Transcriptional consequences of mutations in genes associated with Autism Spectrum Disorder

by

Christoph Dotter

August, 2022

A thesis submitted to the
Graduate School
of the
Institute of Science and Technology Austria
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Committee in charge:

Mario de Bono, Chair
Gaia Novarino
Sandra Siegert
Kyung-Min Noh

The thesis of Christoph Dotter, titled *Transcriptional consequences of mutations in genes associated with Autism Spectrum Disorder*, is approved by:

Supervisor: Gaia Novarino, ISTA, Klosterneuburg, Austria

Signature: _____

Committee Member: Sandra Siegert, ISTA, Klosterneuburg, Austria

Signature: _____

Committee Member: Kyung-Min Noh, EMBL, Heidelberg, Germany

Signature: _____

Defense Chair: Mario de Bono, ISTA, Klosterneuburg, Austria

Signature: _____

Signed page is on file

ISTA Thesis, ISSN: 2663-337X

I hereby declare that this thesis is my own work and that it does not contain other people's work without this being so stated; this thesis does not contain my previous work without this being stated, and the bibliography contains all the literature that I used in writing the dissertation.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee, and that this thesis has not been submitted for a higher degree to any other university or institution.

I certify that any republication of materials presented in this thesis has been approved by the relevant publishers and co-authors.

Signature: _____

Christoph Dotter
August, 2022

Signed page is on file

Abstract

Autism spectrum disorders (ASDs) are a group of neurodevelopmental disorders characterized by behavioral symptoms such as problems in social communication and interaction, as well as repetitive, restricted behaviors and interests. These disorders show a high degree of heritability and hundreds of risk genes have been identified using high throughput sequencing technologies. This genetic heterogeneity has hampered efforts in understanding the pathogenesis of ASD but at the same time given rise to the concept of convergent mechanisms. Previous studies have identified that risk genes for ASD broadly converge onto specific functional categories with transcriptional regulation being one of the biggest groups. In this thesis, I focus on this subgroup of genes and investigate the gene regulatory consequences of some of them in the context of neurodevelopment.

First, we showed that mutations in the ASD and intellectual disability risk gene *Setd5* lead to perturbations of gene regulatory programs in early cell fate specification. In addition, adult animals display abnormal learning behavior which is mirrored at the transcriptional level by altered activity dependent regulation of postsynaptic gene expression. Lastly, we link the regulatory function of Setd5 to its interaction with the Paf1 and the NCoR complex.

Second, by modeling the heterozygous loss of the top ASD gene *CHD8* in human cerebral organoids we demonstrate profound changes in the developmental trajectories of both inhibitory and excitatory neurons using single cell RNA-sequencing. While the former were generated earlier in *CHD8*$^{+/-}$ organoids, the generation of the latter was shifted to later times in favor of a prolonged progenitor expansion phase and ultimately increased organoid size.

Finally, by modeling heterozygous mutations for four ASD associated chromatin modifiers, *ASH1L*, *KDM6B*, *KMT5B*, and *SETD5* in human cortical spheroids we show evidence of regulatory convergence across three of those genes. We observe a shift from dorsal cortical excitatory neuron fates towards partially ventralized cell types resembling cells from the lateral ganglionic eminence. As this project is still ongoing at the time of writing, future experiments will aim at elucidating the regulatory mechanisms underlying this shift with the aim of linking these three ASD risk genes through biological convergence.

Acknowledgements

The work presented in this thesis was carried out at the Life Science department of the Institute of Science and Technology Austria (ISTA) and financially supported by the Simons Foundation Autism Research Initiative (SFARI, grants 401299 and 707964 awarded to Gaia Novarino), the European Research Council (grant 715508 awarded to Gaia Novarino, grant 825759 to Giuseppe Testa), the Austrian Science Fund (FWF, grant 4205-B awarded to Gaia Novarino), the Deutsche Forschungsgemeinschaft (DFG, SPP1738 grant 1249 awarded to Kyung-Min Noh), the Fondazione Cariplo (grant 2017-0886 awarded to Alejandro López-Tóbon), and the E-Rare consortium (E-Rare-3 JTC 2018 awarded to M. Gabriele).

I want to thank my supervisor Gaia Novarino for giving me the opportunity to not just work on my own project but also on whatever else I could get my hands on. This allowed me to work on many different questions and really develop my skill set and passion for data analysis and visualization. I can't even begin to describe how much I grew both professionally and as a person over the last years. I also thank the other members of my thesis committee, Sandra Siegert and Kyung Min Noh for the valuable feedback over the years and the comments and discussion at my defense. This work would not have been possible without the many people who worked on these projects with me. For the Setd5 project I want to especially thank my co-first authors Elena Deliu, Niccolò Arecco, and Jasmin Morandell who made this one of the most enjoyable endeavors of my time here. I also thank all other co-authors: Ximena Contreras, Charles Girardot, Eva-Lotta Käsper, Alena Kozlova, Kasumi Kishi, and Ilaria Chiaradia for their valuable contributions. For the CHD8 project I want to highlight Roberto Sacco and Jasmin Morandell, who I had the pleasure of working with on this project during my first year at IST, and Bárbara Oliveira who was right there in the thick of it with me during the long process of getting it published. In addition, I thank all other co-authors and collaborators: Carlo "Ema" Villa, Cristina Cheroni, Alejandro López-Tóbon, Aysan Çerağ Yahya, Michele Gabriele, Mojtaba Tavakoli, Julia Lyudchik, Christoph Sommer, Mariano Gabitto, Johann Danzl, and Giuseppe Testa. Lastly I want to thank everyone involved in the convergence project so far: Bernadette Basilico, Jessica Leff, Nathalie Gruber, Denise Haslinger, Jessica Kirchner, Aysan Çerağ Yahya, and Roberto Sacco.

I want to thank all the past and present members of the Novarino group who made it a joy to work here and made it so much easier to handle the various setbacks that happened over the years. I treasure all the lengthy discussions about whatever topic popped into our heads, especially at lunch. So, Jasmin, Roberto, Eva, Dora, Gintare, Hagar, Zoë, Barbara, Bernadette, Elena, Luis, Francesca, Emily, Lena, Lisa, Ximena, Jessica and Jessica, Katalin, Albert, Romina, Matej, Sarah, and Alena, thank you for all the good times, the emotional and technical support, the many scientific discussions and just for

overall creating a great work environment. I also want to especially thank Anna for all the great support over the years. Special thanks also to Gintare for proofreading parts of this thesis.

Next, I want to thank my good friend Denise for the great music sharing sessions we spent in the cell culture, for never tiring of all the tings I suggest, your truly never-ending emotional support and positive attitude, especially over these last taxing months, and just overall for being a great person. Thanks for keeping me sane while writing this thing.

I want to thank my parents for their never wavering support, for having my back no matter what I decide to do in life and arguably being my biggest fans. I also thank my brother and sister in law and everyone else in my family who kept rooting for me.

Last but by no means least, I want to thank my partner Eva for your emotional support through it all, the patience to weather all the late nights and weekends I had to spend working, for handling much of our settling into our new flat while I was stuck writing. None of this would have been possible without you. Thank you.

About the Author

Christoph Dotter first received a BSc in Biomedical Informatics at the UMIT in Hall in Tirol. Afterwards he changed fields to study Chemistry at the University of Innsbruck (BSc) followed by an MSc in Biological Chemistry at the University of Vienna. Next he worked as a bioinformatician in the groups of Denise Barlow at CeMM, Vienna and the group of Simon Hippenmeyer at IST Austria. He joined IST Austria in 2015 as a PhD student in the group of Gaia Novarino. Over the course of his PhD he worked on a total of five different projects related to autism spectrum disorders. Besides the three main studies covered in this thesis this included work on a study on the large neutral amino acid transporter Slc7a5 published in 2016 in Cell and minor work on a study on the role of the E3-ubiquitin ligase Cul3 published in 2021 in Nature Communications. His main research focus during this time were the transcriptional perturbations caused by mutations in ASD risk genes.

List of Collaborators and Publications

Elena Deliu*, Niccolò Arecco*, Jasmin Morandell*, **Christoph P. Dotter***, Ximena Contreras, Charles Girardot, Eva-Lotta Käsper, Alena Kozlova, Kasumi Kishi, Ilaria Chiaradia, Kyung-Min Noh, and Gaia Novarino. "Haploinsufficiency of the intellectual disability gene SETD5 disturbs developmental gene expression and cognition". In: *Nature Neuroscience* 21.12 (Nov. 2018), pp. 1717–1727

* Shared first-authorship

Carlo Emanuele Villa, Cristina Cheroni*, **Christoph P. Dotter***, Alejandro López-Tóbon, Bárbara Oliveira, Roberto Sacco, Aysan Çerağ Yahya, Jasmin Morandell, Michele Gabriele, Mojtaba R. Tavakoli, Julia Lyudchik, Christoph Sommer, Mariano Gabitto, Johann G. Danzl, Giuseppe Testa, and Gaia Novarino. "CHD8 haploinsufficiency links autism to transient alterations in excitatory and inhibitory trajectories". In: *Cell Reports* 39.1 (Apr. 2022), p. 110615

* Shared second-authorship

Christoph P. Dotter, Bernadette Basilico, Jessica Leff, Nathalie Gruber, Denise Haslinger, Jessica Kirchner, Aysan Çerağ Yahya, Roberto Sacco, and Gaia Novarino. "Convergence of Causal Paths across Epigenetic Regulators in ASD" (ongoing)

Table of Contents

List of Figures

List of Supplementary Figures

List of Tables

List of Abbreviations

ASD autism spectrum disorder.

CA Cornu Ammonis.

CFC contextual fear conditioning.

ChIP chromatin immunoprecipitation.

CREB cAMP response element-binding protein.

CRISPR clustered regularly interspaced short palindromic repeats.

DEG differentially expressed gene.

DSM Diagnostic and Statistical Manual of Mental Disorders.

EB embryoid body.

EN excitatory neuron.

FACS fluorescence activated cell sorting.

FC fold change.

fEPSP field excitatory postsynaptic potential.

GFP Green Fluorescent Protein.

gRNA guide RNA.

H3K27ac histone 3 lysine 27 acetylation.

H4ac histone 4 acetylation.

HDAC3 Histone deacetylase 3.

hESC human embryonic stem cell.

ICD International Classification of Diseases.

ID intellectual disability.

IN inhibitory neuron.

IP intermediate progenitor.

iPSC induced pluripotent stem cell.

LTP long term potentiation.

NCoR Nuclear receptor co-repressor.

NDD neurodevelopmental delay.

NPC neural progenitor cell.

NSC neural stem cell.

PBS phosphate buffered saline.

PDD Pervasive Developmental Disorders.

PDD-NOS PDD not otherwise specified.

Pol II RNA Polymerase II.

PTV protein-truncating variant.

RG radial glia.

RUV removal of unwanted variation.

SFARI Simon's Foundation Autism Research Initiative.

TF transcription factor.

TSS transcription start site.

UMAP uniform manifold approximation and projection.

WGCNA weighted gene coexpression network analysis.

WHO World Health Organization.

1

Introduction

1.1 Autism spectrum disorder

ASD, the subject of this thesis, is described as an umbrella term for a group of neurodevelopmental disorders. Characteristic symptoms include problems in social communication and interaction on the one hand and repetitive, restricted behaviors and interests on the other [3]. The prevalence of autism in the US is estimated at 1 in 44 according to CDC data from 2018 [4] a number which has been steadily growing over the years (also reviewed in a recent book by Spectrum [5]). Whether this is due to an increase in actual case numbers or a consequence of raised awareness and better diagnosis - the number still stands as testimony for the importance of studying this condition.

In this section I'd first like to take a step back and start with a perspective onto the beginnings of the study of autism as a disorder. The struggle in first defining autism as a separate diagnosis, followed by repeated adjustments to the diagnostic criteria, subcategorization and overall definition illuminates one of the big questions in autism research today - what is autism exactly and how should it be categorized? I will introduce this question first from a phenotypic standpoint before diving into the biological underpinnings with a strong focus on the genetic basis. I am of the opinion that providing a - still lacking - strong biological foundation will aid in clearly defining and categorizing ASD.

1.1.1 Diagnosis of Autism - a historical perspective

The history of autism as a disorder dates back to the work of Hans Asperger in 1938, on what later came to be known as Asperger syndrome [6, 7], as well as Leo Kanner's work in 1943 on what he termed "infantile autism" [8]. Asperger emphasized the general isolation of autistic patients from the outside world, which formed the link to the term autism (from Greek: autos for self). He described autistic patients as having difficulty with social integration and interaction, good verbal skills as well as a limited number of highly specific interests. The latter can take the form of exceptional skills in these areas of interest which is where the idea of the autistic savant originated from.

Kanner's account classified autism as a disorder of affective contact with two kinds of behavioral abnormalities at its core. First, autistic patients present with severe difficulties in the area of social communication and interaction with others, a behavior that Kanner

refers to as a "desire for aloneness", reflective of the previously mentioned concept of isolation of Asperger. The second main characteristic of autism in this account referred to the response of patients to change. Autism is characterized by a "need for sameness" which manifests itself in multiple ways. On the one hand, patients showed a strong adverse reaction to changes in their environment. On the other hand, patients also displayed stereotyped and repetitive behaviors, which Kanner interpreted as a way for these children to themselves create the sameness they require. This facet can also be found in Asperger's mention of restricted interests.

These historical first accounts painted a picture of the autistic patient as someone disconnected from the world around them, supplied with only a limited toolkit to establish this connection. However, these studies diverged on other aspects of the disorder. For example, Kanner put emphasis on autism as an infantile disorder and developmental condition while Asperger described autism more in the context of a personality disorder also present in adults. Kanner treated autism as a narrowly defined rare disorder, while Asperger's view is more in line with autistic traits being much more common but existing across a wide spectrum of severity. This is also exemplified by what both authors remarked in regard to the parents of the evaluated children. Asperger highlighted that some traits were shared between parents and their offspring, also putting forth the hypothesis of a heritable component, and how this could lead to better mutual understanding. Kanner, on the other hand, saw issue in the lack of emotional contact between parents and children - in his words, he observed "very few really warmhearted fathers and mothers" - to the point of questioning whether the parents' behavior influenced the child's condition or not [9, 10].

While it is debated whether Kanner was aware of Asperger's work at the time, given that it was only available in German, Kanner undoubtedly had the bigger impact on how autism came to be defined since then [9]. Many of the features described in his seminal work are still relevant in the diagnosis and study of autism today. However, early versions of the Diagnostic and Statistical Manual of Mental Disorders (DSM) did not list autism as a separated diagnosis. Rather, cases as the ones described above would be diagnosed as childhood schizophrenia [10]. Further research concerning the nature of autism aided in delineating autism from schizophrenia and other psychiatric illnesses as well as providing evidence for a biological basis of autism (for example reviewed in [11]). Especially findings from twin studies showing a strong genetic component in autism [12, 13] and connection to other genetic conditions such as Fragile X syndrome [14, 15]. In 1978 Rutter suggested a redefinition of autism to aid in this distinction [16]. He listed four criteria that should be employed in autism diagnosis. First, the age of onset - a critical difference to schizophrenia - had to be before 30 months of age. Second, patients had to show specific signs of social impairment and this had to be independent of intellectual level. This accounted for the observation that autism symptoms presented in patients across the whole intelligence spectrum. Third, language development had to deviate in certain ways, again independent of intelligence. The fourth criterion was the aforementioned insistence on sameness, which presented through stereotyped behavior and overall resistance to change [16].

Eventually these studies culminated in the inclusion of autism as a separate diagnosis in version 3 of the DSM with diagnostic criteria very similar to the ones listed above [17]. It was included as Pervasive Developmental Disorders (PDD), a group which, for the first time, also included a category for sub-threshold cases (atypical PDD). The revision of the DSM-III (DSM-III-R) marked a conceptual shift in the definition of autism as "infantile

autism" became "autistic disorder", thereby expanding the scope beyond childhood. The very rigid DSM-III criteria were markedly expanded to a set of sixteen, eight of which had to be present for the diagnosis. The diagnostic criteria were grouped into what became the main domains of autism symptoms: 1) qualitatively impaired reciprocal social interaction, 2) communication deficits and 3) repetitive behavior, restricted interests and resistance to change [10, 18]. In line with this more inclusive approach the sub-threshold diagnosis was renamed into "PDD not otherwise specified (PDD-NOS)".

The release of the 10th edition of the International Classification of Diseases (ICD) by the World Health Organization (WHO) saw a divergence in terms of autism diagnosis between the DSM used in the US and the ICD used in most other countries in the world [10, 19]. While the DSM-III-R went into the direction of a broad and all-encompassing diagnosis, the ICD-10 included several separate disorders, such as Asperger syndrome, Rett syndrome and childhood disintegrative disorder. To ameliorate this divergence between ICD and DSM, the fourth edition of the DSM presented a streamlined version of the ICD-10 criteria. It also listed the aforementioned disorders as separate from autistic disorder and stuck to the three main categories of symptoms [20].

This sub-categorization of autism into closely related disorders presented new challenges. Diagnostic practice needed to be consistent in delineating the different subtypes as this had a strong effect on availability of treatment [10]. The effort to define those differences (e.g. between autism and Asperger's syndrome [21] or finding a more accurate definition of PDD-NOS [22]) was complicated by the high variability in symptom severity observed both across but also within subtypes. Multiple studies compared clinical parameters between matched patients with different sub-diagnoses and found very little difference [23, 24]. Other studies found that the association between clinical phenotypes and diagnosis was not reproducible across multiple sites [25] and that there was little difference in prognosis between autism disorder and PDD-NOS [26]. In general, it was seen that the discretization of the diagnosis into subcategories required "necessarily arbitrary" [27] thresholds on continuous behavioral traits. A broadening of the diagnostic criteria would allow for more consistent diagnosis. Taken together these arguments led to a move from different categories to one all-encompassing multi-faceted category - Autism spectrum disorder - in the DSM-5 (Fig. 1.1).

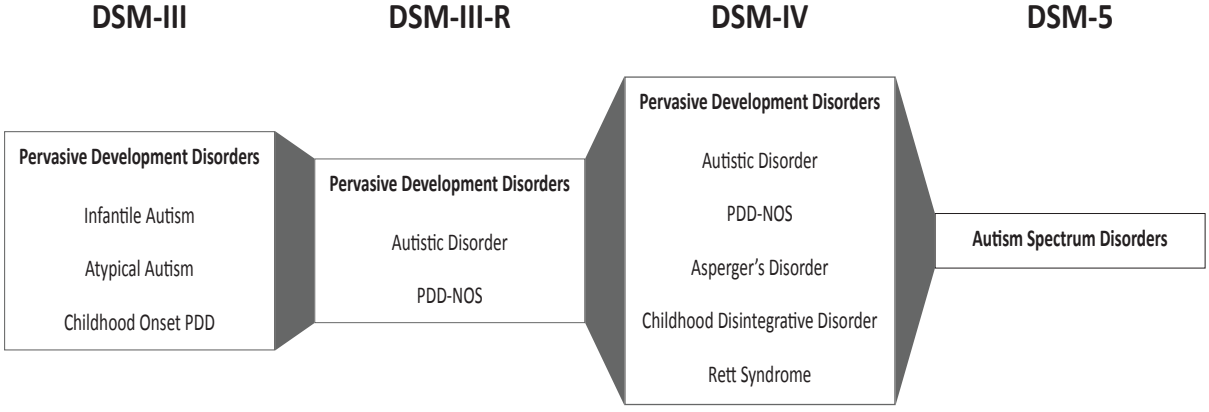


Figure 1.1: **Autism diagnoses in the Diagnostic and Statistical Manual of Mental Disorders (DSM)** since its inclusion as a separate diagnosis in DSM-III.

The term ASD carries the notion that autism is a condition of many faces and serves as an umbrella term for the previously separately diagnosed subtypes such as Asperger's

syndrome. The three diagnostic categories of DSM-IV were reduced to two in the DSM-5 by combining social interaction and communication deficits into one. [3]. In line with the move from a categorical to a more dimensional approach the criteria were broadened. This allowed for more heterogeneous individual manifestations of a specific diagnostic criterion such as deficits in social-emotional reciprocity. ASD in the DSM-5 is presented as a purely behavioral diagnosis which can be detailed further using neurobiological specifiers, such as genetic conditions, or modifiers, i.e. additional behavioral diagnoses, such as intellectual disability or ADHD. For example, in the case of ASD symptoms in Rett syndrome, which is a genetic condition caused by mutations in the gene *MECP2* [28] the diagnosis is now "ASD and Rett Syndrome" [29]. Lastly, the DSM-5 includes specifiers for severity of symptoms. These are based on the level of support per symptom category and, despite shortcomings in consistency [30], are a step in the right direction towards stratification of individuals diagnosed with ASD by need of treatment.

In conclusion, the history of autism diagnosis and definition is a story fraught with changes and adaptations around the very core of what autism is and what it should encompass. This back and forth between more or fewer categories is in large part owed to the fact that autism, with the exception of some syndromic forms with known etiologies, has always been a purely behavioral diagnosis. This is also reflected in the separation of known genetic etiologies from the ASD category in the DSM-5. The challenge that remains is to find a way to stratify individuals diagnosed with ASD based on their needs and design suitable interventions to meet those needs.

1.1.2 The genetic basis of ASD

As illuminated above, autism or ASD has been recognized as a heritable condition with a biological basis for a long time. Nonetheless, we have yet to catch a glimpse of the pathophysiology of ASD. Bearing in mind the purely behavioral nature of ASD core symptoms, studies of its biological foundation have been hampered by the lack of a clearly defined biological substrate to use as an entry point. By taking advantage of ASD's heritable nature these entry points can be found in the study of genetic risk factors, such as ASD candidate genes.

Modern estimates of ASD heritability range from 50 to 90% [31–34] with some evidence that co-occurrence of intellectual disability reduces this estimate to 33% [35]. Another recent study also reported the heritability of quantitative autistic traits rather than the previously investigated qualitative trait of ASD diagnosis [36] with estimates ranging from 24 to 77%.

Even though common variation is estimated to contribute the most to this estimate by far, the focal points of study have been rare inherited and *de novo* mutations which together account for only around 6% of variability in ASD liability [31]. This is due to the fact that these rare variants likely have much larger effect sizes than common variations where single variants have only small effects [37]. Initial array-based studies had great success in associating copy number variations of larger chromosomal regions, often encompassing multiple genes, with ASD [38–44]. Over the last decade, whole exome and whole genome sequencing methods enabled researchers to investigate the presence of genetic variation genome-wide and at single nucleotide resolution in large cohorts of individuals diagnosed with non-syndromic ASD and neurotypical controls [45–56]. These studies reported association of genes with ASD by integrating the gathered information

on protein-truncating and missense variants including *de novo* and inherited mutations from family-based samples as well as mutations identified in case-control populations. The study by Satterstrom et al. [55] represents the largest such study to date encompassing around 35500 samples, 12000 of which are from ASD cases. Their analysis identified a total of 9345 rare *de novo* variants in protein coding exons. The authors further stratified these based on inheritance pattern (*de novo*, inherited, or case-control) as well as variant type (protein truncating (PTV), missense, or synonymous). PTVs and missense variants were further divided based on functional severity. They found significantly higher proportions of PTVs and severe missense mutations in *de novo* variants compared to inherited variants. Overall *de novo* variation was estimated to account for 1.9% of ASD variance but represented the main source for the identification of ASD risk genes. Using an updated version of the Transmission And De novo Association (TADA) method [57] to integrate variant severity scores the authors identified 102 genes associated with ASD, 30 of which they deemed truly novel [55, 58]. Lastly, the study divided the ASD risk genes into ASD-predominant and "ASD and neurodevelopmental delay (NDD)" based on the relative frequency of disruptive *de novo* mutations in their data compared to data from studies on severe NDD.

The results of these studies, together with patient reports, formed the basis for the curation of a list of ASD risk genes by the Simon's Foundation Autism Research Initiative[1] [59]. As of release 2021 Q4 the list comprises 427 genes with at least strong evidence, 207 of which received a score of 1, representing highest confidence in the association. 111 of those 427 genes are listed as "syndromic" genes as they were also linked to other, non-ASD symptoms. One example here is the aforementioned *MECP2* linked to Rett syndrome which has a score of 1S. In addition, the site lists 515 genes with a score of 3 indicating "suggestive evidence" and hinting at the possibility of yet even more candidate genes to consider. Figure 1.2 illustrates the heterogeneity of ASD by displaying all genes with scores 1 and 2, and overlaying the information about predominant ASD genes and ASD & NDD genes from the Satterstrom study.

Given that by definition the perturbation of these factors can lead to ASD it is fair to assume that studying these perturbations will set researchers on the trail to the elusive etiology of ASD. However, the vast heterogeneity identified in even this small sliver of the genetic landscape of ASD presents a difficult challenge. An important question that needs to be answered is whether all these different genetic origins directly translate into phenotypic differences or whether it is possible to stratify ASD by genetic etiology. This issue is reminiscent of the sub-categorization problem encountered at the diagnostic level. But here, in the realm of genetic causes, it is possible to address this issue and possibly provide stratification of ASD cases based on better quantifiable measures.

1.2 Convergence in Autism spectrum disorder pathogenesis

Each mutation in one of the candidate genes associated with ASD through one mechanism or another increases the likelihood for developing ASD or - in more categorical terms - leads to the outcome of ASD. In this setting, addressing the issue of genetic heterogeneity as a sub-classification problem becomes synonymous with the task of finding convergence

[1] https://www.sfari.org/resource/sfari-gene/

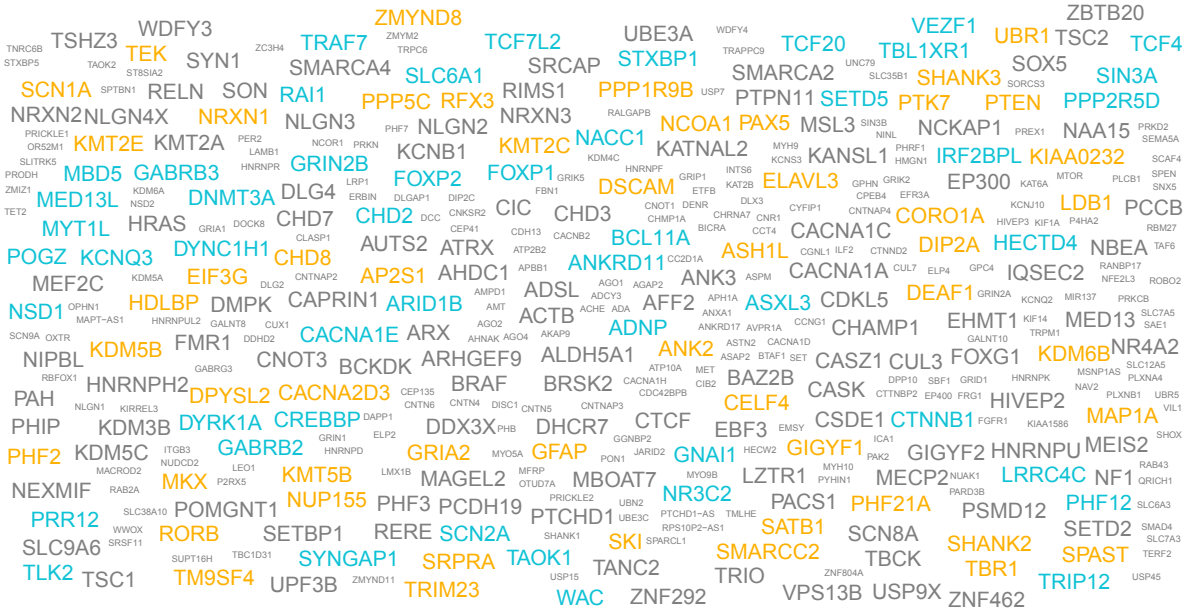


Figure 1.2: **The genetic heterogeneity of ASD.** Shown are ASD risk genes with a SFARI score of 1 (larger) or 2 (smaller). Genes in blue (ASD + NDD) and orange (ASD) belong to the 102 genes from the Satterstrom study.

among these different unknown mechanisms to group genetic causes by their pathogenesis. The large number of ASD risk genes identified by the studies outlined above highlights the vast scale of this task. It is therefore paramount to find sub-groups of these genes based on common features to divide and conquer ASD pathogenesis.

1.2.1 Convergence at the level of candidate gene function

To gain a first glimpse at whether there could be convergence among ASD risk genes previous studies investigated whether already existing knowledge about the candidate genes, such as functional annotation or gene co-expression over time and brain regions could be leveraged to find common themes and perform a crude first sub-division of these genes.

The first point of convergence speculated about was at the level of synaptic function [60]. However, the plethora of genes identified since then have complicated the image of ASD as a synaptopathy [61]. Most of the aforementioned studies also reported trends observed regarding annotated functions of the genes they identified. For example, the study by de Rubeis et al. [50] showed enrichment for synaptic genes and genes involved in transcriptional regulation by expanding their list of genes using protein-protein-interaction data and then functionally annotating sub-clusters of the identified interaction network. Enrichment for chromatin regulators was also found in the set of genes identified by Pinto et al. [62] and Sanders et al. [52] and expansion of the gene list via protein-protein interactions again revealed synaptic genes and chromatin regulation as the two main subgroups, a result also replicated in an analysis by Chang et al. [63]. Lastly, these two categories of genes were also confirmed in the Satterstrom study [55] and form two of the core themes among ASD risk genes. Further evidence was also provided by a study looking at enrichment of ASD genes in modules derived from the whole human protein interactome [64].

With functional themes emerging another question that arose was where and when this functional convergence would manifest itself over the course of brain development. For this reason studies used human gene expression data to derive gene co-expression networks and used various methodologies to embed ASD candidate genes into these networks. The first of these studies clustered genes by weighted gene co-expression network analysis (WGCNA, [65]) and identified one cluster with high expression during development and enriched for transcriptional regulators [66]. In 2013 one study expanded upon this by first generating a general co-expression network from an expression dataset of human cortical development and defining modules of co-expressed genes showing similar developmental trajectories [67]. They then looked at enrichment of ASD risk genes within these modules and thereby linked a total of five modules to ASD. Two of these were enriched for *de novo* mutations, showed higher prenatal expression and were enriched for transcriptional regulators. The other three displayed increasing expression over time and enrichment for genes connected to synaptic transmission. In order to add a spatial dimension to their analysis they also checked for cortical layer specific expression at both fetal and adult stages. In fetal development their analysis highlighted laminae with post-mitotic neurons, with the exception of one module with enrichment in the ventricular and sub-ventricular zone. In the adult cortex they observed enrichment for more superficial layers (Layer 2 - 4). Lastly, enrichment for cell-type specific markers connected upper-layer glutamatergic neurons with both prenatal modules.

Another study from the same year employed the same brain expression dataset but used a different approach to network construction. Rather than generating genome-wide networks, they used a set of nine high confidence ASD risk genes as a seed for the generation of time window specific co-expression networks [68]. They then identified ASD-related networks via enrichment for probable ASD genes and highlighted mid-fetal human cortical development (10-24 post-conceptional weeks). Incorporating layer specific expression data connected the networks to the inner cortical plate and enrichment for cell-type specific markers implicated cortical glutamatergic neurons of deep layers 5 and 6 in ASD. Due to the increase of ASD risk genes since then a more recent study by the same group updated the analysis of layer specific convergence. They confirmed the inner cortical plate, but in addition could highlight the inner subventricular zone and the subplate as a point of spatial convergence thereby also implicating early neurogenesis [69]. Taken together, these studies revealed mid-fetal development as a point of temporal convergence and cortical glutamatergic neurons from both deep and superficial layers as well as neurogenic progenitor cells in the subventricular zone as a point of spatial convergence.

A different branch of analyses done at the level of ASD genes was to check whether gene regulatory proteins encoded by ASD genes themselves regulated other genes connected to ASD. For this purpose studies looked at enrichment of gene targets of, for example, the Fragile X mental retardation protein (FMRP). This is an RNA-binding protein encoded by *FMR1* and mutations in this gene are the cause of Fragile X syndrome, a form of intellectual disability which often also presents with ASD symptoms [70]. FMRP targets identified by cross-linking immunoprecipitation (CLIP) in both mouse brain and human embryonic kidney cells (HEK293) showed strong enrichment for ASD targets [71, 72] highlighting FMRP regulation as a potential point of convergence. Another RNA binding protein associated with ASD is the alternative splicing regulator RNA Binding Fox-1 Homolog 1, encoded by the *RBFOX1* gene. Targets of this gene identified by CLIP [73] have also been shown to be enriched in ASD risk genes [50] indicating alternative splicing

as an additional layer of gene regulation that might be affected in ASD.

At the level of transcriptional regulation the chromatin remodeling factor chromodomain helicase DNA binding protein 8, encoded by the *CHD8* gene, has garnered a lot of attention in recent years. It is one of the genes most frequently found to be mutated in individuals diagnosed with ASD to date [55] and mutations were found to be highly penetrant [74, 75]. Two studies identified CHD8 targets by chromatin immunoprecipitation (ChIP) in human neural progenitor cells (hNPCs) [76] as well as in human neural stem cells (hNSCs) and human midfetal brain [77]. They showed that genes with CHD8 binding at their promoter were enriched for ASD genes indicating that regulation by CHD8 might present another point of convergence. Furthermore, differential gene expression upon knockdown of CHD8 in hNPCs showed enrichment for genes in co-expression modules from Parikshak et al. Down-regulated genes were found to be enriched for synaptic gene modules while up-regulated genes in turn were associated with developmental modules. Lastly, down-regulated genes, specifically those not bound by CHD8, were enriched for ASD genes providing an additional facet of CHD8-mediated ASD gene regulation [76].

In summary, gene level analyses revealed robust connection between ASD and modules of transcriptional regulation and synaptic transmission, both at the level of gene co-expression and protein-protein-interaction networks. Furthermore, they showed the importance of the cerebral cortex in ASD and highlighted its mid-fetal development as a potential period for convergent mechanisms. At the cell type level the strongest connections were found with cortical glutamatergic neurons.

1.2.2 Convergence at the level of causal paths

In order to address the challenge of uncovering convergence in ASD pathogenesis Sanders proposed its conceptualization in the form of intersection of causal paths [78]. In this model a causal path is defined as the smallest set of biological processes whose perturbation is both sufficient and necessary to explain how the root cause, in terms of genetic risk the mutation at the DNA level, eventually leads to the observed phenotypic endpoint [78, 79]. Such a path would traverse across multiple levels of abstraction from the molecular level of chromatin regulation, transcription and translation to the pathway level with protein functions and further on towards higher order levels of cell-types and tissues. Given that mutations rarely have singular effects and usually also lead to a number of confounding outcomes many such paths can be drawn but - in a simplified view - only one of them will lead to the phenotype in question, in this case ASD. Each step therefore holds the potential for spawning a multitude of new paths and complicating the picture ever further. Additional complexity comes from the fact that these levels are not necessarily traversed in a linear fashion, one example being regulatory feedback mechanisms.

Convergence in this context would mean a crossing of paths at a level other than the last. This is illustrated in Figure 1.3. For example, let's consider the two major groups outlined above - synaptic proteins and regulators of gene expression. For the former - which have a more straight-forward association with a specific function - the paths for genes might intersect at the level of protein function, cellular processes or pathways. The primary function of the latter group on the other hand is the regulation of other genes with a potential plethora of different functions. This promiscuity makes it more difficult to connect risk genes with specific cellular functions relevant for ASD pathogenesis. Illustrating this in terms of causal paths also reveals the source of this difference in

functional pleiotropy between the two groups - the function of transcriptional regulators is realized earlier along the path of biological information and this allows confounding paths to spawn many more downstream effects.

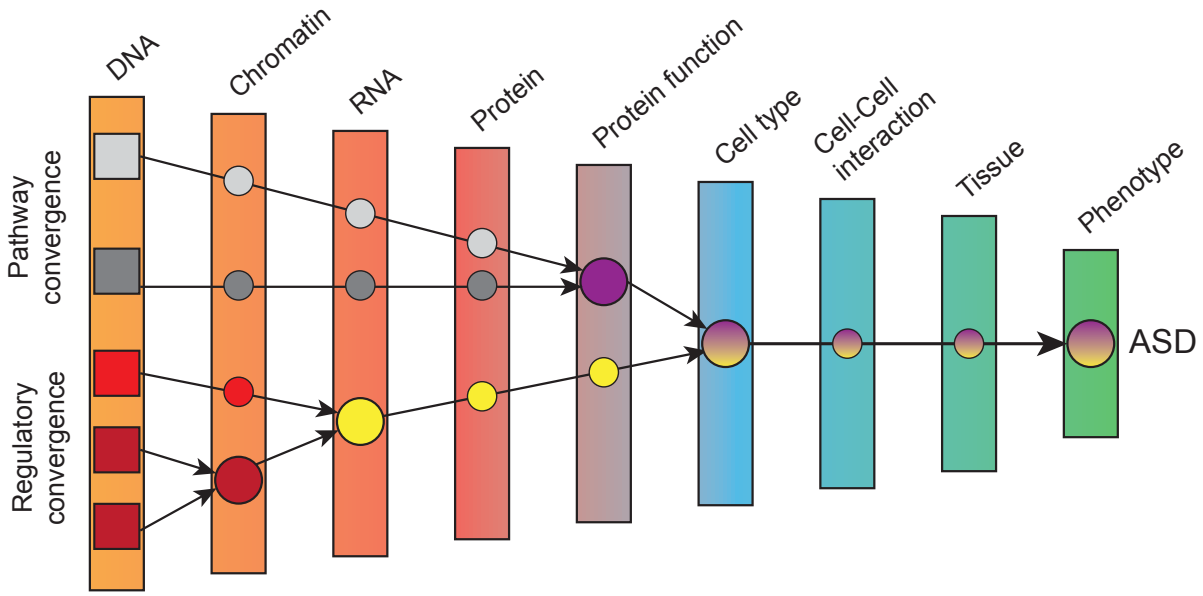


Figure 1.3: **Possible convergence of causal paths in ASD.** Depicted are examples of information flow progression from DNA mutation in risk genes on the left to the phenotypic level on the right. Mutations in ASD risk genes might show direct functional convergence if the risk genes themselves (gray squares) are involved in the same cellular process (big purple circle), for example synaptic function. Regulatory convergence, on the other hand, happens earlier and might arise from convergence on altered transcription of effector genes (big yellow circle). In addition, mutations (dark red squares) might lead to overlapping changes in the chromatin landscape (dark red circle) and subsequent changes in transcription. Ultimately both convergence paths might cross at higher levels, for example by preferentially affecting the same cell types (purple/yellow circle). This figure was adapted from Sanders, 2015 [78].

Thinking of cases of regulatory convergence as an intersection of paths is particularly beneficial. Studying consequences of mutations in transcriptional regulators usually involves the generation of large lists of affected genes, for example by differential gene expression analysis, with no further indication as to which of these changes are of interest for the phenotype under investigation. Intersecting those results from multiple genes with each other - the equivalent to drawing intersecting paths - helps immensely in pruning away confounding effects and enables researchers to prioritize potential relevant mechanisms for the pathogenesis in question. Given the enrichment of transcriptional regulators in ASD risk genes and the accessibility of high-throughput genome-wide readouts in the form of transcriptomic or epigenomic assays, exploiting the benefits of studying regulatory convergence stands out as a prime area of focus for the study of ASD pathobiology.

1.2.3 Regulatory convergence at the transcriptomic and epigenomic level

Evidence for regulatory convergence in ASD came first and foremost from transcriptomic and epigenomic studies of human post-mortem brain tissue [80–90]. Employing first microarray and later RNA-sequencing methods these studies were working towards the common goal of identifying shared changes in gene regulation between samples from individuals diagnosed with ASD and matched neurotypical controls. Through the use of

larger and larger patient/control cohorts they continuously improved the generalizability of their results and furthered our understanding of ASD at the gene regulatory level.

The first transcriptome study of post-mortem brain tissue was carried out on samples of the superior temporal gyrus from six ASD cases and six matched controls [91]. Despite the low statistical power their analysis revealed 221 genes as differentially regulated between ASD samples and controls. Of these 221, the vast majority (186) were found to be upregulated in ASD samples. Overall they manually annotated 72 of their differentially expressed genes (DEGs) as either immune-system related or cytokine responsive and sub-classified them into genes involved in cell communication and motility, cell fate and differentiation, and chaperones. Additionally the study reported increased variability in gene expression levels in ASD samples compared to controls.

In 2011 Voineagu et al. [80] published the first well-powered micro-array study using samples of prefrontal and temporal cortex as well as cerebellar vermis from 19 individuals diagnosed with ASD and 17 controls. The study found 444 significantly dysregulated genes in ASD cortex but only two genes in cerebellum. The two cortical regions were shown to be remarkably similar in terms of ASD related changes. Using co-expression network analysis the study found two gene modules correlated with disease status, one module being enriched for neuronal genes and downregulated in ASD while the other was upregulated and enriched for astrocyte and microglial markers. Interestingly, RBFOX1 was found to be downregulated in ASD samples and one of the hub genes in the neuronal gene module. The authors also identified alternative splicing events connected to RBFOX1 downregulation using RNA-sequencing further strengthening the connection between ASD and alternative splicing. The main findings regarding the two co-expression modules were replicated a few years later by Gupta et al. [82] in an independent set of 104 human brain cortical samples from 32 ASD cases and 40 matched neurotypical controls thereby providing further evidence for the role of immune and synaptic processes in ASD.
In 2016 the Geschwind lab expanded upon their original dataset by increasing their cohort to 48 individuals with ASD and 49 control subjects and a total of 251 samples from cortex and cerebellum [84]. Their analysis revealed a total of 1142 differentially expressed genes in cortex, 558 of which were downregulated in ASD. Through enrichment analyses they showed that downregulated genes were enriched for neuronal processes and neuron- as well as oligodendrocyte-specific genes. Upregulated genes on the other hand were enriched for immune system processes and microglia and astrocyte specific genes. The study also showed that differences between temporal and frontal cortex were attenuated in ASD, that these genes were enriched for Wnt signaling and suggested that the transcription factor SOX5 might play a role in this attenuation. In addition, co-expression analysis revealed evidence for differential age trajectories of two co-expression modules indicating an evolving process in early development. Lastly, they highlighted the dysregulation of 60 long non-coding RNAs (lncRNAs), 9 of which had been shown to interact with FMRP, and used their dataset for a more thorough analysis of alternative splicing. A total of 833 genes showed differential splicing events with most of them involving the exclusion of neuron-specific exons. Notably, alternative splicing was shown to not be driven by differential expression emphasizing its role as an additional level of gene regulation connected to ASD.

Two additional studies in the same year broadened the scope by on the one hand focusing on microRNAs [86] and on the other hand through the analysis of histone acetylation [85] in the same cohort. The latter study was of particular interest given the

enrichment for chromatin regulators in ASD genes since it presented clear evidence of common perturbations at this level of gene regulation. The histone mark investigated was acetylation of lysine 27 of histone 3 (H3K27ac), a mark of active enhancers and promoters generally correlated with gene expression [92, 93]. Around 5000 differentially acetylated (DA) peaks in the prefrontal cortex and 7000 DA peaks in the temporal cortex were identified between ASD and control samples using ChiP-seq. Only 247 DA peaks were found in the cerebellum, again arguing for the preferential vulnerability of cortex. A comparison between histone acetylation in promoter regions and gene expression changes in the same cohort [84] indicated positive correlation in either cortical region. Genes with associated upregulated DA peaks were enriched for neuronal functions as well as ASD risk genes while genes with downregulated peaks were enriched for immune-related terms. More evidence for convergence at the epigenomic level came from a DNA methylation study on largely the same cohort [94]. In line with findings at the transcriptomic level they could show DNA methylation changes associated with ASD. Using co-methylation network analysis they replicated the finding of ASD-associated synaptic and immune response related gene modules at the DNA methylation level.

Given that these results all came from bulk sequencing experiments the question remained whether the observed transcriptional changes were the result of variations in cell-type composition or molecular changes at the single cell level. In an attempt to shed light on this question Yu et al. [95] used single-cell RNA-seq data to deconvolute a published bulk dataset from post-mortem brain samples of individuals diagnosed with ASD and controls [83] and estimate the relative cell type abundances. Their analysis revealed significantly lower estimated proportions of neuronal cells and concurrent increased proportions of glial cells in ASD samples. Through mathematical separation of cell-type composition-dependent from independent effects on gene expression it was shown that cell-type composition was the main driver of observed differences at the bulk level. Nevertheless, a set of 644 genes was highlighted as potentially composition independent changes and was enriched for membrane and synaptic proteins as well as genes with neuron-specific expression. Of interest, these genes, in contrast to genes whose expression changed in a composition dependent manner, were also enriched for ASD risk genes.

In more recent years the focus went towards integrative analyses as well as employing single cell sequencing techniques to further increase the resolution and link back to the spatiotemporal convergence identified in the analyses at the risk gene level. In this context, Velmeshev et al. highlighted changes in upper-layer projection neurons as well as microglia as the most connected to ASD in terms of clinical severity using single nucleus sequencing [88]. Lastly, integrating datasets on mRNA expression, histone acetylation, DNA methylation, and micro RNA expression generated from the aforementioned studies Ramaswami et al. compiled a systems-level overview [90]. Using similarity network fusion they computed overall similarities between samples across modalities and clustered samples accordingly. Thereby they identified a subgroup of ASD samples differing from controls and, by performing subtype specific differential expression analysis, derived an additional 4356 genes differing between ASD samples and controls. This approach highlighted both the presence of a vast amount of similar molecular changes across ASD patients and the usefulness of sub-grouping patients based on information gained from molecular assays. The fact that the second group of ASD samples failed to show consensus molecular dysregulation even led the authors to validate the correctness of the diagnosis.

Linking back to the question of proper diagnosis of ASD, I am of the opinion that the

previous chapters provided evidence for the presence of both divergence at the patient sub-group level as well as convergence within these groups. In order to make use of this information further studies aimed at the sub-classification of ASD patients at the biological rather than behavioral level are needed as the aforementioned studies only scratched the surface in this regard. However, if successful, integration of biological information into the diagnosis of neuropsychiatric disorders holds the promise of enabling the development and delivery of effective interventions where needed.

1.3 Chromatin regulators in neurodevelopment

Epigenetic mechanisms play an integral role in all gene regulatory processes. This section aims to provide an overview of some of these mechanisms and their role in neurodevelopment, especially in a disorder context. As mentioned in the previous chapter, many proteins involved in the regulation of chromatin have been linked with ASD through the identification of mutations in the genes encoding them. In terms of functionality the predominant classes are chromatin remodelers which catalyze the translocation of histones along the genomic DNA, and histone modifiers which catalyze the addition or removal of post-translational modifications to histone tails. Additional important epigenetic processes such as DNA methylation or non-coding RNAs are beyond the scope of this chapter. The most common histone modifications include acetylation, methylation, phosphorylation, and ubiquitination of a variety of residues on multiple histone variants. While in many cases the exact role in the context of gene regulation is still unclear and subject to ongoing research there exist a number of well-studied modifications such as methylation of lysines 4, 9, 27, and 36 or acetylation of lysine 27 on histone 3 [96].

Histone methylation can play a role in creating both a repressive or active chromatin environment, dependent on where it is targeted. For example, histone 3 lysine 4 tri-methylation (H3K4me3) is a classic mark of active promoter regions and associated with actively transcribed genes. Likewise, mono-methylation of the same residue (H3K4me1) is a marker of active enhancers. Histone 3 lysine 36 tri-methylation (H3K36me3) has also been linked to active transcription but is located throughout gene bodies. On the other hand, methylation of histone 3 at lysines 9 or 27 (H3K9me3, H3K27me3) is associated with the formation of compacted, inactive heterochromatin or a repressive promoter state, respectively. Histone acetylation is generally linked to a more open chromatin state with histone 3 lysine 27 acetylation (H3K27ac) also marking active enhancer regions. For detailed reviews on the general function of histone marks see [97, 98].

Chromatin remodeling complexes on the other hand act through ATP-dependent translocation of nucleosomes thereby regulating accessibility of genomic regions. They can consist of many different subunits around the central catalytic Snf2 family ATPase-translocase. The Snf2 family is further divided into the SWI/SNF, ISWI, CHD, and INO80 subfamilies [99].

Histone modification in neurodevelopment

In the context of neurodevelopment, chromatin regulation has been shown to play an important role in a number of different processes starting with the presence of bivalent histone modifications at promoters of key developmental genes in pluripotent embryonic stem cells. These bivalent states are formed by the concurrent presence of both the active

H3K4me3 as well as the repressive H3K27me3 mark. This state prevents the expression of these genes in stem cells but gets resolved into an active state via the removal of H3K27me3 during differentiation down the neural lineage. A similar process happens with neuronal factors which are bivalent (H3K4me2/H3K27me3) in neural stem cells but become active upon terminal differentiation [100, 101]. This process is mediated by the RE1 silencing transcription factor (REST) which silences neuronal genes through the recruitment of histone deacetylases (HDACs) and also acts in long-term silencing of neuronal genes in cells of non-neuronal lineages through the recruitment of H3K9 methyltransferase G9a and H3K4 demethylase LSD1 [102].

Another example is the role of the Polycomb recessive complexes PRC1 and PRC2 which act as transcriptional repressors by H2AK119 monoubiquitination and H3K27me3, respectively [103]. It has been shown that loss of the H3K27 histone methyltransferase subunit Ezh2 at the onset of neurogenesis 12.5 days post conception leads to acceleration of neurogenesis with an eventually reduced neuronal output [104]. Later on, it is also essential in the regulation of the switch from neurogenesis to astrogenesis through the repression of the Wnt-signaling target gene neurogenin 1 *Ngn1/Neurog1* [105]. Mutations in *EZH2* have also been linked to Weaver's syndrome which is characterized by an overall overgrowth phenotype including macrocephaly, further highlighting the role of the PRC2 in neurodevelopment [106].

An interesting example of crosstalk between different histone marks is that H3K36 methylation counteracts Polycomb silencing by inhibiting methylation of the nearby H3K27 residue [107]. This was also demonstrated in human neurons where PRC2 inhibition was necessary for neuronal morphogenesis. Of interest, this was linked to the H3K36 methyltransferase ASH1L, an ASD risk gene. Depletion of ASH1L in this model system led to morphogenesis defects such as decreased neurite outgrowth [108].

Chromatin remodeling in neurodevelopment

Among the different chromatin remodeling complexes the SWItch/Sucrose Non-Fermentable (SWI/SNF) or BRG1/BRM associated factor (BAF) complex stands out as one of the most studied in the context of neurodevelopment. The central component of this complex is one of two catalytic ATPases, BRG1 (SMARCA4) and BRM (SMARCA2), alongside other core subunits like BAF155 (SMARCC1), BAF170 (SMARCC2), and BAF47 (SMARCB1). Its composition has been shown to change over the course of neurodevelopment and both a neural progenitor (npBAF) and neuronal configuration (nBAF) exist. BAF complexes play an integral role across many facets of neurodevelopment. In many ways they act as counterparts to Polycomb complexes, for example in the modulation of Wnt signaling [109, 110]. In a mouse model of complete deletion of all BAF subunits the authors showed that these complexes are essential for proper forebrain development. This study also demonstrated that BAF complexes modulated the activity of the H3K27 demethylases UTX (Kdm6a) and JMJD3 (Kdm6b) which could explain the overall increase in H3K27me3 upon BAF complex loss. Later in development these complexes are involved in the regulation of neurogenesis with the switch between npBAF and nBAF coinciding with the transition from proliferative to neurogenic cell division of neural progenitor cells [111]. Multiple components of the BAF complex have been implicated in neuropsychiatric disorders, primarily Coffin-Siris-Syndrome but also ASD (BAF155, BAF170, BAF180, BAF250b), 6q25 microdeletion syndrome (BAF250b), Nicolaides-Baraitser syndrome (BRM), Kleefstra's syndrome (BAF47), and schizophrenia (BRM) [112].

Next to the BAF complex, chromatin remodelers of the chromodomain helicase DNA-binding (CHD) family form another important regulatory axis in neurodevelopment. The nucleosome remodeling and deacetylase (NuRD) complex contains an ATPase (one of CHD3/4/5) as well as either GATA zinc finger domain containing protein 2A or B (GATAD2A/B) and CDK2 associated protein 1 (CDK2AP1). Its deacetylase subcomplex is comprised of a histone deacetylase (HDAC1 or HDAC2) alongside histone chaperones and two out of three metastasis tumor antigen proteins (RBBP4/7 and MTA1/2/3, respectively). The chromatin remodeling and histone deacetylation subcomplexes are joined by methly-CpG binding proteins 2/3 (MBD2/3) [113]. This complex plays a role in the regulation of gene expression in embryonic stem cells, particularly relating to the balance between self renewal and differentiation [114]. In cortical development heterozygous deletion of MBD3 was shown to induce premature cell cycle exit of neural progenitor cells and reduced neuronal output. The specification of upper layer neurons in this system was also negatively affected [115]. Of note, *CHD3* and *MBD3* are listed in the SFARI database as ASD risk genes.

Other members of the CHD family such as CHD7 (the causative gene for CHARGE syndrome) and CHD8 have also been linked to ASD. CHD7 plays a role in the proliferation of neural progenitors [116] as well as proper development of neural crest stem cells [117] and its loss leads to brain malformations in mice [118]. Lastly, CHD8 is linked to ASD cases with a brain overgrowth phenotype. At the molecular level it has been linked to the regulation of Wnt signaling although the exact nature of this regulation still needs to be clarified. Most studies indicate CHD8 as a negative regulator of Wnt signaling [119–122], however a study in mouse reported upregulation of Wnt signaling upon CHD8 knockdown [123].

These examples are aimed to provide an overview of the many roles that chromatin regulation plays in the context of neurodevelopment. Due to the abundance of chromatin regulating proteins this is only an abridged introduction into this vast field. For further reading I'd like to point out the following reviews: [113, 124–127].

1.4 Organoids as *in vitro* model systems for the study of neurodevelopmental disorders

For the longest time the study of the neurobiology of human neuropsychiatric disorder has for the most part been carried out using various model organisms such as mice. While this circumvents the issue of experimental inaccessibility it produces new issues such as potentially missing human specific biological processes and potential lack of translatability of results between species. On the other hand, *in vitro* differentiation techniques starting from human pluripotent stem cells were first restricted to two-dimensional culture systems with limited ability to capture aspects of the actual *in vivo* situation.

This situation improved dramatically with the advent of sophisticated three-dimensional differentiation systems, so called organoids, which brought a big improvement of how faithfully *in vivo* development could be modeled *in vitro*. Pioneering work by Eiraku et al. [128] based on previous work by Watanabe et al. in the same research group [129, 130] showed for the first time that cortical tissue could be generated in three-dimensional aggregation culture which self-organized to form rosette-like structures and that the

regional identity of this tissue can be controlled by patterning factors such as BMP, Wnt, or FGF.

In 2013, Lancaster et al. published their method for the generation of cerebral organoids [131]. This marked a big step forward from previous methods in that it for the first time presented a model system that could contain multiple brain regions such as the dorsal cortex, ventral forebrain, hippocampus, retina, choroid plexus and midbrain/hindbrain. In addition, by embedding embryoid bodies in Matrigel to provide a scaffold these organoids enabled the formation of large neuroepithelial buds and overall more complex tissue growth. In the same year, Kadoshima et al. published the next generation of the Eiraku et al. protocol marking a big improvement in modeling the development of the cerebral cortex *in vitro* [132].

These protocols both represented unguided approaches which mostly leveraged the self-organizing capacity while at the same time not enforcing a particular lineage. In contrast, patterned protocols are specifically designed to guide differentiation down defined paths. One of the most widely used of these systems is the cortical spheroid method developed by Paşca et al. in 2015 [133]. This method uses the dual SMAD inhibition paradigm developed by Chambers et al. [134] which uses small molecules to inhibit both BMP and TGFβ signaling to induce a neuroepithelial fate in embryoid bodies. Further factors used are growth factors FGF and EGF as well as neurotrophic factors BDNF and NT-3 resulting in the formation of dorsal forebrain tissue which over time generates both upper and lower layer excitatory neurons as well as astrocytes.

Since then, many different organoid protocols for the modeling of various brain regions have been developed such as cerebellar, hippocampal, and midbrain-specific organoids [135–137]. The spheroid toolbox was also greatly expanded by the development of ventral and striatal spheroids [138, 139] as well as the modular multi-region assembloid model system in which spheroids of different lineages are first grown separately and later fused to enable the study of processes such as interneuron migration [138]. Of note, fused organoid systems were also developed in parallel for cerebral organoids [140].

These sophisticated model systems were also characterized in detail by single cell RNA-sequencing or sequencing of sorted cell populations. This demonstrated the remarkable ability of cerebral organoids [141–143] as well as cortical spheroids [144, 145] to reproduce processes of early brain development with unprecedented faithfulness. This provided a great basis for the study of neuropsychiatric disorders with an early developmental component.

At this level one of the first proof of concepts came from the original Lancaster paper [131]. In this paper the authors demonstrated the possibility to model neurodevelopmental disorders such as microcephaly by using patient specific induced pluripotent stem cells (iPSCs) reprogrammed from fibroblasts. Closely related, later studies also investigated microcephaly caused by prenatal exposure to the Zika virus using organoid model systems [146, 147]. Furthermore organoids were used for the study of lissencephaly/Miller-Dieker syndrome [148, 149] and Birey et al. used the assembloid system to study interneuron migration defects in the context of Timothy syndrome [138, 150].

Multiple studies also focused on studying ASD using cerebral organoids or cortical spheroids. At the level of idiopathic ASD this was done using patient specific iPSCs in the context of ASD cases who also presented with macrocephaly [151]. The study found increased FOXG1 expression which led to an overproduction of inhibitory neurons and also

providing another piece of evidence for convergence in ASD. A later study also included both ASD patients with and without macrocephaly and found differences in molecular and cellular changes between the two cohorts [152]. Another recent study used a candidate gene based approach to model mutations of ASD related genes in organoids. They found changes in cell fate determination with an overproduction of inhibitory neurons to varying degrees dependent on genomic context [153]. Other studies focused on the genes *TSC1* and *TSC2* [154, 155] causative of tuberous sclerosis, a neurodevelopmental disorder which includes epilepsy, ASD, and intellectual disability.

Lastly, besides the study of neurodevelopmental disorders organoids also enable the study of human specific brain evolutionary processes. Through the use of organoids derived from other primates a recent study identified ZEB2 as a key factor in the expansion of the human neocortex [156]. Altogether these studies demonstrated the added value that more sophisticated *in vitro* models of human brain development such as cerebral organoids and cortical spheroids can bring and we will likely see further improvements in both culture systems as well as available assays in the future.

2

Outline

Genes linked to transcription or its epigenetic regulation make up a large portion of genes linked to autism spectrum disorder (ASD), as discussed in the introduction. Such risk genes provide a convenient entry point for the study of complex disorders such as ASD along the causal path from mutation to eventual phenotype. Given the complexity of both the neurodevelopmental context and the role of gene transcription in it, this creates a highly interesting area of study, which I have chosen as the overarching theme of my PhD.

Transcriptional and epigenetic regulation play a role in many dynamic processes where they calibrate the context-appropriate cellular response to intrinsic or extrinsic signals [157–160]. In a neurodevelopmental context one such key task lies in the cellular specification and differentiation from pluripotent stem cells to generate the plethora of cell types in the brain [161]. In addition, long-term responses to neuronal activity also depend on the precise regulation of activity-dependent transcription to enable neuronal plasticity, e.g. through long term potentiation [162, 163]. Haploinsufficiency, or the intolerance to the loss of one allele of a gene [164], has been suggested as a key mechanism of heterozygous mutations associated with developmental disorders connected to histone modifiers [165]. Furthermore, transcriptional regulators and epigenetic complexes have been found to be enriched in sets of haploinsufficient/loss-of-function genes [164, 166] indicating a particular vulnerability of these processes to gene dosage effects.

Taking into account both the importance of transcriptional regulation in dynamic processes, its vulnerability to gene dosage effects, and its connection to ASD, a neurodevelopmental disorder, my hypothesis is that ASD cases linked to mutations in transcriptional regulators might arise not just from a change in steady state RNA levels but rather from disproportionate responses to intrinsic/extrinsic signals. This hypothesis predicts that phenotypes caused by such mutations will likely be observed in dynamic processes such as cell differentiation during development or in response to neuronal activity.

Over the course of my doctoral work I studied transcriptional consequences of ASD associated mutations in multiple brain-specific contexts. First, I worked on a project employing mouse and *in vitro* models to investigate the ASD candidate gene *SETD5*, a predicted histone methyltransferase enzyme. This collaborative project published in 2018 [167] was comprised of an *in vivo* part handled by our group and an *in vitro* part

performed by our collaborators. The *in vivo* part was further divided among three co-first authors with me being one of them. My contribution to this project was handling the transcriptomic dimension of the *in vivo* part starting from performing RNA-sequencing experiments to the analysis, visualization, as well as interpretation of the resulting data. This was done both in a developmental context, 9.5 days old mouse embryos, and in a functional setting - to measure the difference in transcriptional response in the contextual fear conditioning behavioral paradigm. Lastly, I was heavily involved in the integrative analysis of the datasets produced throughout this project.

Second, I was part of a project using cerebral organoids as an *in vitro* model system of human brain development to study the transcriptional perturbations resulting from heterozygous loss of the gene *CHD8* [168]. My role as a shared second author involved performing bulk RNA-seq experiments and analyzing, visualizing, and interpreting the resulting data as well as contribute to the analysis, visualization, and interpretation of the single cell RNA-sequencing data from this study.

Finally, the main project of my PhD, which at the time of writing this thesis is still ongoing, focuses on the study of regulatory convergence in ASD using cortical spheroids as a model system. Over the course of this project I generated heterozygous mutant human embryonic stem cell lines for a total of four ASD associated chromatin regulators (*ASH1L, SETD5, KDM6B* and *KMT5B*), established and employed cortical spheroid culture using these lines and performed and analyzed transcriptomic experiments to investigate the developmental perturbations caused by these mutations. Leveraging the use of multiple mutant lines within one study I could demonstrate convergent changes shared between three of the four risk genes.

3

Haploinsufficiency of the intellectual disability gene *Setd5* disturbs developmental gene expression and cognition

3.1 Context

In this project we employed mice as a model system for the study of the intellectual disability and autism spectrum disorder (ASD) gene *SETD5*. A mouse line with heterozygous disruption of *Setd5* was created and assessed using behavioral, electrophysiological and molecular assays. For the scope of this thesis I will highlight my contribution as well as the contact points of these with the rest of the paper. This meant a partial restructuring and refocusing of the published manuscript as well as its supplementation with additional previously unreported data analysis. Given the role of Setd5 as a chromatin modifier and therefore transcriptional modulator I performed RNA-sequencing experiments in two contexts where dynamic responses to external signals play a significant role: cell differentiation in neurodevelopment and activity-dependent transcription during learning. My analysis of embryonic samples was further intersected with RNA-sequencing and ChIP-seq experiments on *in vitro* samples from our collaborators in Heidelberg to gain a more complete understanding of the role of Setd5 in transcriptional regulation. Additionally, the data on activity-dependent transcription was put in the larger context of behavioral and electrophysiological differences. This work was published in 2018 (Deliu, Arecco, Morandell, and Dotter et al, [167]) and the figures presented here were largely adapted from this publication.

3.2 Introduction

Intellectual disability has been estimated to affect around 2% of the general population [169, 170] and is one of the main comorbidities of ASD [171]. Similar to ASD, genetic studies have identified a number of candidate genes implicated in the risk for intellectual disability [172], to a large part based on *de novo* mutation rates. Due to this heterogeneity the etiology of intellectual disability as well as any commonalities or differences with ASD etiology remain unknown.

SET-domain containing 5 (*SETD5*) is a gene encoding for a protein predicted to act as a histone methyltransferase based on sequence similarity. *De novo* mutations in this gene have been identified to cause intellectual disability and ASD alongside a number of craniofacial abnormalities, a condition also referred to as intellectual developmental disorder, autosomal dominant 23 (OMIM #615761) [173, 174]. Its association with ASD is also recognized by its high SFARI score of 1. The prevalence of heterozygous loss of *SETD5* among patients with intellectual disability was estimated at 0.7% [173]. A previous study provided first glimpses into the molecular functions of SETD5. They showed that the protein interacts with the Polymerase-associated factor 1 (PAF1) and Nuclear receptor co-repressor (NCoR) complexes and demonstrated the lethality of full *Setd5*$^{-/-}$ knockout in mice [175]. However, a thorough study of the effects of heterozygous Setd5 loss, mimicking the situation found in patients with *de novo* mutations, is still lacking.

The PAF1 complex plays a role at all stages of transcriptional regulation, i.e. initiation, elongation, and termination, mainly through its interaction with RNA polymerase 2 (Pol II) [176]. In particular, it has been connected to the regulation of Pol II pausing, a process in which Pol II is recruited to a target gene, transcription is initiated but stalls at a pausing site 20 to 120 nucleotides downstream of the transcription start site [177]. Pol II pausing has been shown to be a wide spread and important mechanism in the regulation of transcription [178]. While there is general agreement about the connection between Paf1c and Pol II pausing the exact nature of this regulation is unclear. Some studies suggested that Paf1c acts in maintaining the paused state [179] while others argued its function lies in the release of paused Pol II into productive transcriptional elongation [180, 181]. In the context of mammalian development polymerase pausing has been linked to cell cycle related genes and differentiation potential through signal transduction in embryonic stem cells [182]. In neural progenitor cells and their subsequent differentiation into neurons it was found that Pol II pausing was more connected with the expression of genes important for the current cell state. This went against the assumption that Pol II pausing might prime genes of downstream lineage for expression upon differentiation. However, this was true for bivalent chromatin marks, i.e. co-occurrence of repressive H3K27me3 and active H3K4me3, at promoters [183]. For a comprehensive review of transcriptional pausing and it's connection to neurodevelopmental disorders see Eigenhuis et al. [177]. In humans the PAF1 complex is made up of six subunits PAF1, LEO1, CDC73, CTR9, RTF1, and SKI8/WDR61. This complex is highly conserved and, with the exception of the human-specific SKI8 subunit, has the same composition in mouse [177].

NCoR complexes on the other hand act as transcriptional repressors. They consist of the scaffolding protein NCOR1 (or its homolog NCOR2) and its major catalytic component histone deacetylase 3 (HDAC3), as well as two WD40 domain containing proteins TBL1X

and TBL1XR1 and GPS2 [184, 185]. Of particular interest, this complex is strongly connected to both intellectual disability and autism [186]. NCOR1, TBL1X, and TBL1XR1 are listed in the SFARI database as high risk ASD genes. In addition, MECP2, the main risk gene for Rett syndrome, interacts with the NCoR complex and this interaction has been shown to be abolished by Rett syndrome associated mutations [187].

Here, we study the effects of Setd5 haploinsufficiency in mice using a multi-faceted approach. At the level of transcription we discovered that Setd5 is involved in gene regulation programs across the lifespan of the animals. On the one hand, we investigated the role of Setd5 in a developmental context and found that *Setd5*$^{+/-}$ mouse embryos exhibit dysregulation of cell type specification as early as 9.5 days post conception (E9.5). On the other hand, we found that partial loss of Setd5 interferes with the consolidation of learning at the transcriptional level, particularly of genes encoding for post-synaptic proteins. In addition, our results show that *Setd5*$^{+/-}$ neurons display increased long term potentiation. Lastly, we also confirmed and extended the current knowledge on the interaction partners of Setd5 through proteomics and chromatin immunoprecipitation followed by sequencing (ChIP-seq) experiments.

3.3 Methods

Experiments/Analysis performed by researchers other than Christoph Dotter are indicated. For a complete list of all methods used in this study please refer to the publication.

Mice

Performed by the Preclinical facility

Setd5$^{tm1a(EUCOMM)Wtsi}$ mice from the International Mouse Phenotyping Consortium were crossed with mice homozygous for a Flip cassette (Jackson Laboratories) in order to remove the NeoStop cassette. In order to obtain *Setd5*$^{+/-}$ mice CMVCre (B6.C-Tg(CMV-cre)1Cgn/J) males were mated with heterozygous *Setd5*$^{+/loxP}$ females. C57BL/6J mice were used to backcross *Setd5*$^{+/-}$ mice to the N10 generation after which they were used for experiments. Cages housed 3-4 animals and the mice were kept on a 12h light/dark cycle with the light period starting at 7:00 a.m. Unless explicitly mentioned otherwise, food and water were available to the animals ad libitum. The Setd5GFP mouse model [175] was kindly provided by M.A. Magnuson from the Vanderbilt University in Nashville, TN.

Hippocampal slice preparation and electrophysiology

Performed by Elena Deliu

Animals were anesthetized using isoflurane, brains were dissected and rapidly placed in ice-cold cutting solution (equilibrated with carbogen (95% O_2/5% CO_2); 120 mM sucrose, 64 mM NaCl, 25 mM $NaHCO_3$, 10 mM glucose, 2.5 mM KCl, 0.5 mM $CaCl_2$, 1.25 mM NaH_2PO_4, 7 mM $MgCl_2$). Both hippocampi were dissected, fixed against an agarose (4%) block and sliced on a VT 1200S vibratome (Leica Microsystems). Slices were first recovered at 33°C for 1 hour and RT thereafter in carbogenated artificial cerebrospinal fluid (aCSF; 130 mM NaCl, 28.82 mM $NaHCO_3$, 11 mM glucose, 2.5 mM $CaCl_2$, 5 mM

sodium ascorbate, 3 mM sodium pyruvate, 2.75 mM KCl, 1.43 mM $MgSO_4$ and 1.1 mM NaH_2PO_4; 320 mOsm, 7.2–7.4 pH).

Roughly 90 minutes after slicing a single dorsal hippocampal slice was attached to a poly-D-lysine coated round coverslip (12 mm; cleaned with 1 N HCl, sonicator, 15 min; washed with water, sonicator, 3 x 15 min; maintained in 96% EtOH prior to drying and coating) and transferred to a recording chamber (RC-41LP, Warner Instruments). Slices were visualized using a BX-51WI microscope (Olympus) and infrared-differential interference contrast (IR-DIC) with a QIClick™ charge-coupled device camera (Q Imaging Inc, Surrey, BC, Canada). Borosilicate glass recording microelectrodes (World Precision Instruments) were pulled on a P-1000 puller (Sutter Instruments) and backfilled with aCSF. Stimulating and recording electrodes were placed roughly 500 µm apart in the CA1 stratum radiatum to evoke field excitatory postsynaptic potentials (fEPSP) from the Schaffer collateral/commissural-CA1 synapses. For fEPSP recording (measured as its initial slope function) the recording electrode (1.6 - 2.4 MΩ) was placed in the CA1 apical dendritic layer, a Multiclamp 700B amplifier was used for signal amplification, and the signal was digitized using Digidata 1550A (Molecular Devices). After 15 min of pre-incubation a synaptic input-output curve (afferent stimulation vs fEPSP slope) was generated. The test stimulation intensity was adjusted to elicit fEPSP slopes of 40% of the maximal EPSP response. Long term potentiation (LTP) was induced via high frequency stimulation (4 bursts of 100 pulses for 1 s (100 Hz) delivered every 5 s). Pulse width during tetanization was 0.2 ms and was of equal length as the test pulse width. Baseline recording was obtained by delivering 0.2 ms long constant-current pulses every 30 s (0.03 Hz). Stable baseline was monitored for 30-60 min upon applying the high frequency stimulation protocol, and the fEPSP was recorded after induction, every 30 sec, for 3-5 hours using pCLAMP 10.4 (Axon Instruments/Molecular Devices). For paired-pulse facilitation experiments an input-output curve was obtained and stimulus intensity set as described above. The slice was then left to rest for 10-15 min and three paired-pulse facilitation protocols (inter-pulse intervals of 40/80/20 ms) were applied sequentially (5 min each). The resulting sweeps were averaged, the two fEPSPs in the averaged trace calculated and plotted as a ratio. All slices showed stable baseline (fEPSP slope of first pulse) for the entire duration.

Novel object location test

Performed by Elena Deliu

The novel object location test was done according to a published protocol [188]. In preparation for the test individual mice were handled for 2 min every day for 5 days. On the following 5 days they were exposed to an empty arena (30x30x40 cm^3, Phenotyper, Noldus). The day after (day 11) mice were trained with identical objects placed 7 cm apart in defined locations for 3 min in the same arena. After 24 hours mice were tested again for 5 min with the same objects and in the same arena but with one object having been relocated to a new position 11 cm away from the other. Normally, mice show preference for the relocated object but a 3 min training paradigm is not enough for long-term memory retention in wildtype mice. For the evaluation of the test the cumulative time spent exploring the objects was recorded. From this data a discrimination index (DI) was calculated as $(time_{exploring\ relocated\ obj.} - time_{expl.\ other\ obj.}) * 100/time_{total\ expl.}$. DI > 30% was deemed as an indicator of strong memory retention.

Place aversion test (IntelliCage)

Performed by Elena Deliu

A glass-covered transponder with unique IDs (Datamars SA) was implanted subcutaneously into female animals (age: 2-4 months) under isoflurane anesthesia to allow for radio-frequency identification (RFID) of animals. The IntelliCage (NewBehavior AG) is a automated testing apparatus for the analysis of behavior of RFID-tagged mice [189]. It consists of a standard plastic cage (55 x 37.5 x 20.5 cm^3) with four triangular learning chambers (15 x 15 x 21 cm^3) in each corner, RFID readers and other sensors for the simultaneous monitoring of up to 16 tagged mice. Mice could enter each corner through a tunnel functioning as an RFID antenna. Only one mouse can enter a corner at any time. Inside the mice find two nose poke holes equipped with an infrared beam-break detector. "Correct" nose pokes trigger the opening of an access gate to water-bottle nipples (henceforth referred to as gate). Each behavioral event (corner visit, nose poke, and lick) was automatically recorded using RFID readers, infrared sensors and lickometers. Intellicage software 3.2.3 was used.

First, mice were allowed to adapt to the IntelliCage environment for 2.5 days with free access to water (any corner visit would open the gate). Subsequently the nose poke behavior was trained by only granting access to water upon any nose poke in a corner (gate open for 7 s). After the adaptation phase a place aversion paradigm was introduced where each mouse was randomly assigned an "incorrect" corner where nose pokes would trigger an aversive air-puff (1 bar) stimulus for 1 s. Nose pokes in "correct" corners provided access to water as before. Mice were trained in this task for 48 hours before a 24 hour trial period during which no air puffs were given. Performance was quantified in form of the ratio of nose pokes per visit in "correct" and "incorrect" corners both during training and during the trial.

Contextual fear conditioning

Performed by Elena Deliu

Mice were subjected to the contextual fear conditioning (CFC) task in three separate sessions, with 24 hours between sessions. The first session (day 1) was a training session followed by two re-exposure sessions on days 2 and 3.

Strong training paradigm: The training session lasted for one round of 5 minutes during which the mice learned to associate the conditioned stimulus, in this case the context, with the unconditioned stimulus, a foot shock. Each mouse was placed in a chamber (18 cm^2, Noldus) with an electrified grid floor. After two minutes of free exploration, a foot shock (0.5 mA, 2 s) was applied three times with one minute between shocks. One minute after the last shock the mice were removed from the conditioning chamber. During re-exposure sessions the mice were placed back in the same conditioning chamber without delivery of further foot shocks for ten minutes (strong extinction paradigm) on day 2 and five minutes on day 3 in order to test memory retention and extinction, respectively. A camera mounted above the cage ceiling was used in tandem with the Ethovision XT software to record the behavior. The relative amount of time spent freezing (i.e., absence of all but respiratory movements for $\geqslant$ 3 s) was used as a proxy for emotional reactivity during training and fear memory during retention (day 2, first three minutes) and extinction

(day 3, first three minutes). Behavioral parameters were scored blind to experimental condition.

Weak training paradigm: Duration of the training session was reduced to 150 s with only a single foot-shock (0.5 mA, 2 s) applied after 120 s of exploration and the session on day 3 was reduced to 150 s. Otherwise the same general procedure was applied.

CFC in preparation for RNA-seq: A separate batch of animals, 15 per genotype, was used for the preparation of RNA-seq libraries over the course of the training session on the first day to measure the transcriptional response. For this purpose *Setd5*$^{+/+}$ and *Setd5*$^{+/-}$ females were housed individually for 6 days and pairs of littermates were randomly assigned to one of three experimental groups (5 animals per group per genotype): "home cage", "1 hour after CFC", "3 hours after CFC". All animals were accommodated to the behavioral room overnight. Animals of the "home cage" group were sacrificed upon isoflurane anesthesia, the brain removed and transferred to ice-cold artificial cerebrospinal fluid (125 mM NaCl, 25 mM $NaHCO_3$, 25 mM glucose, 2.5 mM KCl, 2 mM $CaCl_2$, 1.25 mM NaH_2PO_4, 1 mM $MgCl_2$), the hippocampi removed and the CA regions dissected using a 30G needle [190], snap-frozen using liquid nitrogen and stored at -80°C until further processing. Mice in the other two groups were subjected to the CFC strong training session on day 1 and sacrificed for sample collection as described one or three hours after. Littermate pairs were randomly tested either in the morning or the afternoon to account for differences due to circadian rhythm.

RNA isolation and library generation from mouse samples

Mouse dissection/genotyping performed by either Elena Deliu or Jasmin Morandell

Mouse tissue was rapidly dissected on ice, snap-frozen in liquid nitrogen and stored at -80°C until further processing. Genotyping for E9.5 embryos was carried out on placental tissues. Either whole embryos or hippocampal CA region from the left hemisphere were used as input for RNA isolation and preparation of libraries for RNA-sequencing.

RNA was isolated using Trizol/Chloroform extraction. Snap-frozen samples were homogenized in 700 µL Trizol (Invitrogen) after which 140 µL chloroform (Sigma) were added, samples mixed by vortexing for 15 sec, and incubated at room temperature for 2-3 min before centrifugation at 12,000g for 15 min at 4°C. The aqueous upper phase was transferred to a new tube, mixed with 1.5 volumes of 100% ethanol (EtOH) before loading onto a Zymo-Spin™ IC column (Zymo Research). Afterwards the column was washed with 400 µL 70% EtOH and the sample was treated with RQ1 DNAseI (Promega; 40 µL 70% EtOH + 5 µL 10X DNAse reaction buffer + 5 µL DNAseI) for 15 min at RT. After two more washes with 700 µL EtOH the final elution was done with DEPC-treated H_2O. RNA concentration was measured on a NanoDrop spectrophotometer (Thermo Scientific).

For RNA-sequencing library preparation either 1 µg (hippocampus CA) or 500 ng (E9.5 whole embryo) RNA were used as input for the QuantSeq 3' mRNA-Seq FWD Library Prep Kit (Lexogen) and libraries were prepared according to the manual. Quality and size distribution were analyzed using the High Sensitivity DNA Analysis Kit (Agilent) for the Bioanalyzer 2100 (Agilent). Concentration measurements for multiplexing were done using the Qubit® dsDNA HS Assay Kit for the Qubit® 2.0 fluorometer and libraries were pooled in equimolar amounts. Libraries were sequenced as 50 bp single-end reads on an

Illumina HiSeq 2500. Sequencing and demultiplexing was performed by the NGS facility of the Vienna Biocenter Core Facilities.

Transcriptomic analysis - mouse data

Trimming and filtering of demultiplexed raw sequencing reads was performed according to the guidelines provided by Lexogen. In short, BBDuk (BBMAP package[1]) was used to remove/filter out adapter contamination, random primer sequences and low quality tails. The trimmed reads were aligned to the mouse genome (genome: mm10/GRCm38, gene annotation: Gencode M15) using STAR (version 2.5.1, [191]). Only uniquely aligned reads were considered (–outFilterMultimapNmax 1) and STAR was also used to produce read counts per gene (–quantMode GeneCounts). All subsequent analysis was done in R 3.5.0.

Hippocampal CA region from CFC animals

General coherence across replicates was determined using sample distance clustering and principal component analysis (PCA). Based on these results, one sample (wild type, 1 h) was excluded from further analysis due to clustering away from all remaining samples. For differential expression analysis the package RUVSeq was employed, as previously described [192]. In short, initial data exploration and upper quartile normalization was done using the EDASeq package (version 2.14.0, [193]). All samples were normalized together. Thus normalized data was used as input for the RUVs function of RUVSeq (version 1.14.0, [194]) to estimate unwanted sources of variation. Success of normalization was evaluated for a range of k values through visualizing remaining variation through relative log expression (REL) plots and sample clustering through PCA. The chosen k value (k = 10) represented the minimum value that yielded visible separation of samples into groups in the PCA plot. All expressed genes were used as negative controls since the vast majority were not expected to change across conditions (see Peixoto et al. [192]). Following confounding factor estimation differential gene expression was assessed using the edgeR package (version 3.22.2, [195]) including the RUV factors in the model as described and collapsing genotype and time-point into a single variable (type) for ease of pairwise comparisons ($\sim 0 +$ type + [RUVfactors]). Genes with an adjusted p-value (FDR) $\leqslant 0.05$ were selected as significantly differentially expressed. Immediate early genes for visualization were taken from Mews et al, 2017. [196].

To derive genes which showed a significantly different response across the time course edgeR was used with a different model ($\sim$ genotype + time point + [RUVfactors] + genotype:time point) to test for significant genotype:time point at either 1 hour or 3 hours. For trajectory clustering read counts were normalized and log transformed (edgeR::cpm), before averaging across biological replicates. To put expression changes on a uniform scale gene-wise z-scores were calculated. Genes were clustered using k-means clustering with Pearson correlation as distance measure. The final cluster number was chosen by evaluating the Davies Bouldin index across different ks and requiring a minimum cluster size of 10 genes. This clustering was performed for two sets of differentially expressed genes: 1) including only genes with significantly different response (selected k = 11) and 2) all genes which were differentially expressed in at least one pairwise comparison (selected k = 22). Cluster trajectories were visualized using ggplot2 [197].

[1] http://jgi.doe.gov/data-and-tools/bbtools/

Embryonic samples

Differential gene expression analysis was done using RUVSeq/edgeR as described above. In addition to the RUV factor (k = 1) sex was included in the model (∼ genotype + [RUVfactor] + sex) as a confounding factor since both male and female embryos were included. The overlap between the embryo DEGs and DEGs from the *in vitro* experiments was visualized using a modified version of the chordDiagram function from the circlize package (version 0.4.4, [198]).

Enrichment analyses

Gene Ontology functional enrichment analysis for either mouse or *in vitro* results was done using the GOstats package (version 2.46.0, [199]) with a P value cutoff of 0.001, all expressed genes as background, and conditional testing enabled. For the E9.5 data a random sampling strategy (100 random gene lists of the same length) was implemented to exclude terms enriched in overall expressed genes. For E9.5, GO term enrichment results were visualized with a modified version of the GOCircle function from the GOplot package (version 1.0.2, [200]).

Cell type specific gene expression enrichment in CFC results was assessed using the EWCE package (version 0.99.2, [201]) with the scRNA-seq data from Zeisel et al. [202]. Enrichment analyses for sets like transcription factor targets, pathways and disease genes (OMIM) were done using the Enrichr web tool [203, 204]. Additional gene set enrichment analyses were all done in R using a one-sided Fisher's exact test.

The source for the Hdac3 target genes used for enrichment in the CFC data was published ChIP-seq data from mouse hippocampus [205]. For the E9.5 data the ChIP-seq data generated in this study was used. CREB1 target genes were derived from the Ches2016 dataset from Enrichr.

Western blot for histone acetylation

Performed by either Elena Deliu or Jasmin Morandell

Hippocampal CA tissue was homogenized in 1x pre-lysis buffer, centrifuged, resuspended in lysis buffer and incubated on ice for 30 min. Lysates were centrifuged (12,000 rpm, 4°C) and the supernatant was transferred to a new vial and mixed with 0.3 volumes Balance-DTT buffer. Protein concentration was measured with the Pierce™ BCA Protein Assay Kit (Thermo Fisher, #23225). Aliquoted lysates were stored at -80°C.

For Western blot 2.5 µg protein were mixed with 6x Laemmli buffer (375 mM Tris-HCl pH 6.8, 12% SDS, 60% glycerol, 600 mM DTT, 0.06% bromophenol blue) heated to 95°C and resolved on a 15% SDS-PAGE gel in running buffer (3.03 g Tris base, 14.1 g glycine, 1 g SDS in 1 liter MilliQ water). Transfer to a nitrocellulose membrane was done in transfer buffer (running buffer without SDS) for 30 min at 4°C using a Micro Cell Western blotting apparatus (Bio-Rad). After blocking for 1 hour with 5% BSA in 1x Tris buffered saline (TBS) with 1% Tween (TBST) membranes were incubated with primary antibody, H4K8ac (Abcam, ab45166) or Histone 4 (ab10158) overnight at 4°C. The next day, membranes were washed and incubated with horseradish peroxidase coupled anti IgG secondary antibody for 1 hour at room temperature. Detection was done using Pierce

ECL Western Blotting Substrate (Thermo Scientific, #33209), image acquisition was done using an Amersham™ Imager 600 and Fiji/ImageJ was used for quantification.

In vitro experiments

Experiments designed and performed, and data analyzed by Niccolò Arecco, Charles Girardot, Eva-Lotta Käsper and Kyung-Min Noh

Cell culture

Mouse embryonic stem cells (mESCs) were cultured in Knock-Out DMEM containing 15% fetal bovine serum (FBS, EmbryoMax, Millipore), 1% non-essential amino acids, 1% penicillin/streptomycin, 1% GlutaMax, and 20 ng/ml leukemia inhibitory factor (LIF). Medium changes were performed every day and cells were passaged every second day. Cells were routinely tested for mycoplasma contamination.

HEK293T cells were cultured in regular medium (DMEM high glucose medium with 10% regular FBS, 1% penicillin/streptomycin, 1% GlutaMax, and 1% sodium pyruvate) and passaged every four days.

mESCs were differentiated into neural progenitor cells (NPCs) as previously described [206]. In short, mESCs were deprived of LIF and plated in low-adhesion petri dishes in regular medium (described above) on culture day zero. Media was changed every second day and from day four until day eight the medium was supplemented with 5 µm retinoic acid.

RNA isolation and library generation from *in vitro* samples

RNA was extracted with the RNAeasy kit (Qiagen) and contaminating genomic DNA was removed using the Turbo DNAse kit (Ambion). RNA quality was assessed using the Bioanalyzer (Agilent). RNA-sequencing libraries were prepared from samples with RIN $\geqslant 9$ using the NEBNExt Ultra II kit with the oligo-dT capture add-on (NEB). Eight to twelve libraries were multiplexed and sequenced together on an Illumina HiSeq 2000 at the EMBL, Heidelberg Gene Core facility.

Differential expression analysis - *in vitro* data

Fastq files were aligned to the mm10 reference genome using bowtie2 [207] and reads per gene were counted using HTseq (version 0.6.1, [208]). Coherence between biological replicates was checked by PCA. Differential gene expression analysis was done with the DESeq2 R package (version 1.16, [209]). For ESC samples in the heterozygous comparison the variance estimated by DESeq2 was deemed too high (based on P value distribution) and therefore P values were re-estimated using fdrtool (version 1.2.15, [210]).

Chromatin IP coupled with sequencing (ChIP-seq)

Cross-linking: Hdac3, Pol II and Setd5: 100 million ESCs per replicate were harvested and cross-linked in 1.5 mM EGS (ethylene glycol bis(succinimidyl succinate)) in PBS for 1 hour followed by 10 min in 1.5% paraformaldehyde (PFA) in PBS at room temperature (RT, 21-23°C). H3K27ac and H4 panacetylation: 20 million ESCs per replicate were harvested and cross-linked in 1% PFA.

Cross-linking was quenched with 125 mM glycine for 5 min. Cells were pelleted at 3000 rpm for 10 min at 4°C and washed twice with ice-cold PBS supplemented with complete protease inhibitor cocktail. Samples were snap-frozen with liquid nitrogen and stored at -80°C until further processing.

Chromatin was prepared by resuspending cross-linked cells in hypotonic buffer for 10 min on ice, followed by 35 strokes of douncing with a tissue homogenizer. The cytoplasmic fraction was removed by gentle centrifugation and the nuclear pellets resuspended in MNase digestion buffer. Samples were digested with MNase (Worthington) using 15 µL per 50 million cells (25 U/µL MNase) for 5 min at 37°C. Digestion was quenched using an EDTA/EGTA based quenching solution.

Digested chromatin was fragmented via sonication in three cycles on a Bioruptor Pico sonicator (Diagenode) and insoluble chromatin was pelleted by centrifugation at 15,000 g for 10 min at 4°C. Soluble fragmented chromatin was used as input for immunoprecipitation (IP). Both soluble input and insoluble pellet were loaded onto an agarose gel to check the digestion pattern across replicates. 5% of each input were set aside before the IP. Samples were incubated overnight at 4°C with 70 µL of M-280 beads (Invitrogen) precoupled with the specific antibody (6 h at 4°C). On the following day beads were washed four times with LB3-100 buffer (100 mM NaCl, 10 mM TRIS, 1 mM EDTA, 0.5 mM EGTA, 0.5% sarcosine, and 0.1% sodium-deoxycholate), one wash with LB3-500 buffer (500 mM NaCl), and four washes with RIPA-250 buffer (10 mM HEPES, 250 mM LiCl, 1 mM EDTA, 1% IGEPAL CA-630, and 0.7% sodium-deoxycholate). Finally, beads were rinsed with 10 mM Tris/1 mM EDTA buffer and IP chromatin was eluted with elution buffer (1% SDS, 50 mM Tris, 10 mM EDTA) for 30 min at 65°C.

Eluted IP chromatin and ChIP inputs were reverse cross-linked and proteins digested with proteinase K overnight at 65°C. DNA was purified using a PCR purification kit (Qiagen) and used as input for library preparation with the NEBNext Ultra II kit. For each IP target two biological replicates per condition (WT, ΔSET Setd5 het or hom) and one input per genotype (Setd5 WT and Setd5 HA in the case of Setd5 ChIP) were multiplexed together and sequenced on an Illumina HiSeq 2000 at the EMBL, Heidelberg Gene Core facility.

ChIP-seq analysis

ChIP-seq analysis was performed using a local instance of Galaxy [211] maintained by the EMBL Genome Biology Computational Support and R (version 3.5.0, Bioconductor version 3.7). Data was visualized using custom R scripts. Sequencing reads were aligned to mouse genome version mm10 with Bowtie2 [207] using standard options (Galaxy tool version 0.2, sensitive preset). Only uniquely aligned reads were kept for further analysis. Duplicates were removed with Picard's[2] Mark Duplicates. Sequencing quality was checked with Deeptools2 [212] and FastQC[3].

Peaks were called using MACS (version 2, [213]) and the following parameters: Hdac3 (–gsize mm –bw 250 –mfold 5 80 –pvalue 0.0025), Pol II (–gsize mm –bdg –bw 250 –mfold 30 80 –pvalue 0.0025), H3K27ac (–bw 150 –mfold 15 80 –qvalue 0.001), and H4 pan acetylation (–gsize mm –bdg –bw 150 –mfold 5 80 –pvalue 0.001). Peaks from all conditions were merged using bedtools [214] into unique non-overlapping sets of 25,234

[2]http://broadinstitute.github.io/picard

[3]https://www.bioinformatics.babraham.ac.uk/projects/ fastqc/

(Hdac3, Q<0.001), 129,203 (Pol II, Q<0.001), 45,842 (H3K27ac, Q<0.00025), and 53,778 peaks (H4 pan-acetylation, Q<0.1). Enhancers from Shen et al. [215] were adjusted to 2 kb around the middle point and lifted from mm9 to mm10 using the UCSC liftover tool. TSS regions were defined using ENSEMBL version 91 as 2.5 kb regions centered on the TSS of genes longer than 1,250 bp (total: 34,159 genes). Intersecting peaks with these regions gave sets of 14,015 (Hdac3), 16,152 (Pol II), and 8,804 (H3K27ac).

Differential comparisons between mutant and WT were done using DiffBind [216] and edgeR [195] and differentially bound peaks with adjusted $P < 0.05$ were included for further analysis. The Deeptool2 multiBamSummary tool was used to normalize read counts from deduplicated bam files as RPKM. Replicates were averaged and all samples background-normalized by dividing them by the merged input RPKM. Signal files (bigwig) were generated using the Deeptools2 bamCoverage tool and correcting for library size ("Normalize coverage to 1x" option). Replicates were averaged and input was subtracted using the Deeptools2 bigwig-Compare tool.

The definition of the Pol II pausing index was adapted from Chen et al. [179]. For each Setd5 merged replicate the input-subtracted bigwig sample of the average signal at the TSS (-100 to +300 bp) was divided by the gene-body signal downstream of the TSS (+301 to +3,301 bp). Only genes longer than 3.5 kb and with signal > 1 at both TSS and gene-body were included in the analysis. This led to the inclusion of 9,490 genes for the pausing index calculations.

Correlation between biological replicates was verified (Pearson's $r \approx 0.98$ for Pol II at TSS and gene-body; $r \approx 0.89$ for Hdac3 at TSS) and samples were merged for visualization. For box plots in figure 3.4 data points < 5th percentile or > 95th percentile of mean RPKM were removed in a, b, and e. Whiskers in box plots extend to the most extreme data point within 1.5x the interquartile range. Outliers are not shown.

Setd5 immunoprecipitation followed by liquid chromatography–mass spectrometry.

Immunoprecipitation of endogenously tagged Setd5 was performed using either mouse embryonic stem cells or neural progenitor cells. In each case 200 million cells were collected and snap-frozen using liquid nitrogen. Nuclei were extracted using a published lysis method [217] and pellets were resuspended in a 420 mM salt buffer. After incubation for 15 min at 4°C, insoluble material was removed by pelleting. The soluble fraction was further diluted to reach a salt concentration of 150 mM. Samples were incubated overnight at 4°C with 15 µL anti-FLAG M2 beads (Sigma). The next day beads were washed three times using a 300 mM wash buffer and sample was eluted via two 1 hour elutions at 4°C with 40 µL FLAG elution buffer (0.25 mg/ml triple-FLAG peptide). Eluted immunoprecipitate was checked by Western blot and sent to the EMBL Heidelberg Proteomics facility for sample preparation using the SP3 protocol. Digested peptides were analyzed by liquid chromatography and either label free or tandem mass tag (TMT)-labeled mass spectrometry.

For the analysis Setd5-knockin samples were compared against matching WT negative controls lacking any FLAG-HA tagged proteins. For each time point samples were processed in duplicates. Peptides from label-free data were identified/mapped using the Isobarquant software [218]. For data from TMT-labeled runs the MaxQuant software (1.5.6.5, [219]) was used. In both cases the resulting raw peptide counts were normalized

through variance stabilization normalization in R (package vsn, 3.44, [220]). Batch effects were removed by fitting a linear model using the limma package (3.3270, [221]) explaining the variance between replicates. Coherence across samples, time points, and replicates was checked by PCR. The limma package was used for differential protein IP analysis using cutoffs of adj. P value < 0.05 and log2(fold change) > 0 to determine significant events. For the visualization log2(fold change) values derived from Setd5-knockin over control comparisons in mESCs was plotted against the same data from NPCs.

Protein-protein interactions (PPI) were analyzed using Cytoscape [222] with STRING database (version 10, [223]) information channels in addition to data from both label-free and TMT-labeled runs in this study. PPI with a combined score ≤ 300 were discarded. For the visualization an interaction score was derived from log2(fold change), -log10(adj. P value) from both label-free and TMT data were scaled each from 0 to 1 and the scaled values summed up. Only proteins with scaled fold change ≥ 0.048 and scaled -log10(adj. P value) ≥ 0.04 were used for the network. Lastly, log2(fold change) and -log10(adj. P value) from ESCs and NPCs were summed together for a combined interaction score.

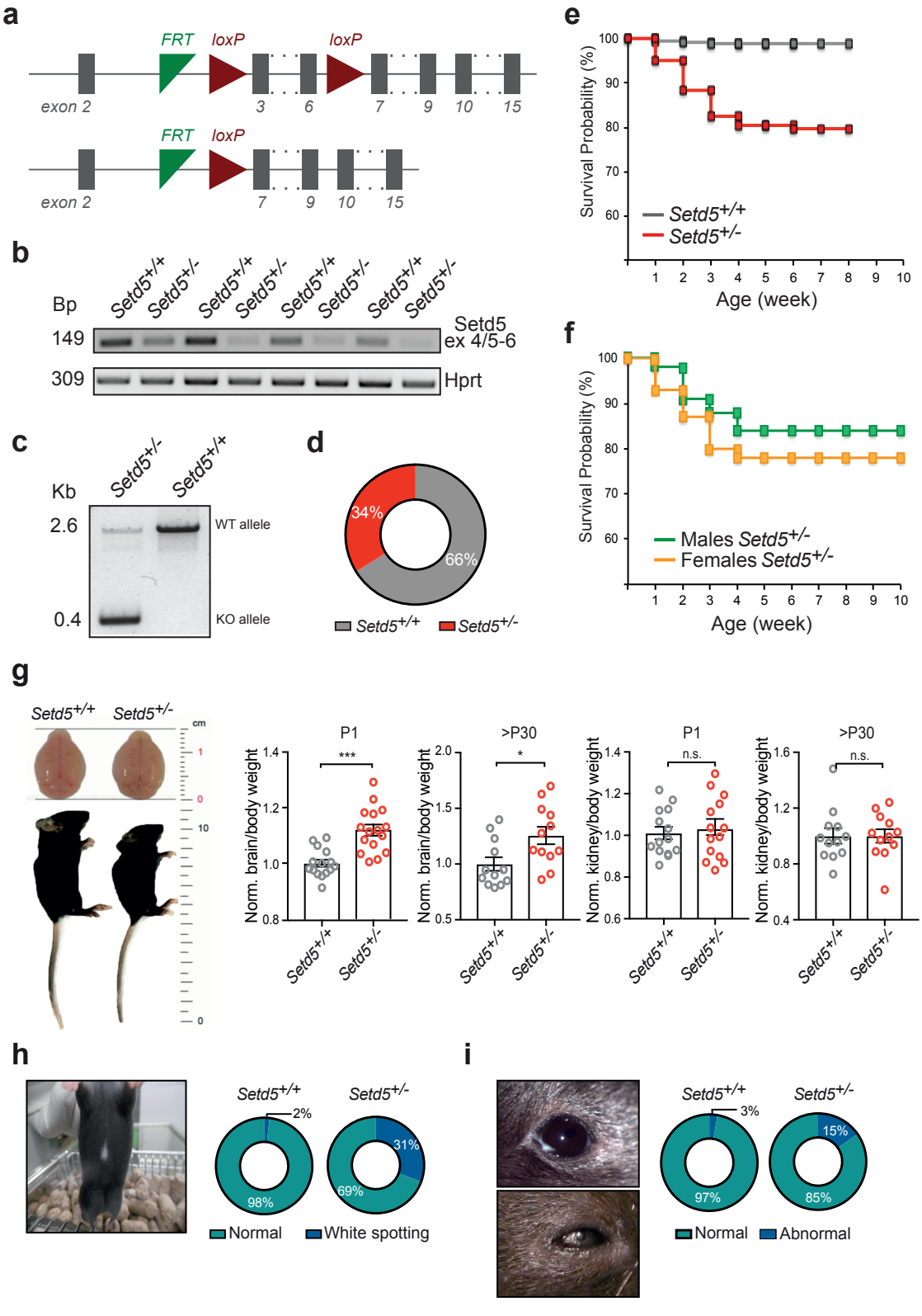
3.4 Results

3.4.1 Setd5 haploinsufficiency leads to early developmental defects and transcriptional dysregulation

In order to study the role of *Setd5* in an *in vivo* context we used a mouse model, *Setd5*$^{+/-}$, where exons 3-6 had been deleted (Fig. S3.1a-c). In contrast to homozygous loss of Setd5, which is embryonically lethal [175], mice heterozygous for the deletion are viable. However, we observed them to be born at a sub-Mendelian rate and they displayed reduced overall survivability (Fig. S3.1d-f). Both newborn and adult *Setd5*$^{+/-}$ mice showed an increase in brain to body weight ratio (Fig. S3.1g) and appeared overall smaller than their *Setd5*$^{+/+}$ counterparts. In addition, craniofacial abnormalities, eye problems as well as white spotting of the belly suggested issues with head and trunk neural crest differentiation and proliferation [224, 225] (Fig. S3.1h-i). These observations indicated that Setd5 might play a role in early developmental processes.

Taking into account the predicted role of SETD5 as a chromatin modifying enzyme prompted us to investigate the transcriptional consequences of its heterozygous loss in early development. For this purpose we performed RNA sequencing on both *Setd5*$^{+/+}$ and *Setd5*$^{+/-}$ whole mouse embryos at 9.5 days of age. Our analysis revealed 487 genes as differentially expressed with 221 and 266 genes upregulated and downregulated in *Setd5*$^{+/-}$ mutants, respectively (Fig. 3.1a). Functional enrichment analysis of these genes showed brain and head development processes preferentially enriched in upregulated genes while terms related to other tissues such as neural crest, limb, skeleton and heart development were enriched in downregulated genes (Fig. 3.1b). We next studied which pathways were likely disrupted by *Setd5* haploinsufficiency and found enrichment for genes involved in Wnt signaling (adjusted P = 0.02) and beta catenin signaling (adj. P = $8 * 10^{-4}$). This finding was supported by a concurrent enrichment for β-catenin and transcription factor 3 (TCF3) target genes in our set of downregulated genes (adj. P = 0.02 and P = 0.003, respectively). Unbiased enrichment analysis revealed two sets of transcription factor with targets enriched exclusively in upregulated or downregulated genes (Fig. 3.1c). Notably, of 12 such factors with downregulated targets 4 are listed in the SFARI database (Nfib, score S; Ctnnb1, 1; Chd7, 1S; Pbx, 3; enrichment P value = 0.05) and an additional two SFARI genes (Ep300, score 1S; Erg, 3) are found for upregulated genes (enrichment P value for 6/21 SFARI genes = 0.03). In addition, DEGs were enriched for genes associated with microphthalmia (such as *Otx2*, *Aldh1a3*, *Stra6*, *Sox2*, adj. P = $4 * 10^{-5}$), one of the dysmorphic features observed in *Setd5*$^{+/-}$ mice as well as patients with *SETD5* mutations.

In order to further delineate differential expression within cell populations from differences in relative cell type abundances we employed *in vitro* differentiation of *Setd5*$^{+/-}$ and *Setd5*$^{+/+}$ mouse embryonic stem cells (mESCs) into neural progenitor cells (NPCs). We investigated gene expression differences at three time points: mESC (d0), embryoid bodies (d4, EBs) and NPCs (d8). We identified 68 (mESC), 242 (EBs) and 548 (NPCs) differentially expressed genes. A total of 67 genes of our embryo DEGs were also found in at least one of those sets (P = $6 * 10^{-15}$, Fig. 3.2b).



Supplementary Figure S3.1: **Generation of *Setd5*$^{+/-}$ mice. a)** Diagram of the *Setd5*-floxed allele (top) with exons 3-6 flanked by loxP sites, and the deletion of these exons after Cre-recombination (bottom). **b)** Representative PCR genotyping results for *Setd5*$^{+/-}$ and *Setd5*$^{+/+}$ mice. **c)** RT-PCR results demonstrating reduction of *Setd5* expression in brain samples (n = 4 mice per genotype); Hypoxanthine

Phosphoribosyltransferase (Hprt) was used as control. **d)** *Setd5*$^{+/-}$ mice were born at a sub-Mendelian ratio (n = 585 mice, 387 *Setd5*$^{+/+}$ and 198 *Setd5*$^{+/-}$). **e-f)** Kaplan-Meier survival curves illustrating the lowered survival probability of *Setd5*$^{+/-}$ mice (n = 387 *Setd5*$^{+/+}$ and 198 *Setd5*$^{+/-}$) which appears to be more noticeable in female animals (n = 98 male / 100 female). **g)** Left: Representative images from female *Setd5*$^{+/+}$ and *Setd5*$^{+/-}$ adult females (>P30) showing comparable brain size but drastically reduced overall body size. Right: Quantification of brain/body weight and kidney/body weight ratios in P1 pups (n = 15) and adult females (>P30, n = 12). Measurements are normalized to *Setd5*$^{+/+}$; * $P < 0.05$, two-tailed Mann-Whitney U test; *** $P < 0.001$, one-way ANOVA and two-tailed t-test; n.s., not significant. Bar plots and error bars show mean ± s.e.m. **h)** White belly spotting as in the representative photo was observed in 31% of *Setd5*$^{+/-}$ mice while almost completely absent in *Setd5*$^{+/+}$ mice. (n = 222 *Setd5*$^{+/+}$ and 157 *Setd5*$^{+/-}$; females and males pooled.) **i)** Eye abnormalities including corectopia, mydriasis and microphthalmia were observed in 15% of *Setd5*$^{+/-}$ but only % of *Setd5*$^{+/+}$ animals. Example images are from our own animals (bottom) and taken with permission from the International Mouse Phenotyping Consortium; Quantifications are based on our own observations. (n = 222 *Setd5*$^{+/+}$ and 157 *Setd5*$^{+/-}$; females and males pooled.) Figure adapted from [167].

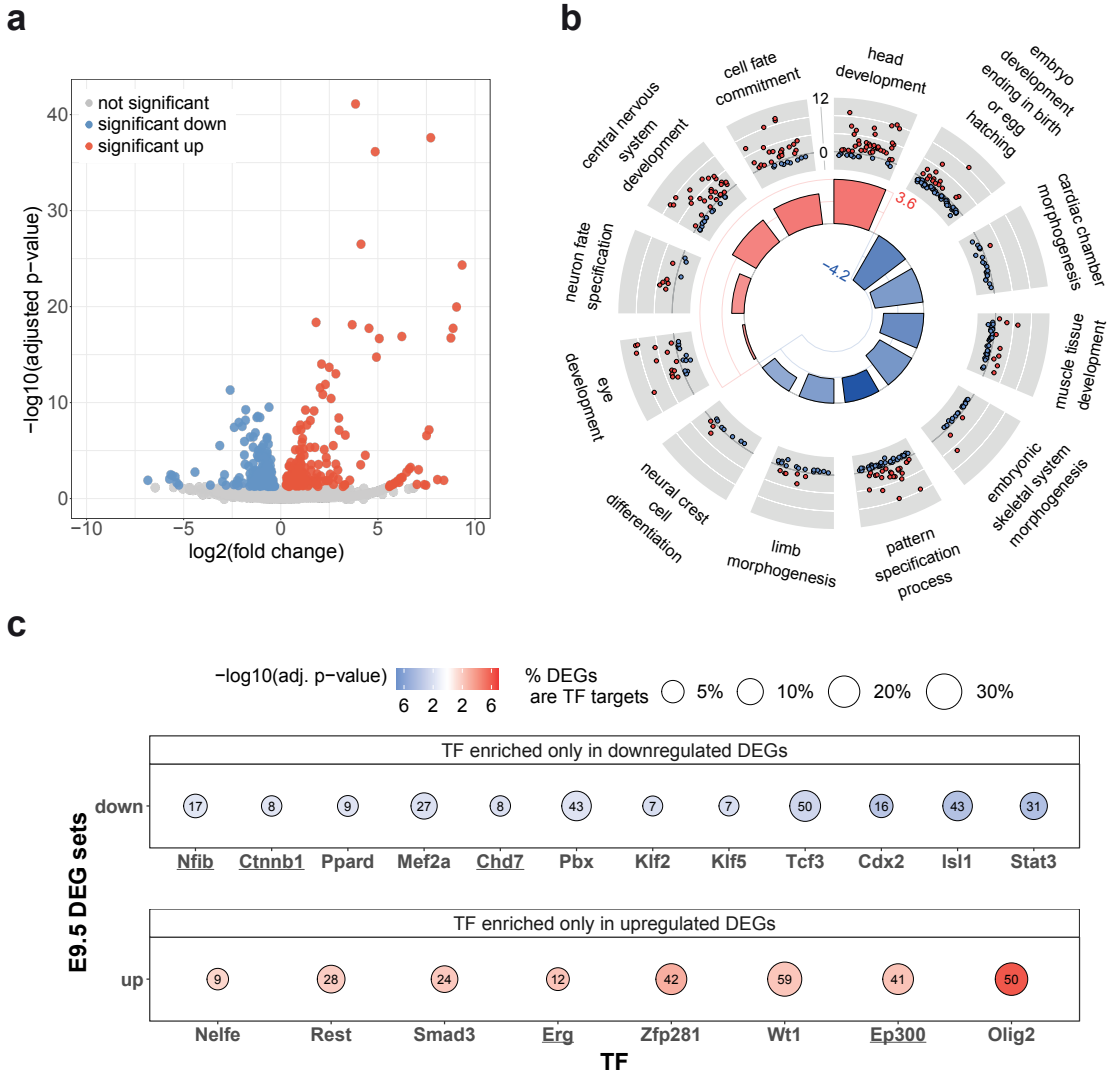


Figure 3.1: ***Setd5*$^{+/-}$ mice show altered gene expression at E9.5. a)** Differential expression analysis results between E9.5 *Setd5*$^{+/-}$ and *Setd5*$^{+/+}$ whole embryo samples (n = 3 embryos per genotype). Significantly changed DEGs are shown in red (upregulated) or blue (downregulated). FDR <= 0.05, likelihood ratio test, edgeR. Non-significant changes in gray. **b)** Selected biological processes from

the Gene Ontology database enriched in our E9.5 DEGs. Inner circle: Enrichment scores for either upregulated (red) or downregulated (blue) DEGs to indicate preferential enrichment in either subset. Numbers show overall maximum enrichment scores and grid lines indicate maxima and half-maxima. Color intensity scales with overall (up+down) enrichment P value with darker colors highlighting lower P values. Outer circle: Expression changes for up- and downregulated DEGs per term are shown as log2(fold changes) (y-axis). Scale is indicated at the top of the plot; White grid lines divide plot into four equal parts; gray line shows zero. **c)** Enrichment analysis for transcription factor targets in DEG sets. Only TF whose targets are exclusively enriched in either downregulated or upregulated genes are shown. Circle size scales with percentage of DEGs in the set of TF targets and color indicates adjusted P value of the enrichment (red and blue for enrichment in upregulated and downregulated genes, respectively). Numbers within circles indicate overlap size. TFs listed in the SFARI database are underlined. Figure partially adapted from [167].

Functional enrichment analysis revealed upregulation of genes related to nervous system development in ESCs and EBs and downregulation of genes associated with other tissues. Furthermore, in line with our embryo findings, EBs and NPCs showed downregulation of Wnt-related genes (Fig. 3.2b). In fact, both enrichment for TCF3 and β-catenin targets are replicated in these datasets (Fig. 3.2c, (adj. P = $3 * 10^{-5}$ and P = $6 * 10^{-4}$ for EBs and NPCs, respectively) and for TCF3 targets (adj. P = $1.8 *^1 0{-}7$ and P = $8 * 10^{-20}$, respectively)). Using the *in vitro* results we could replicate 13 of the 21 unidirectional TF target enrichments in our E9.5 data. In addition, five previously bidirectional enrichments could be resolved in at least one of the three datasets. For example, targets of Ring1b and Cbx2, members of Polycomb repressive complex 1 (PRC1), are only enriched in downregulated genes in NPCs. Taken together, both in E9.5 mice and *in vitro* differentiation *Setd5* haploinsufficiency led to a bias towards neuronal lineage progression at the expense of other lineages, as well as negative modulation of signaling pathways such as Wnt.

3.4.2 Setd5 lacks methyltransferase activity and interacts with NCoR and Paf1 complexes

As a SET-domain containing protein Setd5 shares similarities with other histone methyltransferases, e.g. Setd2 [226]. Despite this similarity we did not find any evidence for methyltransferase activity of Setd5 *in vitro* (Fig. S3.2). Following this we hypothesized that Setd5 might act through interacting with other proteins. Due to the lack of reliable Setd5 antibodies we generated a cell line expressing a FLAG-HA tagged version of Setd5. Using this line we conducted a co-immunoprecipitation experiment followed by quantitative mass spectrometry for samples from mESCs and NPCs (Fig. 3.3a). The results revealed a total of 20 interaction partners of which 6 were found in both mESCs and NPCs, 6 were found only in NPCs and 8 only in mESCs.

Remarkably, all of the core subunits of the Paf1 complex were identified in both. In addition, we also found that several components of the Ncor1-Hdac3 complex interact with Setd5 (Fig. 3.3b). The Paf1 complex, containing Paf1, Leo1, Ctr9, Cdc73, and Wdr61 interacts with RNA polymerase II and has been implicated in regulating transcriptional elongation and polymerase pausing [179, 180, 227]. In addition, the known Paf1 interactor Chd1 [228] as well as a member of the TFIID complex Taf4a [229] were also linked to Setd5. On the other hand, the Ncor1-Hdac3 complex consists of Hdac3, Ncor1, Tbl1x and Tbl1xr1, as well as Gps2 and acts as a transcriptional regulator through histone deacetylation [230, 231].

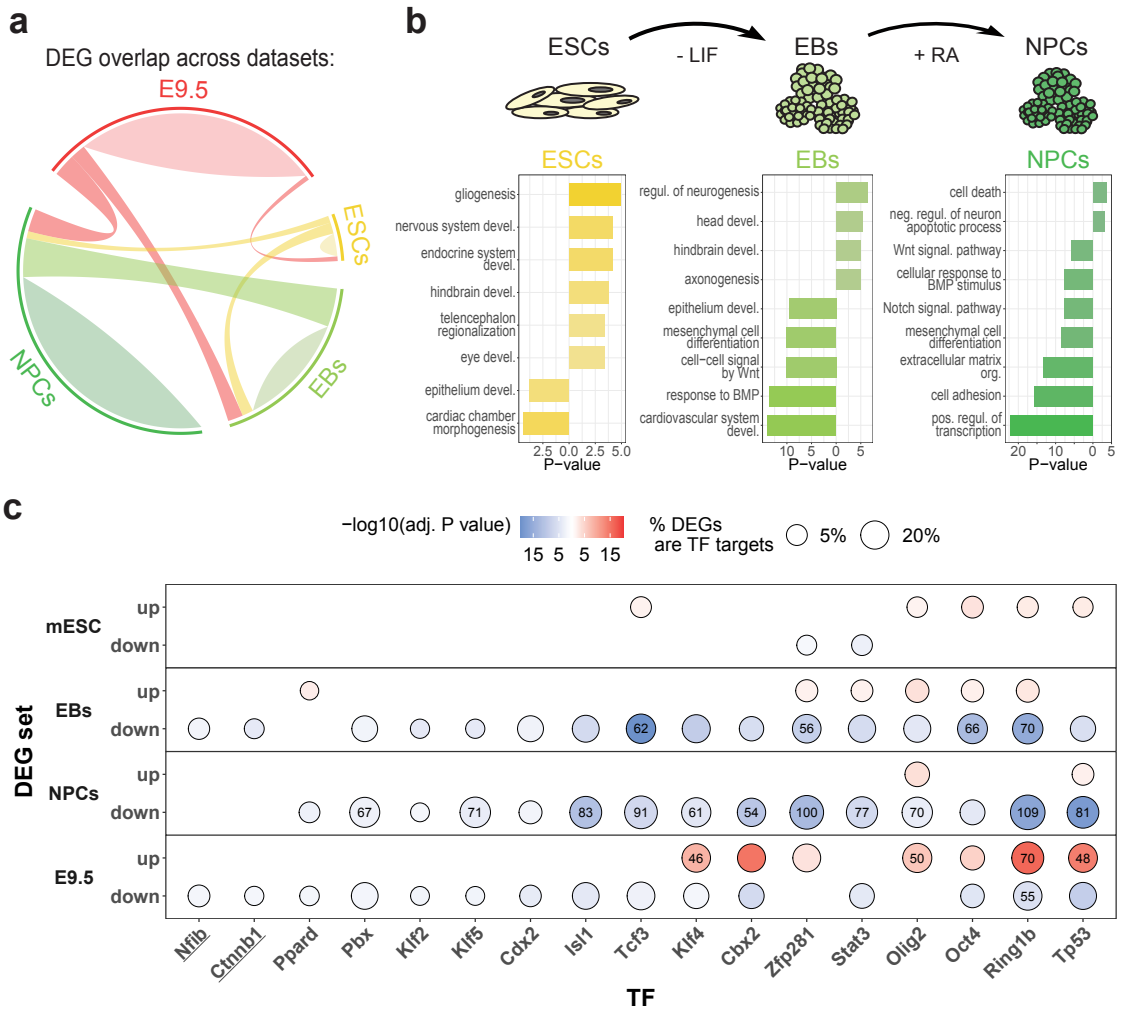
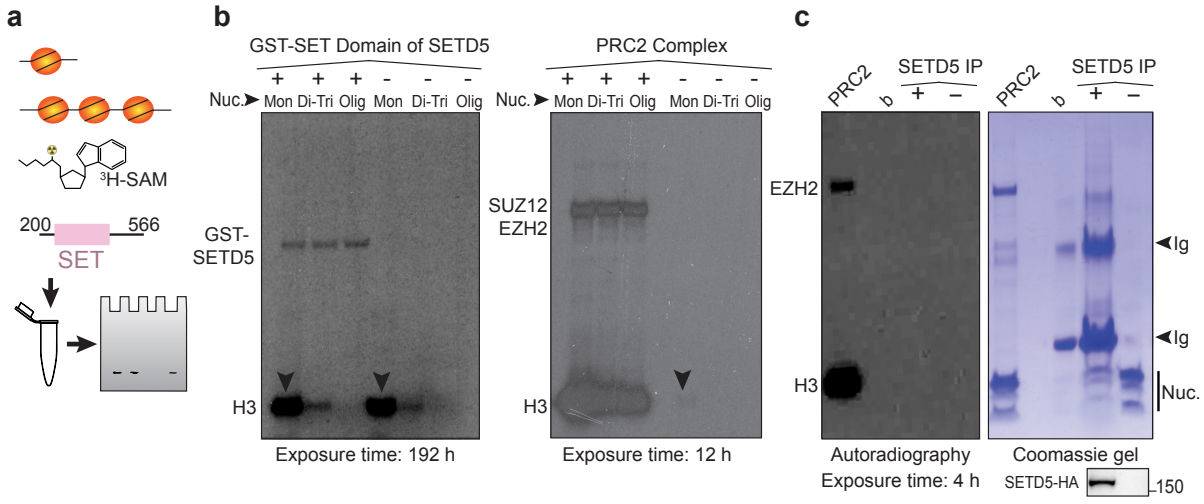


Figure 3.2: **Gene expression changes across *in vitro* development a)** Chord plot showing the overlap between DEGs identified *in vivo* and *in vitro*. Ribbons illustrate overlaps between sets with width scaling with number of overlapping genes. Non-ribbon areas represent DEGs unique for each set. **b)** Top: Schematic of differentiation of mESC to NPCs. Bottom: Gene ontology (GO) terms enriched in DEGs (adjusted P value < 0.05) upregulated (top, bars facing right) and downregulated (bottom, bars facing left) in *Setd5*$^{+/-}$ compared to *Setd5*$^{+/+}$ mESCs, EBs, and NPCs. P values are shown as negative decimal logarithm (one-sided Fisher's exact test). (n = 2 *Setd5*$^{+/+}$, 2 *Setd5*$^{+/+;tr.}$control lines, and 2 *Setd5*$^{+/-}$ cell lines.) **c)** Replication of the enrichment analysis for transcription factor targets in DEG sets. Shown are TFs enriched at E9.5 where either direction-specific enrichment was confirmed or bidirectional enrichment got resolved in the *in vitro* results. Circle size scales with percentage of DEGs in the set of TF targets and color indicates adjusted P value of the enrichment (red and blue for enrichment in upregulated and downregulated genes, respectively). Numbers of overlapping genes are shown where more than 20% of DEGs are TF targets. TFs listed in the SFARI database are underlined. Figure partially adapted from [167].

Of note, both the NCoR complex through Tbl1x/Tbl1xr1 and the Paf1 complex through Cdc73 have been linked with the regulation of Wnt signaling [232–234]. In addition, several members of the NCoR complex, Leo1 of the Paf1 complex and the chromatin modifiers Ankrd11 and Chd1 are all listed in the SFARI database as ASD risk genes.



Supplementary Figure S3.2: **Setd5 does not show methyltransferase activity a)** Schematic outline of the experiment. Mono- or oligonucleosomes purified from HeLa cells were incubated with a GST-tagged SETD5 SET domain and radioactively labeled 3H S-adenosyl methionine (SAM) as substrate. Finally the result is resolved on a gel using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and detected using autoradiography. **b)** Resulting radioactive films for GST-SET domain (left) and PRC2 complex (control, right). Left three lanes show result with tested protein, right three lanes results in its absence. Arrowheads indicate background methylation levels from contaminants in the nucleosome preparation. Note the much longer exposure time for the SET domain film. Experiment was repeated three times with similar results. **c)** Methylation assay using full length FLAG-tagged SETD5 in HEK293T cells. PRC2 was again used as a control. Left: radioactive film. Right: Coomassie staining of SDS-PAGE. Arrows indicate FLAG-bead immunoglobulins (Ig). b, FLAG-beads. Bottom right: Western blot showing presence of FLAG-tagged SETD5. Experiment was performed once. Figure adapted from [167].

We next hypothesized that, if these interactions are functionally important, mutations found in patients with intellectual disability should interfere with the ability of the protein to form these connections. We therefore selected six mutations distributed across the whole gene and generated truncated constructs of human SETD5 (Fig. 3.3c). We then expressed these constructs in HEK293T cells to check overall effects on protein level as well as proper nuclear localization. We found that one mutation (R308*) led to reduced protein levels, one (K399*) failed to enter the nucleus and the remaining four properly located to the nucleus (Fig. 3.3d). Of these four we tested the effect of two (E720* and S1258*) on interactions between SETD5 and members of the NCoR and Paf1 complex. For this purpose we introduced the constructs into a *Setd5*$^{-/-}$ context and tested the interactions via co-immunoprecipitation. We discovered that both mutations virtually abolished interactions with Hdac3 and Tbl1x as well as strongly reduced interactions with Leo1 (Fig. 3.3e).

In summary, these findings suggest that Setd5 acts through interaction with ASD-associated transcriptional regulator complexes NCoR and Paf1. The fact that these interactions are diminished by mutations occurring in patients with intellectual disability further highlights the importance of these links.

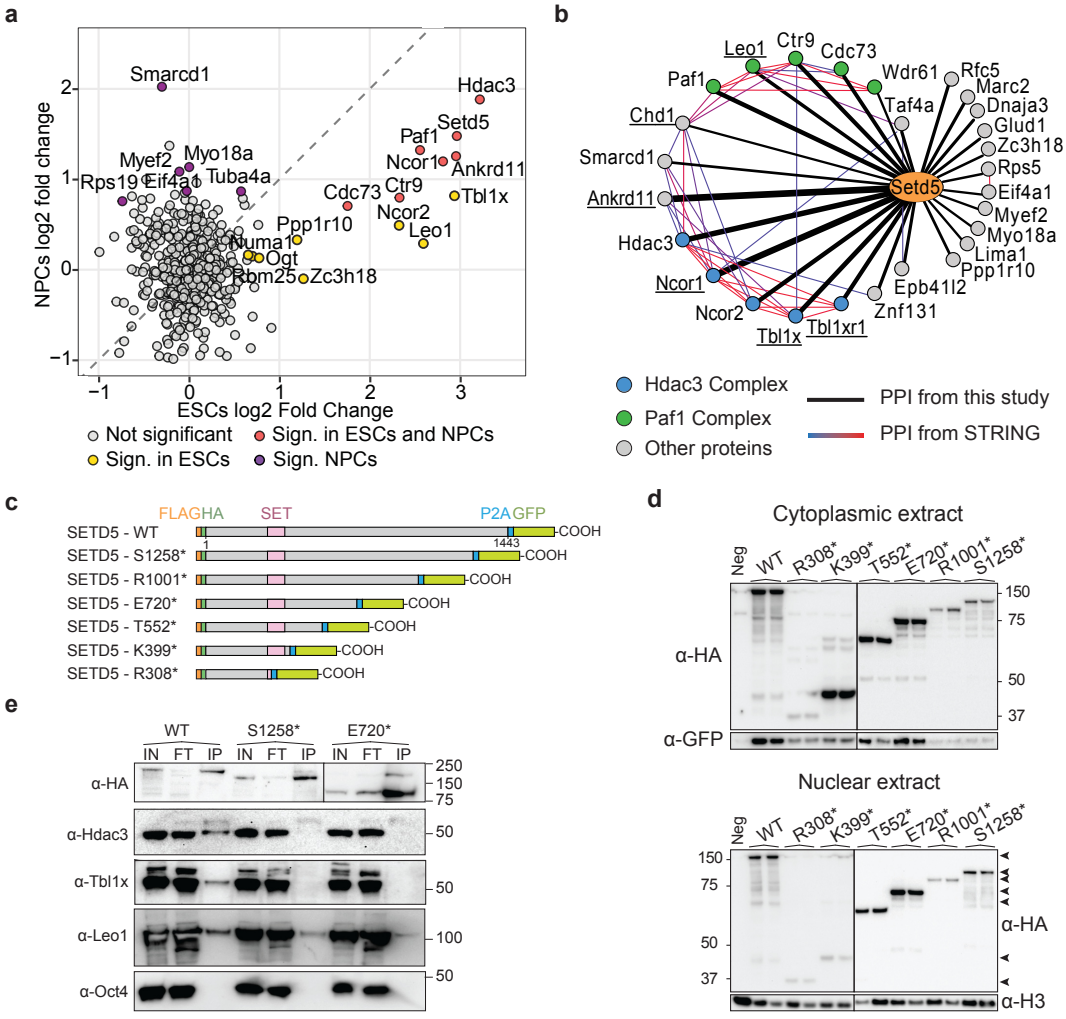


Figure 3.3: **Setd5 interaction with NCoR and Paf1 complex is disrupted by patient mutations.** **a)** Proteins interacting with Setd5 in ESCs and NPCs as identified by quantitative tandem mass tag (TMT) mass spectrometry. Fold changes indicate the comparison between FLAG-tag immunoprecipitation results using tagged Setd5-FLAG-HA and non-tagged WT samples). Color highlights significant enrichment (analysis using limma; adj. P value < 0.05; $\log_2$(fold change) > 0) in both mESC and NPCs (red), only in mESC (yellow) or only in NPCs (purple). n = 2 biological replicates. **b)** Protein interaction network of the Setd5-interactome based on both TMT and label-free mass spectrometry. Interactions identified in this study are shown in black with interaction confidence (see methods) encoded as edge thickness. Published interactions from the STRING database are shown in blue and red for low and high interaction scores, respectively. Highlighted in green are proteins from the Paf1 complex while blue indicates Hdac3-Ncor-Tbl1x complex members. Orange, bait. PPI, protein-protein interaction. Proteins associated with ASD-linked genes are underlined. **c-e)** *SETD5* mutations found in patients disrupt SETD5-interactions with HDAC3 and PAF1. **c)** Schematic of WT and truncated SETD5 constructs. All contain a N-terminal FLAG/HA tag followed by SETD5 cDNA sequence until either endogenous stop codon or stop codons introduced by mutation. All constructs are fused in frame with a P2A sequence and GFP. **d)** Western blot for tagged SETD5 (top) and GFP (bottom) from constructs overexpressed in HEK cells. Two replicates from independent transfections. Samples separated into cytoplasmic and nuclear fractions. Neg, empty vector. Black arrows, expected size for each construct. **e)** Western blot for tagged SETD5 as well as interaction partners for constructs expressed in *Setd5*$^{-/-}$ mESC. Oct4 was used as control. IN, input nuclear extract; FT, flow through; IP, immunoprecipitation with triple-FLAG peptide. 50% IP and 1% IN and FT were loaded. Experiment was repeated three times with similar results. For full-length blots see the supplement of the publication [167]. Figure adapted from [167].

3.4.3 Setd5 regulates Pol II occupancy at primed transcription start sites

Following up on these connections we aimed to gain more insight into changes in the chromatin landscape using chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq). We employed our *Setd5*$^{+/HA}$ cell line to look at genome wide localization of Setd5, with untagged WT cells as background control. Furthermore we used *Setd5*$^{+/-}$, *Setd5*$^{-/-}$, and *Setd5*$^{+/+}$ mESCs to identify changes in occupancy of interaction partners Hdac3 and, via Paf complex, RNA polymerase II (Pol II), as well as changes in histone acetylation using antibodies against histone 3 lysine 27 acetylation (H3K27ac) and histone 4 pan-acetylation (H4ac). We found that Setd5 showed the highest signal at active transcription start sites (TSS) as marked by Pol II occupancy and to a lesser extent at ESC specific enhancers (Fig. 3.4a). Similar to Setd5, Hdac3 binding was detected predominantly at TSS and ESC enhancers (Fig. S3.3a) and colocalized with Setd5 at these loci (Fig. 3.4b). H3K27ac was detected mainly at ESC enhancers (Fig. S3.3b) as well as TSS. For Pol II we used an antibody that detected both unphosphorylated Pol II in the preinitiation complex as well as initiated Ser5 phosphorylated Pol II. We detected the highest signal at TSS but also throughout the gene body and at transcription end sites (TES) (Fig. 3.4c, S3.3c).

Overall Hdac3 occupancy at TSS was only marginally affected by the loss of Setd5 with 21 and 34 loci showing differential binding in *Setd5*$^{+/-}$ and *Setd5*$^{-/-}$ cells, respectively (Fig. S3.3d). This indicated that Setd5 only plays a negligible role in Hdac3 recruitment, in contrast to previous hypotheses [175]. In similar fashion H3K27ac showed no global differences between genotypes (Fig. S3.3b) and only few loci (49 and 103 in *Setd5*$^{+/-}$ and *Setd5*$^{-/-}$ cells) with differential binding (Fig. S3.3e). For histone 4 acetylation we observed a general increase in signal across gene bodies in both *Setd5*$^{+/-}$ and *Setd5*$^{-/-}$ compared to *Setd5*$^{+/+}$ (Fig. S3.3g). Given the role of the Paf1 complex in Pol II pausing we looked into differences in the Pol II pausing index, i.e. the normalized ratio between Pol II density near the promoter compared to the rest of the gene body. We observed 4144 genes with decreased pausing index (44% of genes with Pol II at the TSS) and conversely 2487 genes (26%) with increased pausing index (Fig. 3.4d). In addition and stark contrast to the Hdac3 results Pol II occupancy was significantly altered at 741 TSS in *Setd5*$^{+/-}$ and 593 TSS in *Setd5*$^{-/-}$ cells (Fig. S3.3f). 559 TSS (477 in *Setd5*$^{-/-}$) had increased Pol II occupancy. Furthermore, these TSS showed evidence of Setd5 binding, enrichment for Hdac3 and depletion of H3K27ac and H4ac (Fig. 3.4e). They also showed a shift towards a higher pausing index in both *Setd5*$^{+/-}$ ($P = 1.7 * 10^{-8}$, Kolmogorov-Smirnov test) and *Setd5*$^{-/-}$ cells ($P = 2 * 10^{-4}$). No significant shift was observed for TSS with reduced Pol II signal which presented with H3K27ac and depleted for Hdac3 (Fig. S3.3h). In summary, these results suggested a role for Setd5 in the proper localization of RNA polymerase II but not Hdac3 to transcription start sites.

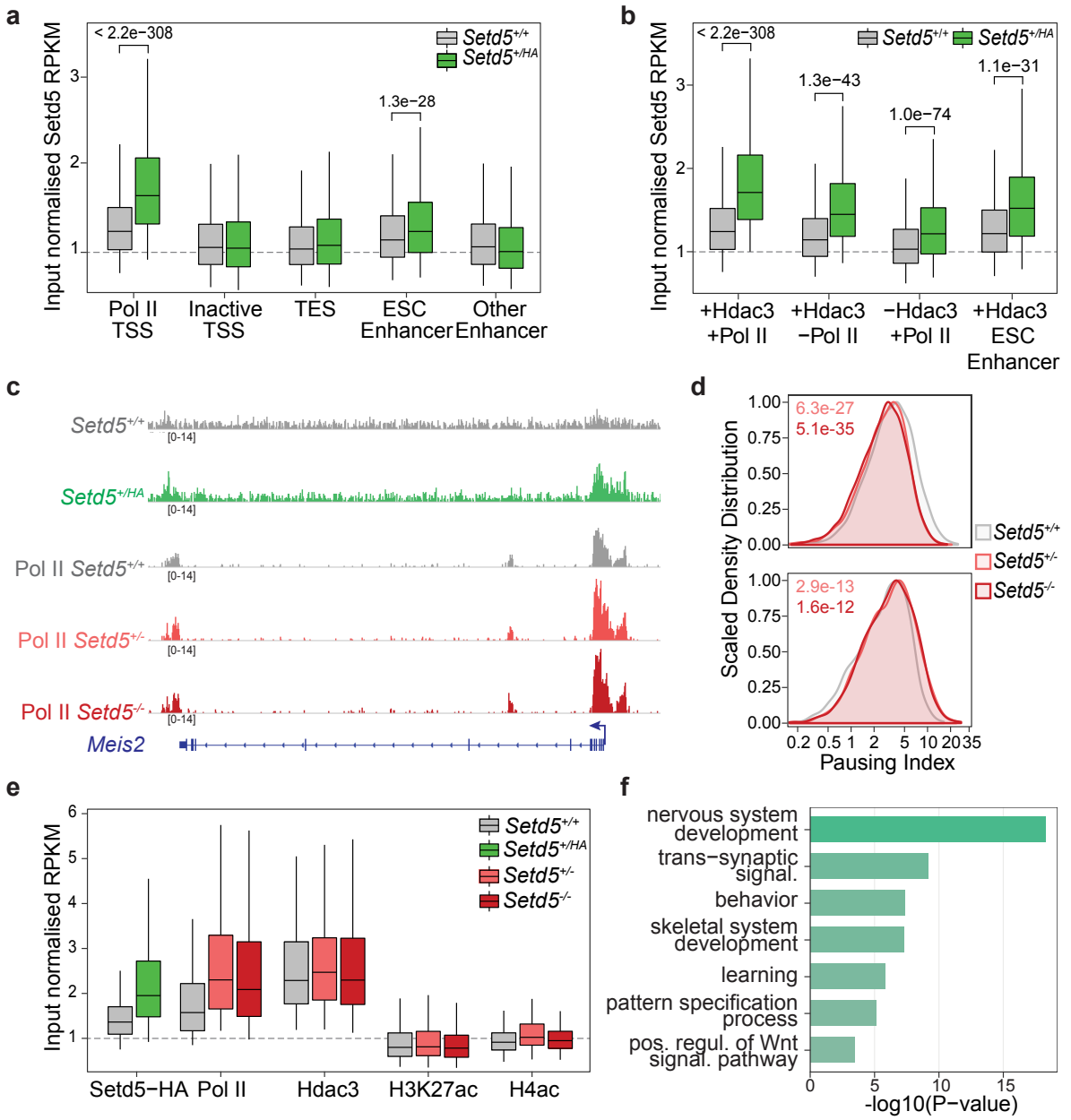
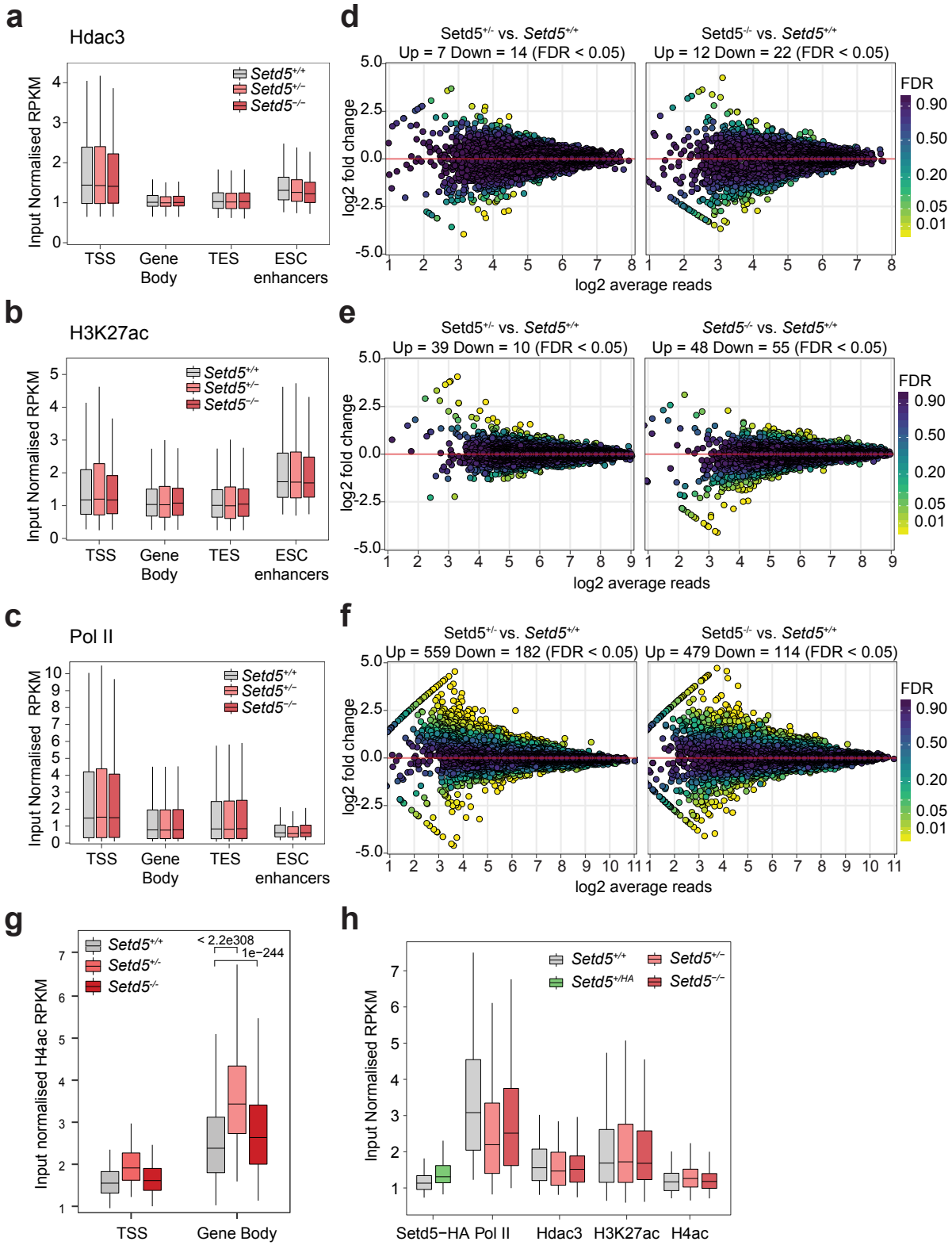


Figure 3.4: **Heterozygous loss of Setd5 in mESC leads to higher Pol II occupancy at the TSS of neurodevelopmental genes a+b)** Setd5 ChIP-seq results using an antibody against the HA tag from a cell line expressing endogenously tagged Setd5 $Setd5^{+/HA}$ and WT cells. One-sided Mann-Whitney U test. RPKM, reads per kilobase per million total reads. **a)** Signal at transcription start sites (TSS) bound (n = 16342 loci, Pol II TSS) or not bound by RNA Polymerase II (n = 17817, Inactive TSS), transcription end sites (n = 34159, TES), ESC specific enhancers (n = 4099), and other enhancers (n = 43657). **b)** Signal at TSS grouped by Hdac3 and Pol II status (Hdac3 and Pol II: n = 13314; Hdac3 only, n = 701; Pol II only, n = 2890) as well as ESC enhancers bound by Hdac3 (n = 688). **c)** Representative tracks of Setd5 ChIP-seq signal in $Setd5^{+/+}$ and $Setd5^{+/HA}$ cells and Pol II ChIP-seq signal in $Setd5^{+/+}$, $Setd5^{+/-}$, and $Setd5^{-/-}$ cells at the *Meis2* locus. Signal from merged replicates is shown with input subtracted and normalized as reads per genomic content. **d)** Distribution of Pol II pausing index for genes with decreased (n = 4144, top) or increased (n = 2487, bottom) pausing index in both $Setd5^{+/-}$ and $Setd5^{-/-}$. x-axis in log10 scale. Kolmogorov-Smirnov test compared to $Setd5^{+/+}$. **e)** ChIP-seq signal for Setd5, Pol II, Hdac3, histone 3 lysine 27 acetylation (H3K27ac), and histone 4 acetylation (H4ac) at genes with significantly higher Pol II occupancy at their TSS (FDR < 0.05, edgeR, n = 559). **f)** GO enrichment analysis for genes in e. Figure adapted from [167].



Supplementary Figure S3.3: **ChIP-seq signal localization and differential binding of Hdac3, Pol II, H3K27ac at the TSS. a-c)** Input-normalized ChIP-seq results in $Setd5^{+/+}$, $Setd5^{+/-}$, and $Setd5^{-/-}$ for Hdac3 (a), H3K27ac (b), and RNA Polymerase II (c) at transcription start sites (n = 34151 loci, TSS) transcription end sites (n = 34159, TES), and ESC specific enhancers (n = 4099). RPKM, reads per

kilobase per million total reads. **d-f** MA plots for differential binding analysis results at TSS for Hdac3 (d), H3K27ac (e), and RNA Polymerase II (f) comparing *Setd5*$^{+/-}$ (left) and *Setd5*$^{-/-}$ (right) to *Setd5*$^{+/+}$ samples. edgeR likelihood ratio test. **g)** Histone 4 acetylation levels in TSS with at least one pan H4ac peak (n = 3018) and across peaks in gene bodies (n = 43448 in 9088 genes). One sided Mann-Whitney U test. **h)** ChIP-seq signal for Setd5, Pol II, Hdac3, histone 3 lysine 27 acetylation (H3K27ac), and histone 4 acetylation (H4ac) at genes with significantly lower Pol II occupancy at their TSS (FDR < 0.05, edgeR, n = 182). Figure adapted from [167].

We next looked at the interplay between chromatin landscape and transcriptional output and intersected our RNA-seq with our ChIP-seq results. We found all DEG sets to be enriched for genes with both Hdac3 and Pol II but no H3K27ac signal at their TSS (mESC: adj. P = 0.04; EBs: $1.6 * 10^{-13}$; NPCs: $1.9 * 10^{-14}$; E9.5: $8.3 * 10^{-22}$). This chromatin configuration suggested a poised state with no active transcription in ESCs which is also reflected in the much stronger enrichment of DEGs from more differentiated samples. To extend our analysis and also include non-expressed genes we revisited the set of 559 TSS with increased Pol II signal. This set is strongly enriched for genes with the same chromatin state ($P = 2.9 * 10^{-97}$) as we found enriched in our DEG sets. Matching our interpretation of non-productive transcription we found very low mRNA levels for these genes in mESC and no concomitant increase of Pol II signal at the gene body. Together these findings were in line with a role of Setd5 in the establishment of a poised transcriptional state in mESC with possible downstream effects throughout differentiation.

At the functional level these 559 genes showed strong enrichment for GO terms related to nervous system development and synaptic signaling, reminiscent of what we observed at the transcriptional level. Wnt signaling pathway genes were also enriched in this set of genes (Fig. 3.4f). This also held when we restricted this set to only genes with a dose-dependent response (higher absolute fold change in *Setd5*$^{-/-}$ than *Setd5*$^{+/-}$) of Pol II (adj. $P = 8 * 10^{-3}$). Of note, while neurodevelopmental genes presented with an overall increase in pausing index (median log2 fold change: 0.53), most genes associated with Wnt signaling in this set had an unchanged pausing index (median log2 fold change: -0.025). In stark contrast with genes with increased Pol II signal, genes with decreased Pol II did not show enrichment for any developmental terms.

In summary our results reveal a role for Setd5 in the establishment of a poised transcriptional chromatin state via the recruitment of Polymerase II to transcription start sites, likely via its interaction with the Paf1 complex. While Hdac3 recruitment itself is not affected by loss of *Setd5* our data suggest that Setd5 preferentially localizes to TSS bound by Hdac3 and might regulate the chromatin state by coordinating Pol II recruitment with Hdac3 mediated deacetylation. Heterozygous or homozygous loss of *Setd5* expression led to global changes in Pol II occupancy with specifically increased binding at the TSS of genes important for neurodevelopment. We hypothesize that this primed them for later expression which might explain the observed upregulation of neurodevelopmental genes in our embryo as well as our in vitro differentiation data.

3.4.4 *Setd5*$^{+/-}$ mice show learning abnormalities and increased long-term potentiation

Following the hypothesis that chromatin modifiers such as Setd5 act as transcriptional modulators in dynamic processes we next focused on behavior in adult mice as a readout

for neuronal function *in vivo*. Given the association of SETD5 with intellectual disability we chose to assess the cognitive abilities with learning tasks.

First, we studied place aversion behavior of group-caged mice. In this paradigm the animals got access to water by nose poking doors placed in front of bottles in four corners of the cage. In the learning phase each mouse is randomly assigned a "wrong" corner in which a nose poke was not rewarded with water but instead resulted in an aversive air puff. We observed that both $Setd5^{+/-}$ and $Setd5^{+/+}$ preferentially visited the correct corners more often than the incorrect one. However, on average $Setd5^{+/-}$ mice administered a higher number of nose pokes per visit, thereby showing signs of increased repetitive behavior [235] (Fig. 3.5a). Intriguingly, the distribution of nose pokes over the course of the whole experiment was markedly different between genotypes. $Setd5^{+/+}$ mice appeared to have learned the task during the first active session and reduced the number of nose pokes to the wrong corner in the second. In contrast, $Setd5^{+/-}$ mice showed deficits in adaptive behavior with the same number of incorrect corner nose pokes as in the first trial (Fig. 3.5b).

Given the problems in adaptation we next aimed to more directly investigate the memory capabilities of $Setd5^{+/-}$ mice. For this purpose we used the novel object location memory test as a hippocampus-dependent test independent of aversive stimuli. We found that, 24 hours after receiving sub-threshold training (3 min), $Setd5^{+/-}$ mice were significantly better at discriminating between novel and old location (Fig. 3.5c). This result was reminiscent of results using mice with a conditional deletion of Hdac3 in the hippocampus which presented with similar increased response to sub-threshold training. It should also be noted that, as we observed, this level of training usually does not produce a memory (discrimination index > 30) in wildtype animals [236].

Last, we used the contextual fear conditioning (CFC) test to study fear memory acquisition and consolidation. We found that $Setd5^{+/-}$ females more strongly memorized the association of context with averse stimulus and showed increased freezing behavior compared to $Setd5^{+/+}$ females upon context exposure 24 hours after acquisition. In addition, while $Setd5^{+/+}$ females largely extinguished this association after an additional 24 hours, $Setd5^{+/-}$ females failed to do so as indicated by no significant difference in freezing behavior between retention and extinction trials (Fig. 3.5d). Of note, both $Setd5^{+/+}$ and $Setd5^{+/-}$ females acquired the memory in the same fashion (Fig. 3.5e). The same test paradigm (3 shocks) in males initially yielded no difference in memory retention with very high % freezing in both genotypes (Fig. 3.5f). In order to circumvent potential dynamic range issues we applied a weaker testing paradigm (1 shock). Under these conditions $Setd5^{+/-}$ males formed a strong fear memory and failed to extinguish it while $Setd5^{+/+}$ males formed only a weak fear memory in the first place (Fig. 3.5g). These findings replicated the increased performance in sub-threshold tasks that we observed in the novel object location test.

In summary, we observed that $Setd5^{+/-}$ mice showed abnormal memory formation and processing. On the one hand, they appeared to more easily retain memory in sub-threshold learning paradigms as indicated by the results of the novel object location test and the weak CFC paradigm. On the other hand, they had trouble in adapting previously formed memories based as demonstrated by the lack of fear memory extinction.

In order to look deeper into the underpinnings of the observed behavioral phenotypes we investigated synaptic plasticity. We used hippocampal slices from $Setd5^{+/+}$ and $Setd5^{+/-}$ littermates to test for differences in long term potentiation at CA3-CA1 synapses.

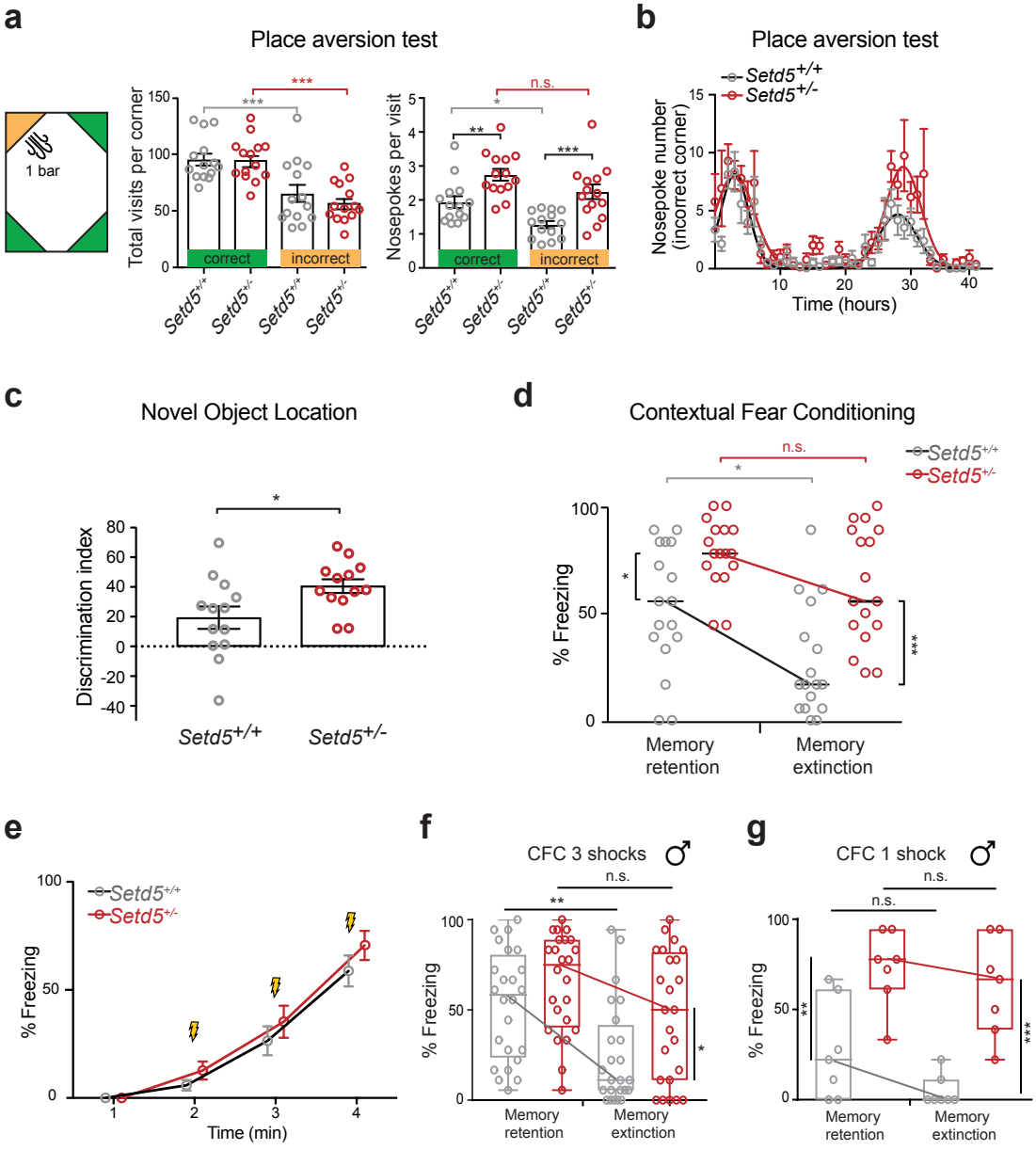


Figure 3.5: ***Setd5*$^{+/-}$ mice show learning defects a+b)** Place avoidance task using the IntelliCage. Trial lasted 48 hours. n = 14 females per genotype. **a)** Left: Schematic of the test paradigm. Nose pokes in three corners (green) were rewarded with water while nose pokes in the wrong corner (red) were punished with an air puff. Middle: Both genotypes preferentially visit correct corners. Right: *Setd5*$^{+/-}$ mice administer more nose pokes per visit. **b)** Distribution of nose pokes in the incorrect corner across whole trial duration. *Setd5*$^{+/+}$ mice appeared to have learned the task during the first active period and reduced the number of nose pokes in the second active period. *Setd5*$^{+/-}$ mice showed no such reduction. Peaks of fitted curves (mean ± s.e.m.) - first 24 h: *Setd5*$^{+/+}$ 7.25 ± 0.6, *Setd5*$^{+/-}$ 8.72 ± 0.8; last 24 h: *Setd5*$^{+/+}$ 4.63 ± 0.4, *Setd5*$^{+/-}$ $8,75 \pm 0.7$. **c)** *Setd5*$^{+/-}$ mice show enhanced memory retention in the sub-threshold training paradigm (3 min) of the novel object location memory test (n = 13 female mice per genotype; data presented as mean ± s.e.m.). **d)** *Setd5*$^{+/-}$ mice display increased fear memory formation and decreased fear memory extinction in the contextual fear condition test (strong paradigm). Y-axis shows percent time freezing during 3 min context exposure. (n = 18 female mice per genotype; lines connect medians) **e)** No difference in freezing behavior was observed during memory acquisition on training day (3 shocks). **f+g)** Contextual fear conditioning test results for male mice using the (f) strong and (g) weak training paradigm. Weak paradigm results replicate the results for females with both stronger memory retention and reduced extinction. n(strong) = 15 males per genotype, n(weak) = 7 males per genotype. Figure adapted from [167].

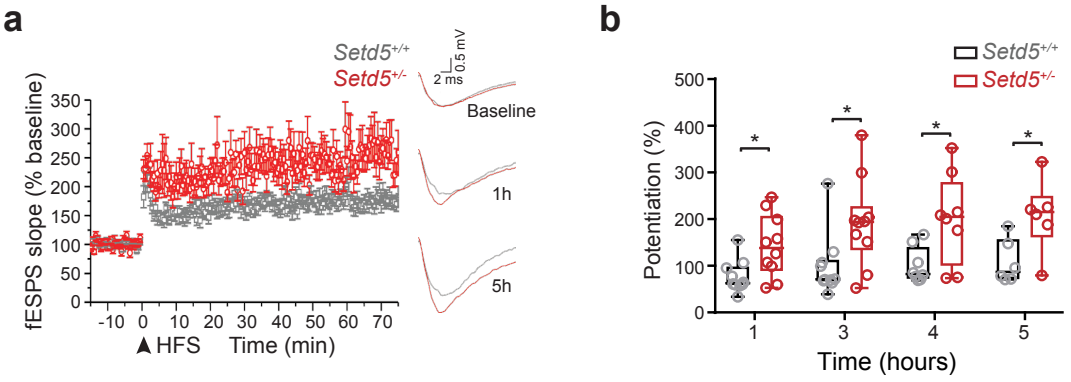


Figure 3.6: ***Setd5*$^{+/-}$ mice present with increased long term potentiation.** **a)** Left: Shown are field excitatory postsynaptic potentials recordings of post-tetanic potentiation and early long term potentiation (LTP) in CA3-CA1 synapses in the stratum radiatum of acute dorsal hippocampal slices (n = 10 male mice per genotype, one slice per mouse, P21-P24). Stimuli were applied every 30 s. LTP was induced via high frequency stimulation (HFS, 4x 100 stimuli of 0.2 ms, 100 Hz, every 5 s) at time 0 (arrow). Data presented as mean ± s.e.m. Right: Representative fEPSP traces at baseline as well as 1 hour (early LTP) and 5 hours (late LTP) after HFS. **b)** Quantification of potentiation at selected time points after HFS (n = 10 (3 h), 8 (4 h), 6 (5 h)). Figure adapted from [167].

We observed that in line with their propensity for strong memory formation synapses of *Setd5*$^{+/-}$ mice showed increased potentiation following high frequency stimulation (Fig. 3.6a-b). Interestingly, this phenotype was also previously observed in a different ASD mouse model [237]. This provided evidence for a cellular substrate and perturbed mechanism connected to the observed learning differences in *Setd5*$^{+/-}$ mice.

3.4.5 Learning abnormalities in *Setd5*$^{+/-}$ mice are paralleled by perturbed transcriptional dynamics

Linking back to the role of Setd5 in transcriptional regulation we next investigated neuronal activity dependent transcription *in vivo*. For this purpose we again employed the CFC paradigm and prepared samples for RNA-sequencing from *Setd5*$^{+/-}$ and *Setd5*$^{+/+}$ animals one and three hours after the initial memory acquisition session. As a control we also included samples from naive homecage animals of both genotypes. We observed 280 (185 upregulated) and 169 (79 up) differentially expressed genes between homecage and 1 hour for *Setd5*$^{+/+}$ and *Setd5*$^{+/-}$, respectively. Contrastingly, for the 1 to 3 hours transition we derived 229 (173 down) and 273 (105 down) DEGs for the two genotypes (Fig. 3.7a). Our differential expression analysis results revealed that we captured the fast induction of immediate early response genes such as *Fos* and *Egr2* [196]. Of note, we did not see mirrored up- and downregulation of such genes to the same extent in *Setd5*$^{+/-}$ samples. This was also indicated by the absence of a bias towards downregulation at the 1h→3h transition observed in wildtype samples (75% of DEG in *Setd5*$^{+/+}$ vs only 38% in *Setd5*$^{+/-}$).

Next we focused in on the differences between the two genotypes via pairwise comparisons at each time point. Surprisingly, we only found four genes - one of them *Setd5* itself - differentially expressed in homecage samples indicating that Setd5 does not play a strong role in the regulation of steady state gene expression in adult animals. In stark contrast, we found 180 and 212 genes differentially expressed at the 1 hour and 3 hour

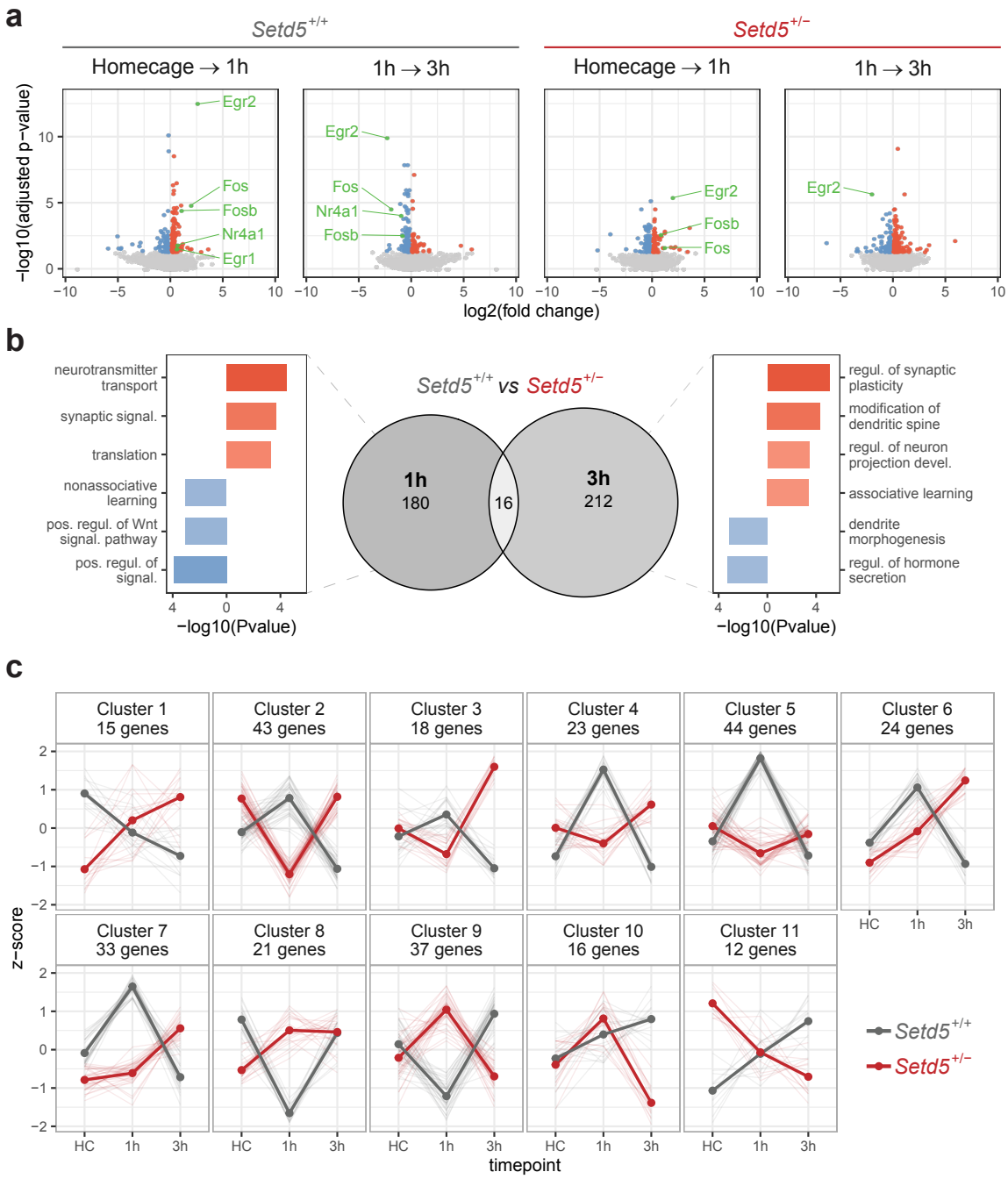


Figure 3.7: ***Setd5*$^{+/-}$ mice present with altered gene expression dynamics after Contextual Fear Conditioning a)** Gene expression changes in *Setd5*$^{+/+}$ (left) and *Setd5*$^{+/-}$ (right) mice one hour after conditioning (compared to homecage baseline) and three hours after conditioning (compared to one hour). n = 5 animals per time point per genotype, except n(*Setd5*$^{+/+}$ 1 hour) = 4 animals. Significantly changed genes are shown in blue and red for down- and upregulation at the later time point, respectively (FDR <= 0.05, likelihood ratio test, edgeR). Immediate early genes are shown in green. **b)** Differential gene expression results between *Setd5*$^{+/-}$ and *Setd5*$^{+/+}$ samples one hour and three hours after conditioning. The central Venn diagram shows size of each DEG set and the overlap between the two sets. Bar plots at either side highlight GO terms enriched in upregulated (red) and downregulated DEGs (blue) at one hour (left) and three hours (right). Bars scale with negative decimal logarithm of the enrichment P value. **c)** Gene expression trajectories for genes with significantly different responses to CFC in *Setd5*$^{+/-}$ compared to *Setd5*$^{+/+}$ (n = 286) clustered by trajectory. Each plot shows mean expression in dots connected by thick lines and data for each gene in a cluster in thin lines. Data normalized as gene-wise z-score. n(*Setd5*$^{+/+}$) = 5 (homecage), 4 (1 hour), 5 (3 hours); n(*Setd5*$^{+/-}$) = 5 (all time points). Figure partially adapted from [167].

time point, respectively (Fig. 3.7b). Functional enrichment of these two sets showed similar results with genes connected with synaptic signaling enriched in upregulated genes. Downregulated genes showed enrichment for fewer and more general terms. Interestingly, despite some agreement at the functional enrichment level only 16 genes were shared between the two sets indicating different temporal trajectories.

To gain a clear picture of differing gene expression dynamics upon CFC training we next determined genes that showed a different response across time points. This revealed 286 genes which we k-means clustered by their genotype specific trajectories into 11 clusters (Fig. 3.7c). We used functional enrichment to annotate these clusters and highlight potentially interesting dynamics. Clusters 4 and 6 were enriched for synaptic genes with cluster 6 showing especially strong enrichment for genes of the postsynaptic density (Fig. 3.8a-b). Of note, the cluster 6 genes *Camk2n1*, *Cpeb1*, *Fxr2*, *Shank1*, and *Synpo* had all been previously linked to neuronal plasticity, memory and intellectual disability [238–242]. Likewise, cluster 4 contained postsynaptic density genes *Lrrc4*, *Srcin1* and *Syngap1* [243–245].

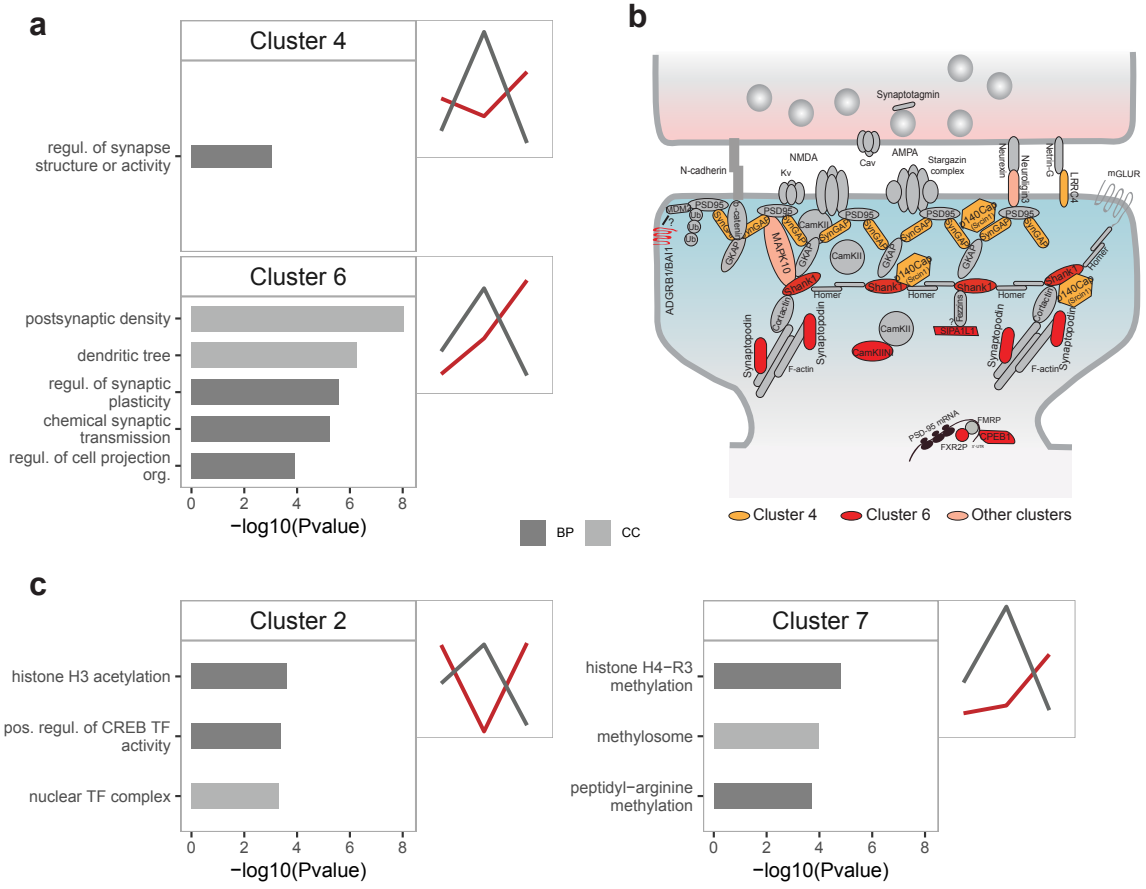
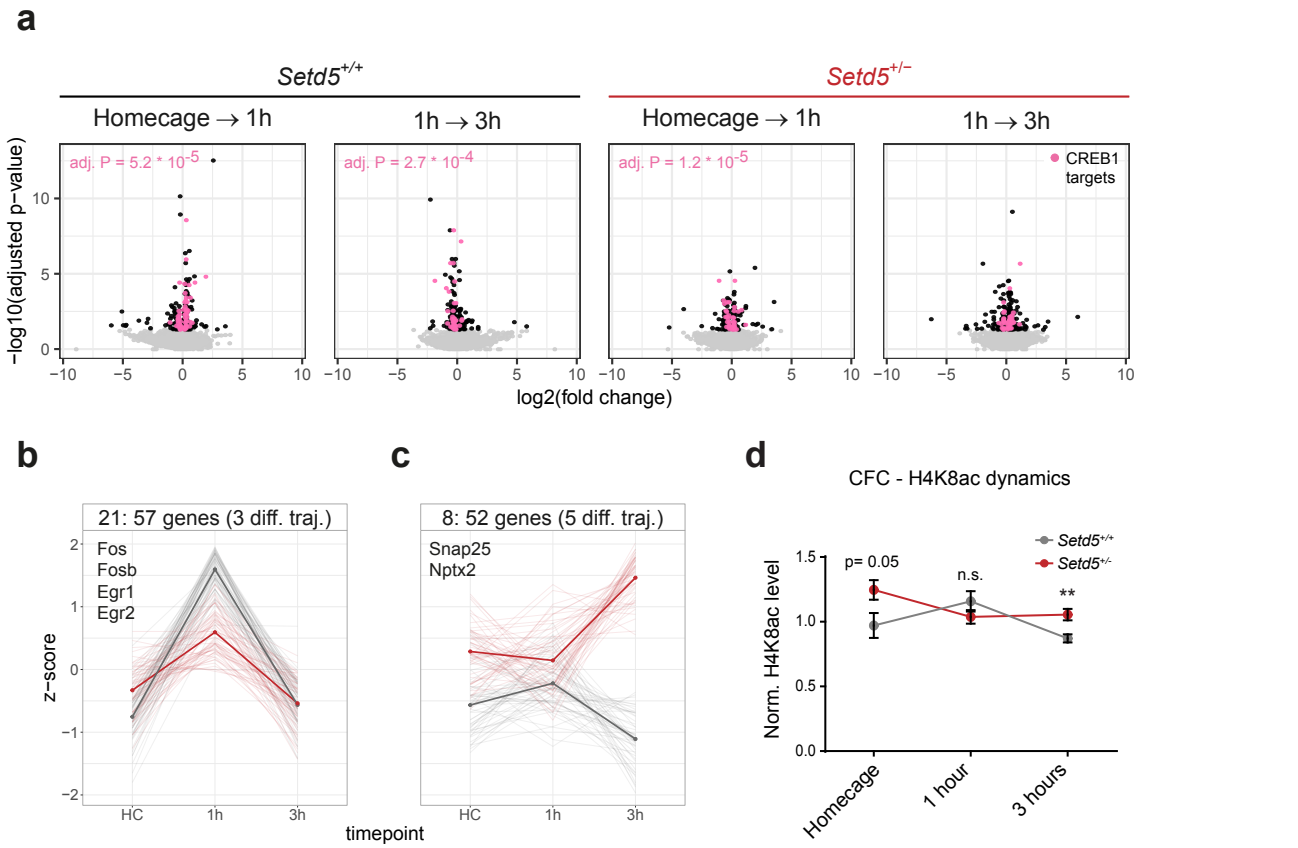


Figure 3.8: ***Setd5*$^{+/-}$ mice show perturbed expression of synaptic genes upon CFC a)** Selected GO term enrichment results for biological processes (BP, dark gray) and cellular components (CC, light gray) demonstrating enrichment for synaptic terms in clusters 4 and 6. Schematics on the right show expression trajectories (*Setd5*$^{+/-}$ in red, *Setd5*$^{+/+}$ in gray). **b)** Illustration of a neuronal synapse highlighting the abundance of cluster 4 and 6 genes in the postsynaptic compartment. **c)** GO term enrichment results for clusters 2 and 7 associated with histone modification and transcriptional regulation. Figure partially adapted from [167].

Both clusters were comprised of genes which showed a clear transient response in wildtype samples (up after 1 hour, down again after 3 hours). In *Setd5*$^{+/-}$ samples however cluster

4 genes only started to go up at the 3 hour time point while cluster 6 genes showed an ongoing rather than transient response. Both clusters shared the characteristic of higher expression in *Setd5*$^{+/-}$ at the 3h time point. In fact, five of the aforementioned genes - *Camk2n1*, *Cpeb1*, *Synpo*, *Lrrc4*, and *Srcin1* - were significantly upregulated at the 3 hour time point in *Setd5*$^{+/-}$ samples. To substantiate our cluster annotation we additionally looked into enrichment for cell type specific genes. This showed a stronger connection of clusters 4 and 6 with excitatory (adj. P 0.06 and 0.006) rather than inhibitory neurons (adj. P > 0.8). These results pointed at perturbed dynamics of postsynaptic gene expression in response to a learning paradigm which might underlie the observed behavioral and electrophysiological differences.



Supplementary Figure S3.4: **Transcriptional response to CFC in *Setd5*$^{+/+}$ and *Setd5*$^{+/-}$ mice.** **a)** CREB target genes in the transcriptional responses after one hour and between the one and three hour time point for *Setd5*$^{+/+}$ and *Setd5*$^{+/-}$. Significantly differentially expressed CREB target genes are shown in pink. Adj. P-value for enrichment of CREB targets in whole (up+down) DEG sets, Fisher's exact test. No such enrichment for *Setd5*$^{+/-}$ 1h $\rightarrow$ 3h. Other DEGs (sign.) are shown in black and other genes are in gray (non sign.). **b-c)** Gene clusters derived from clustering all DEGs by trajectory independent of whether the trajectory differs between genotypes. Number of genes with significantly different trajectories (diff. traj.) in each cluster is listed in the label. **b)**DEG cluster containing the immediate early genes listed. **c)** Cluster of genes with increased expression at the 3 hour time point containing the two genes listed. **d)** Quantification of western blot results for the measurement of histone 4 lysine 8 acetylation in hippocampus CA region samples in the context of contextual fear conditioning. Levels normalized to histone 4 levels, averaged per animal (n = 5 per genotype) over three experiments per time point. For the visualization levels are normalized to average wildtype levels across all time points. Figure adapted from [167].

In addition to altered expression dynamics of synaptic genes we also observed two clusters enriched for terms related to histone modification and transcriptional regulation. Cluster 2, associated with histone acetylation, showed an opposite response in *Setd5*$^{+/-}$ to the up-down in *Setd5*$^{+/+}$ while the histone arginine methylation cluster 7 showed a delayed response pattern. Of further interest was the enrichment for CREB regulators in cluster 2 given that we found CREB target genes enriched in HC→1h DEG sets for both genotypes but only *Setd5*$^{+/+}$ and not *Setd5*$^{+/-}$ 1h→3h DEGs (Fig. S3.4a). Furthermore, histone 4 arginine 3 methylation by the cluster 7 gene product PRMT1 has been linked to histone acetylation via p300/CBP [246, 247].

Lastly, clustering all DEGs independent of trajectory differences resulted in clusters with identical trajectories containing immediate early genes (Fig. S3.4b) but also revealed additional clusters with prolonged expression of synaptic genes such as *Snap25*, a gene with both presynaptic and postsynaptic functions [248] (Fig. S3.4c).

It should also be noted that we found evidence for altered histone acetylation dynamics in the same behavioral context (Fig. S3.4d). Specifically, we observed higher (fold change = 1.3) and significantly higher (fold change = 1.2) histone 4 lysine 8 acetylation at home cage baseline and 3 hours post conditioning, respectively.

In summary, our results highlight an important role for Setd5 in the dynamic regulation of genes involved in synaptic plasticity and structure in the hippocampus in response to fear memory formation. Of note is the prolonged expression of several of these genes at the 3 hour time point where their levels had already gone back to baseline in wildtype mice.

3.5 Discussion

In this study we chose the intellectual disability gene *Setd5* as a starting point to investigate the role of a proposed chromatin modifier in neurodevelopment and neuronal plasticity. We chose these two broad areas under the hypothesis that the transcriptional modulation exerted by chromatin modifying enzymes is likely most important in the fine-tuning of dynamic processes. We employed a heterozygous mouse model in our study as this mimics the heterozygous *SETD5* mutations found in patients [173, 174, 249].

Setd5 haploinsufficiency impacts early lineage specification

Homozygous loss of *Setd5* is embryonically lethal in mice [175] which was our first hint at an early developmental role of *Setd5*. This was further substantiated by our observation of developmental defects in *Setd5*$^{+/-}$ mice such as drastically reduced body size and several craniofacial abnormalities.

Based on these findings we first investigated the transcriptional landscape in early mouse embryo development using RNA-seq at embryonic day 9.5. We found a strong propensity for the expression of neurodevelopmental genes with a concurrent decrease in genes for other lineages. Given that our results are based on bulk whole embryo samples this likely at least in part reflects a shift in relative lineage proportions. This finding together with the observed body size differences raises the possibility that Setd5 is important for the fine tuning of lineage determination in early development. Specifically, we found that Setd5 plays an important role in preventing aberrant expression of neuronal lineage genes. Our results expand earlier findings of perturbed expression of lineage specifiers in homozygous knockout mESC [175] and downregulation of primordial germ cell related genes in *Setd5* knockdown mESC [95]. Our analysis consistently highlighted perturbed Wnt signaling in both *in vivo* and *in vitro* experiments. This is of particular interest given that this pathway plays an essential role throughout neurodevelopment starting with the specification of the neural plate where it exerts a posteriorizing effect and its inhibition is required for the formation of anterior neuroectoderm (reviewed in [250, 251]). As drastic examples, strong reduction of β-catenin levels in early mouse development leads to severe underdevelopment of structures posterior to the thalamic region [252] while ectopic activation of Wnt signaling causes an absence of anterior neural structures [253]. Further studies will be needed to clarify the importance of this pathway in perturbed lineage specification induced by *Setd5* haploinsufficiency.

Delving deeper into the function of Setd5 at the molecular level we found that Setd5 exhibits no methyltransferase activity, contrary to predictions based on the presence of a SET domain in the protein, a common feature of many methyltransferase enzymes [254]. This lack of activity had also been suggested by earlier ChIP-seq results for histone methylation marks upon *Setd5* knockout [175]. Furthermore, both yeast and Drosophila orthologs of Setd5 were both shown to have no methyltransferase activity [255, 256] and KMT2E, a paralog of SETD5 also failed to methylate histones [257]. In the time since our manuscript was published this point was contested by a study in 2019 [226] which reported Setd5 as a histone 3 lysine 36 methyltransferase. This result was however not confirmed by a later study which again found no direct methyltransferase activity of the purified Setd5 protein. They did find evidence for physical interaction of SETD5 with the known histone 3 lysine 9 methyltransferases G9a and GLP and provided the first, albeit indirect, connection of SETD5 and histone methylation [258]. To date the evidence is

therefore in favor of Setd5 acting through its interaction with other enzymes rather than via its own enzymatic activity.

Setd5 haploinsufficiency interacts with the NCoR and Paf1 complexes in a neurodevelopmental context

We probed the Setd5 interactome in mESC and NPCs and found interactions with two gene regulatory complexes, the Polymerase associated factor 1 complex (Paf1c) and the Nuclear receptor co-repressor/Histone deacetylase 3 (NCoR/Hdac3) complex. Our results confirm and expand previous reports of these interactions in HEK293T and mES cells [95, 175] by revealing additional interaction partners, such as Wdr61 for the Paf1 complex and Ncor2 a member of the silencing mediator of retinoic acid and thyroid hormone receptor (SMRT) corepressor complex closely related to NCoR as well as confirming these interactions in a cell type en route to neurogenesis.

In light of the observed perturbation of early cell fate determination it is of particular interest that both these complexes have previously been linked to the regulation of Wnt signaling, a pathway we found altered in almost all assayed samples. NCoR subunits Tbl1x and Tbl1xr1 were both shown to be required for mediating the transcriptional activation of Wnt effector genes by β-catenin [232, 259] and a mutation in Tbl1xr1 linked to schizophrenia was associated with perturbed Wnt signaling activity [260]. Furthermore, the ability of MeCP2 to recruit the NCoR complex to DNA has been highlighted as one of the protein's key functions the loss of which leads to Rett syndrome [261]. For a review of the numerous connections between NCoR and intellectual disability/ASD see Kong et al. [186].

The Paf1 complex on the other hand is connected to Wnt signaling via its parafibromin (*Cdc73*) subunit which was shown to interact with β-catenin and positively regulate the expression of Wnt effector genes [233]. In fact, parafibromin has been demonstrated to play a role in additional developmental pathways such as Hedgehog and Notch signaling with a proposed role as an integrator of multiple morphogen signals [234]. Taken together these links provide evidence that the observed perturbations we found in $Setd5^{+/-}$ mice are possibly mediated by its interaction with one or both of these regulatory complexes in the context of developmental signal regulation.

In addition, we report interactions with other chromatin modifiers such as Ankrd11, Chd1, and Smarcd1 - two of which have been linked with ASD [262, 263] - which could potentially add additional dimensions to the role of Setd5 in chromatin regulation.

Setd5 haploinsufficiency impacts recruitment of RNA Polymerase II at paused genes

We performed ChIP-seq experiments to first characterize Setd5 itself and found it to be preferentially located at either transcription start sites or, to a lesser extent, enhancers. We performed additional experiments to further investigate the collaboration of Setd5 with the two aforementioned regulatory complexes and potentially highlight one over the other. Indeed, while found very little changes in genomic occupancy of Hdac3, member of the NCoR complex, in either $Setd5^{+/-}$ or $Setd5^{-/-}$ cells. This result indicates that Setd5 does not play a substantial role in the recruitment of this complex to target loci. We did however find evidence of Setd5 binding at loci bound by Hdac3, even when Pol II was not present. This opens up the possibility that the reverse might be true and Setd5 might

be recruited to specific loci via its interaction with the NCoR complex. Interestingly, another study since then has investigated the connection between Setd5, Hdac3 and histone acetylation in the context of pancreatic cancer and found that Hdac3 in complex with NCoR/Setd5 has increased substrate specificity (only H3K9 and H3K27) compared to Hdac3 alone [258]. This adds an additional facet to this interaction although it remains to be shown whether it is in fact Setd5 or other components of the NCoR complex which impart this specificity. In addition we also found little overall changes in histone 3 lysine 27 acetylation but did not check histone 3 lysine 9 methylation which was highlighted as the main target in the aforementioned study [258]. We observed changes in histone 4 acetylation in gene bodies but not TSS. Given the absence of Hdac3 from these regions we speculate that these changes are Hdac3 independent.

In contrast to Hdac3, we found substantial reorganization of RNA Polymerase II, which is regulated by Paf1c in numerous ways throughout transcriptional initiation, elongation, and termination [264]. Initially identified as a factor connected to transcriptional elongation it has since then been implicated in promoter-proximal pausing of Pol II, which is defined as the binding and stalling of Pol II shortly downstream of the promoter without further transcription (reviewed in [265]). In our study we identified a subset of 557 genes with increased Pol II binding at their transcription start site. Genes in this set were also bound by Setd5 and Hdac3.and were overall enriched for genes involved in neurodevelopment and, to a lesser extent, genes related to Wnt signaling. We found that the majority of these genes only showed very low levels of expression and in general exhibited a paused Pol II state in wildtype cells. This was further shifted towards higher pausing index in $Setd5^{+/-}$ and $Setd5^{-/-}$ cells. This paints a suggestive picture of "biasing" these genes for transcription which would go in line with the observed propensity towards neural differentiation. However, further experiments will be needed to test this hypothesis as the role of Pol II pausing in mammalian development is still subject to investigation [266]. It should be noted that genes related to Wnt signaling did not show this shift towards more pausing and might therefore be exempt from the creation of a biased state. This might explain why this pathway is rather downregulated later. While we can only speculate on any changes in occupancy of members of the Paf1 complex we think it a reasonable assumption that changes in RNA Polymerase II are at least in part mediated by this interaction.

It should be acknowledged that, given the complexity of gene regulation by these two complexes, we have only begun to scratch the surface of their connection with Setd5. Further experiments also targeting specific Paf1c subunits are needed to clarify this point. In our study we primarily focused on changes at transcriptional start sites and left enhancers largely unexplored. In my opinion this would be a direction worth pursuing further, also given a more recent study linking Setd5 and the NCoR-Hdac3 complex with enhancer regulation during adipogenesis [267]. Given the dynamic nature of both Pol II pausing as well as histone modifications, especially poised bivalent states, throughout development it would be of great interest to characterize these factors throughout neurodevelopment to get a more complete picture of the developmental impact of *Setd5* haploinsufficiency.

Setd5 haploinsufficiency leads to learning abnormalities and enhanced LTP potentially linked to histone deacetylation.

The second part of this study focused on the role of Setd5 in neuronal function. Through the use of a variety of behavioral tests, of which I present here only those relevant for the

scope of this thesis, we established that *Setd5*$^{+/-}$ mice display abnormal learning behaviors. On the one hand they have an increased ability to form and retain memory in tasks such as novel object location and contextual fear conditioning, even when sub-threshold learning paradigms are used. On the other hand, this is accompanied by problems in adaptation as the mice do not properly extinguish fear memory upon exposure to the now "safe" context. These results hint at an inflexibility which is also in line with the increase in repetitive behavior we observed in the place aversion test.

Placing our behavioral results in the context of Setd5 interactions at the protein level strongly highlights histone deacetylation as a likely mechanistic angle. The connection of HDACs to memory formation has been demonstrated in multiple studies employing HDAC inhibitors [268–272] and has given rise to the "molecular brake pad" hypothesis [273]. In short, under this paradigm HDACs act constitutively to keep certain genes in a silent state which is only lifted by sufficiently strong signaling, i.e. through neuronal activity. Another prediction of this model is that histone deacetylases are important to temporally restrict responses by applying the "brakes" to reset the permissive state back to the repressive baseline. This was demonstrated in a study from 2007 where treatment with HDAC inhibitor trichostatin A (TSA) was shown to convert the response to a weak stimulus from a transient transcription-independent form of LTP to prolonged long term potentiation dependent on CREB-mediated transcription. Furthermore, they reported prolonged expression of the immediate early genes *Nr4a1* and *Nr4a2* in response to TSA treatment in the context of contextual fear conditioning [270].

HDAC3 itself has also been identified as an important negative regulator of long-term memory formation [236, 274–276]. Reminiscent of our results in *Setd5*$^{+/-}$ mice, conditional deletion of Hdac3 in the dorsal hippocampus of adult mice resulted in increased performance in the novel object recognition task, especially in a sub-threshold training paradigm [236]. Regarding the effects of Hdac3 inhibition on fear memory formation a study found that infusion with a dominant negative mutant Hdac3 (Y298H) into the dorsal hippocampus led to increased freezing indicative of enhanced fear memory formation [275]. To my knowledge no study has yet assessed the impact of Hdac3 deletion on fear memory extinction. However, treatment with HDAC inhibitors just before re-exposure to the now safe context has been shown to lead to increased extinction during the test the next day [269]. It should be noted that, in the context of cued fear conditioning a study observed no such effect on memory extinction when using an Hdac3 specific inhibitor [277].

In stark contrast with the role of Hdac3 as a negative regulator of memory formation another study employing conditional deletion of *Hdac3* restricted to excitatory forebrain neurons (Camk2a-Cre) reported impaired performance in the novel object location test, albeit with only 90 minutes instead of 24 hours waiting time before moving one of the objects and stronger training [205]. Furthermore, they also report reduced rather than increased freezing behavior in the contextual fear conditioning task. Considering the differences in *Hdac3* deletion timing these contrary results might indicate that while Hdac3 is a negative regulator of memory formation in fully developed animals its lack during critical developmental periods earlier might overrule this effect and lead to more generalized cognitive impairments. Considering our results, it is therefore plausible to assume a model where the modulation of Hdac3 by Setd5 is more important in an adult than a developmental context. This hypothesis is however disputed by the results of a study published shortly after ours in which the authors report impaired performance of *Setd5*$^{+/-}$ animals in the novel object location test after 24 hours with a stronger learning

paradigm (5 min) [226]. Future studies will be needed to resolve this discrepancy and produce a clearer picture of behavioral changes in *Setd5*$^{+/-}$ mice and their potential connection with Hdac3 function.

Lastly, an interesting aspect that warrants further exploration is the possibility that synaptic phenotypes observed in *Setd5*$^{+/-}$ mice might be connected with the suggested impairment in Wnt signaling we observed. Besides its well studied developmental roles the Wnt pathway also plays a role in synapse formation and plasticity (reviewed in [278, 279]). For example, postsynaptic Shank proteins have been shown to regulate synapse development through the modulation of Wnt signaling [280]. While this connection is purely speculative at this point I believe it further highlights that the modulation of Wnt signaling through Setd5 is a future direction worth pursuing.

Setd5 haploinsufficiency causes an altered transcriptional response during learning behavior

Using RNA sequencing following contextual fear conditioning we found evidence of an altered response to this learning paradigm at the transcriptional level. While we observed expected responses such as the induction of immediate early genes and CREB dependent transcription in general, we also find distinct differences between *Setd5*$^{+/-}$ and wildtype mice. Specifically our results indicate a less transient response with the upregulation observed at 1h not mirrored by a concurrent downregulation of 3h as in the wildtype. We further substantiate this hypothesis with an analysis of altered temporal expression trajectories. Of particular interest is the prolonged upregulation of a cluster of synaptic genes, such as *Shank1*, which only show a transient response in wildtype mice. In addition we found upregulation at the three hour time point of the genes *Snap25* and *Nptx2*, two genes both linked to hippocampus dependent memory formation [281–283].

It should be acknowledged that this experiment came with significant caveats. First, we used a 3' library preparation protocol which means that we cannot investigate isoform differences, which for example are observed for *Snap25* [281]. Second, we prepared RNA from the whole CA-region of the hippocampus but only a small number of neurons actually respond to the behavioral test (compare Nr4a1/2 expression analysis in [270]). There's also further possibilities for confounding factors in the precision of the dissection, for example due to contamination with varying amounts of dentate gyrus tissue. We chose to tackle these issues at the computational level using removal of unwanted variation (RUV)-seq, a method published for use in exactly this context [192]. To gain a more detailed picture of the altered transcriptional response in *Setd5*$^{+/-}$ neurons upon activity further studies will be needed, employing either a reduced *in vitro* system or a combination of *in vivo* behavior with RNA sequencing at single-cell resolution.

Despite these shortcomings our results detailing an altered transcriptional response to neuronal activity in a learning paradigm point at an intriguing avenue of study to pursue further. The modulation of activity-dependent transcription is a possible link between chromatin regulators and their dysfunction and higher order differences observed at the level of neuronal function and behavior. Given the abundance of genes encoding for such regulators connected with neuropsychiatric disorders such as ASD or intellectual disability, it is not unlikely that this could present a point of convergence in their pathogenesis.

4

CHD8 haploinsufficiency links autism to transient alterations in excitatory and inhibitory trajectories.

4.1 Context

For this study we chose to model heterozygous loss of the autism spectrum disorder (ASD) gene *CHD8* in a human model system of neurodevelopment, cerebral organoids [131]. This choice was influenced by the fact that mouse models of CHD8 mostly reported mild phenotypes that largely did not replicate across studies ([284–292], reviewed in [293]). We argue that this might be in part due to differences between human and mouse development and that a human model system is needed to gain further insight.

My part in this project focused on the side of transcriptomic experiments and analysis. I performed and analyzed the bulk RNA-sequencing experiment in young organoids as well as contributed in the analysis, visualization and interpretation of the single cell RNA-sequencing experiment. This work was published in 2022 (Villa et al. [168]) and the figures presented here were adapted from this publication.

4.2 Introduction

Chromodomain Helicase DNA Binding Protein 8 (*CHD8*), is one of the genes most strongly associated with ASD [55]. Patients with *de novo* mutations in this gene often present with additional symptoms such as intellectual disability, enlarged head size (macrocephaly) as well as gastrointestinal problems [74, 294].

CHD8 belongs to the CHD family of ATP-dependent chromatin remodelers characterized by two chromodomains upstream of their catalytic Snf2 helicase domain where it is part of the CHD7 subfamily [120, 295]. These factors play important roles in transcriptional regulation through their ability to control DNA accessibility (reviewed in [296]). Previous studies have proposed a role for CHD8 in multiple contexts. It was first identified, then named Duplin, as an interactor of beta-catenin, the main component of canonical Wnt-signaling. It was shown that this interaction inhibits signaling via beta-catenin through blocking it from interacting and thereby activating Tcf/Lef [119]. The connection between CHD8 and Wnt-signaling has since then been detailed further by showing association of CHD8 with the promoter regions of beta-catenin target genes [120] as well as indicating that CHD8 mediated repression works via recruitment of histone H1 [121], a mechanism also proposed for its regulation of apoptosis via p53 regulation [122]. In spite of these results, a more recent study of CHD8 in the context of murine neurodevelopment reported that Chd8 acts as a positive regulator of Wnt signaling. They showed that its down-regulation leads to increased Tcf/Lef activity in neural progenitor cells [123]. To further complicate the picture the results of a study on heterozygous loss of *CHD8* in human cerebral organoids aligned with the previous findings of CHD8 as a negative regulator of Wnt signaling, indicating potential species specific differences in this context [297].

In the context of ASD, CHD8 has been studied largely through the use of mouse models [284–292]. Phenotypes varied across studies with increase in head size or other measures of brain size present but as small as 2.7% [287]. At the behavioral level, assays of sociability did not replicate well across studies either. Lastly, changes at the transcriptomic level also varied strongly across studies [298]. While these issues are likely in part due to differences in mouse line generation, they may also indicate that CHD8 haploinsufficiency might be difficult to model in mice. For this reason, studies using human model systems of brain development are needed.

Cerebral organoids are a great *in vitro* system for the study of human brain development *in vitro*. They are 3D culture systems which faithfully recapitulate early developmental processes in the context of brain region specific development [131, 138, 141, 299]. For this reason organoids have been previously used as models for the study of neuropsychiatric disorders including ASD [131, 150, 151, 300–303]. Prior to this study only two studies employed iPSC-derived *in vitro* models such as cerebral organoids for the study of CHD8 in human cortical development [297, 304]. However, they only focused on single time points thereby losing potential changes in developmental trajectories.

Here, we employ cerebral organoids [131] as a model system of human cortical development to study the heterozygous loss of *CHD8*. We use an isogenic approach based on introducing mutations in human embryonic stem cell (hESC) and investigate the transcriptional landscape using both bulk and single cell RNA-sequencing at multiple time points during development. We found that cerebral organoids replicate the overgrowth phenotype observed as macrocephaly in patients. At the cell population level we observed altered dynamics in developmental trajectories with inhibitory neurons being produced earlier

while the generation of excitatory neurons was delayed in favor of a prolonged phase of progenitor proliferation for this lineage. At the molecular level our analysis uncovered cell-type specific and temporally restricted effects. Overall, our results characterize the outcome of CHD8 haploinsufficiency across development and highlight the importance of longitudinal studies to gain a full picture of the underlying mechanisms.

4.3 Methods

Experiments/Analysis performed by researchers other than Christoph Dotter are indicated
For an exhaustive description of all methods from this study please refer to the publication [168].

Generation of heterozygous mutant hESC

Done by Roberto Sacco, refer to publication for more details

Human embryonic stem cells (hESC) at a passage smaller than 35 were treated with StemPro Accutase (Gibco) for 4 min at 37°C to create a single cell suspension. Dissociated cells were resuspended in mTeSR1 containing 10 µM of the ROCK inhibitor Y-27632 (Stemcell Technologies) or RevitaCell (Gibco). Cells were transfected using Nucleofection buffer (Human Stem Cell Nucleofection Kit 1, Lonza) containing the respective plasmid DNA and the Nucleofector 2B (Lonza), program A-023. For the generation of $CHD8^{+/-}$ hESC 1.2 million cells were used per transfection and 5 µg each of two gRNA/Cas9 expression plasmids containing gRNAs targeting sequences in *CHD8* exon 17 and 19, respectively. For $CHD8^{+/GFP}$ hESC 3 million cells were transfected using 2.5 µg of exon 17 (for $CHD8^{+/GFP}$ and E1114X) or exon 2 (for S62X) and 7.5 µg of the linearized homology arm donor plasmid. Successfully transfected cells were selected either through neomycin selection (using 50 µg /ml G418) or picked based on GFP fluorescence using an EVOS FL imaging system (ThermoFisher). Selected colonies were expanded, genotyped by PCR, checked for off-targets and sub-cloned. Off-targets were predicted using the CRISPR design tool and sites in coding regions were checked by Sanger sequencing.

Southern Blot

Done by Aysan Çerağ Yahya

For Southern blotting genomic DNA was digested with the following restriction enzymes (all NEB): $CHD8^{+/-}$/$CHD8^{+/E1114X;GFP}$: EcoRV HF and NsiI HF; $CHD8^{+/S62X;GFP}$: EcoRV HF and EcoRI HF. Digested samples were resolved on a 0.8% agarose gel before transferring to a positively charged nylon membrane (Roche #11209272001). Membranes were hybridized with DIG labeled probes (Roche #11277065910) recognizing part of the GFP sequence, treated with Anti-DIG-AP, Fab fragments (Roche #11093274910) and CDP Star (Roche #11685627001) and imaged with a Molecular Imaging GE Healthcare Amersham 600.

Generation of cerebral organoids from hESC

Done by Roberto Sacco and Farnaz Freeman

CHD8$^{+/+}$, *CHD8*$^{+/-}$ or *CHD8*$^{+/GFP}$ hESC were grown until a confluency of 70-80% at which point they were dissociated using Accutase (see above). Single cells were seeded in mTeSR1 media with 50 µM Y-27632 ROCK inhibitor into ultra-low-attachment 96-well plates (Corning) at a density of 9000, later optimized to 2500, cells per well. Cells aggregated to EBs and media was changed every other day. On day 3 ROCK inhibitor was omitted and from day 6 cerebral organoid generation was done following the protocol in Lancaster and Knoblich, 2014 [299]. In short, EBs were transferred to low-adhesion 24-well plates (Corning) in neural induction medium (NIM) and medium was changed every other day for five days until formation of neuroepithelial tissue was observed. At this point, denoted as day 0 of organoid generation, organoids were embedded in Matrigel droplets (Corning, #356234) and cultured in cerebral organoid medium (COM) containing B27 (without vitamin A, Gibco) which was changed every other day. After four days the medium was changed to COM with B27 with vitamin A and the plates were placed on an orbital shaker at 70-100 rpm. From this point on medium was changed twice a week. For the purpose of EdU incorporation experiments, 10 µM EdU were added to the culture medium and organoids were incubated for 1 h at 37°C. Afterwards, medium was changed to COM without EdU and organoids were cultured for another 16 hours before processing them for immunostaining.

Immunohistochemistry

Organoids were fixed using 4% paraformaldehyde (PFA) in 1X phosphate buffered saline (PBS) (-/-) at 4°C for 20 min. After washing two times with PBS (-/-) for 10 min each, cryoprotected with 30% sucrose, and embedded in Optimal Cooling Temperature compound (O.C.T.), frozen on dry ice and stored at -20°C. Organoids up to an age of 50 days were cryosectioned into 14-18 mm thick slices, while 18-20 mm were used for day 60-62 organoids.

For immunostainings, sections were washed with 1X PBS (three times, 5 min each) and blocked/permeabilized using 4% normal goat serum (NGS) and 0.5% Triton X-100 in 1X PBS (-/-). Primary antibody incubation was done with antibodies diluted in 0.3% Triton X-100, 4% NGS in PBS at 4°C overnight. After washing with PBS slices were incubated with secondary antibodies for one hour at room temperature. For EdU the Click-iT™ EdU Alexa Fluor™ 647 (Thermo Fisher, C10419) was used. Nuclei were stained using DAPI (Thermo Fisher) in 1X PBS before mounting using DAKO fluorescent mounting medium. Images were acquired using an LSM700 or LSM800 confocal system (Zeiss). Primary antibodies used: SOX2 (rabbit, Millipore, AB5603, 1:300), MAP2 (mouse, Sigma, M4403-50UL, 1:500), Ki67 (rat, Abcam, ab156956, 1:300), FOXG1 (rabbit, Abcam, ab18259), TBR2 (chicken, Millipore, AB15894, 1:300), CTIP2 (rat, Abcam, ab18465). Parvalbumin (mouse, Millipore, MAB1572, 1:250), N-cadherin (mouse, Thermo Fisher 33-3900, 1:500), Laminin (rat, Abcam, ab44941, 1:500), SATB2 (mouse, Santa Cruz sc-81376). Secondary antibodies used were either donkey or goat Alexa Fluor A488, 555, 594, or 647 (Thermo Fisher, 1:1000).

RNA isolation and library preparation

RNA isolation was done with cerebral organoids at day 10 after embedding in Matrigel droplets. Three replicates, each consisting of a pool of three organoids were used per cell line from two independent wildtype lines and two independent *CHD8*$^{+/-}$ clones for a total of twelve samples. RNA was isolated by dissociating the organoids in 700 µL TRIzol™ reagent (Thermo Fisher) using a 20G needle. 140 µL Chloroform were added, samples mixed by vortexing for 15 s, and incubated at room temperature for 2-3 min before centrifugation at 12,000 g for 15 min at 4°C. The aqueous upper phase was transferred to a new tube, mixed with 1.5 volumes of 100% ethanol (EtOH) before loading onto a Zymo-Spin™ IC column (Zymo Research). After subsequent centrifugation (30 s, 10,000 g, room temperature) the column was washed with 400 µL 70% EtOH and the sample was treated with RQ1 DNAseI (Promega; 40 µL 70% EtOH + 5 µL 10X DNAse reaction buffer + 5 µL DNAseI) for 15 min at RT. After two more washes with 700 µL EtOH the final elution was done with DEPC-treated H_2O. RNA concentration was measured on a NanoDrop spectrophotometer (Thermo Scientific) and quality was assessed using the RNA 6000 Nano kit (Agilent) on a Bioanalyzer 2100 (Agilent).
1 µg of RNA was used as input for cDNA library preparation with the QuantSeq 3' mRNA-Seq FWD Library Prep Kit (Lexogen) according to the manual. Success of library preparation and size distribution were checked with the High Sensitivity DNA kit (Agilent). For multiplexing purposes the concentration was measured using the Qubit® dsDNA HS Assay Kit for the Qubit® 2.0 fluorometer and libraries were pooled in equimolar amounts. Sequencing and demultiplexing was performed by the NGS facility of the Vienna Biocenter Core Facilities.

Single-cell RNA-seq library preparation

Done by Roberto Sacco, Alejandro López-Tóbon and Michele Gabriele

Cerebral organoids were collected 20, 60 and 120 days after embedding in matrigel and were dissociated by treatment with Papain/DNAseI (30 U/ml Papain, Worthington LS03126; 3 U/µL DNAseI, Zymo Research) for 30 - 45 minutes dependent on organoid size. Cell suspensions were filtered using 0.4 mm Flowmi™ cell strainers and resuspended in PBS. Cells were counted automatically using a TC20 cell counter (Biorad).
Single-cell suspensions were resuspended in RT-PCR master mix at a concentration of 1000 cells/µL and loaded with partitioning oil and Chromium Single-Cell 3' gel beads (10x Genomics) into a Chromium Single Cell 3' Chip (10x Genomics). Gel beads were coated with primers containing 10x cell barcodes, unique molecular identifiers (UMIs) and poly(dT) sequences. The loaded chip was run on a Chromium controller (10x Genomics) after which cDNAs were fragmented and adapter and sample indices were added to generate sequencing ready libraries. The Qubit® system (Thermo Scientific) calibrated with an in-house control sequencing library was used for library quantification. Size distribution of cDNA and finished libraries were checked using the High Sensitivity DNA kit (Agilent). Two libraries were pooled at an equimolar ratio and sequenced together on an Illumina NovaSeq 6000 using the v2 kit (Illumina) using dual indexing as per the recommendations by 10x Genomics. Sequencing yield was around 250 million reads per sample (2000 - 5000 cells per sample) which corresponded to $\geqslant$ 50,000 reads per cell.

Single-cell RNA-seq analysis

Clustering and Differential Expression analysis

Done by Carlo Emanuele Villa and Cristina Cheroni with input from Christoph Dotter on cluster annotation and relative population frequency visualization.

Cell Ranger (10x Genomics) was used to align raw sequencing reads to the hg38 genome. The following thresholds were used to select high-quality cells: $\leqslant 5\%$ mitochondrial reads and $\geqslant 450$ genes detected. For clustering purposes the data from all 21 samples was integrated using Conos [305] and the resulting clusters were considered as shared populations for scGen [306]. The integrated dataset was used as input for dimensionality reduction using the UMAP functionality of Scanpy (version 1.6.0, [307]). Clustering was done using the Leiden algorithm (Scanpy, [308]) with an empirically chosen resolution of 0.7 yielding a total of 10 clusters.

For cluster annotation three strategies were employed: a) Cluster identifying genes were derived using rank_genes_groups from Scanpy. b) A semi-supervised annotation approach using SCINA incorporating three published datasets [309–311]. c) Overlap of cluster marker genes (log2FC > 0.5, adj. P value < 0.05 in differential expression analysis of this cluster vs all others) with cell type markers identified in a published single cell sequencing study [310]. Statistical significance of overlaps was determined using Fisher's exact test. d) Projection of single cell fetal cortex dataset into our UMAP coordinate system using the ingest algorithm. The collected information was manually curated to derive the final annotation. Visualization of expression of cell population specific genes across clusters was done using Scanpy (Dotplot function).

For pseudotime analysis clusters were first stratified by genotype and age before aggregation via their median into 'supercells'. Supercells with low read contribution from any of the mutant lines were removed. The supercells (CTL) were used as input for the Palantir algorithm [312]. Visualization was done using the STREAM algorithm [313].

Sample-frequencies per cluster were visualized as stacked bar plots per group of 'progenitor clusters', 'inhibitory neuron cluster' and 'excitatory neuron clusters'. Differential abundance analysis was performed according to the workflow outlined in OSCA[1]. Briefly, edgeR was used for testing differential abundance across time points in control lines or to compare mutant (excluding S62X) to controls per time point. 'Unknown' cluster was excluded from this analysis. Abundance differences with a nominal P value < 0.05 were highlighted. Density plots visualizing differences in abundances were created by creating a grid of sub-regions across the UMAP plot [314] and then using python sklearn (version 0.23.2) adding a Gaussian kernel to derive cell density per sub-region for control and mutant cells. Differential plots are calculated by subtracting control from mutant cell density.

Differential gene expression analysis was done using diffxpy (version 0.7.4) employing Mann-Whitney U tests and cells as replicates (after removing outliers identified by the Leiden algorithm). Cutoffs for significant genes were adjusted P value < 0.05 and $|\text{log2FC}| > 1$ unless otherwise specified. The R package TopGO (version 2.42.0, [315]) was used for Gene Ontology functional enrichment analysis for the Biological Component category and using Fisher statistics and the Weight01 algorithm. A set of highly variable genes

[1] http://bioconductor.org/books/release/OSCA/

was used as background for this analysis. GO terms were deemed significantly enriched if they had a P value < 0.01 and an enrichment value $\geqslant 1.75$.

Bulk RNA-seq analysis

Trimming and filtering of demultiplexed raw sequencing reads was performed according to the guidelines provided by Lexogen. In short, BBDuk (BBMAP package[2]) was used to remove/filter out adapter contamination, random primer sequences and low quality tails. The trimmed reads were aligned to the human genome (genome: hg38/GRCh38, gene annotation: Gencode release 24) using STAR (version 2.5.1, [191]). Only uniquely aligned reads were considered (–outFilterMultimapNmax 1) and STAR was also used to produce read counts per gene (–quantMode GeneCounts).
Differential gene expression analysis was done using DESeq2 (version 1.12.4, [209]) using an adjusted P value cutoff of 0.05 and local fit for dispersion estimation. Comparison was done at the genotype level (6 vs 6 samples) treating samples from independent cell lines as replicates. To increase stringency regarding between-line variability we also determined genes differentially expressed between the two wildtype lines and removed those from the DEGs between wildtype and mutant.

Enrichment analyses

Gene Ontology functional enrichment analysis was performed with the GOstats package (version 2.36.0, [199]) with a P value cutoff of 0.001 and conditional testing enabled. Visualization of enriched GO terms was done with the GOTreeVis package[3]. For the segregation of DEGs based on whether they are CHD8 targets or not a published list of genes with CHD8 ChIP-seq signal near their transcription start site [76] was used (5601 genes after filtering duplicate gene symbols).
For overlap analyses between bulk and scRNA-seq DEGs from each cluster with adjusted P value $\leqslant 0.05$ were used, no fold-change filter was applied.
For enrichment analysis of ZEB2 targets ChIP-seq data for ZEB2 (HEK293T cells) was downloaded as called peaks in bed file format from GEO (GSE91749) [316]. Peaks were connected to genes using HOMER (annotatePeaks function, version 4.11, [317]) with default parameters. This resulted in a set of 8209 genes after filtering for duplicate gene symbols.
For the enrichment analysis of additional transcription factor targets the R enrichR package[4] (version 3.0) was used to interface the Enrichr gene set enrichment analysis web server [204]. The dataset used was the ENCODE_and_ChEA consensus set. In the case of multiple significant results from different sets for the same transcription factor the result with the lowest adjusted P value was reported. To increase the stringency and remove enrichments only connected to the genes detected in a certain cell type we calculated enrichments for 1000 permutation gene lists of the same size as the respective DEG set and selected from all detected genes per cluster. From this we calculated a permutation P value as the fraction of background results with enrichment of the same or greater magnitude based on comparisons of adjusted P value. For the figure we used a cutoff of 0.05 for both the adjusted enrichment P value of the DEG set as well as the

[2]http://jgi.doe.gov/data-and-tools/bbtools/
[3]https://github.com/dottercp/GOTreeVis
[4]https://github.com/wjawaid/enrichR

permutation P value. For the visualization the R package ggh4x[5] (version 0.2.1) was used as an add-on for ggplot2.
Statistical testing for gene set overlaps was done using Fisher's exact test in R and P values were corrected for multiple testing using the Benjamini and Hochberg method. Background gene sets were restricted to only detected genes. All visualization was done in R using the ggplot package [197].

Quantitative PCR

Total RNA was isolated from hESC, EBs and organoids as described above. cDNA was synthesized using the SuperScript III kit (Life Technologies). qPCR was done with the LightCycler® 480 SYBR Green I Mastermix (Roche) on a LightCycler® 480 (Roche) according to the manual with primers for hCHD8 and hTBP for normalization. Results were visualized in R.

Western blot

Done by Roberto Sacco.

For hESC protein extracts were prepared by cell lysis in RIPA buffer (150 mM NaCl, 50 mM Tris-HCl, pH 7.6, 2 mM EDTA, 1% Triton X-100, 0.5% Na-deoxycholate, 0.1% SDS, 1x Protease inhibitor cocktail (Roche)). Cerebral organoids were dissociated and cells lysed by trituration in lysis buffer (140 mM NaCl, 50 mM Tris-HCl pH 8, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, 1x Protease inhibitor cocktail (Roche)). Lysates were denatured with Laemmli buffer and boiling for 5 min at 95°C. Aliquots of 20 µg and 50 µg for hESC and organoids, respectively, were resolved on a 6% SDS-PAGE (GAPDH: 10%) and transferred to 0.45 µM Immobilon® -P PVDF membranes (Millipore). Blocking was done with 5% milk/0.05% Tween-PBS for 6 hours at 4°C followed by primary antibody incubation in blocking solution overnight at 4°C. Primary antibodies used were: CHD8 C-terminus (rabbit, Novus Biologicals NB100-60418, 1:2000), CHD8 N-terminus (rabbit, Cell Signaling #77694, 1:2000), P70 S6 Kinase (rabbit, Cell Signaling 2708, 1:1000), and GAPDH (rabbit, Millipore, 1:1000). Membranes were subsequently washed, incubated for one hour with anti-Rabbit IgG (H+L)-conjugated secondary antibody (donkey, Pierce SA1-200, 1:1000), followed by another wash before signal detection via enhanced chemiluminescence (Pierce, #32209 and #34095).

[5]https://github.com/teunbrand/ggh4x

4.4 Results

4.4.1 Modeling human cortical development in cerebral organoids

For modeling CHD8 haploinsufficiency in a human model of early brain development we employed the cerebral organoid method first published by Lancaster et al. [131] with some modifications (Fig. 4.1a, for details see methods). We observed a high rate of successful differentiation with over 90% of embedded embryoid bodies forming cerebral organoids. We verified the success and reproducibility of our model system by immunostaining and found consistency in overall morphology (Fig. S4.1b). Organoids formed rosette structures of SOX2 positive progenitor cells reminiscent of the ventricular zone (Fig. S4.1c-d). These regions eventually also developed a zone with TBR2 positive intermediate progenitor cells surrounded by a layered structure of lower and upper layer cortical excitatory neurons as exemplified by staining for markers CTIP2 and SATB2, respectively (Fig. S4.1e). We also verified that CHD8 itself was expressed throughout the course of organoid development (Fig. 4.1b). Of all time point surveyed we found the highest expression in day 36 and lowest at day 110 hinting that CHD8 might be most important at earlier developmental stages.

4.4.2 Generation of *CHD8*$^{+/-}$ human embryonic stem cells with isogenic controls

We chose to model *CHD8* mutations through the use of Clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 mediated genome editing. This provided us with the advantage of having an isogenic background between mutants and control, thereby eliminating possible confounding effects from genetic differences. In humans the CHD8 protein exists in two main variants, one from a long isoform and one from a N-terminally truncated shorter isoform (Fig. 4.1c). We generated four different heterozygous knockout lines for the scope of this study. First, to model general *CHD8* haploinsufficiency, we deleted a section of the gene spanning from exon 17 to exon 19 as well as a green fluorescent protein (GFP) knock-in mutant disrupting the gene at the same locus by insertion of a GFP cassette (henceforth referred to as *CHD8*$^{+/-}$ and *CHD8*$^{+/GFP}$). In addition we decided to also model two premature stop codon mutations found in patients, S62X and E1114X (*CHD8*$^{+/S62X;GFP}$ and *CHD8*$^{+/E1114X;GFP}$, henceforth referred to as *CHD8*$^{+/E1114X}$ and *CHD8*$^{+/S62X}$). In both cases we introduced the mutation and inserted the GFP cassette downstream in order to facilitate selection of successfully edited clones. Of particular importance was the fact that the S62X mutation, in contrast with E1114X and many other *CHD8* mutations, did not lead to macrocephaly in the patient [74]. This mutation is also located within the N-terminal section exclusive to the long isoform (Fig. 4.1d). Successfully edited heterozygous clones as verified by Sanger sequencing (Fig. S4.2a-d) were selected and expanded. We verified reduction of either *CHD8* mRNA or CHD8 protein levels via qPCR and Western blot (Figs. 4.1e-f and S4.2e) and checked for both CRISPR/Cas9 off-target effects using Sanger sequencing, as well as random integration of the GFP cassette via Southern blot (Fig. S4.2f).

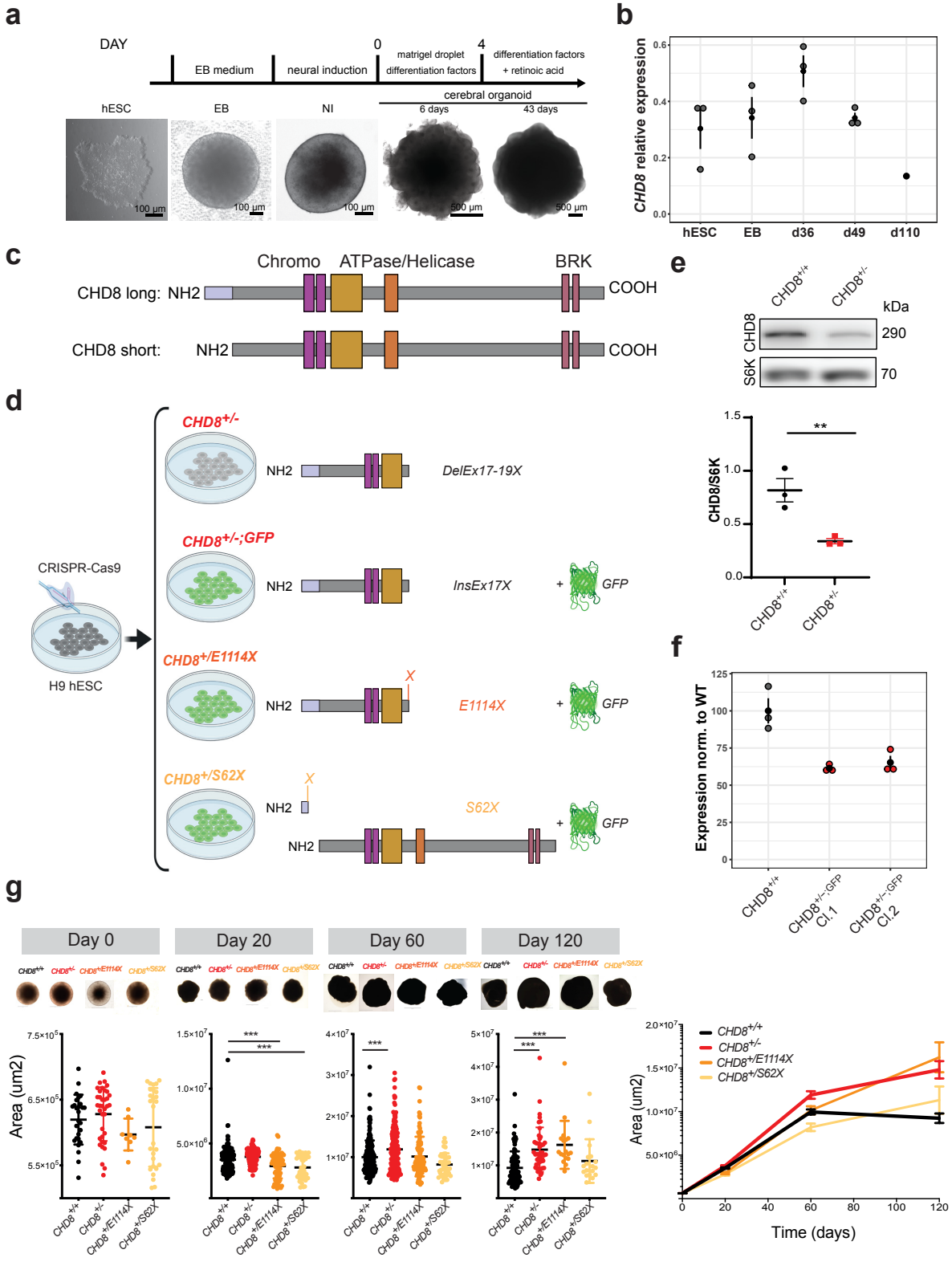
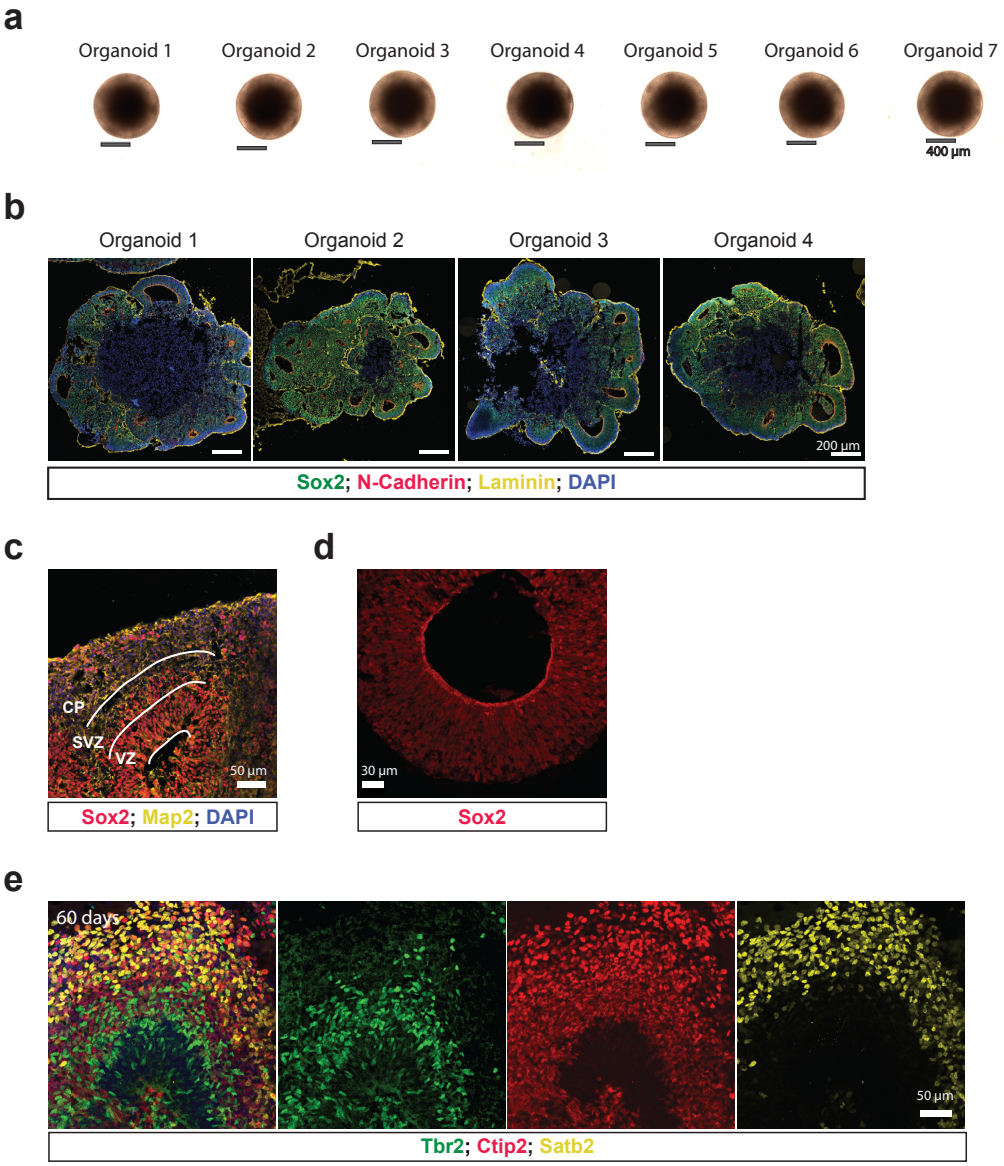


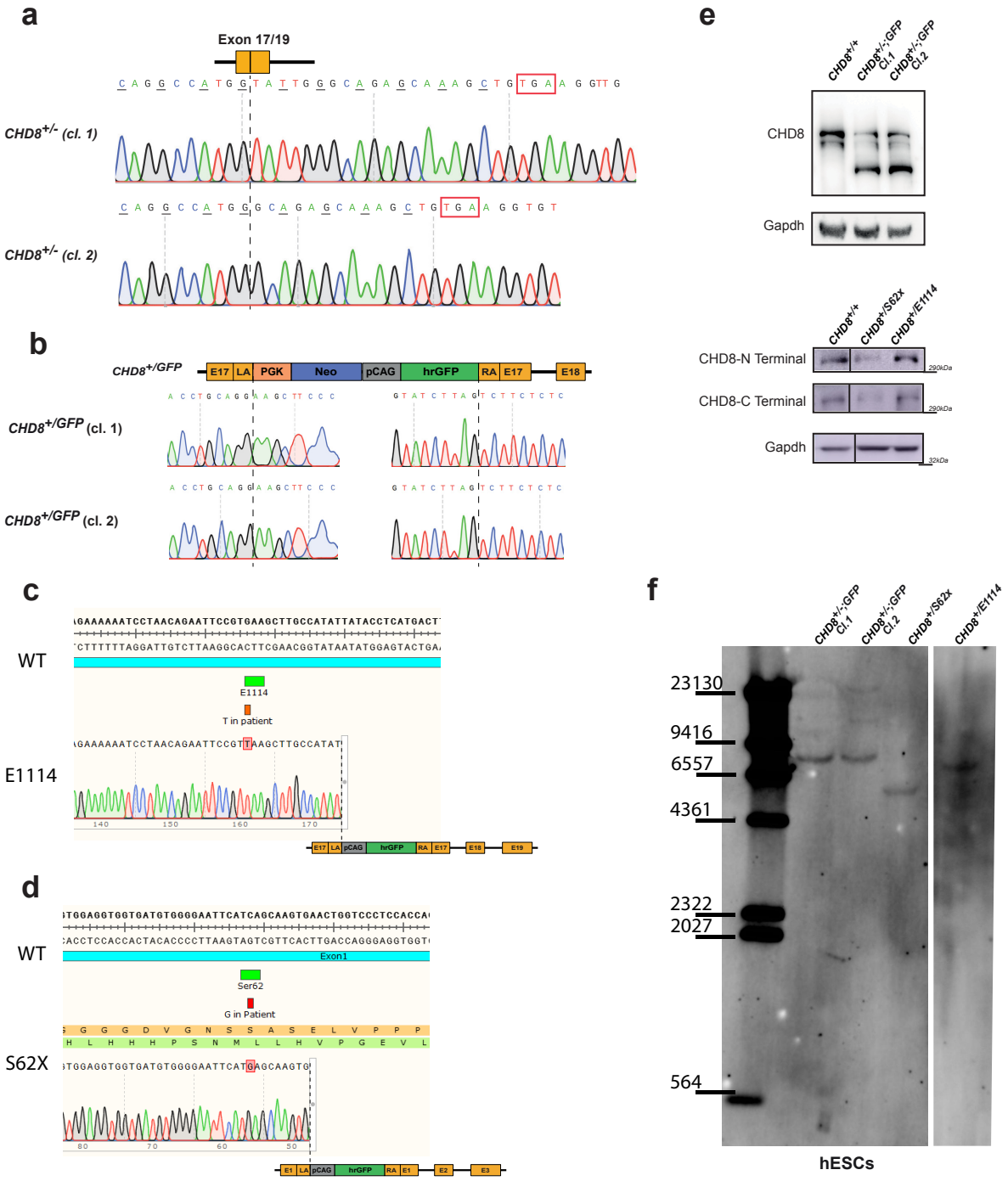
Figure 4.1: ***CHD8*$^{+/-}$ and *CHD8*$^{+/GFP}$ organoids present with increased size. a)** Outline of the cerebral organoid generation protocol. **b)** *CHD8* is expressed across organoid development. qPCR results were normalized to *TBP* and presented as mean ± SEM. **c)** Illustration of human CHD8 isoforms, long (L) and short (S). The long isoform specific N-terminal part is colored in light purple. **d)** Overview of the mutant hESC lines generated in this study using CRISPR/Cas9 including predicted (truncated) proteins. *CHD8*$^{+/-}$: deletion from exon 17 to exon 19 leading to a premature stop codon. *CHD8*$^{+/GFP}$: insertion of

hrGFP into exon 17 introducing a premature stop codon. $CHD8^{+/E1114X}$: glutamate at position p.1114 was mutated to a stop codon, followed by insertion of hrGFP. $CHD8^{+/S62X}$: serine at position p.62 was mutated to a stop codon, followed by insertion of hrGFP. In this case only the long isoform is affected as demonstrated by the presence of the full-length protein encoded by the short isoform. **e)** Western blot for CHD8 in $CHD8^{+/+}$ and $CHD8^{+/-}$ samples revealing reduction in CHD8 protein levels upon deletion of exon 18. Results are shown normalized to S6K and presented as mean ± SEM. p = 0.013, unpaired t-test. **f)** qPCR results for day 60 $CHD8^{+/GFP}$ organoid samples showing reduction of *CHD8* mRNA levels in relation to $CHD8^{+/+}$. Error bars show variation from technical triplicates. Results presented as mean ± SEM. **g)** Organoid growth as measured at four different time points. Top: Representative images of cerebral organoids at day 0, 20, 60 and 120 after embedding. Bottom: Quantification of organoid size as area in bright-field images. $CHD8^{+/-}$ and $CHD8^{+/GFP}$ were pooled. Number of organoids per time point and genotype - Day 0: $CHD8^{+/+}$ = 27, $CHD8^{+/-}$ = 36, $CHD8^{+/E1114X}$ = 11, $CHD8^{+/S62X}$ = 27. Day 20: $CHD8^{+/+}$ = 196, $CHD8^{+/-}$ = 119, $CHD8^{+/E1114X}$ = 80, $CHD8^{+/S62X}$ = 45. Day 60: $CHD8^{+/+}$ = 202, $CHD8^{+/-}$ = 163; $CHD8^{+/E1114X}$ = 106, $CHD8^{+/S62X}$ = 41. Day 120: $CHD8^{+/+}$ = 85, $CHD8^{+/-}$ = 46, $CHD8^{+/E1114X}$ = 18, $CHD8^{+/S62X}$ = 18. **** P < 0.001, ordinary one-way ANOVA, Dunnett's multiple comparisons test. Results presented as mean ± SEM. Figure adapted from [168].



Supplementary Figure S4.1: **Characterization of cerebral organoids.**

a) Representative bright field images of wildtype cerebral organoids at the day of embedding. Neuroectoderm formation can be seen as brightening. **b)** Representative immunostainings for Sox2, N-Cadherin and Laminin highlighting the overall internal structure of wildtype organoids 10 days after embedding. **c-e)** Representative images of wildtype organoids 60-62 days after embedding. Cerebral organoids form rosette structures with Sox2 positive progenitors at their center in a ventricular zone (VZ) like region. Radially outwards this is followed by a zone of intermediate progenitors akin to the subventricular zone (SVZ) and neurons (MAP2+) organized into regions of first lower (CTIP2+) and then upper layer (SATB2+) identity. Stainings are indicated at the bottom of the panels. CP, cortical plate. Figure adapted from [168].



Supplementary Figure S4.2: ***CHD8*$^{+/-}$ mutant hESC line generation.**

a-d) Sanger sequencing results shown as sequences and chromatograms for *CHD8*$^{+/-}$ (a), *CHD8*$^{+/GFP}$ (b), *CHD8*$^{+/E1114X;GFP}$ (c), and *CHD8*$^{+/S62X;GFP}$ (d). **a)** The *CHD8*$^{+/-}$ line was generated through deleting exon 18 as well as parts of exon 17 and 19, employing two guide RNAs. The shown sequence spans the regions flanking the deletion for two independent clones. **b)** The *CHD8*$^{+/GFP}$ line was generated by inserting a cassette containing neomycin under the control of a PGK promoter and hrGFP under the control of a pCAG promoter into the targeting site in exon 17. The shown sequence spans the insertion sites. **c)** For the *CHD8*$^{+/E1114X;GFP}$ line we introduced a patient specific mutation (E1114X) followed by a GFP cassette through homology directed repair. Shown are sequences for wildtype (WT) and *CHD8*$^{+/E1114X;GFP}$ highlighting the introduced mutation as well as the insertion site. **d)** The *CHD8*$^{+/S62X;GFP}$ line was generated in the same fashion but for the S62X mutation. **e)** Western blot checking CHD8 levels in hESC lines. Top: Two independent *CHD8*$^{+/GFP}$ clones. Bottom: *CHD8*$^{+/S62X;GFP}$ and *CHD8*$^{+/E1114X;GFP}$. GAPDH was used for normalization. For the bottom blots the three samples per antibody come from non-consecutive lanes of the same membrane and each antibody result comes from a separate membrane. **f)** Southern blot with a probe against the inserted GFP confirmed absence of random integration and correct insertion at the target site. All results are from the same gel and transfer but the membrane was later cut as indicated. Figure adapted from [168].

In summary we generated four different lines modeling *CHD8* haploinsufficiency covering both general loss of function as well as patient specific truncations.

4.4.3 Cerebral organoids derived from *CHD8*$^{+/-}$ cells display an overgrowth phenotype

We next generated organoids from both *CHD8*$^{+/+}$ control lines and *CHD8*$^{+/-}$ mutant lines (*CHD8*$^{+/-}$, *CHD8*$^{+/GFP}$, *CHD8*$^{+/E1114X}$, and *CHD8*$^{+/S62X}$). Monitoring their growth at the macroscopic level we observed that early on, e.g. between 0 and 20 days after embedding *CHD8*$^{+/-}$ organoids were similar in size or even slightly smaller than *CHD8*$^{+/+}$ organoids, contrary to our expectations given that macrocephaly is an often observed phenotype in patients with *CHD8* mutations [74]. However, this changed dramatically over time with *CHD8*$^{+/-}$, *CHD8*$^{+/GFP}$, and *CHD8*$^{+/E1114X}$ organoids catching up by two months of age with comparable or increased size. Finally, at four months of age they exhibited a clear overgrowth phenotype with around 50% size increase (Fig. 4.1g). Importantly, *CHD8*$^{+/S62X}$ organoids did not show this phenotype and remained similar to *CHD8*$^{+/+}$ organoids which reflected the lack of connection of this mutation with the macrocephaly phenotype in patients. This highlighted the ability of our *in vitro* model system to capture mutation specific effects at the macroscopic level.

4.4.4 *CHD8* mutations lead to altered developmental trajectories in cerebral organoids

Based on the observed differences in growth trajectory between *CHD8*$^{+/+}$ and *CHD8*$^{+/-}$ mutant organoids we speculated that these results might reflect differences in developmental trajectories at the cellular level. Therefore we performed single cell RNA sequencing across multiple time points covering the breadth of growth phenotypes we observed: slightly reduced/equal at day 20, equal/slightly increased at day 60 and clearly increased at day 120. We profiled a total of 75,060 cells across 21 samples (3 control lines, 4 mutant lines) with a median of 3783 cells per sample at day 20, 4241 cells at day 60 and 2218 cells at day 120.

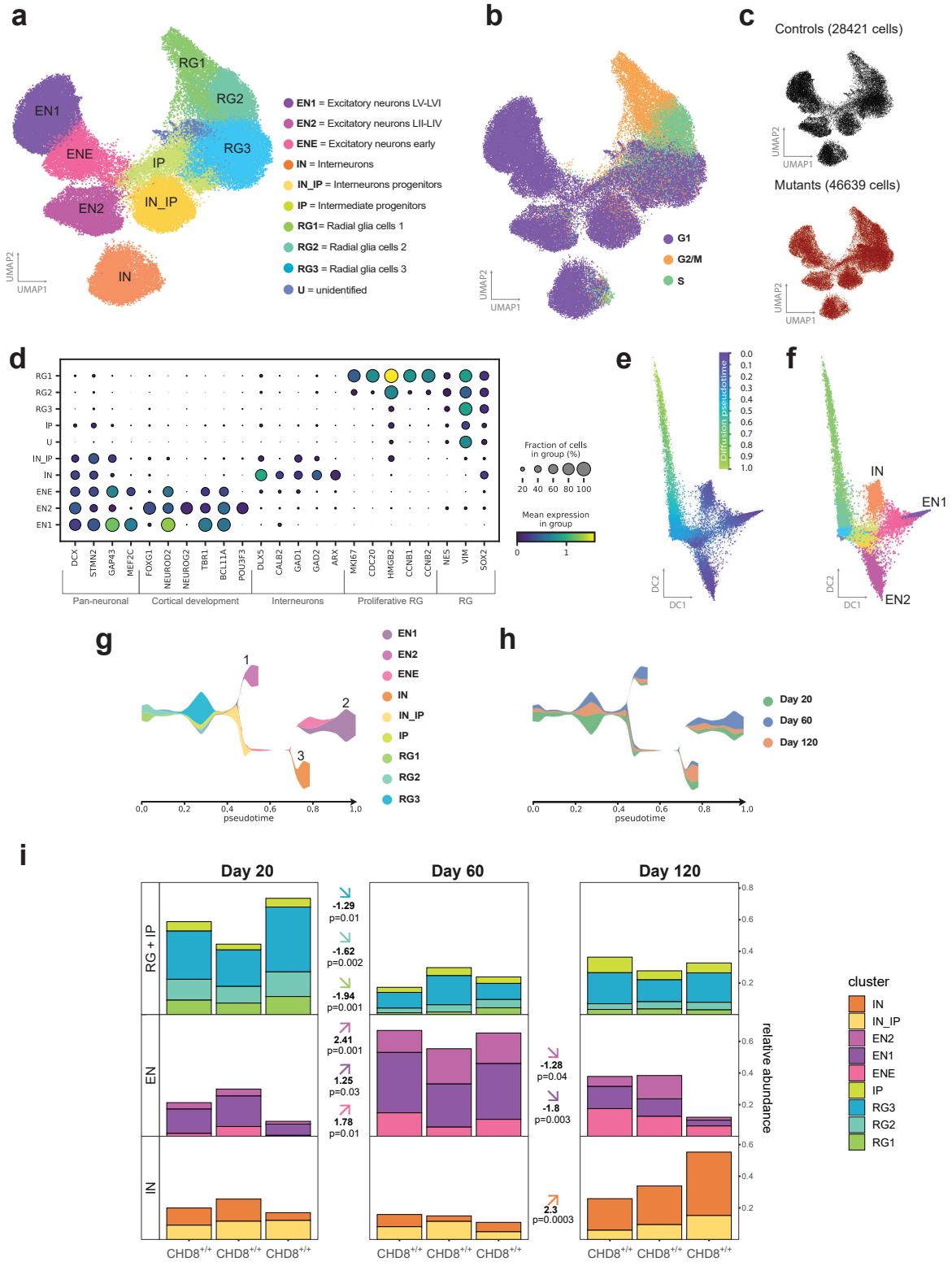
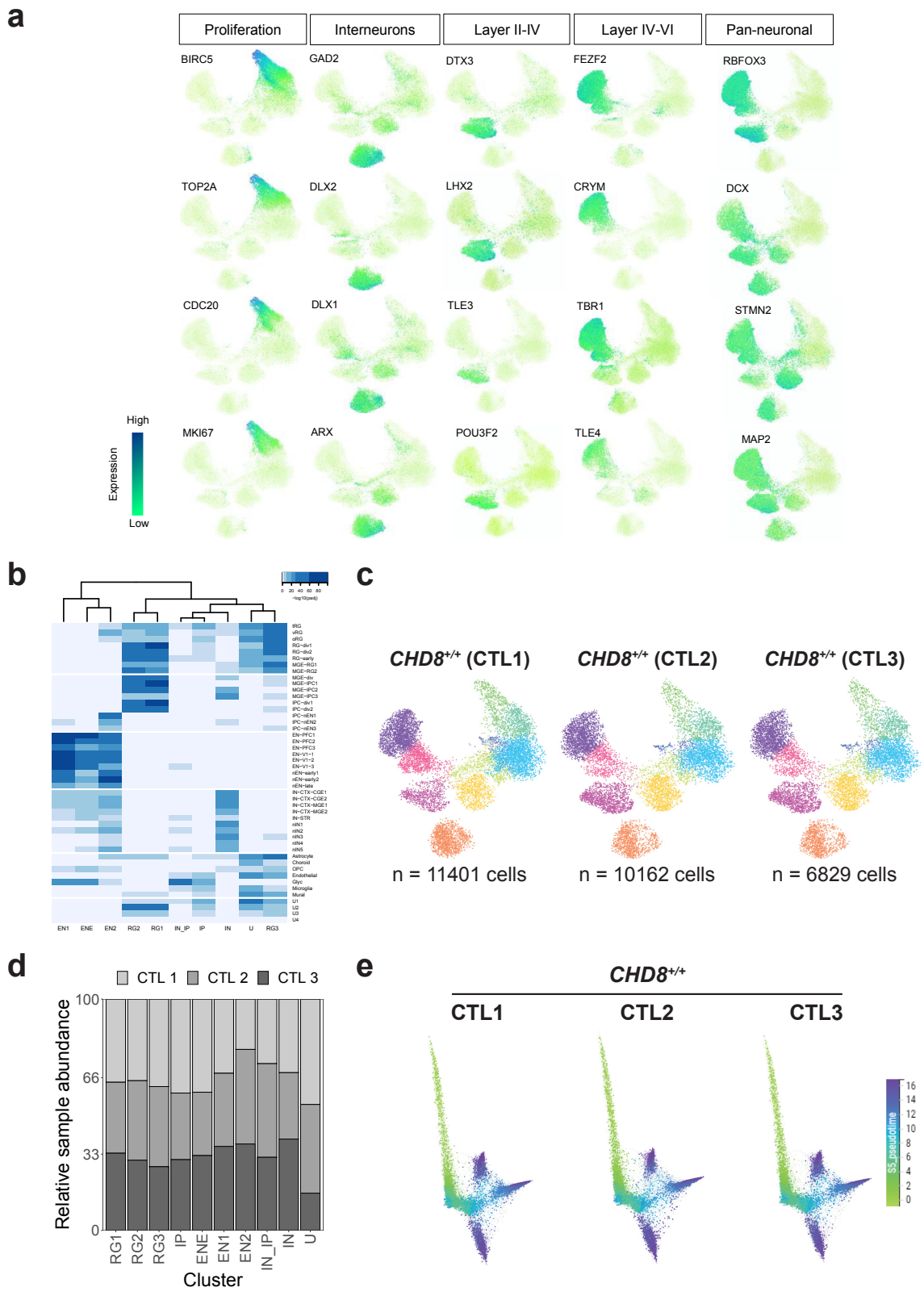


Figure 4.2: **Cerebral organoids recapitulate early neurogenesis. a)** UMAP dimensionality reduction of scRNA-seq data from cerebral organoids at day 20, day 60, and day 120 after matrigel embedding. Colors indicate cell type clusters; EN1, layer V-VI excitatory neurons; EN2, layer II-IV excitatory neurons; ENE, early excitatory neurons; IN, inhibitory neurons; IN_IP, inhibitory neurons intermediate progenitors; RG1-3, radial glial cells 1-3. **b)** UMAP plot colored by cell cycle phase. **c)** UMAP plot showing either just cells from control lines (three lines) or just mutant lines (four lines) and indicating

total cell numbers. Each sample consisted of a pool of three cerebral organoids generated in at least two independent differentiation batches. **d)** Cell type marker expression across the identified clusters. Dot size scales with percentage of cells in a cluster expressing the marker gene. Color indicates mean normalized expression level. **e+f)** Diffusion maps illustrating differentiation trajectories. Colors show (e) pseudotime trajectory information indicating three terminal outcomes (EN1, EN2, and IN) in dark blue and one origin in green or (f) cell type clusters as in a). **g+h)** Tree graphs showing differentiation trajectories across pseudotime (x-axis) as calculated in control lines. The three labeled end points are 1) excitatory neurons LII-IV 2) excitatory neurons LV-VI and 3) interneurons. Branch thickness scales with percentage of cells in this pseudotime frame. Color code relates to **g)** cell clusters and **h)** stage with the scale normalized per time point to number of cells. **i)** Relative proportions of cells per cluster are shown for each stage and for three control cell lines. Clusters are roughly grouped into progenitors (top), excitatory neurons (middle), and interneurons (bottom). Colored arrows between time points indicate significant changes in proportions for the specific clusters. For each such transition the log2 fold change and the p value is shown. All fold changes are calculated as later over earlier time point. Figure adapted from [168].

We identified common cell types by first integrating all samples followed by clustering using the Leiden algorithm and visualization via Uniform Manifold Approximation and Projection (UMAP, Fig. 4.2a). Next we chose a multi-pronged approach to annotate these clusters and associate them with cell types: visualization of known cell type marker expression (Fig. S4.3a), enrichment for markers from a published single cell RNA-seq study on fetal human cortex ([310], Fig. S4.3b), and integration of our data with a different human fetal dataset (see methods). This resulted in the annotation of 10 clusters shown in Fig. 4.2a and used throughout the study: three radial glia cell clusters (RG1-3) corresponding to different cell cycle phases (RG1 - G2M phase, RG2 - S phase, RG3 - G1 phase, Fig. 4.2b), a cluster of intermediate progenitor cells (IP), clusters of excitatory neurons split into early (ENE), upper layer (EN1), and lower layer (EN2) neurons, and clusters of interneuron progenitors (IN_IP) and interneurons (IN). Lastly, one cluster) remained unidentified and was removed from further analyses. At a global level we observed no obvious genotype specific effects with cells of either wildtype or *CHD8*$^{+/-}$ distributed across all clusters (Fig. 4.2c). When we checked the reproducibility of cell type frequencies observed in controls we found comparable results across all three lines (S4.3c-d). Cell type marker expression in the annotated clusters is visualized in Fig. 4.2d.

In order to investigate developmental trajectories we applied a diffusion pseudotime (DPT) analysis modeling transitions between different cell states. We identified three main trajectories characterized by three distinct end points (Fig. 4.2e). Overlaying cluster identities revealed that the common trajectory starting point was formed by proliferating progenitor cells and the three end points corresponded to interneurons as well as upper and lower layer excitatory neurons (IN, EN1, and EN2; Fig. 4.2f). We again confirmed the reproducibility of our organoid generation and found no differences of the identified trajectories between control lines (S4.3e). Additionally, we complemented our diffusion pseudotime approach with STREAM (Single-cell Trajectories Reconstruction, Exploration And Mapping) using control lines only (n = 28,421 cells), reproducing the three main lineages we observed before (Fig. 4.2g). This visualization also nicely demonstrated the temporal changes in each cluster (Fig. 4.2h) with later pseudotime cells predominantly originating from day 60/120 samples. Projecting expression levels of lineage transcription factors onto this graph revealed the importance of NEUROG2/FOXG1/POU3F3(BRN1) and NEUROD2/TBR1 in the specification of upper (EN2) and lower layer neurons (EN1), respectively. In addition, expression of factors such as DLX2/ARX/CALB2 marked the definition of interneurons in our model system (S4.4a).



Supplementary Figure S4.3: **Cluster annotation and reproducibility. a)** Expression of marker genes projected onto the UMAP plot. **b)** Results of hypergeometric tests for the overlap of cluster marker genes with a published human fetal brain single cell dataset [310]. Color encodes negative decimal logarithm of the multiple-testing adjusted P value. **c-e)** Reproducibility of clusters and developmental trajectories across wildtype replicates. c) UMAP per replicate, d) relative abundance of cells per sample and cluster(normalized to cells/sample), and e) diffusion maps per replicate. Figure adapted from [168].

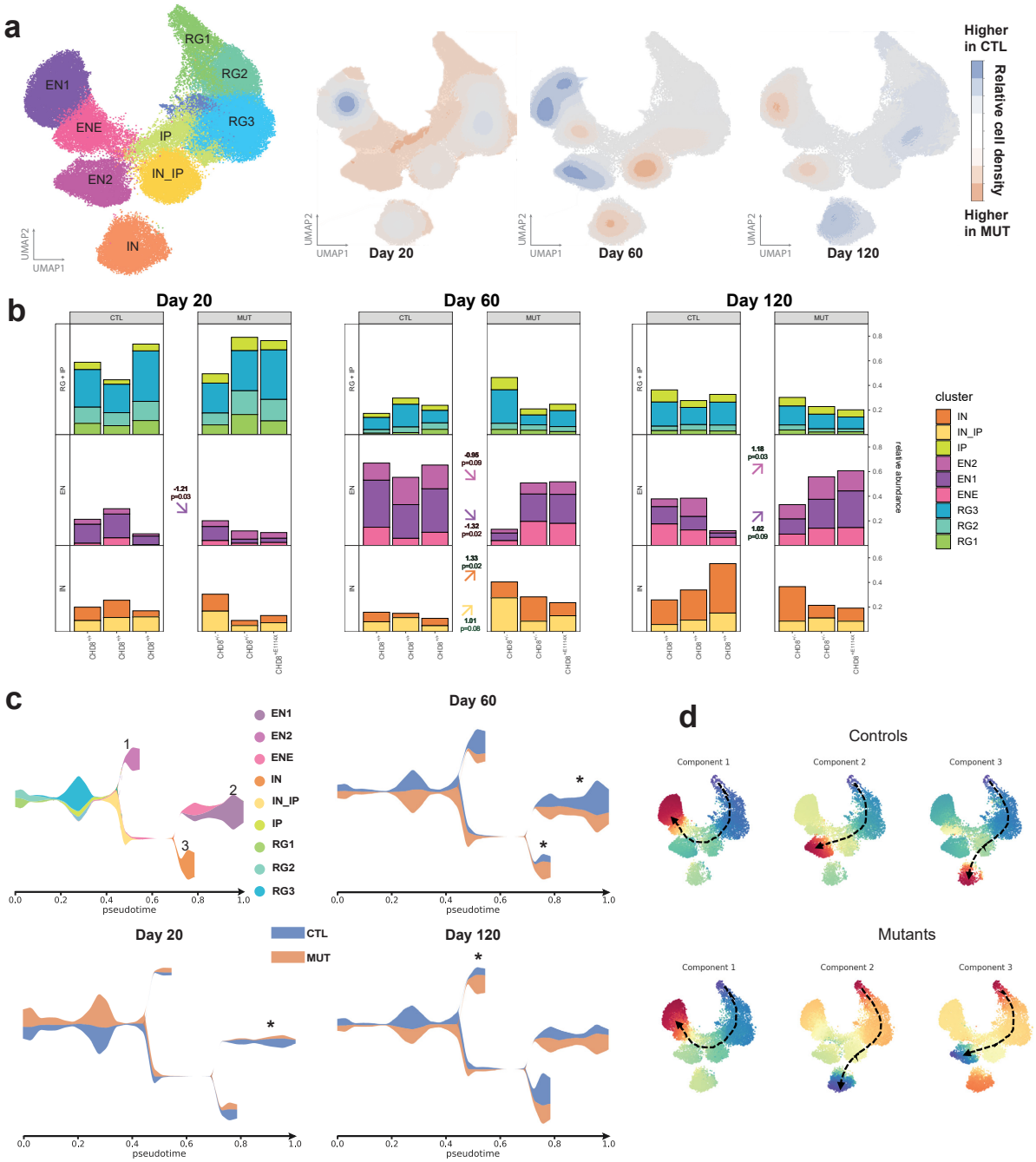
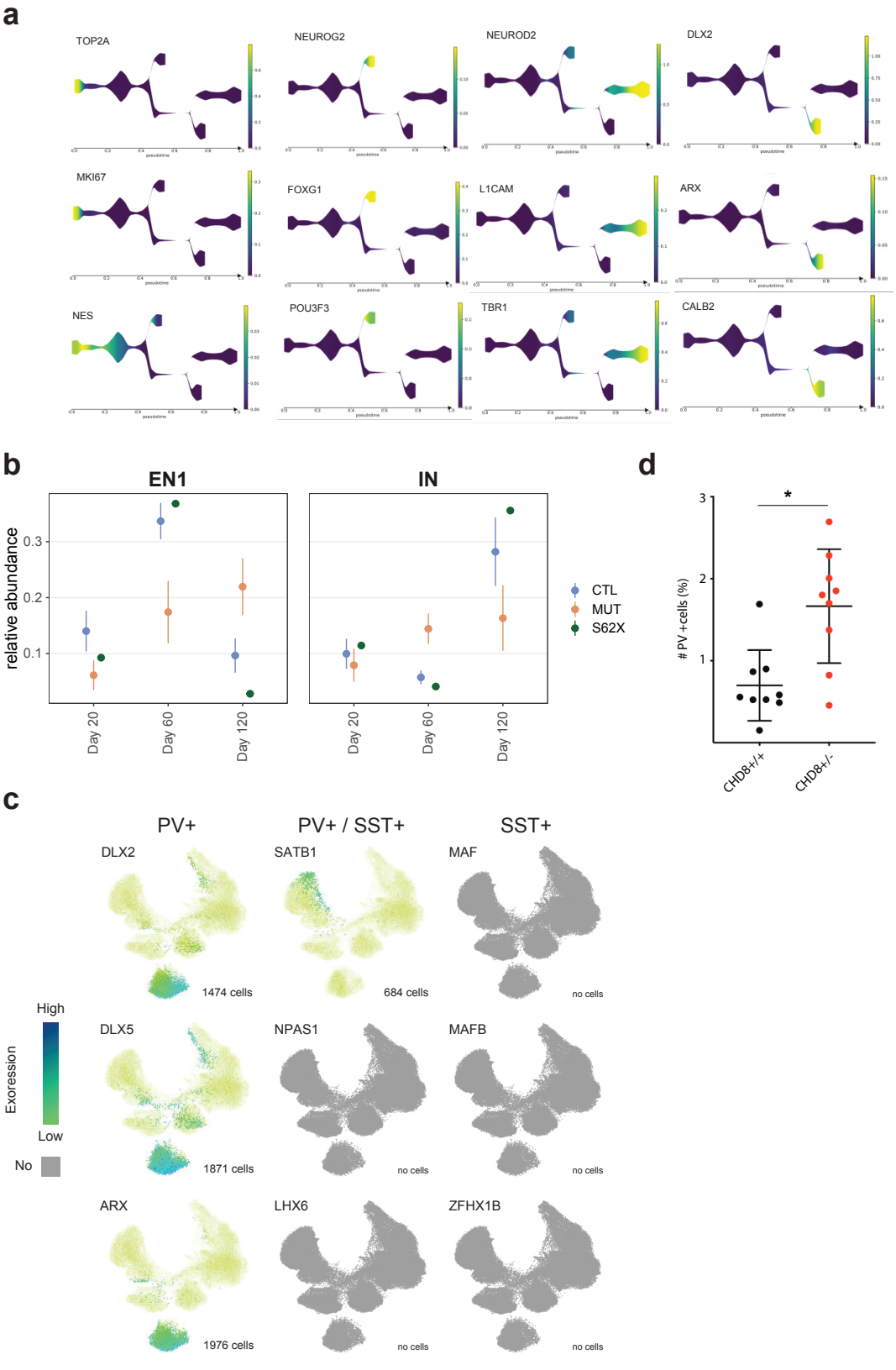


Figure 4.3: **Heterozygous loss of *CHD8* causes transient cell population changes and premature interneuron differentiation. a)** Changes in relative cell abundance across the UMAP plot shown in Fig. 4.2a (left) are shown for the three stages as density contour plots (right). Red and blue indicate increased relative cell abundance in mutant (MUT) and wildtype (CTL), respectively. **b)** Cell type proportions per stage and sample for mutant and wildtype samples are shown as stacked bar plots. Arrows denote changes in abundance between mutants and wildtype samples per cluster per stage as revealed by differential abundance analysis (P value < 0.1; see methods). In addition, log2 fold changes (bold) and P value are shown for those changes. Positive/negative log2 fold changes represent increase/reduction of that population in the mutant. **c)** Relative abundances of cells in mutant (orange) and controls (blue) are projected onto the tree graph from Fig. 4.2g (top left) for each time point. Significantly changed populations (P value < 0.05, as identified in b) are highlighted by asterisks. **d)** The first three diffusion pseudotime components, i.e. the three main trajectories in order of likelihood are shown for day 60 control (top) and mutant samples (bottom). Trajectories represented by each component are shown by arrows between the two ends (blue/red, direction might flip). Note the difference in trajectory order. Figure adapted from [168].



Supplementary Figure S4.4: **Additional results and validation for single cell sequencing. a)** Projection of marker gene expression onto the developmental trajectories visualized as a tree graph. **b)** Comparison of relative cell type abundances between *CHD8*$^{+/+}$ (blue), mutant (*CHD8*$^{+/-}$, *CHD8*$^{+/GFP}$, and *CHD8*$^{+/E1114X}$; orange) and *CHD8*$^{+/S62X}$ (dark green) organoids for clusters EN1 and IN across time points. **c)** Projection of gene expression levels onto the UMAP plot for all cells for markers of interneuron subtypes [318]. **d)** Quantification of immunostaining for parvalbumin in day 60 *CHD8*$^{+/+}$

and *CHD8*$^{+/-}$ cerebral organoids. Mann-Whitney test, * P $<$ 0.01. n(*CHD8*$^{+/+}$) = 9, n(*CHD8*$^{+/-}$) = 9, number of non-consecutive sections from $\geq$ three cerebral organoids. Black lines indicate mean ±s.d. Figure partially adapted from [168].

After our characterization of the control condition we next investigated how heterozygous loss of *CHD8* might affect the observed cell types and their specification. For this we chose to focus on the three cell lines which shared the overgrowth phenotype (*CHD8*$^{+/-}$, *CHD8*$^{+/GFP}$, and *CHD8*$^{+/E1114X}$) in order to potentially unearth its origin. We first focused our attention on changes in relative cell type abundance and observed trends of more proliferating cells and intermediate progenitors at day 20, a reduction of excitatory and concurrent increase of inhibitory neurons at day 60 which got reversed at day 120 with increases in excitatory and relative decrease in inhibitory neurons (Fig. 4.3a). When investigating these differences at the level of specific clusters we observed statistically significant reduction of lower layer excitatory neurons (EN1) at day 20 and 60 as well as significant increase of upper layer neurons (EN2) at day 120. Contrastingly we observed the relative increase in interneurons (IN) at day 60, the largest change overall (Fig. 4.3b). Of note, when checking abundances of clusters EN1 and IN in the *CHD8*$^{+/S62X}$ samples we found dynamics much closer to wildtype than to the other mutants, again suggesting that this mutation differs from the others in terms of developmental impact (Fig. S4.4b). We further confirmed based on marker expression that interneurons at day 60 mainly expressed markers of parvalbumin positive neurons (Fig. S4.4c) and confirmed the abundance difference in our single cell RNA sequencing data using immunostaining in day 60 organoids (Fig. S4.4d).

These changes were also clearly visible when plotting relative abundances for controls and mutants per stage across pseudotime (Fig. 4.3c). In order to gain further insight into how developmental trajectories had changed, we repeated our diffusion pseudotime analysis for both mutant and wildtype cells separately for each stage. Calculating the first three pseudotime components revealed the most likely trajectories for each condition. We found that at the middle time point, day 60, this order was altered in mutant cells. While both excitatory neuron trajectories took precedence over the interneuron trajectory in control cells, it was ranked second in mutant cells, preceding the EN2 lineage (Fig. 4.3d). In summary, these results indicate that *CHD8*$^{+/-}$ mutations lead to a temporal perturbation of neuron specification with interneurons being generated earlier while the production of excitatory neurons was shifted to later time points.

4.4.5 *CHD8* mutations induce cell-type specific transcriptional changes

Next we checked for cell type specific changes at the transcriptomic level by using differential gene expression analysis per cluster. The number of identified differentially expressed genes (DEGs) varied greatly across clusters at each single time point as well as across time within single clusters. At day 20 as well as day 60 we found the most DEGs in the cluster of "interneuron progenitors" (IN_IP) followed by other immature clusters such as early excitatory neurons (ENE), intermediate progenitors (IP) and radial glia (RG3). At day 120 the vast majority of DEGs were found in the ENE cluster. On the contrary, IN_IP and IP showed almost no differential expression at this time point. Interestingly, interneurons (IN), the cell type we found most affected at the population level, almost showed no differential expression either (Fig. 4.4a).

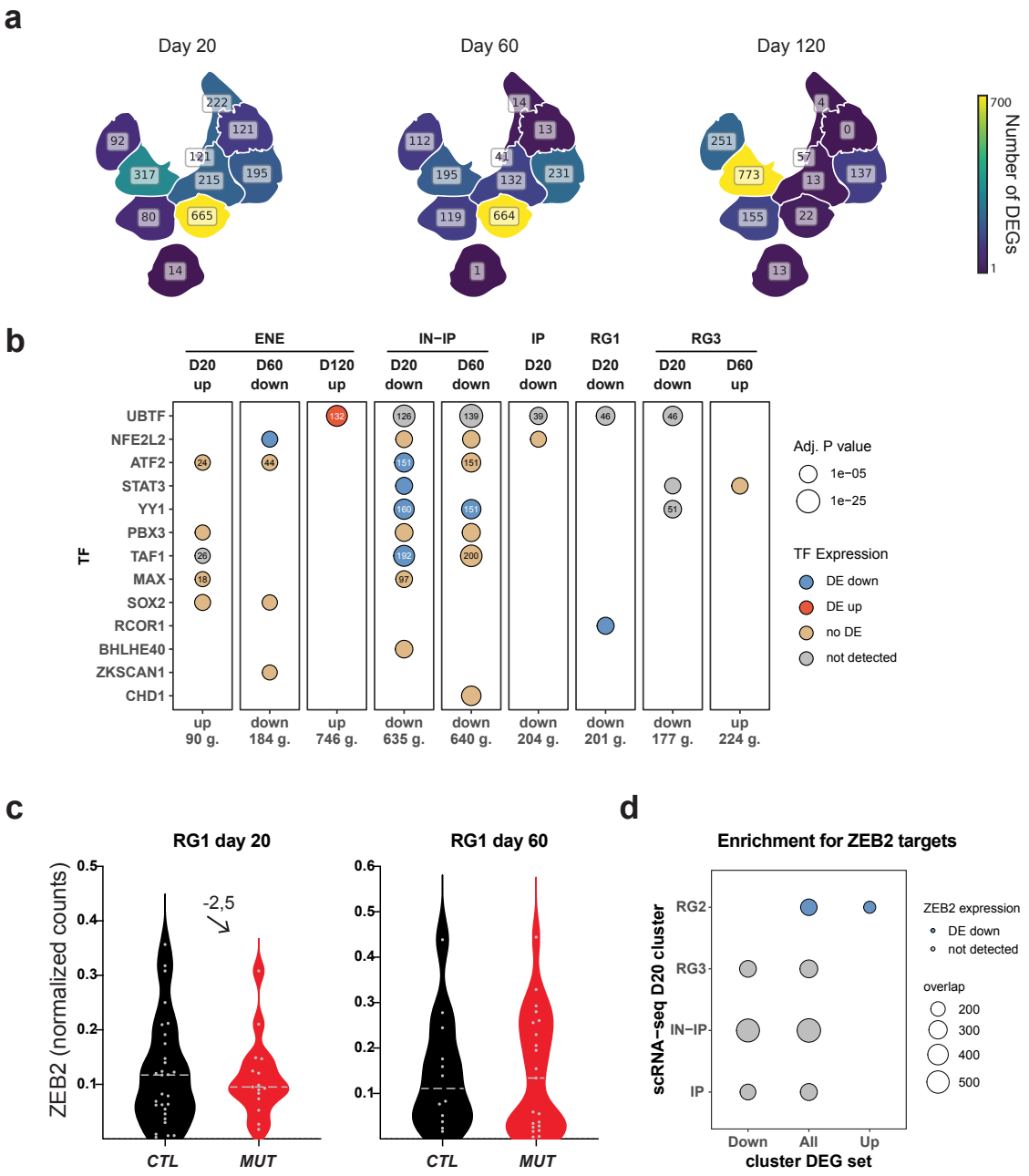


Figure 4.4: **Cell type specific gene expression changes in *CHD8*$^{+/-}$ mutant organoids. a)** Number of differentially expressed genes (false discovery rate (FDR) < 0.05, $|\log2FC| > 1$) between mutants and wildtype per cluster projected onto the UMAP visualization. **b)** Enrichment analysis for transcription factor (TF) targets in DEG sets of progenitor cell clusters. Only significant enrichments (adj. P value ≤ 0.05) are shown. Circle size scales with -log10(adj. P value) of the enrichment (larger size = stronger enrichment) and color indicates differential expression status of the TF gene (red and blue for significant up- or downregulation, respectively; otherwise beige/gray for detected/not detected). Numbers within circles indicate overlap size and are shown where more than 10% of DEGs are targets of the respective TF. **c)** Gene expression levels for *ZEB2* expressed as normalized counts per cell in *CHD8*$^{+/-}$ and *CHD8*$^{+/+}$ cells of the RG1 cluster at day 20 and day 60. log2 fold change of reduction at day 20 is indicated. **d)** Enrichment of ZEB2 target genes in single cell RNA-seq DEGs at day 20 (no FC cutoff) per cluster (y-axis) and split into total, downregulated, and upregulated genes (x-axis). Only significant overlaps are shown (Fisher's exact test, adjusted P value ≤ 0.05). Figure partially adapted from [168].

To gain insight into the nature of these gene expression changes we looked for enriched Gene Ontology terms focusing on the most vulnerable clusters and time points. For the interneuron progenitor cluster at day 20 this revealed a downregulation of genes linked to chromatin organization such as histone 3 lysine 9 and DNA methylation as well as genes connected with the MAPK cascade. Concurrently, upregulated genes were mainly enriched for terms connected to cell cycle regulation. These results were partially recapitulated at day 60 which showed similar enrichment for chromatin organization in downregulated genes in addition to genes connected to Wnt signaling, a pathway linked to CHD8 in multiple studies [120, 121, 123]. Cell cycle genes were enriched in downregulated genes at this time point. Overall, changes in this cluster were the most conserved between day 20 and 60 with 30% of DEGs identified at either stage shared between the two sets and dysregulated in the same direction. The second most affected cluster, ENE, showed distinct differences in terms related to neuronal maturation at day 20. While genes connected with synaptic function were enriched in downregulated genes, upregulated genes were associated with more general neuron differentiation terms. Cell cycle terms were restricted to certain transitions with G1/S and G2/M enriched in upregulated and downregulated genes, respectively. Of note, day 120 showed contrasting results with synaptic function terms enriched in upregulated genes.

In order to potentially link these gene expression changes to specific regulators we looked for enrichment of transcription factor targets for each DEG set. For this analysis we focused on progenitor or immature populations and selected only factors detected as expressed in at least one of the clusters. The results revealed several cases of target enrichment where the respective factor was differentially expressed in that cluster as well (Fig. 4.4b). Of note was the enrichment of targets of the downregulated factor *RCOR1* in the RG1 cluster at day 20 given its role in balancing proliferation and differentiation together with INSM1 [319], which we also found downregulated in this cluster. In addition, two factors connected with intellectual disability, YY1 and TAF1 [320, 321], showed the same pattern of target enrichment and decreased transcription factor expression in the IN-IP cluster at both day 20 and 60.

In addition, we also looked for potentially missed transcription factors not included in the resource of the previous analysis. Since radial glia clusters yielded the least results in the previous analysis, we focused on clusters RG1-3 and checked the top 20 differentially expressed genes for potentially relevant factors. The gene that caught our eye was *ZEB2* which was strongly downregulated (log2(fold change) < -2) in both RG1 and RG2 *CHD8*$^{+/-}$ cells at day 20 but not later time points (Fig. 4.4c). The *ZEB2* gene encodes a known regulator of the epithelial-mesenchymal transition which had recently been linked to regulating the switch from neuroepithelial to radial glia cells [156]. Using a dataset of ZEB2 targets we found enrichment in upregulated genes in the actively proliferating cluster RG2. We identified downregulated genes in early progenitor clusters IP and IN-IP as well as the radial glia cluster RG3 to also be enriched for ZEB2 targets. It should be noted, however, that *ZEB2* expression itself was not detected in these clusters. Overall these results suggest that *CHD8*$^{+/-}$ mutations primarily affect differentiation by perturbing gene expression in early progenitor stages while only marginally affecting mature neuronal populations. Functional enrichment largely supports the observed differences in trajectories by suggesting altered maturation states of progenitor or immature stages. Additionally, perturbed expression of transcription factors and their targets was also most prevalent in these clusters.

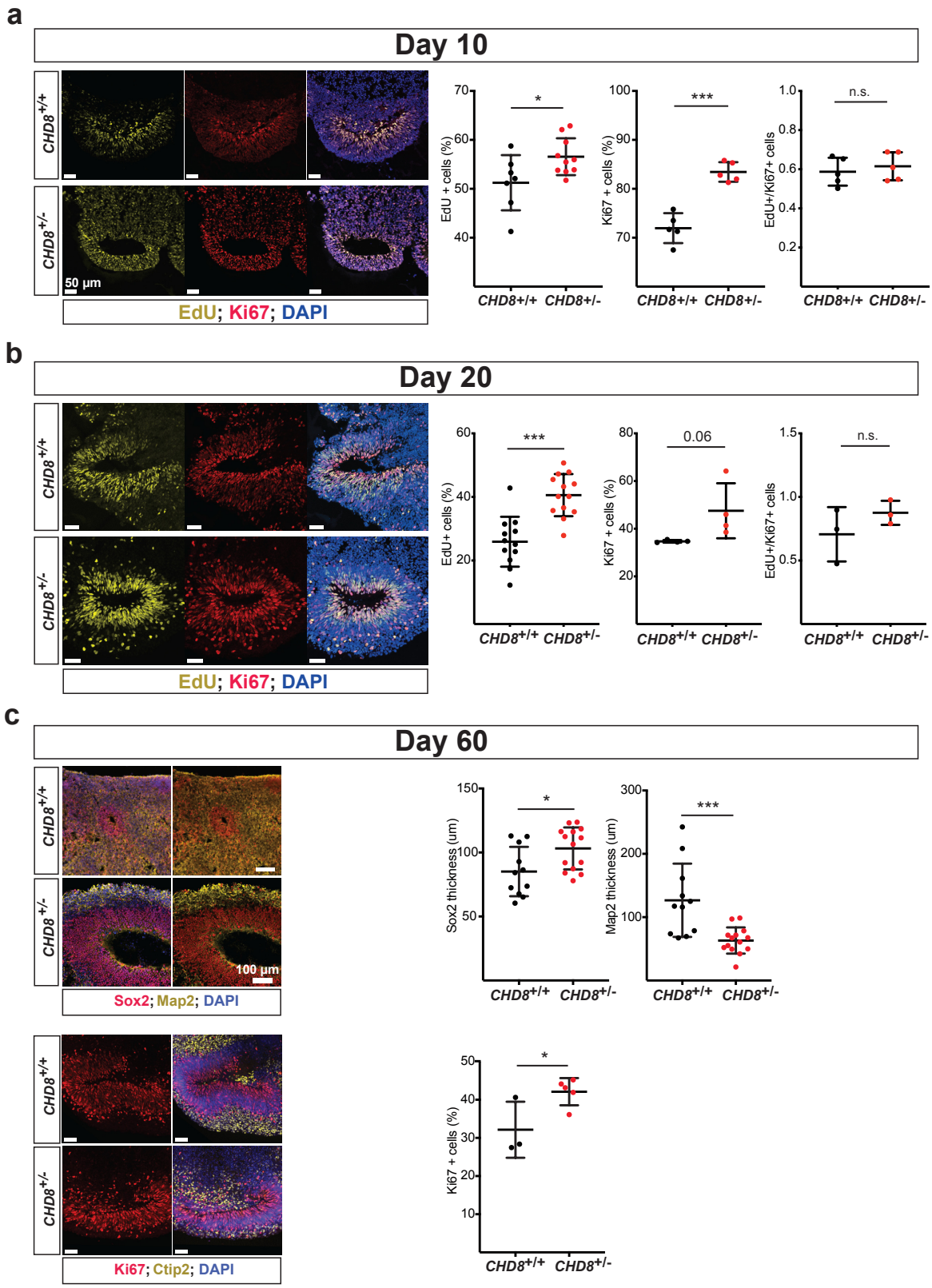


Figure 4.5: ***CHD8$^{+/-}$* organoids show increased and prolonged proliferation. a-b)** Representative images (left) and quantification of cell numbers (right) of immunostainings for Ki67 (red), EdU (yellow), and DAPI in ventricular zone like rosette structures of day 10 (a) and 20 (b) *CHD8*$^{+/+}$ and *CHD8*$^{+/-}$ organoids. EdU was imaged 16 hours after a 1 hour pulse. EdU and Ki67 positive cell counts are shown as % of DAPI counts. n(day 10): EdU *CHD8*$^{+/+}$ = 7, *CHD8*$^{+/-}$ = 10; Ki67 *CHD8*$^{+/+}$ = 5, *CHD8*$^{+/-}$ = 5;

n(day 20): EdU *CHD8*$^{+/+}$ = 12, *CHD8*$^{+/-}$ = 13; Ki67 *CHD8*$^{+/+}$ = 4, *CHD8*$^{+/-}$ = 4. **c)** Representative images (left) and quantifications (right) of immunostainings in ventricular zone like rosette structures of day 60 *CHD8*$^{+/+}$ and *CHD8*$^{+/-}$ organoids. Stainings are shown for Ki67/CTIP2/DAPI (red/yellow/blue, top) and SOX2/MAP2/DAPI (red/yellow/blue, bottom). Quantifications are given for number of Ki67+ cells, as well as thickness of SOX2 and MAP2 positive layers in mm. n(Ki67): *CHD8*$^{+/+}$ = 3, *CHD8*$^{+/-}$ = 5; n(SOX2/MAP2): *CHD8*$^{+/+}$ = 11, *CHD8*$^{+/-}$ = 13. Unpaired t-test, *P < 0.01, ***P < 0.001. n = number of organoids; multiple rosette structure per organoid; at least two batches. Results presented as mean ±s.d. Scale bars are 100 µm for SOX2/MAP2 in c), 50 µm otherwise. Figure adapted from [168].

4.4.6 *CHD8* organoids show sustained proliferation across development

In order to validate our findings, we investigated proliferation at early stages of development given the enrichment for related terms in immature populations. We found a statistically significant increase in proliferating cells (Ki67+) in *CHD8*$^{+/-}$ organoids at day 10 and a non-significant increase at day 20 (Fig. 4.5a-b). In addition, we used an 5-Ethynyl-20-deoxyuridine (EdU)-incorporation assay treating organoids for 1 hour with EdU and assessing its integration 16 hours later. In line with the results observed for Ki67 we found significant increases in EdU positive cells at both time points. However, the relative proportion of EdU+ cells within the pool of proliferating cells (Ki67+) was unchanged, indicating no change in the cell cycle itself.

We next assayed *CHD8*$^{+/-}$ and *CHD8*$^{+/+}$ organoids at day 60 and observed that the increase in proliferative Ki67+ cells was still present at this more advanced time point (Fig. 4.5c). In addition we also confirmed the relative decrease in excitatory neurons at this time using the decreased thickness of the Map2 positive cell layer around rosettes in *CHD8*$^{+/-}$ organoids as a proxy. Conversely, the ventricular zone like Sox2 positive proliferative section of the rosettes was increased (Fig. 4.5c).

Taken together these results indicate a prolonged proliferative phase of progenitors at the expense of neuron generation. This finding is in line with the observed delayed overgrowth phenotype given that *CHD8*$^{+/-}$ end up with more overall progenitors which then, with some delay, go on to produce a higher number of neurons giving rise to *CHD8*$^{+/-}$ organoids eventually outgrowing their *CHD8*$^{+/+}$ counterparts.

4.4.7 Early dysregulation of proliferative genes in *CHD8* mutants

Both single cell RNA-sequencing and immunostainings indicated a potential dysregulation of the balance between differentiation and proliferation in early development. Therefore, we extended the scope of our transcriptional investigation to an earlier time point and performed bulk RNA-seq of *CHD8*$^{+/-}$ and *CHD8*$^{+/+}$ organoids 10 days after embedding. We found 868 genes to be significantly differentially expressed with 441 being up- and 427 downregulated in *CHD8*$^{+/-}$ samples (Fig. 4.6a). Of note, we did not find *ZEB2* differentially expressed in this dataset. Overall, fold changes were small which would fit with either a role in transcriptional modulation or subtle changes in cell type composition.

Functional enrichment analysis revealed an interesting dichotomy based on the direction of change. Gene Ontology terms related to neurogenesis were exclusively found in downregulated genes while cell cycle terms dominated the results for upregulated genes

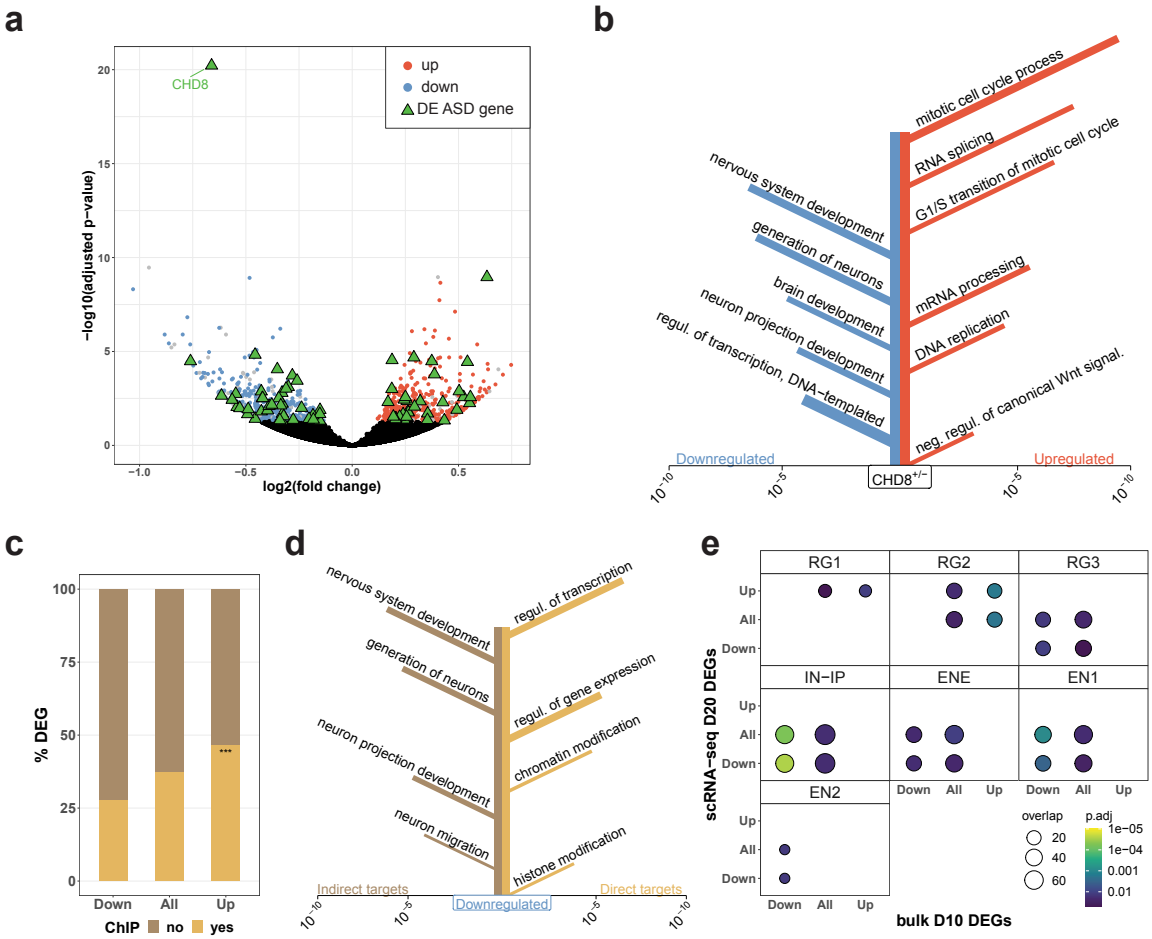


Figure 4.6: **Early transcriptional changes in *CHD8*$^{+/-}$ organoids indicate increased proliferation. a)** Differential expression results for *CHD8*$^{+/-}$ and *CHD8*$^{+/+}$ cerebral organoid samples at day 10 after embedding. Significantly changed genes (adjusted P value ≤ 0.05, 868 genes) are shown in blue and red for down- and upregulation in the mutant, respectively. DEGs filtered because of significant difference between wild-type controls are in gray and green triangles highlight ASD risk genes based on the Simon's Foundation Autism Research Initiative (SFARI) list. **b)** Biological processes (Gene Ontology) enriched in genes downregulated (left) and upregulated (right) in the mutant. The length of the branches scales with significance of the enrichment (-log10(P value)) and branch thickness with number of DEGs linked to a term. **c)** Relative overlap between DEGs, separated by direction of change, with CHD8 targets identified by ChIP [76]. Overlap with upregulated genes is significant (Fisher's exact test, *** $P < 0.001$). **d)** Same as b) but for downregulated genes split into indirect (left) and direct targets (right). **e)** Results for enrichment test of bulk day 10 DEGs (x-axis) across day 20 DEGs per cluster in our single cell dataset (adjusted P value ≤ 0.05, no fold change cutoff). For each DEG set the whole list as well as just up- and downregulated genes were used. Only significant overlaps are shown (Fisher's exact test, adjusted P value ≤ 0.05). Figure adapted from [168].

alongside terms connected to gene regulation. Especially the transition between G1 and S phase was singled out within this set of cell cycle terms. We also found multiple genes associated with the negative regulation of Wnt-signaling to have increased expression in *CHD8*$^{+/-}$ samples (Fig. 4.6b). When repeating our transcription factor target analysis on this dataset we found a large disparity between up- and downregulated DEGs. While the former were enriched for targets of a sizable number of transcription factors, the latter barely showed any such result, despite almost identical size of the two sets (Table 4.1). Among TF for upregulated genes we could confirm 11/13 transcription factors reported in Fig. 4.4b. Additionally, we also recapitulated the enrichment for ZEB2 targets in

upregulated genes of this time point (Adj. P value = $5 * 10^{-5}$), further strengthening this link.

We also observed enrichment for ASD risk genes in these DEG sets (Fig. 4.6a, adjusted P value: $8.3 * 10^{-5}$, $1.6 * 10^{-4}$, 0.04 for all, up-, and downregulated DEGs, respectively). Among these 71 genes we found two genes identified via ChIP as direct targets of CHD8, the NCoR complex member *TBLX1R1* (log2 fold change: -0.15) [231, 322] and *ARID1B* (log2 fc: 0.26), a component of SWI/SNF complexes and associated with Coffin-Siris syndrome [323].

When looking at the proportion of DEGs directly bound by CHD8 we observed a total of 325 (37%) genes falling into this category. In fact, the overlap between CHD8 targets and upregulated genes (206, 47%) was statistically significant (Fig. 4.6c). We next investigated whether segregating DEGs based on CHD8 binding would likewise segregate the functional enrichment and highlight specific roles for direct over indirect targets. While upregulated showed little distinction with cell cycle enrichment either way, the results for downregulated genes painted a different picture. Here, the direct targets were enriched for terms associated with gene regulation and histone modification. Indirect targets, on the other hand, were enriched for neuronal differentiation terms (Fig. 4.6d). Furthermore, restricting our DEGs to just direct CHD8 targets improved the enrichment for ZEB2 targets as well (Adj. P value = $6.6 * 10^{-9}$ for upregulated genes) while this enrichment was absent for indirect target DEGs. Overall, these results highlight a role for CHD8 as a potential key regulator of chromatin and transcriptional regulation. On the other hand, the observed indirect effects, with their split between neuronal differentiation and cell cycle, might reflect relative changes in cell composition.

Given that this dataset represented an extension of our single cell results towards earlier development we next investigated how strongly day 10 bulk gene expression changes were shared with specific cell populations at day 20. We found significant overlap with DEG sets from multiple clusters. Of interest, genes upregulated at day 10 were only enriched in proliferative clusters RG1 and RG2, representing G2/M and S phase, respectively, where they were likewise upregulated. On the other hand, genes downregulated at day 10 were enriched in multiple clusters, most strongly in IN-IP and EN1 and within genes with reduced expression in these clusters (Fig. 4.6e). Notably, the genes shared between IN-IP and bulk downregulated genes were again enriched for chromatin remodeling genes while such genes shared between ENE or EN1 and bulk were enriched for terms related to neurogenesis instead.

Overall, these results indicate significant dysregulation of cerebral organoid development already at 10 days after embedding. Notably, we observed an imbalance between proliferation and differentiation in line with our findings from immunostainings which might set the stage for the changed lineage specification dynamics throughout development.

DEG set	TF set	TF status	Overlap	P value	Adj. P value	Perm. P value
up	CHD1 ENCODE	DE up	40/655	6.7e-09	1.2e-07	0.000
up	GABPA ENCODE	DE up	68/2082	6.2e-04	2.3e-03	0.006
up	E2F4 ENCODE	no DE	59/710	1.2e-18	1.3e-16	0.000
up	E2F6 ENCODE	no DE	129/3245	3.1e-12	1.6e-10	0.000
up	TAF1 ENCODE	no DE	126/3346	2.5e-10	8.6e-09	0.000
up	MAX ENCODE	no DE	86/2073	5.3e-09	1.1e-07	0.000
up	E2F1 CHEA	no DE	45/859	7.4e-08	1.1e-06	0.000
up	SOX2 CHEA	no DE	41/775	2.3e-07	2.9e-06	0.000
up	ATF2 ENCODE	no DE	102/2852	3.2e-07	3.6e-06	0.000
up	NFYB ENCODE	no DE	123/3715	8.7e-07	9e-06	0.000
up	BRCA1 ENCODE	no DE	106/3218	8.5e-06	7.1e-05	0.003
up	SIN3A ENCODE	no DE	48/1131	1.1e-05	8e-05	0.000
up	ZNF384 ENCODE	no DE	35/730	1.5e-05	1.1e-04	0.000
up	USF2 ENCODE	no DE	41/965	4.8e-05	2.9e-04	0.000
up	NELFE ENCODE	no DE	16/234	6.8e-05	3.9e-04	0.000
up	NFYA ENCODE	no DE	76/2250	9.9e-05	5.4e-04	0.004
up	UBTF ENCODE	no DE	58/1631	2e-04	9.8e-04	0.004
up	PML ENCODE	no DE	57/1596	2e-04	9.8e-04	0.000
up	TCF3 ENCODE	no DE	35/840	2.5e-04	1.1e-03	0.002
up	NRF1 ENCODE	no DE	64/1882	3.2e-04	1.3e-03	0.003
up	ZMIZ1 ENCODE	no DE	36/914	5.8e-04	2.2e-03	0.000
up	TCF3 CHEA	no DE	38/1006	9e-04	3.2e-03	0.002
up	BHLHE40 ENCODE	no DE	17/348	2e-03	6.8e-03	0.004
up	TP53 CHEA	no DE	16/319	2e-03	6.8e-03	0.001
up	SALL4 CHEA	no DE	17/355	2.4e-03	7.8e-03	0.000
up	SMAD4 CHEA	no DE	24/584	2.7e-03	8.6e-03	0.000
up	KAT2A ENCODE	no DE	7/90	3.8e-03	1.1e-02	0.000
up	YY1 ENCODE	no DE	81/2753	3.8e-03	1.1e-02	0.023
up	EGR1 ENCODE	no DE	22/537	4.2e-03	1.2e-02	0.002
up	SP1 ENCODE	no DE	27/707	4.2e-03	1.2e-02	0.004
up	NFE2L2 CHEA	no DE	35/1022	6.8e-03	1.8e-02	0.000
up	ZBTB33 ENCODE	no DE	26/702	7.2e-03	1.8e-02	0.003
up	CTCF ENCODE	no DE	55/1790	7.5e-03	1.8e-02	0.001
up	BCLAF1 ENCODE	no DE	30/851	8e-03	1.9e-02	0.006
up	ELF1 ENCODE	no DE	72/2483	8.9e-03	2.1e-02	0.037
up	ZEB1 ENCODE	no DE	7/106	9.1e-03	2.1e-02	0.002
up	ZBTB7A ENCODE	no DE	64/2184	1.1e-02	2.5e-02	0.024
down	TCF3 CHEA	no DE	39/1006	2.5e-04	2.6e-02	0.015
up	RCOR1 ENCODE	no DE	25/702	1.3e-02	2.8e-02	0.011
up	PBX3 ENCODE	no DE	40/1269	1.5e-02	3.1e-02	0.022
up	TAF7 ENCODE	no DE	23/640	1.5e-02	3.2e-02	0.003
up	FOXA1 ENCODE	no DE	10/205	1.6e-02	3.2e-02	0.002
down	AR CHEA	no DE	40/1095	6.7e-04	3.5e-02	0.000

Table 4.1: **Transcription factor target enrichment in CHD8 bulk DEGs**

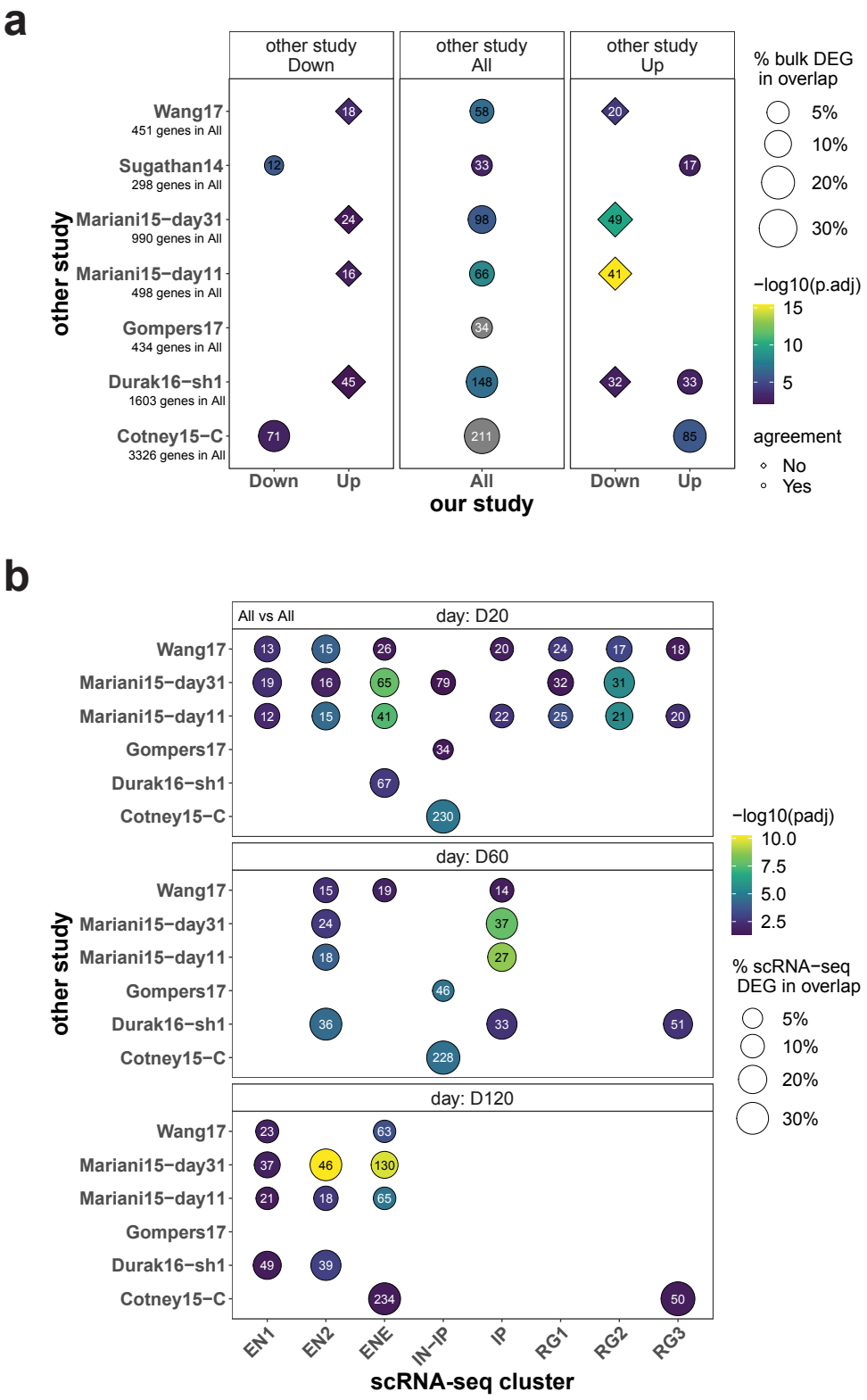
4.5 Discussion

CHD8 is among the genes most commonly found to be mutated in patients with ASD. For this study we chose to model heterozygous loss of this gene in a human model system of early cortical development. We decided on an isogenic approach by generating multiple $CHD8^{+/-}$ mutant lines from the same H9 origin cell line. This allowed us to avoid potential confounding effects from varying genetic backgrounds as is more common when using patient-specific induced pluripotent stem cell (iPSC) lines [324]. Using this approach we modeled *CHD8* haploinsufficiency in three different ways. First, we disrupted the gene within its catalytically important helicase domain [325] via introducing a deletion or inserting a GFP cassette ($CHD8^{+/-}$ and $CHD8^{+/GFP}$). Of note, this domain was also reported to be a hotspot for missense variants found in ASD patients [326]. Second, we added a patient-specific aspect by introducing two specific loss of function mutations, S62X and E1114X ($CHD8^{+/S62X}$ and $CHD8^{+/E1114X}$), both of which result in premature stop codons. E1114X is also located within the helicase domain and, as other mutations in this region, has been linked to macrocephaly in patients. Contrastingly, S62X is not connected to macrocephaly [74] which allowed us to shed some light on the genotype-phenotype spectrum of *CHD8* mutations.

Heterozygous loss of *CHD8* alters the proliferation/differentiation balance.

Our first observation was a delayed onset overgrowth of $CHD8^{+/-}$, $CHD8^{+/GFP}$, and $CHD8^{+/E1114X}$ cerebral organoids compared to both wildtype and $CHD8^{+/S62X}$. This was both reflective of the macrocephaly phenotype widely observed in patients [294, 326–328], as well as the specific lack thereof in the patient with the S62X mutation. Given the link between progenitor proliferation and brain overgrowth [329] we hypothesized that the observed growth differences might be caused by an altered balance between sustained proliferation and differentiation. In support of this hypothesis, we report an increased number of proliferative cells in ventricular zone like rosette structures which persisted until later stages. In addition, our bulk RNA-sequencing data supports this shift towards a more proliferative state by showcasing a split between upregulated cell cycle and downregulated neuronal differentiation related genes.

A previously published meta-analysis reported that while differential gene expression results were largely study-specific the consensus on identified genomic targets of CHD8 was much higher [298]. In our analysis we show that inclusion of this information could delineate direct from indirect effects at the functional level. These results could be interpreted as CHD8 directly affecting cell cycle regulation and the promotion of a proliferative state. This in turn led to the indirect effect of reduced neuronal differentiation at this early developmental time point. In order to place these results into the greater context of the *CHD8* literature we compiled differential gene expression results from previously published studies on CHD8 [76, 77, 123, 286, 297], as well as a study on organoids from iPSCs derived from patients with idiopathic ASD and an overgrowth phenotype [151]. We found significant gene-wise overlap with most of the studies in comparisons ignoring direction of change (Fig. 4.5a). However, direction-specific overlap was largely restricted to studies focusing on early progenitors, such as neural progenitor or neural stem cells, which might reflect the early time point we investigated here. Of interest, more direction-specific overlaps were observed when restricting DEGs to direct targets indicating that these changes are more conserved across studies than secondary effects.



Supplementary Figure S4.5: **DEG overlap with published studies. a-b)** Summary of the overlap analysis between our identified differentially expressed genes and results from published studies. The size of each circle scales with the percentage of DEGs in our respective set found in the overlap. Numbers within the circles give the absolute number of overlapping genes. The color indicates the P value of the hypergeometric test corrected for multiple testing. The y-axis lists the studies used in this comparison while the sets from our bulk (a) and single-cell RNA-seq analysis (b) are shown on the x-axis. To provide a frame of reference the number of DEGs reported by each of these studies is given in (a). In (a) we

investigated both complete (All, middle) as well as direction-specific (e.g. down- vs downregulated) overlaps (left and right). For completeness, non-significant overlaps for All vs All comparisons are shown in gray. For (b) only significant overlaps between complete sets are reported. Sources: [76, 77, 123, 151, 286, 297]

In the context of proliferation contra differentiation we specifically highlight a connection between CHD8 haploinsufficiency and ZEB2. This was motivated by a study from 2021 linking ZEB2 function to the neocortical expansion in humans compared to great apes [156]. They identified ZEB2 as a driver of the transition between neuroepithelial and radial glia cells and showed that both its expression as well as the cell morphological transition are delayed in human organoids compared to gorilla and chimpanzee. This led to a prolonged tangential progenitor expansion and subsequent increase in neuroepithelial bud and overall organoid size. ZEB2 can act as either transcriptional activator or repressor [330] which could explain why we see enrichment for its targets in upregulated genes despite ZEB2 itself showing reduced expression.

CHD8$^{+/-}$ mutations affect the temporal dynamics of neuronal fate specification.

Our second main finding were altered subtype dynamics in neurogenesis. Specifically, using scRNA-seq, we observed an accelerated generation of inhibitory relative to excitatory neurons in *CHD8*$^{+/-}$ cerebral organoids. However, leveraging our longitudinal approach we uncovered that this represented only a transient bias as the latest time point conversely showed an increased number of excitatory neurons. Taken together with the shifted proliferation/differentiation balance discussed above, our findings could be interpreted as follows: Heterozygous loss of *CHD8* keeps excitatory neuron progenitors in a proliferative state which delays the production of such neurons but leads to an eventually larger number due to an increased progenitor pool. On the other hand, *CHD8*$^{+/-}$ organoids instead preemptively initiate terminal differentiation of inhibitory neurons. Given the relative nature of cluster abundances as measured by scRNA-seq these hypotheses are speculative to a certain degree but warrant further investigation.

Of note, our main findings in regard to fate specification were largely reproduced by a study published at around the same time [153]. The authors report a similar bias towards the production of inhibitory neurons and also evidence for a relative reduction of excitatory neurons. In addition, this study, which also targeted the helicase domain, could replicate the increased size of *CHD8*$^{+/-}$ organoids. However, since they only reported the results for either scRNA-seq or organoids size for just one time point, no statement can be made about altered temporal dynamics in their system, highlighting the benefits of our longitudinal approach. The study also reports overproduction of interneurons for organoids derived from *ARID1B*$^{+/-}$ iPSCs. Interestingly, a similar phenotype had previously been linked to ASD with macrocephaly in two independent *in vitro* studies [151, 329] hinting at the possibility of a common axis in ASD pathogenesis.

Leveraging the cell-level resolution of scRNA-seq allowed us to uncover the cell-type specific consequences of heterozygous loss of *CHD8*. Our findings reveal distinct temporal as well as cell-type specific patterns of vulnerability with a general bias towards immature populations showing the most DEGs. Remarkably, the identity of the most affected cluster shifted, in line with the temporal lineage specification dynamics, from interneuron progenitors (IN-IP) at day 20 and 60 to early excitatory neurons (ENE) at day 120.

Comparing these results with previous studies we found most overlap with differential expression results derived using organoids highlighting the choice of model system as a key determinant of the observable outcomes (Fig. 4.5b). Overlaps became more and more cluster restricted over the course of development with more ubiquitous enrichment at day 20 and restriction to excitatory neuron clusters at day 120. These results from later time points were also more concordant in terms of direction as they could mostly be reproduced when the analysis considered directionality.

Lastly, we highlight differential effects of *CHD8* haploinsufficiency on these two lineages by means of functional enrichment analysis. This included transcription factors linked to intellectual disability, Yin Yang 1 (YY1) and TATA-box binding protein associated factor 1 (TAF1) [320, 321]. Of further interest, YY1 had been linked to the regulation of GABAergic neuron differentiation [331]. These factors as well as a significant proportion of their targets were downregulated exclusively in the interneuron lineage, providing evidence for the pathological significance of these changes.

Overall, this study provides insight into how a reduction in CHD8, a top autism risk chromatin remodeler, perturbs the neurodevelopmental transcriptional landscape and reshuffles the probabilities of lineage determination. It should be acknowledged that this study has a number of caveats. First and foremost, our single cell RNA sequencing analysis suffers from a low number of biological replicates. Specifically, we only included single samples of pooled organoids per line and time point. While pooling of multiple organoids aids in ameliorating potential issues stemming from variabilities between single organoids it fails to provide sufficient statistical power for detailed analyses. Despite this, the inter-line variability was sufficiently small to make the analysis performed here possible. Secondly, the single-cell experiment is based on organoids derived from a single differentiation batch and hence does not cover this additional source of biological variability. Lastly, our study overall is limited in its scope since it does not contain data describing changes in the chromatin landscape which will be key in understanding the role of CHD8 in neurodevelopment.

Altogether, we see these results as a first stepping stone upon which future studies of larger scope can build. Our study highlights the importance of using a longitudinal approach to capture developmental dynamics. This is particularly paramount for the study of chromatin regulators which act as modulators of dynamic processes involving the transcriptional dimension. Especially under a heterozygous paradigm it is conceivable that what we observe is a shift of probabilities between competing outcomes and we will need to cover as much ground as possible to uncover these effects.

5

Convergence of Causal Paths across Epigenetic Regulators in ASD

5.1 Context

The aim of this study is to tackle the question of regulatory convergence in the pathogenesis of autism spectrum disorder (ASD). It encompasses multiple genes which were selected because of their function in chromatin organization and transcriptional regulation. As mentioned above, genes related with these processes form a major subgroup of autism risk genes. At the same time, we are in many cases still at the beginning in terms of understanding the role these genes play in the context of neurodevelopment. Given that these genes likely regulate a plethora of different targets the task of delineating the key effectors in any given setting becomes non-trivial. In the context of ASD this is further exacerbated by the fact that in most cases patients present with heterozygous mutations leading to a reduction rather than complete loss of the proteins in question. This is where the concept of convergence shines as a way to find the proverbial needles in the haystack.

The design of this study makes use of this concept in that it assumes a common outcome for the selected genes/mutations, i.e. an ASD diagnosis, and at the base level a shared level of action, i.e. chromatin and transcriptional regulation. Our hypothesis is that under these assumptions and employing a model system of the process of interest, i.e. neurodevelopment, any shared phenotype between genes/mutations has a greater likelihood of being relevant in the genesis of the common outcome. Therefore, we employ convergence to distinguish gene-specific from more general effects in an attempt to shed light on how mutations in these genes affect neurodevelopmental processes and eventually led to ASD.

5.2 Introduction

ASD is a group of neurodevelopmental disorders characterized by large genetic heterogeneity. A large number of risk genes have been identified via findings of increased mutational burden in ASD patients compared to controls [45–56]. These studies also revealed that the identified genes broadly converge at the level of protein function onto transcriptional regulation and synaptic function. Additionally, through the use of co-expression networks a spatiotemporal convergence onto mid-fetal cortical development was found [67, 68]. Further evidence for shared transcriptional regulation, henceforth referred to as regulatory convergence, came from studies on post-mortem brain samples from ASD patients and controls [80–90]. The promise of convergent mechanisms is that it would allow a sub-categorization of the vastly heterogeneous set of ASD based on biological rather than behavioral criteria which could potentially open up new avenues for treatment options in severe cases. However, we are only beginning to understand direct connections between genotype, i.e. risk genes, and gene regulatory phenotypes. Previous studies largely used either shRNA knockdown [332], *in vitro* clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 screens [333], *in vivo* CRISPR/Cas9 screen methodologies using the Pertub-seq method [334, 335], or human pluripotent stem cell lines generated using the CRISPR/Cas9 system [153, 336, 337]. One of the most recent of these studies selected three ASD risk genes, *CHD8*, *ARID1B*, and *KMT5B*, and investigated the effect of heterozygous loss using cerebral organoids [153]. Focusing on a single time point, they reported changes in cell fate determination leading to an overproduction of inhibitory neurons to varying degrees dependent on genomic context. However, they found little overlap at the level of cell-type specific transcription, suggesting different regulatory pathways at play.

A large proportion of ASD risk genes connected to transcriptional regulation encode for proteins involved in the modification of chromatin, such as histone modifiers or chromatin remodeling enzymes [55]. Given the finding of shared changes at the epigenomic level in ASD patients [85] these make for prime targets for the study of regulatory convergence. For this study we select four genes encoding for proteins involved in histone modification: ASH1-Like Histone Lysine Methyltransferase (*ASH1L*/*KMT2H*), Lysine Demethylase 6B (*KDM6B*/*JMJD3*), SET Domain Containing 5 (*SETD5*), and Lysine Methyltransferase 5B (*KMT5B*/*SUV420H1*). ASH1L is a Trithorax histone methyltransferase catalyzing the formation of mono- or di-methylated histone 3 lysine 36 (H3K36) [338, 339] and has been linked to activity dependent transcription [340] and the counteracting of Polycomb silencing [341]. Its loss caused neuronal hyperactivity and various behavioral changes in mouse models [342, 343]. KDM6B is a demethylase targeting the repressive histone 3 lysine 27 (H3K27) mark [344] and has been shown to play a role in neuroectoderm and neural progenitor differentiation [345, 346]. KMT5B is a methyltransferase that was shown to target mono-methylated lysine 20 on histone 4 (H4K20) catalyzing its dimethylation [347, 348]. It was also demonstrated to play a role in neuroectoderm differentiation [349]. Lastly, SETD5 was initially proposed to be a histone methyltransferase but mounting evidence produced by our group as well as others suggests that it possesses no such ability. Rather, it likely exerts its regulatory function through interactions with other complexes, such as the NCoR complex, linked to histone deacetylation [167, 258].

Here, we investigate the effect of haploinsufficiency of these four genes on human neurodevelopment. For this reason, we established heterozygous knockout lines on an isogenic H9 human embryonic stem cell (hESC) background and modeled early human cortical

development using the cortical spheroid model system which largely enforces a dorsal excitatory neuron fate [133, 144]. We found a remarkable propensity for *ASH1L*$^{+/-}$, *KDM6B*$^{+/-}$, and *SETD5*$^{+/-}$ lines but not *KMT5B*$^{+/-}$ lines to partially escape this fate and instead produce cells resembling the ventral forebrain, in particular progenitors of the lateral ganglionic eminence.

5.3 Methods

Human embryonic stem cells

Human H9 human embryonic stem cells (hESCs) were acquired from WiCell (WA09) at passage 24. The cells were cultured on Matrigel (Corning, either growth-factor reduced #356230 or hESC-qualified #354277) using either mTeSR™ 1 (Stemcell Technologies #85850) or StemFlex complete medium (Gibco A3349401) for maintenance and Essential 8 complete medium (Gibco A1517001) prior to cortical spheroid generation. Passaging was done using 0.5 mM EDTA (pH 8.0) in phosphate buffered saline (PBS, without magnesium or calcium chloride) supplemented with 1.8% sodium chloride (0.9 g per 500 ml) adapting a previously published method [350]. In short, cells were washed with 1X PBS and then incubated in EDTA/NaCl/PBS for 2-3 minutes at room temperature. EDTA was removed, cells resuspended in fresh culture medium and split to new matrigel coated wells. All lines used in this study were derived from this cell line.

Plasmids

Cloning for the generation of the WT-GFP line done by Aysan Çerağ Yahya and Roberto Sacco

Plasmids containing guide RNA (gRNA) sequences under the control of a U6 promoter as well as Cas9 from *Streptococcus pyogenes* were generated by cloning the respective gRNA sequences into the pSpCas9(BB)-2A-Puro (PX459) V2.0 plasmid (Addgene 62988, Feng Zhang lab) as described in the protocol by Ran et al. [351]. In short, top and bottom gRNA oligonucleotides were annealed and phosphorylated using T4 polynucleotide kinase (NEB, M0201) by incubation for 30 min at 37°C, followed by 5 min at 95 °C. The product was diluted 1:200 and used as input for the ligation reaction: 100 ng plasmid, 2 µL diluted oligo, 2 µL Tango buffer (Thermo, BY5), 1 µL 10 mM DTT (Thermo, R0861), 1 µL 10 mM ATP (NEB, P0756), 1 µL FastDigest BpiI (Thermo, FD1014), 0.5 µL T7 ligase (NEB, M0318). Ligation was done through 6 cycles of 5 min 37°C/5 min 21°C. Ligated plasmids were transformed into competent Stbl3 cells via heat shock, and plasmids from colony PCR positive clones were purified using the Qiagen Spin Miniprep Kit (Qiagen #27104). gRNA sequences (see Table 5.1 g1a-g6b) were designed using the CRISPOR[1] tool [352]. The plasmid used as donor for the template sequence needed for the insertion of the Green Fluorescent Protein (GFP) cassette using homology directed repair was derived from the CHD8_E19KO_PGK-Neo_pCAG-GFP plasmid without the PGK-Neo cassette (henceforth referred to as HDR-base plasmid) previously generated in our group [168]. First homology arms including overhangs overlapping GFP cassette and plasmid backbone were synthesized by PCR (primers see Table 5.1 h1-24) using the CloneAmp™ HiFi PCR Premix (Takara Bio #639298) resolved on an agarose gel and

[1] http://crispor.tefor.net/

purified with the Promega Wizard® SV Gel and PCR Clean-Up System (A9281). The HDR-base plasmid was digested with restriction enzymes HindIII HF and NsiI HF (NEB R3104 and R3127), the two fragments of sizes 2.8 and 3.8 kb separated using agarose gel electrophoresis and purified the same way as the homology arms. The two fragments were ligated with the left and right homology arms in a single reaction using the In-Fusion® HD Cloning Kit (Takara Bio #639649) with a molar ratio between backbone, GFP cassette, left and right homology arm of 1:1:2:2 in a reaction volume of 20 µL. In-Fusion reactions were transformed into Stbl3 cells, and the plasmids purified using the Qiagen Spin Miniprep Kit (Qiagen #27104). Correct ligation was confirmed by Sanger sequencing. Plasmid preparations for transfections were done using the Nucleobond Xtra Midi kit (Macherey-Nagel #740410.50) followed by Trizol/Chloroform extraction.

Generation of hESC heterozygous knockout lines

WT-GFP line generated by Aysan Çerağ Yahya and Roberto Sacco

For transfections H9 hESCs at a passage < 30 were dissociated into single cells by treatment with StemPro™ Accutase™ cell dissociation reagent (Gibco, #A1110501) for 4 minutes at 37°C, followed by resuspension in either mTeSR1 containing 10 µM Y-27632 Rock inhibitor (Stemcell Technologies #72304) for KMT5B SET lines or StemFlex complete medium containing 1X RevitaCell™ supplement (Gibco, #A2644501) for all other lines. Cells were counted using an automated cell counter (Biorad Tc-20). 1 million live cells were used per transfection and resuspended in 82 µL Nucleofection solution 1 and 18 µL Supplement 1 from the Human Stem Cell Nucleofector™ Kit 1 (Lonza, VPH-5012) containing 7.5 µg HDR plasmid and 2.5 µg gRNA/Cas9 plasmid DNA. Nucleofection was done with a Nucleofector™ 2b Device (Lonza, AAB-1001) using program A-023. Cells were afterwards plated spread over 6 wells of a 6-well plate in either mTeSR1/Y-27632 or StemFlex/RevitaCell and left to recover for 72h before medium was changed to mTeSR1/StemFlex complete. Subsequently cells originating from different wells after transfection were treated as independent clones.

For selection of edited clones a fluorescence activated cell sorting (FACS) based strategy was used. In detail, transfected cells were grown until nearly confluent, dissociated into single cells as described above, and GFP positive singlet live cells were sorted using a BD FACSAria flow cytometer (100 µm nozzle). Propidium iodide was used as a viability dye. Sorted cells were plated on rhLaminin-521 (0.5 µg /cm^2) at low density (300 cells per well of a 6-well plate) in StemFlex complete medium containing 1X RevitaCell, 100 U/ml Penicillin and 100 µg /ml Streptomycin (Gibco, #15140-122). Cells were left to recover for 72-96h and maintained in StemFlex complete medium afterwards. Single colonies with homogeneous GFP expression were picked using an EVOS FL imaging system (Thermo Fisher), screened for insertion of the cassette by PCR, and expanded. To verify the correct genotype both junctions between cassette and genomic DNA were checked by Sanger sequencing. To confirm heterozygosity, the region flanking the gRNA target site on the WT allele was also sequenced (primers see Table 5.1 gt1-19). The absence of random integration was verified by Southern blot (see below). Regions containing potential exonic off-target sites, as predicted by CRISPOR to have a CFD score > 0.02 (cutoff based on [353]), were amplified by PCR (primers see Table 5.1 ot1-20) and checked by Sanger sequencing.

ID	Sequence	Description
g1a	CACCgTGTGTGGATGCCCGTACTTT	gRNA SETD5 SET domain - FWD for cloning
g1b	AAACAAAGTACGGGCATCCACACAc	gRNA SETD5 SET domain - REV for cloning
g2a	CACCgTCAGTGTCATGTACTCCACA	gRNA KMT5B SET domain - FWD for cloning
g2b	AAACTGTGGAGTACATGACACTGAc	gRNA KMT5B SET domain - REV for cloning
g3a	CACCgTACCACAGCGCCCTTCGATA	gRNA KDM6B coding exon 3 - FWD for cloning
g3b	AAACTATCGAAGGGCGCTGTGGTAc	gRNA KDM6B coding exon 3 - REV for cloning
g4a	CACCgCCTATGACCAACGTTCAAGT	gRNA ASH1L coding exon 2 - FWD for cloning
g4b	AAACACTTGAACGTTGGTCATAGGc	gRNA ASH1L coding exon 2 - REV for cloning
g5a	CACCgTAGACTTCATCGGCGGAAGC	gRNA SETD5 coding exon 4 - FWD for cloning
g5b	AAACGCTTCCGCCGATGAAGTCTAc	gRNA SETD5 coding exon 4 - REV for cloning
g6a	AAACATCCTGTCCCTAGTGGCCCC	gRNA AAVS1 locus - FWD for cloning
g6b	CACCGGGGCCACTAGGGACAGGAT	gRNA AAVS1 locus - REV for cloning
h1	TGATTACGCCAAGCTCTGGGGAGGCAATTTCACTAGATGT	SETD5 SET domain - left homology arm - FWD
h2	TAGGAACTTCAAGCTCATCCACACACATCTCTACACCATTGA	SETD5 SET domain - left homology arm - REV
h3	AGTATAGGAACTTCATAATGATGCTCGGTTCATCAGAAGATCAT	SETD5 SET domain - right homology arm - FWD
h4	ATTGGGCCCTCTAGAGGCGTGATCTCGGCTCAC	SETD5 SET domain - right homology arm - REV
h5	TGATTACGCCAAGCTAAAAAAAATGAGCTAGGTGTGGTGGC	KMT5B SET domain - left homology arm - FWD
h6	TAGGAACTTCAAGCTATGACACTGAAGTCGTTTTCTCCATG	KMT5B SET domain - left homology arm - REV
h7	AGTATAGGAACTTCACTGTGCTCAACTCTGGCTG	KMT5B SET domain - right homology arm - FWD
h8	ATTGGGCCCTCTAGAATATTAAGATTTTAAAGGGAAATTTTAAATTTATTGAGGGACATTCC	KMT5B SET domain - right homology arm - REV
h9	TGATTACGCCAAGCTCCACTGCTGACCTGGTAAGG	KDM6B coding exon 3 - left homology arm - FWD
h10	TAGGAACTTCAAGCTAGGGCGCTGTGGTAGC	KDM6B coding exon 3 - left homology arm - REV
h11	AGTATAGGAACTTCAAAGCTTCGCTGAGCTGGG	KDM6B coding exon 3 - right homology arm - FWD
h12	ATTGGGCCCTCTAGATGTAGGTGAGAGGTGAAAGGCA	KDM6B coding exon 3 - right homology arm - REV
h13	TGATTACGCCAAGCTAGGAATTGGTATTGTTCCAGGTT	ASH1L coding exon 2 - left homology arm - FWD
h14	TAGGAACTTCAAGCTTACTTTCAGAAGTAAACTGGTTCTTTCCAACAG	ASH1L coding exon 2 - left homology arm - REV
h15	AGTATAGGAACTTCATCATAGGTCAGTTGGTCATAGTATAAGTATTGAATG	ASH1L coding exon 2 - right homology arm - FWD
h16	ATTGGGCCCTCTAGAGACCTTGTTTGAAGGAAGGGATTTC	ASH1L coding exon 2 - right homology arm - REV
h17	TGATTACGCCAAGCTCCCTTTTGCACTTCCCTGACTCCA	SETD5 coding exon 4 - left homology arm - FWD
h18	TAGGAACTTCAAGCTGATGAAGTCTAATAACCTTCCCCCTGCT	SETD5 coding exon 4 - left homology arm - REV
h19	AGTATAGGAACTTCAACAACATATCAGGTGAGCGGAAG	SETD5 coding exon 4 - right homology arm - FWD
h20	ATTGGGCCCTCTAGATAGCTTATTCTTTTAGGTGTGCCCC	SETD5 coding exon 4 - right homology arm - REV
h21	TGATTACGCCAAGCTTGCTTTCTCTGACCAGCATTCTCTCC	AAVS1 locus - left homology arm - FWD
h22	TAGGAACTTCAAGCTAAGCTTAGTACTGTCGACCCATAGAG	AAVS1 locus - left homology arm - REV
h23	AGTATAGGAACTTCAACTAGGGACAGGATTGGTGACAG	AAVS1 locus - right homology arm - FWD
h24	ATTGGGCCCTCTAGAAGAGCAGAGCCAGGAACCCCTG	AAVS1 locus - right homology arm - REV
gt1	CCGGAACCCTTAATATAACTTCGT	general genotyping primer for left HA side; inside cassette
gt2	AGAGAGAGAGATGACTTAGTATGGG	left HA side genotyping primer for SETD5 SET domain
gt3	AGAAGCGCTGTAAGTGTGCTA	left HA side genotyping primer for KMT5B SET domain
gt4	GAAGAGCGTGTATTCCACTCATAA	left HA side genotyping primer for SETD5 cod. exon 4
gt5	AAATGACAAATTGGTGGATGCC	left HA side genotyping primer for KDM6B cod. exon 3
gt6	GGCTCTGTAGCTGGAATAATACAT	left HA side genotyping primer for ASH1L cod. exon 3
gt7	AGCAATAGCATCACAAATTTCACAA	general genotyping primer for right HA side; inside cassette
gt8	GAAACTATTATACTGCACTGGCTTG	right HA side genotyping primer for SETD5 SET domain
gt9	GAAACCAAAATTCCAGAAGACACA	right HA side genotyping primer for KMT5B SET domain
gt10	ACAGTTAAGGTGATGCTTGTAGG	right HA side genotyping primer for SETD5 cod. exon 4
gt11	TCTCTAAGGTCACTTCCGGG	right HA side genotyping primer for KDM6B cod. exon 3
gt12	TCCTGAAGTTGATCCTCGCT	right HA side genotyping primer for ASH1L cod. exon 3
gt13	TTTCCCAAGTTTAGGCGTTGC	Sanger sequencing primer for WT allele - SETD5 SET domain
gt14	AGATAAAGGTTTGACTGAGGTTGAC	Sanger sequencing primer for WT allele - KMT5B SET domain
gt15	GGCGCCCACTCCAAGA	Sanger sequencing primer for WT allele - SETD5 cod. exon 4
gt16	CAGTGCAGTAGGAGAAAGCAATC	Sanger sequencing primer for WT allele - KDM6B cod. exon 3
gt17	GGCCAAGAGGGGTTCTTAAAATG	Sanger sequencing primer for WT allele - ASH1L cod. exon 3
gt18	CCCCTTACCTCTCTAGTCTGTGC	left HA genotyping primer for WT GFP
gt19	CTCAGGTTCTGGGAGAGGGTAG	right HA genotyping primer for WT GFP
ot1	GAGATCGTAGAGGGTGTCCCA	SETD5 SET - off target 1 - FWD primer
ot2	GCAGATTGACTCGCCAGGAC	SETD5 SET - off target 1 - REV primer
ot3	GCCCTCTATTCTGCCTTCCC	SETD5 SET - off target 2 - FWD primer
ot4	CCACCCACTGCTTGTTCGTA	SETD5 SET - off target 2 - REV primer
ot5	CCTCTCCCTTGCTCTTAATCATCT	SETD5 SET - off target 3 - FWD primer
ot6	CGATCATCACGAGCTTCAGC	SETD5 SET - off target 3 - REV primer
ot7	TGGGGCTAGGGCCTAATCTT	SETD5 SET - off target 4 - FWD primer
ot8	GTGTCCAGGGAGTACAAGGC	SETD5 SET - off target 4 - REV primer
ot9	CTTTTGCCGACCTCAAAGCC	SETD5 SET - off target 5 - FWD primer
ot10	TAAACCTGGCCAGCAGTCTG	SETD5 SET - off target 5 - REV primer
ot11	ACTTCACAATCACTGCAGCTTTC	SETD5 SET - off target 1 - Sanger sequencing primer
ot12	CCTGCTTCCTTCCTCACTGG	KMT5B SET - off target 1 - FWD primer
ot13	AGACAGCAACATGGAGGCAT	KMT5B SET - off target 1 - REV primer
ot14	TCCTTCCTTCAACACCTGCC	KMT5B SET - off target 2 - FWD primer
ot15	CCGTTCTTCCCTCAAACCCA	KMT5B SET - off target 2 - REV primer
ot16	GAGGACGAGGAGCCTGAGT	KMT5B SET - off target 3 - FWD primer
ot17	CTTCGTCCTCCCATGGAACC	KMT5B SET - off target 3 - REV primer
ot18	CATGCAGCAGGGTTAAGGATTG	SETD5 coding exon 4 - off target 1 - FWD primer
ot19	GTCCTTCCAAGGCCCACG	SETD5 coding exon 4 - off target 1 - REV primer
ot20	TGAAGGTACTCCCCAGTCTCC	KDM6B coding exon 3 - off target 1 - FWD primer
ot21	GAAGATCAGCCCCGCACG	KDM6B coding exon 3 - off target 1 - REV primer
ot22	GGCTGGGACTCCTTCCTCTA	KDM6B coding exon 3 - off target 2 - FWD primer
ot23	GTTGGGGTTGGACTGAGGAC	KDM6B coding exon 3 - off target 2 - REV primer
ot24	AGCTAGAAATGTCCTGGCACT	ASH1L coding exon 2 - off target 1 - FWD primer
ot25	TAGTTTCAAGCACACCGGCC	ASH1L coding exon 2 - off target 1 - REV primer
ot26	TGACTAATCACACGTAAGACACCT	ASH1L coding exon 2 - off target 2 - FWD primer
ot27	TCGTAGGGTTCACAGGGGAT	ASH1L coding exon 2 - off target 2 - REV primer
sb1	TTCGAGGTGGTGTACATGAACGACGGCGTGC	FWD Primer for Southern Blot Probe synthesis
sb2	TTACACCCACTCGTGCAGGC	REV Primer for Southern Blot Probe synthesis

Table 5.1: **Oligonucleotide sequences**

All lines were sent for molecular karyotyping (Illumina GSA v3.0+MD) by Life & Brain GmbH and routinely tested for mycoplasma contamination using the MycoAlert™ PLUS Mycoplasma Detection Kit (Lonza LT07-705). Biorender.com was used for creating the schematic depiction in Fig. 5.1.

Southern blot

Done either by Christoph Dotter or Aysan Çerağ Yahya

A Digoxigenin (DIG) labeled probe against the GFP sequence was synthesized using the Roche DIG DNA Labeling mix (Roche, #11277065910) and the primers listed in table 5.1 (sb1-2). For each Southern blot sample 8 µg of genomic DNA isolated from hESCs per sample were digested overnight at 37°C with 30 units of one of the following restriction enzymes: NsI HF (NEB, R3127) for SETD5-SET and KMT5B-SET, KpnI HF (NEB, R3142) for SETD5-ex4, EcoRV HF (NEB, R3195) for ASH1L-ex2, and EcoRI HF (NEB, R3101) for KDM6B-ex3 lines. The digests were resolved on a 0.8% agarose gel, depurinated with 0.25 M HCl for 15 min and denatured with 0.4 M NaOH for two times 15 min. The DNA was then transferred overnight to a positively charged nylon membrane (Roche, #11209272001). The next day the DNA was UV cross-linked (UVP UV Crosslinker, 120 mJ/cm^2) and the DIG-labeled probe was hybridized to the membrane overnight after a 4 hour prehybridization step without probe. On the fourth day the membrane was washed and detected after treatment with Anti-DIG-AP, Fab fragments (Roche, #11093274910) and CDP Star (Roche, #11685627001) in a Molecular Imaging GE Healthcare Amersham 600 (20-50 min exposure dependent on whether one or two membranes were imaged at once).

Cortical spheroid culture

Cortical spheroids were derived from feeder-free hESCs via a protocol based on the method from Yoon et al. [144] with some modifications. In preparation for seeding cells were cultured in Essential 8 complete medium. hESCs at a passage < 50 were dissociated into single cells by treatment with Accutase™ as described above followed by trituration with a 1000 µL pipette. Dissociated cells were spun down at 300 g for 5 min at room temperature, resuspended in Essential 8 complete medium supplemented with 10 µM Y-27632 Rock inhibitor (Stemcell Technologies #72304), and counted twice using an automated cell counter (Biorad Tc-20). Cell suspensions were diluted with more supplemented Essential 8 medium and seeded into ultra-low attachment 96-well plates (either Corning #7007 or faCellitate #F202003) at 10,000 cells in 140 µL per well. Plates were spun down at 200 g (300 g for faCellitate plates) for 2 min to collect cells at the bottom of the well and then cultured at 37°C and 5% CO_2.

On the day after seeding (= day 0) cells had aggregated into embryoid bodies (EBs) and were switched to Neural Induction medium consisting of Essential 6 medium (Gibco, A1516401) supplemented with dual SMAD inhibitors 10 µM SB-431542 (Tocris, #1614) and 2.5 µM dorsomorphin (Sigma, P5499). Medium was not changed on day 1 and daily afterwards on day 2-5. For media changes in 96-well plates old medium was removed down to 10-20 µL to avoid aspiration of spheroids. 140 µL fresh medium were added per well. On days where the medium was switched to a new one 200 µL were used to further dilute out the old medium. On day 6, medium was switched to Neural medium (Neurobasal

A (Gibco, #10888-022), 2% B27 supplement without vitamin A (Gibco, #12587-010), 2 µM (1X) GlutaMAX (Gibco, #35050-061), and 100 U/ml Penicillin and 100 µg /ml Streptomycin) supplemented with 20 ng/ml EGF (Peprotech, AF-100-15) and 20 ng/ml FGF2 (Peprotech, #100-18C). Medium was changed daily until day 14 after which it was changed every other day. On day 13 spheroids were transferred from 96-well plates to 10 cm dishes (either ultra-low attachment or non-treated; Corning, #3262 or #430591) containing 12 ml Neural medium with growth factors with a maximum of 40 per dish. 10 cm dishes were incubated on an orbital shaker at 42 rpm at 37°C and 5% CO_2. On day 25 the medium was switched to Neural medium supplemented with 20 ng/ml BDNF (Peprotech #450-02) and 20 ng/ml NT-3 (Peprotech #450-03) which was changed every 2-3 days. On day 43 the medium was switched to Neural medium without supplements which was changed every 3-4 days.

Cortical spheroid size measurement

For quantifying spheroid growth bright-field images were taken using an EVOS FL imaging system (Thermo Fisher) with brightness set so that the whole spheroid is visible but that background is as bright as possible. Images were then processed using a Fiji/ImageJ script for automatic quantification by converting to grayscale, thresholding by brightness and then measuring using the "Analyze Particles" function. Afterwards data was processed and using R and visualized using the ggplot2 package [197].

RNA isolation

For RNA isolation, spheroids were snap-frozen using liquid nitrogen in pools as follows; day 6: 8 per replicate, day 25/day 60: 2 per replicate. RNA was isolated from snap-frozen samples using Trizol/Chloroform extraction. In short, cell pellets or frozen spheroids were suspended in 500-700 µL (dependent on size of pellet / age of spheroids) TRIzol™ reagent (Thermo Fisher #15596026). Spheroids were crudely dissociated by pipetting before passing three times through a 30G needle. For cell pellets the pipetting step was skipped. Next, 100-140 µL chloroform (0.2x TriZOL volume) were added before samples were mixed by vortexing for 15 s, incubated at room temperature for 2-3 min, and centrifuged for 15 min at 4 °C at 12,000 g to separate the two phases. The aqueous upper phase was transferred to a new tube and 1.5x its volume was added in 100% ethanol (EtOH). The mixture was loaded onto a Zymo-Spin™ IC column (Zymo Research C1004) and passed through by centrifugation for 30 s at 10,000 g and room temperature. After a subsequent wash with 400 µL 70% EtOH the sample was treated in column with 5 µL RQ1 DNAseI (Promega M6101) mixed with 5 µL 10X DNAse reaction buffer and 40 µL 70% EtOH for 15 min. Finally, the column was washed two times with 700 µL 70% EtOH and the sample eluted in 25-50 µL DEPC-treated H_2O. Concentration was measured using a NanoDrop spectrophotometer and RNA integrity was checked with the RNA 6000 Nano kit (Agilent #5067-1511) on a Bioanalyzer 2100 instrument (Agilent).

Quantitative PCR

After RNA isolation from hESCs as described above complementary DNA (cDNA) was synthesized using either the SuperScript™ III kit (Thermo Fisher #18080044, for SETD5 SET and KMT5B SET lines) or the RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher K1622, for all other lines) using oligo(dT) primers and 0.5 - 0.75 µg total RNA

as input. Quantitative PCR was done using the the LightCycler ® 480 SYBR Green I Mastermix (Roche #04707516001) on a LightCycler® 480 (Roche) according to the manual with primers for human SETD5, KMT5B, ASH1L, and KDM6B. HPRT was used for normalization (see Table 5.1 for sequences). Results were visualized using R.

RNA-sequencing - library generation

Libraries for RNA-sequencing were prepared as biological triplicates per line. 400 ng (day 6) or 1 µg (others) of total RNA were used as input for cDNA library preparation using the NEBNext® Ultra™ II Directional RNA Library Prep Kit (NEB E7765) with unique dual indexing primers (NEBNext Multiplex Oligos for Illumina, NEB E6440S) and poly(A) selection with the NEBNext® Poly(A) mRNA Magnetic Isolation Module (NEB E7490) according to the manual. For libraries made from 1 µg RNA the adaptor was diluted 1:5 and 9 PCR cycles were used. For libraries from 400 ng a 1:8 or 1:9 dilution was used and the number of cycles was increased to 10-11. Size distribution of finished libraries was assessed using the High Sensitivity DNA kit (Agilent #5067-4626) on a Bioanalyzer 2100. Concentration was measured with the Qubit® dsDNA HS Assay Kit (Thermo Fisher Q32851) using a Qubit® 2.0 fluorometer and libraries were pooled in equimolar amounts to generate multiplexes with molarities of 15-20 nM. Paired end sequencing of 50 bp reads on a NovaSeq 6000 and demultiplexing was performed by the NGS facility of the Vienna Biocenter Core Facilities. Average sequencing depth per sample was 38 million (day 25), and 26 million (day 60) raw reads.

Immunostaining and imaging

Done by Farnaz Freeman, Nathalie Gruber, Jessica Kirchner, Jessica Leff, and Julia Senkiv.

Cortical spheroids were collected in a 12-well plate, washed with 1X PBS (-/-) and fixed with 4% paraformaldehyde (PFA) in 1X PBS (-/-) first for 15 min at room temperature, then for a minimum of 2 hours (young, small spheroids), 4 hours (bigger spheroids), or up to overnight (older spheroids). After washing with 1X PBS (-/-) spheroids were cryoprotected in 30% sucrose at 4°C. After washing again spheroids were embedded in Optimal Cooling Temperature compound (O.C.T.), frozen on dry ice and stored at -20°C. 20 µm sections were prepared on a cryotome and stored at -20°C until staining.

For immunostaining slices mounted on glass slides were washed with 1X PBS three times for 10 min. In case of the Ki67 staining antigen retrieval was performed by incubating for 10 min in 10X Dako Antigen Retrieval Solution (1:10 in milliQ water) in a water bath at 95°C. After cooling for 30 min following antigen retrieval or directly after washing (if no antigen retrieval was used) slices were block with 2% normal goat serum and 0.2% Triton X-100 in 1X PBS for one hour. Primary antibody incubation was done in 0.2$ Triton X-100 and 2% normal goat serum in PBS overnight at 4°C. After washing three times with PBS slices were incubated for 2 hours at room temperature with secondary antibody. After washing again, slices were incubated with Hoechst solution (1:5000 in 1X PBS) for 10 minutes. Finally, slides were washed with PBS and coverslipped with Dako fluorescence mounting medium. Images were acquired using an LSM700 or LSM800 confocal system (Zeiss). Primary antibodies used: KI67 (1:300, Abcam, ab156956); PAX6 (Thermo Fisher, MA1-109); TBR2 (Sigma, AB2283); CTIP2 (Abcam, ab18465). Secondary antibodies

used (1:500): Alexa 647 goat anti-rat IgG, Life Tech, A-21247; Alexa 488 goat anti-Mouse IgG, Life Tech, A-32723; Alexa 594 goat anti-rabbit IgG, Life Tech, A-11037.

RNA sequencing - mouse embryo

Done by Bernadette Basilico.

Mouse tissue was rapidly dissected on ice, snap-frozen in liquid nitrogen and stored at -80°C until further processing. RNA isolation and RNA-seq library preparation was done as described above and sequenced on an Illumina HiSeq V4 by the NGS facility of the Vienna Biocenter Core Facilities. Raw reads were processed and aligned as described above and differential expression was assessed via Galaxy [211] using DESeq2 [209]. For comparison with spheroid data the same cutoffs (FDR ≤ 0.05, |log2 fold change| ≥ 0.5) were used.

RNA-seq - data analysis

Preprocessing and quality control

Demultiplexed sequencing reads were processed using cutadapt (version 3.4, [354]) to remove adapter sequences and low-quality bases (–nextseq-trim=20) towards the ends of the reads. Trimmed reads were aligned to the human genome (genome: hg38/GRCh38, gene annotation: Gencode release 24)) using STAR (version 2.5.1b, [191]). Reads aligning more than once were removed (–outFilterMultimapNmax 1) and counts of uniquely aligned reads per gene were collated by STAR (–quantMode GeneCounts). All subsequent analysis was done in R 4.1.2. Data for each time point was processed separately. Coherence between replicate samples was assessed using sample distance clustering and principal component analysis (PCA, pcatools 2.6.0, [355]) after variance stabilizing transformation (DESeq2 1.34.0, [209]). Coherence between replicates and batches of each line was high but some differences between replicate lines, most notably WT and WT-GFP, were found.

In order to adjust for these differences for subsequent analysis this unwanted variability was corrected using the RUVSeq package (version 1.28.0, [194]) following the manual. In short, genes with very low read counts were filtered out by considering only genes with at least 10 raw reads in more than 6 samples (corresponding to the number of replicates per line). Subsequently, the data was normalized via upper quartile normalization (EDASeq, 2.28.0, [193]) before estimating sources of variation using the RUVs function. Given the stark differences between some mutant lines and wildtype lines an empirical set of negative control genes was used since the assumption that almost all genes are not differentially expressed did not hold. This set was determined by performing differential gene expression analysis using edgeR (see below), selecting the 6,000-7,000 non-significant genes with highest P values from each mutant/wildtype comparison, and compiling a list of genes present in all four of these sets as the least changed across all conditions. Samples were placed into groups based on genotype (i.e. wildtype, *SETD5*$^{+/-}$, *KMT5B*$^{+/-}$, *ASH1L*$^{+/-}$, *KDM6B*$^{+/-}$). The number of factors, k, was selected by evaluating a range of k values through visualizing remaining variation as relative log expression (RLE) plots and sample clustering through PCA. The select k value (k = 2 for day 25, k = 3 for day 60, k = 1 for the combined day 25/60 PCA plot in Fig. 5.4a) was chosen as the lowest value where lines for the same genotype clustered together. PCA plots shown in the manuscript were made from the normalized count output of RUVSeq.

Differential gene expression analysis

After estimation of factors of unwanted variation they were used as input as confounding variables in the model used for differential gene expression analysis besides genotype, the variable of interest. The edgeR package (version 3.36.0, [195]) was used for this analysis, using the likelihood ratio test and an FDR cutoff of 0.05. All samples per time point were processed together and single genotype results were extracted via coefficients. For subsequent analyses only differentially expressed genes with an absolute log2(fold change) ≥ 0.5 were chosen to decrease the amount of spuriously called differentially expressed genes. Heatmap for expression of marker genes was created using the ComplexHeatmap package (version 2.13.1, [356]).

Enrichment analyses

Gene Ontology functional enrichment analysis was done using the GOstats package (version 2.60.0, [199]) with a P value cutoff of 0.001, all expressed genes after filtering (see above) as background list, and conditional testing enabled. Enrichment for transcription factor targets was done using the enrichR package[2] (version 3.0) to interface the Enrichr gene set enrichment analysis web server [204]. The datasets used were the ENCODE_and_ChEA consensus set, the ENCODE_TF_ChIP-seq_2015 set, and the ChEA 2016 set . In the case of multiple significant results from different sets for the same transcription factor the result with the lowest adjusted P value was reported. For the figure we used a cutoff of 0.05 for the adjusted enrichment P value and transciption factors related to signaling or ones that showed differential expression themselves were selected. Visualization was done using the R package ggh4x[3] (version 0.2.1) as an add-on for ggplot2.
Fisher's exact test was used to test the significance of gene set overlaps with P values corrected for multiple testing with the Benjamini/Hochberg method. Expressed genes (post filtering) were used as background sets. All visualization was done in R using the ggplot package [197]. Cell type markers for the marker enrichment analysis presented in Fig. 5.4b were derived from the supplement of a study on the generation of striatal organoids by Miura et al. [139]. The top 100 genes for each cluster were filtered for cluster specific unique markers yielding a set of 88 excitatory neuron markers and 90 interneuron markers. For the comparison with the mouse data the genes from the mouse dataset were converted to their human equivalents using the getLDS function from the biomaRt package (version 2.50.3, [357], host: https://dec2021.archive.ensembl.org/).

Cell type deconvolution

For the deconvolution of our bulk RNA-seq data into cell types the SCDC package (version 0.0.0.9000, [358]) was used. We used the single cell RNA-sequencing dataset from Yoon et al. [144] as reference as it was generated from the same experimental model and should therefore be closest in terms of cell type identity. For this purpose the cells x genes matrix was downloaded from GEO (GSE107771). As no clustering information was provided the data was reprocessed using the Seurat package (version 4.1.0, [359]) with the following parameters: CreateSeuratObject(..., min.cells = 3, min.features = 200); FindVariableFeatures(..., selection.method = "vst", nfeatures = 2000); FindNeighbors(..., dims = 1:20); FindClusters(...,resolution = 0.6); runTSNE(..., dims = 1:20); FindAllMarkers(..., only.pos = TRUE, min.pct = 0.25, logfc.threshold = 0.25). A total of 13

[2]https://github.com/wjawaid/enrichR
[3]https://github.com/teunbrand/ggh4x

clusters were identified and annotated based on comparing marker genes with the list of markers provided in Birey et al. ([138]). For use in the deconvolution analysis the four excitatory neuron and the two interneuron clusters were collapsed into one EN and one IN cluster, respectively. After benchmarking against pseudobulk datasets, we decided to only include the EN, IN, radial glia (RG), choroid plexus, and astrocyte clusters in the deconvolution of our bulk data. For the SCDC analysis the following parameters were used: SCDC_qc(..., qcthreshold = 0.7). Choroid plexus and astrocytes were omitted from the visualization as they made up only a small percentage and the percentages of the remaining three groups were rescaled to add up to 100%.

5.4 Results

5.4.1 Modeling human cortical development using cortical spheroids

In order to study human cortical development *in vitro*, we chose to employ the cortical spheroid method developed by the Paşca group [133, 144, 360] with some adaptations (see methods). This method was demonstrated to recapture the early stages of corticogenesis at both the transcriptional and chromatin level [145] which made it an ideal candidate for this study. Starting from human embryonic stem cells (hESCs) this guided differentiation protocol is designed to enforce a dorsal forebrain fate. Cells progress through an early neuroepithelial stage towards the formation of ventricular zone like neural progenitor zones termed rosettes. Eventually, these progenitors generate excitatory cortical neurons (Fig. 5.1a). Using immunostainings we confirmed the formation of rosette structures and the high degree of proliferative progenitor cells positive for the active proliferation marker KI67 during early stages (Fig. S5.3a). At later stages we observed a layer of TBR2 positive intermediate progenitor cells around the PAX6 positive centers of rosette structures, reminiscent of the sub-ventricular zone. Overall our results showed that the majority of cells differentiated into CTIP2 positive deep layer neurons at all later stages assayed S5.3b-c) .

5.4.2 Generation of heterozygous mutant hESC for four ASD risk genes

For the scope of this study we selected four genes based on their strong association with autism spectrum disorder (ASD) (Simon's Foundation Autism Research Initiative (SFARI) score 1) and their involvement in chromatin regulation as we deemed these most likely to exert their function at the gene regulatory level. This led to the selection of the genes *ASH1L* (*KMT2H*), *KDM6B* (*JMJD3*), *SETD5*, and *KMT5B* (*SUV420H1*) (Fig. S5.1). In order to investigate convergence in a human model system we first established hESC carrying heterozygous disruptions of these genes of interest. For this we chose the H9 hESC line as an isogenic starting point and used the Clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 system to insert a Green Fluorescent Protein (GFP) cassette into one allele of our target genes. For this we targeted either a known/predicted functional domain, as for *KMT5B* and *SETD5* line 1, or a region in the beginning of the gene, as for *ASH1L*, *KDM6B*, and *SETD5* line 2 (Fig. 5.1b). The additional targeting for *SETD5* was in part motivated by the findings of our previous study that SETD5 does not exhibit methyltransferase activity. Correct insertion and absence of off-target events was confirmed by Sanger sequencing and clones were checked for the absence of random integration events using Southern blot (Fig. S5.2a-c) This GFP knock-in strategy for the creation of heterozygous knockouts enabled us to use fluorescence activated cell sorting (FACS) based selection (Fig. 5.1c, details see methods). We checked reduction of target gene expression at the mRNA level and observed a consistent 50% reduction for *SETD5*, *KMT5B*, and *ASH1L* while the results for *KDM6B* were inconclusive due to high variability in both mutant and wildtype lines (Fig. S5.2d). In summary, we used a total of eight lines in this study: two wildtype control lines, WT and WT-GFP (a WT line with the GFP cassette inserted into the AAVS1 locus); two $SETD5^{+/-}$ lines, SETD5 1 and 2, targeting the SET domain and coding exon 4, respectively; two $KMT5B^{+/-}$ lines; one $ASH1L^{+/-}$ line; one $KDM6B^{+/-}$ line (Table 5.2).

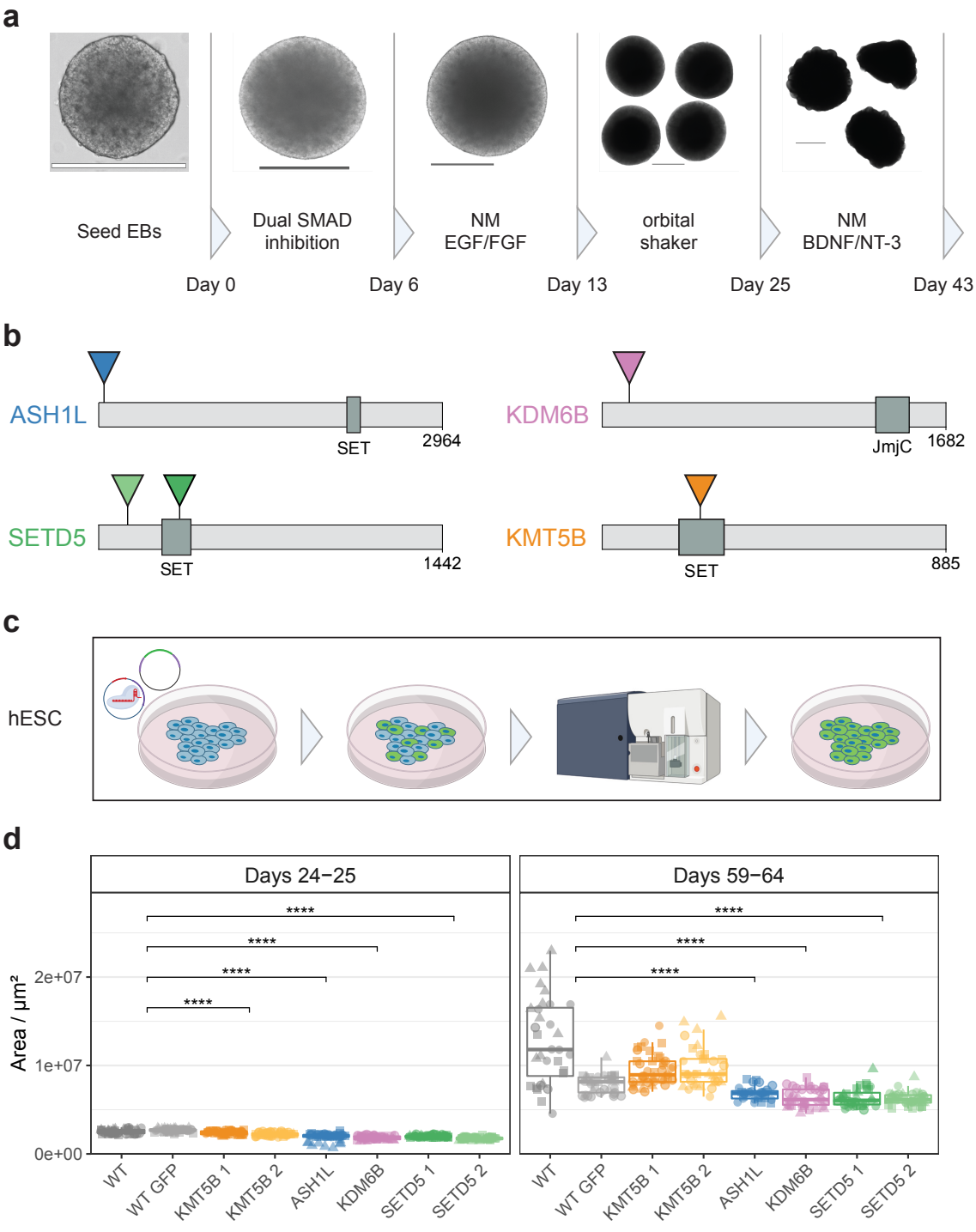
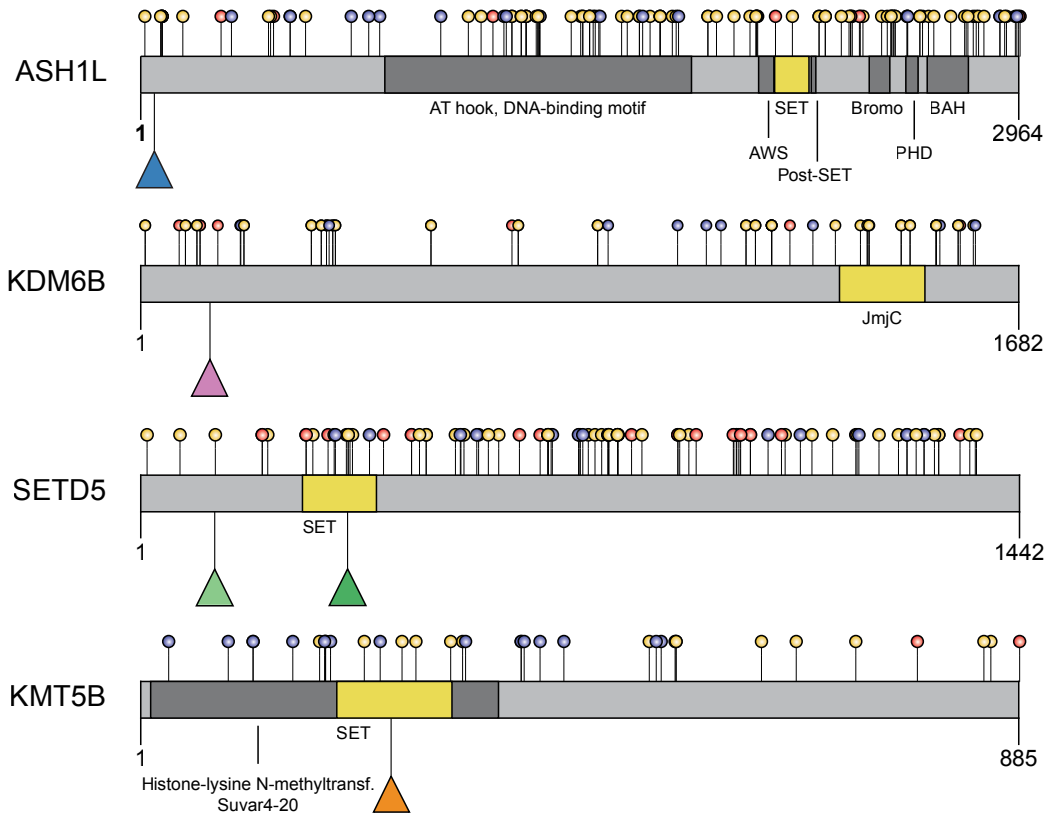


Figure 5.1: **Modeling heterozygous loss of ASD genes in cortical spheroids. a)** Overview of the cortical spheroid differentiation protocol used in this study. Images show representative bright-field images of wildtype spheroids in the indicated age periods. Timing of switches between media as well as start of orbital shaker use are indicated at the bottom. NM, neural medium. **b)** Schematic depiction of the four selected ASD genes. Triangles indicate sites targeted by CRISPR/Cas9. Catalytic domains for the respective enzyme classes are highlighted in dark gray. Protein sizes in amino acids are shown on the right. **c)** Schematic overview of the generation of mutant hESC lines. Starting with transfection of plasmids containing gRNA and Cas9 as well as the HDR template for insertion of the GFP cassette, followed by FACS-enabled selection of transfected cells. FACS image from BioRender.com. **d)** Spheroid size measured at two different time points using area in bright-field images as a proxy. n(Days 24-25): WT = 48, WT-GFP = 46, KMT5B 1 = 45, KMT5B 2 = 46, ASH1L = 44, KDM6B = 45, SETD5 1 =

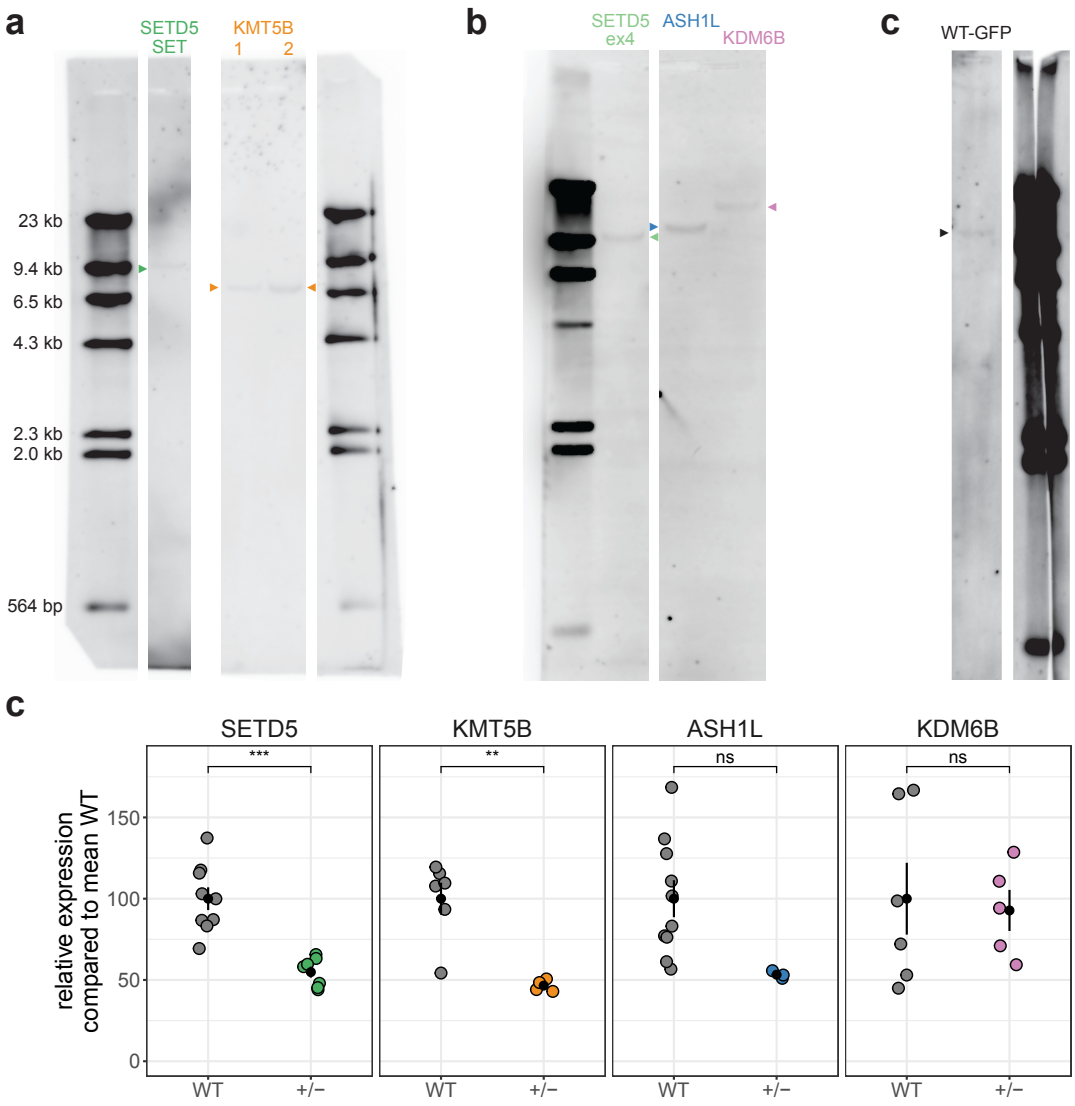
46, SETD5 2 = 46; n(Days 59-64): WT = 32, WT-GFP = 36, KMT5B 1 = 41, KMT5B 2 = 34, ASH1L = 27, KDM6B = 37, SETD5 1 = 34, SETD5 2 = 37 spheroids. 4 differentiation batches for all except 3 batches for ASH1L, Days 59-64. Welch two sample t-test between pooled lines per genotype; ****: Benjamini Hochberg adjusted P value ≤ 0.0001.



Supplementary Figure S5.1: **Overview of targeted genes.** Schematic depiction of the proteins encoded for by the four target genes. On top, lollipops indicate reported mutations in ASD patients; red: premature stop codon, yellow: missense, blue: frame-shift. Annotated proteins are visualized as boxes and important catalytic domains for histone methyltransferases (SET domain) and demethylases (JmJC) are highlighted in yellow. Targeting sites are shown as triangles at the bottom. Numbers of the right note protein length in amino acids. For further information on cell lines see Table 5.2 below.

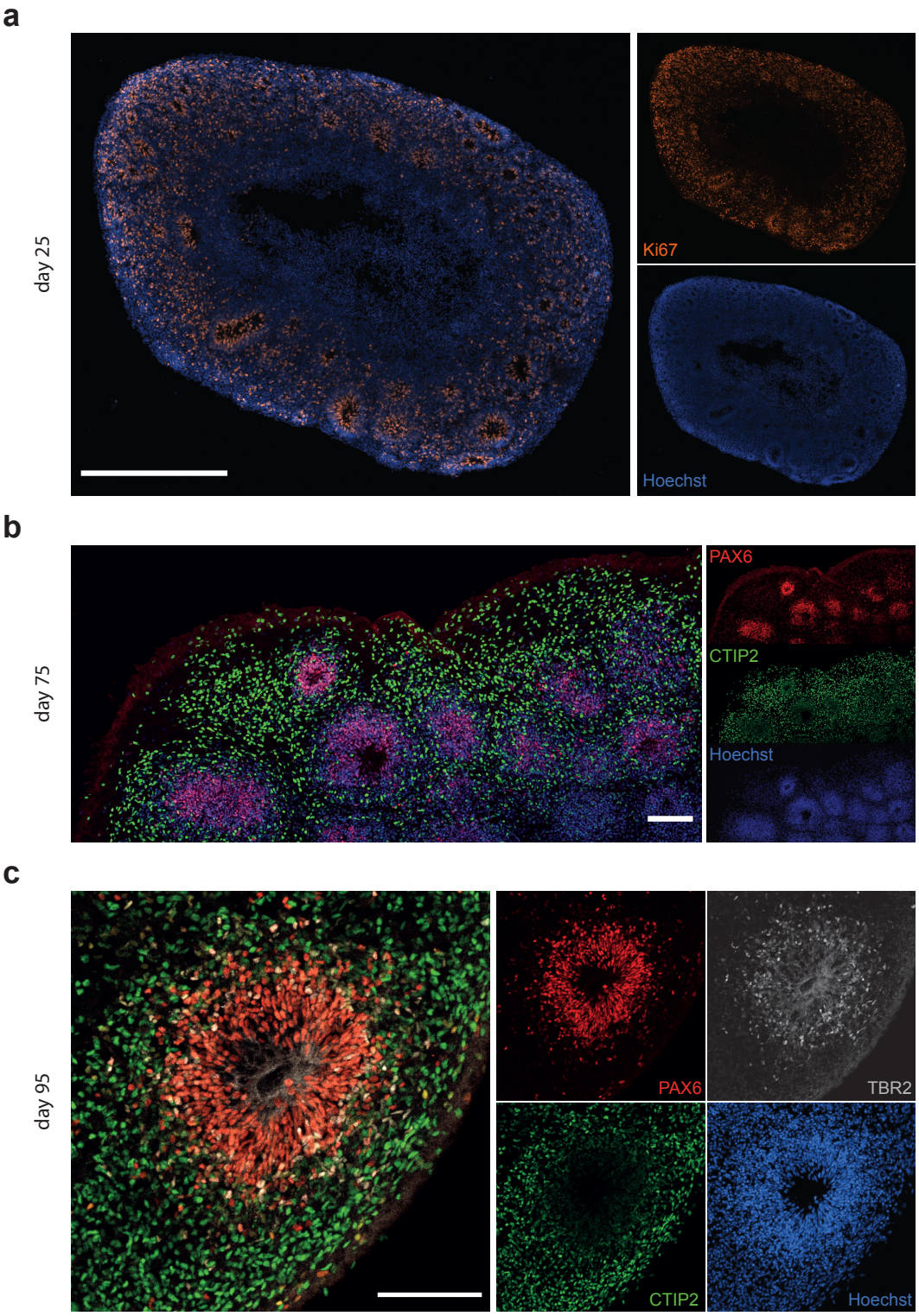
Cell line	Genotype	Target gene	RefSeq Transcript ID	Targeted coding exon	First changed aa	Targeted domain
ASH1L	ASH1L +/-	ASH1L	NM_018489	2	Asp48	-
KDM6B	KDM6B +/-	KDM6B	NM_001080424	3	Leu134	-
KMT5B 1	KMT5B +/-	KMT5B	NM_017635	6	Met253	SET
KMT5B 2	KMT5B +/-	KMT5B	NM_017635	6	Met253	SET
SETD5 1	SETD5 +/-	SETD5	NM_001080517	8	Ala340	SET
SETD5 2	SETD5 +/-	SETD5	NM_001080517	4	Arg122	-
WT-GFP	+/+	AAVS1 locus	-	-	-	-
WT	+/+	-	-	-	-	-

Table 5.2: **Cell lines in this study**



Supplementary Figure S5.2: **Generation of heterozygous mutant hESC lines a+c)** Southern blot results for a probe against GFP confirmed absence of random integration and correct insertion at the target site. Results in a and b are each from the same gels but non-consecutive lanes. Membrane in blot a was cut in the middle (as indicated by the wider gap) for hybridization and imaging. **d)** qPCR results for the four target genes in the respective mutant lines (+/-) and wild type controls (WT). Data is shown first normalized to a housekeeping gene (*TBP* for *KDM6B*, *HPRT* for the others) and then normalized to mean WT expression. Black dots show mean values, lines indicate standard error.

When generating cortical spheroids from these lines we observed divergent development at the macroscopic level. This was especially apparent after terminal neuronal differentiation was started at day 25. While spheroids from wildtype or *KMT5B*$^{+/-}$ lines expanded in size and became more bulbous in appearance, spheroids from *ASH1L*$^{+/-}$, *KDM6B*$^{+/-}$, and *SETD5*$^{+/-}$ lines exhibited stunted growth throughout (Fig. 5.1d). We observed size reductions of 24 - 33% at days 24-25 and 33% - 39% at days 59-64 in *ASH1L*$^{+/-}$, *KDM6B*$^{+/-}$, and *SETD5*$^{+/-}$ compared to pooled WT and WT-GFP data. In comparison, *KMT5B*$^{+/-}$ lines differed by 8 - 15% at the earlier and 4 - 7% at the later time point.



Supplementary Figure S5.3: **Generation of cortical spheroids. a)** Representative image of a whole section of a wild type spheroid at day 25. Ki67 staining shows widespread proliferation and many small rosette structures are visible. Scale bar: 500 µm . **b)** Representative overview of a region in a day 75 organoid shows multiple rosette structures positive for PAX6 surrounded by an abundance of CTIP2 positive neurons. Scale bar: 100 µm . **c)** Representative close-up of a rosette structure at day 95 highlights the layering of PAX6 positive radial glia on the inside in the ventricular zone like region, surrounded by a layer of TBR2 positive intermediate progenitor cells reminiscent of a subventricular zone. Outside these structures CTIP2 positive neurons form the majority of cells. Scale bar: 100 µm .

5.4.3 Divergent development of cortical spheroids as a shared phenotype

Given our observation of differing growth trajectories starting from day 25 we decided to investigate spheroid development at the transcriptional level. For this reason we performed bulk RNA-sequencing of whole spheroids at two different time points, day 25 and day 60. While we initially observed high concordance of replicates of the same cell line, even across batches (min. $R^2 = 0.97$, Spearman correlation), we noticed increased variability across different cell lines within genotype groups, especially the WT and WT-GFP lines (min. $R^2 = 0.92$). Therefore we chose to remove this variation computationally (see methods).

We first investigated day 25 as the time point prior to the exposure to neurotrophic factors where spheroids would largely consist of progenitor cells. We observed a clear separation of samples by genotype (Fig. 5.2a). Remarkably, we found that samples from the lines of three genes, *ASH1L*$^{+/-}$, *KDM6B*$^{+/-}$, and *SETD5*$^{+/-}$, continued to cluster together even after correction for unwanted variation, indicating a high degree of similarity. This was also apparent at the level of differential gene expression where we found a large proportion of differentially expressed genes (DEGs) shared (61 - 75% of total DEGs per set) between those three mutants (Fig. 5.2b). Furthermore these genes showed highly concordant directionality of regulation with over 99% (1994) of these genes dysregulated in the same way in all three sets. Overall we found a high number of differentially expressed genes (FDR ≤ 0.05 and $|\log2(\text{fold change})| \geq 0.5$), especially in *ASH1L*$^{+/-}$, *KDM6B*$^{+/-}$, and *SETD5*$^{+/-}$ samples (2050 - 3274 DEGs).

Next, we aimed to understand these changes at the functional level. We reasoned that the genes shared between *ASH1L*$^{+/-}$, *KDM6B*$^{+/-}$, and *SETD5*$^{+/-}$ were likely the most relevant within these sets and therefore focused on this overlap instead of individual results. Through functional enrichment analysis using the Gene Ontology database we discovered very strong enrichment for genes related to synaptic signaling among the up-regulated genes of the three-way overlap. We also found that not only genes encoding proteins of glutamatergic but also GABAergic synapses were enriched at the cellular component level (Fig. 5.2c).

This was of particular interest given that the cortical spheroid system is designed to enforce a dorsal, rather than ventral, cortical fate and to produce predominantly excitatory neurons. For down-regulated genes we instead observed enrichment for genes related to developmental signaling pathways such as Wnt and BMP signaling.

Furthermore, we found enrichment for terms like neurogenesis related to earlier neurodevelopment which, together with the aforementioned upregulation of synaptic genes, painted a picture of premature neuronal differentiation in these three mutants. The results for *KMT5B*$^{+/-}$ showed a similar albeit weaker enrichment for synaptic genes among up-regulated genes. However, at the cellular component level we only found enrichment for excitatory but not inhibitory synapse components. Down-regulated genes also showed similarities to the results of the other three lines with the exception of no enrichment of terms related to Wnt signaling (Fig. 5.2d).

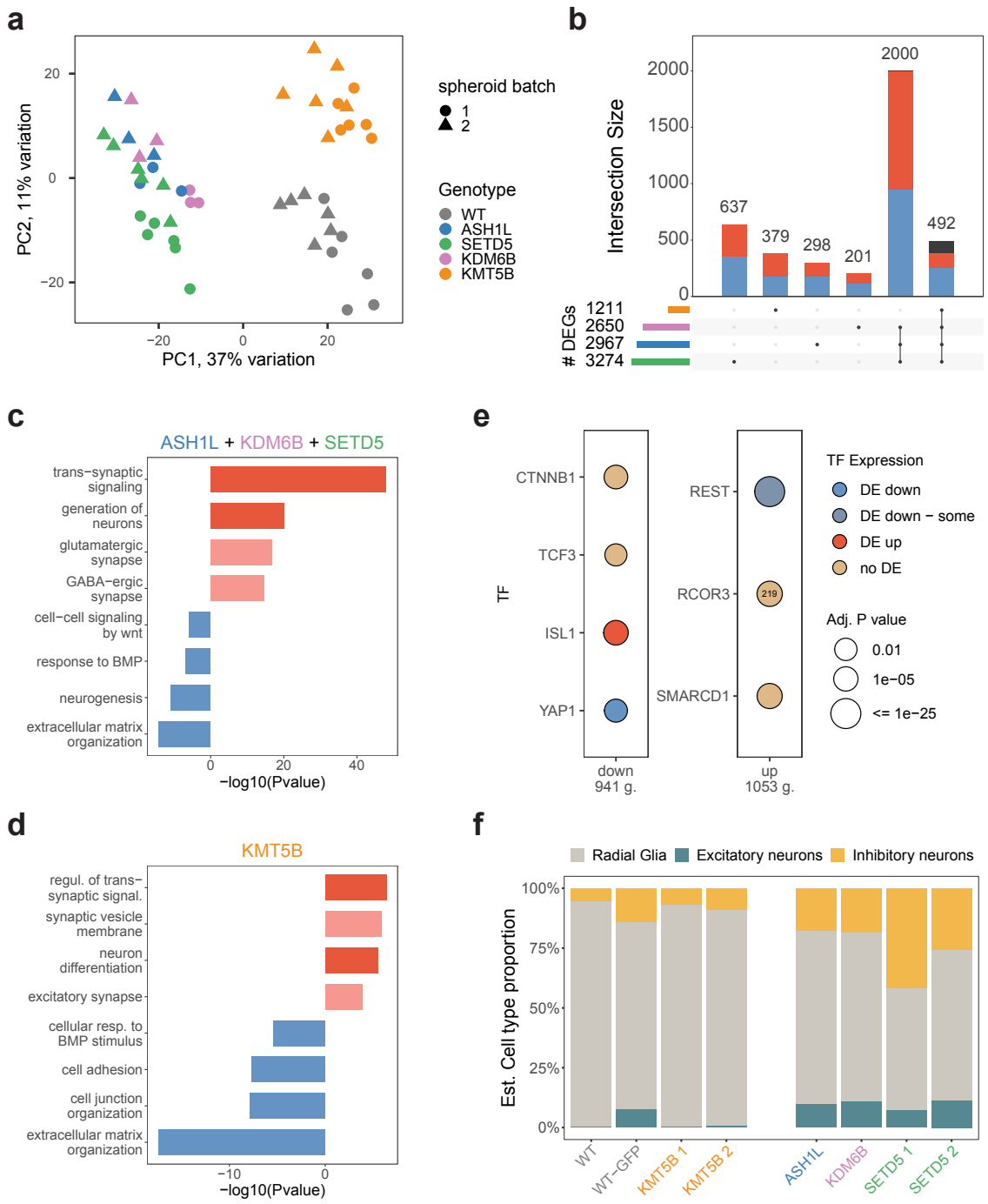


Figure 5.2: **Shared transcriptional changes across *ASH1L*$^{+/-}$, *KDM6B*$^{+/-}$, and *SETD5*$^{+/-}$ cortical spheroids at day 25. a)** Principal component analysis of RNA-sequencing results from cortical spheroids collected at day 25. Data from two independent differentiation batches with three replicates per line per batch, corrected for unwanted variation (see methods). **b)** Upset plot showing the overlap between sets of DEGs (FDR ≤ 0.05, $|\log2(\text{fold change})| \geq 0.5$). Colored bars at the bottom left show the total amount of DEGs for each line and dots to the right of it indicate the respective intersections with single dots indicating DEGs unique to each set. Bars on the top right show the number of DEGs in each intersection colored by direction of change and consistency: red/blue indicate consistent up-/down-regulation across all sets in the intersection while black shows shared DEGs with divergent directions of change. **c+d)** Selected Gene Ontology enrichment results for the three-way DEG intersection between *ASH1L*$^{+/-}$, *KDM6B*$^{+/-}$, and *SETD5*$^{+/-}$ (c) and the set of DEGs for *KMT5B*$^{+/-}$ (d). Red and light-red bars indicate enrichment in up-regulated genes for biological processes and cellular components, respectively. Blue bars show enrichment for biological processes in down-regulated genes. Of note is the

strong enrichment for synaptic genes and genes related to GABAergic synapses in the three-way overlap. Only genes with shared direction of change were considered. **e)** Selected results of transcription factor (TF) target enrichment in the three-way overlap for down-regulated (left) and up-regulated genes (right). Color indicates (differential) expression (DE) of the respective transcription factors themselves with red and blue representing up- and down-regulation respectively. *REST* was significantly down-regulated in all three comparisons but made the fold change cutoff only in $SETD5^{+/-}$. Circle size scales with negative decimal logarithm of the adjusted enrichment P value and is capped at 10^{-25}. Total sizes of each set are shown beneath the x-axis labels. Only genes with shared direction of change were considered. Number of genes is indicated in cases where more than 20% of the DEGs are transcription factor targets. **f)** Cell type deconvolution results for cortical spheroids at day 25 (see methods for details). Stacked bar plots indicate mean relative proportions of radial glia, excitatory neurons, and inhibitory neurons per cell line (n = 6 samples per line).

Next, we supplemented our functional enrichment analysis by investigating potential regulators directing these changes. For this, we checked enrichment for transcription factor targets among the DEGs shared between $ASH1L^{+/-}$, $KDM6B^{+/-}$, and $SETD5^{+/-}$ using the Enrichr database. Among these results a number of factors caught our eye (5.2e). In line with the down-regulation of Wnt signaling indicated by our GO term enrichment analysis we found down-regulated genes to be enriched for both targets of β-catenin as well as TCF3. Besides that, we found two transcription factors whose targets were enriched in down-regulated genes and which themselves were differentially regulated in these three mutants. ISL1, a marker of the lateral ganglionic eminence [361, 362], was previously shown to play a role in striatal development [363, 364]. YAP1 on the other hand had been linked to proliferation of radial glia cells [365] and its down-regulation might therefore indicate a relative reduction of this cell type.
On the other hand, upregulated genes were strongly enriched for targets of REST, itself downregulated, and RCOR3, one of its corepressors. As REST plays a role in repressing neuronal genes [366] these results also support the hypothesis of premature neuronal differentiation. Lastly, we also found enrichment for targets of SMARCD1 (BAF60A), a member of the SWI/SNF chromatin remodeling complex. This factor has been linked to a neurodevelopmental disorder including intellectual disability [367] and has been suggested to play a role in the exit from pluripotency [368]. Of note, our previous study on the role of *Setd5* in mouse identified Smarcd1 as a physical interactor of Setd5 [167].

In order to further investigate the hypothesis that changes in cell type proportions, i.e. due to premature production of neurons, might be responsible for the vast number of differentially expressed genes we applied a cell type deconvolution algorithm. This approach uses cell-type specific expression profiles prepared from single cell RNA-seq datasets to estimate the relative proportions of cell types that make up a bulk sample. This analysis, using a cortical spheroid dataset [144] as reference, revealed a remarkable increase in the estimated proportion of inhibitory neurons while excitatory neurons were only slightly increased compared to wildtype and $KMT5B^{+/-}$ samples (5.2f). The strongest such effect was seen in one of the $SETD5^{+/-}$ lines with estimates per sample ranging from 39 to 53%.

Taken together these results suggest a shared phenotype among $ASH1L^{+/-}$, $KDM6B^{+/-}$, and $SETD5^{+/-}$ lines in terms of accelerated production of neurons at a time point before treatment with neurotrophic factors. Of special interest is the evidence hinting at an escape from targeted excitatory neuron differentiation and acquirement of an alternative inhibitory neuron identity.

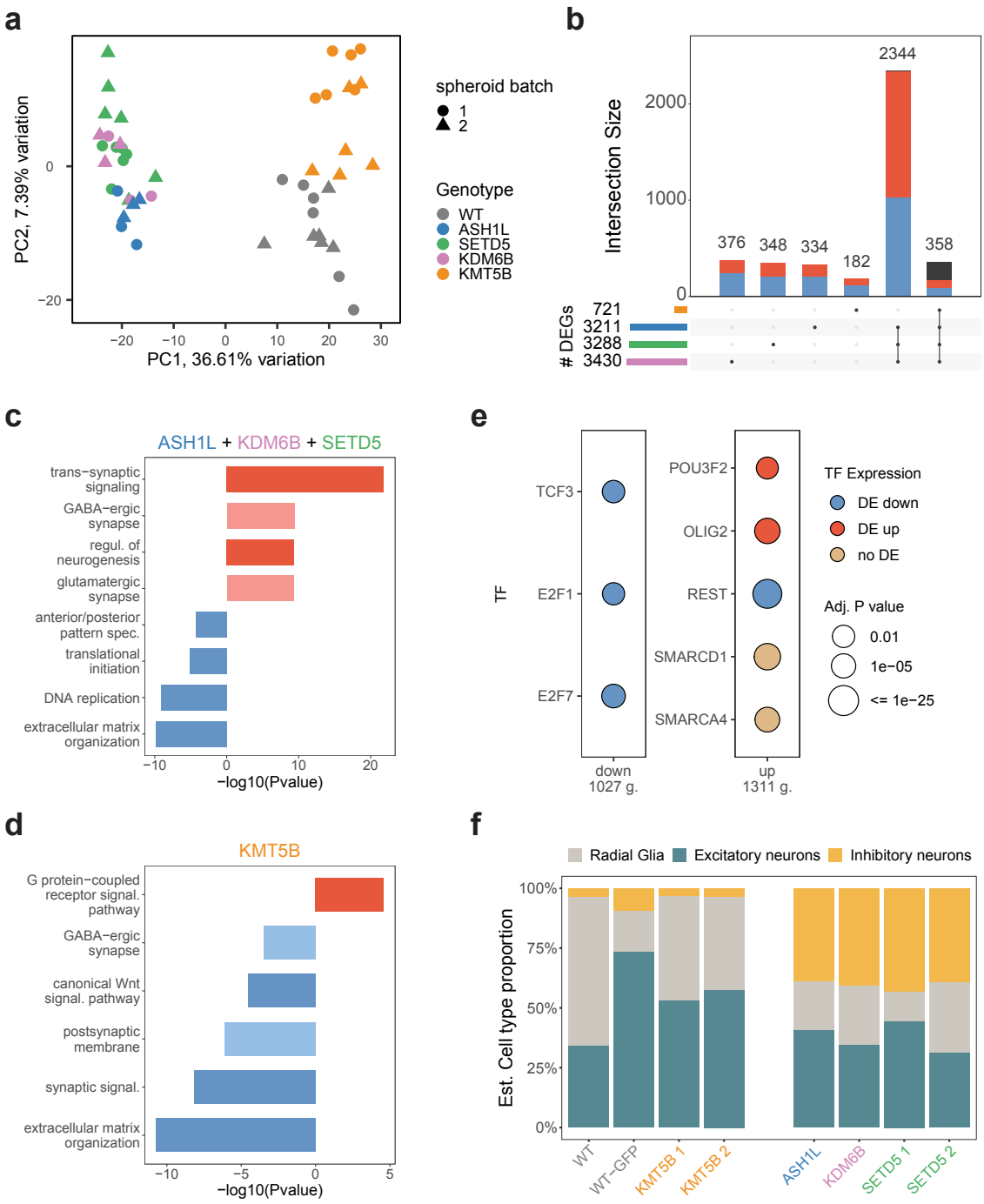


Figure 5.3: **Shared transcriptional changes across *ASH1L*$^{+/-}$, *KDM6B*$^{+/-}$, and *SETD5*$^{+/-}$ cortical spheroids at day 60. a)** Principal component analysis of RNA-sequencing results from cortical spheroids collected at day 60. Data from two independent differentiation batches with three replicates per line per batch, corrected for unwanted variation (see methods). **b)** Upset plot showing the overlap between sets of DEGs (FDR ≤ 0.05, $|\log2(\text{fold change})| \geq 0.5$). Colored bars at the bottom left show the total amount of DEGs for each line and dots to the right of it indicate the respective intersections with single dots indicating DEGs unique to each set. Bars on the top right show the number of DEGs in each intersection colored by direction of change and consistency: red/blue indicate consistent up-/down-regulation across all sets in the intersection while black shows shared DEGs with divergent directions of change. **c+d)** Selected Gene Ontology enrichment results for the three-way DEG intersection between *ASH1L*$^{+/-}$, *KDM6B*$^{+/-}$, and *SETD5*$^{+/-}$ (c) and the set of DEGs for *KMT5B*$^{+/-}$ (d). Red and light-red bars (blue and light-blue bars) indicate enrichment in up-regulated (down-regulated) genes for biological processes and cellular components, respectively. Only genes with shared direction

of change were considered. **e)** Selected results of transcription factor (TF) target enrichment in the three-way overlap for down-regulated (left) and up-regulated genes (right). Color indicates (differential) expression (DE) of the respective transcription factors themselves with red and blue representing up- and down-regulation respectively. Circle size scales with negative decimal logarithm of the adjusted enrichment P value and is capped at 10^{-25}. Total sizes of each set are shown beneath the x-axis labels. Only genes with shared direction of change were considered. **f)** Cell type deconvolution results for cortical spheroids at day 25 (see methods for details). Stacked bar plots indicate mean relative proportions of radial glia, excitatory neurons, and inhibitory neurons per cell line (n = 6 samples per line). Of note, proportion estimates for inhibitory neurons reach almost 50% in *ASH1L*$^{+/-}$, *KDM6B*$^{+/-}$, and *SETD5*$^{+/-}$ lines. Proportions of excitatory neurons increased in all samples compared to day 25.

We then turned to the later time point, day 60, to see whether the observed changes persisted or diminished over time once the scheduled production of excitatory neurons took hold. We found that the results were highly reminiscent of the earlier time point with similar segregation of mutant samples in the principal component analysis (Fig. 5.3a). Also at the level of differential expression the results showed a remarkable resemblance with a strong overlap between DEGs identified in *ASH1L*$^{+/-}$, *KDM6B*$^{+/-}$, and *SETD5*$^{+/-}$ samples. Numbers of DEGs were similar or increased for these three mutants while they decreased by 40% for *KMT5B*$^{+/-}$. The degree of discordance in the four-way overlap between all sets also increased from 22% to 54% indicating further divergence in development between these mutants (Fig. 5.3b).

The results for the three-way overlap at the functional level were consistent with day 25. We found strong enrichment for Gene Ontology terms related to both GABAergic and glutamatergic synapses in up-regulated genes (Fig. 5.3c). The results for down-regulated genes differed and more specifically highlighted a decrease in cell cycle related gene expression indicative of an extended depletion of progenitor cells. The results for *KMT5B*$^{+/-}$ on the other hand showed very few terms enriched in up-regulated genes (Fig. 5.3d). In contrast to the earlier time point synaptic genes, particularly those related to GABAergic synapses, were now specifically enriched in down-regulated genes.

The slight differences in functional enrichment for the overlap between *ASH1L*$^{+/-}$, *KDM6B*$^{+/-}$, and *SETD5*$^{+/-}$ were also apparent at the level of over-representation of transcription factor targets (Fig. 5.3e). Down-regulated genes were mostly enriched for targets of transcription factors related to the cell cycle, such as those of the E2F family [369], as well as targets of TCF3 (but not β-catenin), which itself was down-regulated. For up-regulated genes we again found enrichment for REST, as well as SMARCD1 targets. In addition, SMARCA4, a score 1 SFARI gene, was also among the transcription factors identified at this time point expanding the potential link to the SWI/SNF complex. Lastly, we found enrichment for two factors previously linked to psychiatric disorders, POU3F2 (BRN2) and OLIG2 [370, 371]. It should be noted that, although significantly differentially expressed according to our criteria, OLIG2 overall expression was still low across samples (max cpm = 2.5).

Application of the same cell type deconvolution approach before revealed two important findings. First, spheroids across all cell lines assayed initiated the production of excitatory neurons as indicated by an increased proportion of this cell type compared to day 25. Second and most importantly, overproduction of inhibitory neurons in *ASH1L*$^{+/-}$, *KDM6B*$^{+/-}$, and *SETD5*$^{+/-}$ mutant spheroids had further progressed since day 25 and now reached estimates of 21 - 64% across all four lines, compared to 1 - 14% in wildtype or *KMT5B*$^{+/-}$ lines (Fig. 5.3f). The increase in both these cell types resulted in a marked

decrease of the estimate for proliferative progenitors. This was in line with both our functional enrichment results and observations of spheroid growth at the macroscopic level.

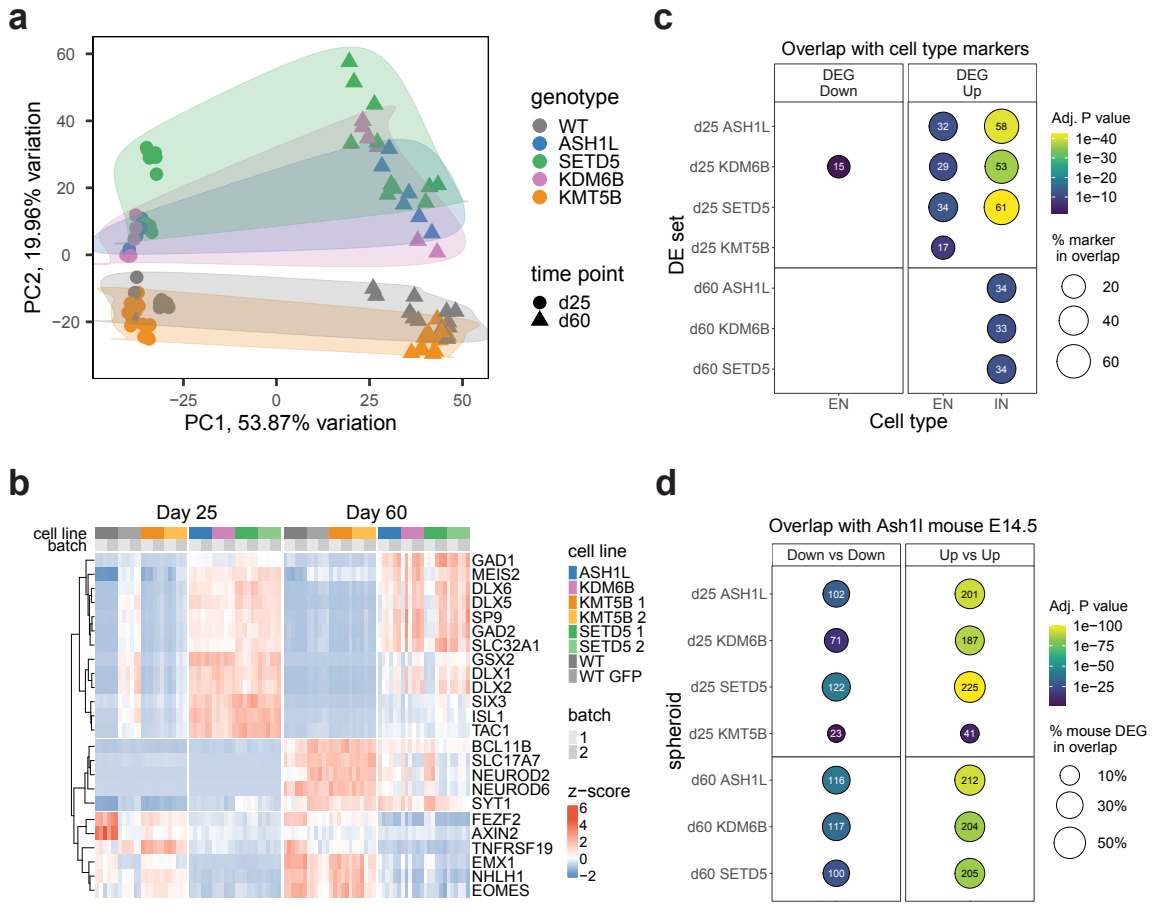
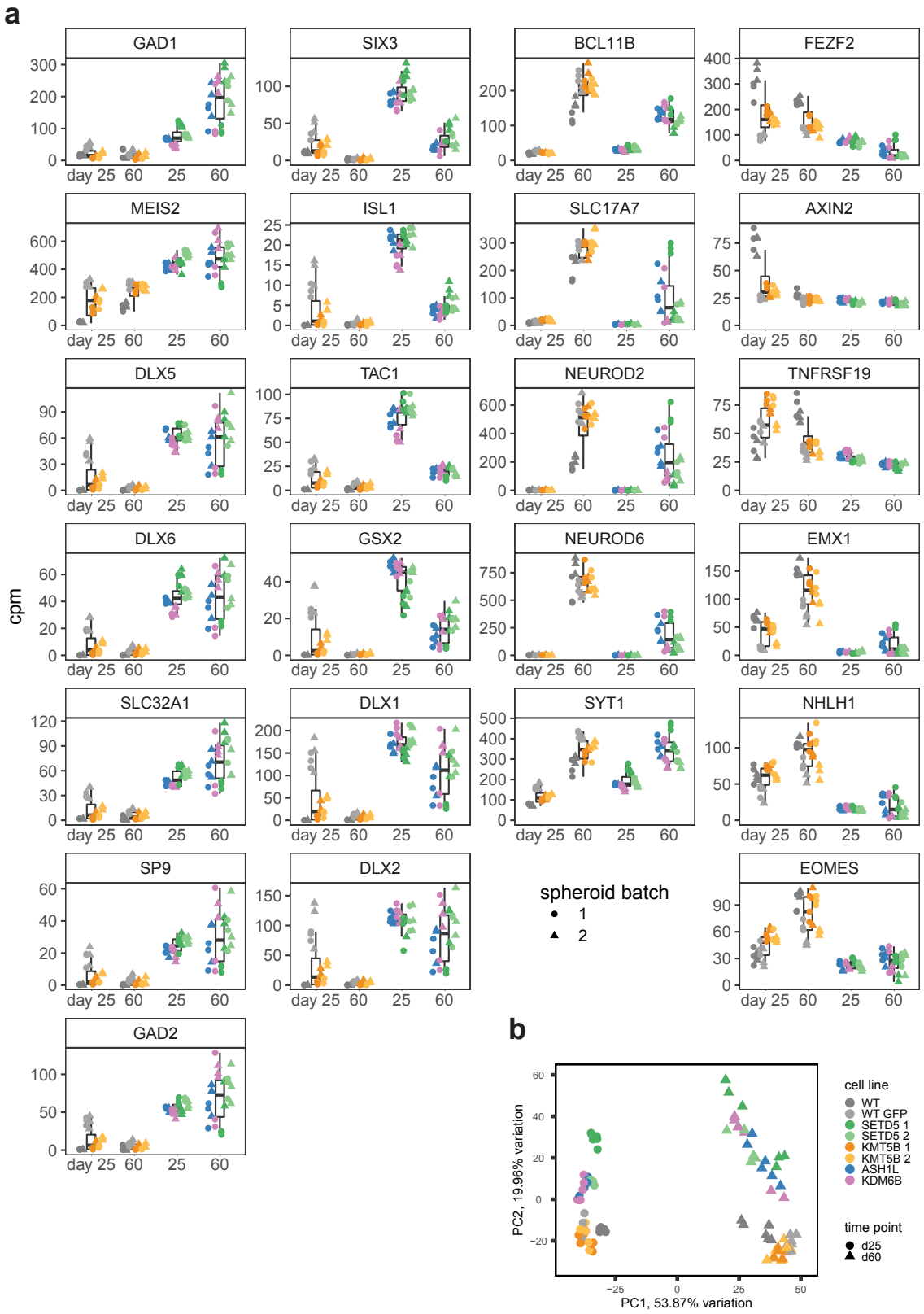


Figure 5.4: **Divergent development across multiple time points a)** Combined principal component analysis including all samples from both day 25 and day 60. Data corrected to remove unwanted variation (see methods). **b)** Enrichment for excitatory (EN) and inhibitory neuron (IN) markers from human striatal organoids. Color indicates multiple-testing adjusted enrichment P value, circle size scales with the percentage of marker genes in the overlap. Numbers of overlapping genes is shown inside the circles. **c)** Heatmap of normalized expression for the listed marker genes. Data is presented as gene-wise z-scores calculated from read counts normalized as counts per million reads (cpm). **d)** Comparison of cortical spheroid data with DEGs between whole brain samples of wildtype and *Ash1l*$^{+/-}$ E14.5 mouse embryos. Color indicates multiple-testing adjusted enrichment P value, circle size scales with the percentage of mouse DEGs in the overlap. Numbers of overlapping genes is shown inside the circles. Note the difference in color scale between b and d. For information per sample/cell line for panels a) and c) please refer to Fig. S5.4.

In summary, our results show that the divergence in development persisted and was even exacerbated over time. Using principal component analysis of all samples together (Fig. 5.4a) we managed to visualize this summary. This analysis highlighted that, after separating by time (principal component (PC) 1), samples of both time points separated along a shared second axis (PC2) in a remarkably consistent and genotype-specific fashion reminiscent of the clustering seen at each separate time point. To get a more fine-grained understanding of the nature of these diverging fates, we looked into the expression of cell type markers across time points, focusing on differences between *ASH1L*$^{+/-}$, *KDM6B*$^{+/-}$, and *SETD5*$^{+/-}$ mutants and the other lines (Fig. 5.4b).



Supplementary Figure S5.4: **Extended information for Figure 5.4. a)** Visualization of the same data that went into the heatmap in Fig. 5.4b in a cell line and batch specific fashion. Cell lines are grouped into the three-way overlap group of *ASH1L*$^{+/-}$, *KDM6B*$^{+/-}$, and *SETD5*$^{+/-}$ on the right and *KMT5B*$^{+/-}$ and wild type lines on the left. Box plots show summary statistics for each of those groups. Color legend see b). **b)** Principal component analysis plot as in Fig. 5.4a but colored by cell line.

Starting from day 25, we observed consistent up-regulation of early interneuron markers of the DLX family. In addition we found increased expression of markers of GABAergic neurons such as glutamate decarboxylases 1 and 2 (*GAD1/GAD67* and *GAD2/GAD65*) and the vesicular GABA transporter (*SLC32A1/VGAT*). In stark contrast, we found consistent down-regulation of markers of the excitatory neuron fate such as *NEUROD2*, *NEUROD6*, vesicular glutamate transporter 1 (*SLC17A7*), markers of deep layer cortical neurons (*BCL11B/CTIP2* and *FEZF2*) [372], and intermediate progenitor cells (*TBR2/EOMES*).

However, despite the mounting evidence for the acquisition of an inhibitory neuron fate by a subpopulation of cells in spheroids derived from $ASH1L^{+/-}$, $KDM6B^{+/-}$, and $SETD5^{+/-}$ lines we did not find expression for the medial ganglionic eminence marker NKX2-1, which was abundantly expressed in ventralized spheroids resembling the subpallium [138]. For this reason we looked into other possible inhibitory neuron fates. Taking our earlier finding of potential involvement of Islet-1 (ISL1) in the regulation of DEGs at day 25 as an indication, we considered the possibility of a lateral ganglionic eminence (LGE) fate. We consistently found increased expression of several LGE markers such as *GSX2* and *MEIS2* [373, 374], as well as *SIX3*, *SP9*, and *TAC1* related to striatal medium spiny neurons, which originate from the LGE [375–377]. In line with the results discussed above suggesting dysregulation of Wnt signaling, we also found, especially at day 25, increased expression of Wnt readout genes *AXIN2* and *TNFRSF19* [378–381].

To further underline the connection between our observed phenotype and LGE/striatal cell fates, we compared our DEGs with excitatory (EN) and inhibitory neuron (IN) markers compiled from a study on the generation of striatal organoids [139]. We found strong enrichment for IN markers in up-regulated genes of $ASH1L^{+/-}$, $KDM6B^{+/-}$, and $SETD5^{+/-}$ mutants, especially at day 25. No IN marker enrichment was found for $KMT5B^{+/-}$ DEGs. In contrast, excitatory neuron markers showed enrichment with higher adjusted P values at day 25 and none at day 60.

Lastly, in order to validate our findings in an *in vivo* context, we compared our differential expression results with genes dysregulated in mouse brain samples of E14.5 $Ash1l^{+/-}$ mouse embryos. Despite the difference in model system we found strong enrichment of mouse DEGs in the DEGs we identified in the spheroid system, especially between up-regulated genes in both models. This enrichment was much stronger for $ASH1L^{+/-}$, $KDM6B^{+/-}$, and $SETD5^{+/-}$ sets than for $KMT5B^{+/-}$ again underlining the phenotypic differences between these mutants.

Taken together, our results show divergent development in three of four mutants. However, the perturbations only partially divert cells towards the alternative fate, which we identified as reminiscent of cell fates connected to the lateral ganglionic eminence. These cells following down this new lineage path make up the foundation of the shared phenotype we observe in $ASH1L^{+/-}$, $KDM6B^{+/-}$, and $SETD5^{+/-}$ spheroids.

5.5 Discussion

In this study we investigated regulatory convergence among risk genes for ASD. For this, we employed cortical spheroids, a human *in vitro* model system of early cortical development. We generated heterozygous mutant embryonic stem cell lines on an isogenic background for four chromatin modifiers implicated in ASD: *ASH1L*, *KDM6B*, *KMT5B*, and *SETD5*; and focused on the transcriptional perturbations resulting from these mutations.

Acquisition of an alternate inhibitory neuron identity as a shared phenotype between *ASH1L*$^{+/-}$, *KDM6B*$^{+/-}$, and *SETD5*$^{+/-}$ mutant spheroids.

Generating cortical spheroids from these lines, we first observed differences in growth trajectories at the macroscopic level which preliminarily segregated three of the candidate genes, *ASH1L*, *KDM6B*, and *SETD5*, into their own subgroup. The core of the study so far is formed by an investigation of changes to the transcriptional landscape at multiple stages of cortical spheroid development. This analysis recapitulated our target gene grouping and revealed a remarkably high overlap between *ASH1L*$^{+/-}$, *KDM6B*$^{+/-}$, and *SETD5*$^{+/-}$ lines at the transcriptional level. We found that at day 25 spheroids had acquired a partial inhibitory, ventralized identity, contrary to the design of the guided differentiation protocol, which should enforce a dorsal excitatory neuron identity [144]. At day 60 we observed that this divergence was further exacerbated, with relative abundance of inhibitory neurons, as estimated by cell type deconvolution, reaching up to 64%. In addition multiple results pointed towards a reduction of the proliferative capabilities of the spheroids, e.g. from a relative depletion of cycling progenitors, which could explain the differences in growth trajectories. However, it should be emphasized that these spheroids also developed along the main route as exemplified by the increase in the proportion of excitatory neurons across all lines. This parallel presence of both identities within the same system suggests the possibility of an underlying stochastic process in cell fate determination whose parameters were altered by heterozygous loss of either of these three genes.

In contrast to the cell types observed in a ventral spheroid protocol developed by the same group [138], our spheroids showed a distinct lack of *Nkx2-1* expression at any time point, indicating an identity other than medial ganglionic eminence (MGE) derived interneurons. This led us to investigate the gene expression profiles of our spheroids in more detail to potentially unearth the nature of this additional differentiation path. Based on our finding that ISL1, an LGE marker, was up-regulated and its targets enriched in our differentially expressed genes, we hypothesized that a proportion of cells might have acquired an LGE identity. We substantiated this hypothesis by checking the expression of related marker genes such as *MEIS2* and *GSX2*. Additional evidence was provided by a comparison with spheroids derived from a protocol specifically designed to generate striatal organoids of LGE origin. The overlap between cell type markers and differentially expressed genes in our system was highly significant and even more so than the same analysis done using markers derived from the original cortical spheroid protocol (data not shown). Based on the data available so far, we conclude from this that the degree of ventralization is high enough for a transition between dorsal cortex and ganglionic eminences but not strong enough to reach an MGE associated identity.

The fact that this is data generated from an *in vitro system* begs the question how transferable these results are to *in vivo* development. To answer this question we compared

our results with data derived from a comparison between brain samples of *Ash1l*$^{+/-}$ and wild type mouse embryos. We were surprised by the substantial overlap of DEGs, given the multitude of differences between the model systems. Further work will be needed to unravel this connection, but this first glimpse is very promising.

Speculation on potential signaling pathways involved in bringing forth the shared phenotype

Despite this intriguing overlap with the mouse model we refrain from suggesting that the potential switch from dorsal excitatory to more ventral inhibitory identities is akin to the actual phenotype one would observe *in vivo*, i.e. in patients. Rather, we propose that these results be taken as pointers towards developmental pathways that might be particularly affected by these mutations. Many such pathways are very tightly regulated, both temporally and spatially, and even slight deviations might lead to very strong effects, particularly in a reduced *in vitro* system lacking many compensatory mechanisms. This viewpoint opens up a number of interesting avenues of interpretation for the results presented here. One way to tackle this question is to investigate *in vitro* differentiation protocols designed to induce a similar identity as the one we potentially observed here. Information of the pathway modulations applied in these protocols in conjunction with our observations might highlight potential key players.

Previous studies on *in vitro* differentiation of stem cells towards LGE or striatal identities followed two main strategies: activation of sonic hedgehog (SHH) signaling with or without the inhibition of Wnt, or treatment with activin A. SHH signaling is a major player in the specification of the ventral forebrain identity [382] and its activation has been shown to induce differentiation towards an MGE/LGE identity [383, 384]. In addition, the inhibition of Wnt signaling has been shown to work in a synergistic manner together with SHH activation [385, 386] and the combination of the two has been previously used for *in vitro* differentiation into LGE as well as MGE progenitors and derivatives [138, 386–388]. On the other hand, treatment with activin A, a member of the family of transforming growth factor β proteins [389], has been shown to specifically target differentiation towards an LGE (or caudal ganglionic eminence, dependent on the timing) rather than MGE identity in an SHH-independent manner [139, 390, 391]. This provides an interesting alternative pathway to an LGE identity. However, given the fact that this specification is inhibited by SB431542, a factor used in our protocol during the initial phase of dual SMAD inhibition, further data from an earlier time point is required to gain more insight.

In our dataset we observed down-regulation of downstream Wnt signaling targets such as *AXIN2* and *TNFRSF19*, as well as changes in *WNT3A* and *WNT8B* at day 25. Further evidence for the involvement of Wnt signaling, at least at the earlier time point, comes from the enrichment of beta-catenin and TCF3 targets in down-regulated genes. On the other hand, we only detected very low levels of *SHH* but some significant changes in SHH-signaling related genes [392] such as *PTCH1* (max 2.5-fold up), *GAS1* (max 2-fold down) at both time points, and *BOC* and *CDON* down-regulated at day 60 only. While these results implicate Wnt signaling as a focus point, the role of Sonic hedgehog signaling is less conclusive. However, crosstalk between the two pathways through the regulation of GLI3 [386] provides an intriguing possible explanation. This study reports that either SHH-signaling activation or weak SHH activation in addition to Wnt inhibition induce an MGE identity as measured by *NKX2-1* expression. Interestingly, for the trial performed using only Wnt inhibition in this setting they observed absence of *NKX2-1* and

presence of *PAX6*, a marker combination also agreeing with a (dorsal) LGE identity [393]. Finally, the ventralizing effect of Wnt inhibition alone in the absence of SHH activation was demonstrated by Nicoleau et al. [394]. They showed that prolonged Wnt inhibition by the small molecule XAV-939 induces expression and detectable immunostaining for *GSX2* while no such staining was detected for NKX2-1.

In summary, the results collected thus far in this study indicate an impact of heterozygous loss of *ASH1L*, *KDM6B*, and *SETD5* on the specification of the dorsal-ventral axis in an *in vitro* model system of cortical development. Cross-referencing *in vitro* protocols targeting cells towards ventral identities with our results implicate Wnt signaling as a potential agent in this dysregulation while Sonic hedgehog signaling might play a lesser role. Therefore the ventralization phenotype we see could be conceptualized in terms of negative modulation of Wnt signaling in these mutants. Likewise it could potentially be rescued by the well-timed and properly scaled re-activation of this pathway. Of note, all three of these genes have been linked to the regulation of Wnt signaling in varying contexts such as neurodevelopment for *SETD5* in our previous study [167], mesendoderm development for *KDM6B* [395], and human pluripotent stem cells for *ASH1L* [337]. Of interest, the latter study screened a number of ASD risk genes for modulation of the Wnt response in stem cells and found additional genes shared the hyporesponsiveness seen for *ASH1L*, thereby hinting at the possibility that this might be a more widespread phenotype. However, given the context dependency of Wnt signaling [396, 397], this will need to be confirmed in a neurodevelopmental context. Lastly, the only other gene from this study they investigated, *KMT5B*, did not share this phenotype, supporting the distinction we observe in our model.

Based on previous studies, this intervention should be timed after the initial SMAD inhibition phase since early Wnt activation extinguishes anterior in favor of posterior cell identities [398]. Previous studies have shown that modulation of Wnt signaling can rescue increased ventralization in directed differentiation paradigms through the application of the Wnt activator CHIR-99021 [378]. Furthermore, it was applied in the rescue of overall organoid growth in the context of Miller-Dieker syndrome, a cortical malformation [149]. These precedents provide an exciting basis for rescue experiments in the near future.

Conclusion

At the time of writing this study is a still ongoing project and more experiments and analyses will be needed to bring it to a close. However, I am of the opinion that the evidence presented here already provides compelling insights into a potentially shared neurobiological phenotype across several risk genes for ASD. *In vitro* model systems like cortical spheroids come with a number of caveats, with incompleteness of the system being at the forefront. The lack of certain cell types as well as regional guidance cues likely perturbs the system's compensatory mechanisms leading to potentially exaggerated phenotypes. Despite or rather because of that, systems like these might provide us with a magnifying glass which could reveal mechanistic insights otherwise too subtle or confounded by other factors to be detected.

6

Conclusions and future perspectives

Autism spectrum disorder (ASD) is characterized by a high degree of heterogeneity, both in diagnosis and etiology. On the one hand, the difficulty in delineating different conditions based on behavioral diagnosis alone led to the agglomeration of autism related conditions under the umbrella term autism spectrum disorders. On the other hand, the advent of next-generation sequencing ushered in a new era for the field of ASD genetics. A large number of genes has since been linked to ASD risk, making it difficult if not impossible to find "the" cause of ASD. However, the fact that there is still convergence onto similar diagnoses, heterogeneous as they may be, has ignited hope that we might yet find common mechanisms for at least subgroups of genetic causes. Previous studies have identified transcriptional regulation to be one of the biggest groups of ASD risk genes.

Over the course of my PhD, I studied multiple of these transcriptional regulators in various contexts. In studies focusing on single genes we uncovered important functions of chromatin regulators. We found that heterozygous mutations in *Setd5* in mice, as well as *CHD8* in cerebral organoids, led to changes in cell fate determination in early development. Furthermore, using behavioral assays linked to transcriptomic readouts, we found that heterozygous loss of *Setd5* alters the gene regulatory program triggered by neuronal activity. These exciting results indicate that loss of chromatin modifiers is most visible in dynamic processes, such as differentiation or reaction to stimuli. In both these contexts the tight regulation of signaling pathways and their transcriptional output plays an important role. It is therefore plausible that chromatin modifiers act through the modulation of these pathways and that perturbation of these factors results in an altered relationship between stimuli and response. Furthermore, our findings indicate a role for ASD related chromatin modifiers in neuronal function in adults. If these findings can be replicated in human cells and generalized over multiple ASD risk genes, this will open up new possibilities for potential treatments of the debilitating ASD cases that are the focus of our research.

In my main project I focused explicitly on the topic of regulatory convergence by studying a group of chromatin modifiers related to ASD. Modeling early neurodevelopment using cortical spheroids, this study revealed a shared alteration of cell fate determination in a remarkably high proportion of three out of four genes assayed. We observed that these mutants partially escape the guided differentiation laid out by the protocol in favor of more ventral cell fates. While this process likely does not happen at the same scale *in vivo*

my results could also be interpreted as a readout for alterations in developmental signaling pathways. At the time of writing this thesis this project is still ongoing with a shift in focus towards earlier development. This will allow me to get to the root of the divergence in development we found and potentially shed light on the pathway perturbation at play here. In this regard, our results so far suggest a role for Wnt signaling. Given the accessibility of modulators of this pathway, its involvement opens up the exciting possibility of rescue experiments. In the end, the results from this study could potentially link three ASD risk genes through common biology and create a first stepping stone towards an era of ASD diagnosis aided by biological criteria.

Altogether, altered differentiation and the resulting effects on cell fate determination emerged as a common thread across multiple ASD risk genes in both *in vivo* and *in vitro* model systems. This is in line with the hypothesis stated at the beginning of this thesis that such processes are particularly vulnerable to dosage alterations in their regulators. Additional evidence for this hypothesis was also provided from our results that heterozygous loss of *Setd5* lead to an altered transcriptional response to neuronal activity in a behavioral paradigm. However, in order to gauge the importance and translatability of such results to a patient context several points have to be considered.

First, the phenotypes observed in model systems likely don't fully align with the situation observed in patients. I would argue that this is especially true for results observed in organoids due to the reduced nature of these models which potentially lack vital compensatory mechanisms present *in vivo*. It is therefore paramount to discern the exact form of the observed phenotypes, particularly in the context of studying convergence across multiple genetic etiologies. Rephrasing this question in terms of the concept of causal paths introduced above, one needs to define the earliest point of intersection of such paths. This is important as each step has the potential to introduce model system specific effects so convergence at later stages has a lower chance of translating to the *in vivo* human context. In the case of altered differentiation trajectories, as observed here, this point could either be found at the cell type level - in that the altered abundances are the phenotype - or further upstream, for example at the level of signaling pathways. In the former case the observed convergence is more likely to be model specific, which diminishes the probability of these findings translating to a patient context. On the other hand, if the point of convergence were to be found at the pathway level, for example Wnt signaling, this finding would have a greater likelihood of translating across contexts. Despite that, downstream phenotypes such as the observed ventralization will probably manifest in a context-specific manner due to differences in compensatory mechanisms. Additional experiments targeted at earlier time points and generating data on both mRNA levels and open chromatin will shed further light on the regulatory underpinnings of the phenotype and will ideally help us reveal the shared mechanism.

Second, the goal of finding translatable results is ultimately to use this knowledge for the development of treatments for such severe cases of ASD/ID. For this, the uncovered phenotypes are ideally amenable to pharmacological or other interventions. Given that treatments are usually only administered after diagnosis this precludes many developmental and cell type specification phenotypes from proper treatment as the time window for intervention is likely already closed at this point. However, despite the assumption of neurodevelopmental disorders being irreversible in adulthood, there are examples of rescue experiments reversing some of the observed deficits in mouse models [399]. Examples include rescue of neurological symptoms in models of Rett syndrome [400–404], synaptic

plasticity in Down syndrome [405], and behavioral symptoms in Fragile X syndrome [406–409]. More recently, the treatment of mice with a conditional knockout of *Tbr1* in layer 5 neurons with a Wnt agonist could rescue synaptic defects found in this model [410, 411]. The restoration of Wnt signaling has also been shown to be efficacious in a mouse model of Rett syndrome [412].

In summary, while the neurodevelopmental aspects involving early differentiation might not be accessible for treatment yet, exciting results like those mentioned here suggest the possibility that symptoms related to neuronal function are in fact amenable to treatment. The finding that Wnt modulation is included in the list of potential treatments is of particular interest for the results presented here given that the potential involvement of this pathway runs like a red thread throughout. Further studies will be needed to elucidate connections between mutations and signaling pathways, and also delineate both their developmental as well as functional consequences. In the end, this will enable us to categorize severe ASD/ID cases based on biological information as well as give us an idea to which extent these detrimental conditions may be targeted by treatment.

Bibliography

[1] Elena Deliu* et al. “Haploinsufficiency of the intellectual disability gene SETD5 disturbs developmental gene expression and cognition”. In: *Nature Neuroscience* 21.12 (Nov. 2018), pp. 1717–1727.

[2] Carlo Emanuele Villa et al. “CHD8 haploinsufficiency links autism to transient alterations in excitatory and inhibitory trajectories”. In: *Cell Reports* 39.1 (Apr. 2022), p. 110615.

[3] American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, 5th ed.* American Psychiatric Association, May 2013.

[4] Matthew J. Maenner et al. “Prevalence And Characteristics Of Autism Spectrum Disorder Among Children Aged 8 Years — Autism And Developmental Disabilities Monitoring Network, 11 Sites, United States, 2018”. In: *MMWR. Surveillance Summaries* 70.11 (Dec. 2021), pp. 1–16.

[5] Spectrum. *Autism by the numbers: Explaining its apparent rise*. Simons Foundation, 2021.

[6] Hans Asperger. “Das psychisch abnorme Kind”. In: *Wiener Klinische Wochenschrift* 49 (1938), pp. 1314–1317.

[7] Hans Asperger. “Die „Autistischen Psychopathen” im Kindesalter”. In: *Archiv für Psychiatrie und Nervenkrankheiten* 117.1 (1944), pp. 76–136.

[8] Leo Kanner. “Autistic disturbances of affective contact”. In: *Nervous child* 2.3 (1943), pp. 217–250.

[9] Nick Chown and Liz Hughes. “History and First Descriptions of Autism: Asperger Versus Kanner Revisited”. In: *Journal of Autism and Developmental Disorders* 46.6 (Feb. 2016), pp. 2270–2272.

[10] Nicole E. Rosen, Catherine Lord, and Fred R. Volkmar. “The Diagnosis of Autism: From Kanner to DSM-III to DSM-5 and Beyond”. In: *Journal of Autism and Developmental Disorders* 51.12 (Feb. 2021), pp. 4253–4270.

[11] Michael Rutter. “CONCEPTS OF AUTISM: A REVIEW OF RESEARCH”. In: *Journal of Child Psychology and Psychiatry* 9.1 (Oct. 1968), pp. 1–25.

[12] Susan Folstein and Michael Rutter. “Genetic influences and infantile autism”. In: *Nature* 265.5596 (Feb. 1977), pp. 726–728.

[13] E. R. Ritvo et al. “Concordance for the syndrome of autism in 40 pairs of afflicted twins.” In: *The American journal of psychiatry* 142 (1 Jan. 1985), pp. 74–77. ppublish.

[14] Jan Wahlström et al. “Infantile autism and the fragile X. A swedish multicenter study”. In: *American Journal of Medical Genetics* 23.1-2 (Jan. 1986), pp. 403–408.

[15] Christopher Gillberg and Jan Wahlström. “Chromosome abnormalities in infantile autism and other childhood psychoses: a population study of 66 cases”. In: *Developmental Medicine & Child Neurology* 27.3 (Nov. 1985), pp. 293–304.

[16] Michael Rutter. “Diagnosis and definition of childhood autism”. In: *Journal of Autism and Childhood Schizophrenia* 8.2 (June 1978), pp. 139–161.

[17] American Psychiatric Association. *Diagnostic and statistical manual of mental disorders, 3rd ed.* American Psychiatric Association Washington, DC, 1980.

[18] American Psychiatric Association. *Diagnostic and statistical manual of mental disorders, 3rd ed, revised.* American Psychiatric Association Washington, DC, 1987.

[19] World Health Organization. *The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines*. World Health Organization, 1992.

[20] American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, 4th Ed.* Vol. 3. American Psychiatric Association Washington, DC, 1994.

[21] Kathleen E. Macintosh and Cheryl Dissanayake. “Annotation: The similarities and differences between autistic disorder and Asperger's disorder: a review of the empirical evidence”. In: *Journal of Child Psychology and Psychiatry* 45.3 (Mar. 2004), pp. 421–434.

[22] Darlene R. Walker et al. “Specifying PDD-NOS: A Comparison of PDD-NOS, Asperger Syndrome, and Autism”. In: *Journal of the American Academy of Child & Adolescent Psychiatry* 43.2 (Feb. 2004), pp. 172–180.

[23] Andrea N. Witwer and Luc Lecavalier. “Examining the Validity of Autism Spectrum Disorder Subtypes”. In: *Journal of Autism and Developmental Disorders* 38.9 (Mar. 2008), pp. 1611–1624.

[24] Anne V. Snow and Luc Lecavalier. “Comparing Autism, PDD-NOS, and Other Developmental Disabilities on Parent-Reported Behavior Problems: Little Evidence for ASD Subtype Validity”. In: *Journal of Autism and Developmental Disorders* 41.3 (June 2010), pp. 302–310.

[25] Catherine Lord. “A Multisite Study of the Clinical Diagnosis of Different Autism Spectrum Disorders”. In: *Archives of General Psychiatry* 69.3 (Mar. 2012), p. 306.

[26] Marianne Mordre et al. “Is long-term prognosis for pervasive developmental disorder not otherwise specified different from prognosis for autistic disorder? Findings from a 30-year follow-up study”. In: *Journal of Autism and Developmental Disorders* 42.6 (June 2011), pp. 920–928.

[27] Daniel H. Geschwind. “Advances in Autism”. In: *Annual Review of Medicine* 60.1 (Feb. 2009), pp. 367–380.

[28] Ruthie E. Amir et al. “Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2”. In: *Nature Genetics* 23.2 (Oct. 1999), pp. 185–188.

[29] Catherine Lord and Somer L. Bishop. “Recent Advances in Autism Research as Reflected in DSM-5 Criteria for Autism Spectrum Disorder”. In: *Annual Review of Clinical Psychology* 11.1 (Mar. 2015), pp. 53–70.

[30] Amy S. Weitlauf et al. “Brief Report: DSM-5 “Levels of Support:” A Comment on Discrepant Conceptualizations of Severity in ASD”. In: *Journal of Autism and Developmental Disorders* 44.2 (June 2013), pp. 471–476.

[31] Trent Gaugler et al. “Most genetic risk for autism resides with common variation”. In: *Nature Genetics* 46.8 (July 2014), pp. 881–885.

[32] Beata Tick et al. “Heritability of autism spectrum disorders: a meta-analysis of twin studies”. In: *Journal of Child Psychology and Psychiatry* 57.5 (Dec. 2015), pp. 585–595.

[33] Sven Sandin et al. “The Heritability of Autism Spectrum Disorder”. In: *JAMA* 318.12 (Sept. 2017), p. 1182.

[34] Dan Bai et al. “Association of Genetic and Environmental Factors With Autism in a 5-Country Cohort”. In: *JAMA Psychiatry* 76.10 (Oct. 2019), p. 1035.

[35] Sherlly Xie et al. “The Familial Risk of Autism Spectrum Disorder with and without Intellectual Disability”. In: *Autism Research* 13.12 (Oct. 2020), pp. 2242–2250.

[36] Sara C. Taylor et al. “Heritability of quantitative autism spectrum traits in adults: A family-based study”. In: *Autism Research* 14.8 (July 2021), pp. 1543–1553.

[37] Daniel H. Geschwind and Jonathan Flint. “Genetics and genomics of psychiatric disease”. In: *Science* 349.6255 (Sept. 2015), pp. 1489–1494.

[38] Jonathan Sebat et al. “Strong Association of De Novo Copy Number Mutations with Autism”. In: *Science* 316.5823 (Apr. 2007), pp. 445–449.

[39] M-L Jacquemont et al. “Array-based comparative genomic hybridisation identifies high frequency of cryptic chromosomal rearrangements in patients with syndromic autism spectrum disorders”. In: *Journal of Medical Genetics* 43.11 (Nov. 2006), pp. 843–849.

[40] Lauren A. Weiss et al. “Association between Microdeletion and Microduplication at 16p11.2 and Autism”. In: *New England Journal of Medicine* 358.7 (Feb. 2008), pp. 667–675.

[41] Christian R. Marshall et al. “Structural Variation of Chromosomes in Autism Spectrum Disorder”. In: *The American Journal of Human Genetics* 82.2 (Feb. 2008), pp. 477–488.

[42] Autism Genome Project Consortium et al. “Mapping autism risk loci using genetic linkage and chromosomal rearrangements.” In: *Nature genetics* 39 (3 Mar. 2007), pp. 319–328. ppublish.

[43] R. A. Kumar et al. “Recurrent 16p11.2 microdeletions in autism”. In: *Human Molecular Genetics* 17.4 (Nov. 2007), pp. 628–638.

[44] Dalila Pinto et al. “Functional impact of global rare copy number variation in autism spectrum disorders”. In: *Nature* 466.7304 (June 2010), pp. 368–372.

[45] Brian J O'Roak et al. "Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations." In: *Nature genetics* 43 (6 June 2011), pp. 585–589.

[46] Brian J O'Roak et al. "Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations." In: *Nature* 485 (7397 Apr. 2012), pp. 246–250.

[47] Benjamin M Neale et al. "Patterns and rates of exonic de novo mutations in autism spectrum disorders." In: *Nature* 485 (7397 Apr. 2012), pp. 242–245.

[48] Ivan Iossifov et al. "De Novo Gene Disruptions in Children on the Autistic Spectrum". In: *Neuron* 74.2 (Apr. 2012), pp. 285–299.

[49] Ivan Iossifov et al. "The contribution of de novo coding mutations to autism spectrum disorder." In: *Nature* 515 (7526 Nov. 2014), pp. 216–221.

[50] Silvia De Rubeis et al. "Synaptic, transcriptional and chromatin genes disrupted in autism." In: *Nature* 515 (7526 Nov. 2014), pp. 209–215.

[51] Stephan J. Sanders et al. "De novo mutations revealed by whole-exome sequencing are strongly associated with autism." In: *Nature* 485 (7397 Apr. 2012), pp. 237–241.

[52] Stephan J Sanders et al. "Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci." In: *Neuron* 87 (6 Sept. 2015), pp. 1215–1233.

[53] Ryan K C Yuen et al. "Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder". In: *Nature Neuroscience* 20.4 (Mar. 2017), pp. 602–611.

[54] Pamela Feliciano et al. "Exome sequencing of 457 autism families recruited online provides evidence for autism risk genes". In: *npj Genomic Medicine* 4.1 (Aug. 2019).

[55] F. Kyle Satterstrom et al. "Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism". In: *Cell* 180.3 (Feb. 2020), 568–584.e23.

[56] Behrang Mahjani et al. "Prevalence and phenotypic impact of rare potentially damaging variants in autism spectrum disorder". In: *Molecular Autism* 12.1 (Oct. 2021).

[57] Xin He et al. "Integrated Model of De Novo and Inherited Genetic Variants Yields Greater Power to Identify Risk Genes". In: *PLoS Genetics* 9.8 (Aug. 2013). Ed. by Scott M. Williams, e1003671.

[58] Christina A. Castellani and Dan E. Arking. "High-Risk, High-Reward Genetics in ASD". In: *Neuron* 105.3 (Feb. 2020), pp. 407–410.

[59] Brett S Abrahams et al. "SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs)". In: *Molecular Autism* 4.1 (Oct. 2013).

[60] Huda Y. Zoghbi. "Postnatal Neurodevelopmental Disorders: Meeting at the Synapse?" In: *Science* 302.5646 (Oct. 2003), pp. 826–830.

[61] Josefa M Sullivan, Silvia De Rubeis, and Anne Schaefer. "Convergence of spectrums: neuronal gene network states in autism spectrum disorder". In: *Current Opinion in Neurobiology* 59 (Dec. 2019), pp. 102–111.

[62] Dalila Pinto et al. "Convergence of Genes and Cellular Pathways Dysregulated in Autism Spectrum Disorders". In: *The American Journal of Human Genetics* 94.5 (May 2014), pp. 677–694.

[63] Jonathan Chang et al. "Genotype to phenotype relationships in autism spectrum disorders". In: *Nature Neuroscience* 18.2 (Dec. 2014), pp. 191–198.

[64] Jingjing Li et al. "Integrated systems analysis reveals a molecular network underlying autism spectrum disorders". In: *Molecular Systems Biology* 10.12 (Dec. 2014), p. 774.

[65] Bin Zhang and Steve Horvath. "A General Framework for Weighted Gene Co-Expression Network Analysis". In: *Statistical Applications in Genetics and Molecular Biology* 4.1 (Jan. 2005).

[66] E Ben-David and S Shifman. "Combined analysis of exome sequencing points toward a major role for transcription regulation during brain development in autism". In: *Molecular Psychiatry* 18.10 (Nov. 2012), pp. 1054–1056.

[67] Neelroop N Parikshak et al. "Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism." In: *Cell* 155 (5 Nov. 2013), pp. 1008–1021.

[68] A Jeremy Willsey et al. "Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism." In: *Cell* 155 (5 Nov. 2013), pp. 997–1007.

[69] Helen Rankin Willsey et al. "Parallel in vivo analysis of large-effect autism genes implicates cortical neurogenesis and estrogen in risk and resilience". In: *Neuron* 109.5 (Mar. 2021), 788–804.e8.

[70] R. Nick Hernandez et al. “Autism spectrum disorder in fragile X syndrome: A longitudinal evaluation”. In: *American Journal of Medical Genetics Part A* 149A.6 (June 2009), pp. 1125–1137.

[71] Jennifer C. Darnell et al. “FMRP Stalls Ribosomal Translocation on mRNAs Linked to Synaptic Function and Autism”. In: *Cell* 146.2 (July 2011), pp. 247–261.

[72] Manuel Ascano et al. “FMRP targets distinct mRNA sequence elements to regulate protein expression”. In: *Nature* 492.7429 (Dec. 2012), pp. 382–386.

[73] Sebastien M. Weyn-Vanhentenryck et al. “HITS-CLIP and Integrative Modeling Define the Rbfox Splicing-Regulatory Network Linked to Brain Development and Autism”. In: *Cell Reports* 6.6 (Mar. 2014), pp. 1139–1152.

[74] Raphael Bernier et al. “Disruptive CHD8 Mutations Define a Subtype of Autism Early in Development”. In: *Cell* 158.2 (July 2014), pp. 263–276.

[75] Ivan Iossifov et al. “Low load for disruptive mutations in autism genes and their biased transmission”. In: *Proceedings of the National Academy of Sciences* 112.41 (Sept. 2015), E5600–E5607.

[76] Aarathi Sugathan et al. “CHD8 regulates neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors”. In: *Proceedings of the National Academy of Sciences* 111.42 (Oct. 2014), E4468–E4477.

[77] Justin Cotney et al. “The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment”. In: *Nature Communications* 6.1 (Mar. 2015).

[78] Stephan J Sanders. “First glimpses of the neurobiology of autism spectrum disorder”. In: *Current Opinion in Genetics & Development* 33 (Aug. 2015), pp. 80–92.

[79] Veronica B. Searles Quick, Belinda Wang, and Matthew W. State. “Leveraging large genomic datasets to illuminate the pathobiology of autism spectrum disorders”. In: *Neuropsychopharmacology* 46.1 (July 2020), pp. 55–69.

[80] Irina Voineagu et al. “Transcriptomic analysis of autistic brain reveals convergent molecular pathology.” In: *Nature* 474 (7351 May 2011), pp. 380–384.

[81] Matthew R. Ginsberg et al. “Brain Transcriptional and Epigenetic Associations with Autism”. In: *PLoS ONE* 7.9 (Sept. 2012). Ed. by Nicoletta Landsberger, e44736.

[82] Simone Gupta et al. “Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism”. In: *Nature Communications* 5.1 (Dec. 2014).

[83] Xiling Liu et al. “Disruption of an Evolutionarily Novel Synaptic Expression Pattern in Autism”. In: *PLOS Biology* 14.9 (Sept. 2016). Ed. by David Penny, e1002558.

[84] Neelroop N. Parikshak et al. “Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism.” In: *Nature* 540 (7633 Dec. 2016), pp. 423–427.

[85] Wenjie Sun et al. “Histone Acetylome-wide Association Study of Autism Spectrum Disorder.” In: *Cell* 167 (5 Nov. 2016), 1385–1397.e11.

[86] Ye E. Wu et al. “Genome-wide, integrative analysis implicates microRNA dysregulation in autism spectrum disorder.” In: *Nature neuroscience* 19 (11 Nov. 2016), pp. 1463–1476.

[87] Michael J. Gandal et al. “Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder”. In: *Science* 362.6420 (Dec. 2018).

[88] Dmitry Velmeshev et al. “Single-cell genomics identifies cell type–specific molecular changes in autism”. In: *Science* 364.6441 (May 2019), pp. 685–689.

[89] Jillian R. Haney et al. “Broad transcriptomic dysregulation across the cerebral cortex in ASD”. In: *bioRxiv* (Dec. 2020).

[90] Gokul Ramaswami et al. “Integrative genomics identifies a convergent molecular subtype that links epigenomic with transcriptomic differences in autism”. In: *Nature Communications* 11.1 (Sept. 2020).

[91] Krassimira Garbett et al. “Immune transcriptome alterations in the temporal cortex of subjects with autism”. In: *Neurobiology of Disease* 30.3 (June 2008), pp. 303–311.

[92] Zhibin Wang et al. “Combinatorial patterns of histone acetylations and methylations in the human genome”. In: *Nature Genetics* 40.7 (June 2008), pp. 897–903.

[93] M. P. Creyghton et al. “Histone H3K27ac separates active from poised enhancers and predicts developmental state”. In: *Proceedings of the National Academy of Sciences* 107.50 (Nov. 2010), pp. 21931–21936.

[94] Chloe C Y Wong et al. "Genome-wide DNA methylation profiling identifies convergent molecular signatures associated with idiopathic and syndromic autism in post-mortem human brain tissue". In: *Human Molecular Genetics* 28.13 (Mar. 2019), pp. 2201–2211.

[95] Qianhui Yu and Zhisong He. "Comprehensive investigation of temporal and autism-associated cell type composition-dependent and independent gene expression changes in human brains". In: *Scientific Reports* 7.1 (June 2017).

[96] Tony Kouzarides. "Chromatin Modifications and Their Function". In: *Cell* 128.4 (2007), pp. 693–705.

[97] C. David Allis and Thomas Jenuwein. "The molecular hallmarks of epigenetic control". In: *Nature Reviews Genetics* 17.8 (2016), pp. 487–500.

[98] Vicky W. Zhou, Alon Goren, and Bradley E. Bernstein. "Charting histone modifications and the functional organization of mammalian genomes". In: *Nature Reviews Genetics* 12.1 (2010), pp. 7–18.

[99] Rakesh Kumar Sahu, Sakshi Singh, and Raghuvir Singh Tomar. "The mechanisms of action of chromatin remodelers and implications in development and disease". In: *Biochemical Pharmacology* 180 (2020), p. 114200.

[100] Deborah Roidl and Christine Hacker. "Histone methylation during neural development". In: *Cell and Tissue Research* 356.3 (2014), pp. 539–552.

[101] Fabio Mohn et al. "Lineage-Specific Polycomb Targets and De Novo DNA Methylation Define Restriction and Potential of Neuronal Progenitors". In: *Molecular Cell* 30.6 (2008), pp. 755–766.

[102] Nurit Ballas and Gail Mandel. "The many faces of REST oversee epigenetic programming of neuronal genes". In: *Current Opinion in Neurobiology* 15.5 (2005), pp. 500–506.

[103] Jehnna L. Ronan, Wei Wu, and Gerald R. Crabtree. "From neural development to cognition: unexpected roles for chromatin". In: *Nature Reviews Genetics* 14.5 (2013), pp. 347–359.

[104] João D. Pereira et al. "Ezh2, the histone methyltransferase of PRC2, regulates the balance between self-renewal and differentiation in the cerebral cortex". In: *Proceedings of the National Academy of Sciences* 107.36 (2010), pp. 15957–15962.

[105] Yusuke Hirabayashi et al. "Polycomb Limits the Neurogenic Competence of Neural Precursor Cells to Promote Astrogenic Fate Transition". In: *Neuron* 63.5 (2009), pp. 600–613.

[106] William T. Gibson et al. "Mutations in EZH2 Cause Weaver Syndrome". In: *The American Journal of Human Genetics* 90.1 (2012), pp. 110–118.

[107] Hitomi Miyazaki et al. "Ash1l Methylates Lys36 of Histone H3 Independently of Transcriptional Elongation to Counteract Polycomb Silencing". In: *PLoS Genetics* 9.11 (2013). Ed. by Giacomo Cavalli, e1003897.

[108] Seonhye Cheon et al. "Counteracting epigenetic mechanisms regulate the structural development of neuronal circuitry in human neurons". In: *Molecular Psychiatry* 27.4 (2022), pp. 2291–2303.

[109] Lena Ho et al. "esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function". In: *Nature Cell Biology* 13.8 (2011), pp. 903–913.

[110] N. Barker. "The chromatin remodelling factor Brg-1 interacts with beta-catenin to promote target gene activation". In: *The EMBO Journal* 20.17 (2001), pp. 4935–4943.

[111] Julie Lessard et al. "An Essential Switch in Subunit Composition of a Chromatin Remodeling Complex during Neural Development". In: *Neuron* 55.2 (2007), pp. 201–215.

[112] Godwin Sokpor et al. "Chromatin Remodeling BAF (SWI/SNF) Complexes in Neural Development and Disorders". In: *Frontiers in Molecular Neuroscience* 10 (2017).

[113] Anke Hoffmann and Dietmar Spengler. "Chromatin Remodeling Complex NuRD in Neurodevelopment and Neurodevelopmental Disorders". In: *Frontiers in Genetics* 10 (2019).

[114] Susanne Bornelöv et al. "The Nucleosome Remodeling and Deacetylation Complex Modulates Chromatin Structure at Sites of Active Transcription to Fine-Tune Gene Expression". In: *Molecular Cell* 71.1 (2018), 56–72.e4.

[115] Erin Knock et al. "The methyl binding domain 3/nucleosome remodelling and deacetylase complex regulates neural cell fate determination and terminal differentiation in the cerebral cortex". In: *Neural Development* 10.1 (2015).

[116] Hui Yao et al. "CHD7 promotes neural progenitor differentiation in embryonic stem cells via altered chromatin accessibility and nascent gene expression". In: *Scientific Reports* 10.1 (2020).

[117] Ruchi Bajpai et al. "CHD7 cooperates with PBAF to control multipotent neural crest formation". In: *Nature* 463.7283 (2010), pp. 958–962.

[118] Xuan Jiang et al. “The Mutation in Chd7 Causes Misexpression of Bmp4 and Developmental Defects in Telencephalic Midline”. In: *The American Journal of Pathology* 181.2 (2012), pp. 626–641.

[119] Ikuo Sakamoto et al. “A Novel Beta-Catenin-binding Protein Inhibits Beta-Catenin-dependent Tcf Activation and Axis Formation”. In: *Journal of Biological Chemistry* 275.42 (2000), pp. 32871–32878.

[120] Brandi A. Thompson et al. “CHD8 Is an ATP-Dependent Chromatin Remodeling Factor That Regulates β-Catenin Target Genes”. In: *Molecular and Cellular Biology* 28.12 (June 2008), pp. 3894–3904.

[121] Masaaki Nishiyama, Arthur I. Skoultchi, and Keiichi I. Nakayama. “Histone H1 Recruitment by CHD8 Is Essential for Suppression of the Wnt–β-Catenin Signaling Pathway”. In: *Molecular and Cellular Biology* 32.2 (Jan. 2012), pp. 501–512.

[122] Masaaki Nishiyama et al. “CHD8 suppresses p53-mediated apoptosis through histone H1 recruitment during early embryogenesis”. In: *Nature Cell Biology* 11.2 (2009), pp. 172–182.

[123] Omer Durak et al. “Chd8 mediates cortical neurogenesis via transcriptional regulation of cell cycle and Wnt signaling”. en. In: *Nat. Neurosci.* 19.11 (Nov. 2016), pp. 1477–1488.

[124] Ummi Ciptasari and Hans van Bokhoven. “The phenomenal epigenome in neurodevelopmental disorders”. In: *Human Molecular Genetics* 29.R1 (2020), R42–R50.

[125] Taiping Chen and Sharon Y. R. Dent. “Chromatin modifiers and remodellers: regulators of cellular differentiation”. In: *Nature Reviews Genetics* 15.2 (2013), pp. 93–106.

[126] Michele Gabriele et al. “The chromatin basis of neurodevelopmental disorders: Rethinking dysfunction along the molecular and temporal axes”. In: *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 84 (2018), pp. 306–327.

[127] Britt Mossink et al. “The emerging role of chromatin remodelers in neurodevelopmental disorders: a developmental perspective”. In: *Cellular and Molecular Life Sciences* 78.6 (2020), pp. 2517–2563.

[128] Mototsugu Eiraku et al. “Self-Organized Formation of Polarized Cortical Tissues from ESCs and Its Active Manipulation by Extrinsic Signals”. In: *Cell Stem Cell* 3.5 (2008), pp. 519–532.

[129] Kiichi Watanabe et al. “Directed differentiation of telencephalic precursors from embryonic stem cells”. In: *Nature Neuroscience* 8.3 (2005), pp. 288–296.

[130] Kiichi Watanabe et al. “A ROCK inhibitor permits survival of dissociated human embryonic stem cells”. In: *Nature Biotechnology* 25.6 (2007), pp. 681–686.

[131] Madeline A. Lancaster et al. “Cerebral organoids model human brain development and microcephaly”. In: *Nature* 501.7467 (Aug. 2013), pp. 373–379.

[132] Taisuke Kadoshima et al. “Self-organization of axial polarity, inside-out layer pattern, and species-specific progenitor dynamics in human ES cell–derived neocortex”. In: *Proceedings of the National Academy of Sciences* 110.50 (2013), pp. 20284–20289.

[133] Anca M Paşca et al. “Functional cortical neurons and astrocytes from human pluripotent stem cells in 3D culture”. In: *Nature Methods* 12.7 (May 2015), pp. 671–678.

[134] Stuart M Chambers et al. “Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling”. In: *Nature Biotechnology* 27.3 (2009), pp. 275–280.

[135] Keiko Muguruma et al. “Self-Organization of Polarized Cerebellar Tissue in 3D Culture of Human Pluripotent Stem Cells”. In: *Cell Reports* 10.4 (2015), pp. 537–550.

[136] Hideya Sakaguchi et al. “Generation of functional hippocampal neurons from self-organizing human embryonic stem cell-derived dorsomedial telencephalic tissue”. In: *Nature Communications* 6.1 (2015).

[137] Junghyun Jo et al. “Midbrain-like Organoids from Human Pluripotent Stem Cells Contain Functional Dopaminergic and Neuromelanin-Producing Neurons”. In: *Cell Stem Cell* 19.2 (2016), pp. 248–257.

[138] Fikri Birey et al. “Assembly of functionally integrated human forebrain spheroids”. In: *Nature* 545.7652 (Apr. 2017), pp. 54–59.

[139] Yuki Miura et al. “Generation of human striatal organoids and cortico-striatal assembloids from human pluripotent stem cells”. In: *Nature Biotechnology* 38.12 (Dec. 2020), pp. 1421–1430.

[140] Joshua A Bagley et al. “Fused cerebral organoids model interactions between brain regions”. In: *Nature Methods* 14.7 (2017), pp. 743–751.

[141] J. Gray Camp et al. “Human cerebral organoids recapitulate gene expression programs of fetal neocortex development”. In: *Proceedings of the National Academy of Sciences* 112.51 (2015), pp. 15672–15677.

[142] Sabina Kanton et al. “Organoid single-cell genomic atlas uncovers human-specific features of brain development”. In: *Nature* 574.7778 (2019), pp. 418–422.

[143] Jonas S. Fleck et al. “Inferring and perturbing cell fate regulomes in human cerebral organoids”. In: (2021).

[144] Se-Jin Yoon et al. “Reliability of human cortical organoid generation”. In: *Nature Methods* 16.1 (Dec. 2018), pp. 75–78.

[145] Alexandro E. Trevino et al. “Chromatin accessibility dynamics in a model of human forebrain development”. In: *Science* 367.6476 (Jan. 2020).

[146] Xuyu Qian et al. “Brain-Region-Specific Organoids Using Mini-bioreactors for Modeling ZIKV Exposure”. In: *Cell* 165.5 (2016), pp. 1238–1254.

[147] Momoko Watanabe et al. “Self-Organized Cerebral Organoids with Human-Specific Features Predict Effective Drugs to Combat Zika Virus Infection”. In: *Cell Reports* 21.2 (2017), pp. 517–532.

[148] Marina Bershteyn et al. “Human iPSC-Derived Cerebral Organoids Model Cellular Features of Lissencephaly and Reveal Prolonged Mitosis of Outer Radial Glia”. In: *Cell Stem Cell* 20.4 (2017), 435–449.e4.

[149] Vira Iefremova et al. “An Organoid-Based Model of Cortical Development Identifies Non-Cell-Autonomous Defects in Wnt Signaling Contributing to Miller-Dieker Syndrome”. In: *Cell Reports* 19.1 (2017), pp. 50–59.

[150] Fikri Birey et al. “Dissecting the molecular basis of human interneuron migration in forebrain assembloids from Timothy syndrome”. In: *Cell Stem Cell* 29.2 (2022), 248–264.e7.

[151] Jessica Mariani et al. “FOXG1-Dependent Dysregulation of GABA/Glutamate Neuron Differentiation in Autism Spectrum Disorders”. In: *Cell* 162.2 (July 2015), pp. 375–390.

[152] Alexandre Jourdon et al. “ASD modelling in organoids reveals imbalance of excitatory cortical neuron subtypes during early neurogenesis”. In: (2022).

[153] Bruna Paulsen et al. “Autism genes converge on asynchronous development of shared neuron classes”. en. In: *Nature* 602.7896 (Feb. 2022), pp. 268–273.

[154] John D. Blair, Dirk Hockemeyer, and Helen S. Bateup. “Genetically engineered human cortical spheroid models of tuberous sclerosis”. In: *Nature Medicine* 24.10 (2018), pp. 1568–1578.

[155] Oliver L. Eichmüller et al. “Amplification of human interneuron progenitors promotes brain tumors and neurological defects”. In: *Science* 375.6579 (2022).

[156] Silvia Benito-Kwiecinski et al. “An early cell shape transition drives evolutionary expansion of the human forebrain”. In: *Cell* 184.8 (Apr. 2021), 2084–2102.e19.

[157] Dawn X Zhang and Christopher K Glass. “Towards an understanding of cell-specific functions of signal-dependent transcription factors”. In: *Journal of Molecular Endocrinology* 51.3 (2013), T37–T50.

[158] Agustina D’Urso and Jason H. Brickner. “Mechanisms of epigenetic memory”. In: *Trends in Genetics* 30.6 (2014), pp. 230–236.

[159] Yongchao Gou, Tingwei Zhang, and Jian Xu. “Transcription Factors in Craniofacial Development”. In: *Current Topics in Developmental Biology*. Elsevier, 2015, pp. 377–410.

[160] Irfan A. Qureshi and Mark F. Mehler. “Epigenetic mechanisms underlying nervous system diseases”. In: *Neurogenetics, Part I*. Elsevier, 2018, pp. 43–58.

[161] Yusuke Hirabayashi and Yukiko Gotoh. “Epigenetic control of neural precursor cell fate during development.” In: *Nature reviews. Neuroscience* 11 (6 June 2010), pp. 377–388.

[162] Peter V. Nguyen, Ted Abel, and Eric R. Kandel. “Requirement of a Critical Period of Transcription for Induction of a Late Phase of LTP”. In: *Science* 265.5175 (1994), pp. 1104–1107.

[163] Ram Madabhushi and Tae-Kyung Kim. “Emerging themes in neuronal activity-dependent gene expression”. In: *Molecular and Cellular Neuroscience* 87 (2018), pp. 27–34.

[164] Adam F. Johnson, Ha T. Nguyen, and Reiner A. Veitia. “Causes and effects of haploinsufficiency”. In: *Biological Reviews* 94.5 (2019), pp. 1774–1785.

[165] Víctor Faundes et al. “Histone Lysine Methylases and Demethylases in the Landscape of Human Developmental Disorders”. In: *The American Journal of Human Genetics* 102.1 (2018), pp. 175–187.

[166] R.A. Veitia, S. Caburet, and J.A. Birchler. "Mechanisms of Mendelian dominance". In: *Clinical Genetics* 93.3 (2017), pp. 419–428.

[167] Elena Deliu* et al. "Haploinsufficiency of the intellectual disability gene SETD5 disturbs developmental gene expression and cognition". In: *Nature Neuroscience* 21.12 (Nov. 2018), pp. 1717–1727.

[168] Carlo Emanuele Villa et al. "CHD8 haploinsufficiency links autism to transient alterations in excitatory and inhibitory trajectories". In: *Cell Reports* 39.1 (Apr. 2022), p. 110615.

[169] Vincent des Portes. "Intellectual disability". In: *Neurocognitive Development: Disorders and Disabilities*. Elsevier, 2020, pp. 113–126.

[170] Orla McBride et al. "Prevalence estimation of intellectual disability using national administrative and household survey data: The importance of survey question specificity". In: *International Journal of Population Data Science* 6.1 (2021).

[171] Catherine Lord et al. "Autism spectrum disorder". In: *The Lancet* 392.10146 (2018), pp. 508–520.

[172] Hans van Bokhoven. "Genetic and Epigenetic Networks in Intellectual Disabilities". In: *Annual Review of Genetics* 45.1 (2011), pp. 81–104.

[173] Detelina Grozeva et al. "De Novo Loss-of-Function Mutations in SETD5, Encoding a Methyltransferase in a 3p25 Microdeletion Syndrome Critical Region, Cause Intellectual Disability". In: *The American Journal of Human Genetics* 94.4 (Apr. 2014), pp. 618–624.

[174] Alma Kuechler et al. "Loss-of-function variants of SETD5 cause intellectual disability and the core phenotype of microdeletion 3p25.3 syndrome". In: *European Journal of Human Genetics* 23.6 (Aug. 2014), pp. 753–760.

[175] Anna B. Osipovich et al. "Setd5 is essential for mammalian development and the co-transcriptional regulation of histone acetylation." In: *Development (Cambridge, England)* 143 (24 Dec. 2016), pp. 4595–4607. ppublish.

[176] Alex M. Francette, Sarah A. Tripplehorn, and Karen M. Arndt. "The Paf1 Complex: A Keystone of Nuclear Regulation Operating at the Interface of Transcription and Chromatin". In: *Journal of Molecular Biology* 433.14 (2021), p. 166979.

[177] Kristel N. Eigenhuis, Hedda B. Somsen, and Debbie L. C. van den Berg. "Transcription Pause and Escape in Neurodevelopmental Disorders". In: *Frontiers in Neuroscience* 16 (May 2022).

[178] Daniel S. Day et al. "Comprehensive analysis of promoter-proximal RNA polymerase II pausing across mammalian cell types". In: *Genome Biology* 17.1 (2016).

[179] Fei Xavier Chen et al. "PAF1, a Molecular Regulator of Promoter-Proximal Pausing by RNA Polymerase II". In: *Cell* 162.5 (Aug. 2015), pp. 1003–1015.

[180] M. Yu et al. "RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II". In: *Science* 350.6266 (Dec. 2015), pp. 1383–1386.

[181] Lipeng Wu et al. "H2B Ubiquitylation Promotes RNA Pol II Processivity via PAF1 and pTEFb". In: *Molecular Cell* 54.6 (June 2014), pp. 920–931.

[182] Lucy H. Williams et al. "Pausing of RNA Polymerase II Regulates Mammalian Developmental Potential through Control of Signaling Networks". In: *Molecular Cell* 58.2 (Apr. 2015), pp. 311–322.

[183] Jiancheng Liu et al. "Dynamics of RNA Polymerase II Pausing and Bivalent Histone H3 Methylation during Neuronal Differentiation in Brain Development". In: *Cell Reports* 20.6 (Aug. 2017), pp. 1307–1318.

[184] H.-G. Yoon. "Purification and functional characterization of the human N-CoR complex: the roles of HDAC3, TBL1 and TBLR1". In: *The EMBO Journal* 22.6 (2003), pp. 1336–1346.

[185] Matthew J. Emmett and Mitchell A. Lazar. "Integrative regulation of physiology by histone deacetylase 3". In: *Nature Reviews Molecular Cell Biology* 20.2 (2018), pp. 102–115.

[186] Yan Kong, Wenjun Zhou, and Zheng Sun. "Nuclear receptor corepressors in intellectual disability and autism". In: *Molecular Psychiatry* 25.10 (Feb. 2020), pp. 2220–2236.

[187] Matthew J Lyst et al. "Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor". In: *Nature Neuroscience* 16.7 (2013), pp. 898–902.

[188] Annie Vogel-Ciernia and Marcelo A. Wood. "Examining Object Location and Object Recognition Memory in Mice". In: *Current Protocols in Neuroscience* 69.1 (2014).

[189] Toshihiro Endo et al. "Automated test of behavioral flexibility in mice using a behavioral sequencing task in IntelliCage". In: *Behavioural Brain Research* 221.1 (2011), pp. 172–181.

[190] Hideo Hagihara et al. “Dissection of Hippocampal Dentate Gyrus from Adult Mouse”. In: *Journal of Visualized Experiments* 33 (Nov. 2009).

[191] Alexander Dobin et al. “STAR: ultrafast universal RNA-seq aligner”. In: *Bioinformatics* 29.1 (Oct. 2012), pp. 15–21.

[192] Lucia Peixoto et al. “How data analysis affects power, reproducibility and biological insight of RNA-seq studies in complex datasets”. In: *Nucleic Acids Research* 43.16 (July 2015), pp. 7664–7674.

[193] Davide Risso et al. “GC-Content Normalization for RNA-Seq Data”. In: *BMC Bioinformatics* 12.1 (Dec. 2011).

[194] Davide Risso et al. “Normalization of RNA-seq data using factor analysis of control genes or samples”. In: *Nature Biotechnology* 32.9 (Aug. 2014), pp. 896–902.

[195] M. D. Robinson, D. J. McCarthy, and G. K. Smyth. “edgeR: a Bioconductor package for differential expression analysis of digital gene expression data”. In: *Bioinformatics* 26.1 (Nov. 2009), pp. 139–140.

[196] Philipp Mews et al. “Acetyl-CoA synthetase regulates histone acetylation and hippocampal memory”. In: *Nature* 546.7658 (May 2017), pp. 381–386.

[197] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.

[198] Zuguang Gu et al. “circlize implements and enhances circular visualization in R”. In: *Bioinformatics* 30.19 (June 2014), pp. 2811–2812.

[199] S. Falcon and R. Gentleman. “Using GOstats to test gene lists for GO term association”. In: *Bioinformatics* 23.2 (Nov. 2006), pp. 257–258.

[200] Wencke Walter, Fátima Sánchez-Cabo, and Mercedes Ricote. “GOplot: an R package for visually combining expression data with functional analysis: Fig. 1.” In: *Bioinformatics* 31.17 (May 2015), pp. 2912–2914.

[201] Nathan G. Skene and Seth G. N. Grant. “Identification of Vulnerable Cell Types in Major Brain Disorders Using Single Cell Transcriptomes and Expression Weighted Cell Type Enrichment”. In: *Frontiers in Neuroscience* 10 (Jan. 2016).

[202] Amit Zeisel et al. “Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq”. In: *Science* 347.6226 (Mar. 2015), pp. 1138–1142.

[203] Edward Y Chen et al. “Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool”. In: *BMC Bioinformatics* 14.1 (Apr. 2013).

[204] Maxim V. Kuleshov et al. “Enrichr: a comprehensive gene set enrichment analysis web server 2016 update”. In: *Nucleic Acids Research* 44.W1 (May 2016), W90–W97.

[205] Alexi Nott et al. “Histone deacetylase 3 associates with MeCP2 to regulate FOXO and social behavior”. In: *Nature Neuroscience* 19.11 (July 2016), pp. 1497–1505.

[206] Miriam Bibel et al. “Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells”. In: *Nature Protocols* 2.5 (Apr. 2007), pp. 1034–1043.

[207] Ben Langmead and Steven L Salzberg. “Fast gapped-read alignment with Bowtie 2”. In: *Nature Methods* 9.4 (Mar. 2012), pp. 357–359.

[208] S. Anders, P. T. Pyl, and W. Huber. “HTSeq–a Python framework to work with high-throughput sequencing data”. In: *Bioinformatics* 31.2 (Sept. 2014), pp. 166–169.

[209] Michael I Love, Wolfgang Huber, and Simon Anders. “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2”. In: *Genome Biology* 15.12 (Dec. 2014).

[210] K. Strimmer. “fdrtool: a versatile R package for estimating local and tail area-based false discovery rates”. In: *Bioinformatics* 24.12 (Apr. 2008), pp. 1461–1462.

[211] Jeremy Goecks et al. “Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences”. In: *Genome Biology* 11.8 (2010), R86.

[212] Fidel Ramírez et al. “deepTools: a flexible platform for exploring deep-sequencing data”. In: *Nucleic Acids Research* 42.W1 (May 2014), W187–W191.

[213] Yong Zhang et al. “Model-based Analysis of ChIP-Seq (MACS)”. In: *Genome Biology* 9.9 (Sept. 2008).

[214] Aaron R. Quinlan. “BEDTools: The Swiss-Army Tool for Genome Feature Analysis”. In: *Current Protocols in Bioinformatics* 47.1 (Sept. 2014).

[215] Yin Shen et al. “A map of the cis-regulatory sequences in the mouse genome”. In: *Nature* 488.7409 (July 2012), pp. 116–120.

[216] Caryn S. Ross-Innes et al. “Differential oestrogen receptor binding is associated with clinical outcome in breast cancer”. In: *Nature* 481.7381 (Jan. 2012), pp. 389–393.

[217] J. D. Dignam, R. M. Lebovitz, and R. G. Roeder. “Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei”. In: *Nucleic Acids Research* 11.5 (1983), pp. 1475–1489.

[218] Zhongping Liao et al. “IsoQuant: A Software Tool for Stable Isotope Labeling by Amino Acids in Cell Culture-Based Mass Spectrometry Quantitation”. In: *Analytical Chemistry* 84.10 (2012), pp. 4535–4543.

[219] Stefka Tyanova, Tikira Temu, and Juergen Cox. “The MaxQuant computational platform for mass spectrometry-based shotgun proteomics”. In: *Nature Protocols* 11.12 (2016), pp. 2301–2319.

[220] W. Huber et al. “Variance stabilization applied to microarray data calibration and to the quantification of differential expression”. In: *Bioinformatics* 18.Suppl 1 (2002), S96–S104.

[221] Matthew E. Ritchie et al. “limma powers differential expression analyses for RNA-sequencing and microarray studies”. In: *Nucleic Acids Research* 43.7 (2015), e47–e47.

[222] Paul Shannon et al. “Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks”. In: *Genome Research* 13.11 (2003), pp. 2498–2504.

[223] L. J. Jensen et al. “STRING 8–a global view on proteins and their functional interactions in 630 organisms”. In: *Nucleic Acids Research* 37.Database (2009), pp. D412–D416.

[224] Dwight R. Cordero et al. “Cranial neural crest cells on the move: Their roles in craniofacial development”. In: *American Journal of Medical Genetics Part A* 155.2 (Dec. 2010), pp. 270–279.

[225] Laura L. Baxter et al. “Spotlight on Spotted Mice: A Review of White Spotting Mouse Mutants and Associated Human Pigmentation Disorders”. In: *Pigment Cell Research* 17.3 (June 2004), pp. 215–224.

[226] Alessandro Sessa et al. “SETD5 Regulates Chromatin Methylation State and Preserves Global Transcriptional Fidelity during Brain Development and Neuronal Wiring”. In: *Neuron* 104.2 (Oct. 2019), 271–289.e13.

[227] Jaehoon Kim, Mohamed Guermah, and Robert G. Roeder. “The Human PAF1 Complex Acts in Chromatin Transcription Elongation Both Independently and Cooperatively with SII/TFIIS”. In: *Cell* 140.4 (Feb. 2010), pp. 491–503.

[228] Yaelim Lee, Daechan Park, and Vishwanath R. Iyer. “The ATP-dependent chromatin remodeler Chd1 is recruited by transcription elongation factors and maintains H3K4me3/H3K36me3 domains at actively transcribed and spliced genes”. In: *Nucleic Acids Research* 45.14 (July 2017), pp. 8646–8646.

[229] Adrian Brunkhorst et al. “A specific role for the TFIID subunit TAF4 and RanBPM in neural progenitor differentiation”. In: *Molecular and Cellular Neuroscience* 29.2 (June 2005), pp. 250–258.

[230] P. Karagianni and J. Wong. “HDAC3: taking the SMRT-N-CoRrect road to repression”. In: *Oncogene* 26.37 (Aug. 2007), pp. 5439–5449.

[231] Valentina Perissi et al. “TBL1 and TBLR1 Phosphorylation on Regulated Gene Promoters Overcomes Dual CtBP and NCoR/SMRT Transcriptional Repression Checkpoints”. In: *Molecular Cell* 29.6 (Mar. 2008), pp. 755–766.

[232] Jiong Li and Cun-Yu Wang. “TBL1-TBLR1 and beta-catenin recruit each other to Wnt target-gene promoter for transcription activation and oncogenesis”. In: *Nature Cell Biology* 10.2 (Jan. 2008), pp. 160–169.

[233] Christian Mosimann, George Hausmann, and Konrad Basler. “Parafibromin/Hyrax Activates Wnt/Wg Target Gene Transcription by Direct Association with beta-catenin/Armadillo”. In: *Cell* 125.2 (Apr. 2006), pp. 327–341.

[234] Ippei Kikuchi et al. “Dephosphorylated parafibromin is a transcriptional coactivator of the Wnt/Hedgehog/Notch pathways”. In: *Nature Communications* 7.1 (Sept. 2016).

[235] Sheryl S. Moy et al. “Development of a mouse test for repetitive, restricted behaviors: Relevance to autism”. In: *Behavioural Brain Research* 188.1 (Mar. 2008), pp. 178–194.

[236] S. C. McQuown et al. “HDAC3 Is a Critical Negative Regulator of Long-Term Memory Formation”. In: *Journal of Neuroscience* 31.2 (Jan. 2011), pp. 764–774.

[237] Kim Nuytens et al. “Haploinsufficiency of the autism candidate gene Neurobeachin induces autism-like behaviors and affects cellular and molecular processes of synaptic plasticity in mice”. In: *Neurobiology of Disease* 51 (Mar. 2013), pp. 144–151.

[238] Fabio Antonio Vigil et al. “Prevention of long-term memory loss after retrieval by an endogenous CaMKII inhibitor”. In: *Scientific Reports* 7.1 (June 2017).

[239] Juan M. Alarcon et al. “Selective Modulation of Some Forms of Schaffer Collateral-CA1 Synaptic Plasticity in Mice With a Disruption of the *CPEB-1* Gene”. In: *Learning & Memory* 11.3 (May 2004), pp. 318–327.

[240] C. J. M. Bontekoe. “Knockout mouse model for Fxr2: a model for mental retardation”. In: *Human Molecular Genetics* 11.5 (Mar. 2002), pp. 487–498.

[241] Jing Zhang et al. “Altered Hippocampal Synaptic Plasticity in the Fmr1 Gene Family Knockout Mouse Models”. In: *Journal of Neurophysiology* 101.5 (May 2009), pp. 2572–2580.

[242] A. Y. Hung et al. “Smaller Dendritic Spines, Weaker Synaptic Transmission, but Enhanced Spatial Learning in Mice Lacking Shank1”. In: *Journal of Neuroscience* 28.7 (Feb. 2008), pp. 1697–1708.

[243] Seho Kim et al. “NGL family PSD-95 interacting adhesion molecules regulate excitatory synapse formation”. In: *Nature Neuroscience* 9.10 (Sept. 2006), pp. 1294–1301.

[244] Annalisa Alfieri et al. “Synaptic Interactome Mining Reveals p140Cap as a New Hub for PSD Proteins Involved in Psychiatric and Neurological Disorders”. In: *Frontiers in Molecular Neuroscience* 10 (June 2017).

[245] Fadi F. Hamdan et al. “Mutations in SYNGAP1 in Autosomal Nonsyndromic Mental Retardation”. In: *New England Journal of Medicine* 360.6 (Feb. 2009), pp. 599–605.

[246] Michael Litt, Yi Qiu, and Suming Huang. “Histone arginine methylations: their roles in chromatin dynamics and transcriptional regulation”. In: *Bioscience Reports* 29.2 (Feb. 2009), pp. 131–141.

[247] Xingguo Li et al. “H4R3 methylation facilitates beta-globin transcription by regulating histone acetyltransferase binding and H3 acetylation”. In: *Blood* 115.10 (Mar. 2010), pp. 2028–2037.

[248] Flavia Antonucci et al. “SNAP-25, a Known Presynaptic Protein with Emerging Postsynaptic Functions”. In: *Frontiers in Synaptic Neuroscience* 8 (Mar. 2016).

[249] Milena Crippa et al. “SETD5 Gene Haploinsufficiency in Three Patients With Suspected KBG Syndrome”. In: *Frontiers in Neurology* 11 (July 2020).

[250] Yusuke Marikawa. “Wnt/beta-catenin signaling and body plan formation in mouse embryos”. In: *Seminars in Cell & Developmental Biology* 17.2 (Apr. 2006), pp. 175–184.

[251] Kimberly A. Mulligan and Benjamin N.R. Cheyette. “Neurodevelopmental Perspectives on Wnt Signaling in Psychiatry”. In: *Molecular Neuropsychiatry* 2.4 (2016), pp. 219–246.

[252] Stefan Rudloff and Rolf Kemler. “Differential requirements for β-catenin during mouse development”. In: *Development* 139.20 (Oct. 2012), pp. 3711–3721.

[253] Cheol-Hee Kim et al. “Repressor activity of Headless/Tcf3 is essential for vertebrate head formation”. In: *Nature* 407.6806 (Oct. 2000), pp. 913–916.

[254] Bing Xiao, Jonathan R. Wilson, and Steven J. Gamblin. “SET domains and histone methylation”. In: *Current Opinion in Structural Biology* 13.6 (Dec. 2003), pp. 699–705.

[255] W.W.M. Pim Pijnappel et al. “The S. cerevisiae SET3 complex includes two histone deacetylases, Hos2 and Hst1, and is a meiotic-specific repressor of the sporulation gene program”. In: *Genes & Development* 15.22 (Nov. 2001), pp. 2991–3004.

[256] Hector Rincon-Arano et al. “UpSET Recruits HDAC Complexes and Restricts Chromatin Accessibility and Acetylation at Promoter Regions”. In: *Cell* 151.6 (Dec. 2012), pp. 1214–1228.

[257] Sarah Mas-y Mas et al. “The Human Mixed Lineage Leukemia 5 (MLL5), a Sequentially and Structurally Divergent SET Domain-Containing Protein with No Intrinsic Catalytic Activity”. In: *PLOS ONE* 11.11 (Nov. 2016). Ed. by Albert Jeltsch, e0165139.

[258] Zhentian Wang et al. “SETD5-Coordinated Chromatin Reprogramming Regulates Adaptive Resistance to Targeted Pancreatic Cancer Therapy”. In: *Cancer Cell* 37.6 (June 2020), 834–849.e13.

[259] Hyo-Kyoung Choi et al. “Reversible SUMOylation of TBL1-TBLR1 Regulates β-Catenin-Mediated Wnt Signaling”. In: *Molecular Cell* 43.2 (July 2011), pp. 203–216.

[260] Akira Nishi et al. “De novo non-synonymous TBL1XR1 mutation alters Wnt signaling activity”. In: *Scientific Reports* 7.1 (June 2017).

[261] Rebekah Tillotson et al. “Radically truncated MeCP2 rescues Rett syndrome-like neurological defects”. In: *Nature* 550.7676 (Oct. 2017), pp. 398–401.

[262] Marjolein H Willemsen et al. “Identification of ANKRD11 and ZNF778 as candidate genes for autism and variable cognitive impairment in the novel 16q24.3 microdeletion syndrome”. In: *European Journal of Human Genetics* 18.4 (Nov. 2009), pp. 429–435.

[263] Genay O Pilarowski et al. “Missense variants in the chromatin remodeler CHD1 are associated with neurodevelopmental disability”. In: *Journal of Medical Genetics* 55.8 (Sept. 2017), pp. 561–566.

[264] S. Branden van Oss, Christine E. Cucinotta, and Karen M. Arndt. “Emerging Insights into the Roles of the Paf1 Complex in Gene Regulation”. In: *Trends in Biochemical Sciences* 42.10 (Oct. 2017), pp. 788–798.

[265] Karen Adelman and John T. Lis. “Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans”. In: *Nature Reviews Genetics* 13.10 (Sept. 2012), pp. 720–731.

[266] Abderhman Abuhashem, Vidur Garg, and Anna-Katerina Hadjantonakis. “RNA polymerase II pausing in development: orchestrating transcription”. In: *Open Biology* 12.1 (Jan. 2022).

[267] Yoshihiro Matsumura et al. “Spatiotemporal dynamics of SETD5-containing NCoR–HDAC3 complex determines enhancer activation for adipogenesis”. In: *Nature Communications* 12.1 (Dec. 2021).

[268] Jonathan M. Levenson et al. “Regulation of Histone Acetylation during Memory Formation in the Hippocampus”. In: *Journal of Biological Chemistry* 279.39 (Sept. 2004), pp. 40545–40559.

[269] K. Matthew Lattal, Ruth M. Barrett, and Marcelo A. Wood. “Systemic or intrahippocampal delivery of histone deacetylase inhibitors facilitates fear extinction.” In: *Behavioral Neuroscience* 121.5 (2007), pp. 1125–1131.

[270] C. G. Vecsey et al. “Histone Deacetylase Inhibitors Enhance Memory and Synaptic Plasticity via CREB: CBP-Dependent Transcriptional Activation”. In: *Journal of Neuroscience* 27.23 (June 2007), pp. 6128–6140.

[271] Ruth M. Barrett and Marcelo A. Wood. “Beyond transcription factors: The role of chromatin modifying enzymes in regulating transcription required for memory”. In: *Learning & Memory* 15.7 (June 2008), pp. 460–467.

[272] Timothy W. Bredy and Mark Barad. “The histone deacetylase inhibitor valproic acid enhances acquisition, extinction, and reconsolidation of conditioned fear”. In: *Learning & Memory* 15.1 (Jan. 2008), pp. 39–45.

[273] Susan C. McQuown and Marcelo A. Wood. “HDAC3 and the molecular brake pad hypothesis”. In: *Neurobiology of Learning and Memory* 96.1 (July 2011), pp. 27–34.

[274] G. A. Rogge et al. “HDAC3 Is a Negative Regulator of Cocaine-Context-Associated Memory Formation”. In: *Journal of Neuroscience* 33.15 (Apr. 2013), pp. 6623–6632.

[275] Janine L Kwapis et al. “Context and Auditory Fear are Differentially Regulated by HDAC3 Activity in the Lateral and Basal Subnuclei of the Amygdala”. In: *Neuropsychopharmacology* 42.6 (Dec. 2016), pp. 1284–1294.

[276] Yasaman Alaghband et al. “Distinct roles for the deacetylase domain of HDAC3 in the hippocampus and medial prefrontal cortex in the formation and extinction of memory”. In: *Neurobiology of Learning and Memory* 145 (Nov. 2017), pp. 94–104.

[277] Mallory E. Bowers et al. “The Class I HDAC inhibitor RGFP963 enhances consolidation of cued fear extinction”. In: *Learning & Memory* 22.4 (Mar. 2015), pp. 225–231.

[278] Vivian Budnik and Patricia C Salinas. “Wnt signaling during synaptic development and plasticity”. In: *Current Opinion in Neurobiology* 21.1 (Feb. 2011), pp. 151–159.

[279] Carolina A. Oliva, Jessica Y. Vargas, and Nibaldo C. Inestrosa. “Wnts in adult brain: from synaptic plasticity to cognitive deficiencies”. In: *Frontiers in Cellular Neuroscience* 7 (2013).

[280] K. P. Harris et al. “Shank Modulates Postsynaptic Wnt Signaling to Regulate Synaptic Development”. In: *Journal of Neuroscience* 36.21 (May 2016), pp. 5820–5832.

[281] Xinjuan Wang et al. “Phosphorylated SNAP25 in the CA1 regulates morphine-associated contextual memory retrieval via increasing GluN2B-NMDAR surface localization”. In: *Addiction Biology* 23.5 (Sept. 2017), pp. 1067–1078.

[282] Katisha R. Gopaul et al. “Developmental Time Course of SNAP-25 Isoforms Regulate Hippocampal Long-Term Synaptic Plasticity and Hippocampus-Dependent Learning”. In: *International Journal of Molecular Sciences* 21.4 (Feb. 2020), p. 1448.

[283] Zhilin Wang et al. “Retrieval-Driven Hippocampal NPTX2 Plasticity Facilitates the Extinction of Cocaine-Associated Context Memory”. In: *Biological Psychiatry* 87.11 (June 2020), pp. 979–991.

[284] Yuta Katayama et al. “CHD8 haploinsufficiency results in autistic-like phenotypes in mice”. en. In: *Nature* 537.7622 (Sept. 2016), pp. 675–679.

[285] Randall J Platt et al. “Chd8 mutation leads to autistic-like behaviors and impaired striatal circuits”. en. In: *Cell Rep.* 19.2 (Apr. 2017), pp. 335–350.

[286] Andrea L Gompers et al. “Germline Chd8 haploinsufficiency alters brain development in mouse”. en. In: *Nat. Neurosci.* 20.8 (Aug. 2017), pp. 1062–1073.

[287] Philipp Suetterlin et al. “Altered neocortical gene expression, brain overgrowth and functional over-connectivity in Chd8 haploinsufficient mice”. en. In: *Cereb. Cortex* 28.6 (June 2018), pp. 2192–2206.

[288] Hwajin Jung et al. “Sexually dimorphic behavior, neuronal activity, and gene expression in Chd8-mutant mice”. en. In: *Nat. Neurosci.* 21.9 (Sept. 2018), pp. 1218–1228.

[289] Jessica A Jiménez et al. “Chd8 haploinsufficiency impairs early brain development and protein homeostasis later in life”. en. In: *Mol. Autism* 11.1 (Oct. 2020), p. 74.

[290] Samuel W Hulbert et al. “A novel Chd8 mutant mouse displays altered ultrasonic vocalizations and enhanced motor coordination”. en. In: *Autism Res.* 13.10 (Oct. 2020), pp. 1685–1697.

[291] Atsuki Kawamura et al. “Oligodendrocyte dysfunction due to Chd8 mutation gives rise to behavioral deficits in mice”. en. In: *Hum. Mol. Genet.* 29.8 (May 2020), pp. 1274–1291.

[292] Shaun Hurley et al. “Distinct, dosage-sensitive requirements for the autism-associated factor CHD8 during cortical development”. en. In: *Mol. Autism* 12.1 (Feb. 2021), p. 16.

[293] Orly Weissberg and Evan Elliott. “The mechanisms of CHD8 in neurodevelopment and autism spectrum disorders”. en. In: *Genes (Basel)* 12.8 (July 2021), p. 1133.

[294] Jennifer S. Beighley et al. “Clinical Phenotypes of Carriers of Mutations in CHD8 or Its Conserved Target Genes”. In: *Biological Psychiatry* 87.2 (Jan. 2020), pp. 123–131.

[295] A. Flaus. “Identification of multiple distinct Snf2 subfamilies with conserved structural motifs”. In: *Nucleic Acids Research* 34.10 (2006), pp. 2887–2905.

[296] Andrej Alendar and Anton Berns. “Sentinels of chromatin: chromodomain helicase DNA-binding proteins in development and disease”. In: *Genes and Development* 35.21-22 (2021), pp. 1403–1430.

[297] Ping Wang et al. “CRISPR/Cas9-mediated heterozygous knockout of the autism gene CHD8 and characterization of its transcriptional networks in cerebral organoids derived from iPS cells”. en. In: *Mol. Autism* 8 (Mar. 2017), p. 11.

[298] A Ayanna Wade et al. “Common CHD8 genomic targets contrast with model-specific transcriptional impacts of CHD8 haploinsufficiency”. en. In: *Front. Mol. Neurosci.* 11 (2018), p. 481.

[299] Madeline A Lancaster and Juergen A Knoblich. “Generation of cerebral organoids from human pluripotent stem cells”. In: *Nature Protocols* 9.10 (Sept. 2014), pp. 2329–2340.

[300] Timothy S. Catlett et al. “RHOA signaling defects result in impaired axon guidance in iPSC-derived neurons from patients with tuberous sclerosis complex”. In: *Nature Communications* 12.1 (2021).

[301] Job O. de Jong et al. “Cortical overgrowth in a preclinical forebrain organoid model of CNTNAP2-associated autism spectrum disorder”. In: *Nature Communications* 12.1 (2021).

[302] Jorge Urresti et al. “Cortical organoids model early brain development disrupted by 16p11.2 copy number variants in autism”. In: *Molecular Psychiatry* 26.12 (2021), pp. 7560–7580.

[303] Michelle L. Wegscheid et al. “Patient-derived iPSC-cerebral organoid modeling of the 17q11.2 microdeletion syndrome establishes CRLF3 as a critical regulator of neurogenesis”. In: *Cell Reports* 36.1 (2021), p. 109315.

[304] Ping Wang et al. “CRISPR/Cas9-mediated heterozygous knockout of the autism gene CHD8 and characterization of its transcriptional networks in neurodevelopment”. en. In: *Mol. Autism* 6.1 (Oct. 2015), p. 55.

[305] Nikolas Barkas et al. “Joint analysis of heterogeneous single-cell RNA-seq dataset collections”. In: *Nature Methods* 16.8 (July 2019), pp. 695–698.

[306] Mohammad Lotfollahi, F. Alexander Wolf, and Fabian J. Theis. “scGen predicts single-cell perturbation responses”. In: *Nature Methods* 16.8 (July 2019), pp. 715–721.

[307] F. Alexander Wolf, Philipp Angerer, and Fabian J. Theis. “SCANPY: large-scale single-cell gene expression data analysis”. In: *Genome Biology* 19.1 (Feb. 2018).

[308] V. A. Traag, L. Waltman, and N. J. van Eck. “From Louvain to Leiden: guaranteeing well-connected communities”. In: *Scientific Reports* 9.1 (Mar. 2019).

[309] Alejandro López-Tobón et al. "Human Cortical Organoids Expose a Differential Function of GSK3 on Cortical Neurogenesis". In: *Stem Cell Reports* 13.5 (Nov. 2019), pp. 847–861.

[310] Tomasz J. Nowakowski et al. "Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex". In: *Science* 358.6368 (Dec. 2017), pp. 1318–1323.

[311] Alex A. Pollen et al. "Establishing Cerebral Organoids as Models of Human-Specific Brain Evolution". In: *Cell* 176.4 (Feb. 2019), 743–756.e17.

[312] Manu Setty et al. "Characterization of cell fate probabilities in single-cell data with Palantir". In: *Nature Biotechnology* 37.4 (Mar. 2019), pp. 451–460.

[313] Huidong Chen et al. "Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM". In: *Nature Communications* 10.1 (Apr. 2019).

[314] Etienne Becht et al. "Dimensionality reduction for visualizing single-cell data using UMAP". In: *Nature Biotechnology* 37.1 (Dec. 2018), pp. 38–44.

[315] Adrian Alexa, Jorg Rahnenfuhrer, et al. "topGO: enrichment analysis for gene ontology". In: *R package version* 2.0 (2010), p. 2010.

[316] The ENCODE Project Consortium. "An integrated encyclopedia of DNA elements in the human genome". In: *Nature* 489.7414 (Sept. 2012), pp. 57–74.

[317] Sven Heinz et al. "Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities". In: *Molecular Cell* 38.4 (May 2010), pp. 576–589.

[318] Lynette Lim et al. "Development and Functional Diversification of Cortical Interneurons". In: *Neuron* 100.2 (Oct. 2018), pp. 294–313.

[319] Caitlin E. Monaghan et al. "REST corepressors RCOR1 and RCOR2 and the repressor INSM1 regulate the proliferation-differentiation balance in the developing brain". In: *Proceedings of the National Academy of Sciences* 114.3 (Jan. 2017).

[320] Michele Gabriele et al. "YY1 Haploinsufficiency Causes an Intellectual Disability Syndrome Featuring Transcriptional and Chromatin Dysfunction". In: *The American Journal of Human Genetics* 100.6 (June 2017), pp. 907–925.

[321] Sanna Gudmundsson et al. "TAF1, associated with intellectual disability in humans, is essential for embryogenesis and regulates neurodevelopmental processes in zebrafish". In: *Scientific Reports* 9.1 (July 2019).

[322] Jinsong Zhang et al. "The N-CoR-HDAC3 Nuclear Receptor Corepressor Complex Inhibits the JNK Pathway through the Integral Subunit GPS2". In: *Molecular Cell* 9.3 (Mar. 2002), pp. 611–623.

[323] Pleuntje J. van der Sluijs et al. "The ARID1B spectrum in 143 patients: from nonsyndromic intellectual disability to Coffin–Siris syndrome". In: *Genetics in Medicine* 21.6 (June 2019), pp. 1295–1307.

[324] Pierre-Luc Germain and Giuseppe Testa. "Taming Human Genetic Variability: Transcriptomic Meta-Analysis Guides the Experimental Design and Interpretation of iPSC-Based Disease Modeling". In: *Stem Cell Reports* 8.6 (June 2017), pp. 1784–1796.

[325] Benjamin J. Manning and Timur Yusufzai. "The ATP-dependent chromatin remodeling enzymes CHD6, CHD7, and CHD8 exhibit distinct nucleosome binding and remodeling activities". In: *Journal of Biological Chemistry* 292.28 (July 2017), pp. 11927–11936.

[326] Yu An et al. "De novo variants in the Helicase-C domain of CHD8 are associated with severe phenotypes including autism, language disability and overgrowth". In: *Human Genetics* 139.4 (Jan. 2020), pp. 499–512.

[327] Philip J. Ostrowski et al. "The *CHD8* overgrowth syndrome: A detailed evaluation of an emerging overgrowth phenotype in 27 patients". In: *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* 181.4 (Nov. 2019), pp. 557–564.

[328] Sofia Douzgou et al. "The clinical presentation caused by truncating *CHD8* variants". In: *Clinical Genetics* 96.1 (May 2019), pp. 72–84.

[329] M C Marchetto et al. "Altered proliferation and networks in neural cells derived from idiopathic autistic individuals". In: *Molecular Psychiatry* 22.6 (July 2016), pp. 820–835.

[330] Ekaterina Epifanova et al. "Role of Zeb2/Sip1 in neuronal development". In: *Brain Research* 1705 (Feb. 2019), pp. 24–31.

[331] M. Yao et al. “Identification of a Smad4/YY1-Recognizedand BMP2-Responsive Transcriptional Regulatory Module in the Promoter of Mouse GABA Transporter Subtype I (Gat1) Gene”. In: *Journal of Neuroscience* 30.11 (Mar. 2010), pp. 4062–4071.

[332] Samuel Thudium et al. “Identification of a transcriptional signature found in multiple models of ASD and related disorders”. In: *bioRxiv* (2022).

[333] Matthew A. Lalli et al. “High-throughput single-cell functional elucidation of neurodevelopmental disease–associated genes reveals convergent mechanisms altering neuronal differentiation”. In: *Genome Research* 30.9 (2020), pp. 1317–1331.

[334] Xin Jin et al. “In vivo Perturb-Seq reveals neuronal and glial abnormalities associated with autism risk genes”. In: *Science* 370.6520 (2020).

[335] Helen Rankin Willsey et al. “Parallel in vivo analysis of large-effect autism genes implicates cortical neurogenesis and estrogen in risk and resilience”. In: *Neuron* 109.5 (2021), 788–804.e8.

[336] Eric Deneault et al. “Complete Disruption of Autism-Susceptibility Genes by Gene Editing Predominantly Reduces Functional Connectivity of Isogenic Human Neurons”. In: *Stem Cell Reports* 11.5 (2018), pp. 1211–1225.

[337] Gustav Y. Cederquist et al. “A Multiplex Human Pluripotent Stem Cell Platform Defines Molecular and Functional Subclasses of Autism-Related Genes”. In: *Cell Stem Cell* 27.1 (2020), 35–49.e6.

[338] Eric J. Wagner and Phillip B. Carpenter. “Understanding the language of Lys36 methylation at histone H3”. In: *Nature Reviews Molecular Cell Biology* 13.2 (2012), pp. 115–126.

[339] Mohammad S. Eram et al. “Kinetic characterization of human histone H3 lysine 36 methyltransferases, ASH1L and SETD2”. In: *Biochimica et Biophysica Acta (BBA) - General Subjects* 1850.9 (2015), pp. 1842–1848.

[340] Tao Zhu et al. “Histone methyltransferase Ash1L mediates activity-dependent repression of neurexin-1α”. In: *Scientific Reports* 6.1 (2016).

[341] Hitomi Miyazaki et al. “Ash1l Methylates Lys36 of Histone H3 Independently of Transcriptional Elongation to Counteract Polycomb Silencing”. In: *PLoS Genetics* 9.11 (2013). Ed. by Giacomo Cavalli, e1003897.

[342] Yuen Gao, Mohammad B. Aljazi, and Jin He. “Neural Hyperactivity Is a Core Pathophysiological Change Induced by Deletion of a High Autism Risk Gene Ash1L in the Mouse Brain”. In: *Frontiers in Behavioral Neuroscience* 16 (2022).

[343] Yuze Yan et al. “ASH1L haploinsufficiency results in autistic-like phenotypes in mice and links Eph receptor gene to autism spectrum disorder”. In: *Neuron* 110.7 (2022), 1156–1172.e9.

[344] Karl Agger et al. “UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development”. In: *Nature* 449.7163 (2007), pp. 731–734.

[345] Yongli Shan et al. “JMJD3 and UTX determine fidelity and lineage specification of human neural progenitor cells”. In: *Nature Communications* 11.1 (2020).

[346] Yajing Meng et al. “Depletion of Demethylase KDM6 Enhances Early Neuroectoderm Commitment of Human PSCs”. In: *Frontiers in Cell and Developmental Biology* 9 (2021).

[347] Gunnar Schotta et al. “A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse”. In: *Genes and Development* 22.15 (2008), pp. 2048–2061.

[348] Sara Weirich, Srikanth Kudithipudi, and Albert Jeltsch. “Specificity of the SUV4–20H1 and SUV4–20H2 protein lysine methyltransferases and methylation of novel substrates”. In: *Journal of Molecular Biology* 428.11 (2016), pp. 2344–2358.

[349] Dario Nicetto et al. “Suv4-20h Histone Methyltransferases Promote Neuroectodermal Differentiation by Silencing the Pluripotency-Associated Oct-25 Gene”. In: *PLoS Genetics* 9.1 (2013). Ed. by Anne C. Ferguson-Smith, e1003188.

[350] Guokai Chen. “Splitting hESC/hiPSC lines with EDTA in feeder-free conditions”. In: *StemBook* (2014).

[351] F Ann Ran et al. “Genome engineering using the CRISPR-Cas9 system”. In: *Nature Protocols* 8.11 (Oct. 2013), pp. 2281–2308.

[352] Jean-Paul Concordet and Maximilian Haeussler. “CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens”. In: *Nucleic Acids Research* 46.W1 (May 2018), W242–W245.

[353] Maximilian Haeussler et al. "Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR". In: *Genome Biology* 17.1 (July 2016).

[354] Marcel Martin. "Cutadapt removes adapter sequences from high-throughput sequencing reads". In: *EMBnet.journal* 17.1 (May 2011), p. 10.

[355] Kevin Blighe and Aaron Lun. *PCAtools: PCAtools: Everything Principal Components Analysis*. R package version 2.6.0. 2021.

[356] Zuguang Gu, Roland Eils, and Matthias Schlesner. "Complex heatmaps reveal patterns and correlations in multidimensional genomic data". In: *Bioinformatics* 32.18 (2016), pp. 2847–2849.

[357] Steffen Durinck et al. "Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt". In: *Nature Protocols* 4 (2009), pp. 1184–1191.

[358] Meichen Dong et al. "SCDC: bulk gene expression deconvolution by multiple single-cell RNA sequencing references". In: *Briefings in Bioinformatics* 22.1 (Jan. 2020), pp. 416–427.

[359] Yuhan Hao et al. "Integrated analysis of multimodal single-cell data". In: *Cell* 184.13 (June 2021), 3573–3587.e29.

[360] Steven A. Sloan et al. "Generation and assembly of human brain region-specific three-dimensional cultures". In: *Nature Protocols* 13.9 (Sept. 2018), pp. 2062–2085.

[361] Jan Stenman, Håkan Toresson, and Kenneth Campbell. "Identification of Two Distinct Progenitor Populations in the Lateral Ganglionic Eminence: Implications for Striatal and Olfactory Bulb Neurogenesis". In: *The Journal of Neuroscience* 23.1 (Jan. 2003), pp. 167–174.

[362] Faez Siddiqi et al. "Islet1 Precursors Contribute to Mature Interneuron Subtypes in Mouse Neocortex". In: *Cerebral Cortex* 31.11 (July 2021), pp. 5206–5224.

[363] Kuan-Ming Lu et al. "Dual role for *Islet-1* in promoting striatonigral and repressing striatopallidal genetic programs to specify striatonigral cell identity". In: *Proceedings of the National Academy of Sciences* 111.1 (Dec. 2013).

[364] R. R. Waclaw et al. "Foxo1 is a downstream effector of Isl1 in direct pathway striatal projection neuron development within the embryonic mouse telencephalon". In: *Molecular and Cellular Neuroscience* 80 (Apr. 2017), pp. 44–51.

[365] Alfonso Lavado et al. "YAP/TAZ maintain the proliferative capacity and structural organization of radial glial cells during brain development". In: *Developmental Biology* 480 (Dec. 2021), pp. 39–49.

[366] Simon Maksour, Lezanne Ooi, and Mirella Dottori. "More than a Corepressor: The Role of CoREST Proteins in Neurodevelopment". In: *eneuro* 7.2 (Feb. 2020), ENEURO.0337–19.2020.

[367] Kevin C.J. Nixon et al. "A Syndromic Neurodevelopmental Disorder Caused by Mutations in SMARCD1, a Core SWI/SNF Subunit Needed for Context-Dependent Neuronal Gene Regulation in Flies". In: *The American Journal of Human Genetics* 104.4 (Apr. 2019), pp. 596–610.

[368] Adi Alajem et al. "Differential Association of Chromatin Proteins Identifies BAF60a/SMARCD1 as a Regulator of Embryonic Stem Cell Differentiation". In: *Cell Reports* 10.12 (Mar. 2015), pp. 2019–2031.

[369] Desssislava K Dimova and Nicholas J Dyson. "The E2F transcriptional network: old acquaintances with new faces". In: *Oncogene* 24.17 (Apr. 2005), pp. 2810–2826.

[370] Ranjie Xu et al. "OLIG2 Drives Abnormal Neurodevelopmental Phenotypes in Human iPSC-Based Organoid and Chimeric Mouse Models of Down Syndrome". In: *Cell Stem Cell* 24.6 (June 2019), 908–926.e8.

[371] Chao Chen et al. "The transcription factor POU3F2 regulates a gene coexpression network in brain tissue from patients with psychiatric disorders". In: *Science Translational Medicine* 10.472 (Dec. 2018).

[372] Bin Chen et al. "The *Fezf2-Ctip2* genetic pathway regulates the fate choice of subcortical projection neurons in the developing cerebral cortex". In: *Proceedings of the National Academy of Sciences* 105.32 (Aug. 2008), pp. 11382–11387.

[373] K. Yun, S. Potter, and J.L. Rubenstein. "Gsh2 and Pax6 play complementary roles in dorsoventral patterning of the mammalian telencephalon". In: *Development* 128.2 (Jan. 2001), pp. 193–205.

[374] Zihao Su et al. "*Dlx1/2*-dependent expression of *Meis2* promotes neuronal fate determination in the mammalian striatum". In: *Development* 149.4 (Feb. 2022).

[375] Zhejun Xu et al. "*Sp8* and *Sp9* coordinately promote D2-type medium spiny neuron production by activating *Six3* expression". In: *Development* (Jan. 2018).

[376] Xiaolei Song et al. “Homeobox Gene Six3 is Required for the Differentiation of D2-Type Medium Spiny Neurons”. In: *Neuroscience Bulletin* 37.7 (May 2021), pp. 985–998.

[377] Yingchao Shi et al. “Mouse and human share conserved transcriptional programs for interneuron development”. In: *Science* 374.6573 (Dec. 2021).

[378] Alessio Strano et al. “Variable Outcomes in Neural Differentiation of Human PSCs Arise from Intrinsic Differences in Developmental Signaling Pathways”. In: *Cell Reports* 31.10 (June 2020), p. 107732.

[379] Eek hoon Jho et al. “Wnt Beta-Catenin/Tcf Signaling Induces the Transcription of Axin2, a Negative Regulator of the Signaling Pathway”. In: *Molecular and Cellular Biology* 22.4 (Feb. 2002), pp. 1172–1183.

[380] Laura Buttitta et al. “Microarray analysis of somitogenesis reveals novel targets of different WNT signaling pathways in the somitic mesoderm”. In: *Developmental Biology* 258.1 (June 2003), pp. 91–104.

[381] Onur Basak et al. “Troy+ brain stem cells cycle through quiescence and regulate their number by sensing niche occupancy”. In: *Proceedings of the National Academy of Sciences* 115.4 (Jan. 2018).

[382] Stephen W. Wilson and John L.R. Rubenstein. “Induction and Dorsoventral Patterning of the Telencephalon”. In: *Neuron* 28.3 (Dec. 2000), pp. 641–651.

[383] Taisuke Kadoshima, Hideya Sakaguchi, and Mototsugu Eiraku. “Generation of Various Telencephalic Regions from Human Embryonic Stem Cells in Three-Dimensional Culture”. In: *Methods in Molecular Biology*. Springer New York, 2017, pp. 1–16.

[384] Naoya Amimoto et al. “Generation of striatal neurons from human induced pluripotent stem cells by controlling extrinsic signals with small molecules”. In: *Stem Cell Research* 55 (Aug. 2021), p. 102486.

[385] Roberto Alvarez-Medina et al. “Wnt canonical pathway restricts graded Shh/Gli patterning activity through the regulation of Gli3 expression”. In: *Development* 135.2 (2008), pp. 237–247.

[386] Xue-Jun Li et al. “Coordination of sonic hedgehog and Wnt signaling determines ventral and dorsal telencephalic neuron types from human embryonic stem cells”. In: *Development* 136.23 (Dec. 2009), pp. 4055–4063.

[387] Laetitia Aubry et al. “Striatal progenitors derived from human ES cells mature into DARPP32 neurons in vitro and in quinolinic acid-lesioned rats”. In: *Proceedings of the National Academy of Sciences* 105.43 (Oct. 2008), pp. 16707–16712.

[388] Alessia Delli Carri et al. “Developmentally coordinated extrinsic signals drive human pluripotent stem cell differentiation toward authentic DARPP-32+ medium-sized spiny neurons”. In: *Development* 140.2 (Jan. 2013), pp. 301–312.

[389] Mariko Sekiguchi et al. “Neuron type-selective effects of activin on development of the hippocampus”. In: *Neuroscience Letters* 452.3 (2009), pp. 232–237.

[390] Serafí Cambray et al. “Activin induces cortical interneuron identity and differentiation in embryonic stem cell-derived telencephalic neural precursors”. In: *Nature Communications* 3.1 (2012).

[391] Charles Arber et al. “Activin A directs striatal projection neuron differentiation of human pluripotent stem cells”. In: *Development* 142.7 (2015), pp. 1375–1386.

[392] V. Ribes and J. Briscoe. “Establishing and Interpreting Graded Sonic Hedgehog Signaling during Vertebrate Neural Tube Patterning: The Role of Negative Feedback”. In: *Cold Spring Harbor Perspectives in Biology* 1.2 (2009), a002014–a002014.

[393] Marie-Christin Pauly et al. “Organization of the human fetal subpallium”. In: *Frontiers in Neuroanatomy* 7 (2014).

[394] Camille Nicoleau et al. “Embryonic stem cells neural differentiation qualifies the role of Wnt/β-Catenin signals in human telencephalic specification and regionalization”. In: *Stem Cells* 31.9 (2013), pp. 1763–1774.

[395] Wei Jiang, Jinzhao Wang, and Yi Zhang. “Histone H3K27me3 demethylases KDM6A and KDM6B modulate definitive endoderm differentiation from human ESCs by regulating WNT signaling pathway”. In: *Cell Research* 23.1 (2012), pp. 122–130.

[396] F. Ille and L. Sommer. “Wnt signaling: multiple functions in neural development”. In: *CMLS Cellular and Molecular Life Sciences* 62.10 (2005), pp. 1100–1108.

[397] Takamasa Masuda and Tohru Ishitani. "Context-dependent regulation of the β-catenin transcriptional complex supports diverse functions of Wnt/β-catenin signaling". In: *Journal of Biochemistry* 161.1 (2016), pp. 9–17.

[398] Noel Moya et al. "Endogenous WNT Signaling Regulates hPSC-Derived Neural Progenitor Cell Heterogeneity and Specifies Their Regional Identity". In: *Stem Cell Reports* 3.6 (2014), pp. 1015–1028.

[399] E. Castren et al. "Treatment of Neurodevelopmental Disorders in Adulthood". In: *Journal of Neuroscience* 32.41 (2012), pp. 14074–14079.

[400] Christopher M. McGraw, Rodney C. Samaco, and Huda Y. Zoghbi. "Adult Neural Function Requires MeCP2". In: *Science* 333.6039 (2011), pp. 186–186.

[401] Sandra Luikenhuis et al. "Expression of MeCP2 in postmitotic neurons rescues Rett syndrome in mice". In: *Proceedings of the National Academy of Sciences* 101.16 (2004), pp. 6033–6038.

[402] Emanuela Giacometti et al. "Partial rescue of MeCP2 deficiency by postnatal activation of MeCP2". In: *Proceedings of the National Academy of Sciences* 104.6 (2007), pp. 1931–1936.

[403] Daniela Tropea et al. "Partial reversal of Rett Syndrome-like symptoms in MeCP2 mutant mice". In: *Proceedings of the National Academy of Sciences* 106.6 (2009), pp. 2029–2034.

[404] Jacky Guy et al. "Reversal of Neurological Defects in a Mouse Model of Rett Syndrome". In: *Science* 315.5815 (2007), pp. 1143–1147.

[405] Andrea Contestabile et al. "Lithium rescues synaptic plasticity and memory in Down syndrome mice". In: *Journal of Clinical Investigation* 123.1 (2012), pp. 348–361.

[406] Aubin Michalon et al. "Chronic Pharmacological mGlu5 Inhibition Corrects Fragile X in Adult Mice". In: *Neuron* 74.1 (2012), pp. 49–56.

[407] Jill L. Silverman et al. "Negative Allosteric Modulation of the mGluR5 Receptor Reduces Repetitive Behaviors and Rescues Social Deficits in Mouse Models of Autism". In: *Science Translational Medicine* 4.131 (2012).

[408] Emily K. Osterweil et al. "Lovastatin Corrects Excess Protein Synthesis and Prevents Epileptogenesis in a Mouse Model of Fragile X Syndrome". In: *Neuron* 77.2 (2013), pp. 243–250.

[409] Elizabeth M. Berry-Kravis et al. "Drug development for neurodevelopmental disorders: lessons learned from fragile X syndrome". In: *Nature Reviews Drug Discovery* 17.4 (2017), pp. 280–299.

[410] Siavash Fazel Darbandi et al. "Enhancing WNT Signaling Restores Cortical Neuronal Spine Maturation and Synaptogenesis in Tbr1 Mutants". In: *Cell Reports* 31.2 (2020), p. 107495.

[411] Siavash Fazel Darbandi et al. "LiCl treatment leads to long-term restoration of spine maturation and synaptogenesis in adult Tbr1 mutants". In: *Journal of Neurodevelopmental Disorders* 14.1 (2022).

[412] Wei-Lun Hsu et al. "Restoring Wnt6 signaling ameliorates behavioral deficits in MeCP2 T158A mouse model of Rett syndrome". In: *Scientific Reports* 10.1 (2020).