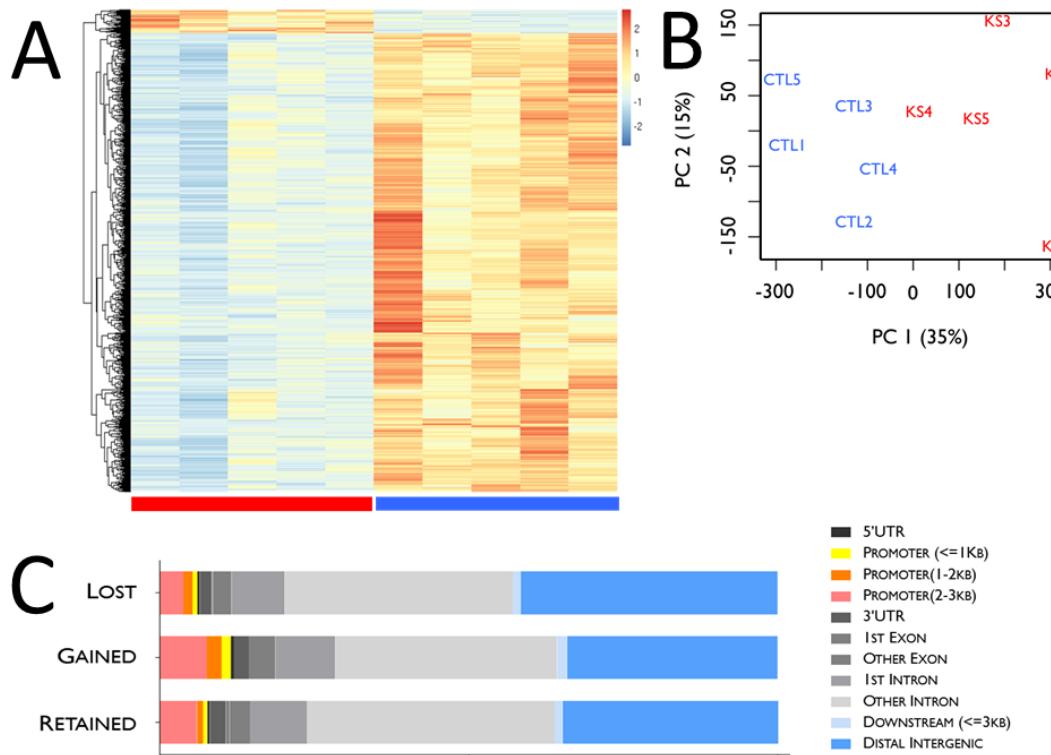
Figure 15 Genome wide distribution of H3K4me1 in primary fibroblasts of KS individuals and matched controls. A: Differentially methylated regions detected using EdgeR paired analysis. Conditions: Red: Kabuki; Blue: Controls.

According to previous research (Lee et al., 2013a) that describe KMT2D with a key role in enhancer regulation, the peak distribution is mainly lost at distal intergenic regions (Figure 15C), while regions proximal to gene promoters are less affected. To further valorise our H3K4me1 ChIP-seq we crossed it with the Roadmap Epigenomics public data (Kundaje et al., 2015) using dermal fibroblast histone modification data (Figure 6). Here we can validate that KS have a defect in H3K4me1 in regions normally marked with this histone modification in Roadmap epigenomics datasets. Moreover, regions marked by H3K27me3, in Roadmap epigenomics

44

datasets, are aberrantly marked by H3K4me1 in KS primary fibroblasts. This observation might be caused by an aberrant compensation of the other COMPASS complexes. To validate this observation, we performed H3K27me3 ChIP-seq in fibroblast samples, the current result remains inconclusive. Therefore, this data needs further analysis.

	ATT4CNOK (1929)	ALLCMAX1K (991)	INALLCONTROLS (12627)	INALLKABUKI (19069)	ALLKMAX1C (1851)	ALLK4NOC (4271)	ALLPEAKS (1137677)	
	1238	752	10715	16507	1460	2502	348952	E126-H3K4ME2 (153218)
	665	393	5359	9427	1013	1810	255212	E126-H2A.Z (289127)
	620	403	5927	10427	1045	1619	215271	E126-H3K4ME3 (226309)
	592	323	4764	7659	682	1449	298054	E126-H3K79ME2 (135114)
	698	522	8508	13090	977	1454	239089	E126-H3K9AC (207500)
	751	356	5827	10125	1078	2575	538609	E126-H4K20ME1 (405231)
	1194	747	10540	14286	929	1392	324771	E126-H3K27AC (203183)
	310	120	1738	2557	247	650	228029	E126 –H3K36ME3 (265238)
	1750	944	12223	17217	1319	2172	478131	E126-H3K4ME1 (211261)
	75	32	268	452	73	222	92831	E126-H3K9ME3 (444612)
	101	34	305	1604	380	1052	181853	E126-H3K27ME3 (342338)

COLOR KEY AND HISTOGRAM

Figure 16 H3K4me1 ChIP-seq data overlap with Roadmap epigenomics data. Numbers among brackets: total numbers of peaks. Numbers in the boxes: overlapping peaks; Colors: Z-score of ratio between columns and rows. Columns: genomic regions of H3K4me1 fibroblast ChIP-seq in set of different stringency

Generation of Kabuki syndrome iPSCs

To model the pathogenesis and the impact of KMT2D in non accessible tissues, starting from 3 patients and their matched controls we generated iPSCs with a non-integrative self-replicating mRNA (Yoshioka et al., 2013) that contains *GLIS1*, a maternal factor highly expressed only in unfertilized egg and one-cell stage embryos, instead c-myc, and was shown to achieve superior reprogramming quality in terms of ES-cell-like colonies number and survival of chimaeric mice (Maekawa et al.,

2011). 2 additional samples that were refractory to the reprogramming with this method were reprogrammed using the Sendai virus, a well-established reprogramming platform (Sochacki et al., 2016) that permits to reprogram with the classical OKSM factor with a single infection. First, iPSCs were characterized by immunofluorescence stainings for pluripotency defining markers NANOG, OCT4, TRA-1-60 (Figure 17).
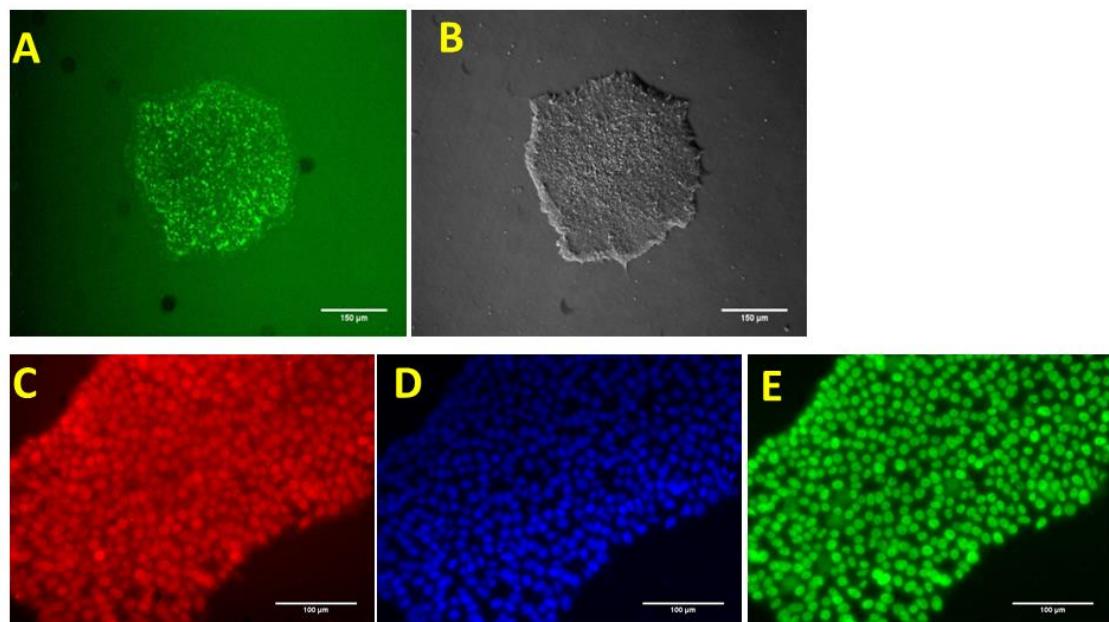
Figure 17 Representative staining for pluripotency defining markers in KS iPSCs. A: Anti TRA-1-60 4X, live staining; B: Phase-contrast; C: Anti NANOG 20X; D: DAPI 20X; E: OCT4 20X
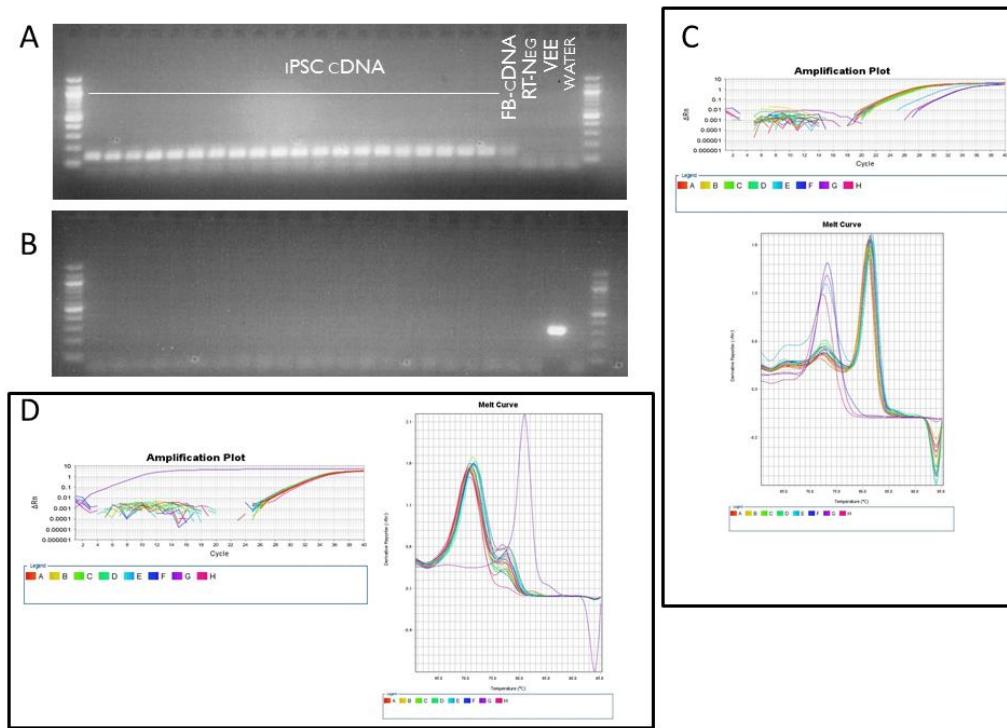
Figure 18 iPSCs samples reprogrammed with self-replicating RNA method are negative for the VEE construct. A: TBP qPCRgel is positive only for iPSCs and fibroblasts;B: VEE nsP2 qPCR gel is negative for iPSCs and fibroblasts; C: TBP qPCR is positive only for iPSCs and fibroblasts, purple and close samples: positive control of VEE mRNA, water, negative for RT, no cDNA in reaction D: VEE nsP2 qPCR is negative for iPSCs and fibroblasts, purple: positive control of VEE mRNA Melt curves of iPSCs are equivalent to negative RT, water, no cDNA in reaction, and fibroblast sample.

Then, to verify the complete depletion of the self-replicating mRNA from the reprogrammed colonies, as a consequence of the removal of B18R from the reprogramming medium (which was added to suppress the innate immune system), qPCR with primers designed to detect the construct was performed on all samples. iPSCs and a fibroblast sample were positive for TBP and completely negative for the self-replicating mRNA compared with a positive control (Figure 18).

iPSCs obtained with self-replicating mRNA were profiled with RNA-seq to assess both the impact of *KMT2D* mutation at the pluripotent stage. As depicted in Figure 19, KMT2D heterozygous mutations impact the pluripotency stage only marginally. With 0,05 FDR 295 DEGs were identified; 907 with an FDR of 0,1. Although the number of DEGs were comparable to the ones detected in fibroblasts, no Gene Ontology enrichment were identified. Indeed, clones of each individual, family and condition cluster together in unsupervised clustering (Figure 20), which shows a

high homogeneity among clones, indicating that at the pluripotency stage the genetic background overrides the differences imputed to KMT2D mutations. To assess whether the mutation could impact the enhancer activation status in pluripotency, iPSCs generated with the self-replicating mRNA were analyzed for H3K4me1 and H3K27Ac ChIP-seq. Consistent with the transcriptomic analysis, also the ChIP-seq results show an alteration at very few loci and PCA analysis are not able to detect any clear separation between KS and control samples both for H3K4me1 and H3K27Ac (Figure 21 and Figure 22).
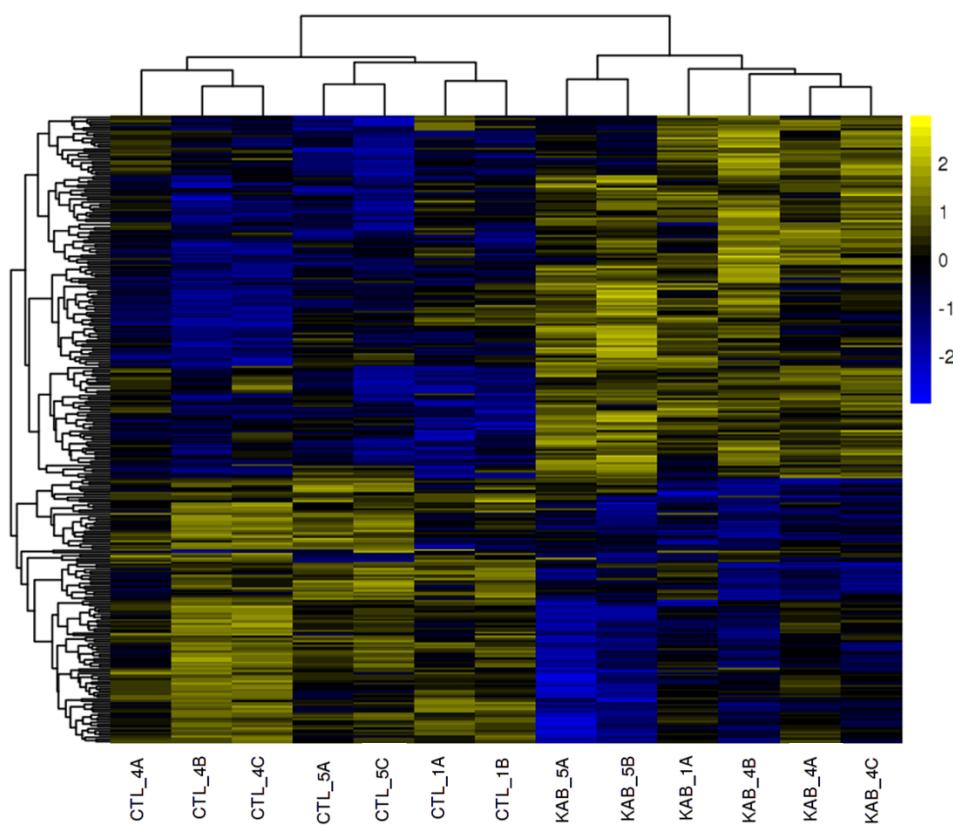
Figure 19 Transcriptome analysis of KS iPSCs and matched controls reprogrammed with self-replicating mRNA. KMT2D mutations do not dramatically alter the transcriptome in pluripotency Scalebar: Log norm gene counts FDR: 0,05
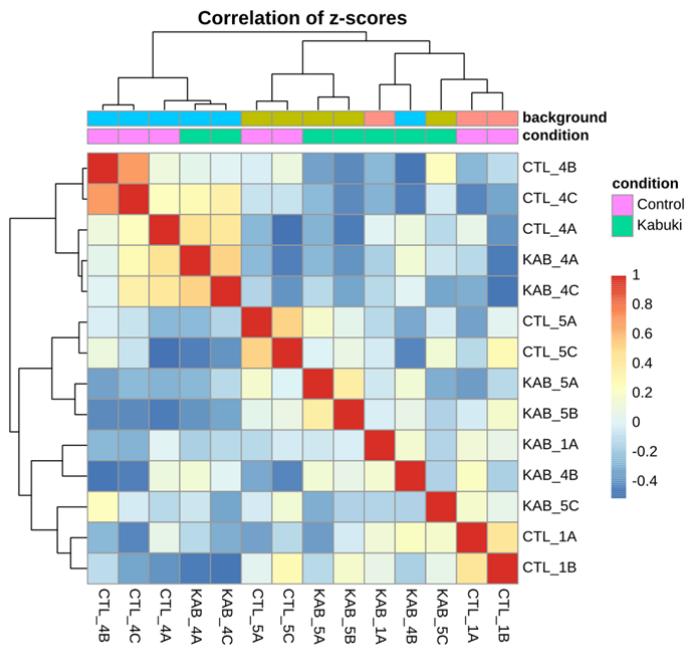
Figure 20 Unsupervised clustering of iPSC transcriptome shows high correlation among clones of each patient and individuals of each family. Color code: Pearson correlation of z-scores across all samples.
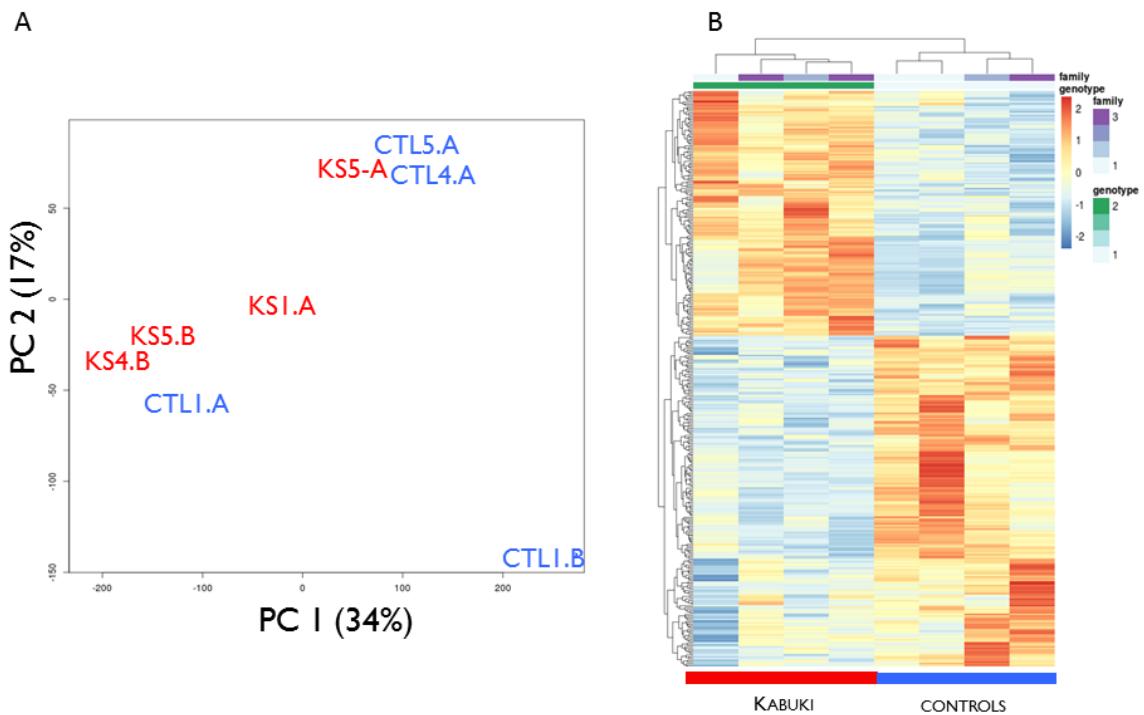
Figure 21 H3K4me1 in KS iPSCs reprogrammed with self replicating mRNA. A: PCA analysis; B: Differentially methylated regions
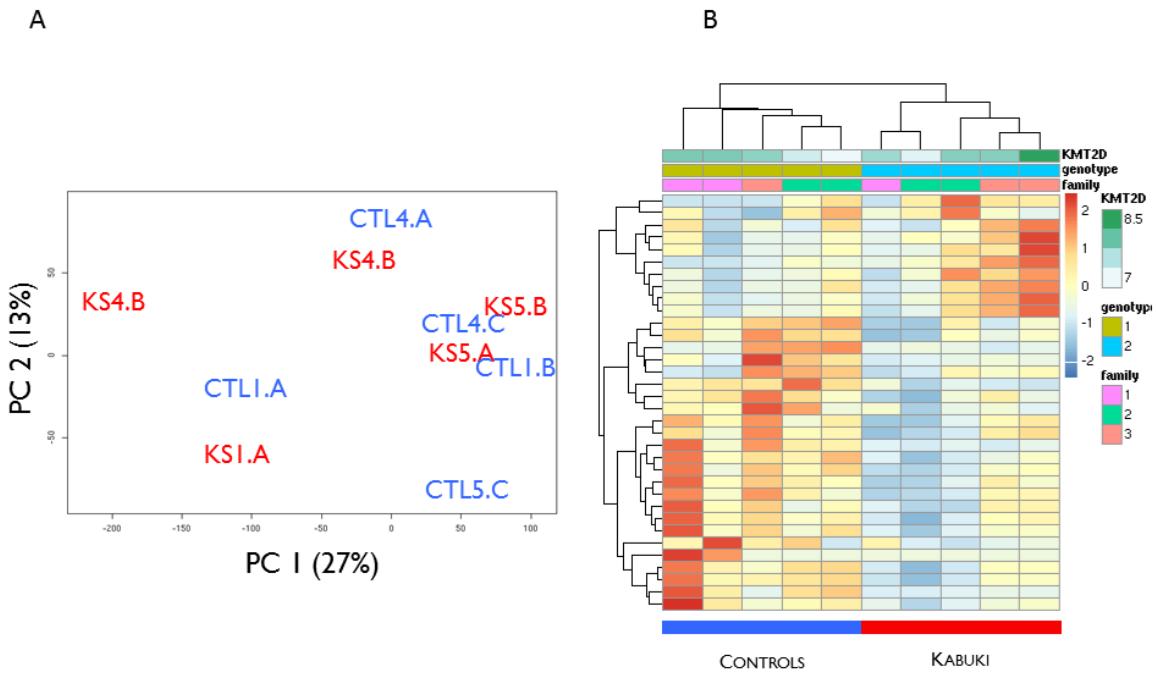
Figure 22 H3K27Ac in KS and half-mathced controls reprogrammed with self-replicating mRNA. H3K27Ac ChIP-seq in iPSCs A: PCA analysis; B: Differentially methylated regions

It is important to underscore that at the pluripotent stage RPKM values for *KMT2D* do not show a statistical difference between controls and KS (Figure 23). To increase

logFC	logCPM	LR	PValue	FDR
0.2964276	5.770919	1.5722	0.2098878	0.4689536
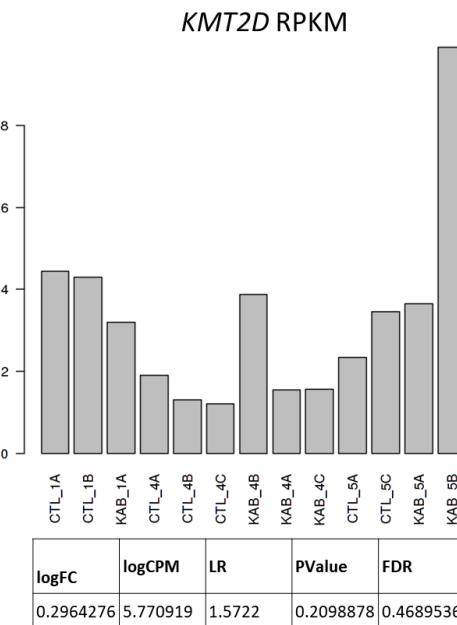
the sample number size, using the Sendai Virus, two additional KS samples were reprogrammed into iPSCs. The following table summarizes the complete cohort of obtained KS iPSCs.

Figure 23 KMT2D expression values in iPSCs. Values in the barplot: RPKM

Clone names	Reprogramming method
CTL_1A	Self-replicating mRNA VEE
CTL_1B	Self-replicating mRNA VEE
CTL_1C	Self-replicating mRNA VEE
CTL_4A	Self-replicating mRNA VEE
CTL_4B	Self-replicating mRNA VEE
CTL_4C	Self-replicating mRNA VEE
CTL_5A	Self-replicating mRNA VEE
CTL_5C	Self-replicating mRNA VEE
KAB_1A	Self-replicating mRNA VEE
KAB_3A	Sendai Virus
KAB_3B	Sendai Virus
KAB_3C	Sendai Virus
KAB_4A	Self-replicating mRNA VEE
KAB_4B	Self-replicating mRNA VEE
KAB_4C	Self-replicating mRNA VEE
KAB_5A	Self-replicating mRNA VEE
KAB_5B	Self-replicating mRNA VEE
KAB_5C	Self-replicating mRNA VEE
KAB_7A	Sendai Virus
KAB_7B	Sendai Virus
KAB_7C	Sendai Virus
KAB_7D	Sendai Virus
KAB_7E	Sendai Virus
KAB_7F	Sendai Virus
KAB_7G	Sendai Virus

Differentiation of KS iPSCs into disease-relevant cell types

For each individual, several iPSCs clones have been selected for expansion and used to analyze the pluripotent stage. To decide the number of iPSC clones to be differentiated in disease-relevant cell types we took in account several factors. From RNA-seq and ChIP-seq data, each clone has been found to be homogenous with the other iPSCs clones from the same individual. The major source of variability was identified to be the familial genetic background, as each clone of individuals of the same families segregated together. Moreover, genetic variation within iPSC clones, that was initially thought to be induced by the reprogramming process has been demonstrated to arise from somatic chimerism already existing in the parental cell from which the iPSC clone arose (Abyzov et al., 2012; Young et al., 2012). Also, it has been described that the primary source of variability among iPSCs clones

coming from different subjects arises from inter-individual genetic differences (Kilpinen et al., 2017). Accordingly, given the genetic heterogeneity between individuals, to obtain a robust experimental design able to flatten differences coming from such genetic backgrounds, and to outstand the variability driven by shared genetic lesions, recent *in silico* analysis suggested to increase the sample cohort by including multiple biological replicates (defined as individuals with unrelated genetic background) rather than increasing the number of clones of fewer subjects: a practice that increases the probability of identifying false positives *as bona* fide differentially expressed genes (Germain and Testa, 2017). Indeed, permutation analysis showed that the use of multiple clones per individual can be detrimental for transcriptomic studies, while the use of a single clone and multiple individuals benefit the specificity and sensitivity (Germain and Testa, 2017). Therefore, given the fact that in this project we obtained a number of individuals' iPSCs compatible with the scenario that permits to identify the highest number of *bona fide* DEGs, keeping low the number of false positives, we decided to differentiate only one clone from each individual into disease-relevant cell types.

Differentiation in cranial neural crest stem cells

NCSCs are involved in the development of multiple tissues, organs, and systems affected in KS, such as the immune system, the atrial septa, the gastrointestinal tract, craniofacial structures, and many others (see Chapter " The Neural Crest Stem Cells"). To study how KMT2D heterozygosity is responsible for craniofacial dysmorphisms, we decided to differentiate iPSCs to cranial NCSC using a two weeks protocol (Menendez et al., 2013) already established in the laboratory (Adamo et al., 2015). The purity of the differentiation was assessed by FACS analyses for surface markers NHK-1 and NGFR (Menendez et al., 2013), as depicted in Figure 24.
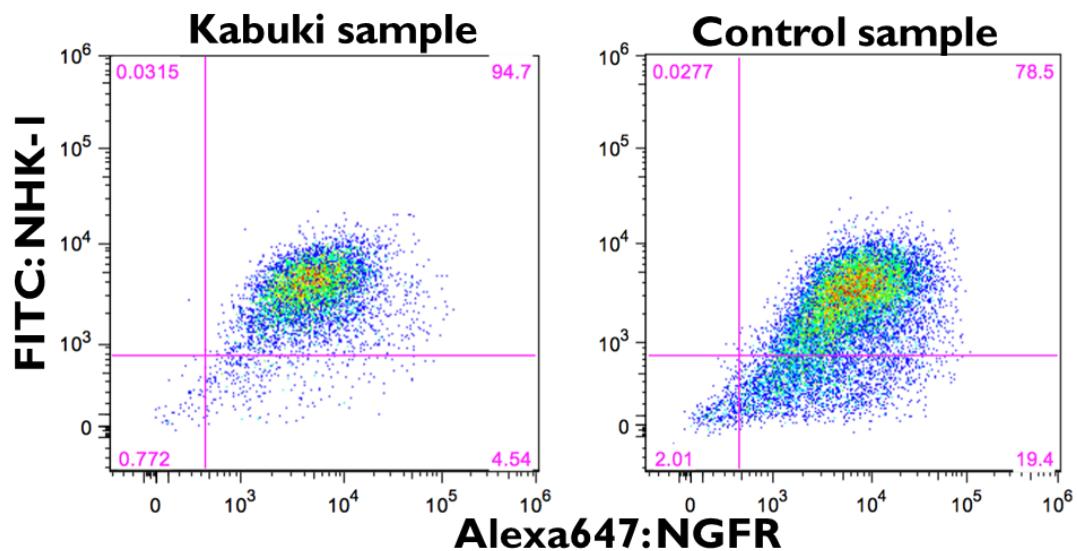
Figure 24 FACS analysis show positivity for NGFR and NHK-1

Transcriptional characterization of KS neural crest stem cells

NCSCs were subjected to transcriptome analysis and genome-wide characterization of H3K4me1 and H3K27Ac. PCA of transcriptome does not show any clear separation between KS and controls (Figure 25).
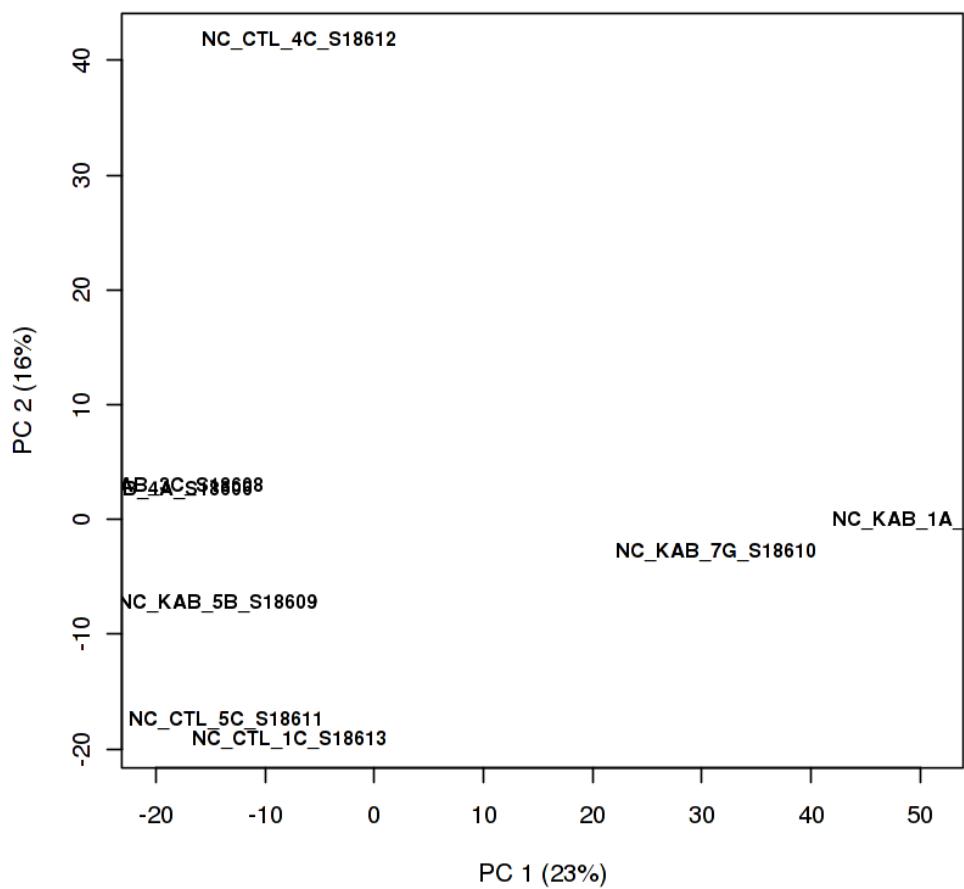
Figure 25 PCA of KS NCSC and matched controls

With an FDR threshold of 0,05 only 42 DEGs were identified (Figure 26).
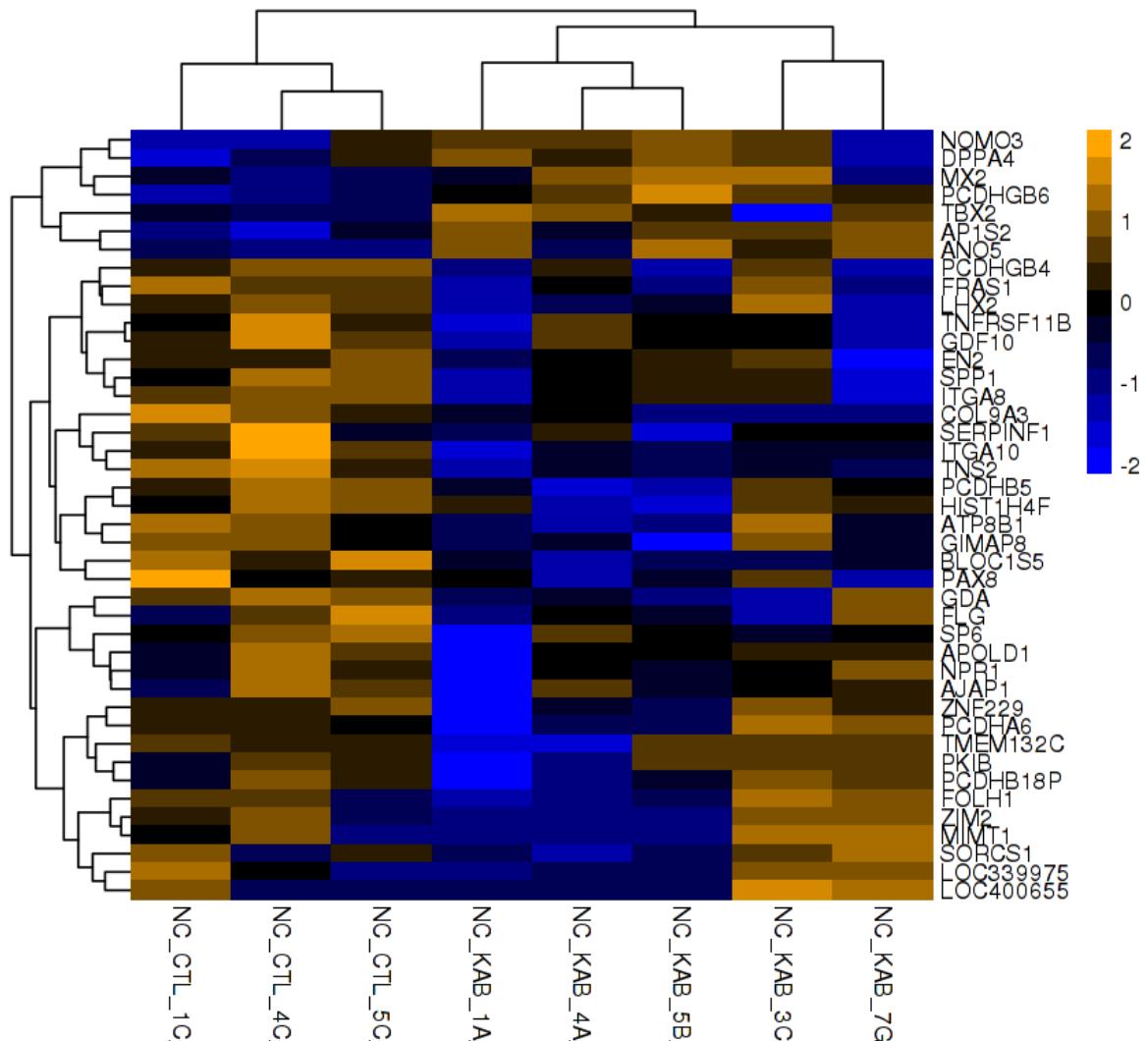
Figure 26 RNAseq of KS NCSCs and controls. FDR: 0,05. Scale log norm gene counts. Clustering: Pearson distance

Given the fact that two KS do not have an half matched healthy control to be matched with, the DEA was repeated without keeping in consideration the family covariate. In this case, the DEGs with an FDR of 0,05 were reduced to 31 (Figure 27). Lowering the FDR threshold to 0,1 and intersecting the between the two analyses, the following genes were overlapping: *'GDA', 'ITGA10', 'PCDHB5', 'MX2', 'PCDHGB4', 'BLOC1S5', 'COL9A3', 'AP1S2', 'PCDHGB6', 'ANO5', 'FRAS1', 'TNS2',*

'FGF2', 'KBTBD11', 'ABCG1'. These genes are enriched for the biological process "cell adhesion" with an FDR of 4.18E-05.
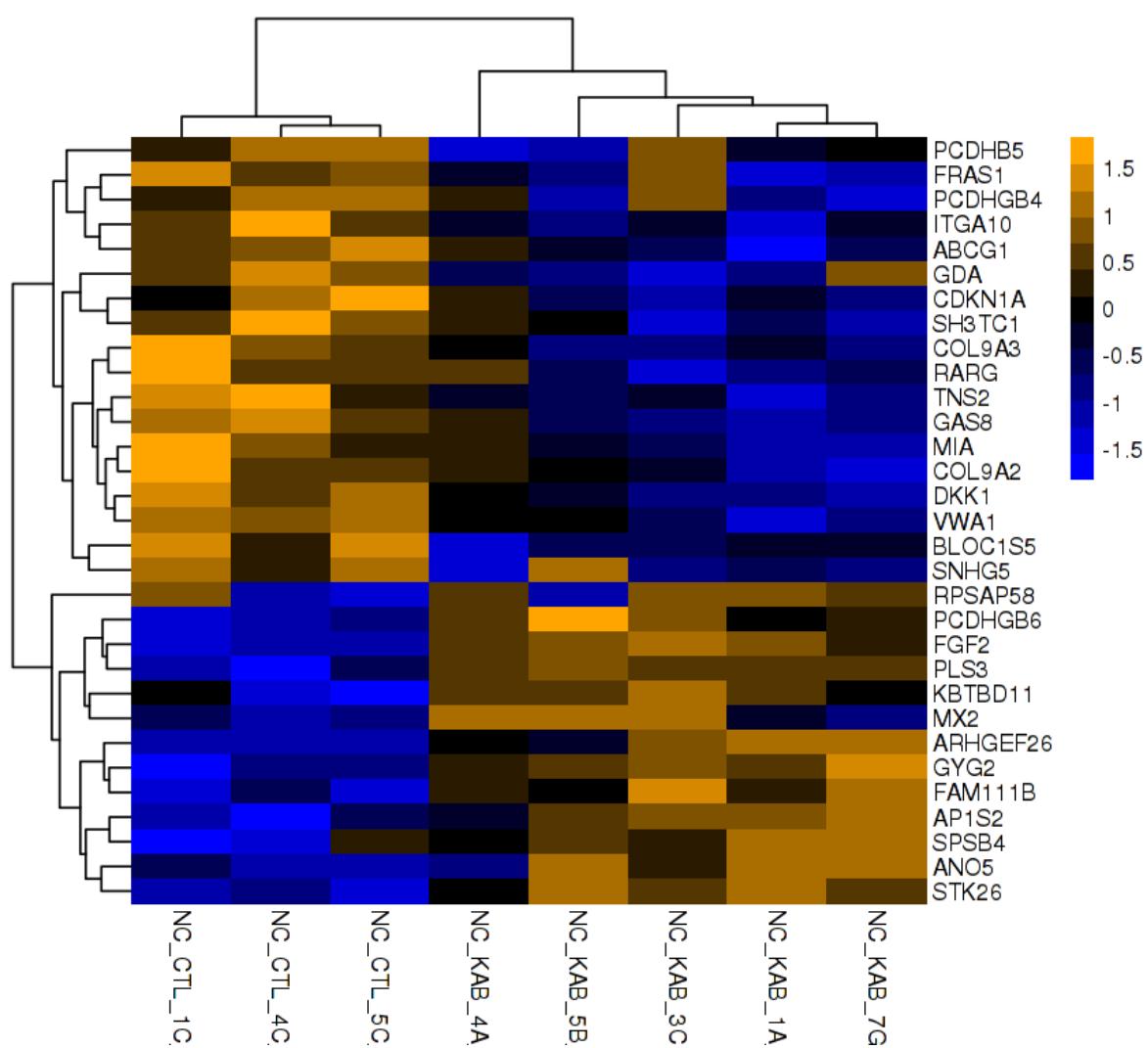
Figure 27 RNAseq of KS NCSCs and controls. FDR: 0,05. Scale log norm gene counts. Family was not included as covariate. Clustering: Pearson distance

The expression level of KMT2D, evaluated as RPKM, shows a reduction trend in KS when compared to the controls but it is not statistically significant (Figure 28). Worth of mention, in the analysis without the family as covariate, KDM6A has a logFC of 0,59, resulting overexpressed, with a P-value of 0,005, showing a potentially very relevant compensatory trend, despite the fact that its differential expression does not pass multiple test correction.

KMT2D RPKM

	logFC	logCPM	LR	PValue	FDR
KMT2D	-0.330029	6.975962	3.38345	0.06585406	0.9998489
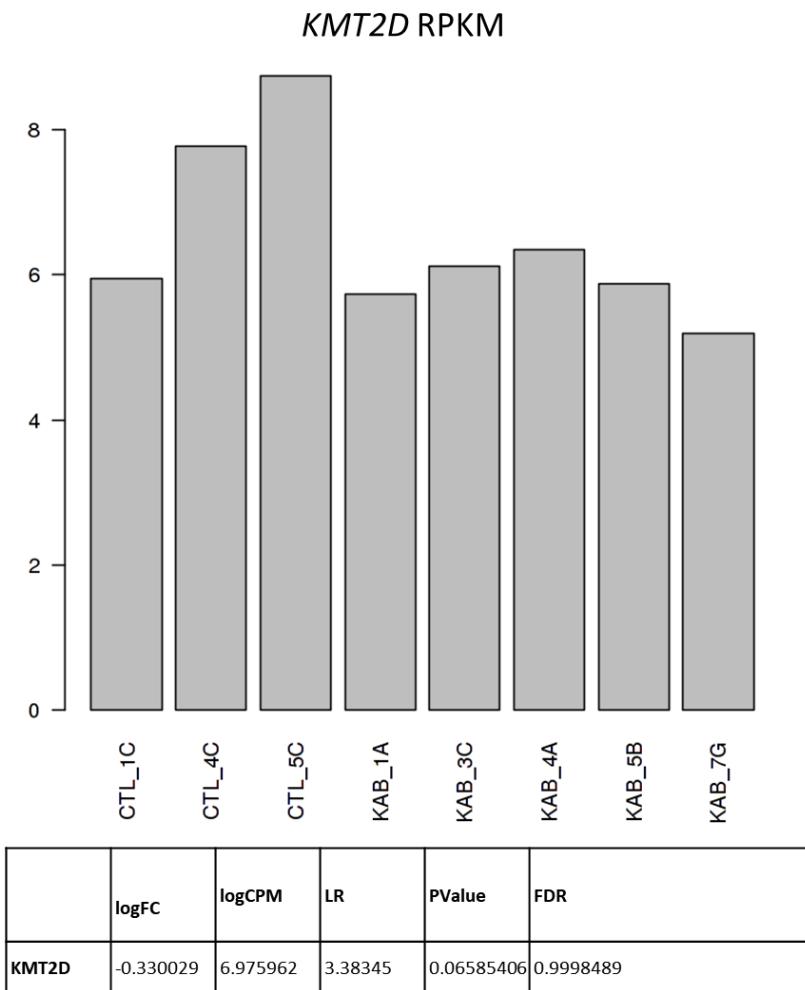
Figure 28 KMT2D RPKM in KS NCSC and controls

By selecting DEGS with a pValue < 0,05 it is possible to find enrichment for biological processes relevant for neural crest and the KS clinical traits such as "cardiac septum morphogenesis", "eyes morphogenesis", "kidneys morphogenesis" , and "central nervous system development" (Figure 29).

Figure 29 biological processes enrichment for DEGS (pValue <0,05) in KS and control NCSCs

H3K4me1 and H3K27Ac analysis of KS NCSCs

Following H3K4me1 ChIP-seq analysis on KS NCSC and controls, a Kabuki sample was excluded given the failure of immunoprecipitation (Figure 30). By normalizing on library size 32 differentially methylated regions (DMR) between controls and KS samples were identified, with an FDR of 0,1. To identify the putative target genes of these DMRs, given the lack of HiC dataset in human NCSC, this data was crossed with an internal set of presumptive enhancers, identified by high-throughput profiling of a large cohort of 32 NCSCs lines by means of both RNA-seq and ChIPseq experiments, thus generated by overlapping regions positive for H3K27Ac,

H3K4me1, excluding promoter regions (positive for H3K4me3), and choosing the closest expressed gene (without any distance limit).
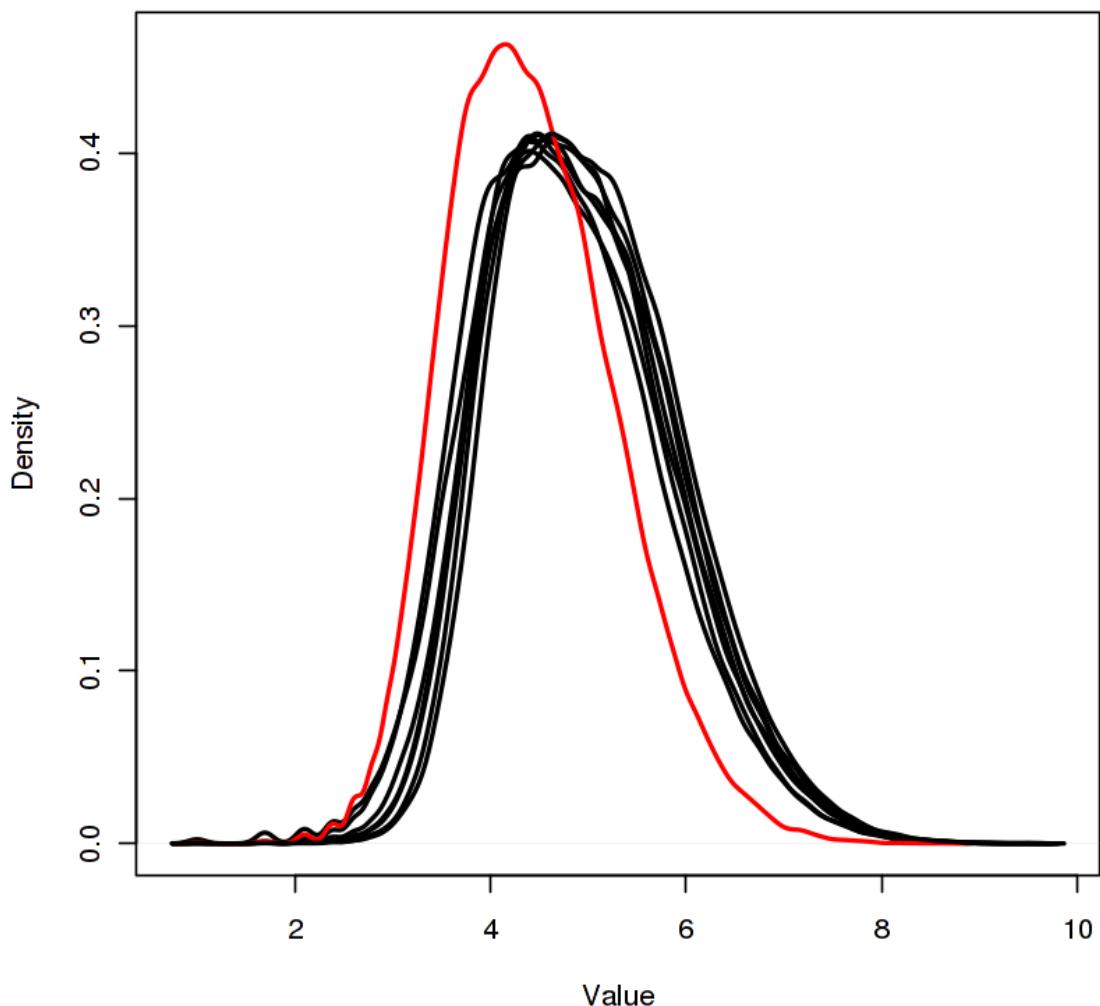
Figure 30 genomic density distribution of H3K4me1 in KS and controls NCSC. The red sample was excluded given the anomalous distribution

By doing so, a subset of putative active enhancers with diminished H3K4me1 signal was identified and crossed with the DEGs identified in NCSC. Without filtering for FDR, the genes *NEDD4L, PEG3, PHTF2, SLC15A4, SLITRK1, TMEM132B, TMEM132C, TXNL1, UNC5D, WSCD2,* and *ZIM2* resulted to have a putative active enhancer with a reduction in H3K4me1. The function of these genes is summarized in the following table:

Table 2 genes with a reduction of H3K4me1 in putative enhancers in NCSC

	protein class	OMIM disease	Phenotype MIM number
NEDD4L	E3 ubiquitin ligase	Periventricular nodular heterotopia 7	617201
PEG3	KRAB box transcription factor	Undifferentiated Embryonal Sarcoma Of The Liver and Glioma	NA
PHTF2	homeodomain transcription factor	NA	NA
SLC15A4	histidine transporter	Cri-Du-Chat Syndrome	123450
SLITRK1	receptor with neurite-modulating activity	Trichotillomania and Tourette syndrome	613229 and 137580
TMEM132B	transmembrane protein	NA	NA
TMEM132C	transmembrane protein	NA	NA
TXNL1	redox activity	Thymus Adenocarcinoma	NA
UNC5D	Netrin Receptor	Anisometropia	
WSCD2	NA	Porokeratosis	NA
ZIM2	KRAB box transcription factor	NA	NA

The same procedure was applied to H3K27Ac data. Quantitative analysis of H3K27Ac identified 11 regions differentially acetylated with FDR < 0,05. Increasing the DEGs FDR threshold to 0,1 and intersecting these differentially acetylated regions with our internal database, to identify regulated genes of these putative deregulated active enhancers, *PAX8* and *TMEM132C* were identified. Importantly, PAX8 is a transcription factor mutated in congenital hypothyroidism (OMIM: 218700).

Structural and physiological characterization of upper layer cortical neurons obtained from KS iPSCs

To probe the spontaneous electrophysiological activity and to characterize the neuronal network structure, upper layer cortical neurons were obtained by overexpressing Ngn2 (Y. Zhang et al., 2013) (hereafter referred to as iNeurons) from 4 KS patients, 3 half-matched healthy controls, and one unmatched control. Immunofluorescence showed positivity for general neuronal markers MAP2, TUJ1 (Figure 31), and for glutamatergic neurons VGLUT1 (Figure 32) already after 12 days after doxycycline administration.
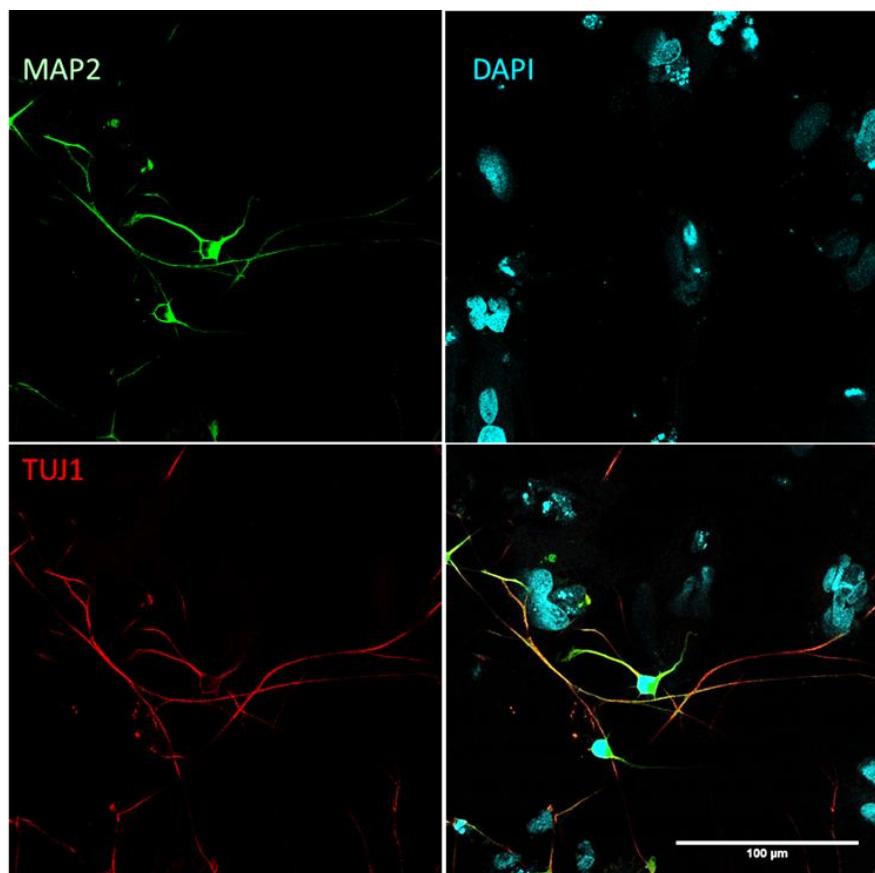
Figure 31 TUJ1 and MAP2 stainings of Ngn2 neurons. Day 12 after Dox induction. Confocal 63x magnification. Co-culture with astrocytes
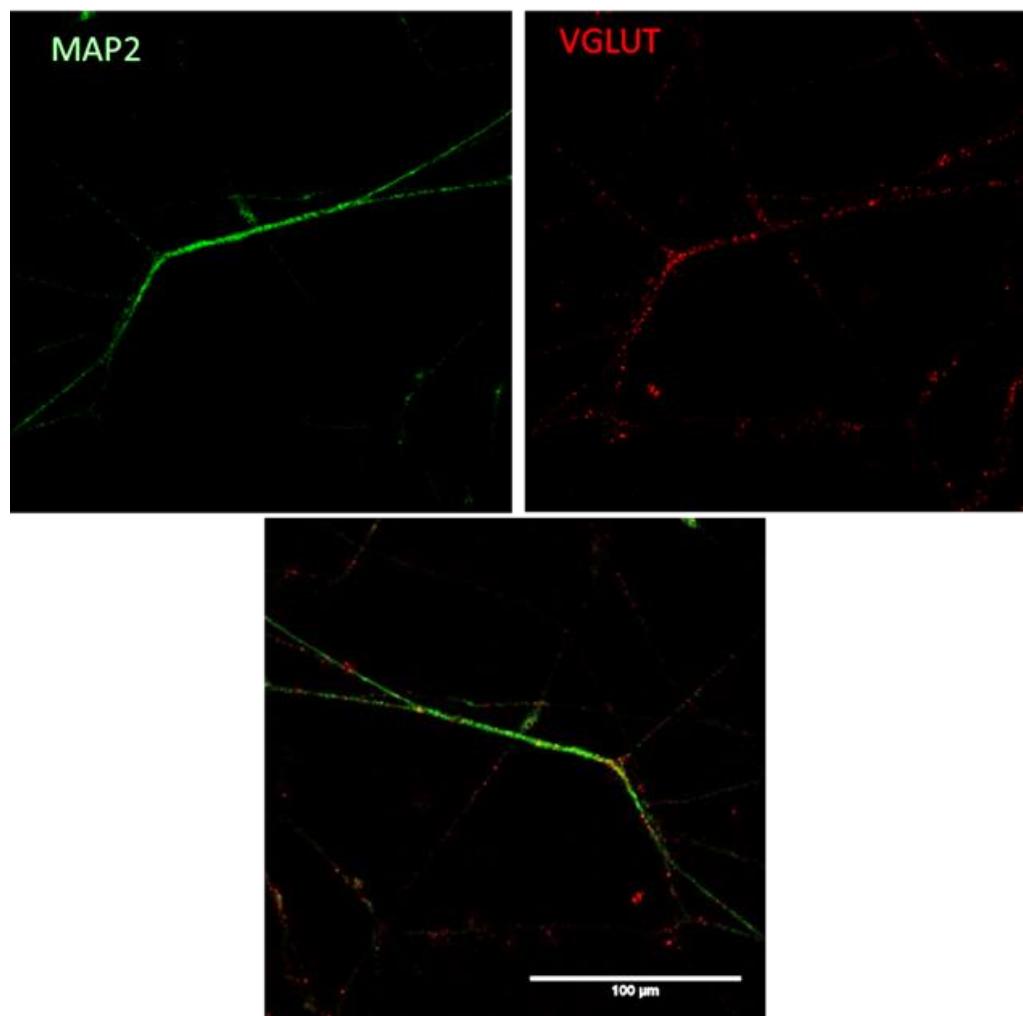
Figure 32 VGLUT1 and MAP2 stainings of Ngn2 neurons. Day 12 after Dox induction. Confocal 63x magnification. Co-culture with astrocytes.

At day 23 of differentiation iNeurons showed mature morphology and morphological parameters, such as number of primary dendrites, nodes, and primary endings, did not show any statistical difference between KS and control samples (Figure 33).
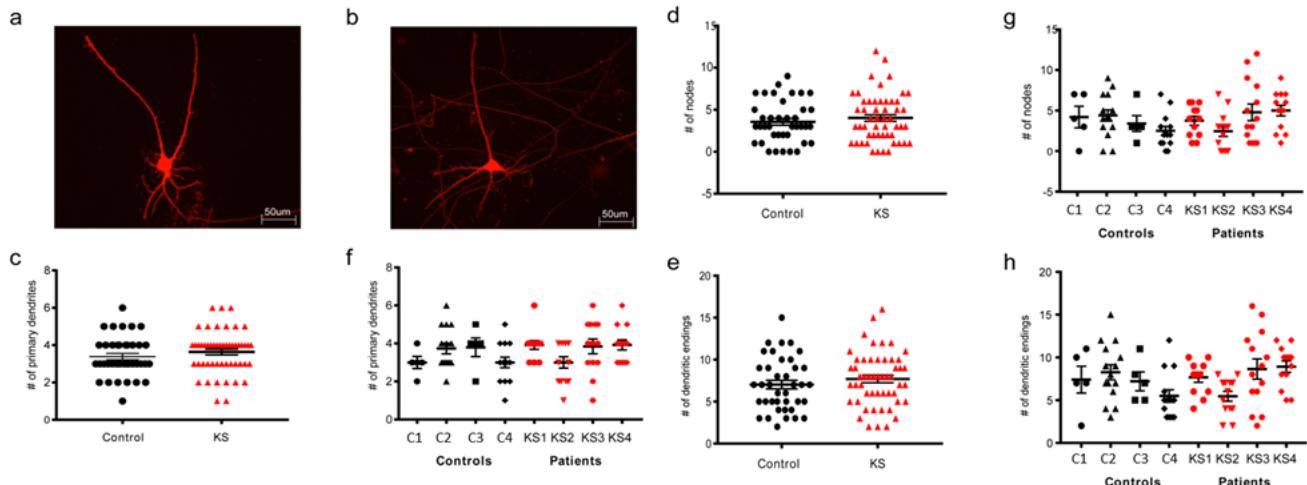
Figure 33 Structural analysis of neurons. a-b) Representative images of a) control and b) KS neurons transfected with dsRed (scale: 50 μm). Graphs c-e) showing the c) number of primary dendrites, d) nodes and e) dendritic endings in control (black, n=39) and Kabuki neurons (red, n=52) derived from hiPSCs (i.e. pooled results). Graphs f-h) showing differences in f) number of primary dendrites, g) nodes and h) endings in neurons derived from 4 control (black, C1 n=5, C2 n=15, C3 n=5, C4 n=14) and 4 Kabuki (red, KS1 n=12, KS2 n=13, KS3 n=14, KS4 n=13) hiPSCs lines. Data represent means ± SEM. Statistics: normality test, Kruskal-Walis Test, post-hoc Bonferroni correction.

From micro-electrode arrays (MEAs) analysis it is possible to record spontaneous spikes, burst, and network bursts, which are indicative of a mature organized neuronal network Figure 34.
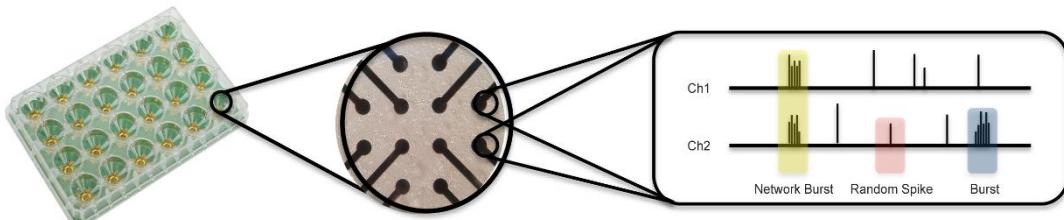
Figure 34 Schematic depiction of a MEA with 12 electrodes and definition of spike, burst and network burst.

Already during the second week in vitro, which start is defined as the moment of plating on MEAs, the neurons derived from healthy subjects formed functionally active neuronal networks, showing spontaneous events (i.e. spike and burst, Figure 34). Late in development (i.e. fifth week in vitro) the neuronal network showed high level of spontaneous activity as well as regular network bursting pattern (i.e. synchronous events involving almost all channels of the MEAs, highlighted in grey in Figure 35A). Neurons derived from KS patients established spiking activity during

early network development as well as network burst involving most of the channels of the MEAs later in development (Figure 35B). During the fifth week in vitro, at the population level KS patient-derived neuronal networks exhibited a global level of activity that was slightly higher than controls (i.e. firing rate Figure 35C, p=0.049). The level of synchronous activity exhibited by KS-derived neuronal network was comparable to control (Figure 35D, p=0.02), too. Although the level of both global and synchronous activity was only slightly impaired, the pattern of activity exhibited by KS patient-derived neuronal networks was very different compared to control condition. First, the percentage of random spike (i.e. events not organized within a burst) was statistically lower in KS-derived neuronal compared to control (Figure 35E, p<0.00025). Furthermore, the network burst appeared with longer durations compared to controls (Figure 35F, p<0.00025). The interval occurring between two consecutive network bursts was also longer compared to control condition (i.e. inter burst interval, Figure 35G, statistically different before Bonferroni correction). Finally, the pattern of synchronous activity (i.e. in terms of timing and organization) was impaired in KS-derived neuronal networks. In particular, KS patient-derived neuronal networks exhibited an irregular bursting pattern, as shown in the raster plots in Figure 35A-B. Then, the organization of the burst involved in a synchronous event was different. In control condition, a network burst was composed but single burst appearing simultaneously in most of the channel (see raw data highlighted in black in Figure 35A). Instead, the network burst exhibited by KS-derived neuronal networks was composed by "mini-bursts" (i.e. 4 "mini-bursts", see raw data highlighted in red in Figure 35A and Figure 35H). The results indicate that KS-derived neuronal cultures showed an altered pattern of activity in excitatory neuronal networks.
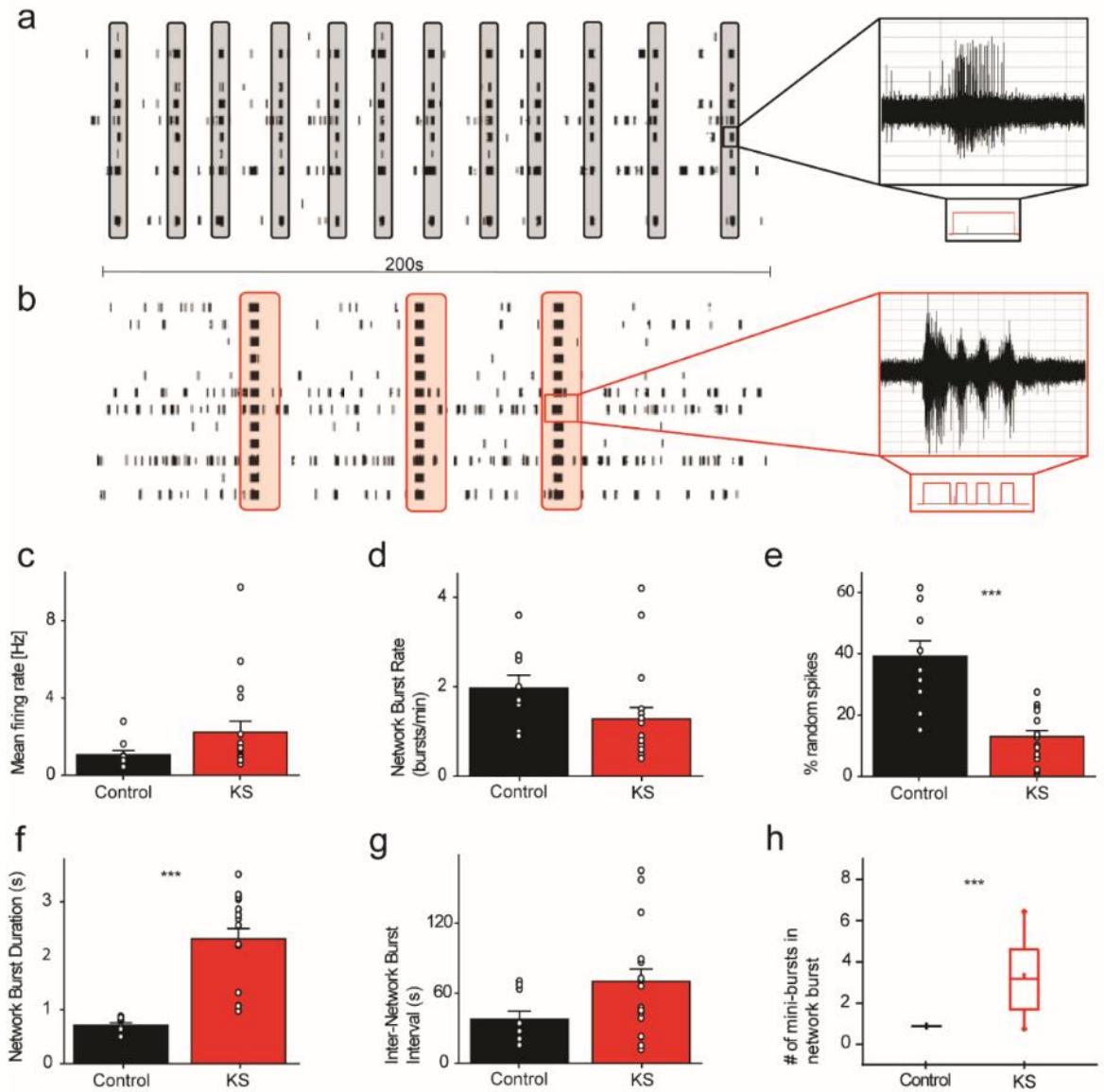
Figure 35 Spontaneous electrophysiological activity of neuronal networks grown on MEAs. A) raw data of recordings from representative control; B) raw data of recordings from representative KS; C) mean firing rate; D) network burst rate; E) percentage of random spikes; F) network burst duration; G) inter-network burst interval; H) average number of mini-burst observed.

Transcriptional characterization of upper layer cortical neurons obtained from KS iPSCs

iNeurons at day 31 were subjected to RNA-seq to identify molecular dysfunction that could explain the phenotype observed by MEA analysis, which was not explained by the neuronal network architecture. First, to confirm neuronal identity of cortical neurons, the expression level of a panel of genes, representative of glutamatergic

neurons, glycinergic and GABA-ergic interneurons, dopaminergic, and cholinergic neurons, housekeeping genes, pluripotency genes, several markers specific of cortical layers I-VI, markers of neurogenesis, post-mitotic neurons, glia, and expression for chromatin remodelers relevant for neurodevelopmental syndromes, was investigated in this iNeuron cohort as depicted in Figure 36. Expression levels of selected markers, listed in Table 3, indicate neuronal identity of upper layer cortical excitatory neurons.
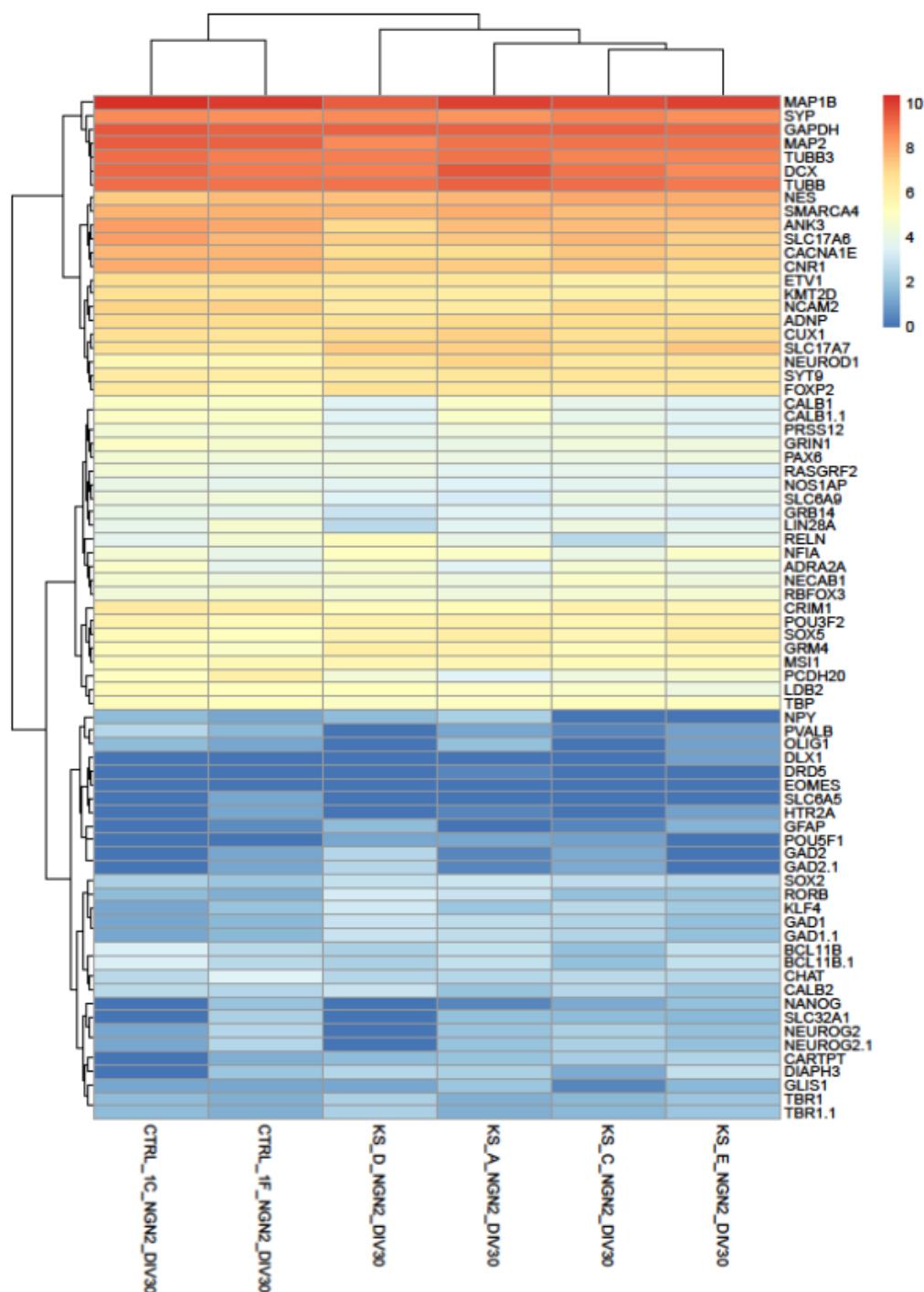
Figure 36 Expression of genes used to check the identity of neurons at day 30 of differentiation. Blue genes appear as not expressed and are filtered out during differential expression analysis.

Table 3 Gene selected to assess transcriptional identity of Ngn2 neurons

Function		gene	OMIM
Cortical layers	Layer I	*RELN*	**600514**
	Layer II-IV	*POU3F2*	600494
	Layer II-III	*CUX1*	116896
		RASGRF2	606614
	Upper Layer neuron generation	*CACNA1E*	601013
		PRSS12	606709
		CARTPT	602606
	Layer IV	*PCDH20*	614449
	Layer IV	*GRM4*	604100
		CALB1	114050
		RORB	601972
		NECAB1	618130
	Layer V	*ETV1*	600541
		SOX5	604975
		BCL11B (CTIP2)	606558
		LDB2	603450
		GRB14	601524
		SYT9	613528
		CRIM1	606189
	Layer VI	*TBR1*	604616
		FOXP2	605317
		BCL11B (CTIP2)	606558
		NFIA	600727
		NOS1AP	605551
		ADRA2A	104210
Neurogenesis	Early in neuronal development	*NES*	600915
		NCAM2	602040
		MSI1	603328
	Neuronal precursor cells and immature neurons	*DCX*	300121
	Dendrite structures	*MAP2*	157130
		MAP1B	157129
	Involved in neurogenesis, axon guidance, and maintenance	*TUBB3 (Tuj1)*	602661
	Axon formation	*ANK3*	600465

Post-mitotic neurons		*RBFOX3 (NeuN)*	616999
Post-synaptic marker	Synaptic vesicle endocytosis	*SYP*	313475
Glutamatergic neurons	Vesicular glutamate transporters	*SLC17A6 (VGLUT2)*	607563
		SLC17A7 (VGLUT1)	605208
	NMDA glutamate receptor	*GRIN1*	138249
	Glutamatergic neurogenesis	*NEUROG2 (Ngn2)*	606624
		PAX6	607108
		EOMES (TBR2)	604615
Glycinergic and GABAergic neurons	Glycinergic neurons	*SLC6A9*	601724
		SLC6A5	604159
	Gabaergic neurons	*GAD1*	605363
		GAD2	138275
	Transports both GABA and glycine into synaptic vesicles	*SLC32A1*	616440
Dopaminergic neurons	Differentiation and survival of midbrain dopaminergic neurons	*NEUROG2*	606624
		DRD5	126453
Cholinergic neurons		*CHAT*	118490
Interneurons	Cortical interneuron subclasses	*CALB1*	114050
		CALB2	114051
	Inhibitory axon terminals impinging upon hippocampal interneurons	*CNR1*	114610
	Cortical and hippocampal interneurons	*DLX1*	600029
	somatostatin-containing interneurons in stratum oriens alveus	*NPY*	162640
	striatal interneuron	*PVALB*	168890
	neocortical interneurons	*HTR2A*	182135
Pluripotency and neuronal progenitors	Pluripotency	*NANOG*	607937
		LIN28A	611043
		KLF4	602253

68

		GLIS1	610378
		POU5F1 (OCT4)	164177
	Pluripotency and neural stem cells	*SOX2*	184429
	Neuronal progenitors	*DIAPH3*	614567
Housekeeping genes	highly expressed HK	*GAPDH*	138400
		TUBB	191130
	lowly expressed HK	*TBP*	600075
Markers of oligodendrocytes and glia		*GFAP*	137780
		OLIG1	606385
Chromatin remodelers involved in ND		*KMT2D*	602113
		SMARCA4	603254
		ADNP	611386

The male patient was excluded from the analysis given the low quality of the sequencing and possible cross-contamination during flow-cell preparation. Consequently, to have the cohort balanced for the gender, also the male half-matched unaffected control has been excluded. Therefore, to increase the robustness of the Differential expression analysis (DEA), four independent control samples were included in the analysis. In this cell type, 1318 DEGs were identified (Figure 37), with an FDR of 0,05. These genes are enriched in categories that are strictly related to neuronal function (Figure 38) and to translation. Solute carriers and axonal morphogenesis genes appear to be deregulated. Consistent with the considerable transcriptional alteration, here *KMT2D* results more than halved (logFC -0,79, FDR: 0.0004853473). On the other hand, KDM6A here is significantly upregulated (logFC: 0.9506504, FDR: 7.488772e-05).
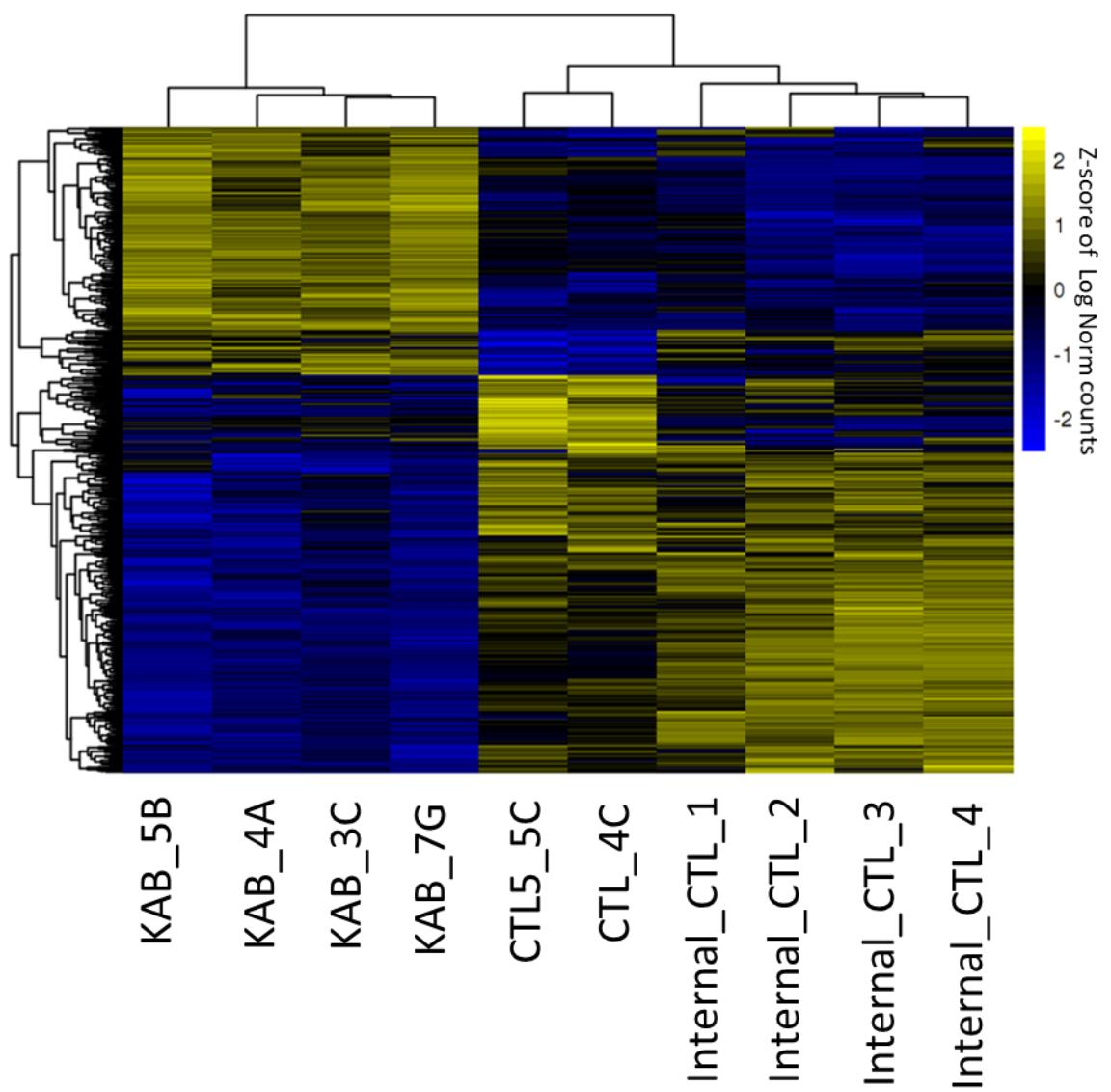
Figure 37 DEGs of KS and controls cortical neurons at day 30. FDR: 0,05; Scale: log norm gene counts.

Figure 38 Gene Ontology for biological processes iNeurons DEGs.
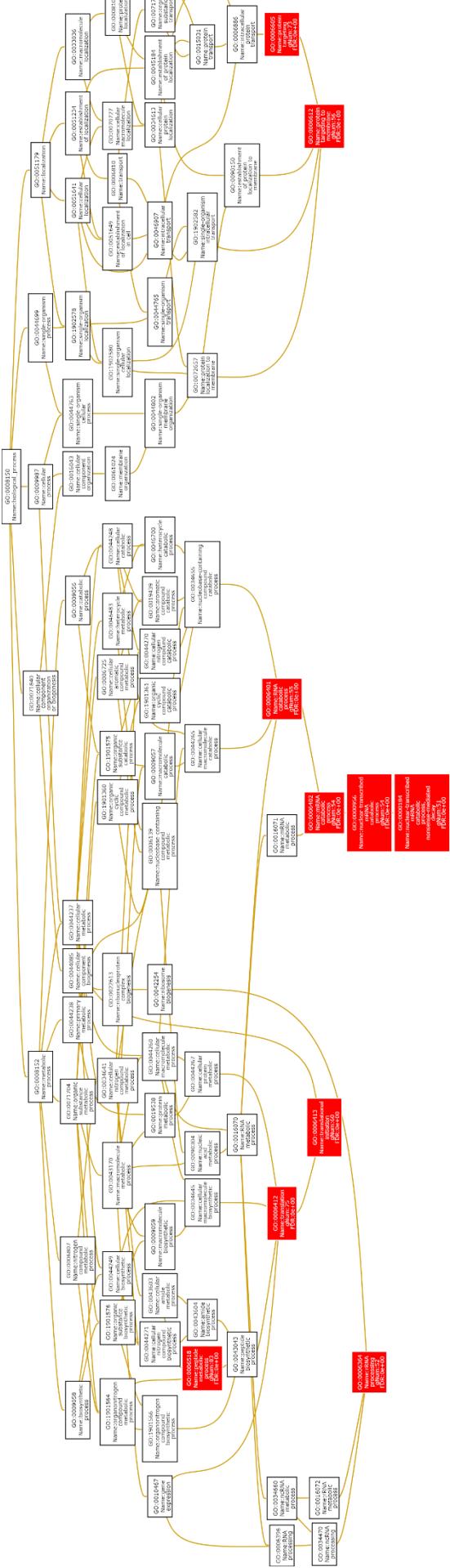
To better understand transcriptional alteration in iNeurons, several tools for gene ontology analysis were adopted. Moreover, overexpressed and downregulated genes were also analyzed separately. By doing so, it is possible to see that overexpressed genes are enriched in categories related to translation initiation and to nonsense-mediated decay (Figure 39), downregulated genes are specifically enriched for categories specific for neuronal function and signaling (Figure 40).

Figure 39 Enrichment results for biological processes in upregulated genes. Overrepresentation Enrichment Analysis using Webgestalt. Reference gene list: genes expressed in Ngn2 neurons.
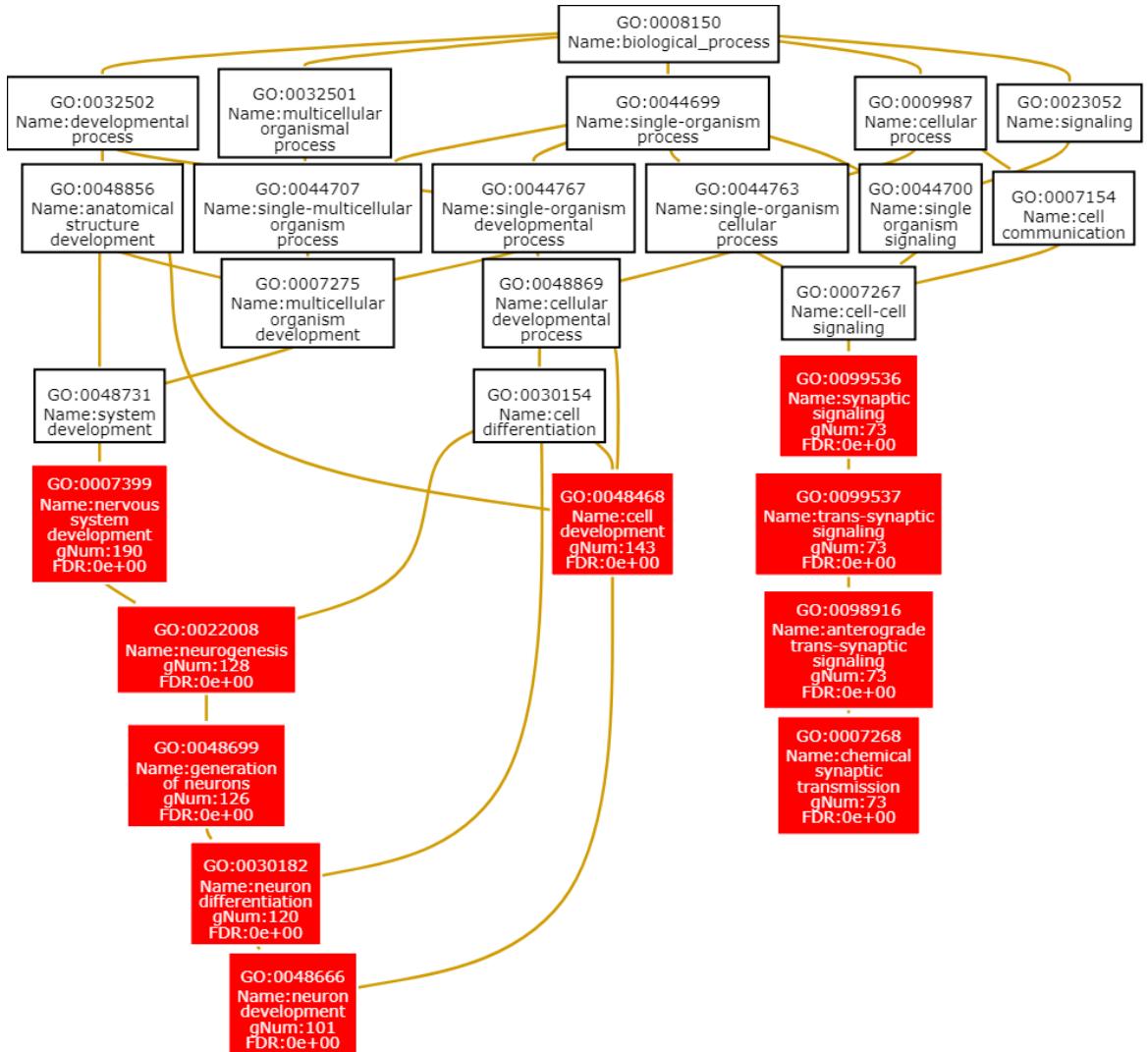
Figure 40 Enrichment results for biological processes in downregulated genes. Overrepresentation Enrichment Analysis using Webgestalt. reference gene list: genes expressed in Ngn2 neurons

Upregulation of translation and translation initiation was maintained also using the IPA tool (Qiagen). Moreover, categories enriched for downregulated genes are again enriched for neuron activity for excitatory neurons Figure 41.
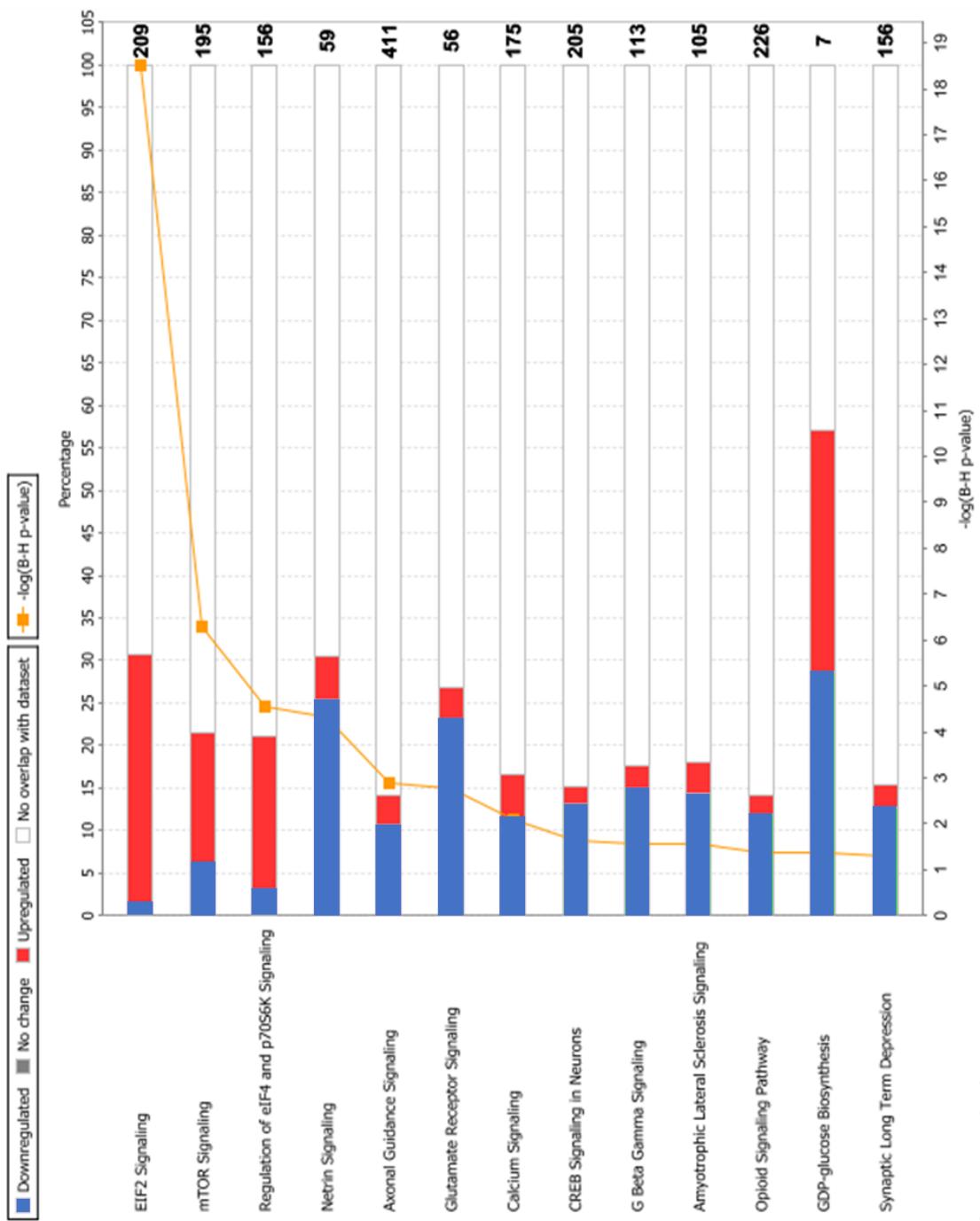
Figure 41 IPA analysis of iNeurons DEGs (FDR 0,05). Colors were manually changed to fit color blind visualization. Yellow line: multiple-test correction significance

Also, using IPA is possible to identify enrichment for disease in gene categories. By separating upregulated and downregulated genes, it is possible to uncover an interesting divergence between the two DEGs subsets. Upregulated genes are enriched for neuronal-related diseases, and tumor suppressor genes such as *TP53*, *ATM*, *BAX*, KLF10, and *CDK1A* (known as p21). On the other hand, downregulated genes are enriched for melanoma and gastrointestinal tumors, as showed in the

Table 4, in particular, RET, MET, WNT5A, INO80D, KRAS, and RAS-related genes KSR2, RAB27B, RAB30, RAB31, RAB33A, RAB3B, RAB3IP, RAB6B, RABGAP1L, RAC3, RALGPS1, RAP1GAP2, RAPGEF2, RAPH1, RASSF5, and the transcription factors POU5F1B, POU4F1, SOX11, SOX4. Also, the tumor suppressor CHD5, which protects from neuroblastoma, is downregulated.

Table 4 Top scoring "Disease and function" of IPA (Qiagen) analysis

downregulated genes		upregulated genes	
Diseases or Functions Annotation	p-Value	Diseases or Functions Annotation	p-Value
Cutaneous melanoma	1,29E-32	Cell death of osteosarcoma cells	2,81E-28
Intestinal carcinoma	1,29E-32	Cell death of cancer cells	2,23E-18
Large intestine adenocarcinoma	1,29E-32	Cell death of tumor cells	5,47E-18
Large intestine carcinoma	1,42E-32	Translation	9,73E-18
Gastrointestinal adenocarcinoma	9,96E-32	Translation of mRNA	3,79E-17
Gastrointestinal carcinoma	2,06E-31	Translation of protein	8,38E-17
Skin cancer	4,60E-31	Expression of mRNA	9,27E-16
Gastrointestinal tumor	6,46E-31	Metabolism of protein	4,05E-15
Intestinal tumor	6,57E-31	Synthesis of protein	2,29E-14
Skin lesion	6,57E-31	Expression of protein	6,71E-14
Skin tumor	7,30E-31	Cancer	7,20E-13
Melanoma	8,84E-31	Solid tumor	1,52E-12
Large intestine neoplasm	8,84E-31	Non-melanoma solid tumor	2,18E-12
Liver carcinoma	1,08E-30	Tumorigenesis of tissue	2,88E-12
Liver tumor	1,35E-30	Cell death	6,14E-12
Intestinal cancer	6,37E-30	Chemoresistant acute myeloid leukemia	6,84E-12
Development of neurons	7,18E-30	Abdominal neoplasm	7,53E-12

Given the high number of DEGs is possible to divide the genes into clusters according to their trend across samples. By applying k-means clustering, 6 clusters were identified. A master regulator analysis was performed in each cluster to identify putative regulators upstream of the DEGs composing each cluster. Interestingly, the cluster 3, included RBBP5 as upstream regulators of 27 genes of this cluster (FDR ~ 0,05). Among these 27 genes, there is an enrichment of RNA binding proteins (panther ORA, FDR: 2.35E-02), given by the five genes: NSA2, G3BP2, RPF1, ZC3H15, and ALKBH8.

Genome-wide H3K4me1 and H3K27ac analysis of iNeurons

From the very same dishes used to collect RNA, chromatin was collected for ChIP-seq analysis. There is no clear separation in genomic density distribution between control and KS samples (Figure 42) To link putative unregulated active enhancers to target genes, for this cell type it has been used the dataset for hippocampal neurons of the 4D project (Dekker et al., 2017). Of all expressed genes, the differential Hek4me1 regions are associated to 496 genes; the overlap with the DEGs (FDR 0,05) is of 47 genes, of which 30 are predicted to be bound at the enhancer by RBBP5, a partner of KMT2D (hypergeometric test p-Value: 2.42356254688917e-81).
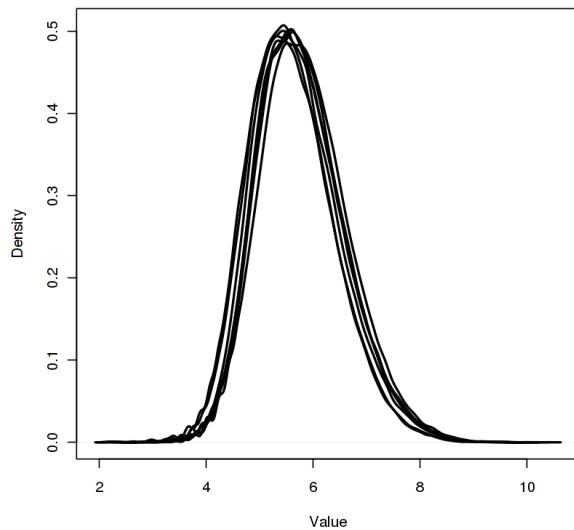
Figure 42 genomic density of H3K4me1 in KS and control iNeurons, DIV 30.

Within the target genes of predicted enhancers to have H3K4me1 dysregulation, there is a biological process enrichment for biological categories (Figure 43)
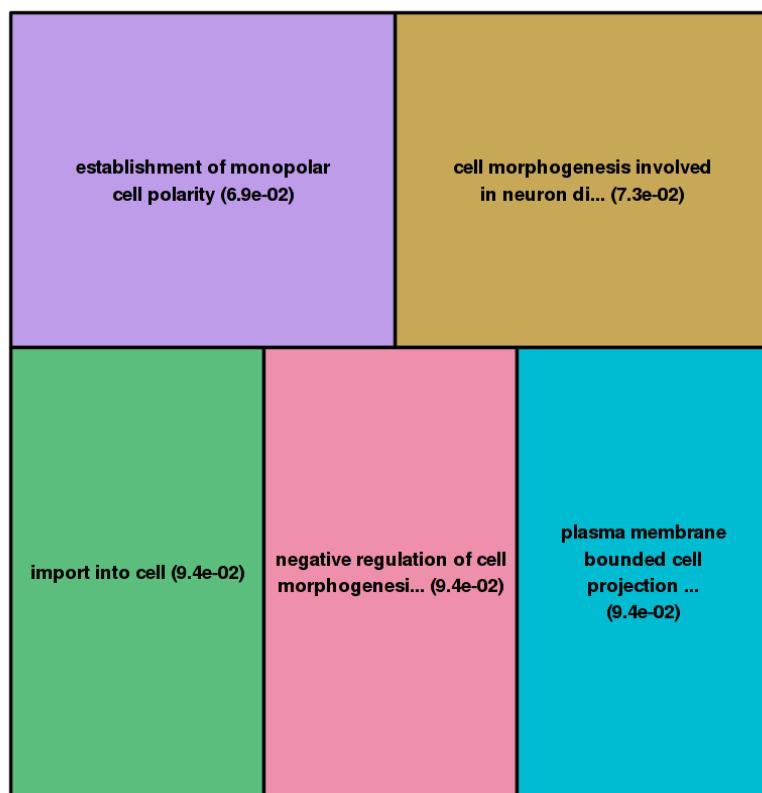
Figure 43 Biological processes treemap for genes of which putative enhancers have a dysregulation in H3K4me1 in Ks iNeurons at day 30. Values: FDR 0,05

	upregulated genes	downregulated genes	In all expressed genes
gain in H3K4me1	29		640
loss of H3K4me1		8	104
total DEGs FDR 0,5	565	753	

After normalizing on library size, with the exception of one sample, KS samples show an increase of H3K27Ac (Figure 45
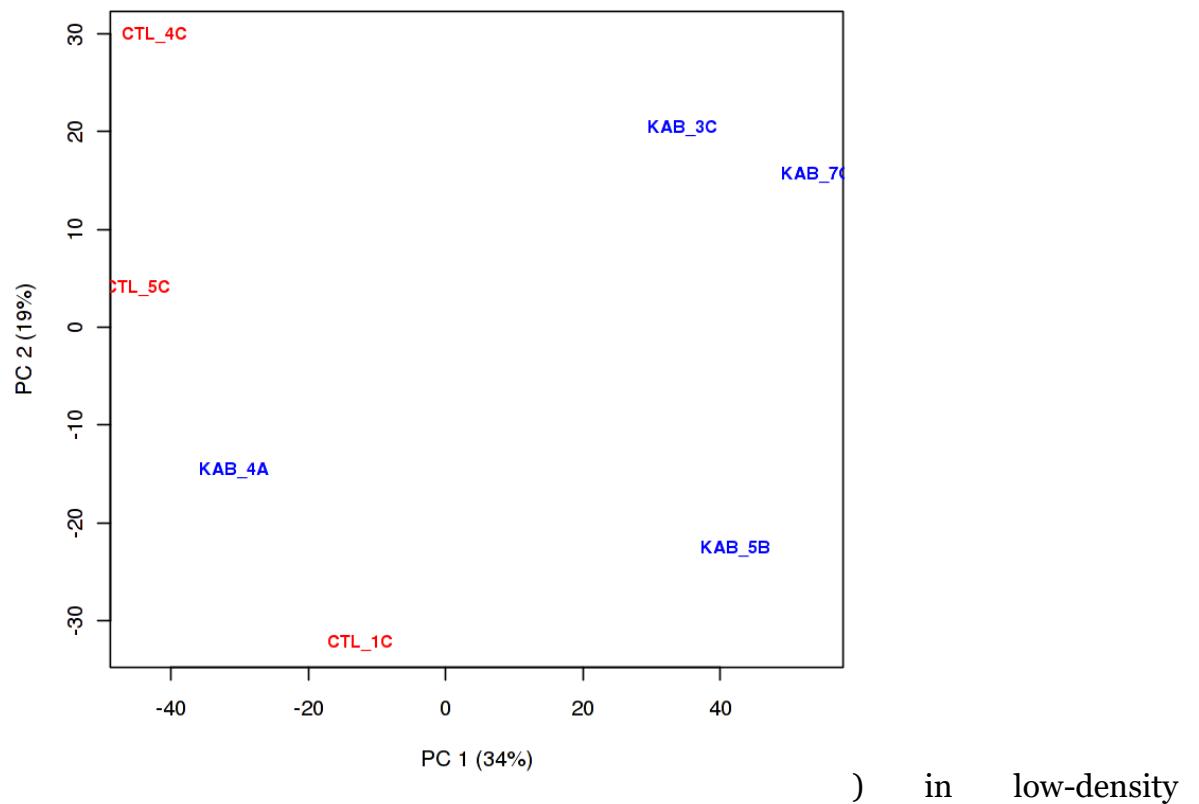
) in low-density regions, while showed a reduction in the regions with the highest density. The outlier KS sample also cluster together with controls in PCA, which with the exception of this sample, separated the samples on PC1 according to the genotype.
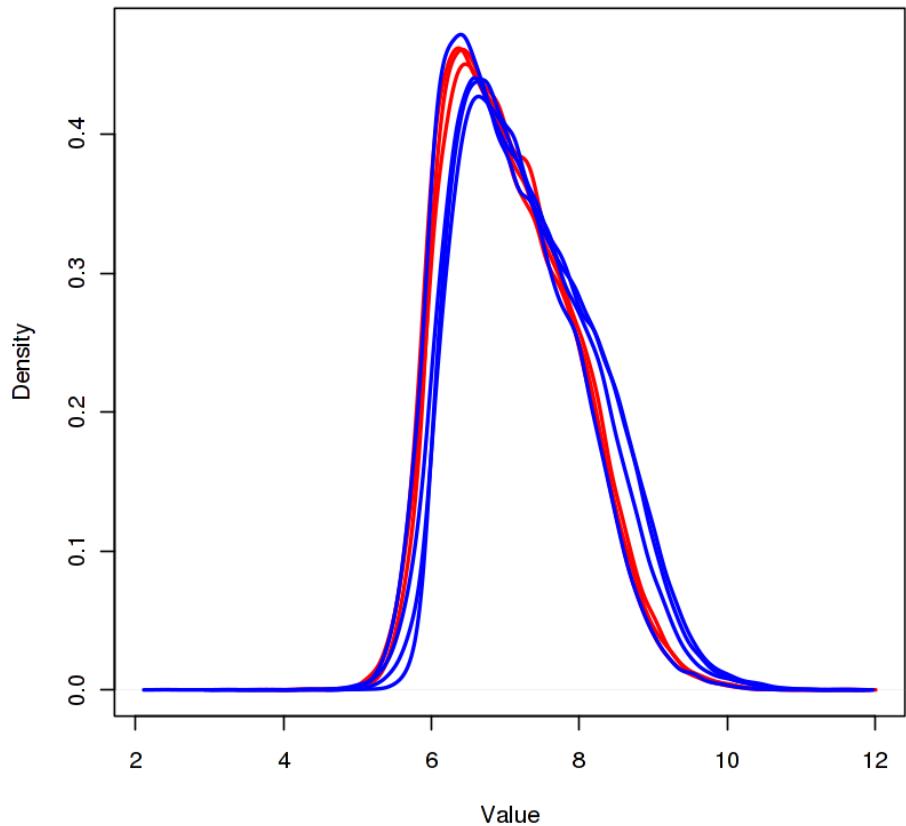
Figure 44 genomic density of H3K27Ac in KS and control iNeurons, DIV 30. Red: control samples; Blue: KS samples.
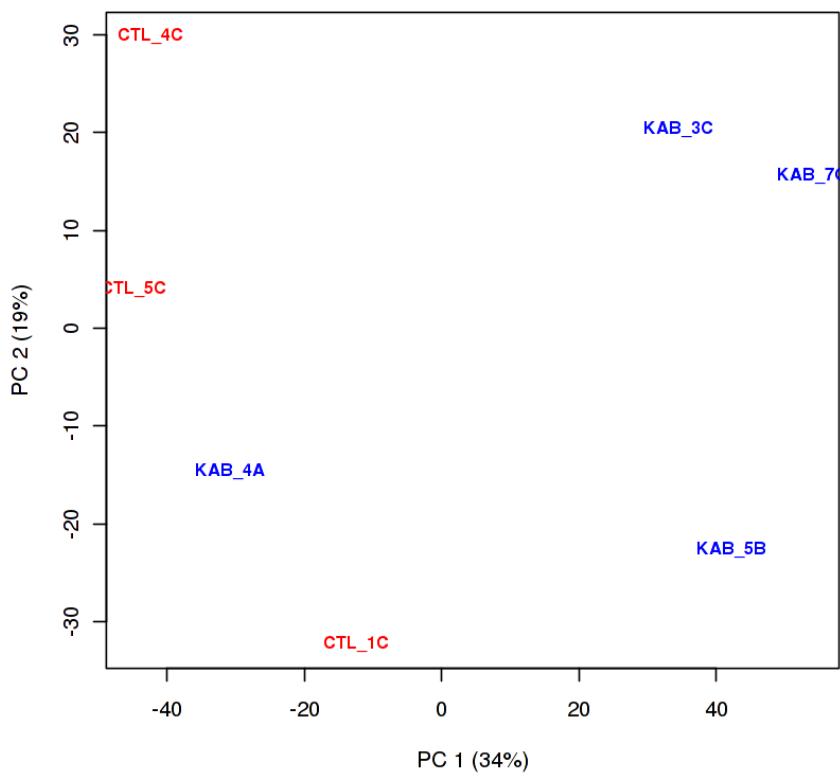
Figure 45 PCA analysis of KS and controls H3K27Ac in iNeurons at day 30. Red: Control samples, Blue: KS samples

Given the distribution of density regions with gain and loss of H3K27Ac have been analyzed separately. Gene identified to be target genes of putative dysregulated enhancers were intersected with DEGs (FDR 0,1). The following table summarizes the results of the intersections:

Table 5 Intersection with target regions of altered H3K27Ac with differentially expressed genes (FDR < 0,1)

	upregulated genes	downregulated genes	In all expressed genes
gain in H3K27Ac	221		1874
loss of H3K27Ac		141	963
total DEGs FDR 0,1	877	1055	

Genes target with a loss of H3K27Ac are enriched for biological processes specific for neuronal function (Figure 46). Conversely, putative target genes with a gain of H3K27Ac are enriched in biological processes unrelated to neuronal physiology (Figure 47).
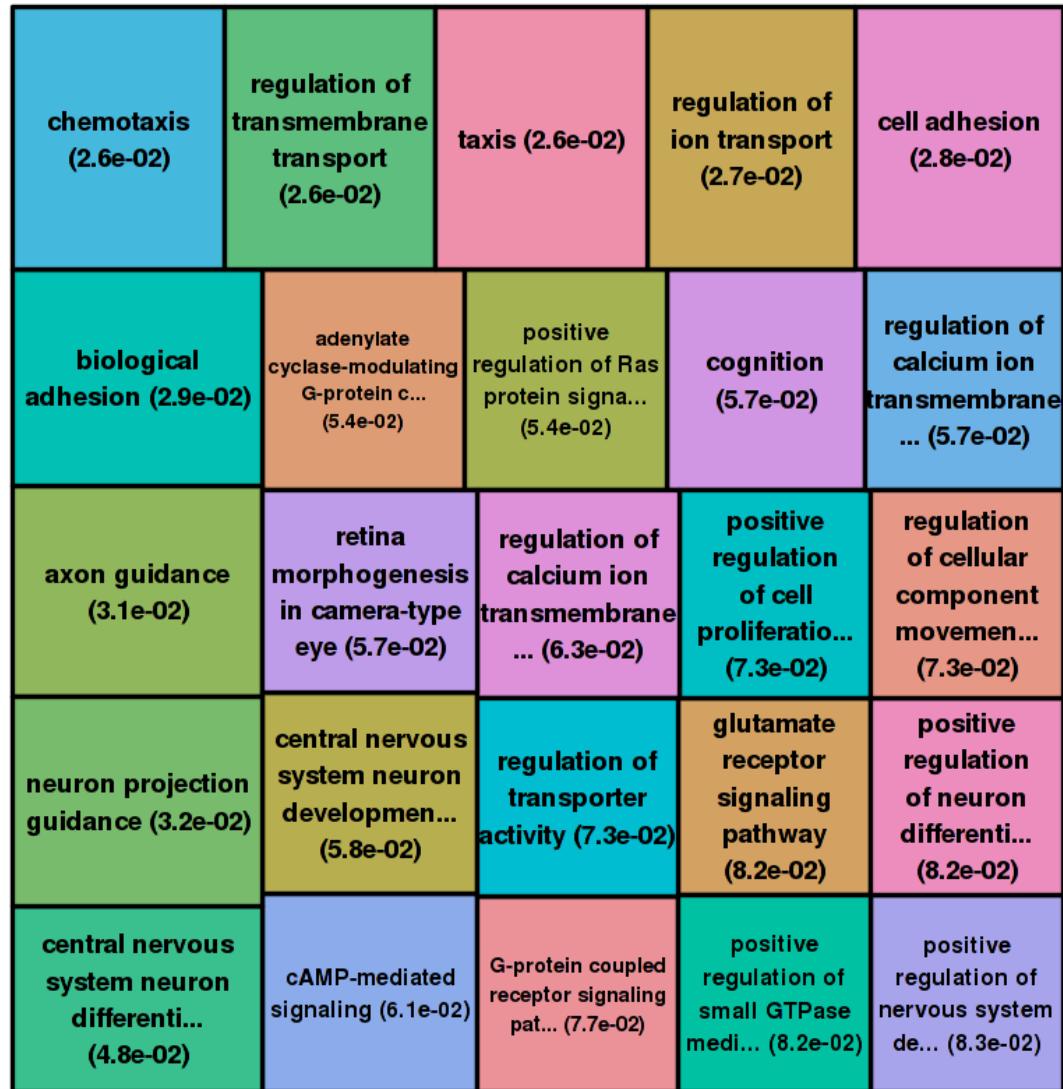
Figure 46 Biological processes treemap for genes of which putative enhancers have a loss in H3K27Ac. Values: FDR

Figure 47 Biological processes treemap for genes of which putative enhancers have a gain in H3K27Ac. Values: FDR

The intersection between regions regulated by RBBP5 and DEGS (FDR < 0,1) with matched putative enhancer with H3K27ac alterations, with identified 98 genes with a loss of H3K27Ac and 140 genes with a gain in H3K27Ac in the enhancer regions. The genes with a loss of H3K27ac are enriched for biological processes specific for neuronal physiology (Figure 48). Genes with a gain of H3K27ac are enriched for "cargo loading into COPII-coated vesicle" (FDR 3.3e-02).
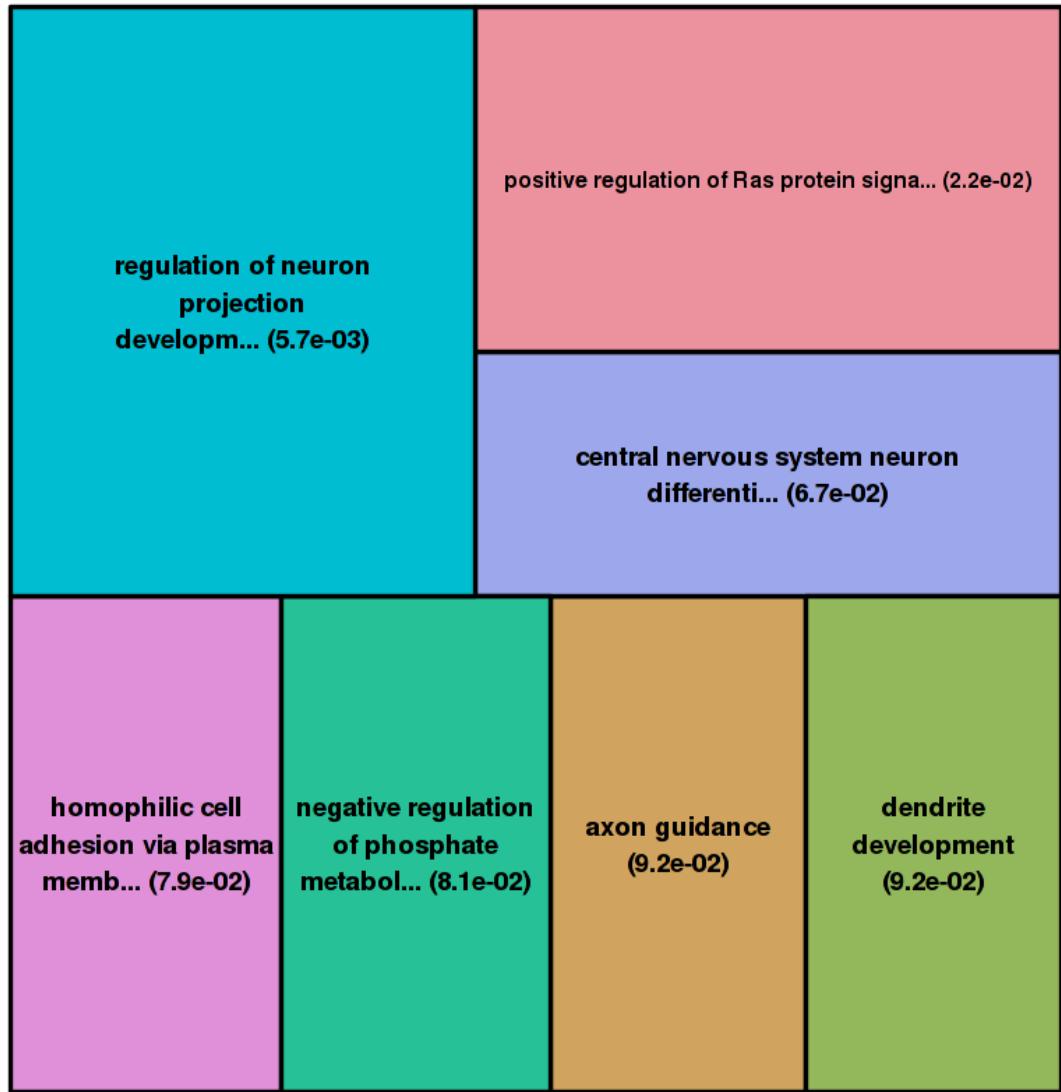
Figure 48 Biological processes treemap for the intersection of genes regulated by RBBP5 with putative enhancers featuring a loss in H3K27Ac. Values: FDR

On the other hand DEGs with a gain in H3K27Ac but not target of RBBP5 lose the enrichment for categories related to muscles and sarcoplasm (Figure 49). Also, DEGs with a loss in H3K27Ac but not target of RBBP5 completely lose any enrichment.
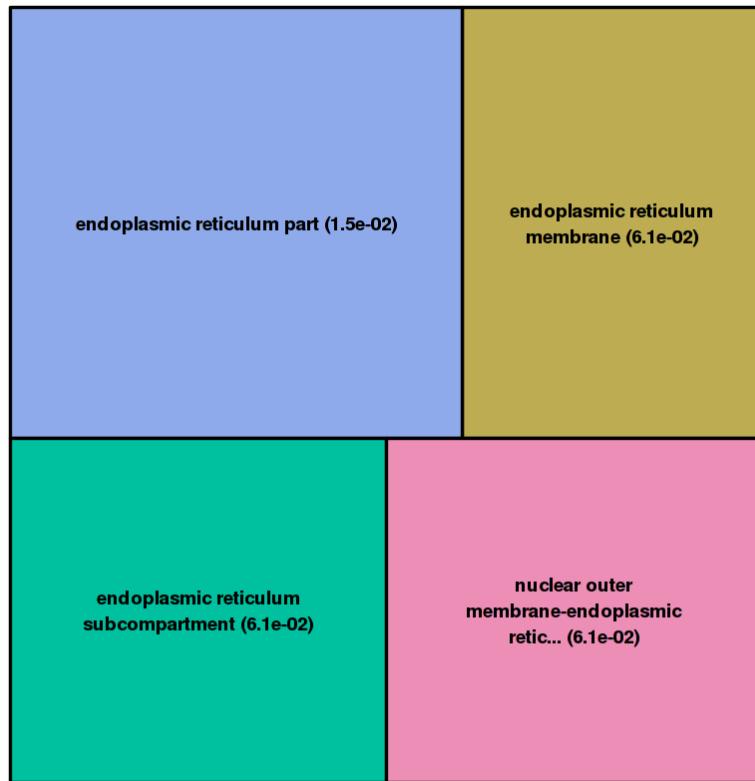
Figure 49 Biological processes enrichment for upregulated DEGs with a change in H3K27Ac in the enhancer region deployed by RBBP5 target genes

Gabriele-deVries syndrome from discovery to modeling

Heterozygous YY1 mutations cause intellectual disability with facial dysmorphisms

After the first identification of *YY1* mutations in a trio exome-sequencing (Lisenka E. L. M. Vissers et al., 2010), in an international effort, we identified ten *de novo* mutations in a cohort of 14,969 individuals with an undiagnosed intellectual disability. The probability of observing this proportion of mutations in *YY1* in a cohort of this size, after Bonferroni-adjustment, is $p = 2.8 \times 10^{-6}$ (Gabriele et al., 2017). These ten individuals, represented in Figure 50A, shared borderline to moderate intellectual disability and craniofacial dysmorphisms including facial asymmetry with a broad forehead, fullness of the upper eyelids, and an upper-lip indentation shaped like a Gingko leaf. In addition, were recurrently observed behavioral problems, movement abnormalities, intrauterine growth restriction, feeding problems, and eye abnormalities.

Moreover, one or two individuals were also featuring abnormalities such as cleft palate, craniosynostosis, esophageal atresia, Epstein anomaly, and hydronephrosis. The genetic and phenotypic characterization of individuals featuring *YY1* point mutations is summarized in Table 6. Mutations position is represented in Figure 50B.
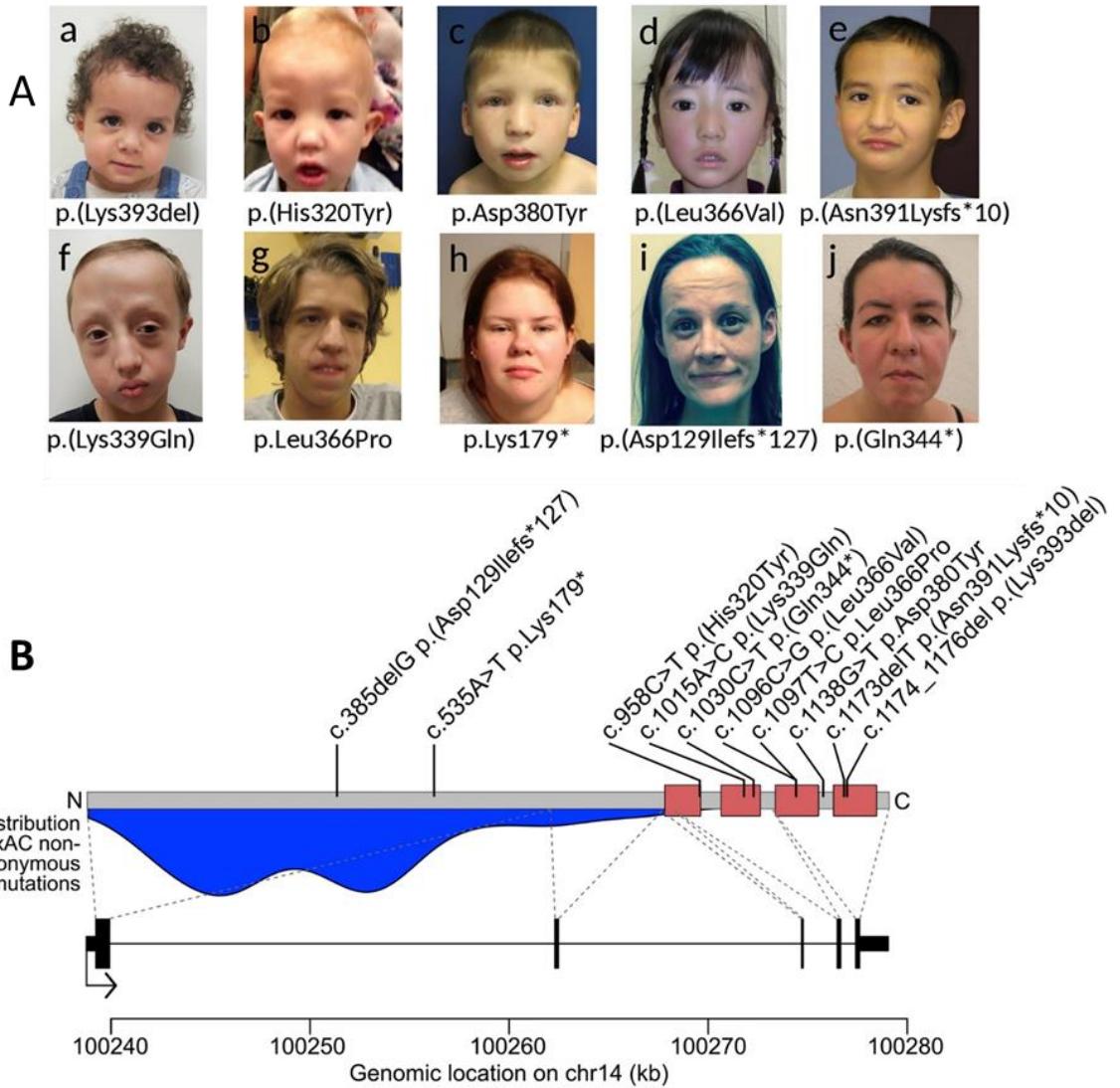
Figure 50 A) representation of GADEVS individuals sorted by age. B) Distribution of *YY1* mutations. Pink: zinc finger domains; blue: frequency of non-pathogenic mutations in ExAC. Adapted from Gabriele M., et al. 2017.

Moreover, we identified 13 individuals, with an overlapping phenotype of the individuals with *YY1* point mutations, with genomic deletions of the size range from 75Kb to 13Mb, which were completely deleting *YY1*, and of which half encompass also the maternal/paternal UPD(14) gene cluster (Gabriele et al., 2017). To study the impact of YY1 mutations, lymphoblastoid cell lines (LCLs) were derived from four heathy donors, two individuals with missense mutations located in the zinc finger domain (Individual 1 and 2 from now on referred as p.Asp380Tyr and p.Leu366Pro, respectively); an individual with a nonsense mutation (individual 5 referred as

p.Lys179∗); and an individual with the complete *YY1* deletion (referred as deletion), encompassing also the two genes *SLC25A29* and *SLC25A47*. Sanger sequencing of *YY1* cDNA with and without cycloheximide treatment of LCL confirmed NMD in p.Lys179∗, as shown in   Figure 51A. mRNA levels where halved in p.Lys179∗ and deletion, while   *YY1* mRNA levels of individuals with missense variation were unaltered, as represented in Figure 51B. Protein abundance reflects mRNA levels (Figure 51C).
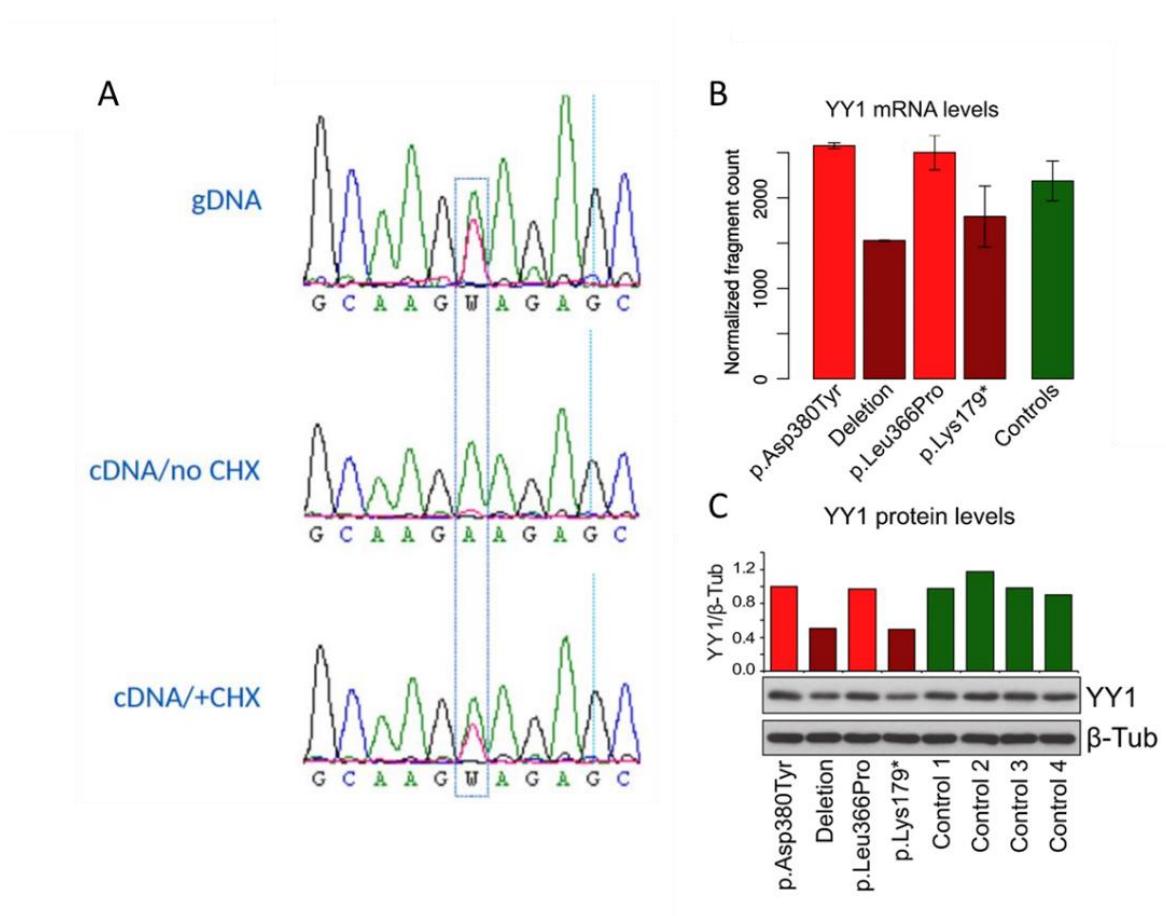
Figure 51 A) patient LCLs carrying mutation p.Lys179* show complete nonsense mediated decay. B) *YY1* mRNA levels in LCL RNA-seq. C) YY1 protein levels normalized to β-tubulin. Adapted from Gabriele M., et al. 2017.

Table 6 individuals with de novo YY1 mutations

		Individual 1	Individual 2	Individual 3	Individual 4	Individual 5	Individual 6	Individual 7	Individual 8	Individual 9	Individual 10
Mutation	cDNA change[#]	c.1138G>T	c.1097T>C	c.1096C>G	c.1030C>T	c.535A>T	c.1173delT	c.1174_1176del	c.385delG	c.1015A>C	c.958C>T
	Protein change*	p.Asp380Tyr	p.Leu366Pro	p.(Leu366Val)	p.(Gln344*)	p.Lys179*	p.(Asn391Lysfs*10)	p.(Lys393del)	p.(Asp129Ilefs*127)	p.(Lys339Gln)	p.(His320Tyr)
	Chromosome position[∞]	g.100743830G>T	g.100743789T>C	g.100743788C>G	g.100742953C>T	g.100706116A>T	g.100743865del	g.100743869-110743871del	g.100239629delG	g.100742938A>C	g.100742881C>T
	Inheritance	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*
	PhyloP[†]	6.18	5.13	2.14	6.10	0.12	1.01	4.51 [‡]	2.87	4.97	6.26
	Mutation Taster	damaging	damaging	damaging	NA	NA	NA	NA	NA	damaging	damaging
	Polyphen-2	damaging	damaging	damaging	NA	NA	NA	NA	NA	damaging	damaging
	Cohort size	10	500	500	350	100	5,500±	5,500±	6,709±	1,300	6,709±
Growth	Age	2y9m	15y10m	5y1m	39y	17y6m	7y10m	1y3m	35y	9y3m	1y5m
	Gender	M	M	F	F	F	M	M	F	M	F
	Birth weight (g; SD)	2010 (-2.5)	2220 (-3)	2290 (-1.6)	NA	2600 (-2)	2050 (-1.3)	2409 (-1.8)	3.5	3.83	3062
	Height (cm; SD)	84 (-3)	170 (-1)	105.2 (-0.5)	153.5 (-1.5)	159 (+0.7)	124 (0)	77.2 (-0.8)	154.4 (-2.5)	125.5 (-1.5)	79.2 (-0.8)

	Weight (kg; SD)	NA	47.5 (-1.5)	12.5 (-2.1)	51 (-1)	65 (+0.7)	24.6 (0)	8.6 (-2.6)	52.8 (+0.3)	23 (-1.7)	9.1 (-1.5)
	HC (cm; SD)	48.5 (-1)	57 (+0.4)	50.5 (+0.6)	54 (-1.5)	56 (+0.7)	52.5 (0)	47.3 (0)	52.5 (-2)	50 cm (-1.3)	44.8 (-1.8)
Development	Motor development	sitting 1y; walking2y9m	walking 2y	sitting 1y; walking4y8m	mild delay	walking 15m	walking 3y	sitting 1y	walking 15m	walking 6.5y	sitting 5m; walking 22m
	Speech development	first words 2y	first words 2y	mild delay	mild delay	first words 2y	delayed, 2-3 word phrases	first words 13m	delay	non-verbal	first words 14m
	Intellectual disability	moderate	moderate	mild	mild	mild/learning difficulties	mild/moderate	mild (DQ 68)	special education	moderate/ severe	NA
Neurologic	Hypotonia	-	-	-	+	-	-	+	-	moderate	-
	Behavioral abnormality	anxiety	ASD		schizoaffective disorder	ADHD	-	-	-	autism	-
	Sleep disturbance	-	-		+	+	-	-	-	-	-
	Abnormality of movement	-	toe walking		tremor	progressive dystonia	waddling gait	-	progressive dystonia	-	-
	Brain MRI	NA	normal	normal	frontal gliosis, enlarged SA spaces	normal	delayed myelination, cortical dysplasia, diffuse white matter loss	minimal prominence of the right lateral ventricle	subcortical bifrontal white matter foci	focal areas of encephalomalacia	normal

Facial dysmorphisms										
Facial asymmetry	+	+	+	+	-	-	mild	mild	+	-
Broad forehead	+	+	+	+	+	+	+	+	+	+
Simple posteriorly rotated ears	+	+	+	+	+	+	low-set	simple		protruding
Periorbital fullness	+/-	+	+	+	+	-	-	-	suborbital fullness	+
Downslant	-	+	+	+	-	+	-	-	+	-
Full nasal tip	+	+	+	+	+	-	+	-	+	+
Malar flattening	+	+	+	+	-	+	-	+	+	+
Indentation upper lip	+	+	+	+	-	-	+	-	-	-
Thick lower lip	+	+	mild	+	+	+	-	-	+	+
Pointed chin	+	+	+	-	-	+	-	-	-	-
Other	telecanthus	-		epiblepharon, high palate		Pierre-Robin sequence with CP	-		micrognathia, ptosis, sparse eyebrows	dolichocephaly, hypotonic facies, frontal upsweep

Other	Extremities	-	patella luxations, finger joint laxity	long fingers	finger joint laxity, Sydney line, sandal gap	hemihypotrophy distal left leg	laterally deviated halluces	-		distal arthrogryposis	-
	Feeding problems	+	+	+ (G-tube till 2y5m)	-	+	+ (G-tube)	+ (G-tube)	difficulty with chewing and swallowing	consistently underweight	+
	Recurrent infections	-	+	-	-	-	-	-	-	-	-
	Eye abnormalities	-	strabismus	-	strabismus	hypermetropia	-	strabismus	-	mild astigmatism	strabismus
	Renal abnormalities	HN	HN, UPJ stenosis		-	-	-	-	-	NA	NA
	Cardiac abnormalities	-	-	-	-	-	ebstein's anomaly	PFO, small aorto-pulmonary collateral	-	-	-

Legend: +, present; −, absent; +/−, minor; ADHD, attention deficit hyperactivity disorder; ASD, autism spectrum disorder; CP, cleft palate; DQ, developmental quotient; GH, growth hormone; G-tube, gastrostomy tube; HC, head circumference; HN, hydronephrosis; NA, not available; NK, not known; PFO, patent foramen ovale; SA, subarachnoid; TE, trachea-oesophageal; and UPJ, ureteropelvic junction.

YY1 deletions, nonsense or missense mutations affect YY1 chromatin binding in TSS proximal and distal regions

To test how YY1 missense mutations in the zinc finger domains and YY1 halved dosage could impact genomic localization, ChIP-seq was performed employing two independent antibodies, recognizing both the NTD and the CTD fragments of YY1. ChIP-seq results, represented in Figure 52, show overlap of the peak calling and a high Pearson correlation of fold enrichments over input in the union of enriched regions (>0.93). Peaks lost or retained are consisted across samples, irrespectively of the mutation. This underscore that both deletion, nonsense, and missense mutations reduce in the same measure YY1 capability of binding chromatin. The read density (left panel Figure 16A) defined high-occupancy peaks, which were retained both in controls and in mutated individuals' LCLs. Regions with a minor read density over the input were preferentially lost in mutants. The sample p.Asp380Tyr only shows higher number of peaks retained than the other samples due to its high coverage.

Distribution of conserved YY1 peaks (Figure 52B) showed preferential genome occupancy (>77%) in regions close to promoters, defined as regions close less than 1 kb upstream of the transcription start sites (TSSs). Gene ontology of these TSS YY1-bound genes revealed and enrichment in genes regulating transcription, mRNA processing, and translation Figure 52C). Regions with reduced YY1 binding are enriched in distal intergenic regions (Figure 52B).
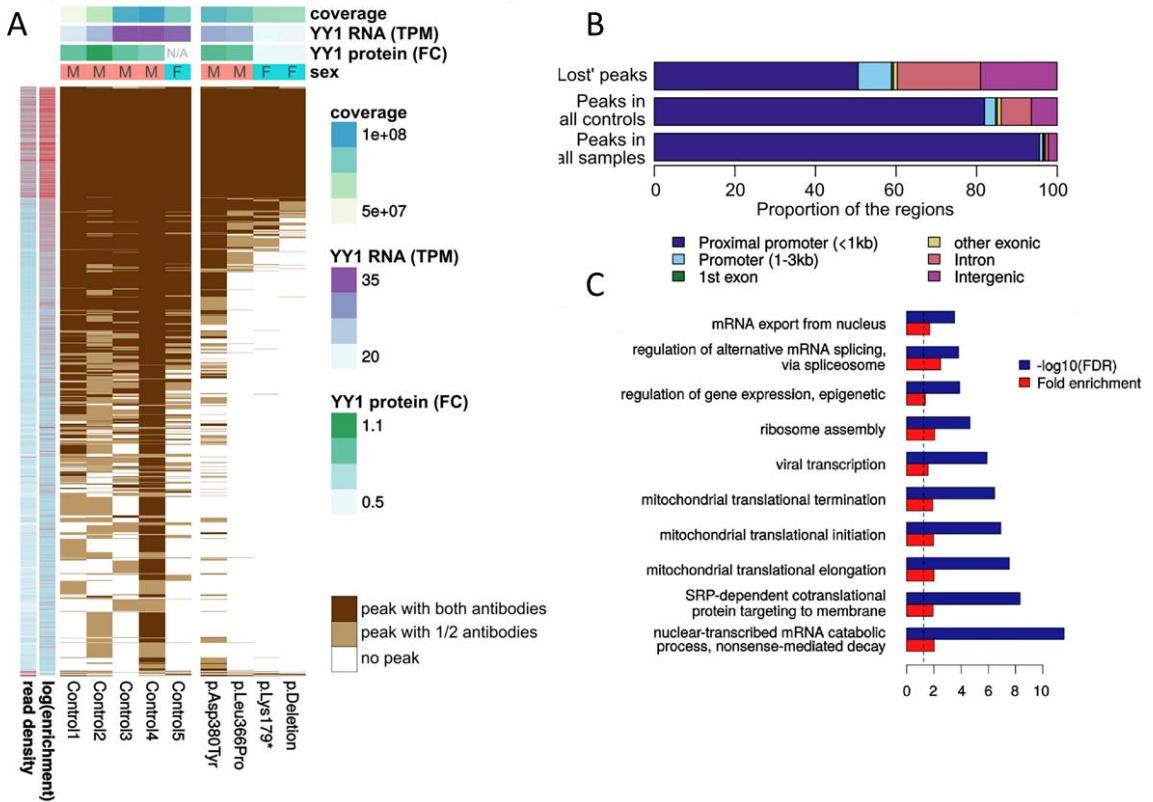
Figure 52 YY1 ChIP-seq with double antibody. Horizontal lines represent peaks; white: undetectable YY1 signal; *YY1* expression level addressed by RNA-seq; TPM: transcripts per million; FC: fold change in relation to the control samples. B) Distribution of conserved and "lost" YY1 peaks across genomic features. C) Gene Ontology (GO) enrichments of the YY1-bound TSS. Adapted from Gabriele M., et al. 2017.

YY1 Haploinsufficiency leads to differential expression of only a minority of its targets

To estimate the impact of YY1-binding gene occupancy loss on gene expression, RNAseq was performed on *YY1* mutated LCL individuals and including an additional control. Principal component analysis of these transcriptome shows a clear separation between controls and mutated sample across component 2 (Figure 53A). The source of variability responsible for principal component 1 was not identified with any biological or technical factor probed (sex, age, coverage, sequencing lane, RNA extraction efficiency, or proportion of ribosomal or Epstein-Barr virus RNA). Following differential expression analysis 152 DEGs were identified (Figure 53B). In this dataset, genes related to chromatin silencing were

significantly enriched with genes related to chromatin silencing (p $\sim$ 9 $\times$ 10^{-10}). Figure 53C shows DEGs directly bound by YY1, which are about 21% of DEGs. Of all the YY1-bound genes, only <1% are differentially expressed.
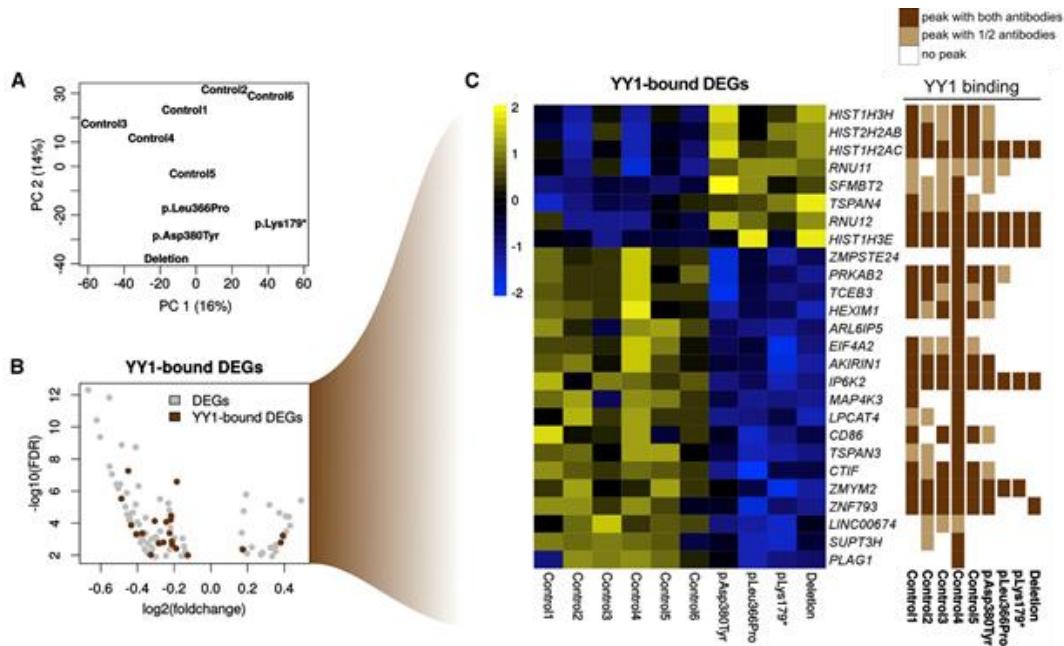
Mutation on transcription. A) PCA of LCLs' RNA-seq.). B) Representation of differentially expressed genes in Volcano plot: brown dots represent genes which are directly bound by YY1. C) Heatmap of Z.-score of YY1-bound DEGs and YY1 binding pattern. Adapted from Gabriele et. Al, 2017.

To reduce technical variability and improve the robustness, RNA was extracted again from the same independently cultured LCLs samples, and including data from 73 apparently healthy individuals LCLs transcriptional profiles available in the HapMap database (Montgomery et al., 2010). A large number of permutation analysis was performed to identify genes variable in healthy individuals, to find genes which expression was associated with *YY1* expression levels, and to identify high-confidence DEGs between proband and controls (Gabriele et al., 2017). The set of genes associated with *YY1* levels is composed of 6,687 genes, of which 2248 are directly bound by *YY1*. This set of genes is enriched in categories relevant for translation and transcription control, as well as wnt canonical pathway (Figure 54A and Figure 54B). Differentially expressed genes coming from this analysis are enriched in categories relevant for the disease such as synapse organization and

cardiac function (Figure 54C). Moreover, a subset of high-confidence DEGs composed of 50 genes, present in both analyses, was identified (Figure 54D).
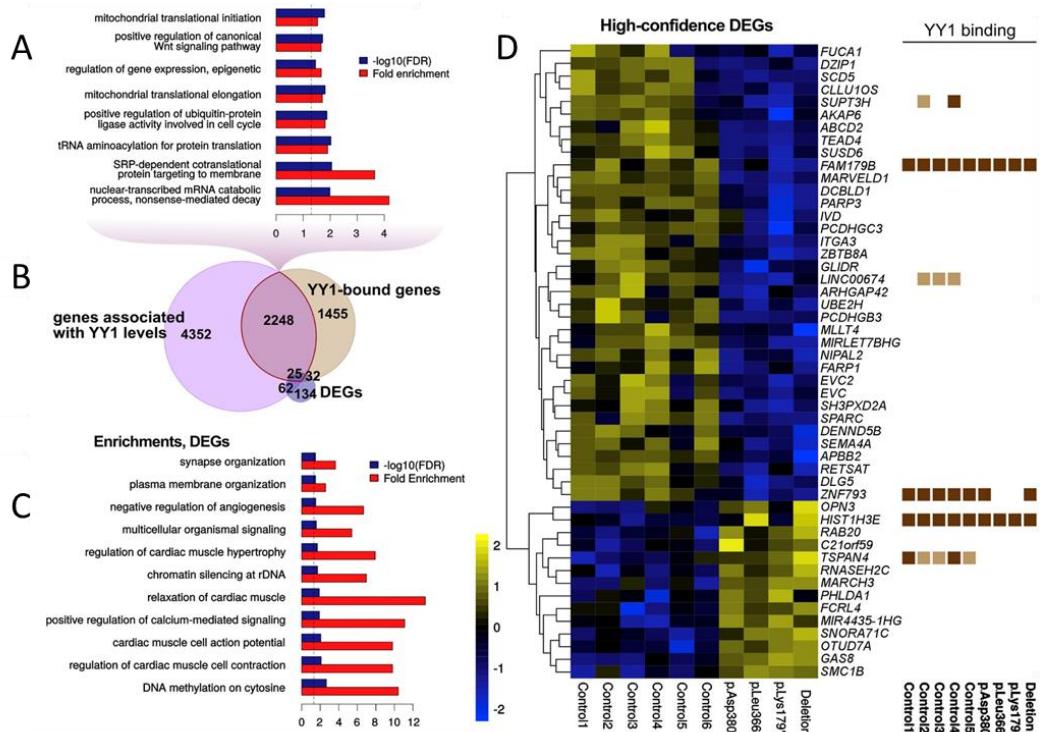
Figure 54 A) GO enrichments of the YY1-bound genes that show significant association with YY1 expression levels across HapMap LCLs data. B) Overlap between YY1-bound and YY1-associated genes, and differentially expressed genes. C) GO enrichment for all differentially expressed genes. D) High-confidence DEGs (left) with YY1 bindings at their TSS (right).

Regardless of the low overlap between DEGs and YY1-bound genes, DEGs had a greater decrease in YY1 binding in comparison to genes without a statistically significant difference in expression (p ∼ 1 × 10−4). A large overlap between the two distributions (Figure 55) suggests that the presence of binding partners is likely to distinguish YY1-dosage-dependent, transcriptionally affected genes in any given cellular state.
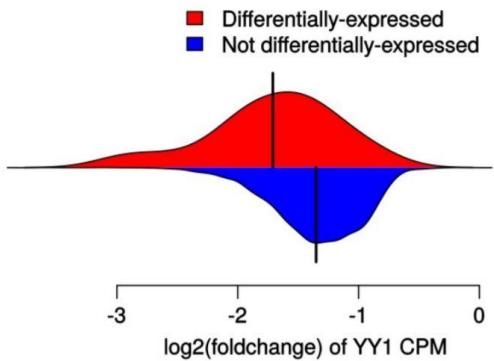
Figure 55 Distribution of YY1 binding foldchanges at the TSS of DEGs and not significantly DEGs. CPM: Counts per Million reads mapped

YY1 Haploinsufficiency Is Associated with Widespread Loss of H3K27 Acetylation

Since it is known from previous literature that YY1 interacts with several histone modifiers, especially lysine acetylases and deacetylases, such as HDAC1, HDAC2, p300, and CBP, and most DEGs are downregulated, the impact of YY1 haploinsufficiency on H3K27Ac was evaluated by ChIP-seq. The distribution of H3K27Ac separates probands apart from controls (Figure 56A). Given the fact that different normalization methods may lead to different interpretation, H3K27ac data were analyzes with the most conservative analysis, focusing on genomic sites that were showing reduced H3K27Ac even under the assumption of no difference in the global distribution. This analysis potentially underestimates regions with H3K27Ac loss, but the possibility that spurious regions will be identified as artifacts of normalization is excluded (Gabriele et al., 2017). As a result of this analysis, 39% of YY1-bound DEGs lose H3K27Ac at their proximal promoter. However, the intersection between H3K4me1 and H3K27Ac, to distinguish active and poised enhancers, showed that 99% of regions with reduced H3K27Ac was on active enhancers. A marked decrease of H3K27Ac on YY1-bound active enhancer was

identified in probands, and YY1 fold changes with H3K27Ac in active enhancer were positively correlated ($p < 2.2 \times 10^{-16}$) (Figure 56B) Moreover, after normalization on library size, 82% of YY1-bound enhancers showed a statistically significant decrease in H3K27ac (Figure 56C), while only 25% of the other active enhancer show a reduction in H3K27ac.
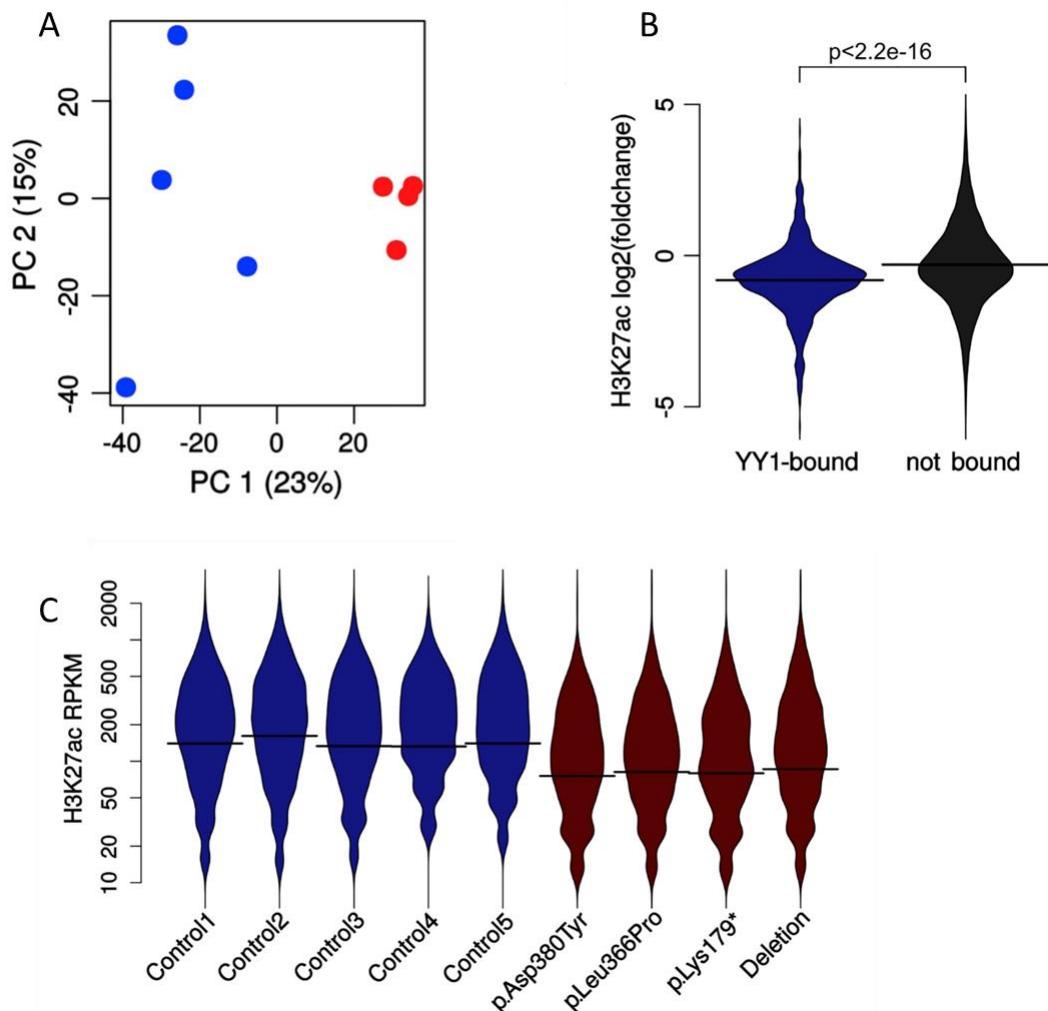
Figure 56 YY1 Haploinsufficiency causes enhancer dysregulation. A) PCA of the H3K27ac read-count distribution. B) Loss of H3K27ac at YY1-bound compared to non-YY1-bound enhancers. P-value calculated with two-tailed test. C) Distribution of H3K27ac read densities at YY1-bound enhancer. RPKM: reads per kilobase pair per million reads mapped. Adapted from Gabriele et Al. 2017.

These results are consistent with the already present evidence that YY1 mediates long-range interactions (Gerasimova et al., 2015; Medvedovic et al., 2013), subsequently confirmed by Young group (Weintraub et al., 2017). Consequently, the possibility that YY1 binding at distal enhancers might have a stronger impact on

gene expression, was tested. Following this assumption, of the YY1 binding sites not mapped in the proximal promoters, 65% were located on active enhancer, while only <4% were located on poised enhancers (identified by the simultaneous presence of H3K27me3 and H3K4me1). Also, 82% of YY1-bound enhancers showed a statistically significant decrease in H3K27ac, in contrast with merely 25% of the other active enhancers. Moreover, while no global increase of H3K27me3 was found, across YY1-bound regions the loss of H3K27ac is associated with an increase in H3K27me3, especially in regions with enrichment for these two histone modifications (Figure 57).
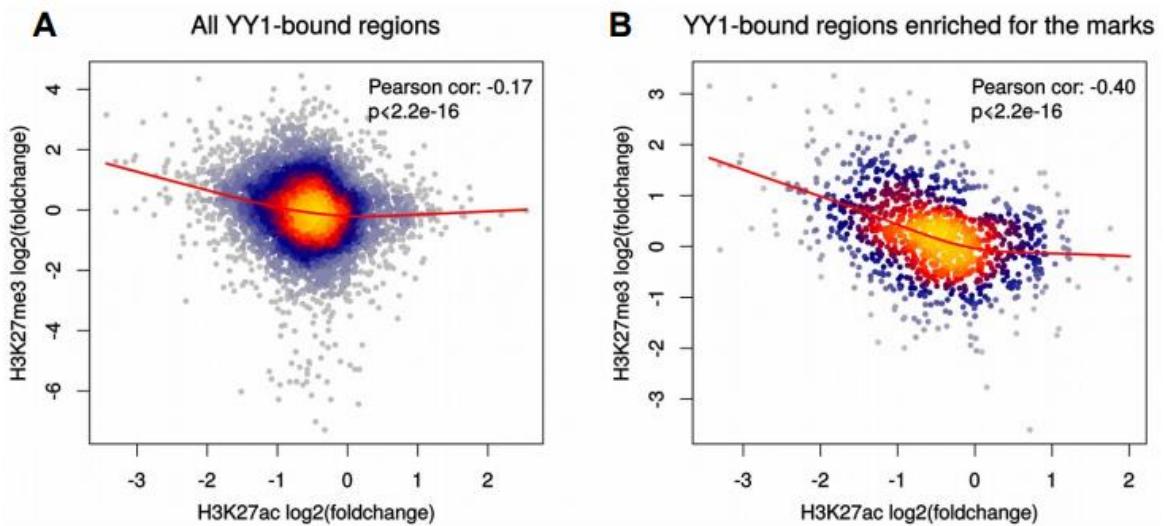
Figure 57 A) Foldchanges of H3K27ac and H3K27me3 across all YY1-enriched regions. B) H3K27ac and H3K27me3 foldchanges across YY1- bound regions with an enrichment for H3K27me3 and H3K27ac in at least one sample. The red line: smooth spline fitted on the data.

Taking in account the 500 YY1 peaks least affected by its dosage, only 15% of overlapping enhancers showed a reduction of H3K27Ac, highlighting its *cis* role in regulating enhancer activation. For this reason, publicly available Hi-C profiles of LCL (Rao et al., 2014) were used to identify chromatin loops responsible for linking distal regulatory sites to TSS of putative genes. Following this analysis, 545 genes distally bound by YY1 were identified, of these set 206 are also bound at the TSS

(Gabriele et al., 2017). Interestingly, when compared with proximally genes, distally YY1-bound genes were much more likely to be differentially expressed (p ~ 1.5 × 10^{-5}, chi-square test). In addition, distally YY1-bound genes were enriched with differentially expressed genes (p ~ 6 × 10^{-8} with respect to the whole genome and p ~ 7 × 10^{-4} with respect to the expressed genes, hypergeometric test) (Gabriele et al., 2017). These observation underscore that regardless of its preferential enrichment at the TSS, YY1 affects disproportionately more its distal targets.

Treatment with histone deacetylase SAHA increases the expression of differentially acetylated genes in YY1 mutated samples

Given the widespread loss of H3K27Ac in *YY1* probands, the histone deacetylase inhibitor SAHA was adopted to treat controls and probands LCLs to test its capability of recovery gene expression. Following overnight treatment with SAHA 0,1 μM gene expression was evaluated in two differentially H3k27 acetylated, differentially expressed, and differentially YY1-bound genes (*FAM179B* and *SUPT3H*), and two classical housekeeping genes (*TUBB* and *TBP*) neither differentially acetylated nor differentially expressed but bound by YY1. RT-qPCR results shown in Figure 58 show that only probands are sensitive to this treatment and, compared with DMSO, expression levels are increased for all four tested genes.
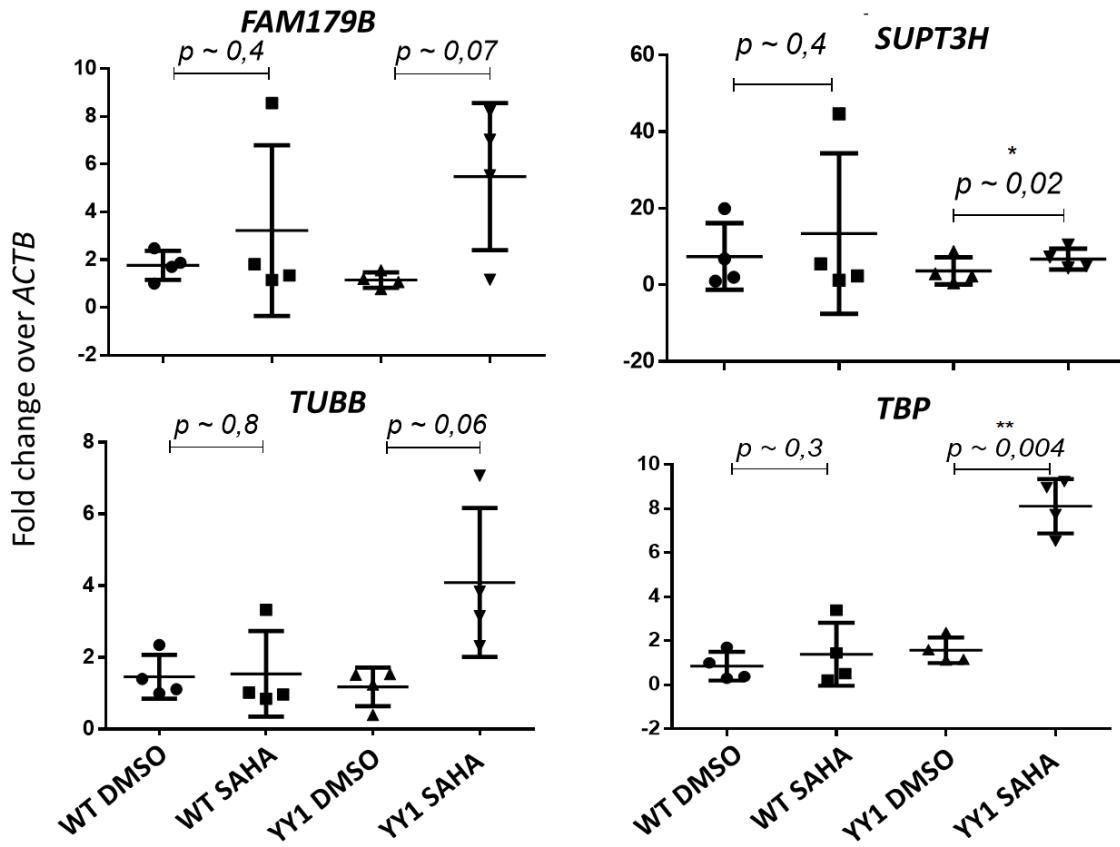
Figure 58 YY1 mutated samples show increase expression levels after treatment with histone deacetylase inhibitor SAHA.

YY1 patient-derived iPSCs show a selected deregulation

To study the impact of *YY1* mutations in more appropriate disease-relevant cell types, LCL of probands and controls, and a fibroblast samples of individual eight were subjected to reprogramming into iPSCs using Sendai virus. Sendai reprogramming was successful only in the fibroblast sample and in one proband and controls' LCL. After reprogramming from LCL, Epstein Barr virus (EBV) expression in iPSCs was evaluated. Both samples completely turned off EBV expression compared to a positive control cDNA taken from LCL, showing irrelevant Ct, and with melting curves comparable with the negative (no cDNA), as shown in Figure 59.
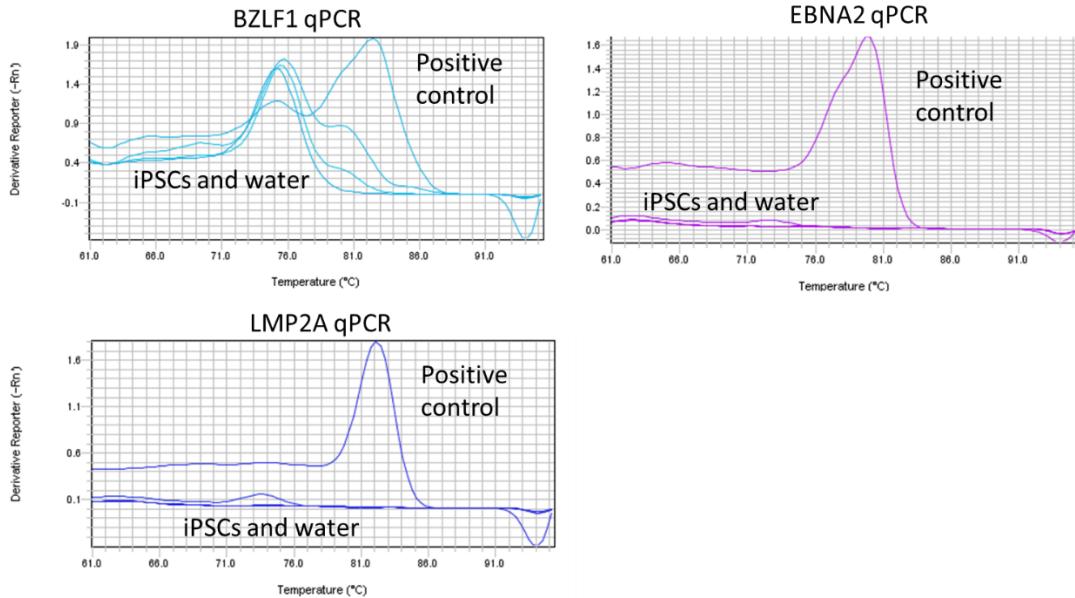
Figure 59 iPSCs derived from LCL turn off EBV genes. Melting courve for a positive control cDNA taken from LCL, the two iPSCs derived from LCL, and no cDNA control.

The two *YY1* iPSCs, the control iPSC LCL-derived, and one unmatched control reprogrammed previously with an independent reprogramming method (Adamo et al., 2015) were subjected to RNA-seq for genome wide-transcription analysis. First, the level of pluripotency genes was evaluated. All samples express comparable levels of *NANOG*, *POU5F1* (*OCT4*), and *SOX2* as depicted in Figure 60.
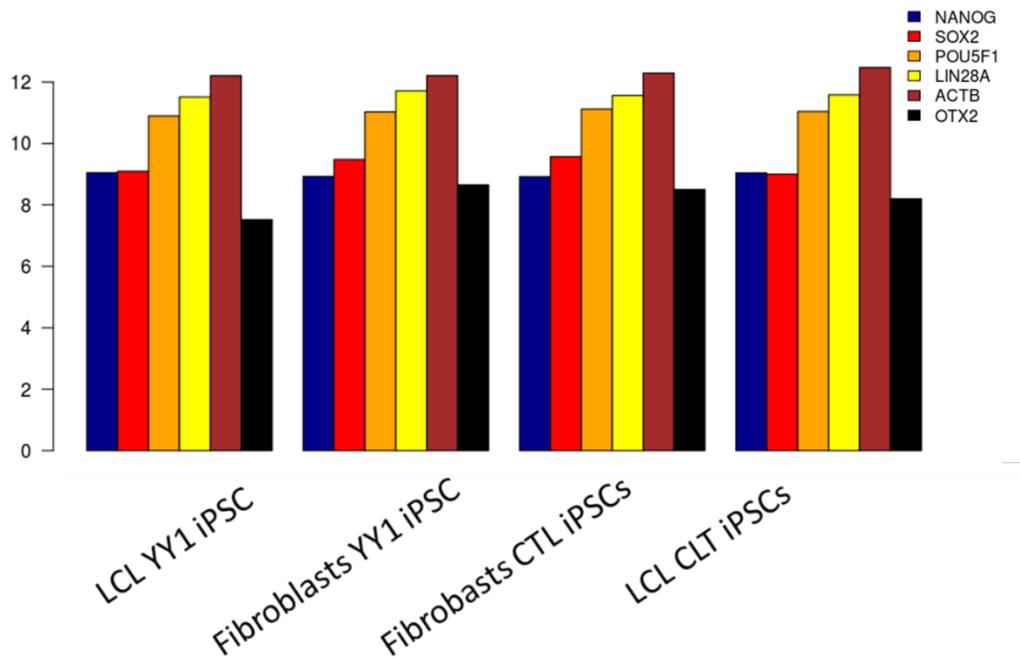
Figure 60 iPSC derived from LCL and fibroblasts of controls and YY1 probands express a similar level of pluripotency gene and ACTB as housekeepings. Data showed Log normalized read counts.

Differential expression analyses identified 795 differentially expressed genes (DEGs) (FDR <0,25); 271 DEGs (FDR <0,1); 131 DEGs (FDR <0,0,5), represented in Figure 61

To identify gene with high robustness the differentially expressed genes in GADEVS iPSCs,

Differentially expression analysis between probands and control lines idenfied 236 DEGs (Figure 61).
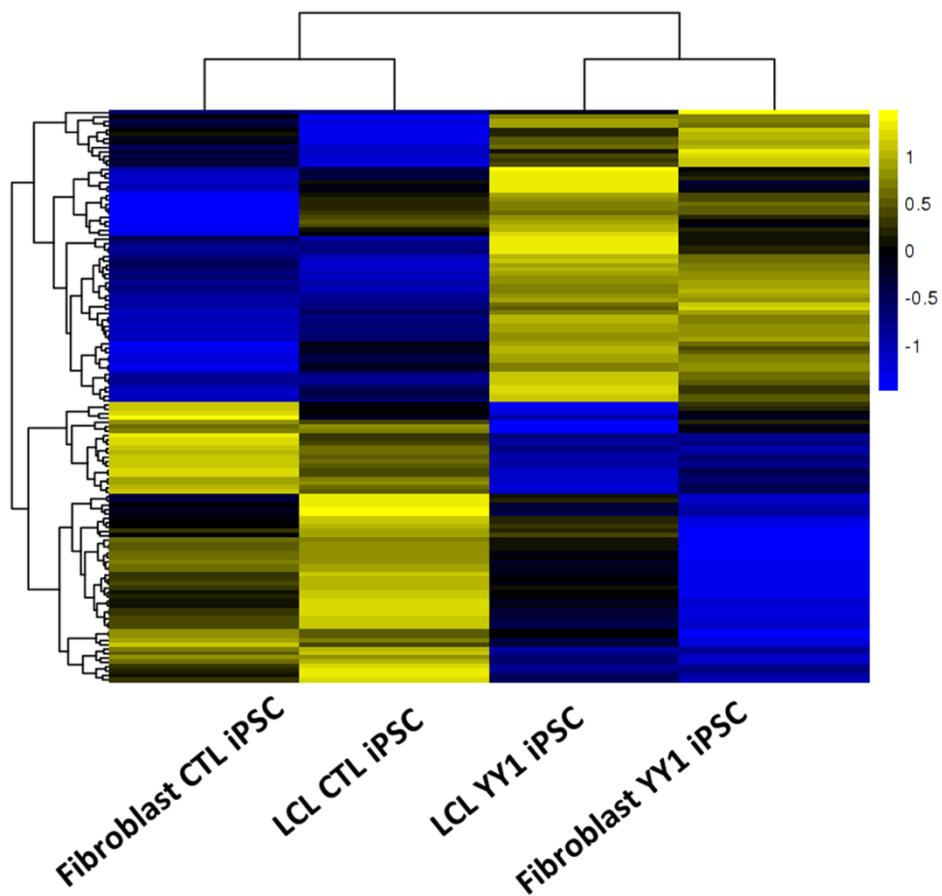
Figure 61 DEA of YY1 iPSCs and CTL iPSCs. 131 differentially expressed genes with FDR <0,05. Scalebar represent lognormalized gene counts.

DISCUSSION

KMT2D haploinsufficiency unveil transcriptional vulnerability in disease-relevant cell types

Heterozygous mutations on *KMT2D* causes KS, of which clinical features are supposed to have origins mostly from altered development of cortical neurons (intellectual disability) and neural crests (systemic abnormalities and facial features). Here we describe for the first time a human *KMT2D* heterozygous modeling platform that appears to recapitulate molecular deregulations underlying KS clinical traits. We assembled a uniquely informative patient cohort composed of fibroblast biopsies from seven KS patients and five half-matched healthy controls. The first characterization of primary fibroblasts did not show any alteration of the bulk H3K4 methylation statuses, in line with previous reports that showed how KMT2D KO does not affect bulk deposition of H3K4me1 in the presence of two copies homolog KMT2C. Nevertheless, it was only showed a partial redundancy, with two very distinct phenotypes when one of the two was constitutively KO. (Lee et al., 2013b). Indeed, to identify specific deregulation caused by KMT2D heterozygosity, fibroblast were characterized transcriptionally. Already at this stage was possible to identify transcription dysregulation enriched for processes relevant to KS. Afterward, genome-wide H3K4me1 profiling was able to separate patients and controls according to the genotype, and a major defects on 7455 regions was identified. The loci with H3K4me1 loss were enriched in intergenic regions, pointing to the involvement of KMT2D in enhancer regulation. To validate this defect in

independent fibroblast datasets, we crossed our data with Roadmap epigenomics (Kundaje et al., 2015). This intersection confirmed a loss of monomethylation in regions commonly marked with this PTM, and we also identified the presence of an excess of H3K4me1 in regions commonly marked by H3K27me3. Although we have found a molecular scar indicative of KMT2D heterozygosity, dermal fibroblasts are not a cell type majorly affected in KS, and not suitable for extensive analysis given the limited cell division before they undergo senescence. For this reason, fibroblast samples were reprogrammed into iPSCs. KMT2D heterozygosity did not apparently interfere with the reprogramming process. In the iPSCs, we identified a comparable number of DEGs found in fibroblast, but differently from these, at the pluripotency stage, no enrichment was detected. Indeed, unsupervised clustering showed that each iPSCs clone was correlating with other clones of the same individual and the same family, suggesting that the differences among individual due to genetic background were overriding the effect of KMT2D heterozygosity. Genome-wide characterization of H3K4me1 and H3K27ac showed a separation between controls and mutated individuals in the former one, while no clear separation with very selective deregulated regions in the latter one. This evidence is compatible with the fact that *KMT2D* does not appear to be differentially deregulated in pluripotency and is in line with previous research that showed KMT2D to be necessary after gastrulation (Lee et al., 2013b). The results obtained *from* iPSCs characterization support evidence showing that KMT2D is needed later in development to an epigenetic stated required to establish or maintain differentiated cell state (Siang-Yun Ang et al., 2016; Kim et al., 2016; Lee et al., 2013b) and that to the fact that somatic reprogramming is able to delete previous epigenetic signature derived from genetic and environmental factors (Mertens et al., 2015) *de facto* creating a new starting point. For this reason, iPSCs were differentiated to NCSC and cortical neurons to reproduce the recapitulate developmental stages and to identify the cell

stage at which deregulation responsible for the clinical phenotype occur. Transcriptome KS and controls NCSCs does not segregate by genotype samples. Regardless of this, it was possible to identify a set of deregulated genes, with a subset of genes recurrently identified changing analysis parameters. This small subset of genes was significantly enriched for the biological process cell adhesion. Worth of mention is that several of these genes are important for NC biology and KS clinical symptoms, For instance, FGF2, necessary for NCSC proliferation; ITGA10, an adhesion protein highly expressed in chondrocytes, COL9A3, which is associated with a developmental syndrome (OMIM # 600969), AP1S2, associated with mental retardation (OMIM #304340), ANO5, mutated in muscular dystrophies (OMIM # 613319, .611307) and gnathodiaphyseal dysplasia (OMIM # 166260); FRAS1, known to be responsible for a recessive multisystem disorder characterized by craniofacial, urogenital and respiratory anomalies (OMIM # 219000); TNS2, required for kidney function (Cho et al., 2006). Also, the whole set of DEGs is enriched for several categories relevant for KS pathogenesis associated with NC-derived tissues. At the stage of NCSC *KMT2D* appears slightly downregulated when compared to controls, but the variability among controls did not permit to make statistical claims. Profiling of chromatin modification H3K4me1 and H3K27Ac did not identify any dramatic alteration. Nevertheless, some deregulated enhancers were targeting genes fundamental for NCSC development and differentiation such as PAX8 (Blake and Ziman, 2014); NEDD4L, responsible for an autosomal dominant ND caused by a defect in neuronal migration(OMIM: #617201), SLITRK1, found mutated in a subset of Tourette syndrome patients (OMIM: # 137580) and a neuropsychiatric disorder (OMIM: #613229). Remarkably, the mild phenotype that we observed in NCSC is in line with a recent line of research that shows that in KS the fate of NCSC is mainly affected only in the post-migratory stage, when NCSC further differentiate in other cell types (Shpargel et al., 2017)

To understand how KMT2D mutations lead to intellectual disability we differentiated iPSCs into upper cortical layer excitatory neurons. Morphology of neuronal network was not found to be altered when compared to control samples. The spontaneous electrical activity of network burst showed a very specific phenotype distinct from controls and any other ND to when compared to Kleefstra syndrome, caused by mutations in EHMT1 but also KMT2C (Frega et al., 2018; Gabriele et al., 2018). This unique, spontaneous electrical network activity was identified by the presence of multiple firing bursts for each network burst, characteristic that lacks healthy individuals. Given the fact that neuronal architecture is not responsible for this striking phenotype, we investigated the molecular alterations of transcription and chromatin regulation. In iNeurons *KMT2D* is significantly downregulated with a logFC of -0,79. In a phenomenon that seems a clumsy attempt of compensating, *KDM6A* results upregulated with a logFC: 0.9506504. Cortical neurons appear to degrade the mutant KMT2D allele through NMD fully. The lack of an efficient antibody does not permit to verify this observation ad the protein level. In cortical neurons we identified 1318 DEGs with an FDR < 0,05. These genes are enriched for categories relevant for neuronal electrophysiology, signal transduction, and translation. Importantly, by separating up and downregulated genes is possible to describe a specific impact of these two categories. While downregulated genes are responsible for the enrichment specific for neuronal function, upregulated genes are responsible for upregulation of mRNA metabolism and translation initiation. Remarkably, a flawed protein translation is an hallmark of intellectual disabilities such as autism (Rogozin et al., 2018; Torre-Ubieta et al., 2016). Using a tool of the Ingenuity pathway analysis is possible to investigate diseases related to a gene subset. Remarkably, while upregulated genes were associated with categories related to cell death of cancer death and protein translation, the top categories in downregulated genes are associated with

melanoma and gastrointestinal cancer. Specifically, the tumor suppressor genes TP53, ATM, BAX, KLF10, and CDK1A are upregulated. It is relevant to discuss in detail a number of oncogenes resulted downregulated found in several disease ontology related to cancer from IPA analysis. These genes are listed in the following table:

Table 7 oncogenes downregulated in KS iNeurons

gene	role	cancer involvment	Germline mutations	Reference
RET	receptor tyrosine kinase	Constitutively activated in multiple endocrine neoplasia type 2;	Hirschsprung disease	(Plaza-Menacho et al., 2006)
MET	tyrosine-protein kinase	Constitutively activated by somatic and germline mutations in a broad variety of tumors	Downregulated in Autism	(Campbell et al., 2006; Comoglio et al., 2018)
WNT5A	Wnt Ligand	Upregulated in a variety of tumors	Robinow syndrome	(Asem et al., 2016; Robinow et al., 1969)
INO80D	Chromatin remodeller	Oncogenic in non-small cell lung cancer		(Zhang et al., 2017)
KRAS	RAS GTPase-activating protein	Oncogenic in non-small cell lung cancer and a variety of tumors	Noonan syndrome, cardio-facio-cutaneous syndrome	(Haigis, 2017; Niihori et al., 2006; Román et al., 2018; Schubbert et al., 2006)
KSR2	Scaffold regulator for MAPK	Overexpressed in melanoma and sustain tumor metabolism	Early-onset obesity and severe insulin resistance	(Bottiglione and Giurisato, 2015; Fernandez et al., 2012; Guo et al., 2017)
Rab27B	Ras superfamily of monomeric G proteins	Breast cancer progression		(Chua and Tang, 2015)
RAB31	Ras superfamily of monomeric G proteins	Upregulated in breast, and cervical cancer, glioblastoma progression		(Chua and Tang, 2015)
POU5F1B	OCT4 pseudogene	Amplified in aggressive gastric cancer phenotype		(Hayashi et al., 2015)
SOX11	Transcription factor	Promotes breast cancer, and suppresses gastric cancer progression	Mild Coffin-Siris syndrome	(Oliemuller et al., 2017; QU et al., 2014; Shepherd et al., 2016; Tsurusaki et al., 2014)
SOX4	Transcription factor	Oncogenic role in acute myeloid and lymphoblastic leukemia		(Ramezani-Rad et al., 2013; H. Zhang et al., 2013)

From this data it appears that MAPK pathway result deregulated. A number of oncogenes, listed in Table 7, is aligned with the theory of the inverse comorbidity (Tabarés-Seisdedos and Rubenstein, 2013) and while somatic overexpression leads to cancer, germline loss of function mutations lead to NDs, of which Robinow syndrome and Noonan syndrome share abnormalities that affect similar anatomical systems also affected in KS. Moreover, germline alteration of the MEK/ERK pathway were previously found responsible for KS. Indeed, homozygous mutations in *RAP1A* and heterozygous mutations *RAP1B* have been found responsible for KS, and RBBP5 was found to bind the promoter of RAP1B (Bögershausen et al., 2015). Moreover, biochemical studies involving the mutated RAP1A showed a reduced level of activation of MEK/ERK through BRAF activation and a reduced level of ERK pathway thorough RAF1 inhibition (Bögershausen et al., 2015). Remarkably, in our iNeuron datasets RAP1GAP2, a RAP1 activating GTPase (Schultess et al., 2005), results downregulated. Importantly, *KMT2D* mutations have been found tumorigenic in melanoma (Bossi et al., 2016), associated with poor prognosis in Non-small-cell lung cancer (Ardeshir-Larijani et al., 2018), frequently mutated in Small Cell Lung Cancer and diffuse large B-cell lymphoma (Augert et al., 2017; Pasqualucci et al., 2011), inactivated in blood, large intestine, and brain tumors but associated with poor prognosis of breast cancer (Rao and Dou, 2015).

Contrary to what expected, with the exception of very few cases of KS with tumors, it does not seem that KS confers an increased cancer incidence when compared to the general population (Casanova et al., 2011). Therefore, it is tempting to speculate that KMT2D mutations impact cell identity with a very time-specific effect. While somatic mutations lead to cancer, it seems that germline mutations create an altered molecular environment that leads to KS but compensate preventing cancer onset. Our data from iNeurons supports this hypothesis since several oncogenes appear to be downregulated, and essential tumor suppressors are upregulated. Among these

108

upregulated tumor suppressors, the most probable candidate responsible for this compensation is KDM6A, not only because it is a direct interactor of KMT2D, but also because it was recently identified as a tumor suppressor protein, given its capability of counteracting PRC2 (Wang et al., 2018). Consequently, it is appealing to suggest that overexpression of *KDM6A* in terminally differentiated post-mitotic cell types, together with the downregulation of oncogenes and MAPK pathway, might be responsible for the lack of increased cancer susceptibility in KS patients.

The density of genome-wide distribution of H3K4me1 did not appear to be strikingly affected. Nevertheless, quantitative analyses identified genomic regions corresponding to putative enhancers, which target genes are enriched for biological categories relevant for the neuronal biology. On the other hand, the analysis of H3K27Ac, with the exception of a sample, segregates KS individuals and controls according to their genotype. Genomic density distribution for this PTM shows a reduction of H3K27Ac in regions with high density of this mark and enrichment in some regions with a low density. Quantitative analysis of target genes of putative enhancer regions with loss or gain of H3K27Ac showed that hypoacetylated enhancers regulate genes related to KS neurological phenotype. On the other hand, hyperacetylated enhancers seem to regulate genes important for melanocytes or associated with a muscular function, such as contractile fiber and sarcoplasm. Similar results were recently shown by an independent line of research of our laboratory, in which it was demonstrated that the lack of KMT2B during trans-differentiation from mouse embryonic fibroblast to neurons lead to imperfect transdifferentiation with overexpression of genes related to muscular function (Barbagiovanni et al., 2018).

Using an internal database to identify master regulator upstream to gene networks is possible to use RBBP5 as a proxy to identify the putative direct target of KMT2D

for DEGs. By doing this analysis, we identified that 98 of the DEGs with a loss of H3K27Ac are putative targets of RBBP5, while 140 DEGs have a gain in H3K27Ac. The putative direct targets with a defect of enhancer activation are enriched in specific genes for neuronal function. On the other hand, the putative direct target of RBBP5 of the regions with a gain of H3K27Ac appears to be directly responsible for the enrichment for contractile fiber and sarcoplasm categories, since the subtraction of these genes to the set of genes erases such enrichment. Taken together, all these observations across several cell types are in line with recent evidence that shows that KMT2D catalytical activity is not necessary for its role in transcription regulation(Dorighi et al., 2017; Rickels et al., 2017). Indeed, only in the cell type with a complete *KMT2D* downregulation, in which the protein level is presumably halved, the most robust phenotype was detected both at functional and molecular level. This advancement of the understanding of the KS molecular pathology, the identification of strong molecular and physiological phenotype, and the possibility of reproducing disease-relevant cell types, permits us now to design realistic strategy to intervene in actionable pathways, to find molecular intervention able to revert the phenotype and, in parallel, to create a better understanding for the tumorigenic processes that underly somatic *KMT2D* mutations.

Discussion of results for YY1 haploinsufficiency

We described for the first time a new neurodevelopmental disorder caused by haploinsufficiency of *YY1*, later named Gabriele-deVries syndrome (OMIM: # 617557), which can result from its deletion or from single nucleotide mutations causing either missense mutations in the DNA-binding domain, or truncating mutations. The latter ones trigger NMD, leading to the degradation of the mutant transcript. By ChIP-seq with two independent antibodies, we showed that YY1 deletion and missense and nonsense mutations equally impair YY1 chromatin binding, affecting mostly the low-occupancy sites supporting the proposed model for which YY1 haploinsufficiency cause the syndrome.

According to the capability of YY1 to mediate chromatin loops between enhancers and promoters (Gabriele et al., 2017; Gerasimova et al., 2015; Medvedovic et al., 2013; Weintraub et al., 2017), individuals harboring YY1 mutations and deletions display an extensive loss of H3K27Ac on YY1-bound enhancers. These line of evidence supports the classification of Gabriele-de Vries syndrome as an enhanceropathy.

In agreement with the observations that mutations in chromatin and transcriptional regulators severely impacts the neurodevelopment leading to a variety of neurodevelopmental disorders, given the vulnerability of this developmental stage and cell types to transcription regulation (Gabriele et al., 2018), and in agreement with the role in enhancer-promoter looping, YY1 has been shown to be involved in mediating enhancer-promoter looping in genes specific for mouse neuronal precursor (Beagan et al., 2017). In addition, YY1 has been shown to interact with the lncRNA Sox2ot, coded in the same locus of Sox2, and mediates Sox2 repression during the transition from neuronal precursor to neuronal differentiation. In this

context, YY1 downregulation was found to increase neural precursor population over early- and late-born neurons (Knauss et al., 2018).

These pieces of evidence further support the haploinsufficiency of YY1 to be causative of intellectual disability, a foundational phenotype shared across all Gabriele-deVries patients, that failures thus likely cause it in orchestrating the stages of neural development. Given the higher vulnerability of YY1 mutated individuals to respond to histone deacetylase inhibitors (HDACs) it becomes pertinent to test this class of compounds in disease-relevant cell types such as cortical neurons.

In this study we adopted LCL and iPSCs which are a good model to study YY1 haploinsufficiency but not the most appropriate cell type of studying neurodevelopment. Indeed, shortly iPSCs will be differentiated in cortical neurons, as done for Kabuki syndrome, to address the specific impact of YY1 mutations in neurodevelopment. Nevertheless, since it was previously shown that the dysregulation found at the pluripotency stage can be further amplified and propagated in more differentiated cell types (Adamo et al., 2015), and given the fact that Gabriele de-Vries and Kabuki syndromes are enhanceropaties, it has been investigated the possibility that genes already dysregulated at the pluripotency stage in GADEVS could be deregulated in Kabuki neurons. The intersection between KS iPSCs DEGS (907, FDR < 0,1) and GADEVS iPSCs DEGs (271, FDR < 0,1) identified 51 overlapping genes between the two conditions (hypergeometric test FDR ~ 0,008). Moreover, the intersection of this subset of genes with DEGs (FDR < 0,1) in KS iNeurons identifies 10 genes that area also DEGs in this differentiated cell type, of which some are already known to be involved in neurodevelopmental diseases (Figure 62).
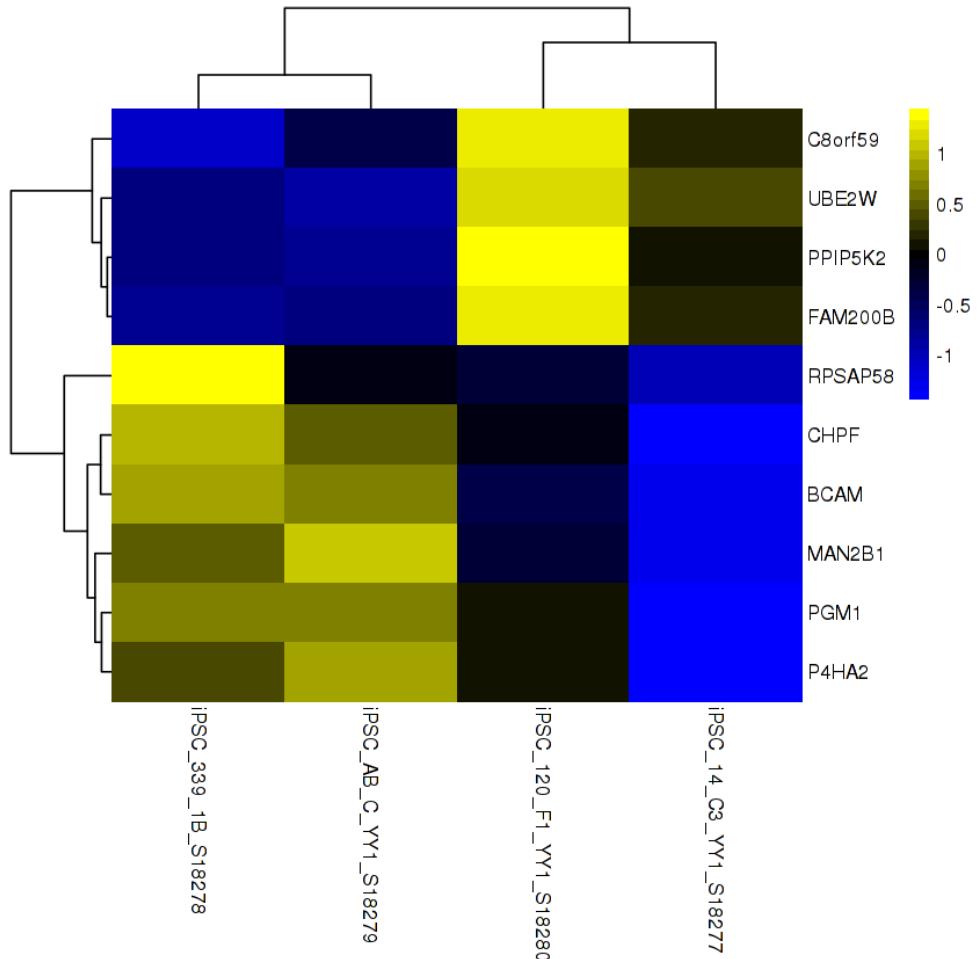
Figure 62 Expression profile of deregulated genes across disorders and cell types. Scalebar: log norm counts.

These genes are *PPIP5K2*, *PGM1*, *FAM200B*, and *MAN2B1*. It is relevant to underscore that *PPIP5K2, PGM1,* and *MAN2B1* are known to be associated with ASD, developmental delay, and a neurodevelopmental disorder, respectively. This preliminary analysis supports a model in which enhanceropaties caused by mutations in different genes converge towards regulatory pathways that may impact a subset of downstream genes responsible for the shared clinical manifestation.

BIBLIOGRAPHY

Abyzov, A., Mariani, J., Palejev, D., Zhang, Y., Haney, M.S., Tomasini, L., Ferrandino, A., Belmaker, L.A.R., Szekely, A., Wilson, M., Kocabas, A., Calixto, N.E., Grigorenko, E.L., Huttner, A., Chawarska, K., Weissman, S., Urban, A.E., Gerstein, M., Vaccarino, F.M., 2012. Somatic copy-number mosaicism in human skin revealed by induced pluripotent stem cells. Nature 492, 438–442. https://doi.org/10.1038/nature11629

Adam, M.P., Hudgins, L., 2005. Kabuki syndrome: a review. TL - 67. Clinical genetics 67 VN-r, 209–219. https://doi.org/10.1111/j.1399-0004.2004.00348.x

Adamo, A., Atashpaz, S., Germain, P.-L., Zanella, M., D'Agostino, G., Albertin, V., Chenoweth, J., Micale, L., Fusco, C., Unger, C., Augello, B., Palumbo, O., Hamilton, B., Carella, M., Donti, E., Pruneri, G., Selicorni, A., Biamino, E., Prontera, P., McKay, R., Merla, G., Testa, G., 2015. 7q11.23 dosage-dependent dysregulation in human pluripotent stem cells affects transcriptional programs in disease-relevant lineages TL - 47. Nature genetics 47 VN-r, 132–41. https://doi.org/10.1038/ng.3169

Adolphs, R., 2015. The unsolved problems of neuroscience. Trends Cogn Sci 19, 173–175. https://doi.org/10.1016/j.tics.2015.01.007

Agger, K., Cloos, P.A.C., Christensen, J., Pasini, D., Rose, S., Rappsilber, J., Issaeva, I., Canaani, E., Salcini, A.E., Helin, K., 2007. UTX and JMJD3 are histone H3K27 demethylases involved in *HOX* gene regulation and development. Nature 449, 731–734. https://doi.org/10.1038/nature06145

Allis, C.D., Berger, S.L., Cote, J., Dent, S., Jenuwien, T., Kouzarides, T., Pillus, L., Reinberg, D., Shi, Y., Shiekhattar, R., Shilatifard, A., Workman, J., Zhang, Y., 2007. New Nomenclature for Chromatin-Modifying Enzymes. Cell. https://doi.org/10.1016/j.cell.2007.10.039

Ang, S.-Y., Uebersohn, A., Spencer, C.I., Huang, Y., Lee, J.-E., Ge, K., Bruneau, B.G., 2016. KMT2D regulates specific programs in heart development via histone H3 lysine 4 di-methylation. Development 143, 810–821. https://doi.org/10.1242/dev.132688

Ang, Siang-Yun, Uebersohn, A., Spencer, C.I., Huang, Y., Lee, J.-E., Ge, K., Bruneau, B.G., 2016. KMT2D regulates specific programs in heart development via histone H3 lysine 4 di-methylation. Development (Cambridge, England) 143, 810–21. https://doi.org/10.1242/dev.132688

Aravind, L., Burroughs, A.M., Zhang, D., Iyer, L.M., 2014. Protein and DNA Modifications: Evolutionary Imprints of Bacterial Biochemical Diversification and Geochemistry on the Provenance of Eukaryotic Epigenetics. Cold Spring Harb Perspect Biol 6. https://doi.org/10.1101/cshperspect.a016063

Ardeshir-Larijani, F., Bhateja, P., Lipka, M.B., Sharma, N., Fu, P., Dowlati, A., 2018. KMT2D Mutation Is Associated With Poor Prognosis in Non-Small-Cell Lung Cancer. Clin Lung Cancer 19, e489–e501. https://doi.org/10.1016/j.cllc.2018.03.005

Asem, M.S., Buechler, S., Wates, R.B., Miller, D.L., Stack, M.S., 2016. Wnt5a Signaling in Cancer. Cancers (Basel) 8. https://doi.org/10.3390/cancers8090079

Augert, A., Zhang, Q., Bates, B., Cui, M., Wang, X., Wildey, G., Dowlati, A., MacPherson, D., 2017. Small Cell Lung Cancer Exhibits Frequent Inactivating Mutations in the Histone Methyltransferase KMT2D/MLL2:

CALGB 151111 (Alliance). J Thorac Oncol 12, 704–713. https://doi.org/10.1016/j.jtho.2016.12.011

Avery, O.T., MacLeod, C.M., McCarty, M., 1944. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii. Journal of Experimental Medicine 79, 137–158. https://doi.org/10.1084/jem.79.2.137

Banerji, J., Rusconi, S., Schaffner, W., 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell 27, 299–308. https://doi.org/10.1016/0092-8674(81)90413-X

Banka, S., Lederer, D., Benoit, V., Jenkins, E., Howard, E., Bunstone, S., Kerr, B., McKee, S., Lloyd, I.C., Shears, D., Stewart, H., White, S.M., Savarirayan, R., Mancini, G.M.S., Beysen, D., Cohn, R.D., Grisart, B., Maystadt, I., Donnai, D., 2015. Novel KDM6A (UTX) mutations and a clinical and molecular review of the X-linked Kabuki syndrome (KS2). Clin. Genet. 87, 252–258. https://doi.org/10.1111/cge.12363

Barbagiovanni, G., Germain, P.-L., Zech, M., Atashpaz, S., Lo Riso, P., D'Antonio-Chronowska, A., Tenderini, E., Caiazzo, M., Boesch, S., Jech, R., Haslinger, B., Broccoli, V., Stewart, A.F., Winkelmann, J., Testa, G., 2018. KMT2B Is Selectively Required for Neuronal Transdifferentiation, and Its Loss Exposes Dystonia Candidate Genes. Cell Reports 25, 988–1001. https://doi.org/10.1016/j.celrep.2018.09.067

Beagan, J.A., Duong, M.T., Titus, K.R., Zhou, L., Cao, Z., Ma, J., Lachanski, C.V., Gillis, D.R., Phillips-Cremins, J.E., 2017. YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. Genome Res. 27, 1139–1152. https://doi.org/10.1101/gr.215160.116

Beby, F., Lamonerie, T., 2013. The homeobox gene Otx2 in development and disease. Exp. Eye Res. 111, 9–16. https://doi.org/10.1016/j.exer.2013.03.007

Beck, D.B., Bonasio, R., Kaneko, S., Li, G., Li, G., Margueron, R., Oda, H., Sarma, K., Sims, R.J., Son, J., Trojer, P., Reinberg, D., 2010. Chromatin in the Nuclear Landscape. Cold Spring Harb Symp Quant Biol 75, 11–22. https://doi.org/10.1101/sqb.2010.75.052

Bernier, F.-E., Schreiber, A., Coulombe, J., Hatami, A., Marcoux, D., 2017. Pilomatricoma Associated with Kabuki Syndrome. Pediatric Dermatology 34, e26–e27. https://doi.org/10.1111/pde.13014

Blake, J.A., Ziman, M.R., 2014. Pax genes: regulators of lineage specification and progenitor cell maintenance. Development 141, 737–751. https://doi.org/10.1242/dev.091785

Blecher-Gonen, R., Barnett-Itzhaki, Z., Jaitin, D., Amann-Zalcenstein, D., Lara-Astiaso, D., Amit, I., 2013. High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. Nature Protocols 8, 539–554. https://doi.org/10.1038/nprot.2013.023

Bögershausen, N., Tsai, I.-C., Pohl, E., Kiper, P.Ö.S., Beleggia, F., Percin, E.F., Keupp, K., Matchan, A., Milz, E., Alanay, Y., Kayserili, H., Liu, Y., Banka, S., Kranz, A., Zenker, M., Wieczorek, D., Elcioglu, N., Prontera, P., Lyonnet, S., Meitinger, T., Stewart, A.F., Donnai, D., Strom, T.M., Boduroglu, K., Yigit, G., Li, Y., Katsanis, N., Wollnik, B., 2015. RAP1-mediated MEK/ERK pathway defects in Kabuki syndrome. J Clin Invest 125, 3585–3599. https://doi.org/10.1172/JCI80102

Bologna, L.L., Pasquale, V., Garofalo, M., Gandolfo, M., Baljon, P.L., Maccione, A., Martinoia, S., Chiappalone, M., 2010. Investigating neuronal activity by SPYCODE multi-channel data analyzer. Neural Netw 23, 685–697. https://doi.org/10.1016/j.neunet.2010.05.002

Boniolo, G., Testa, G., 2012. The Identity of Living Beings, Epigenetics, and the Modesty of Philosophy. Erkenn 76, 279–298. https://doi.org/10.1007/s10670-011-9308-9

Bossi, D., Cicalese, A., Dellino, G.I., Luzi, L., Riva, L., D'Alesio, C., Diaferia, G.R., Carugo, A., Cavallaro, E., Piccioni, R., Barberis, M., Mazzarol, G., Testori, A., Punzi, S., Pallavicini, I., Tosti, G., Giacó, L., Melloni, G., Heffernan, T.P., Natoli, G., Draetta, G.F., Minucci, S., Pelicci, P., Lanfrancone, L., 2016. In Vivo Genetic Screens of Patient-Derived Tumors Revealed Unexpected Frailty of the Transformed Phenotype. Cancer discovery 6, 650–63. https://doi.org/10.1158/2159-8290.CD-15-1200

Bottiglione, F., Giurisato, E., 2015. Scaffold KSR2 overexpression is associated with melanoma A375 cells resistance to Vemurafenib. Medical Research Archives 2.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., Gifford, D.K., Melton, D.A., Jaenisch, R., Young, R.A., 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 122, 947–956. https://doi.org/10.1016/j.cell.2005.08.020

Brown, J.L., Mucci, D., Whiteley, M., Dirksen, M.-L., Kassis, J.A., 1998. The Drosophila Polycomb Group Gene pleiohomeotic Encodes a DNA Binding Protein with Homology to the Transcription Factor YY1. Molecular Cell 1, 1057–1064. https://doi.org/10.1016/S1097-2765(00)80106-9

Burgold, T., Spreafico, F., Santa, F, Totaro, M.G., Prosperini, E., Natoli, G., Testa, G., De Santa, Francesca, Totaro, M.G., Prosperini, E., Natoli, G., Testa, G., 2008. The histone H3 lysine 27-specific demethylase Jmjd3 is required for neural commitment. PLoS One 3 VN-re, e3034. https://doi.org/10.1371/journal.pone.0003034

Burgold, T., Voituron, N., Caganova, M., Tripathi, P.P., Menuet, C., Tusi, B.K., Spreafico, F., Bévengut, M., Gestreau, C., Buontempo, S., Simeone, A., Kruidenier, L., Natoli, G., Casola, S., Hilaire, G., Testa, G., Bevengut, M., Gestreau, C., Buontempo, S., Simeone, A., Kruidenier, L., Natoli, G., Casola, S., Hilaire, G., Testa, G., 2012. The H3K27 demethylase JMJD3 is required for maintenance of the embryonic respiratory neuronal network, neonatal breathing, and survival. Cell Rep 2 VN-re, 1244–58. https://doi.org/10.1016/j.celrep.2012.09.013

Cai, Y., Jin, J., Yao, T., Gottschalk, A.J., Swanson, S.K., Wu, S., Shi, Y., Washburn, M.P., Florens, L., Conaway, R.C., Conaway, J.W., 2007. YY1 functions with INO80 to activate transcription. Nat. Struct. Mol. Biol. 14, 872–874. https://doi.org/10.1038/nsmb1276

Calo, E., Wysocka, J., 2013. Modification of enhancer chromatin: what, how and why? Mol Cell 49. https://doi.org/10.1016/j.molcel.2013.01.038

Campbell, D.B., Sutcliffe, J.S., Ebert, P.J., Militerni, R., Bravaccio, C., Trillo, S., Elia, M., Schneider, C., Melmed, R., Sacco, R., Persico, A.M., Levitt, P., 2006. A genetic variant that disrupts MET transcription is associated with autism. Proc. Natl. Acad. Sci. U.S.A. 103, 16834–16839. https://doi.org/10.1073/pnas.0605296103

Caretti, G., Padova, M.D., Micales, B., Lyons, G.E., Sartorelli, V., 2004. The Polycomb Ezh2 methyltransferase regulates muscle gene expression and skeletal muscle differentiation. Genes Dev. 18, 2627–2638. https://doi.org/10.1101/gad.1241904

Carlén, M., 2017. What constitutes the prefrontal cortex? Science 358, 478–482. https://doi.org/10.1126/science.aan8868

Casanova, M., Selicorni, A., Ferrari, A., 2011. Cancer predisposition in children with Kabuki syndrome. Am. J. Med. Genet. A 155A, 1504. https://doi.org/10.1002/ajmg.a.33711

Cho, A.-R., Uchio-Yamada, K., Torigai, T., Miyamoto, T., Miyoshi, I., Matsuda, J., Kurosawa, T., Kon, Y., Asano, A., Sasaki, N., Agui, T., 2006. Deficiency of the tensin2 gene in the ICGN mouse: an animal model for congenital nephrotic syndrome. Mamm. Genome 17, 407–416. https://doi.org/10.1007/s00335-005-0167-z

Cho, Y.-W.W., Hong, T., Hong, S., Guo, H., Yu, H., Kim, D., Guszczynski, T., Dressler, G.R., Copeland, T.D., Kalkum, M., Ge, K., 2007. PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. TL - 282. The Journal of biological chemistry 282 VN-, 20395–20406. https://doi.org/10.1074/jbc.M701574200

Chua, C.E.L., Tang, B.L., 2015. The role of the small GTPase Rab31 in cancer. J Cell Mol Med 19, 1–10. https://doi.org/10.1111/jcmm.12403

Comoglio, P.M., Trusolino, L., Boccaccio, C., 2018. Known and novel roles of the MET oncogene in cancer: a coherent approach to targeted therapy. Nature Reviews Cancer 18, 341–358. https://doi.org/10.1038/s41568-018-0002-y

Craft, A.M., Johnson, M., 2017. From stem cells to human development: a distinctly human perspective on early embryology, cellular differentiation and translational research. Development 144, 12–16. https://doi.org/10.1242/dev.142778

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., Boyer, L.A., Young, R.A., Jaenisch, R., 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences of the United States of America 107, 21931–6. https://doi.org/10.1073/pnas.1016071107

Cutter, A.R., Hayes, J.J., 2015. A brief review of nucleosome structure. FEBS Lett. 589, 2914–2922. https://doi.org/10.1016/j.febslet.2015.05.016

Davies, J.O.J., Oudelaar, A.M., Higgs, D.R., Hughes, J.R., 2017. How best to identify chromosomal interactions: a comparison of approaches. Nat. Methods 14, 125–134. https://doi.org/10.1038/nmeth.4146

Deciphering Developmental Disorders Study, 2017. Prevalence and architecture of de novo mutations in developmental disorders. Nature 542, 433–438. https://doi.org/10.1038/nature21062

Dekker, J., Belmont, A.S., Guttman, M., Leshyk, V.O., Lis, J.T., Lomvardas, S., Mirny, L.A., O'Shea, C.C., Park, P.J., Ren, B., Politz, J.C.R., Shendure, J., Zhong, S., Network, the 4D N., 2017. The 4D nucleome project. Nature 549, 219–226. https://doi.org/10.1038/nature23884

Digilio, M.C., Marino, B., Toscano, A., Giannotti, A., Dallapiccola, B., 2001. Congenital heart defects in Kabuki syndrome. American journal of medical genetics 100, 269–74.

Donohoe, M.E., Zhang, X., McGinnis, L., Biggers, J., Li, E., Shi, Y., 1999. Targeted disruption of mouse Yin Yang 1 transcription factor results in peri-implantation lethality. Molecular and cellular biology 19, 7237–44.

Dorighi, K.M., Swigut, T., Henriques, T., Bhanu, N. V., Scruggs, B.S., Nady, N., Still, C.D., Garcia, B.A., Adelman, K., Wysocka, J., 2017. Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation, Molecular Cell. https://doi.org/10.1016/j.molcel.2017.04.018

Dupin, E., Calloni, G.W., Coelho-Aguiar, J.M., Le Douarin, N.M., 2018. The issue of the multipotency of the neural crest cells. Dev. Biol. https://doi.org/10.1016/j.ydbio.2018.03.024

ENCODE Project Consortium, 2012. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74. https://doi.org/10.1038/nature11247

Eriksson, P.S., Perfilieva, E., Björk-Eriksson, T., Alborn, A.-M., Nordborg, C., Peterson, D.A., Gage, F.H., 1998. Neurogenesis in the adult human hippocampus. Nature Medicine 4, 1313–1317. https://doi.org/10.1038/3305

Evans, M.J., Kaufman, M.H., 1981. Establishment in culture of pluripotential cells from mouse embryos. Nature 292, 154–156. https://doi.org/10.1038/292154a0

Ezponda, T., Dupéré-Richer, D., Will, C.M., Small, E.C., Varghese, N., Patel, T., Nabet, B., Popovic, R., Oyer, J., Bulic, M., Zheng, Y., Huang, X., Shah, M.Y., Maji, S., Riva, A., Occhionorelli, M., Tonon, G., Kelleher, N., Keats, J., Licht, J.D., 2017. UTX/KDM6A Loss Enhances the Malignant Phenotype of Multiple Myeloma and Sensitizes Cells to EZH2 inhibition. Cell Reports 21, 628–640. https://doi.org/10.1016/j.celrep.2017.09.078

Fernandez, M.R., Henry, M.D., Lewis, R.E., 2012. Kinase Suppressor of Ras 2 (KSR2) Regulates Tumor Cell Transformation via AMPK. Mol Cell Biol 32, 3718–3731. https://doi.org/10.1128/MCB.06754-11

Ford, D.J., Dingwall, A.K., 2015. The cancer COMPASS: navigating the functions of MLL complexes in cancer. Cancer genetics 208, 178–91. https://doi.org/10.1016/j.cancergen.2015.01.005

Fragola, G., Germain, P.-L.L., Laise, P., Cuomo, A., Blasimme, A., Gross, F., Signaroldi, E., Bucci, G., Sommer, C., Pruneri, G., Mazzarol, G., Bonaldi, T., Mostoslavsky, G., Casola, S., Testa, G., 2013. Cell reprogramming requires silencing of a core subset of polycomb targets TL - 9. PLoS Genet 9, e1003292. https://doi.org/10.1371/journal.pgen.1003292

Frank, S.R., Schroeder, M., Fernandez, P., Taubert, S., Amati, B., 2001. Binding of c-Myc to chromatin mediates mitogen-induced acetylation of histone H4 and gene activation. Genes & development 15, 2069–82. https://doi.org/10.1101/gad.906601

Frega, M., Selten, M., Mossink, B., Keller, J.M., Linda, K., Moerschen, R., Qu, J., Koerner, P., Jansen, S., Bijvank, E., Oudakker, A., Kleefstra, T., Bokhoven, H. van, Zhou, H., Schubert, D., Kasri, N.N., 2018. Distinct pathogenic genes causing intellectual disability and autism exhibit overlapping effects on neuronal network development. bioRxiv 408252. https://doi.org/10.1101/408252

Froimchuk, E., Jang, Y., Ge, K., 2017. Histone H3 lysine 4 methyltransferase KMT2D. Gene 627, 337–342. https://doi.org/10.1016/j.gene.2017.06.056

Fulco, C.P., Munschauer, M., Anyoha, R., Munson, G., Grossman, S.R., Perez, E.M., Kane, M., Cleary, B., Lander, E.S., Engreitz, J.M., 2016. Systematic mapping of functional enhancer–promoter connections with CRISPR interference. Science 354, 769–773. https://doi.org/10.1126/science.aag2445

Fusaki, N., Ban, H., Nishiyama, A., Saeki, K., Hasegawa, M., 2009. Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. Proc. Jpn. Acad., Ser. B, Phys. Biol. Sci. 85, 348–362.

Gabriele, M., Lopez Tobon, A., D'Agostino, G., Testa, G., 2018. The chromatin basis of neurodevelopmental disorders: Rethinking dysfunction along the molecular and temporal axes. Progress in Neuro-Psychopharmacology and Biological Psychiatry. https://doi.org/10.1016/j.pnpbp.2017.12.013

Gabriele, M., Silfhout, A.T.V., Germain, P.-L., Vitriolo, A., Kumar, R., Douglas, E., Haan, E., Kosaki, K., Takenouchi, T., Rauch, A., Steindl, K., Frengen, E., Misceo, D., Pedurupillay, C.R.J., Stromme, P., Rosenfeld, J.A., Shao, Y., Craigen, W.J., Schaaf, C.P., Rodriguez-Buritica, D., Farach, L., Friedman, J.,

119

Thulin, P., McLean, S.D., Nugent, K.M., Morton, J., Nicholl, J., Andrieux, J., Stray-Pedersen, A., Chambon, P., Patrier, S., Lynch, S.A., Kjaergaard, S., Tørring, P.M., Brasch-Andersen, C., Ronan, A., Haeringen, A. van, Anderson, P.J., Powis, Z., Brunner, H.G., Pfundt, R., Schuurs-Hoeijmakers, J.H.M., Bon, B.W.M. van, Lelieveld, S., Gilissen, C., Nillesen, W.M., Vissers, L.E.L.M., Gecz, J., Koolen, D.A., Testa, G., Vries, B.B.A. de, 2017. YY1 Haploinsufficiency Causes an Intellectual Disability Syndrome Featuring Transcriptional and Chromatin Dysfunction. The American Journal of Human Genetics 100, 907–925. https://doi.org/10.1016/j.ajhg.2017.05.006

Gerasimova, T., Guo, C., Ghosh, A., Qiu, X., Montefiori, L., Verma-Gaur, J., Choi, N.M., Feeney, A.J., Sen, R., 2015. A structural hierarchy mediated by multiple nuclear factors establishes IgH locus conformation. Genes Dev. 29, 1683–1695. https://doi.org/10.1101/gad.263871.115

Germain, P.-L., Testa, G., 2017. Taming Human Genetic Variability: Transcriptomic Meta-Analysis Guides the Experimental Design and Interpretation of iPSC-Based Disease Modeling. Stem Cell Reports 8, 1784–1796. https://doi.org/10.1016/j.stemcr.2017.05.012

Germain, P.-L., Vitriolo, A., Adamo, A., Laise, P., Das, V., Testa, G., 2016. RNAontheBENCH: computational and empirical resources for benchmarking RNAseq quantification and differential expression methods. Nucleic acids research. https://doi.org/10.1093/nar/gkw448

Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., Furlong, E.E.M., 2014. Enhancer loops appear stable during development and are associated with paused polymerase. Nature 512, 96–100. https://doi.org/10.1038/nature13417

Ginalski, K., Rychlewski, L., Baker, D., Grishin, N.V., 2004. Protein structure prediction for the male-specific region of the human Y chromosome. PNAS 101, 2305–2310. https://doi.org/10.1073/pnas.0306306101

Goo, Y.-H., Sohn, Y.C., Kim, D.-H., Kim, S.-W., Kang, M.-J., Jung, D.-J., Kwak, E., Barlev, N.A., Berger, S.L., Chow, V.T., Roeder, R.G., Azorsa, D.O., Meltzer, P.S., Suh, P.-G., Song, E.J., Lee, K.-J., Lee, Y.C., Lee, J.W., 2003. Activating signal cointegrator 2 belongs to a novel steady-state complex that contains a subset of trithorax group proteins. Molecular and cellular biology 23, 140–9. https://doi.org/10.1128/MCB.23.1.140-149.2003

Greenfield, A., Carrel, L., Pennisi, D., Philippe, C., Quaderi, N., Siggers, P., Steiner, K., Tam, P.P.L., Monaco, A.P., Willard, H.F., Koopman, P., 1998. The UTX Gene Escapes X Inactivation in Mice and Humans. Hum Mol Genet 7, 737–742. https://doi.org/10.1093/hmg/7.4.737

Guo, C., Chang, C.-C., Wortham, M., Chen, L.H., Kernagis, D.N., Qin, X., Cho, Y.-W., Chi, J.-T., Grant, G.A., McLendon, R.E., Yan, H., Ge, K., Papadopoulos, N., Bigner, D.D., He, Y., 2012. Global identification of MLL2-targeted loci reveals MLL2's role in diverse signaling pathways. Proc Natl Acad Sci U S A 109, 17603–17608. https://doi.org/10.1073/pnas.1208807109

Guo, C., Chen, L.H., Huang, Y., Chang, C.-C., Wang, P., Pirozzi, C.J., Qin, X., Bao, X., Greer, P.K., McLendon, R.E., Yan, H., Keir, S.T., Bigner, D.D., He, Y., 2013. KMT2D maintains neoplastic cell proliferation and global histone H3 lysine 4 monomethylation. Oncotarget 4, 2144–53. https://doi.org/10.18632/oncotarget.1555

Guo, L., Costanzo-Garvey, D.L., Smith, D.R., Neilsen, B.K., MacDonald, R.G., Lewis, R.E., 2017. Kinase Suppressor of Ras 2 (KSR2) expression in the brain regulates energy balance and glucose homeostasis. Molecular Metabolism 6, 194–205. https://doi.org/10.1016/j.molmet.2016.12.004

Haigis, K.M., 2017. KRAS Alleles: The Devil Is in the Detail. Trends in Cancer 3, 686–697. https://doi.org/10.1016/j.trecan.2017.08.006

Hayashi, H., Arao, T., Togashi, Y., Kato, H., Fujita, Y., De Velasco, M.A., Kimura, H., Matsumoto, K., Tanaka, K., Okamoto, I., Ito, A., Yamada, Y., Nakagawa, K., Nishio, K., 2015. The OCT4 pseudogene POU5F1B is amplified and promotes an aggressive phenotype in gastric cancer. Oncogene 34, 199–208. https://doi.org/10.1038/onc.2013.547

He, Y., Casaccia-Bonnefil, P., 2008. The Yin and Yang of YY1 in the nervous system. J Neurochem 106, 1493–1502. https://doi.org/10.1111/j.1471-4159.2008.05486.x

He, Y., Dupree, J., Wang, J., Sandoval, J., Li, J., Liu, H., Shi, Y., Nave, K.A., Casaccia-Bonnefil, P., 2007. The Transcription Factor Yin Yang 1 Is Essential for Oligodendrocyte Progenitor Differentiation. Neuron 55, 217–230. https://doi.org/10.1016/j.neuron.2007.06.029

He, Y., Kim, J.Y., Dupree, J., Tewari, A., Melendez-Vasquez, C., Svaren, J., Casaccia, P., 2010. Yy1 as a molecular link between neuregulin and transcriptional modulation of peripheral myelination. Nature neuroscience 13, 1472–80. https://doi.org/10.1038/nn.2686

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., Wang, W., Weng, Z., Green, R.D., Crawford, G.E., Ren, B., Hawkins, D.R., Barrera, L.O., Calcar, S., Qu, C., Ching, K.A., Wang, W., Weng, Z., Green, R.D., Crawford, G.E., Ren, B., 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature Genetics 39, 311–8. https://doi.org/10.1038/ng1966

Hershey, A.D., Chase, M., 1952. INDEPENDENT FUNCTIONS OF VIRAL PROTEIN AND NUCLEIC ACID IN GROWTH OF BACTERIOPHAGE. J Gen Physiol 36, 39–56.

Hong, S., Cho, Y.-W., Yu, L.-R., Yu, H., Veenstra, T.D., Ge, K., 2007. Identification of JmjC domain-containing UTX and JMJD3 as histone H3 lysine 27 demethylases. PNAS 104, 18439–18444. https://doi.org/10.1073/pnas.0707292104

Hu, Z., Tee, W.-W., 2017. Enhancers and chromatin structures: regulatory hubs in gene expression and diseases. Biosci Rep 37. https://doi.org/10.1042/BSR20160183

Hughes, H.E., Davies, S.J., 1994. Coarctation of the aorta in Kabuki syndrome. Archives of disease in childhood 70, 512–4.

Hyun, I., Hochedlinger, K., Jaenisch, R., Yamanaka, S., 2007. New Advances in iPS Cell Research Do Not Obviate the Need for Human Embryonic Stem Cells. Cell Stem Cell 1, 367–368. https://doi.org/10.1016/j.stem.2007.09.006

Ilic, D., Ogilvie, C., 2017. Concise Review: Human Embryonic Stem Cells-What Have We Done? What Are We Doing? Where Are We Going? Stem Cells 35, 17–25. https://doi.org/10.1002/stem.2450

Ilyina, H., Lurie, I., Naumtchik, I., Amoashy, D., Stephanenko, G., Fedotov, V., Kostjuk, A., 1995. Kabuki make-up (Niikawa-Kuroki) syndrome in the Byelorussian register of congenital malformations: ten new observations. American journal of medical genetics 56, 127–31. https://doi.org/10.1002/ajmg.1320560202

Inoue, F., Kircher, M., Martin, B., Cooper, G.M., Witten, D.M., McManus, M.T., Ahituv, N., Shendure, J., 2017. A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. Genome Res. 27, 38–52. https://doi.org/10.1101/gr.212092.116

Issaeva, I., Zonis, Y., Rozovskaia, T., Orlovsky, K., Croce, C.M., Nakamura, T., Mazo, A., Eisenbach, L., Canaani, E., 2007. Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. Molecular and cellular biology 27, 1889–903. https://doi.org/10.1128/MCB.01506-06

Jacob, F., Monod, J., 1961. Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. 3, 318–356.

Jang, Y., Wang, C., Zhuang, L., Liu, C., Ge, K., 2017. H3K4 Methyltransferase Activity Is Required for MLL4 Protein Stability. Journal of Molecular Biology 429, 2046–2054. https://doi.org/10.1016/j.jmb.2016.12.016

Jeon, Y., Lee, J.T., 2011. YY1 tethers Xist RNA to the inactive X nucleation center. Cell 146, 119–133. https://doi.org/10.1016/j.cell.2011.06.026

Jin, Z., Liu, Y., 2018. DNA methylation in human diseases. Genes & Diseases 5, 1–8. https://doi.org/10.1016/j.gendis.2018.01.002

Jost, D., Carrivain, P., Cavalli, G., Vaillant, C., 2014. Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. Nucleic Acids Res 42, 9553–9561. https://doi.org/10.1093/nar/gku698

Kaikkonen, M.U., Spann, N.J., Heinz, S., Romanoski, C.E., Allison, K.A., Stender, J.D., Chun, H.B., Tough, D.F., Prinjha, R.K., Benner, C., Glass, C.K., 2013. Remodeling of the Enhancer Landscape during Macrophage Activation Is Coupled to Enhancer Transcription. Molecular Cell 51, 310–325. https://doi.org/10.1016/j.molcel.2013.07.010

Kamakaka, R.T., Biggins, S., 2005. Histone variants: deviants? Genes Dev. 19, 295–316. https://doi.org/10.1101/gad.1272805

Kandoth, C., McLellan, M.D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J.F., Wyczalkowski, M.A., Leiserson, M.D.M., Miller, C.A., Welch, J.S., Walter, M.J., Wendl, M.C., Ley, T.J., Wilson, R.K., Raphael, B.J., Ding, L., 2013. Mutational landscape and significance across 12 major cancer types. TL - 502. Nature 502 VN-, 333–339. https://doi.org/10.1038/nature12634

Kantidakis, T., Saponaro, M., Mitter, R., Horswell, S., Kranz, A., Boeing, S., Aygün, O., Kelly, G.P., Matthews, N., Stewart, A.F., Stewart, A.F., Svejstrup, J.Q., 2016. Mutation of cancer driver MLL2 results in transcription stress and genome instability. Genes & development 30, 408–20. https://doi.org/10.1101/gad.275453.115

Kappetein, A.P., Gittenberger-de Groot, A.C., Zwinderman, A.H., Rohmer, J., Poelmann, R.E., Huysmans, H.A., 1991. The neural crest as a possible pathogenetic factor in coarctation of the aorta and bicuspid aortic valve. J. Thorac. Cardiovasc. Surg. 102, 830–836.

Kawame, H., Hannibal, M.C., Hudgins, L., Pagon, R.A., 1999. Phenotypic spectrum and management issues in Kabuki syndrome. The Journal of Pediatrics 134, 480–485. https://doi.org/10.1016/S0022-3476(99)70207-6

Keyte, A., Hutson, M.R., 2012. The Neural Crest in Cardiac Congenital Anomalies. Differentiation 84, 25–40. https://doi.org/10.1016/j.diff.2012.04.005

Kilpinen, H., Goncalves, A., Leha, A., Afzal, V., Alasoo, K., Ashford, S., Bala, S., Bensaddek, D., Casale, F.P., Culley, O.J., Danecek, P., Faulconbridge, A., Harrison, P.W., Kathuria, A., McCarthy, D., McCarthy, S.A., Meleckyte, R., Memari, Y., Moens, N., Soares, F., Mann, A., Streeter, I., Agu, C.A., Alderton, A., Nelson, R., Harper, S., Patel, M., White, A., Patel, S.R., Clarke, L., Halai, R., Kirton, C.M., Kolb-Kokocinski, A., Beales, P., Birney, E., Danovi, D., Lamond, A.I., Ouwehand, W.H., Vallier, L., Watt, F.M., Durbin, R., Stegle, O., Gaffney, D.J., 2017. Common genetic variation drives molecular heterogeneity in human iPSCs. Nature 546, 370–375. https://doi.org/10.1038/nature22403

Kim, D.-H., Kim, J., Kwon, J.-S., Lee, S.-K., Lee, S., Lee, J.W., Sandhu, J., Tontonoz, P., 2016. Critical Roles of the Histone Methyltransferase MLL4/KMT2D in Murine Hepatic Steatosis Directed by ABL1 and PPARγ2. Cell Reports 17, 1671–1682. https://doi.org/10.1016/j.celrep.2016.10.023

Kim, D.-H., Rhee, J.C., Yeo, S., Shen, R., Lee, S.-K., Lee, J.W., Lee, S., 2015. Crucial roles of mixed-lineage leukemia 3 and 4 as epigenetic switches of the hepatic circadian clock controlling bile acid homeostasis in mice. Hepatology 61, 1012–1023. https://doi.org/10.1002/hep.27578

Kim, J.-H., Sharma, A., Dhar, S.S., Lee, S.-H., Gu, B., Chan, C.-H., Lin, H.-K., Lee, M.G., 2014. UTX and MLL4 coordinately regulate transcriptional programs for cell proliferation and invasiveness in breast cancer cells. Cancer research 74, 1705–17. https://doi.org/10.1158/0008-5472.CAN-13-1896

Knauss, J.L., Miao, N., Kim, S.-N., Nie, Y., Shi, Y., Wu, T., Pinto, H.B., Donohoe, M.E., Sun, T., 2018. Long noncoding RNA Sox2ot and transcription factor YY1 co-regulate the differentiation of cortical neural progenitors by repressing Sox2. Cell Death & Disease 9, 799. https://doi.org/10.1038/s41419-018-0840-2

Kruidenier, L., Chung, C., Cheng, Z., Liddle, J., Che, K., Joberty, G., Bantscheff, M., Bountra, C., Bridges, A., Diallo, H., Eberhard, D., Hutchinson, S., Jones, E., Katso, R., Leveridge, M., Mander, P.K., Mosley, J., Ramirez-Molina, C., Rowland, P., Schofield, C.J., Sheppard, R.J., Smith, J.E., Swales, C., Tanner, R., Thomas, P., Tumber, A., Drewes, G., Oppermann, U., Patel, D.J., Lee, K., Wilson, D.M., 2012. A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. Nature 488, 404–8. https://doi.org/10.1038/nature11262

Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., Amin, V., Whitaker, J.W., Schultz, M.D., Ward, L.D., Sarkar, A., Quon, G., Sandstrom, R.S., Eaton, M.L., Wu, Y.-C., Pfenning, A., Wang, X., Liu, M., Coarfa, C., Harris, A.R., Shoresh, N., Epstein, C.B., Gjoneska, E., Leung, D., Xie, W., Hawkins, D.R., Lister, R., Hong, C., Gascard, P., Mungall, A.J., Moore, R., Chuah, E., Tam, A., Canfield, T.K., Hansen, S.R., Kaul, R., Sabo, P.J., Bansal, M.S., Carles, A., Dixon, J.R., Farh, K.-H., Feizi, S., Karlic, R., Kim, A.-R., Kulkarni, A., Li, D., Lowdon, R., Elliott, G., Mercer, T.R., Neph, S.J., Onuchic, V., Polak, P., Rajagopal, N., Ray, P., Sallari, R.C., Siebenthall, K.T., Sinnott-Armstrong, N.A., Stevens, M., Thurman, R.E., Wu, J., Zhang, B., Zhou, X., Abdennur, N., Adli, M., Akerman, M., Barrera, L., Antosiewicz-Bourget, J., Ballinger, T., Barnes, M.J., Bates, D., Bell, R.J.A., Bennett, D.A., Bianco, K., Bock, C., Boyle, P., Brinchmann, J., Caballero-Campo, P., Camahort, R., Carrasco-Alfonso, M.J., Charnecki, T., Chen, H., Chen, Z., Cheng, J.B., Cho, S., Chu, A., Chung, W.-Y., Cowan, C., Deng, Q., Deshpande, V., Diegel, M., Ding, B., Durham, T., Echipare, L., Edsall, L., Flowers, D., Genbacev-Krtolica, O., Gifford, C., Gillespie, S., Giste, E., Glass, I.A., Gnirke, A., Gormley, M., Gu, H., Gu, J., Hafler, D.A., Hangauer, M.J., Hariharan, M., Hatan, M., Haugen, E., He, Y., Heimfeld, S., Herlofsen, S., Hou, Z., Humbert, R., Issner, R., Jackson, A.R., Jia, H., Jiang, P., Johnson, A.K., Kadlecek, T., Kamoh, B., Kapidzic, M., Kent, J., Kim, A., Kleinewietfeld, M., Klugman, S., Krishnan, J., Kuan, S., Kutyavin, T., Lee, A.-Y., Lee, K., Li, J., Li, N., Li, Y., Ligon, K.L., Lin, S., Lin, Y., Liu, J., Liu, Yuxuan, Luckey, J.C., Ma, Y.P., Maire, C., Marson, A., Mattick, J.S., Mayo, M., McMaster, M., Metsky, H., Mikkelsen, T., Miller, D., Miri, M., Mukame, E., Nagarajan, R.P., Neri, F., Nery, J., Nguyen, T., O'Geen, H., Paithankar, S., Papayannopoulou, T., Pelizzola, M., Plettner, P., Propson, N.E., Raghuraman, S., Raney, B.J., Raubitschek, A., Reynolds, A.P., Richards, H., Riehle, K., Rinaudo, P., Robinson, J.F., Rockweiler, N.B., Rosen, E., Rynes, E., Schein, J., Sears, R., Sejnowski, T., Shafer, A., Shen, L., Shoemaker, R., Sigaroudinia, M., Slukvin, I., Stehling-Sun, S., Stewart, R., Subramanian, S., Suknuntha, K., Swanson, S., Tian, S., Tilden, H., Tsai, L., Urich, M., Vaughn, I., Vierstra, J., Vong, S., Wagner, U., Wang, H., Wang,

Tao, Wang, Y., Weiss, A., Whitton, H., Wildberg, A., Witt, H., Won, K.-J., Xie, M., Xing, X., Xu, I., Xuan, Z., Ye, Z., Yen, C., Yu, P., Zhang, Xian, Zhang, Xiaolan, Zhao, J., Zhou, Y., Zhu, J., Zhu, Y., Ziegler, S., Beaudet, A.E., Boyer, L.A., Jager, P.L., Farnham, P.J., Fisher, S.J., Haussler, D., Jones, S.J.M., Li, W., Marra, M.A., McManus, M.T., Sunyaev, S., Thomson, J.A., Tlsty, T.D., Tsai, L.-H., Wang, W., Waterland, R.A., Zhang, M.Q., Chadwick, L.H., Bernstein, B.E., Costello, J.F., Ecker, J.R., Hirst, M., Meissner, A., Milosavljevic, A., Ren, B., Stamatoyannopoulos, J.A., Wang, Ting, Kellis, M., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., Amin, V., Whitaker, J.W., Schultz, M.D., Ward, L.D., Sarkar, A., Quon, G., Sandstrom, R.S., Eaton, M.L., Wu, Y.-C., Pfenning, A.R., Wang, X., Claussnitzer, M., Liu, Yaping, Coarfa, C., Harris, A.R., Shoresh, N., Epstein, C.B., Gjoneska, E., Leung, D., Xie, W., Hawkins, D.R., Lister, R., Hong, C., Gascard, P., Mungall, A.J., Moore, R., Chuah, E., Tam, A., Canfield, T.K., Hansen, S.R., Kaul, R., Sabo, P.J., Bansal, M.S., Carles, A., Dixon, J.R., Farh, K.-H., Feizi, S., Karlic, R., Kim, A.-R., Kulkarni, A., Li, D., Lowdon, R., Elliott, G., Mercer, T.R., Neph, S.J., Onuchic, V., Polak, P., Rajagopal, N., Ray, P., Sallari, R.C., Siebenthall, K.T., Sinnott-Armstrong, N.A., Stevens, M., Thurman, R.E., Wu, J., Zhang, B., Zhou, X., Beaudet, A.E., Boyer, L.A., Jager, P.L., Farnham, P.J., Fisher, S.J., Haussler, D., Jones, S.J.M., Li, W., Marra, M.A., McManus, M.T., Sunyaev, S., Thomson, J.A., Tlsty, T.D., Tsai, L.-H., Wang, W., Waterland, R.A., Zhang, M.Q., Chadwick, L.H., Bernstein, B.E., Costello, J.F., Ecker, J.R., Hirst, M., Meissner, A., Milosavljevic, A., Ren, B., Stamatoyannopoulos, J.A., Wang, Ting, Kellis, M., 2015. Integrative analysis of 111 reference human epigenomes TL - 518. Nature 518 VN-, 317–330. https://doi.org/10.1038/nature14248

Kuroki, Y., Suzuki, Y., Chyo, H., Hata, A., Matsui, I., 1981. A new malformation syndrome of long palpebralfissures, large ears, depressed nasal tip, and skeletal anomalies associated with postnatal dwarfism and mental retardation. The Journal of Pediatrics 99, 570–573. https://doi.org/10.1016/S0022-3476(81)80256-9

Lai, B., Lee, J.-E., Jang, Y., Wang, L., Peng, W., Ge, K., 2017. MLL3/MLL4 are required for CBP/p300 binding on enhancers and super-enhancer formation in brown adipogenesis. Nucleic Acids Research 45, 6388–6403. https://doi.org/10.1093/nar/gkx234

Lan, F., Bayliss, P.E., Rinn, J.L., Whetstine, J.R., Wang, J.K., Chen, S., Iwase, S., Alpatov, R., Issaeva, I., Canaani, E., Roberts, T.M., Chang, H.Y., Shi, Y., 2007. A histone H3 lysine 27 demethylase regulates animal posterior development. Nature 449, 689–694. https://doi.org/10.1038/nature06192

Lawrence, M., Daujat, S., Schneider, R., 2016. Lateral Thinking: How Histone Modifications Regulate Gene Expression. Trends in Genetics 32, 42–56. https://doi.org/10.1016/j.tig.2015.10.007

Lederer, D., Grisart, B., Digilio, M.C., Benoit, V., Crespin, M., Ghariani, S.C., Maystadt, I., Dallapiccola, B., Verellen-Dumoulin, C., 2012. Deletion of KDM6A, a Histone Demethylase Interacting with MLL2, in Three Patients with Kabuki Syndrome, The American Journal of Human Genetics. https://doi.org/10.1016/j.ajhg.2011.11.021

Lee, J.-E., Wang, C., Xu, S., Cho, Y.-W., Wang, L., Feng, X., Baldridge, A., Sartorelli, V., Zhuang, L., Peng, W., Ge, K., 2013a. H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. eLife 2, e01503. https://doi.org/10.7554/eLife.01503

Lee, J.-E., Wang, C., Xu, S., Cho, Y.-W., Wang, L., Feng, X., Baldridge, A., Sartorelli, V., Zhuang, L., Peng, W., Ge, K., 2013b. H3K4 mono- and di-

methyltransferase MLL4 is required for enhancer activation during cell differentiation. eLife Sciences 2, e01503. https://doi.org/10.7554/eLife.01503

Lee, J.S., Galvin, K.M., See, R.H., Eckner, R., Livingston, D., Moran, E., Shi, Y., 1995. Relief of YY1 transcriptional repression by adenovirus E1A is mediated by E1A-associated protein p300. Genes Dev. 9, 1188–1198. https://doi.org/10.1101/gad.9.10.1188

Lee, M.G., Villa, R., Trojer, P., Norman, J., Yan, K.-P., Reinberg, D., Croce, L.D., Shiekhattar, R., 2007. Demethylation of H3K27 Regulates Polycomb Recruitment and H2A Ubiquitination. Science 318, 447–450. https://doi.org/10.1126/science.1149042

Lee, S., Kim, D.-H., Goo, Y.H., Lee, Y.C., Lee, S.-K., Lee, J.W., 2009. Crucial Roles for Interactions between MLL3/4 and INI1 in Nuclear Receptor Transactivation. Mol Endocrinol 23, 610–619. https://doi.org/10.1210/me.2008-0455

Lee, S., Lee, J.W., Lee, S.-K., 2012. UTX, a histone H3-lysine 27 demethylase, acts as a critical switch to activate the cardiac developmental program. Dev. Cell 22, 25–37. https://doi.org/10.1016/j.devcel.2011.11.009

Li, M., Belmonte, J.C.I., 2017. Ground rules of the pluripotency gene regulatory network. Nature Reviews Genetics 18, 180–191. https://doi.org/10.1038/nrg.2016.156

Li, Y., Han, J., Zhang, Y., Cao, F., Liu, Z., Li, S., Wu, J., Hu, C., Wang, Y., Shuai, J., Chen, J., Cao, L., Li, D., Shi, P., Tian, C., Zhang, J., Dou, Y., Li, G., Chen, Y., Lei, M., 2016. Structural basis for activity regulation of MLL family methyltransferases. Nature 530, 447–452. https://doi.org/10.1038/nature16952

Linda, K., Fiuza, C., Nadif Kasri, N., 2018. The promise of induced pluripotent stem cells for neurodevelopmental disorders. Prog. Neuropsychopharmacol. Biol. Psychiatry 84, 382–391. https://doi.org/10.1016/j.pnpbp.2017.11.009

Long, H.K., Prescott, S.L., Wysocka, J., 2016. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell 167, 1170–1187. https://doi.org/10.1016/j.cell.2016.09.018

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., Richmond, T.J., 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389, 251–260. https://doi.org/10.1038/38444

Maekawa, M., Yamaguchi, K., Nakamura, T., Shibukawa, R., Kodanaka, I., Ichisaka, T., Kawamura, Y., Mochizuki, H., Goshima, N., Yamanaka, S., 2011. Direct reprogramming of somatic cells is promoted by maternal transcription factor Glis1. TL - 474. Nature 474 VN-, 225–229. https://doi.org/10.1038/nature10106

Makhlouf, M., Ouimette, J.-F., Oldfield, A., Navarro, P., Neuillet, D., Rougeulle, C., 2014. A prominent and conserved role for YY1 in *Xist* transcriptional activation. Nature Communications 5, 4878. https://doi.org/10.1038/ncomms5878

Medvedovic, J., Ebert, A., Tagoh, H., Tamir, I.M., Schwickert, T.A., Novatchkova, M., Sun, Q., Huis in 't Veld, P.J., Guo, C., Yoon, H.S., Denizot, Y., Holwerda, S.J.B., de Laat, W., Cogné, M., Shi, Y., Alt, F.W., Busslinger, M., 2013. Flexible Long-Range Loops in the VH Gene Region of the Igh Locus Facilitate the Generation of a Diverse Antibody Repertoire. Immunity 39, 229–244. https://doi.org/10.1016/j.immuni.2013.08.011

Meloni, M., Testa, G., 2014. Scrutinizing the epigenetics revolution. BioSocieties 9, 431–456. https://doi.org/10.1057/biosoc.2014.22

Menendez, L., Kulik, M.J., Page, A.T., Park, S.S., Lauderdale, J.D., Cunningham, M.L., Dalton, S., 2013. Directed differentiation of human pluripotent cells to

neural crest stem cells. TL - 8. Nature protocols 8 VN-re, 203–212. https://doi.org/10.1038/nprot.2012.156

Mertens, J., Paquola, A.C.M., Ku, M., Hatch, E., Böhnke, L., Ladjevardi, S., McGrath, S., Campbell, B., Lee, H., Herdy, J.R., Gonçalves, J.T., Toda, T., Kim, Y., Winkler, J., Yao, J., Hetzer, M.W., Gage, F.H., 2015. Directly Reprogrammed Human Neurons Retain Aging-Associated Transcriptomic Signatures and Reveal Age-Related Nucleocytoplasmic Defects. Cell Stem Cell 17, 705–718. https://doi.org/10.1016/j.stem.2015.09.001

Miller, S.A., Mohn, S.E., Weinmann, A.S., 2010. Jmjd3 and UTX Play a Demethylase-Independent Role in Chromatin Remodeling to Regulate T-Box Family Member-Dependent Gene Expression. Molecular Cell 40, 594–605. https://doi.org/10.1016/j.molcel.2010.10.028

Miyake, N., Mizuno, S., Okamoto, N., Ohashi, H., Shiina, M., Ogata, K., Tsurusaki, Y., Nakashima, M., Saitsu, H., Niikawa, N., Matsumoto, N., 2013. KDM6A Point Mutations Cause Kabuki Syndrome. Human Mutation 34, 108–110. https://doi.org/10.1002/humu.22229

Mo, R., Rao, S.M., Zhu, Y.-J., 2006. Identification of the MLL2 complex as a coactivator for estrogen receptor alpha. J. Biol. Chem. 281, 15714–15720. https://doi.org/10.1074/jbc.M513245200

Mohan, M., Herz, H.-M., Smith, E.R., Zhang, Y., Jackson, J., Washburn, M.P., Florens, L., Eissenberg, J.C., Shilatifard, A., 2011. The COMPASS family of H3K4 methylases in Drosophila. Molecular and cellular biology 31, 4310–8. https://doi.org/10.1128/MCB.06092-11

Montgomery, S.B., Sammeth, M., Gutierrez-Arcelus, M., Lach, R.P., Ingle, C., Nisbett, J., Guigo, R., Dermitzakis, E.T., 2010. Transcriptome genetics using second generation sequencing in a Caucasian population. Nature 464, 773–777. https://doi.org/10.1038/nature08903

Muerdter, F., Boryń, Ł.M., Woodfin, A.R., Neumayr, C., Rath, M., Zabidi, M.A., Pagani, M., Haberle, V., Kazmar, T., Catarino, R.R., Schernhuber, K., Arnold, C.D., Stark, A., 2018. Resolving systematic errors in widely used enhancer activity assays in human cells. Nature Methods 15, 141–149. https://doi.org/10.1038/nmeth.4534

Mumbach, M.R., Rubin, A.J., Flynn, R.A., Dai, C., Khavari, P.A., Greenleaf, W.J., Chang, H.Y., 2016. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nature Methods 13, 919–922. https://doi.org/10.1038/nmeth.3999

Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., Yamanaka, S., 2008. Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nat. Biotechnol. 26, 101–106. https://doi.org/10.1038/nbt1374

Ng, S.B., Bigham, A.W., Buckingham, K.J., Hannibal, M.C., McMillin, M.J., Gildersleeve, H.I., Beck, A.E., Tabor, H.K., Cooper, G.M., Mefford, H.C., Lee, C., Turner, E.H., Smith, J.D., Rieder, M.J., Yoshiura, K., Matsumoto, N., Ohta, T., Niikawa, N., Nickerson, D.A., Bamshad, M.J., Shendure, J., 2010. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome TL - 42. Nature Genetics 42 VN-r, 790–793. https://doi.org/10.1038/ng.646

Niihori, T., Aoki, Y., Narumi, Y., Neri, G., Cavé, H., Verloes, A., Okamoto, N., Hennekam, R.C.M., Gillessen-Kaesbach, G., Wieczorek, D., Kavamura, M.I., Kurosawa, K., Ohashi, H., Wilson, L., Heron, D., Bonneau, D., Corona, G., Kaname, T., Naritomi, K., Baumann, C., Matsumoto, N., Kato, K., Kure, S., Matsubara, Y., 2006. Germline *KRAS* and *BRAF* mutations in cardio-facio-cutaneous syndrome. Nature Genetics 38, 294–296. https://doi.org/10.1038/ng1749

Niikawa, N., Kuroki, Y., Kajii, T., Matsuura, N., Ishikiriyama, S., Tonoki, H., Ishikawa, N., Yamada, Y., Fujita, M., Umemoto, H., 1988. Kabuki make-up (Niikawa-Kuroki) syndrome: a study of 62 patients. TL - 31. American journal of medical genetics 31 VN-r, 565–589. https://doi.org/10.1002/ajmg.1320310312

Niikawa, N., Matsuura, N., Fukushima, Y., Ohsawa, T., Kajii, T., 1981. Kabuki make-up syndrome: a syndrome of mental retardation, unusual facies, large and protruding ears, and postnatal growth deficiency. TL - 99. The Journal of pediatrics 99 VN-r, 565–569. https://doi.org/10.1016/S0022-3476(81)80255-7

Noctor, S.C., Flint, A.C., Weissman, T.A., Dammerman, R.S., Kriegstein, A.R., 2001. Neurons derived from radial glial cells establish radial units in neocortex. Nature 409, 714–720. https://doi.org/10.1038/35055553

Noctor, S.C., Martínez-Cerdeño, V., Ivic, L., Kriegstein, A.R., 2004. Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases. Nat. Neurosci. 7, 136–144. https://doi.org/10.1038/nn1172

Northcott, P.A., Jones, D.T.W., Kool, M., Robinson, G.W., Gilbertson, R.J., Cho, Y.-J., Pomeroy, S.L., Korshunov, A., Lichter, P., Taylor, M.D., Pfister, S.M., 2012. Medulloblastomics: the end of the beginning. Nature Reviews Cancer 12, 818–834. https://doi.org/10.1038/nrc3410

Ohdo, S., Madokoro, H., Sonoda, T., Nishiguchi, T., Kawaguchi, K., Hayakawa, K., 1985. Kabuki make-up syndrome (Niikawa-Kuroki syndrome) associated with congenital heart disease. Journal of medical genetics 22, 126–7.

Okita, K., Ichisaka, T., Yamanaka, S., 2007. Generation of germline-competent induced pluripotent stem cells. Nature 448, 313–317. https://doi.org/10.1038/nature05934

Oliemuller, E., Kogata, N., Bland, P., Kriplani, D., Daley, F., Haider, S., Shah, V., Sawyer, E.J., Howard, B.A., 2017. SOX11 promotes invasive growth and ductal carcinoma in situ progression. J Pathol 243, 193–207. https://doi.org/10.1002/path.4939

On the use of the word 'epigenetic,' 2007. . Current Biology 17, R233–R236. https://doi.org/10.1016/j.cub.2007.02.030

Osumi, N., Shinohara, H., Numayama-Tsuruta, K., Maekawa, M., 2008. Concise review: Pax6 transcription factor contributes to both embryonic and adult neurogenesis as a multifunctional regulator. Stem Cells 26, 1663–1672. https://doi.org/10.1634/stemcells.2007-0884

Park, D.H., Hong, S.J., Salinas, R.D., Liu, S.J., Sun, S.W., Sgualdino, J., Testa, G., Matzuk, M.M., Iwamori, N., Lim, D.A., 2014. Activation of Neuronal Gene Expression by the JMJD3 Demethylase Is Required for Postnatal and Adult Brain Neurogenesis. Cell Rep 8, 1290–9. https://doi.org/10.1016/j.celrep.2014.07.060

Pasqualucci, L., Trifonov, V., Fabbri, G., Ma, J., Rossi, D., Chiarenza, A., Wells, V.A., Grunn, A., Messina, M., Elliot, O., Chan, J., Bhagat, G., Chadburn, A., Gaidano, G., Mulligan, C.G., Rabadan, R., Dalla-Favera, R., 2011. Analysis of the coding genome of diffuse large B-cell lymphoma TL - 43. Nat Genet 43 VN-r, 830–7. https://doi.org/10.1038/ng.892

Patel, S.R., Kim, D., Levitan, I., Dressler, G.R., 2007. The BRCT-Domain Containing Protein PTIP Links PAX2 to a Histone H3, Lysine 4 Methyltransferase Complex. Developmental Cell 13, 580–592. https://doi.org/10.1016/j.devcel.2007.09.004

Patro, R., Duggal, G., Kingsford, C., 2015. Accurate, fast, and model-aware transcript expression quantification with Salmon. bioRxiv.

Patten, D.K., Corleone, G., Győrffy, B., Perone, Y., Slaven, N., Barozzi, I., Erdős, E., Saiakhova, A., Goddard, K., Vingiani, A., Shousha, S., Pongor, L.S.,

Hadjiminas, D.J., Schiavon, G., Barry, P., Palmieri, C., Coombes, R.C., Scacheri, P., Pruneri, G., Magnani, L., 2018. Enhancer mapping uncovers phenotypic heterogeneity and evolution in patients with luminal breast cancer. Nature Medicine 1. https://doi.org/10.1038/s41591-018-0091-x

Plaza-Menacho, I., Burzynski, G.M., Groot, J.W. de, Eggen, B.J.L., Hofstra, R.M.W., 2006. Current concepts in RET-related genetics, signaling and therapeutics. Trends in Genetics 22, 627–636. https://doi.org/10.1016/j.tig.2006.09.005

QU, Y., ZHOU, C., ZHANG, J., CAI, Q., LI, J., DU, T., ZHU, Z., CUI, X., LIU, B., 2014. The metastasis suppressor SOX11 is an independent prognostic factor for improved survival in gastric cancer. Int J Oncol 44, 1512–1520. https://doi.org/10.3892/ijo.2014.2328

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., Wysocka, J., 2011. A unique chromatin signature uncovers early developmental enhancers in humans. TL - 470. Nature 470 VN-, 279–283. https://doi.org/10.1038/nature09692

Rahman, S., Zorca, C.E., Traboulsi, T., Noutahi, E., Krause, M.R., Mader, S., Zenklusen, D., 2017. Single-cell profiling reveals that eRNA accumulation at enhancer–promoter loops is not required to sustain transcription. Nucleic Acids Res 45, 3017–3030. https://doi.org/10.1093/nar/gkw1220

Rakic, P., 1972. Mode of cell migration to the superficial layers of fetal monkey neocortex. J. Comp. Neurol. 145, 61–83. https://doi.org/10.1002/cne.901450105

Ralston, A., Rossant, J., 2005. Genetic regulation of stem cell origins in the mouse embryo. Clin. Genet. 68, 106–112. https://doi.org/10.1111/j.1399-0004.2005.00478.x

Ramezani-Rad, P., Geng, H., Hurtz, C., Chan, L.N., Chen, Z., Jumaa, H., Melnick, A., Paietta, E., Carroll, W.L., Willman, C.L., Lefebvre, V., Müschen, M., 2013. SOX4 enables oncogenic survival signals in acute lymphoblastic leukemia. Blood 121, 148–155. https://doi.org/10.1182/blood-2012-05-428938

Rao, R.C., Dou, Y., 2015. Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. Nature reviews. Cancer 15, 334–46. https://doi.org/10.1038/nrc3929

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., Aiden, E.L., 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159, 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021

Reshef, Y.A., Finucane, H.K., Kelley, D.R., Gusev, A., Kotliar, D., Ulirsch, J.C., Hormozdiari, F., Nasser, J., O'Connor, L., Geijn, B. van de, Loh, P.-R., Grossman, S.R., Bhatia, G., Gazal, S., Palamara, P.F., Pinello, L., Patterson, N., Adams, R.P., Price, A.L., 2018. Detecting genome-wide directional effects of transcription factor binding on polygenic disease risk. Nature Genetics 1. https://doi.org/10.1038/s41588-018-0196-7

Rickels, R., Herz, H.-M., Sze, C.C., Cao, K., Morgan, M.A., Collings, C.K., Gause, M., Takahashi, Y., Wang, L., Rendleman, E.J., Marshall, S.A., Krueger, A., Bartom, E.T., Piunti, A., Smith, E.R., Abshiru, N.A., Kelleher, N.L., Dorsett, D., Shilatifard, A., 2017. Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. Nature Genetics. https://doi.org/10.1038/ng.3965

Robinow, M., Silverman, F.N., Smith, H.D., 1969. A newly recognized dwarfing syndrome. Am. J. Dis. Child. 117, 645–651.

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.

Bioinformatics (Oxford, England) 26, 139–40. https://doi.org/10.1093/bioinformatics/btp616

Rocha-Viegas, L., Villa, R., Gutierrez, A., Iriondo, O., Shiekhattar, R., Croce, L.D., 2014. Role of UTX in Retinoic Acid Receptor-Mediated Gene Regulation in Leukemia. Molecular and Cellular Biology 34, 3765–3775. https://doi.org/10.1128/MCB.00839-14

Rogozin, I.B., Gertz, E.M., Baranov, P.V., Poliakov, E., Schaffer, A.A., 2018. Genome-Wide Changes in Protein Translation Efficiency Are Associated with Autism. Genome Biol Evol 10, 1902–1919. https://doi.org/10.1093/gbe/evy146

Roma, D., Palma, P., Capolino, R., Figà-Talamanca, L., Diomedi-Camassei, F., Lepri, F.R., Digilio, M.C., Marras, C.E., Messina, R., Carai, A., Randi, F., Mastronuzzi, A., 2015. Spinal ependymoma in a patient with Kabuki syndrome: a case report. BMC medical genetics 16, 80. https://doi.org/10.1186/s12881-015-0228-4

Román, M., Baraibar, I., López, I., Nadal, E., Rolfo, C., Vicent, S., Gil-Bazo, I., 2018. KRAS oncogene in non-small cell lung cancer: clinical perspectives on the treatment of an old target. Molecular Cancer 17, 33. https://doi.org/10.1186/s12943-018-0789-x

Ropers, H.H., 2010. Genetics of Early Onset Cognitive Impairment. Annual Review of Genomics and Human Genetics 11, 161–187. https://doi.org/10.1146/annurev-genom-082509-141640

Ruthenburg, A.J., Allis, C.D., Wysocka, J., 2007. Methylation of Lysine 4 on Histone H3: Intricacy of Writing and Reading a Single Epigenetic Mark. Molecular Cell 25, 15–30. https://doi.org/10.1016/j.molcel.2006.12.014

Schubbert, S., Zenker, M., Rowe, S.L., Böll, S., Klein, C., Bollag, G., van der Burgt, I., Musante, L., Kalscheuer, V., Wehner, L.-E., Nguyen, H., West, B., Zhang, K.Y.J., Sistermans, E., Rauch, A., Niemeyer, C.M., Shannon, K., Kratz, C.P., 2006. Germline KRAS mutations cause Noonan syndrome. Nat. Genet. 38, 331–336. https://doi.org/10.1038/ng1748

Schultess, J., Danielewski, O., Smolenski, A.P., 2005. Rap1GAP2 is a new GTPase-activating protein of Rap1 expressed in human platelets. Blood 105, 3185–3192. https://doi.org/10.1182/blood-2004-09-3605

Schuurmans, C., Armant, O., Nieto, M., Stenman, J.M., Britz, O., Klenin, N., Brown, C., Langevin, L.-M., Seibt, J., Tang, H., Cunningham, J.M., Dyck, R., Walsh, C., Campbell, K., Polleux, F., Guillemot, F., 2004. Sequential phases of cortical specification involve Neurogenin-dependent and -independent pathways. EMBO J 23, 2892–2902. https://doi.org/10.1038/sj.emboj.7600278

Shepherd, J.H., Uray, I.P., Mazumdar, A., Tsimelzon, A., Savage, M., Hilsenbeck, S.G., Brown, P.H., 2016. The SOX11 transcription factor is a critical regulator of basal-like breast cancer growth, invasion, and basal-like gene expression. Oncotarget 7, 13106–13121. https://doi.org/10.18632/oncotarget.7437

Shi, Y., Seto, E., Chang, L.S., Shenk, T., 1991. Transcriptional repression by YY1, a human GLI-Krüppel-related protein, and relief of repression by adenovirus E1A protein. Cell 67, 377–88.

Shilatifard, A., 2012. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. Annual review of biochemistry 81, 65–95. https://doi.org/10.1146/annurev-biochem-051710-134100

Short, P.J., McRae, J.F., Gallone, G., Sifrim, A., Won, H., Geschwind, D.H., Wright, C.F., Firth, H.V., FitzPatrick, D.R., Barrett, J.C., Hurles, M.E., 2018. *De novo* mutations in regulatory elements in neurodevelopmental disorders. Nature. https://doi.org/10.1038/nature25983

Shpargel, K.B., Sengoku, T., Yokoyama, S., Magnuson, T., 2012. UTX and UTY Demonstrate Histone Demethylase-Independent Function in Mouse Embryonic Development. PLOS Genetics 8, e1002964. https://doi.org/10.1371/journal.pgen.1002964

Shpargel, K.B., Starmer, J., Wang, C., Ge, K., Magnuson, T., 2017. UTX-guided neural crest function underlies craniofacial features of Kabuki syndrome. PNAS 114, E9046–E9055. https://doi.org/10.1073/pnas.1705011114

Sigova, A.A., Abraham, B.J., Ji, X., Molinie, B., Hannett, N.M., Guo, Y.E., Jangi, M., Giallourakis, C.C., Sharp, P.A., Young, R.A., 2015. Transcription factor trapping by RNA in gene regulatory elements. Science 350, 978–981. https://doi.org/10.1126/science.aad3346

Smith, E., Shilatifard, A., 2014. Enhancer biology and enhanceropathies TL - 21. Nature Structural & Molecular Biology 21 VN-r, 210–9. https://doi.org/10.1038/nsmb.2784

Smith, R.L., Redd, M.J., Johnson, A.D., 1995. The tetratricopeptide repeats of Ssn6 interact with the homeo domain of alpha 2. Genes Dev. 9, 2903–2910. https://doi.org/10.1101/gad.9.23.2903

Sochacki, J., Devalle, S., Reis, M., Mattos, P., Rehen, S., 2016. Generation of urine iPS cell lines from patients with Attention Deficit Hyperactivity Disorder (ADHD) using a non-integrative method. Stem Cell Research. https://doi.org/10.1016/j.scr.2016.05.015

Stiles, J., Jernigan, T.L., 2010. The Basics of Brain Development. Neuropsychol Rev 20, 327–348. https://doi.org/10.1007/s11065-010-9148-4

Tabarés-Seisdedos, R., Rubenstein, J.L., 2013. Inverse cancer comorbidity: a serendipitous opportunity to gain insight into CNS disorders. Nature Reviews Neuroscience 14, 293–304. https://doi.org/10.1038/nrn3464

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., Yamanaka, S., 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131, 861–872. https://doi.org/10.1016/j.cell.2007.11.019

Takahashi, K., Yamanaka, S., 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663–676. https://doi.org/10.1016/j.cell.2006.07.024

Talbert, P.B., Henikoff, S., 2010. Histone variants — ancient wrap artists of the epigenome. Nature Reviews Molecular Cell Biology 11, 264–275. https://doi.org/10.1038/nrm2861

Testa, G., 2011. The time of timing: how Polycomb proteins regulate neurogenesis TL - 33. Bioessays 33 VN-r, 519–28. https://doi.org/10.1002/bies.201100021

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., Jones, J.M., 1998. Embryonic Stem Cell Lines Derived from Human Blastocysts. Science 282, 1145–1147. https://doi.org/10.1126/science.282.5391.1145

Tie, F., Banerjee, R., Conrad, P.A., Scacheri, P.C., Harte, P.J., 2012. Histone demethylase UTX and chromatin remodeler BRM bind directly to CBP and modulate acetylation of histone H3 lysine 27. Mol. Cell. Biol. 32, 2323–2334. https://doi.org/10.1128/MCB.06392-11

Tippens, N.D., Vihervaara, A., Lis, J.T., 2018. Enhancer transcription: what, where, when, and why? Genes Dev. 32, 1–3. https://doi.org/10.1101/gad.311605.118

Torre-Ubieta, L. de la, Won, H., Stein, J.L., Geschwind, D.H., 2016. Advancing the understanding of autism disease mechanisms through genetics. Nature Medicine 22, 345–361. https://doi.org/10.1038/nm.4071

Tsurusaki, Y., Koshimizu, E., Ohashi, H., Phadke, S., Kou, I., Shiina, M., Suzuki, T., Okamoto, N., Imamura, S., Yamashita, M., Watanabe, S., Yoshiura, K.,

Kodera, H., Miyatake, S., Nakashima, M., Saitsu, H., Ogata, K., Ikegawa, S., Miyake, N., Matsumoto, N., 2014. De novo SOX11 mutations cause Coffin-Siris syndrome. Nat Commun 5, 4011. https://doi.org/10.1038/ncomms5011

Tumino, M., Licciardello, M., Sorge, G., Cutrupi, M.C., Di Benedetto, F., Amoroso, L., Catania, R., Pennisi, M., D'Amico, S., Di Cataldo, A., 2010. Kabuki syndrome and cancer in two patients. American Journal of Medical Genetics Part A 152A, n/a-n/a. https://doi.org/10.1002/ajmg.a.33405

van Haaften, G., Dalgliesh, G.L., Davies, H., Chen, L., Bignell, G., Greenman, C., Edkins, S., Hardy, C., O'Meara, S., Teague, J., Butler, A., Hinton, J., Latimer, C., Andrews, J., Barthorpe, S., Beare, D., Buck, G., Campbell, P.J., Cole, J., Dunmore, R., Forbes, S., Jia, M., Jones, D., Kok, C.Y., Leroy, C., Lin, M.-L., McBride, D.J., Maddison, M., Maquire, S., McLay, K., Menzies, A., Mironenko, T., Lee, M., Mudie, L., Pleasance, E., Shepherd, R., Smith, R., Stebbings, L., Stephens, P., Tang, G., Tarpey, P.S., Turner, R., Turrell, K., Varian, J., West, S., Widaa, S., Wray, P., Collins, V.P., Ichimura, K., Law, S., Wong, J., Yuen, S.T., Leung, S.Y., Tonon, G., DePinho, R.A., Tai, Y.-T., Anderson, K.C., Kahnoski, R.J., Massie, A., Khoo, S.K., Teh, B.T., Stratton, M.R., Futreal, P.A., 2009. Somatic mutations of the histone H3K27 demethylase, UTX, in human cancer. Nat Genet 41, 521–523. https://doi.org/10.1038/ng.349

Van Laarhoven, P.M., Neitzel, L.R., Quintana, A.M., Geiger, E.A., Zackai, E.H., Clouthier, D.E., Artinger, K.B., Ming, J.E., Shaikh, T.H., 2015. Kabuki syndrome genes KMT2D and KDM6A: functional analyses demonstrate critical roles in craniofacial, heart and brain development. Human molecular genetics 24, 4443–53. https://doi.org/10.1093/hmg/ddv180

Vandamme, J., Lettier, G., Sidoli, S., Di Schiavi, E., Nørregaard Jensen, O., Salcini, A.E., 2012. The C. elegans H3K27 Demethylase UTX-1 Is Essential for Normal Development, Independent of Its Enzymatic Activity. PLoS Genetics 8, e1002647. https://doi.org/10.1371/journal.pgen.1002647

Vazin, T., Freed, W.J., 2010. Human embryonic stem cells: Derivation, culture, and differentiation: A review. Restor Neurol Neurosci 28, 589–603. https://doi.org/10.3233/RNN-2010-0543

Vega-Lopez, G.A., Cerrizuela, S., Tribulo, C., Aybar, M.J., 2018. Neurocristopathies: New insights 150 years after the neural crest discovery. Dev. Biol. https://doi.org/10.1016/j.ydbio.2018.05.013

Vella, P., Barozzi, I., Cuomo, A., Bonaldi, T., Pasini, D., 2012. Yin Yang 1 extends the Myc-related transcription factors network in embryonic stem cells. Nucleic Acids Res 40, 3403–3418. https://doi.org/10.1093/nar/gkr1290

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., Afzal, V., Ren, B., Rubin, E.M., Pennacchio, L.A., 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457, 854–858. https://doi.org/10.1038/nature07730

Vissers, Lisenka E L M, de Ligt, J., Gilissen, C., Janssen, I., Steehouwer, M., de Vries, P., van Lier, B., Arts, P., Wieskamp, N., del Rosario, M., van Bon, B.W.M., Hoischen, A., de Vries, B.B.A., Brunner, H.G., Veltman, J.A., 2010. A de novo paradigm for mental retardation. Nature genetics 42, 1109–12. https://doi.org/10.1038/ng.712

Vissers, Lisenka E. L. M., de Ligt, J., Gilissen, C., Janssen, I., Steehouwer, M., de Vries, P., van Lier, B., Arts, P., Wieskamp, N., del Rosario, M., van Bon, B.W.M., Hoischen, A., de Vries, B.B.A., Brunner, H.G., Veltman, J.A., 2010. A de novo paradigm for mental retardation. Nat. Genet. 42, 1109–1112. https://doi.org/10.1038/ng.712

Waddington, C.H., 1942. Canalization of Development and the Inheritance of Acquired Characters. Nature 150, 563–565. https://doi.org/10.1038/150563a0

Wai, D.C.C., Shihab, M., Low, J.K.K., Mackay, J.P., 2016. The zinc fingers of YY1 bind single-stranded RNA with low sequence specificity. Nucleic Acids Res 44, 9153–9165. https://doi.org/10.1093/nar/gkw590

Wang, C., Lee, J.-E., Cho, Y.-W., Xiao, Y., Jin, Q., Liu, C., Ge, K., 2012. UTX regulates mesoderm differentiation of embryonic stem cells independent of H3K27 demethylase activity. Proceedings of the National Academy of Sciences of the United States of America 109, 15324–9. https://doi.org/10.1073/pnas.1204166109

Wang, C., Lee, J.-E., Lai, B., Macfarlan, T.S., Xu, S., Zhuang, L., Liu, C., Peng, W., Ge, K., 2016. Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition. Proceedings of the National Academy of Sciences 113, 11871–11876. https://doi.org/10.1073/pnas.1606857113

Wang, L., Zhao, Z., Ozark, P.A., Fantini, D., Marshall, S.A., Rendleman, E.J., Cozzolino, K.A., Louis, N., He, X., Morgan, M.A., Takahashi, Y., Collings, C.K., Smith, E.R., Ntziachristos, P., Savas, J.N., Zou, L., Hashizume, R., Meeks, J.J., Shilatifard, A., 2018. Resetting the epigenetic balance of Polycomb and COMPASS function at enhancers for cancer therapy. Nature Medicine 1. https://doi.org/10.1038/s41591-018-0034-6

Wang, S.-P., Tang, Z., Chen, C.-W., Shimada, M., Koche, R.P., Wang, L.-H., Nakadai, T., Chramiec, A., Krivtsov, A.V., Armstrong, S.A., Roeder, R.G., 2017. A UTX-MLL4-p300 Transcriptional Regulatory Network Coordinately Shapes Active Enhancer Landscapes for Eliciting Transcription. Mol. Cell 67, 308-321.e6. https://doi.org/10.1016/j.molcel.2017.06.028

Watson, J.D., Crick, F.H., 1953. Genetical implications of the structure of deoxyribonucleic acid. Nature 171, 964–967.

Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., Guo, Y.E., Hnisz, D., Jaenisch, R., Bradner, J.E., Gray, N.S., Young, R.A., 2017. YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell 171, 1573-1588.e28. https://doi.org/10.1016/j.cell.2017.11.008

Wu, S., Shi, Yujiang, Mulligan, P., Gay, F., Landry, J., Liu, H., Lu, J., Qi, H.H., Wang, W., Nickoloff, J.A., Wu, C., Shi, Yang, 2007. A YY1-INO80 complex regulates genomic stability through homologous recombination-based repair. Nat. Struct. Mol. Biol. 14, 1165–1172. https://doi.org/10.1038/nsmb1332

Yoo, K.H., Oh, S., Kang, K., Wang, C., Robinson, G.W., Ge, K., Hennighausen, L., 2016. Histone Demethylase KDM6A Controls the Mammary Luminal Lineage through Enzyme-Independent Mechanisms. Molecular and cellular biology 36, 2108–20. https://doi.org/10.1128/MCB.00089-16

Yoshioka, N., Gros, E., Li, H.-R., Kumar, S., Deacon, D.C., Maron, C., Muotri, A.R., Chi, N.C., Fu, X.-D., Yu, B.D., Dowdy, S.F., 2013. Efficient generation of human iPSCs by a synthetic self-replicative RNA. Cell Stem Cell 13, 246–254. https://doi.org/10.1016/j.stem.2013.06.001

Young, M.A., Larson, D.E., Sun, C.-W., George, D.R., Ding, L., Miller, C.A., Lin, L., Pawlik, K.M., Chen, K., Fan, X., Schmidt, H., Kalicki-Veizer, J., Cook, L.L., Swift, G.W., Demeter, R.T., Wendl, M.C., Sands, M.S., Mardis, E.R., Wilson, R.K., Townes, T.M., Ley, T.J., 2012. Background mutations in parental cells account for most of the genetic heterogeneity of induced pluripotent stem cells. Cell Stem Cell 10, 570–582. https://doi.org/10.1016/j.stem.2012.03.002

Zhang, H., Alberich-Jordà, M., Amabile, G., Yang, H., Staber, P.B., DiRuscio, A., Welner, R., Ebralidze, A., Zhang, J., Levantini, E., Lefebvre, V., Valk, P.J.M.,

Delwel, R., Hoogenkamp, M., Nerlov, C., Cammenga, J., Saez, B., Scadden, D.T., Bonifer, C., Ye, M., Tenen, D.G., 2013. Sox4 is a key oncogenic target in C/EBPα mutant Acute Myeloid Leukemia. Cancer Cell 24, 575–588. https://doi.org/10.1016/j.ccr.2013.09.018

Zhang, S., Cui, W., 2014. Sox2, a key factor in the regulation of pluripotency and neural differentiation. World J Stem Cells 6, 305–311. https://doi.org/10.4252/wjsc.v6.i3.305

Zhang, S., Zhou, B., Wang, L., Li, P., Bennett, B.D., Snyder, R., Garantziotis, S., Fargo, D.C., Cox, A.D., Chen, L., Hu, G., 2017. INO80 is required for oncogenic transcription and tumor growth in non-small cell lung cancer. Oncogene 36, 1430–1439. https://doi.org/10.1038/onc.2016.311

Zhang, Y., Pak, C., Han, Y., Ahlenius, H., Zhang, Z., Chanda, S., Marro, S., Patzke, C., Acuna, C., Covy, J., Xu, W., Yang, N., Danko, T., Chen, L., Wernig, M., Südhof, T.C., 2013. Rapid single-step induction of functional neurons from human pluripotent stem cells. TL - 78. Neuron 78 VN-r, 785–798. https://doi.org/10.1016/j.neuron.2013.05.029

ACKNOWLEDGMENTS

I wish to acknowledge my supervisor Giuseppe Testa for having allowed me to work on this stimulating environment and for making me work to improve my weaknesses. Nael Nadif Kasri, who hosted me in his lab in Nijmengen for to perform MEA analysis and for the constructive advices. Bert De Vries for the fruitful collaboration. Monica Frega, Katrin Linda, Britt Mossink, Teun Klein Gunnewiek for the technical help with MEA and friendship in Nijmegen. All my collegues in Milan, in particular Alessandro Vitriolo, for the strict collaboration about the bioinformatic analysis, for the endless constructive scientific conversations, and for his friendship. Pierre-Luc Germain, as an awesome flat-mate, for his initial tutoring regarding the new world of transcription and epigenetic analysis. My older colleagues Giulia Barbagiovanni, Pietro Lo Riso, Elena Signaroldi, Giulia Fragola, Silvia Cristofanon, for their practical help, teachings, and positive and stimulating interaction. Sina Atashpaz for his tutoring activity in my first year. Giuseppe D'Agostino and Alejandro Lopez for the constructive conversations and teachings. Federico Boem, who introduced me to the philosophy of science and made me understand the importance of it, and his friendship. Matteo Zanella, for his teachings regarding neural crest stem cells. The first people I tutored Danila Pallotta, Celeste Franconi, Ludovico Rizzuti, and Marlene Pereira, for stimulating me to become a better tutor and teaching me how to teach. Other colleagues and new friends as Maria Elena Boggio Merlo, Mattia Andreoletti, Raffaele Luongo, for friendship and moral support.