Single cell sequencing as a phenotyping strategy to decipher the molecular mechanisms of developmental disorders

Inaugural-Dissertation
to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry, Pharmacy
of Freie Universität Berlin

by

Jana Maria Henck

Berlin 2023

The work presented in the following was executed at the Max Planck Institute for Molecular Genetics in the research group of Prof. Dr. Spielmann in the Department of Development and Disease from March 2019 to October 2023.

1st reviewer: Prof. Dr. Malte Spielmann

2nd reviewer: Prof. Dr. Sigmar Stricker

Date of defense: 23.04.2024

Declaration of Independence

I herewith certify that I have formulated and written this thesis independently and that I have not used any other sources or references than those indicated within this document. Moreover, I have not applied for an examination procedure anywhere else nor submitted the dissertation in this or any other form to any other faculty.

“It is the brain, the little gray cells on which one must rely.
One must seek the truth within--not without." (Poirot)
-Agatha Christie,1920

Table of Content

Summary
Zusammenfassung

1. Introduction **1**
1.1 Deciphering the genome 1
1.1.1 The non-coding genome and the role of cis-regulatory elements (CREs) 2
1.1.2 Long range enhancer-promoter contact 4
1.1.3 Identification and distinction of promoters and enhancers 4
1.1.4 Examining enhancer function 5
1.2. Deciphering the 3D genome 6
1.2.1 Organization of the genome within the nucleus 6
1.2.2 3C-based methods to study 3D chromatin organization 7
1.2.3 Topologically associating domains (TADs) 8
1.2.4 Emerging role of structural variants in disease 9
1.3 Studying novel mutations 12
1.3.1 CRISPR/Cas9 technology in mouse models 12
1.3.2 Current research and mouse models of ADLD 14
1.3.3 The 3D locus of *Lmnb1* 15
1.3.4 ADLD pathology 17
1.4 Phenotyping methods to decipher mouse models 19
1.4.1. Conventional methods 19
1.4.2 scRNA-seq as a new technology in disease phenotyping 20

2. Aim of this Study **22**

3. Materials **23**
3.1 Antibodies, Enzymes & Probes 23
3.2 Bacterial strains & Plasmids 23
3.3 Buffers & Solutions 23
3.4 Chemicals 24
3.5 Instruments 25
3.6 Kits 25
3.7 Media 26
3.8 Oligonucleotides 26
3.9 Mouse lines 27
3.10 Software, Tools & Programming Language 27

4. Methods **29**
4.1 Standard methods 29
4.1.1 Designing & Cloning of single RNA guides for CRISPR/Cas9 29
4.2 Embryonic stem cell culture & mouse line generation 30
4.2.1 Preparation and seeding of feeder cells 30
4.2.2 Handling of mouse embryonic stem cells 30
4.2.3 CRISPR/Cas9 genome editing of mESC's for generating mutant clones 30
4.2.4 Extraction of genomic DNA 31
4.2.5 Design of Genotyping primers 31
4.2.6 Selecting CRISPR/Cas9 edited clones by genotyping 32
4.2.7 Aggregation and mouse line generation 32
4.3 Single cell RNA library preparation for the ADLD Project 33
4.3.1 Brain extraction and nuclei isolation 33
4.3.2 Library preparation and sequencing 34
4.4 10x scRNA-seq analysis 34
4.4.1 Mapping of sequencing reads, filtering and quality control 34
4.4.2 Integration, cell clustering and annotation 35
4.4.3 Gene Set Enrichment Analysis (GSEA) 35
4.4.4 Cell composition analysis 35
4.4.5 RNA velocity 36
4.5 Capture HiC 36
4.5.1 Capture Probe design 36
4.5.2 Sample fixation and lysis 37
4.5.3 3C-library preparation 37
4.5.4 Capture HiC library generation & sequencing 38
4.5.5 Capture HiC Analysis 39
4.5.6 Subtraction Maps 40
4.5.7 Fan-C Insulation score 40
4.6 Magnetic resonance imaging (MRI) of aging mutant ADLD mice 40
4.6.1 MRI experiment 40
4.6.2 MRI Data analysis of T2-weighted (T2w) and diffusion MRI (dMRI) axial diffusion matrices 41
4. 7 MMCA Methods 42

5. Results **55**
5.1 Mouse mutant cell atlas (MMCA) 55
5.1.1 Quality of the dataset 57
5.1.2 Cell composition analysis 58
5.1.3 LochNESS analysis in the *Gli2* KO mutant 60
5.1.4 *Gli2* KO mutant morphological changes in the neural tube and choroid plexus 63
5.1.5 Combinatorial in-depth analysis of the *Sox9* regulatory INV mutant 64
5.2 Mouse model of acute demyelinating leukodystrophy 68
5.2.1 The 3D chromatin changes upon ADLD based mutations 69
5.2.2 Molecular phenotype of the deletion and duplication mutant 73

5.2.3 Morphological phenotype of the aging mouse mutants 79

6. Discussion 82
6.1 Summary of the studies 82
6.2 Summary of the MMCA scRNA-seq dataset 82
6.2.1 Limitations of the MMCA scRNA-seq dataset 84
6.2.2 Perspectives of the MMCA scRNA-seq dataset 87
6.3 Summary of the findings in the ADLD project 88
6.3.1 Capture HiC results require follow up exploration of the locus 88
6.3.2 Structural variants result in early molecular changes within both mutants 90
6.3.3 Single cell RNA-seq analysis reveals cholesterol and glia differentiation difference in the mutants 91
6.3.4 MRI reveals morphological changes within the duplication mutant 92
6.3.5 Directions of exploration in the ADLD scRNA-seq dataset 92
6.3.6 MRI requires replicates for significance statement 94
6.3.7 Lamin B1 involvement in global 3D genome architecture in disease development 95
6.4 Conclusion and Outlook 96

7. Appendix 98
7.1 Plasmid Maps 98
7.2 Supplementary Figures 99
7.3 Supplementary Tables 101
7.4 List of Abbreviations 104
7.5 List of Figures 106
7.6 List of Tables 107

8. References 108
9. Acknowledgements 125
10. Scientific Publications 127
11. Scientific Conferences 128

Summary

In order to understand and eventually cure genetic disease, it is essential to understand which transcriptional and cellular mechanisms lead to the symptoms experienced by patients. Over the past few decades, the introduction of variants into the genome of laboratory mice as model organisms has been a central part of disease research. The process of generating mouse mutants has been accelerated by the discovery of the genome editing technology CRISPR/Cas9. However, the complete analysis of a mouse mutant with conventional phenotyping methods remains laborious. Another layer of complexity has been the discovery that non-coding mutations can be disease causing. Studies in cancer and rare skeletal phenotypes have demonstrated that variants involving the boundaries of the genome organizational units of topologically associating domains (TADs) can cause disruption of the functional units of cis-regulatory elements (CREs), leading to misexpression of genes within the locus and consequent development of disorders or diseases.

In this work, we have addressed two central cornerstones of disease research. First, the development and application of a new strategy for phenotyping mouse mutants, in which we tested single-cell RNA sequencing as a technology for unbiased and comparative study of whole mouse mutants. Second, we extended the role of TAD variants to the highly prevalent field of neurological disorders.

In the first project, we generated the mouse mutant cell atlas (MMCA) composed of 22 mouse mutants of varying severity. We developed novel analysis tools to detect and visualize mutant phenotypes at different granularities. By applying scRNA-seq to all mutants in a single experiment, we gained new knowledge about previously studied mutants, uncovered new phenotypes, unraveled the cellular dynamics of disease progression, and established the toolkit for comparative and unbiased whole-embryo analysis.

In the second project, we performed an in-depth analysis of the *Lmnb1* related neurodegenerative disorder acute demyelinating leukodystrophy (ADLD). We generated mouse models of a patient derived duplication of the gene itself and a deletion upstream of the gene. We discovered 3D conformational changes within the locus caused by the deletion, early onset of molecular changes, and morphological changes in the adult mice, together indicating differences in disease progression between the two variants.

In summary, both projects contribute to the field of disease research by establishing scRNA-seq as a tool to generate and analyze mutants and by providing an in-depth analysis

of a neurodegenerative disease, challenging the current categorisation of this group of diseases as "late-onset".

Zusammenfassung

Um genetisch bedingte Krankheiten zu verstehen und eventuell zu heilen, müssen zunächst transkriptionelle und zelluläre Mechanismen verstanden werden, die zu klinischen Anomalien und Symptomen führen. In den letzten Jahrzehnten ist die Einführung von Varianten in das Genom von Labormäusen ein zentraler Bestandteil der Krankheitsforschung. Während die Erzeugung von Mausmutanten durch die Entdeckung von CRISPR/Cas9 erheblich beschleunigt wurde, ist die Phänotypisierung einer Mausmutante mit herkömmlichen Methoden nach wie vor ein aufwendiger Prozess. Ein weiterer wichtiger Faktor ist die Entdeckung, dass auch nicht-kodierende Mutationen Krankheiten verursachen können. Studien über Krebs und seltene Skelettphänotypen zeigen, dass Varianten, welche die Grenzen von topologisch assoziierten Domänen betreffen, eine Störung der funktionellen Einheiten von cis-regulatorischen Elementen verursachen können, was zu einer Fehlexpremierung von Genen und folglich zur Entwicklung von Störungen oder Krankheiten führt.

In dieser Arbeit haben wir uns mit zwei zentralen Eckpfeilern der Krankheitsforschung befasst. Erstens, die Entwicklung und Anwendung neuer Strategien zur Phänotypisierung von Mausmutanten, bei der wir Einzelzell-RNA-Sequenzierung als Technologie für unvoreingenommene und vergleichende Untersuchungen ganzer Mausmutanten getestet haben. Zweitens haben wir die Rolle der TAD-Varianten auf das Gebiet der neurologischen Erkrankungen ausgeweitet.

Im ersten Projekt haben wir den “mouse mutant cell atlas” erstellt, der aus 22 Mausmutanten besteht. Wir haben neue Analysestrategien zur Erkennung und Visualisierung von Mutationsphänotypen entwickelt. Durch die Anwendung der Technologie auf alle Mutanten in einem einzigen Experiment haben wir neue Erkenntnisse über bereits untersuchte Mutanten gewonnen, neue Phänotypen entdeckt, zelluläre Dynamiken aufgeklärt und die Methode für vergleichende und unvoreingenommene Analysen etabliert.

Im zweiten Projekt führten wir eine eingehende Analyse der mit *Lmnb1*-assoziierten neurodegenerativen Erkrankung “Akute Demyelinisierende Leukodystrophie” durch. Wir generierten Mausmodelle mit einer von Patienten stammenden Duplikation des Gens und einer Deletion nahe dem Gen. Wir entdeckten durch die Deletion verursachte Konformationsveränderungen, frühe molekulare Veränderungen und morphologische

Veränderungen in den erwachsenen Mäusen, die zusammen auf Unterschiede zwischen den beiden Varianten hinweisen.

Zusammengefasst leisten beide Projekte einen Beitrag zur Krankheitsforschung, indem sie die scRNA-Sequenzierung als Werkzeug zur Erzeugung und Analyse von Mutanten etablieren und eine detaillierte Analyse einer neurodegenerativen Erkrankung liefern, die die derzeitige Einstufung dieser Krankheitsgruppe als "spät" in Frage stellt.

1. Introduction

1.1 Deciphering the genome

Pursuing research in the field of human genetics has been central in promoting knowledge about the development and diversity of the human species, as well as causes of diseases that result from mutations in the genome. In 2000, when the first draft of the human genetic code was published, the president of the White House commented in a press conference on this milestone, stating “(..) It will revolutionize the diagnosis, prevention and treatment of most, if not all, human diseases. In coming years, doctors increasingly will be able to cure diseases like Alzheimer's, Parkinson's, diabetes and cancer by attacking their genetic roots.” (“June 2000 White House Event” n.d.). The human genome project worked on this endeavor from 1990 until published in 2004, half a decade after Franklin, Watson and Crick published their papers about the double helix structure of the DNA (Venter et al. 2001; “A Physical Map of the Human Genome” 2001, “Finishing the Euchromatic Sequence of the Human Genome” 2004). Aside from the human genome, also the genome of model organisms like Drosophila Melanogaster, C. Elegans and Mus Musculus (mouse) were sequenced during that time (Gregory et al. 2002; Adams et al. 2000; C. elegans Sequencing Consortium 1998).

The information provided with the publication of these datasets allowed huge leaps in our understanding of the genetic code and revolutionized our research in the field of evolution, development and diseases. For example, the decryption of the human genome allowed us to decipher bigger scale mutations. Therefore, it made sense that an accumulation of papers reporting about rearrangements in the genome resulting in disease were published during that time (Lupski and Stankiewicz 2005; Mazzarella and Schlessinger 1998; Christian et al. 1999; Ji et al. 2000)

Limited by the available methods (short read sequencing, stitched together computationally) the assembled human genome from 2004 was incomplete until recently. By applying new long read sequencing methods such as Oxford Nanopore and PacBio DNA sequencing, two decades after the first publications, a more complete version of the human genome, including the heterochromatin regions was published by the Telomere to Telomere (T2T) consortium (Nurk et al. 2022; Aganezov et al. 2022).

Contrary to the statement made in the White House in the 2000’s, we know nowadays, that, although this event had been a giant leap in the field of human genetics, solving mutation based diseases is exceeding knowledge about the genetic code at base pair resolution. Deciphering the genetic code enabled researchers to investigate the genome for big chromatin arrangements to coding mutations in risk genes in, for example, Alzheimers or

Parkinsons (Bekris et al. 2010; C. Klein and Westenberger 2012). However, disease-causing variants are discovered in the clinic that are not in the coding sequence of a known gene. Deciphering the mechanism of disease caused by these variants goes beyond our knowledge of the genetic code itself and requires knowledge of the organization and non-coding sequences of the genome. Recent studies on structural variants that disrupt the gene regulatory network have demonstrated that there are other causes besides mutations in the coding genome resulting in disease, introducing the field of human genetics to a new research area.

The following introduction is a summary of the state of knowledge about the composition of the non-coding genome, specifically cis-regulatory elements and the organization of the 3D genome in disease and the methods applied to survey them. The first part (1.1) will further contain an overview of the state of knowledge of cis-regulatory elements, followed by a summary of the current state of knowledge about the 3D genome organization and structural variants affecting these elements (1.2). The third part (1.3) focuses on how to investigate these variants in mouse models using CRISPR/Cas9 technology and deciphering the phenotype of mutations and the state of art of research in the neurodegenerative disorder autosomal dominant leukodystrophy (ADLD) within mouse development. Finally (1.4) we are introducing single cell RNA sequencing as a method for phenotyping mouse models of disease.

1.1.1 The non-coding genome and the role of cis-regulatory elements (CREs)

Cis-regulatory elements (CREs) are non-coding sequences of hundred to thousand base pairs (bp) in the genome that regulate the transcription of the coding genome by transcription factors binding to the respective genomic region. Two of the most well known and studied CREs are promoters and enhancers (Moore et al. 2020; Cramer 2019).

Promoters are positioned upstream of the transcriptional start site (TSS) of the coding sequence of a gene. They consist of two parts, the core and the proximal promoter. The core promoter is a 50-100 bp sequence consisting of transcription factor binding sites, like the TATA-box and frequently contain CpG islands (Bannister and Kouzarides 2011; Roy and Singer 2015). Core promoters are involved in the assembly of the preinitiation complex (PIC), which contains RNA-Polymerase II, required for the initiation and elongation of transcription (X. Chen et al. 2021). In short, promoters regulate the basal transcriptional activity of its endogenous gene.

Gene transcription can be regulated by enhancers which, in contrast to the promoter, can be proximal or distal to the gene with some enhancers reported to regulate a gene from 1

Megabase (Mb) distance (Lettice et al. 2014). Enhancers are regulatory sequences of 0.2-1 kilobases (kb) size that are orientation and relative-distance independent, binding transcription factors needed for tissue specific gene activation (Pennacchio et al. 2013; Levine 2010).

Until now, it remains unclear how orientation, spacing, order and distance of the respective sequences are impacting the affinity and specificity between promoter and enhancer. Deciphering the molecular properties encoded in the sequence of an enhancer is investigated in the field of “enhancer grammar” (Jindal and Farley 2021). Adding to the complexity, it is estimated that over a million enhancers are present in the genome, outnumbering the amount of genes present (human ~ 20.000 genes) by far (Karnuta and Scacheri 2018; Shen et al. 2012). Having multiple enhancers regulating a gene facilitates tissue and developmental stage specific gene expression. In some cases, more than one enhancer can influence the gene expression of a specific gene, crucial for the correct orchestration of the spatio-temporal development of an organism (**Fig. 1.1**) (Uyehara and Apostolou 2023). It has been reported that multiple enhancers regulate the complex expression of one gene in additive, competitive or synergistic manner (**Fig. 1.1**) (J. Huang et al. 2018, 2016). Additional research has revealed that multiple, strong enhancers that are found in clusters to drive certain developmental genes can be defined as ‘super enhancers’ or (partially) compensate for the loss of another enhancer as redundant or ‘shadow enhancers’ (Cannavò et al. 2016; Hnisz et al. 2013).

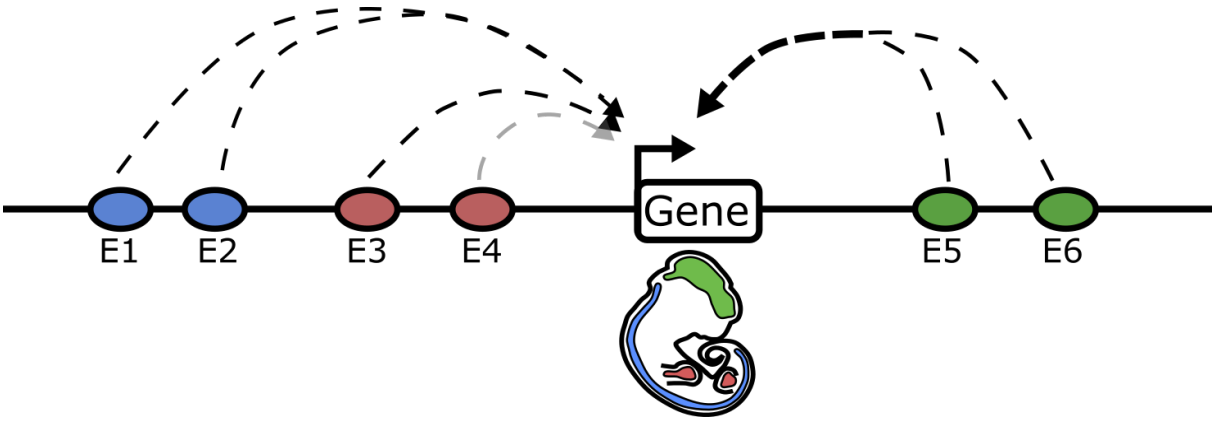


Figure 1.1 The activity of a gene can be regulated by a complex landscape in different tissues

The activity of a single gene can be regulated by multiple enhancers in different tissues. This schematic of a regulatory landscape visualizes the expression of a gene in the brain (green), the neural tube (blue) and the limbs (red). Enhancers dispersed across the landscape can regulate tissue specific gene expression in many ways such as in an additive (E1 and E2), competitive (E3 and E4) or synergistic (E5 & E6) way. The figure is inspired by (Peng and Zhang 2018).

1.1.2 Long range enhancer-promoter contact

One of the many features distinguishing enhancers from promoters is their variability in distance to the respective gene. While promoters are at a defined position upstream of the coding part of the gene, enhancers can be found dispersed throughout the locus. In consequence, enhancers are defined more through their function rather than their position. Exactly how enhancers regulate their targets gene expression remains largely unclear, especially when enhancers are positioned distally from their target promoter, as in the example of the Sonic hedgehog (*Shh*) gene, which is expressed in the proximal and distal limb by its enhancer ZRS, 1Mb away from the gene (Furlong and Levine 2018; Lettice et al. 2014). A common working hypothesis is that physical proximity needs to be established between enhancers and their target promoters to drive expression. Exactly how enhancer-promoter contact is facilitated remains active research. Several models have been proposed: One is the cohesin-mediated loop extrusion model, in which cohesin, a ring-shaped protein complex, is bringing enhancers and promoters into proximity (Banigan and Mirny 2020; Yatskevich, Rhodes, and Nasmyth 2019). Another model postulates that specific transcription factors of the promoter and enhancer together with the mediator complex (Med), a multi-protein complex, act as a bridge in between the two regulatory sequences leading to tissue specific gene expression (Allen and Taatjes 2015; Petrenko et al. 2016).

Recently, the condensate/phase separation model was proposed, in which the contact of enhancers and promoters is facilitated through drop-like aggregates of transcription facilitating elements such as transcription factors, the mediator complex and RNA-Polymerase II (Boija et al. 2018; Sabari et al. 2018; Hnisz et al. 2017). This model proposes a possible mechanism for the simultaneous proximity of several enhancers to a promoter without direct and exclusive contact.

Aside from the exact mechanism, however, it is clear that enhancers need to be brought into physical proximity to the promoter for it to control gene expression in a distinct spatio-temporal manner for precise development of complex organisms.

1.1.3 Identification and distinction of promoters and enhancers

The identification of CRE's has been central in understanding the underlying mechanisms of gene regulation. Promoters and enhancers can be identified and distinguished by their respective epigenetic profiles. A layer of organization is the packaging of DNA which is wrapped around histone octamers, called nucleosomes. These proteins can be post-transcriptionally modified through acetylation, methylation, ubiquitination and

phosphorylation on their N-terminal domains, which are established read outs of the chromatin state of the DNA that is wrapped around (Bannister and Kouzarides 2011).
Assays such as chromatin immunoprecipitation and sequencing (ChIP-seq) have been applied to identify the histone modification and the respective epigenetic properties of a DNA sequence (Mikkelsen et al. 2007). For example, active promoters are enriched for the histone motive H3K4me1 and H3K4me3, while active enhancers only display the H3K4me1 histone mark (Heintzman et al. 2007; Allis and Jenuwein 2016).
Aside from histone profiles, candidate CREs can be profiled with various methods capturing the chromatin profile by, for example, DNaseI-hypersensitive site sequencing (DNase-seq) Assay for Transposase-Accessible Chromatin sequencing (ATAC-seq) or DNA methylation, which is accessible through whole-genome bisulfite sequencing (WGBS) (Bannister and Kouzarides 2011; Buenrostro et al. 2013; Thurman et al. 2012) .

1.1.4 Examining enhancer function

While promoters can be identified due to their position upstream of the gene body, enhancers are dispersed throughout the locus surrounding the gene (Despang et al. 2019; Robson, Ringel, and Mundlos 2019). With the goal to identify potential enhancers and understand their function, several methods have been developed.
In order to examine a potential enhancer, the LacZ reporter assay is a standard method applied to measure the regulatory activity of a respective sequence *in vivo* (O'Kane and Gehring 1987; Kothary et al. 1989). Executed in large scale experiments, databases such as the VISTA enhancer browser offer the possibility to assess the activity of non-coding elements in mouse embryos at embryonic stage E11.5 to reveal the diverse and precise spatiotemporal activity of human and mouse candidate enhancers (Visel et al. 2007). As this endeavor required the generation of hundreds of transgenic mice and thus remained a laborious low-throughput analysis of single non-coding sequences, other recently developed sequencing methods such as Massively Parallel Reporter Assays (MPRAs) present a high throughput alternative to facilitate the analysis of potential enhancer activities of thousands of elements in parallel (Inoue and Ahituv 2015).

1.2. Deciphering the 3D genome

1.2.1 Organization of the genome within the nucleus

As previously touched on and proposed in various models, the need for enhancers to come into proximity with their target promoter is a requirement of precise gene expression. The exact mechanism remains elusive, but the contact between distal enhancers and their promoters has to be facilitated by changes in chromatin structure, amongst others. The organization of chromatin within the nucleus has been a subject to researchers for many decades. In the early 1930's the term heterochromatin was established by Emil Heitz. Looking at plants and Drosophila nuclei, he discovered stronger stained regions within the nucleus hypothesizing that these regions were less active or contained no genes (Heitz 1933; J. Liu, Ali, and Zhou 2020). With the development of technologies such as electron microscopy in the 1960's, the knowledge expanded into the discovery of chromosome territories, which elucidated the non-random organization of the genome within a nucleus (Brown 1966; Cremer et al. 1982). This marked the beginning of decade-long research on chromatin organization and its impact on gene expression, which is still subject to many studies today. Over the years, it has been established that chromatin is organized in several hierarchies. On the macro-level, chromosomes are located in distinct areas, so called chromosome territories (**Fig. 1.2**). These are subdivided into compartments, which can be distinguished into two functional groups: the active A-compartments and the repressed, nuclear envelope associated B-compartments (**Fig. 1.2**) (Lieberman-Aiden et al. 2009).
With the aim to explore the chromatin architecture at higher resolution, several methods have been developed and applied. Recently, the combination of fluorescence in situ hybridization (FISH), modeling the 3D architecture, and 3-C based methods has been employed to provide a comprehensive understanding of chromatin organization within a specific locus and its interacting elements (Mifsud et al. 2015; Ringel et al. 2022; Jerkovic´ and Cavalli 2021).

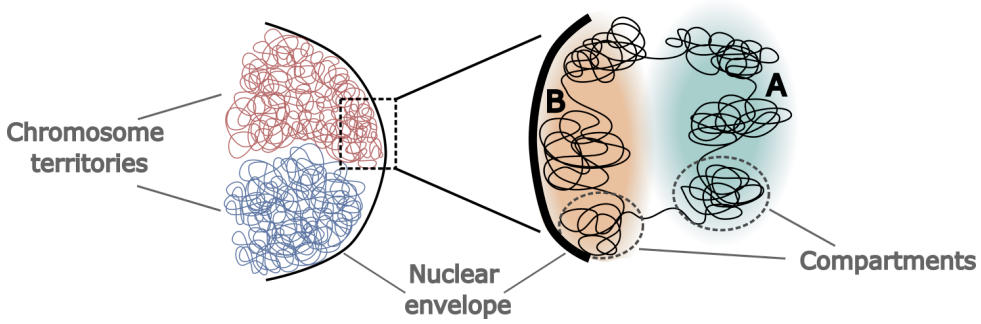


Figure 1.2 Schematic visualization of chromatin organization in the nucleus

Chromosomes are organized into territories in the nucleus. As a functional sub-unit, chromatin organizes into active A-compartments and inactive B-compartments. The B-compartments are likely organized closer to the nuclear envelope. The figure was inspired by (Szabo, Bantignies, and Cavalli 2019)

1.2.2 3C-based methods to study 3D chromatin organization

Apart from imaging and computational modeling based methods, the rise of chromosome conformation capture (3C) - based methods have elevated our understanding of the 3D genome architecture (reviewed in:(Bonev and Cavalli 2016)). Briefly, most 3C based methods include the following steps: reversible crosslinking of chromatin, fragmentation by digestion with a restriction enzyme and ligation of the open DNA ends. Afterwards, it can be followed up with a variability of extensive steps to investigate one specific contact (3C), all contacts from one locus of interest (4C) or all contacts across the genome (HiC) (Dekker et al. 2002; Lieberman-Aiden et al. 2009). Another interesting extension of the 3C-based methods is the development of capture HiC (cHiC) (all contacts within a defined region). This method includes a step of enrichment for a defined genomic region through the introduction of target specific oligonucleotide probes, creating high resolution contact maps as outputs, allowing the analysis of that region in a specific tissue or cell type (**Fig. 1.3**) (Mifsud et al. 2015). This additional step generates output contact maps with extremely high resolution for a region of interest.

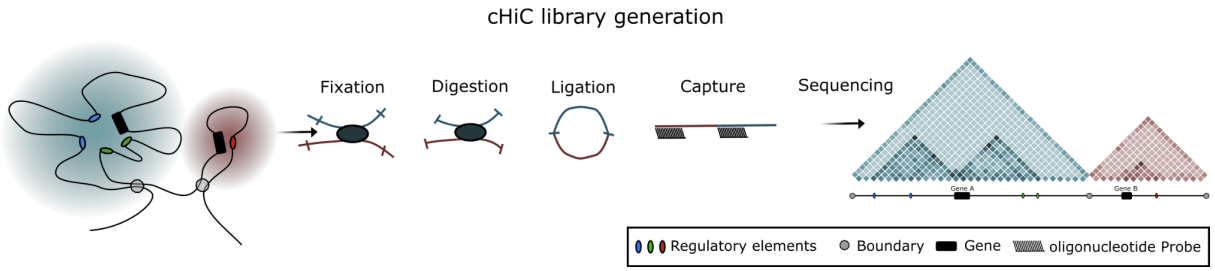


Figure 1.3 Schematic overview of cHiC library generation
Schematic visualization of two topologically associating domains (blue and red) separated by boundaries, how they might be positioned in a nucleus within a specific tissue. Briefly, in the capture HiC workflow, after extracting a specific tissue, the chromatin is fixed, digested with a restriction enzyme and ligated, these steps are general to 3C based methods. Unique to the capture HiC method is the step of capture by locus specific designed probes that allow the pull down of DNA sequences from the desired region. After sequencing and mapping, the output is a high-resolution cHiC map that visualizes the 3D chromatin interactions at that locus. The figure was inspired by (Robson, Ringel, and Mundlos 2019; Bonev and Cavalli 2016)

1.2.3 Topologically associating domains (TADs)

Applying 3C-based methods such as HiC and cHiC enables the analysis of a sub-organization of compartments, namely topologically associating domains (TADs). TADs are self-interactive chromatin regions often overlapping with the functional regulatory interactions between genes and their respective regulatory elements at sub-megabase size (**Fig. 1.4a**) (Bolt et al. 2022; Dekker and Heard 2015). The composition of TADs is known to be tissue specific and to dynamically alter structure throughout the development of a tissue, typically ranging between a few hundred kb and several Mb with a median size of 880 kb in mammalian genomes (Dixon et al. 2012; Dekker and Heard 2015; Winick-Ng et al. 2021).
As mentioned above, TADs are understood as long range self-interactive domains, isolated from contacting the neighboring TAD by boundary elements (Rajderkar et al. 2023; H. Huang et al. 2021). In the example schematics, Gene A is regulated by the green and blue enhancer elements in the brain and spinal cord, marked by the increased signal between the enhancers and the gene (**Figure 1.4a**). Separated by a boundary element, Gene B is regulated by the red enhancer in the limb. The precise mechanism of insulation is still active research, with one commonly cited model proposing that insulation is facilitated by the enrichment of convergently orientated binding sites for the CCCTC-binding factor (CTCF) at boundary elements on which the chromatin extruding structural maintenance of chromosomes (SMC)-complex cohesin halts, creating self interacting regions within these boundaries that are isolated from the neighboring chromatin (Merkenschlager and Nora 2016; Dixon et al. 2012; Rao et al. 2014; Sexton et al. 2012). Other loci are reported to

create TAD boundaries through increased gene expression as in the example of *Zfp608* in neuronal differentiation (Bonev et al. 2017).

TAD boundaries are typically visualized by CTCF ChIP-seq, but recent analysis tools, like the Fan-C tool, allow the prediction of boundaries from HiC and cHiC data by calculating insulation scores from the contact maps (Nanni, Ceri, and Logie 2020; Kruse, Hug, and Vaquerizas 2020).

Supportive of the imposed correlation between TAD organization and functional gene regulation, studies have investigated enhancer positioning within and outside the boundaries of their target gene's TAD. These studies demonstrate that enhancers located within the TAD boundaries exhibit higher contact frequency to their promoters and result in higher gene expression compared to those outside the boundaries (Despang et al. 2019; Zuin et al. 2022; Bolt et al. 2022).

Taken together, research on TADs elucidate their role of genome organization into dynamic self-interacting domains, providing insulation for gene regulation, crucial in development.

1.2.4 Emerging role of structural variants in disease

When searching for disease-causing variants, the coding genome has been the focus of researchers. This changed when mutations in non-coding elements like CREs or TAD boundaries were identified as the cause of diseases (Spielmann, Lupiáñez, and Mundlos 2018). Structural variants that are disease causing can be subcategorized into balanced (translocations, inversions) and unbalanced mutations (deletion, duplication).

They are considered intra-TAD structural variants when the affected regulatory elements lie within a TAD without disruption of the boundary elements insulating the region (**Figure 1.4**). Depending on the type of mutation, these can result in gene dosage effects on the target gene. Duplication of an enhancer element has been shown to result in the upregulation of endogenous gene expression, while the deletion of enhancers has been shown to result in reduction or complete loss of the target gene expression (**Figure 1.4b,c**). Examples of such variants are the duplicated *Shh* enhancer ZRS causing the Laurin-Sandrow syndrome and the deletion of a part of the regulatory landscape of *Sox9* associated with the Pierre-Robin sequence (Benko et al. 2009; Lohan et al. 2014). Inversions have not yet been reported to result in gene expression changes, which could be because of the more difficult detection of these balanced mutations or their more subtle rearrangement of the genomic content(**Figure 1.4d**).

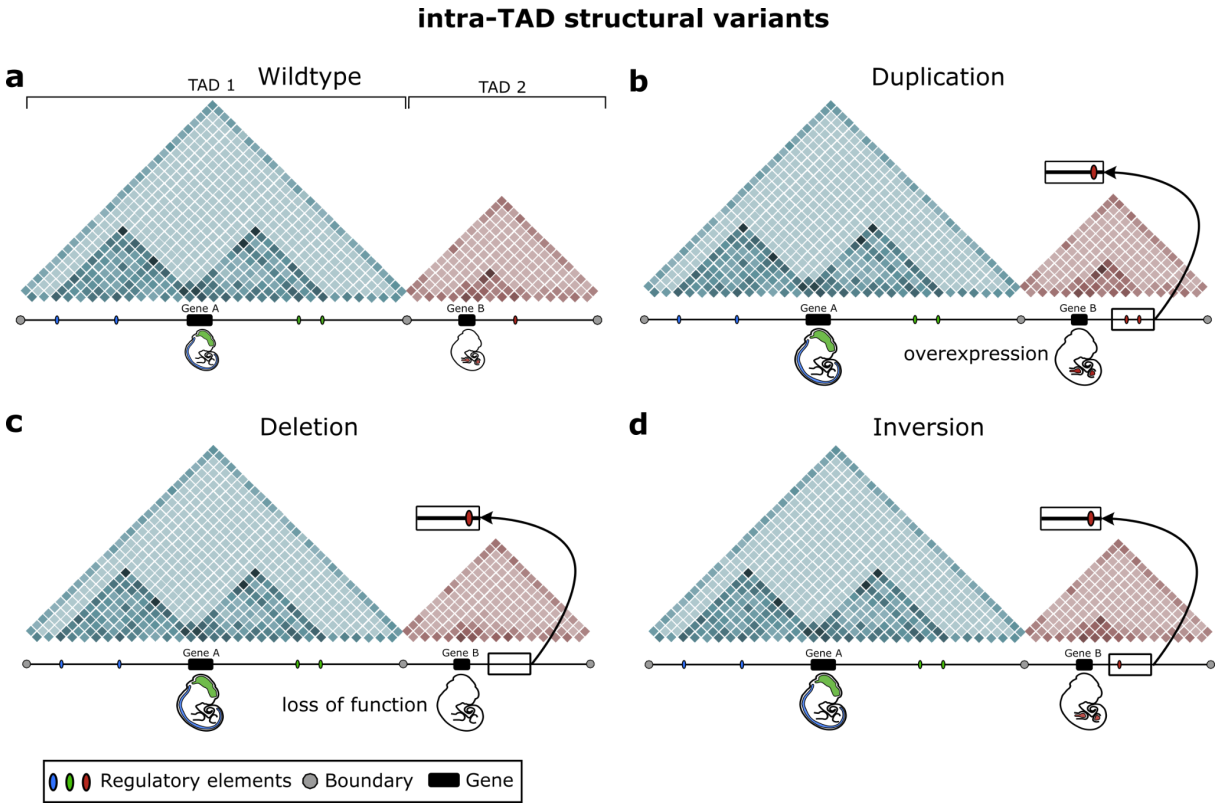


Figure 1.4 Intra-TAD structural variants
a,In the schematic representation of the wildtype genomic locus, gene A is expressed in the developing brain and gene B is expressed in the developing limbs. Both genes are regulated by their own tissue-specific cis-regulatory elements (red and blue, respectively) located in different TADs separated by boundary elements **b,c,d,** Intra-TAD structural variants can result in various gene expression changes. **b,c,** The duplications of enhancers can result in overexpression while the deletion can result in the loss of expression of the target gene's expression. **d,** Balanced mutations such as Inversions have not yet been reported to lead to distinct phenotypes.

When the structural variants are spanning TAD boundaries, they are considered inter-TAD structural variants (**Figure 1.5**). These mutations can result in disrupted or rearranged 3D chromatin organization of a locus by repositioning TAD boundaries and rewiring regulatory elements in the locus (Lupiáñez, Spielmann, and Mundlos 2016; Spielmann, Lupiáñez, and Mundlos 2018). Many of these genomic rearrangements have been found to result in cancer or congenital disorders (Claringbould and Zaugg 2021; Spielmann, Lupiáñez, and Mundlos 2018). Structural rearrangements in a locus spanning a TAD boundary can lead to the separation of an enhancer and its endogenous gene, which results in ectopic contact of the enhancer to another gene, a process that has since been established as 'enhancer adoption' or 'enhancer hijacking' (Lettice et al. 2011; Northcott et al. 2014). For example, duplications spanning TAD boundaries can result in the formation of a "neo" TAD, like in the locus of *Sox9* and *Kcnj2*, which led to the upregulation of *Kcnj2* expression and a malformation of the limb, associated with the Cooks syndrome (**Fig. 1.5a**) (Franke et al. 2016).

Inversions spanning a TAD boundary have previously been shown to result in the ectopic expression of a gene in the pattern of the hijacked enhancer (**Fig. 1.5b**). In a study within the same locus, the *Sox9*_InvC mouse mutant, in which a part of the *Sox9* regulatory landscape is inverted away from its target gene, the enhancers are hijacked by the neighboring TAD's gene *Kcnj2*. This inversion results in *Kcnj2* expression in a *Sox9*-like pattern in the developing limb and malformation of bone development for example of the digits (Despang et al. 2019).

The deletion of boundaries can result in "fused" TADs, which has been described at the locus of *Epha4*, resulting in the ectopic activation of the neighboring TADs gene *Pax6* by a cluster of enhancers and a congenital malformation of the limbs, namely brachydactyly (**Fig. 1.5c**) (Lupiáñez et al. 2015). Most diseases associated with inter-TAD structural variants are rare skeletal phenotypes and cancers (Dixon et al. 2018; Xu et al. 2022; Botten et al. 2023; Lupiáñez et al. 2015; Despang et al. 2019; Kraft et al. 2019). However, given that neurological disorders exhibit notably higher prevalence rates compared to rare skeletal congenital disorders, expanding research to focus on conditions other than rare developmental disorders is a high priority (MacDonald et al. 2000).

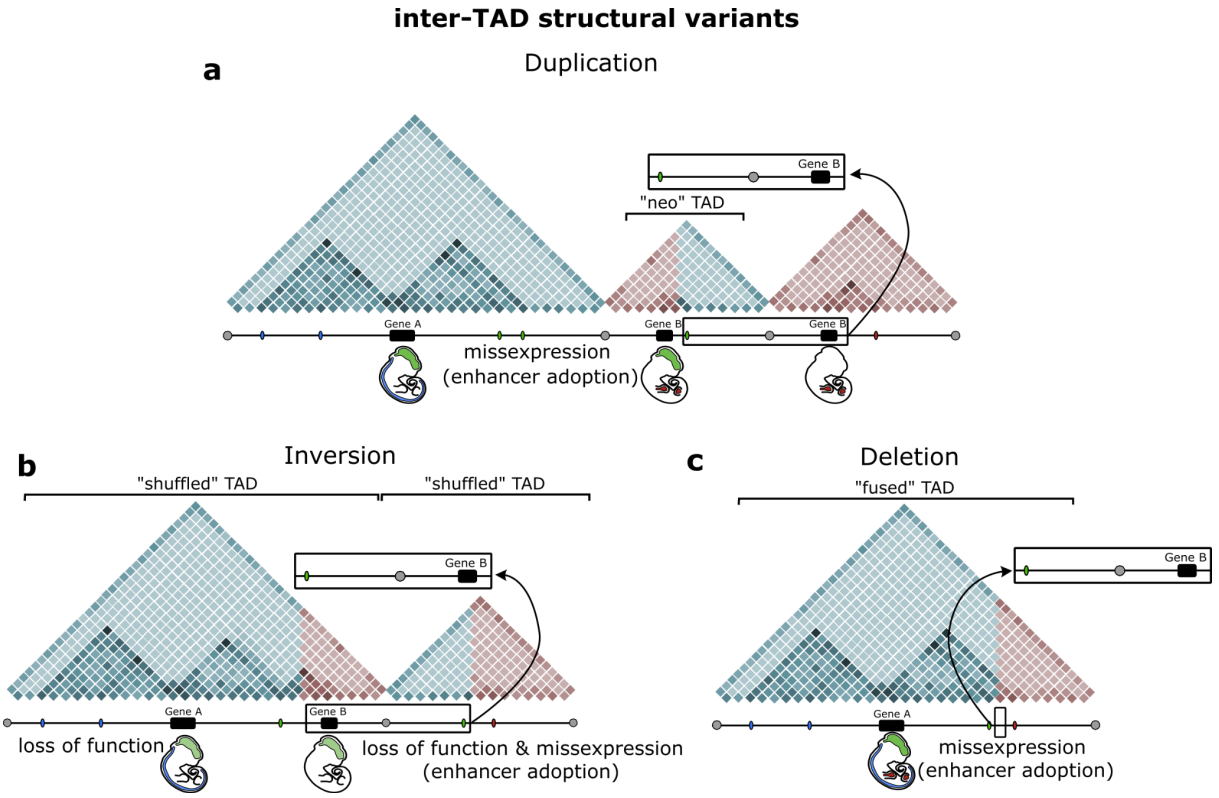


Figure 1.5 Inter-TAD structural variants
Mutations spanning the boundaries of TADs are considered Inter-TAD structural variants. **a,** Duplications of a region involving a TAD boundary can result in the formation of "neo" TADs, creating a new regulatory unit **b,** Inversions can result in a shuffled TAD structure, relocating genes and regulatory elements away from their endogenous functional units, resulting in their misexpression **c,**

When deletions span TAD boundaries they can result in “fused” TADs, meaning the fusion of two neighboring TADs and also the rewiring of the regulatory elements previously separated. All together these structural rearrangements can result in the rewiring of the regulatory landscape and cause loss in function as well as misexpression through enhancer adoption. The figure was inspired by (Spielmann, Lupiáñez, and Mundlos 2018)

1.3 Studying novel mutations

Identifying de novo mutations through clinical research has provided crucial insights into the genetic basis of various diseases, particularly in cases where the pathogenic mechanism requires further investigation. Various animal models and cell culture methods, including induced pluripotent stem cells derived from patient cells and the promising development of organoid systems, have been instrumental in generating new discoveries (Eichmüller and Knoblich 2022). For over a century, a standard animal model for studying human diseases has been the laboratory mouse (Rosenthal and Brown 2007). There are several advantages for choosing a mammalian system, particularly mice: First, the comparative genetics to the human biology in contrast to other model organisms, second, their, for mammals, comparatively fast life cycle, third, the high number of offsprings, and lastly, easy husbandry (Bryda 2013; Vandamme 2014).

1.3.1 CRISPR/Cas9 technology in mouse models

The discovery of the bacterial antiviral immune defense and the consecutive development into the genome editing CRISPR (**c**lustered **r**egularly **i**nterspaced **s**hort **p**alindromic **r**epeats)/Cas9 tool has vastly accelerated the precise generation of mutations in the genome of interest (Jinek et al. 2012; Ishino et al. 1987). Previous genome editing tools, such as zinc finger nucleases and transcription activator–like (TAL) effectors, were outperformed by CRISPR/Cas9 in the design, validation and overall mutation generation time as represented by the fact that it took half a year from the introduction as a genome editing tool to the first applications in mammalian genomes (Jinek et al. 2012; Cong et al. 2013; Mali et al. 2013; Doudna and Charpentier 2014; Gaj, Gersbach, and Barbas 2013).

In short, the bacterial endonuclease Cas9, when combined with a single guide RNA (sgRNA), that is specifically designed to complement the genomic region of interest, can be transfected into a target cell to created double strand breaks (DSB) at the precise site of interest in the genome (**Fig. 1.6a**) (Jinek et al. 2012). These breaks are repaired by one of the target cell’s two pathways. One pathway is error-prone non-homologous end joining (NHEJ), which randomly recombines the two loose DNA ends, creating a few nucleotide alteration of insertions or deletions (indel), also known as repair scars. Alternatively the

breaks are repaired by the high fidelity homology-directed repair (HDR), that utilizes the sister chromatid strand as a template for the repair (Jasin and Rothstein 2013; Hartlerode and Scully 2009). Creating indels by the NHEJ repair pathway can be used to generate gene knockouts when targeting the coding exon of a gene (Shalem, Sanjana, and Zhang 2015; Ran et al. 2013). Moreover, through cotransfection of the target cell with a donor DNA such as plasmids carrying a sequence of interest the open site created by the CRISPR/Cas9 can be utilized for targeted insertion of template DNA into the region of interest (Zhang et al. 2017; Selvaraj et al. 2019).

Further, designing two sgRNA surrounding a region of interest initiated the generation of genomic structural variants, through which one can create kilobase to megabase sized deletions, duplications or inversions (**Fig. 1.6b**) (Kraft et al. 2015). The application of CRISPR/Cas9 to create genomic variants in the mouse genome has since been streamlined to a 10 week protocol, making this approach undeniably the fastest method to create structural variations in the mammalian genome to date (Kraft et al. 2015).

CRISPR has advanced into a toolbox of methods to repress, activate and screen multiple sites in the genome at once (Larson et al. 2013; Konermann et al. 2014; Shalem, Sanjana, and Zhang 2015; Bock et al. 2022). Also, new bacterial endonucleases are investigated for a more flexible implementation of the tool, as the choice of cutting sites is limited by the required presence of the protospacer adjacent motifs (PAM) at the site for Cas9 endonuclease activity (G. Liu et al. 2022).

CRISPR technologies have been put to use in a plethora of areas aside from research conducted in molecular biology and genomics, like agriculture and clinical application (Waltz 2022; Li et al. 2022). Especially in the field of patient targeted approaches, gene editing tools provide encouraging first results in clinical trials editing sickle cell disease in patients, opening up further opportunities of the use of genome editing in clinical application (Zarghamian, Klermund, and Cathomen 2022; Demirci et al. 2019; Urnov 2021).

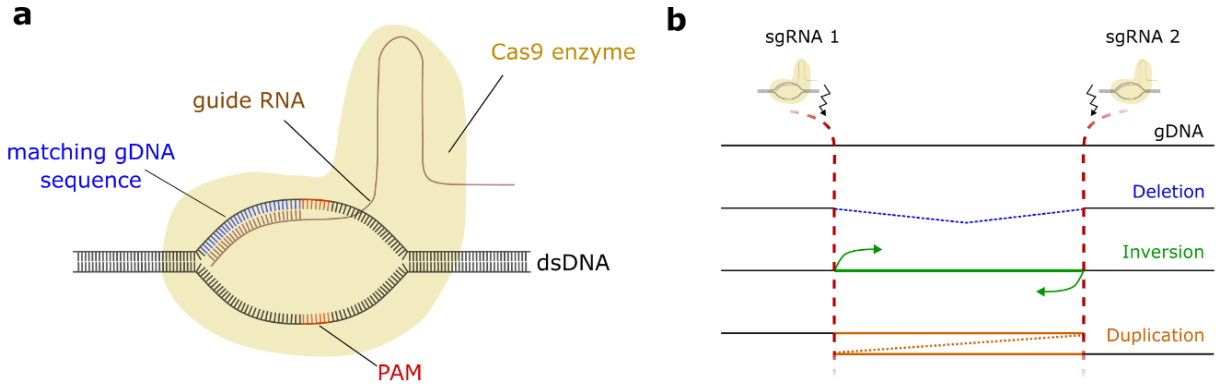


Figure 1.6 CRISPR/Cas9 technology in genome editing

a, CRISPR/Cas9 endonuclease and single guide RNA (sgRNA) genome editing complex assembly at the side of guide RNA complementary genomic DNA. yellow: Cas9-endonuclease, orange: guide

RNA, red: PAM-sequence, blue: sgRNA complementary genomic DNA sequence **b,** Schematic overview of “two-sgRNA-targeted” mutations. At the position, guided through the complementary sgRNA, both Cas9 nucleases induce a double strand break (DSB) which can create large genomic variants. The sequence between the sgRNAs can be deleted (blue), inverted (green) or duplicated and reinserted in tandem (orange). The figure was inspired by (Kraft et al. 2015; Jinek et al. 2012).

In the ADLD project part of this work, we want to utilize CRISPR/Cas9 genome editing to introduce structural variants at the locus of *Lmnb1*/*Zfp608* into mouse embryonic stem cells (mESCs). By generating new mouse lines, we want to investigate changes in the genomic region that result in the neurodegenerative disorder autosomal dominant leukodystrophy (ADLD).

1.3.2 Current research and mouse models of ADLD

For a few reasons it remains complex to research neurodegenerative disorders such as ADLD and to find the link between de novo mutations and the extremely diverse set of cell types. For one, there are regular delays in diagnosis, often by months to years after the first notable symptoms appear. Second, molecular and cellular changes caused by the variant emerge ahead of the clinical symptoms, which are often irreversible at the stage of detection (Aires et al. 2019; Gwathmey et al. 2023).

The search for biomarkers and early, non-invasive diagnosis is of high priority as the rise in neurological disorders rank leading in cause for disability and second in deaths worldwide with numbers increasing due to aging populations (GBD 2016 Neurology Collaborators 2019; C. Ding et al. 2022; Feigin et al. 2020).

Autosomal dominant leukodystrophy is a rare, fatal, late onset disease classically known to be caused by a duplication spanning the gene *Lmnb1*, encoding for the nuclear envelope protein Lamin-B1 (Giorgio et al. 2013; Quasar S. Padiath et al. 2006; Molloy et al. 2012; Brussino et al. 2009; Sundblom et al. 2009; Meijer et al. 2008; Schuster et al. 2011). *Lmnb1* expression is vital in various biological processes, such as the nuclear envelope stability, heterochromatin establishment, cellular senescence, proliferation and mitotic spindle organization (Camps, Erdos, and Ried 2015; Freund et al. 2012; Shimi et al. 2011; Cristofoli et al. 2020; Kaneshiro, Capitanio, and Hetzer 2023). Studies on *Lmnb1* knockout mice have elucidated its function in tissues such as the lung, bone and fibroblasts of the developing embryo (Vergnes et al. 2004).

Additionally, *Lmnb1* expression is distinctively important in the development of the brain, specifically in the nuclear shape, differentiation and migration of neurons, dendrite development, and the organization of cortical layers. Mouse models carrying a knockout of *Lmnb1* present smaller numbers of neuron progenitors, deformed cortical layers and reduced

brain size (Giacomini et al. 2016; Coffinier et al. 2011; Kim et al. 2011; Koufi et al. 2023). Further, *Lmnb1* expression plays a role in glia cell differentiation. Glia are an important population of non-neuronal cells in the brain with a plethora of functions like neuronal support, protection myelination and nutrient supply (Yattah et al. 2020; Mahajani et al. 2017).
Mouse models overexpressing *Lmnb1*, which are designed to mimic the gene duplication found in patients, have shown signs of motor dysfunction, cognitive defects and epilepsy (Heng et al. 2013). Regarding the disease causing cell types, two opposing theories have been proposed, stating the exclusive roles of either of the two glia cell types, namely I. oligodendrocytes specific defects in lipid and II. cholesterol pathways and reactive astrocytes which depict increase in immune related pathways. These results are underlining the importance of further investigations of the exact mechanisms involved in the disease development associated with *Lmnb1* overexpression (Ratti, Rusciano, Mongiorgi, Owusu Obeng, et al. 2021; Rolyan et al. 2015; Ratti, Rusciano, Mongiorgi, Neri, et al. 2021).

1.3.3 The 3D locus of *Lmnb1*

Several patients with ADLD have been found to carry a heterozygous duplication spanning the *Lmnb1* gene, resulting in the overexpression of *Lmnb1*, which is considered the marker gene for the ADLD (**Fig. 1.7, blue boxes**)(Giorgio et al. 2013; Quasar S. Padiath et al. 2006; Molloy et al. 2012; Brussino et al. 2010; Sundblom et al. 2009; Schuster et al. 2011; Meijer et al. 2008). When patients were described with increased *Lmnb1* expression and symptoms of ADLD without alterations at the gene body of *Lmnb1*, clinicians were discussing different mechanisms of disease development (Brussino et al. 2010). Recently, these patients were found to carry a deletion upstream of *Lmnb1* (**Fig. 1.7, red boxes**) (Giorgio et al. 2015; Mezaki et al. 2018; Nmezi et al. 2019; Borja et al. 2022). Because these deletions are not encompassing *Lmnb1* directly but a non-coding region upstream of the gene, one study hypothesizes that the mechanism of disease development in the deletion variants is caused by the disruption of a TAD boundary and as a consequence hijacking of enhancers, ectopically upregulating *Lmnb1* expression (Giorgio et al. 2015).
Capture HiC (cHiC) of the mouse embryonic brain *Lmnb1* region at embryonic stage E11.5 presents two distinct TADs. *Lmnb1* is positioned in a TAD (~1.1Mb) containing several other genes. This TAD shows distinct looping, characteristic for gene activity at the locus in E11.5 mouse brain development (**Fig. 1.7**). The neighboring, bigger TAD (~2.3Mb) is comparatively empty with a single gene, *Zfp608,* positioned in the center known to be involved in neurogenesis, forming boundaries upon expression in neuronal differentiation (Bonev et al. 2017; Ayoub et al. 2011; van den Berg et al. 2017). Supportive for its role in neurogenesis,

multiple VISTA enhancers were identified with specific and/or exclusive activity in the developing brain (**Fig. 1.7**) (Visel et al. 2007).

When converting some known patient mutations into the mouse genome, the duplications all overlap at the position of *Lmnb1* as the common denominator region (**Fig. 1.7, dotted lines, Supplementary Figure 7.2**). The deletion variants do not overlap with the gene but are visually spanning into the neighboring *Zfp608*-TAD, supportive of the hypothesis of an inter-TAD structural variant with potential enhancer hijacking. These hypotheses have to be experimentally validated as a step towards understanding the role of the two mutations in the development of ADLD.

Of note, the genes positioned between *Lmnb1* and *Zfp608* (*Gramd3*, *Aldh7a1* and *Phax*) are not linked to any function within neurogenesis and therefore are not considered relevant to the disease mechanism as discussed in: (Giorgio et al. 2015).

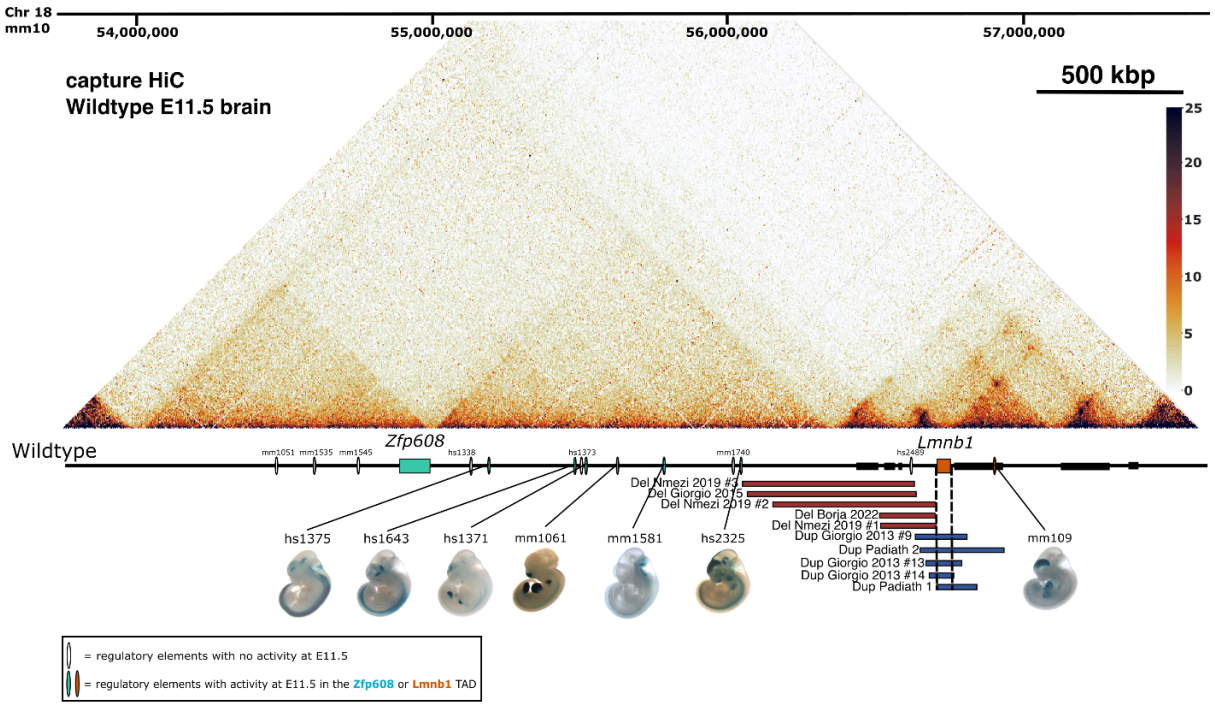


Figure 1.7 Structural Variants from patients in the *Lmnb1*/*Zfp608* mouse brain cHiC

Capture HiC (cHiC) of wildtype E11.5 mouse brains from the genetic region of the *Lmnb1* gene. The contact matrix visualizes two distinct topologically associated domains (TADs), the left which contains *Zfp608* (blue) and the right which contains *Lmnb1* (orange) and some unmarked genes (black blocks in the order of the genes: *Gramd3*, *Aldh7a1*, *Phax*, *March3*, *Megf10*, *Prrc1*). Enhancers that are active at mouse embryonic stage E11.5, sourced from VISTA enhancer browser, are displayed with LacZ reporter mice showing their tissue of activity. Human mutations were converted to the mouse genome (mm10) using UCSC LiftOver. Patient mutations (deletions= red; duplications= blue) are sourced from following publications: (Borja et al. 2022; Nmezi et al. 2019; Giorgio et al. 2015, 2013; Quasar S.

Padiath et al. 2006). Only a few duplications are displayed, a fuller representation is shown in **Supplementary Fig. 7.2**.

1.3.4 ADLD pathology

ADLD was first mistakenly described as an adult onset variant of Pelizaeus-Merzbacher disease (Zerbin-Rüdin and Peiffer 1964). To this date 70 cases of ADLD are described and it is considered a rare disease, although likely underdiagnosed (Finnsson et al. 2015; Raininko, Gosky, and Padiath 2016). ADLD is a fatal, slowly progressive degenerative neurological disorder with no known cure to date (Lin, Ptácek, and Fu 2011; Nmezi et al. 2020). The age of onset in patients varies with symptoms starting on average around the age of 40-50 years (Lin, Ptácek, and Fu 2011). In the majority of cases, patients first display autonomic symptoms, like bladder and bowel dysfunctions, followed by the development of motor, cerebellar and pyramidal symptoms such as spacities and tremors (Quasar S. Padiath 2019; Raininko, Gosky, and Padiath 2016). In the late stages of the disease progression some cases display cognitive impairment, including the development of dementia and psychiatric symptoms (Terlizzi et al. 2016; Raininko, Gosky, and Padiath 2016; Finnsson et al. 2015). Notably, patients carrying the deletion variant were characterized by slight alterations in symptoms including an absence of the autonomic signs, particularly at disease onset (Brussino et al. 2010; Giorgio et al. 2015).

Typically, ADLD is primarily diagnosed and monitored using non-invasive magnetic resonance imaging (MRI) (Sundblom et al. 2009; Melberg et al. 2006; Schuster et al. 2011; Brussino et al. 2009). Additionally, patient fibroblasts are sometimes screened for expression levels of the biomarker *Lmnb1* (Schuster et al. 2011; Quasar S. Padiath et al. 2006). MRI of duplication patients have previously shown white matter hyperintensities within the cortex and the cerebellum (**Fig. 1.8a**) (Finnsson et al. 2015; Brussino et al. 2010; Schuster et al. 2011; Melberg et al. 2006). Strikingly, a patient carrying a deletion variant displayed increased involvement of the cortex white matter inflammation, while the cerebellum was less inflamed, compared to a duplication variant carrying patient (**Fig. 1.8a, red arrows**) (Giorgio et al. 2015; Brussino et al. 2010). These results hint at the role of the cerebellum in the autonomic symptoms of the duplication patients and a stronger involvement of the cortical inflammation within the deletion variants disease development.

Taken together, there is an established connection between the over/misexpression of *Lmnb1* and the onset of the slowly progressive neurodegenerative disorder ADLD in patients, which is caused by different variants close to or encompassing the gene.

Mouse models have been subjected to different experiments to investigate the effects of *Lmnb1* overexpression. In several studies, Plp1-FLAG mouse models were generated to

drive overexpression of *Lmnb1* exclusively in oligodendrocytes (Heng et al. 2013; Rolyan et al. 2015; Lo Martire et al. 2018). One study reported an onset of motor dysfunctions in the mouse model with walking deficits starting at 40 weeks of age and a shortened lifespan. In addition, the mouse model displayed degenerating axons and aberrant myelination in the brain (**Fig. 1.8b**) (Heng et al. 2013). A similar model also reported motor dysfunction and the demyelination of spinal cord white matter, for which the researchers established a link to a defect in lipid synthesis, specifically in the oligodendrocytes (**Fig. 1.8c**) (Rolyan et al. 2015). Using the same mouse model as (Rolyan et al. 2015), another study detected early onset motor changes in 29 week old mice. Interestingly, although some patient related phenotypes were detected, others like cardiac autonomic dysfunction, are missing in oligodendrocyte targeted mouse models, suggesting that oligodendrocytes are not the exclusive drivers of the disease (Lo Martire et al. 2018).

Taken together, these mouse models recapitulate aspects of the disease and present convincing results towards deregulated myelination as a causative factor in ADLD. It has previously been reported that *Lmnb1* expression is an important regulator of cholesterol and lipid biosynthesis, implying a promising but not exclusive role of oligodendrocytes in driving disease progression of ADLD (Yattah et al. 2020; Quasar S. Padiath 2016).

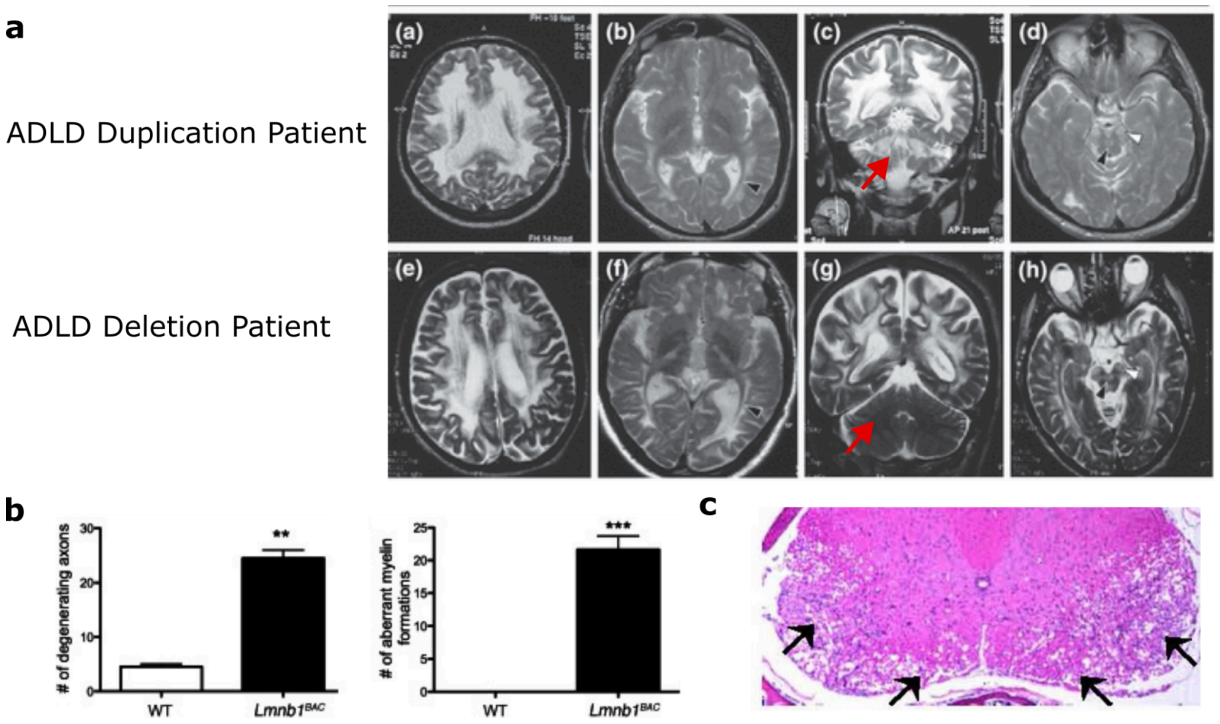


Figure 1.8 ADLD patient MRI and phenotype of mouse *Lmnb1* overexpression a, Brain MRI comparison of white matter inflammation between a patient carrying a duplication spanning the *Lmnb1* gene (a-d) and a patient carrying a deletion (e-h). Panel c and g depictMRI scans of the patients in the

coronal plane with a distinct involvement of the cerebellum in the white matter inflammation in the duplication patient (c, red arrow) but not the deletion patient (g,red arrow)**. b,** In a *Lmnb1*-BAC mouse model, the overexpression of *Lmnb1* resulted in significant increase of degenerating neuronal axons and aberrant myelination of neuron fiber, quantified in the axons of 12 month old mutant and wildtype mice (n=3 per condition) with the values expressed as means ± SEM. **P < 0.01. **c,** A mouse model which specifically overexpressed *Lmnb1* in oligodendrocytes displayed degeneration of myelination (arrows) in H&E stainings of the spinal cord of the transgenic mice. The figures were adapted from (Brussino et al. 2010; Rolyan et al. 2015; Heng et al. 2013).

1.4 Phenotyping methods to decipher mouse models

1.4.1. Conventional methods

Mouse models that are created to recapitulate aspects of human diseases are investigated using various methods. Depending on the question about the disease and previous knowledge about the regulatory elements or genes involved, different aspects are highlighted, often elucidating only a fraction of the complex phenotypes. For example, disorders affecting the brain like inflammation in neurodegenerative disorders are accessible using MRI (Kurniawan 2018). Molecular methods like immunofluorescence imaging or RNAscope have been implemented to visualize tissue malformations, visualize gene expression in certain tissues and quantify fluorescence signals of marker genes to track molecular changes and connect them to alterations in cell compositions in a defined tissue (Malinverno et al. 2019; Dobson et al. 2016). Malformation can be investigated applying hematoxylin-eosin (H&E) stainings when looking into histological malformations of soft tissues and by skeletal preparations to comprehend cartilage and skeletal deformations as in the example of the rare skeletal phenotypes (McDonald et al. 2011; Despang et al. 2019). Nevertheless, the full examination of a mouse mutant is laborious and cost intensive, and the comparability between mutants from different studies often falls short due to difference in experimental set-up or analysis of the mutants as in the example of the Plp1-FLAG mice discussed above. Even when using the same mouse strain, slightly different results may come out depending on the focus of the research (Rolyan et al. 2015; Lo Martire et al. 2018). One strategy to circumvent this issue are mouse clinics, which apply a battery of standardized tests to investigate mouse mutants in order to keep the analysis of phenotypes as standardized as possible (Mossbrugger et al. 2007; Gailus-Durner et al. 2005). Although proven powerful, shortcomings of this approach remain as not everyone has access to mouse clinics, and the bandwidth of experiments which would be needed to analyze every

phenotypic aspect of a mouse mutant is often not available. Therefore there is a need for standardized comparative methods, which allow high throughput of mutant phenotyping.

1.4.2 scRNA-seq as a new technology in disease phenotyping

Over the last few years single-cell sequencing methods such as single cell RNA sequencing (scRNA-seq) have been increasingly applied in the field of developmental and disease research. While the predecessor, bulk RNA sequencing, has offered evaluation of the overall transcriptomic differences of a tissue between healthy and diseased samples, the adaptation to profile the transcriptome of single cells opens the opportunities to investigate granular alterations in a diseased tissue as well as cell type specific changes.

The success of this method has been demonstrated in developmental studies, where the transcriptomic single cell early embryonic mouse development was captured in just four studies and summarized in Qui et al. (Cao et al. 2019; Mohammed et al. 2017; Pijuan-Sala et al. 2019; Mittnenzweig et al. 2021; Qiu et al. 2022). Due to accomplishments like these, single cell technologies have been termed “breakthrough of the year” in 2018 (Science -2018). Recent adaptation of the technology in disease research has shown promising results like in the example of Tirosh et al., who found two malignant cell groups expressing different types of marker genes suggestive for treatment strategies in human melanomas (Tirosh et al. 2016). Since then, the method has been applied in various case studies such as heart diseases, Lupus, Multiple Sclerosis, Alzheimers and Parkinsons, just to name a few (Tirosh et al. 2016; Samad and Wu 2021; Der et al. 2017; Schafflick et al. 2020; Jiang et al. 2020; Smajić et al. 2021).

Of note, several other methods capturing different aspects of single cell identity, like ATAC sequencing, spatial transcriptomics and multi-omic approaches have been successfully established (Cusanovich et al. 2015, 2018; Ståhl et al. 2016; Rodriques et al. 2019; Han et al. 2014; S. Chen, Lake, and Zhang 2019).

To prepare the samples for scRNA-seq, the tissue of interest is extracted, dissociated into cells and optionally (depending on the method) lysed into a suspension of nuclei (**Fig. 1.9**). The extraction of nuclei present advantageous in scenarios where the integrity of the cell membrane might not be guaranteed, for example when the tissue was flash-frozen or if cells are entangled and require physical force to dissociate fully. In general, using nuclei from a heterogeneous tissue allows for a more uniform dissociation of the sample. The sample is made subject to a barcoding strategy, which is the most variable step between the different methods. This step is critical as it assigns a unique barcode to every captured cell, important for connecting the captured transcripts to their cellular identity later in analysis. The main barcoding strategies are droplet or split and pool based (**Fig 1.9**). In brief, the droplet method

is based on the combination of beads and single cells in a specific dilution which obtains one bead to combine with one cell/nucleus in an emulsion droplet (A. M. Klein et al. 2015; Bageritz and Raddi 2019). The split and pool technique, also known as combinatorial indexing, is based on a series of nuclei distribution into wells in which adaptors are being attached, followed by the sample being pooled and distributed again for several rounds of indexing (Cao et al. 2019; Martin et al. 2022).

In order to analyze these datasets, a framework of different analysis tools, such as Seurat, Monocle, Scanpy, Cellranger and many more, was developed, some discussed in this review: (Zappia and Theis 2021). As method development is progressing rapidly, various new technologies and bioinformatic tools will be coming out within the following years to further the analysis of big scale datasets.

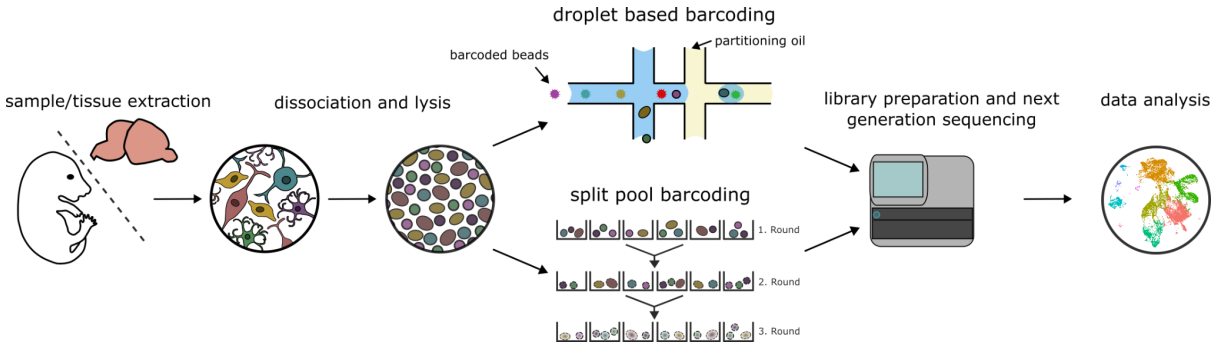


Figure 1.9 Overview of single cell dissection and barcoding methods of the mouse brain
In a generalized example single cell RNA tissue preparation protocol, the organ of interest or whole embryo is extracted, the tissue is dissociated into single cells and optionally lysed to a nuclei suspension. Nuclei are made subject to either a droplet or a split and pool based method to achieve individual barcoding of each cell. Droplet barcoding generally involves the concept of barcoded beads combined with single cells in a dilution, allowing one cell/nucleus to be captured with a single bead in partitioning oil. The split and pool barcoding method generally involves sequential barcoding rounds with samples being mixed and distributed in wells to achieve the highest library complexity. Barcoding is followed by library preparation and sequencing. After the sequence run, the data has to be analyzed with bioinformatic tools.

We want to apply scRNA-seq as a tool for phenotyping mouse mutants, once with the split and pool based method using multiple whole embryo mutant samples and once with a droplet based method to conduct an in-depth analysis of a neurodegenerative disorder using mouse embryonic brain samples from two embryonic stages.

2. Aim of this Study

To address the lack of phenotyping tools that allow large-scale analysis of disease development in mouse models, we hypothesize that the application of scRNA-seq is the logical next step. Our aim is to use this technology in two different settings to test the ability of the tool to phenotype mouse mutants.

In a first project, we will apply the split and pool based sci-RNA-seq3 method to multiple whole embryo mutant samples. Within a single experiment we aim to test the ability of the method to detect changes in mutants ranging from affected pleiotropic genes to small non-coding variants. To do this, we will build a framework of new analytical tools to perform comparative analysis and detect small changes in local gene expression. With this set-up, we aim to help establish single cell technologies as a standard tool for mutant phenotyping and provide a database for further exploration by the scientific community.

In a second project, we will apply the 10x droplet based method in the context of an in-depth analysis of the neurodegenerative disorder ADLD by generating mouse models of two different patient-derived variants leading to the disorder and studying embryonic brain samples. We aim to combine different technologies, namely scRNA-seq and cHiC, to study embryonic samples, as well as MRI in adult mice to dissect the development of the late-onset neurodegenerative disorder.

Our first goal is to decipher how the deletion variant causes the upregulation of *Lmnb1* in the context of ADLD-related disease development. The current working hypothesis is that the deletion spans the TAD boundary to the neighboring gene *Zfp608*, resulting in neighboring regulatory elements ectopically contacting *Lmnb1* and leading to its misexpression. We will extend this hypothesis by generating cHiC maps of the deletion and duplication mutants to compare changes in 3D chromatin organization between the mutants. Further, we also want to challenge the current state of knowledge on the onset of neurodegenerative diseases. Because the clinical symptoms of patients with neurodegenerative disorders develop late in life, these disorders are defined as late-onset. However, we hypothesize that the onset of these disorders may occur at earlier time points, as clinical symptoms are preceded by molecular and cellular changes. To test this hypothesis, we aim to generate single-cell datasets of the developing embryonic mouse brain and analyze potential early-onset changes associated with the disease. In addition, we will analyze potential differences in disease development between the two variants associated with the different mutations. To detect the manifestation of morphological changes, we aim to validate these newly generated mouse models with MRI experiments in the adult mice.

3. Materials

3.1 Antibodies, Enzymes & Probes

Table 3.1 Antibodies, Enzymes and Probes used in the ADLD and MMCA Project

Name	Application	Source
Proteinase K	protein degradation	In-House preparation (by Asita C. Stiege)
T4 Ligase	guide ligation	Thermo Scientific (#EK0032)
DnpII	Restriction Enzyme in cHiC digest	New England BioLabs
Taq Polymerase	Genotyping Polymerase for PCR amplification	In-House preparation (by Asita C. Stiege)
SUPERase In	RNase Inhibitor for single cell RNA preparation	Thermo Scientific (#AM2694)
Mm-Kcnj2	RNAscope™ Probe	Advanced Cell Diagnostics (#476261)
Mm-Sox9-C2	RNAscope™ Probe	Advanced Cell Diagnostics (#401051-C2)
Anti-Pax6 rabbit polyclonal antibody	Immunofluorescence Antibody	Merck-Sigma (#AB2237)
Anti-Prealbumin rabbit monoclonal antibody	Immunofluorescence Antibody	Abcam (#EPR20971)
Goat Anti-Rabbit Alexa Fluor 488-conjugated secondary antibody	Immunofluorescence Antibody	Leica (#A-11008)

3.2 Bacterial strains & Plasmids

For transformation and multiplication of plasmids the chemically competent Escherichia coli (E.Coli) TOP10 strain was used (prepared in-house by Asita C. Stiege), as a plasmid for cloning single guide RNA's px459-pSpCas9(BB)-2A-Puro (Addgene, #62988). For long term storage, glycine stock of transfected bacteria were stored at -80°C.

3.3 Buffers & Solutions

Table 3.2 standard solutions used in this study

Solution	Composition
Lysis Buffer	17 mM Tris(pH 7.5), 17mM EDTA, 170 mM NaCl, 0.85 % SDS, added freshly

	0.08 µg/ µl Proteinase K
DEPC-H_2O	0.1% Diethylpyrocarbonat (DEPC) in ddH_2O
10x PBS-DEPC	1.37 M NaCl, 27 mM KCl, 100 mM Na2HPO4, 20 mM KH2PO4 in DEPC-H_2O, adjust pH to 7.4 with HCl & autoclave
4%PFA	Dissolve 40 mg/ml PFA in 1x PBS-DEPC,dissolve PFA by heating the solution to 55°C, adjust pH to 7.4 with HCl

Table 3.3 Buffers and Solutions for 10x based single cell nuclei preparation

Buffer	Composition
Cell lysis buffer (CSB)	10 mM Tris-HCl(pH 7.4), 10 mM NaCl, 3mM MgCl2, 2% BSA (NEB), 0.1% IGEPAL CA-630 and 1%SUPERase In RNase Inhibitor (20 U/µL, Ambion) in DEPC-H_2O
Nuclei suspension buffer (NSB)	10 mM Tris-HCl(pH 7.4), 10 mM NaCl, 3mM MgCl2, 2% BSA (NEB) and 1%SUPERase In RNase Inhibitor (20 U/µL, Ambion) in DEPC-H_2O

Table 3.4 Buffers and Solutions for sci RNA seq based single cell nuclei preparation

Buffer	Composition
Cell lysis buffer (CSB)	10 mM Tris-HCl(pH 7.4), 10 mM NaCl, 3 mM MgCl2, 2% BSA (NEB), 0.1% IGEPAL CA-630 and 1%SUPERase In RNase Inhibitor (20 U/µL, Ambion) in DEPC-H_2O
Nuclei suspension buffer (NSB)	10 mM Tris-HCl(pH 7.4), 10 mM NaCl, 3 mM MgCl2, 2% BSA (NEB) and 1%SUPERase In RNase Inhibitor (20 U/µL, Ambion) in DEPC-H_2O
Nuclei dilution buffer	10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM MgCl2, 1% SUPERase In and 1% BSA

Table 3.5 Buffers and Solutions for cHiC sample and library preparation

Buffer/Solution	Composition
HiC lysis buffer	250 µl 5 M Tris (pH 8.0), 150 µl 5M NaCl, 50 µl 0.5M EDTA, 250 µl 10% NP-40, 575 µl 10% TritonX-100, 200 µl 25 X PI (protease inhibitor (Roche)) in 3.525 ml H_2O
Digestion Mix	146.4 µl H_2O, 23.6 µl 10X DpnII buffer, 5 µl 10% SDS

3.4 Chemicals

If not stated otherwise, all chemicals are obtained from Merck, Roth, Sigma Aldrich or Thermo and used in analytical grade quality.

3.5 Instruments

Table 3.6 Instruments used in this study

Name	Type	Source
Table Top centrifuge	5414D	Eppendorf
Cooling centrifuge	5417R	Eppendorf
Cooling centrifuge	Avanti J-E	Beckman-Coulter
QuantStudio 7 Flex Real-Time PCR System		Applied Biosystems
Thermocycler	GeneAmp PCR System 9700 & SimpliAmp™	Applied Biosystems
Stereomicroscope	MZ12 Discovery V12	Zeiss
Camera	DFC420	Leica
Light Source	KL1500 LCD	Leica
Nanodrop 2000 Spektralphotometer	ND-2000	Thermo Scientific
HybEZ™II Hybridization System (220V)	321720	Advanced Cell Diagnostics
Sterile Bench, HERASafe		Thermo Scientific
Heracell CO2 Incubator		Thermo Scientific
LSM 980 Airyscan 2		Carl Zeiss AG
NanoZoomer 2.0 HT		Hamamatsu
Chromium Controller		10x Genomics
E220 Focused- Ultrasonicator		Covaris

3.6 Kits

Table 3.7 Kits used in this study

Name	Application	Source
NucleoSpin Plasmid Easy Pure	Plasmid clean up	Macherey&Nagel (#740727.10)
RNAscope® Multiplex Fluorescent Reagent Kit v2	RNAscope in situ Hybridization	Advanced Cell Diagnostics (#323100)

Chromium Next GEM Single Cell 3' GEM, Library & Gel Bead Kit v3.1	10x single cell library preparation	10x Genomics(#PN-1000121)
Chromium Next GEM Chip G Single Cell Kit	10x single cell library preparation	10x Genomics(#PN-1000127)
Single Index Kit T Set A, 96 rxns	10x single cell library preparation	10x Genomics(#PN-1000213)

3.7 Media

Table 3.8 Media used in this study

Medium	Content
mESC medium	Knockout DMEM (Gibco,#10829-018) containing 15% FCS (PANSera ES, #P30-2600), 10 mM Glutamine (100x Lonza,#BE17-605E), 1x penicillin/streptomycin (Lonza, #DE17-603), 1x non-essential amino acids(100x Gibco, #11140-35), 1x nucleosides (100x, Chemicon, ES-008D), 0.1 mM beta-Mercaptoethanol (Gibco, #3150-010)
mESC medium +LIF	ESC Medium with 1000 U/ml Leukemia Inhibitory Factor (LIF) (Chemicon, #ESG1107)
freezing medium	Bicarbonate free DMEM (Gibco, #52100) with 10 mM HEPES pH 7.2, 20% FCS (PANSera ES, #P30-2600) and 10% DMSO (Sigma, #D-2650)

3.8 Oligonucleotides

All genotyping primers, plasmid primers, for the ADLD Project are synthesized by Eurofins Genomics and displayed 5'-3' orientation and are listed below in **Table 3.9**.

Table 3.9 Oligonucleotides used in the ADLD project

Purpose/Name	Sequence
Genotyping Primers	
5_PCR_Del1_F1_Gio15	TGACATCAATGGGAGGAAGG
6_PCR_Del1_R1_Gio15	ACTCAGGCAGGAGAACATGC
7_PCR_Del1_F2_Gio15	AAGACAAAAGGGAGACAGGTTG
8_PCR_Del1_R2_Gio15	ACCAAGGGACCACTGACTTG
9_PCR_Dup9_F1_Gio13	GCATGATCTAACCACTATAAACATCC
10_PCR_Dup9_R1_Gio13	AGGTACCTGCCCACATCTTG
12_PCR_Dup9_R2_Gio13	TGTGAGGGAGTGGCTTTTTC
53_PCR_Dup9_F2_Gio13	AAATCATTCATCACCACTAAGAAC
Plasmid/Vector primers	

KolR	CACGCGCTAAAAACGGACTA
U6	ACTATCATATGCTTACCGTAA
Sp6_fwd	ATTTAGGTGACACTATAG
T7_fwd	TAATACGACTCACTATAGGG
CRISPR/Cas9 Guides	
1_guide_Del1_Gio15_cen_fwd	CACCGACGCAGGCTGAGTTGTAAAA
2_guide_Del1_Gio15_cen_rev	AAACTTTTACAACTCAGCCTGCGTC
3_guide_Del1_Gio15_tel_fwd	CACCGAGAGAGAGTCAAGGGTCAAA
4_guide_Del1_Gio15_tel_rev	AAACTTTGACCCTTGACTCTCTCTC
5_guide_Dup1_Gio13_cen_fwd	CACCGTGACACTAGGAAGAAACTGG
6_guide_Dup1_Gio13_cen_rev	AAACCCAGTTTCTTCCTAGTGTCAC
7_guide_Dup1_Gio13_tel_fwd	CACCGTGGAGAGCAGGGGAGTAAAC
8_guide_Dup1_Gio13_tel_rev	AAACGTTTACTCCCCTGCTCTCCAC

3.9 Mouse lines

Mouse lines collected for the MMCA Project are listed in Supplementary table 7.3.2.
Mouse lines for the ADLD Project using the CRISPR/Cas9 genome editing system are created in accordance with the guidelines of the german government regulations of the “Landesamt für Gesundheit und Soziales” (G0243/18) both were established as lines (G0176/19).

Table 3.10 Transgenic mouse lines generated in the ADLD project

Transgenic mouse line	Alias	Patient	Genomic region (mm10)
Del_Lmnb1	Deletion	Del1_Giorgio_2015	chr18:56066950-56636813
Dup_Lmnb1	Duplication	Dup9_Giorgio_2013	chr18:56632299-56809381

3.10 Software, Tools & Programming Language

Table 3.11 Software, Tools & Programming Languages used in the ADLD & MMCA project

Name	Function	Source
Benchling	CRISPR Guide design	https://www.benchling.com/
Primer3	Primer design	https://primer3.ut.ee/
NCBI primer blast	Primer design	https://www.ncbi.nlm.nih.gov/tools/primer-blast/
Netprimer	Primer design	https://www.premierbiosoft.com/netprimer/

Inkscape	Figure design	https://inkscape.org/de/
Adobe Illustrator	Figure design	https://www.adobe.com/
Fiji 64	Imaging Software	https://imagej.net/
NDP.view2	Imaging Software	https://www.hamamatsu.com/
Definiens Developer XD2	Image Analysis Software	Definiens AG
R/R-Studio	Programming Language	
Python	Programming Language	
Fan-C	Insulation Score Analysis	https://github.com/vaquerizaslab/fanc
Seurat	Single cell analysis toolkit	https://satijalab.org/seurat/ (Butler et al. 2018)
Monocle	Single cell analysis toolkit	https://cole-trapnell-lab.github.io/monocle3/ (Cao et al. 2019)
Scrublet	Doublet removal tool	https://github.com/swolock/scrublet
scVelo	Single cell velocity tool	https://github.com/theislab/scvelo
ggplot2	graphics tool	https://ggplot2.tidyverse.org/
UCSC Genome Browser	Genome Browser	https://genome.ucsc.edu/

4. Methods

Particular parts of the methods are taken from the manuscript of Huang, Henck & Xie et al 2023 and are indicated as such. When an experiment/analysis was carried out in collaboration or with noteworthy participation that is described at the top of the respective section as well.

4.1 Standard methods

All standardized molecular methods (including: Agarose Gel Electrophoresis, DNA cloning, transformation of competent cells (E.Coli) and Polymerase Chain Reaction (PCR)) are performed according to protocols described in Green and Sambrook(Green and Sambrook, 2012).

4.1.1 Designing & Cloning of single RNA guides for CRISPR/Cas9

For designing the right guides surrounding the lifted patient identified mutations, a region close to the mutation was selected and inserted into the Benchling sgRNA design tool. Using the tool, guide sequences with NGG as protospacer adjacent motif (PAM), 20 nucleotides length and px459 vector specific restriction recognition sites as overhangs were designed. Only oligonucleotides with high off target (min: above 80) and high on target scores were selected to achieve minimum off target effects. For both, the duplication and the deletion guides were designed for each cutting site of the target region.The complementary oligonucleotides were annealed, ligated into the BbsI linearized vector using T4 Ligase and transformed into Top10 bacteria. Bacteria were stored for further use as glycerin stocks in -80°C.
Plasmid DNA was isolated and purified using the NucleoSpin Plasmid Easy Pure Kit according to the manufacturer's guidelines and the concentration of the plasmid was measured for transfection of mouse embryonic stem cells (mESC's). Successful integration of the sgRNA's was confirmed with Sanger sequencing (Primers: KolR/U6).

4.2 Embryonic stem cell culture & mouse line generation

4.2.1 Preparation and seeding of feeder cells

Feeder cells, both CD1 and DR4 (carries resistance to puromycin, hygromycin and geneticin) essential for the survival of mouse embryonic stem cells (mESC's) were created in house by Asita C. Stiege. Briefly, mouse embryonic fibroblasts (MEF's) are extracted from DR4 and CD1 mouse embryos respectively. They are expanded until passage 5, mitotically inactivated using Mitomycin C and stored in -80°C or liquid nitrogen (LN) for long term storage with $2.5x10^6$ cells per cryovial. Feeder cells are seeded the day before seeding the mESC's on gelatinized plates or wells (37°C for 15 minutes, in 7.5% CO2, 0.01% gelatin) so they can build a monolayer to support the growth and pluripotency of the mESC's.

4.2.2 Handling of mouse embryonic stem cells

All mESC's used in the ADLD-Project are male G4 stem cells (crossing of 129S6/SvEvTac x C57BL/6Ncr mouse strains). Minimum 12 hours after plating feeder cells, mESC's were plated in ESC-medium + LIF onto the feeder monolayer. The medium was changed every 24 hours and mESC's were split onto new feeder plates every 2-3 days, depending on growth rate and experimental conditions. The cells were frozen in mESC medium + 20% FCS and 20% DMSO in a concentration of $1\text{-}1.2x10^6$ cells per cryovial in LN.

4.2.3 CRISPR/Cas9 genome editing of mESC's for generating mutant clones

The mESC's were plated in the concentration of $3.5\text{-}4 \times 10^6$ cells/plate onto three gelatinized 6 cm feeder plates one day prior to transfection. On the day of transfection, the medium was changed to mESC medium - antibiotics (penicillin/streptomycin) + LIF (2 ml/plate). For each transfection of 3x 6 cm plates, a DNA Mix (4 µg each guide and 125 µl OptiMEM (Gibco, #51985-026)) and a FuGene Mix (25 µl FuGENE HD agent (Promega,#E2311) into 100 µl OptiMEM) were pipetted. Both mixes were combined, vortexed thoroughly and incubated at RT for 15 minutes before added dropwise to the cells. After 16 hours the medium was changed back to mESC-medium+ LIF and cells were split 3-24 hours after, onto 1:3 onto puromycin-resistant DR4 feeder plates for selection in mESC medium +LIF + 2 µg/ml puromycin (Sigma-Aldrich, #P8833). The cells are under selection for 48 hours until medium is changed to mESC medium + LIF to allow recovery for 2-3 days until clones are grown sufficiently for picking. Single mESC colonies were collected with a pipette tip with the help of

binocular microscopes. The respective colonies were transferred into U- 96 well plates containing 0.2% Trypsin-EDTA (Gibco, #25300-054) and incubated for 10 minutes at 37°C. Adding mESC medium +LIF clones were resuspended into single cells and transferred into gelatinized, CD1 feeder flat bottom 96 well plates. Depending on the mutation, 300-500 clones were picked per experiment. The mESC's remaining on the 6 cm plates were lysed, genomic DNA (gDNA) was extracted and used for primer testing.
The clones were cultured for another 2-3 days until confluent for split and freeze. For freezing, the clones were split into 3 plates: most of the clones were split into 2 replicate U-96 well plates containing freezing medium for freezing the clones at -80°C. The remaining cells in the flat bottom 96 well plate were cultured for gDNA extraction.

4.2.4 Extraction of genomic DNA

For genotyping, embryonic tissue, mouse ear clips or cells were incubated in 500 µl lysis buffer at 55°C overnight for tissue dissociation. 250 µl 5M NaCl were added and the samples were incubated another 30 minutes at 55°C while shaking (700 rpm) followed by an incubation of the sample on ice for 10 minutes. For removal of protein content, samples were spun down at least 30 minutes (9000 rpm, 4°C). The supernatant was mixed 1:2 with 100% ice cold ethanol and spun another time for 1-2 hours (maximum speed: >13000 rpm, 4°C). The supernatant was discarded and the precipitated DNA pellets were washed with 70% ethanol, centrifuged for 30 minutes (maximum speed: >13000 rpm, 4°C) and the cleaned pellets were dried at RT for 10 minutes followed by dissolving them in 30 - 300 µl water. Samples were stored in -20°C until used for genotyping PCR.

4.2.5 Design of Genotyping primers

For designing genotyping primers, the mouse genomic region surrounding the sgRNAs was selected in UCSC and pasted into Primer3, the suggested oligonucleotides generated were scanned for off target effects using the NCBI primer blast tool and checked in rating in the premier biosoft tool Netprimer. Oligonucleotides of 20-22 nucleotides length with a score above 85 that showed no off target effects in the size range of a standard PCR were considered for genotyping and tested on the DNA extracted from leftover CRISPR/Cas9 6 cm plates.

4.2.6 Selecting CRISPR/Cas9 edited clones by genotyping

DNA extracted from the 96 well plate clones was screened for successful genome editing. For genotyping, standard PCR using Taq polymerase was performed. The reaction was pipetted on ice as listed in **Table 4.1&4.2**).

Table 4.1 & 4.2 Chemicals and cycle numbers of the standard genotyping PCR

Chemical	**amount in µl**
Fermentas Buffer (10x)	2.5
Fwd/Rvs Primer (100 µM)	0.1
dNTP's (12 mM)	0.1
Taq Polymerase	0.5
Template (DNA)	1
H_2O	20.7

Temp (in °C)	**Time (in min)**	
96	5	
96	0.5	35 x
55	0.5	
72	1	
72	7	
4	∞	

After selecting candidate clones, one replicate U well plate was defrosted and clones were seeded and serially expanded until 1-1.2x10^6 cells could be frozen. Few cells each clone were left to grow feeder depleted to repeat genotyping as described above using standard PCR. Additionally, qPCR was performed with primers inside and outside the mutated region for copy number analysis. Briefly, for a standard qPCR reaction in a 384 well plate 6 µl SYBR Green Master Mix
(Applied Biosystems), 2 µl PrimerMix (1.5 µM each) and 4 µl gDNA (50 ng total) were pipette in triplicates for each primer/sample combination on ice. The efficiency of the qPCR primers was tested by standard dilution series of wildtype gDNA. qPCRs were run on the QuantStudio 7 Flex Real-Time PCR System. Copy number analysis was performed with the 2-ΔΔCt method comparing signals between the mutant and a wildtype control.

4.2.7 Aggregation and mouse line generation

For mutant embryo and adult mouse generation, diploid and tetraploid aggregations were performed. All animal studies in the ADLD Project were conducted under the license

numbers G0243/18 (generation) and G0176/19 (breeding) issued by the “Landesamt für Gesundheit und Soziales” (LaGeSo) Berlin.

The genotyped G4 clone was seeded two days prior in concentrations of 2 - 4x10^5. Female CD1 mice were used as fosters for the morula aggregation. For the establishment and maintenance of the mouse lines, the born genotyped mutants were crossed with wildtype C57Bl6/J mice. The mice were kept at the in house animal care facility in a controlled environment.

4.3 Single cell RNA library preparation for the ADLD Project

4.3.1 Brain extraction and nuclei isolation

The embryos were extracted at the respective time point of development (E11.5 and E18.5) and the brains were immediately extracted, flash frozen and stored in liquid nitrogen until used for sample preparation. Tailtips or other remaining tissues were used in DNA isolation for confirmation of correct genotypes.
In general, for each embryonic time point two independent replicates were created. Within one experimental run, deletion, duplication and wildtype samples were processed in parallel to avoid experiment based batch effects in the sample analysis. Every sample contains 3 brains pooled together to correct for minimal sampling time differences across the dataset.
For nuclei isolation, the protocol for sci-RNA-seq3 nuclei extraction was adapted from Cao *et al.* 2019 and described in a previous publication using the same adapted protocol as used in this thesis (Sreenivasan et al. 2023). In essence, the E18.5 brains were homogenized by cutting them using a sterile razor in 1 ml ice cold cell lysis buffer (CLB, prepared the same day), followed by thorough suspension adding 2 ml CLB to the sample on ice for proper tissue dissociation. For the E11.5 brain samples it was sufficient to resuspend them in 500 µl CLB to be dissociated. After straining the solution through 40µm cell strainers (Thermo Scientific, #229817) into Epis (for E11.5) or precoated 15 ml falcon tubes (for E18.5), the sample was resuspended again allowing full dissociation. After a centrifugation step for 5 minutes (4°C, 0.5 rcf), the supernatant was removed, 2 ml (E18.5) /500 µl (E11.5) nuclei suspension buffer (NSB, prepared the same day) was added and the pellet was dissolved through gentle resuspension. After another centrifugation (5 min, 4°C, 0.5rcf), the supernatant was removed and the pellets of the three biological samples were pooled in 4 ml / 2 ml NSB minimum. For confirmation of sufficient lysis and to determine the concentration of nuclei, an aliquot was mixed 1:1 with trypan blue and counted using a standard counting chamber (Neubauer). The samples were then diluted to a concentration of 1000 nuclei per µl

in a 20 µl aliquot, aiming for roughly 10.000 nuclei captured per sample and directly processed for sequencing.

4.3.2 Library preparation and sequencing

The samples were processed using the Chromium Next GEM single-cell 3' RNA seq protocol (V3.1, 10x Genomics) with dual index chemistry. As recommended in the 10x protocol, the quality of the cDNA and the final library were assessed using Qubit (Qubit dsDNA HS Assay Kit, #Q32854) and Agilent TapeStation (HighSensitivity D5000 ScreenTape, #5067-5592). Sequencing of the samples was performed in-house at the Max Planck Institute Sequencing facility on a NovaSeq 6000 system (Illumina) with a depth of 300 million reads per sample (asynchronous paired end sequencing according to 10x protocol).

4.4 10x scRNA-seq analysis

Bioinformatic analysis for the single cell data of the ADLD Project was done in cooperation with Varun K. A. Sreenivasan, and Saranya Balachandran from UKSH in Lübeck, who set up and adapted the sc-analysis pipeline (Institute of Human Genetics, University Medical Center Schleswig-Holstein, University of Lübeck & Kiel University, Lübeck, Germany).

4.4.1 Mapping of sequencing reads, filtering and quality control

Taken together, 2 replicates each of mutant (duplication, deletion) and wildtype at two developmental time points (E11.5 & E18.5) summed up to 12 individual conditions. Sequencing data was demultiplexed and converted into fastq files using bcl2fastq2 v2.20. These were mapped to the reference transcriptome mm10-2020-A (source: 10x) and processed to count-matrices using cellranger-5.0.1. Introns were included due to the spliced/unspliced ratio of nuclei used in sample preparation, which typically are enriched for unspliced reads. Quality reports and consecutive filtering was done using Seurat-4.0.5 (Hao et al. 2021) and scrublet-0.2.3 (Wolock, Lopez, and Klein 2019). To filter low quality reads and features the lower cut off was set at 400-600 features (transcripts) per cell and 700-800 UMIs (counts), the higher numbers for the two parameters exclusively referring to the second E11.5 replicate which generated fewer nuclei with higher overall counts. Nuclei with mitochondrial and ribosomal reads over 10% and nuclei with a doublet score over 15% were removed from the dataset. In total 128.511 nuclei were left for downstream analysis, 86.300 from E18.5 brain samples, 42.211 from E11.5 brain samples.

4.4.2 Integration, cell clustering and annotation

The 12 individual single-nuclei datasets (two repeats per condition each of the two developmental timepoints) were proceeded for normalization (function: NormalizeData) followed by integration of the main clusters was performed using the harmony-0.1.0 package. For clustering the following parameters were applied to the dataset nhvg=2500 (function:FindVariableFeatures), npcs=30 (function:RunPCA), nclust=30 (function:RunHarmony). Appling the Uniform Manifold Approximation and Projection (UMAP) algorithm(function:RunUMAP), the top 30 principal components were projected in 2D providing a dimensionality reduced visualization of the cells. To extract the marker genes to determine cluster celltype identities, the FindAllMarkers function was applied with parameters set to min.cells.group=5, min.pct=0.05, and logfc_threshold=0.2. The differentially expressed genes (DEG's) were ranked by applying the Wilcoxon rank-sum test. The clusters were annotated utilizing both the marker gene output of the analysis and known marker genes from literature (**Supplementary table 7.3.1**). Subclusters were integrated using Seurat-4.0.5.

4.4.3 Gene Set Enrichment Analysis (GSEA)

For Gene Set Enrichment Analysis (ssGSEA) the escape R-package was applied. For a first look into the data, enrichment of the hallmark gene sets (50 gene sets) was determined for the main clusters followed up by an in-depth enrichment analysis of interesting pathways from the C5 (ontology gene sets) in specific clusters only (functions:msigdbr, getGeneSets, with species set to "Mus musculus"). The gene sets available in the molecular signature database website https://www.gsea-msigdb.org/. The data was analyzed regarding these sampling groups: condition vs condition and condition_timepoint vs condition_timepoint respectively, and withdefault parameters. The respective median enrichment scores were compared between the sampling group using wilcoxon rank sum test and significance of the changes were determined using Wilcoxon rank sum test with the Bonferroni correction applied for multiple comparisons.

4.4.4 Cell composition analysis

The cell composition analysis to calculate the differential abundance of main and subclusters between the wildtype and the Dup/Del mutants per developmental time point was performed similarly to (Huang, Henck, Qui et al. 2023). Firstly, the cell numbers in each main/subcluster were normalized to the total number of cells per sample. To calculate the log2 fold change cellular abundance between each mutant and the wildtype, the mean of the normalized cell

numbers were calculated between repeats for each of the genotypes before calculating the fold change. To identify the significance of this change, beta-binomial regression was performed on the normalized cell numbers (while retaining the repeat information) using the VGAM package of R using the function vglm with the parameter “family = betabinomial”. The scripts are available at the github page https://github.com/SpielmannLab/henck_lmnb1.

4.4.5 RNA velocity

RNA velocity analysis was performed on the radial glia sub-trajectories, applying scVelo/v0.2.5 (https://github.com/theislab/scvelo). Briefly, using scv.pp.filter_and_normalize, the count matrices of the subcluster were filtered and normalized (parameters: min_shared_counts=20 and n_top_genes=2000). The means and variances between 30 nearest neighbors based on the PCA-embedding, imported from the seurat-based analysis described above (“integration cell clustering and annotation”), was then calculated using scv.pp.moments function. The velocities per condition (WT/Dup/Del) were calculated separately and projected onto the UMAP embedding of the respective cells per subcluster exported from the Seurat-based analysis. The scripts are available at the github page https://github.com/SpielmannLab/henck_lmnb1.

4.5 Capture HiC

The protocol for capture HiC library design was established by the Hughes and Oudelaar Lab (Oudelaar et al. 2017, 2020) and adapted and introduced by Mikie (Mai) Phan from the Mundlos Department at Max Planck Institute for Molecular Genetics. Further Mikie assisted in designing the locus specific capture probes.

4.5.1 Capture Probe design

In order to perform capture HiC (cHiC) to decipher the chromatin interaction between the *Lmnb1* and *Zfp608* locus, 100 nucleotide long double-stranded capture probes were designed spanning the region mm10 chr18:53,531,958-58,093,157. The custom probes were ordered from Twist Bioscience.

4.5.2 Sample fixation and lysis

For sample preparation three E11.5 brains were pooled per condition directly after extracting the tissue. For dissociation, 1 ml 0.2% Trypsin-EDTA (Gibco, #25300-054) was added and the sample was incubated at 37 °C for 10 minutes resuspending the sample every 2-3 minutes. To stop the trypsinization, 5 ml 10% FCS-PBS prewarmed was added and the sample was strained through 40 µm cell strainers (Thermo scientific, #229817), washing the strainer with another 4 ml 10% FCS-PBS. Cells were counted and centrifuged for 10 minutes (RT, 0.3rcf). The supernatant was removed, cells were resuspended in 949.5 µl 10% FCS-PBS and 54.1 µl 37% Formaldehyd (Sigma-Aldrich,#252549) was added (end concentration 2%), for a 10 minute incubation while tumbling. The fixation was quenched by adding 100 µl 1.425 M glycine on ice and centrifuged for 15 minutes (4°C, 0.5rcf). The supernatant was removed and cells were resuspended in freshly made ice cold HiC lysis buffer (1 ml for $5x10^6$ cells, 2 ml for $10x10^6$ cells) and incubated for 20 minutes on ice. After centrifugation for 15 minutes (4°C, 0.5rcf), the nuclei were washed and resuspended in PBS. The nuclei were split into $1x10^6$ aliquots, snap frozen in LN and stored in -80°C until library preparation.

4.5.3 3C-library preparation

Samples were defrosted and resuspended in 500 µl 1x DpnII restriction buffer (50 µl 10x DpnII Buffer (NEB, #R0543S) into 450 µl H_2O) and centrifuged for 10 minutes (4°C, 1.0rcf). Samples were resuspended in 175 µl digestion mix and incubated in a thermomixer at 37°C and 1000 rpm. After 1 hour 16 µl 20% Triton X-100 was added and the sample was incubated for an additional hour. An aliquot of 4 µl was taken aside as an undigested control. For sample digestion 15 µl DpnII restriction enzyme (100U/µl) (NEB, #R0543S) was added to the sample, incubated for 1 hour on the shaker, followed by an additional 15 µl DpnII enzyme added and the sample was incubated ON. In the morning, another 15 µl DpnII enzyme was added and incubated for another 2 hours before stopping the digestion by incubating the sample at 65°C. The sample was cooled on ice and an aliquot of 4 µl was taken aside as a digestion control. This is followed by adding 250 µl 1x T4 ligation buffer (36.5 µl 10x T4 ligation buffer added to 88.5 µl nuclease free water) and 4 µl T4 ligase (30 U/µl) (Thermo Scientific, #EL0013) and an incubation in 16°C in a rotating incubator for 22 hours. After ligation 7 µl ligation control was taken from the sample before 2 µl Proteinase K (20 mg/ml) was added and the sample was incubated at 65°C ON. The next day 4 µl RNase A was added to the ligation reaction and incubated for 30 minutes at 37°C followed by adding 375 µl phenol-chloroform. The sample was vortexed thoroughly and transferred into a phase lock

tube in which it was centrifuged for 15 minutes (RT, 15 rcf) for the aqueous phase to be separated. The upper phase (~ 350 µl) was transferred to a new tube and a mix of 3M NaOAc, 30 µg Glycogen and 960 µl 100% ethanol was added and the sample was mixed until incubated at -20°C for at least ON to precipitate. After that the sample was centrifuged for at least 45 minutes (4°C, maximum speed). The resulting pellet was washed twice with 70% ethanol, dried and resuspended in 132 µl nuclease free H_2O. From that 1 µl was taken for measuring the concentration. From the remaining sample 1 µg was taken, filled up to 120 µl with nuclease free H_2O for sonication.

Before proceeding with the capture HiC part of the experiment, the control aliquots were tested for successful digestion and ligation of the sample. The aliquots were filled to 100 µl with 10 mM Tris (pH 7.5), 4 µl RNAse A was incubated and the sample was incubated at 37°C for 30 minutes, followed by adding 5 µl Proteinase K and another incubation at 65°C for 4 hours. Finally 100 µl phenol-chloroform was added to the aliquot, vortexed thoroughly, centrifuged for 10 minutes and the resulting upper phase (~ 25 µl) was analyzed on an agarose gel (1%).

After confirming successful digestion and ligation, the sample was sonicated with a E220 Sonicator (Covaris) (Duty cycle: 10%, Intensity: 5, Cycles per Burst: 200, Time:6 cycles of 240 seconds, Set Mode: Frequency Sweeping). The sheared sample was subjected to a 1.8x SPRI beads clean up, resuspended in 52 µl nuclease free H_2O and subjected to capture library preparation. Successful sonication was confirmed by running a 1 µl aliquot on TapeStation (D5000).

4.5.4 Capture HiC library generation & sequencing

The enrichment and library preparation steps were executed using Twist Biosciences reagents, if not indicated otherwise or already stated to be standard chemicals in house.
For capture HiC library preparation the sonicated sample was subjected to end repair by incubating 50 µl sample with 7 µl End Prep Rxn buffer (NEB) and 3 µl End Prep enzyme mix (NEB) for 30 minutes at 20°C followed by 20 minutes at 65°C. Followed by adding another mix of 30 µl Ligation master mix (NEB), 1 µl ligation enhancer (NEB) and 2.5 µl adaptor (NEB) for Illumina sequencing and an incubation for 15 minutes at 20°C. Finally 3 USER enzyme (NEB) was added and incubated at 37°C for 15 minutes, followed by a 1.8x SPRI clean up.
The sample was resuspended in 15 µl nuclease free H_2O and after adding 5 i5 &i7 indexing primer respectively and 25 Ultra Q5 2x Master Mix (NEB) subjected to PCR library

amplification (7 cycles). The amplified DNA was cleaned using 1.8x SPRI beads and resuspended in 32 µl nuclease free H_2O, from which 1 µl aliquot was checked for quality control and concentration on TapeStation (D5000) and Qubit.

As input for library preparation, 500 ng sample was aliquoted, dried in a SpeedVac system at 45°C and resuspended with 5 µl Blocker Mix 1 and 7µl Blocker Mix 2. Oligo panels were prepared for hybridization as recommended by Twist. 28 µl of the prepared oligos (4 µl Oligo Probes, 20 µl Hybridization Mix and 4 µl H_2O) and 30 µl Hybridization Enhancer were added to the resuspended library and incubated in a thermocycler at 70°C (lid at 85°C) for 16 hours-ON.

The next day, 200 µl Binding Buffer and 100 µl Streptavidin beads were mixed and added to a magnetic stand (Thermo Scientific, DynaMag). The wash was repeated twice and the last 200 µl Binding Buffer were added to the dried beads and vortexed to homogenization. The library was added to the mix and put to the magnetic stand. The supernatant was removed carefully and 200 µl Washing Buffer 1 was added. The solution was put onto the magnetic stand again, the supernatant was removed and 200 µl Washing Buffer 2 (heated to 48°C) was added and resuspended. After 5 minutes incubation at 48°C, the sample was put on the magnetic stand, this washing step was repeated 2x until the beads were resuspended in 45 µl H_2O. The solution was incubated on ice for 1 hour until amplified by PCR. For the PCR the beads mixture was split into two reactions, adding half of the amplification mix (2.5 µl Amplification Primers, 25 µl KAPA Ready Mix), respectively. For the panel size (~ 4Mb) 7 amplification cycles were run. The amplified library was cleaned up using 1.8x SPRI beads and pooled. From the library 1 µl aliquot was checked for fragment distribution and concentration on TapeStation (D5000) and Qubit.

The final cHiC library was sequenced in-house at the Max Planck Institute Sequencing facility on the NovaSeq 6000 (Illumina) with a depth of 50 million reads (100 bp paired-end).

4.5.5 Capture HiC Analysis

Bioinformatic analysis for the capture HiC data was done in cooperation with Josh Kim from the Spielmann Lab at the MPI for Molecular Genetics & UKSH in Lübeck, who set up and applied the Fan-C tool for insulation score analysis.

Sequencing outputs of paired-end reads length 100bp were mapped iteratively using bwa version 0.7.17-r1188 with MAPQ <=3 (default bwa-setting). Only read pairs for region chr18:53716792-57626883 (mm10) were considered for the generation of the cHiC maps.

Chimeric reads were split predictively using the DpnII recognition sequence and filtering of the data was performed with FAN-C(command: fanc pairs), established by the Vaquerizas Lab (https://github.com/vaquerizaslab/fanc) according to default parameters (self-ligated

reads, PCR duplicates, and a maximum distance to restriction site of 10kb)(Kruse, Hug, and Vaquerizas 2020). Additionally, multimapping (-u) and low quality (-q 3) reads were filtered out.
cHiC contact matrices were generated with FAN-C (command: fanc hic) for a resolution of 5kb (Kruse, Hug, and Vaquerizas 2020). Low coverage regions were filtered automatically based on a 10% coverage threshold (-a). The diagonal was filtered up to a distance of 3 (-d 3). The matrices were normalized according to Knight-Ruiz normalization (Knight and Ruiz 2012). The custom map was generated, extracting the deleted region (chr18:56066950-56636813) from the reference genome.

4.5.6 Subtraction Maps

Subtraction maps were generated by first scaling the two matrices (duplication and wildtype of deletion and wildtype) to have the same number of valid pairs. These were generated from unnormalized contact matrices, as the assumptions made in normalization do not hold once subtracting two matrices.

4.5.7 Fan-C Insulation score

The insulation score was called using FAN-C which uses the methods applied in (Crane et al. 2015). Briefly, contacts are summed up and averaged in a sliding window along the diagonal of the contact matrix. The averages are normalized according to the log2-ratio of each individual score and the global mean of the matrix, local extremes of low scores are interpreted as insulating regions or TAD boundaries.

4.6 Magnetic resonance imaging (MRI) of aging mutant ADLD mice

The MRI experiment and analysis was performed in collaboration with Susanne Mueller and Professor Dr. Philipp Böhm-Sturm from the Department of Experimental Neurology at Charité - University Medicine Berlin.

4.6.1 MRI experiment

Anesthesia was induced with 2.5 % and maintained with 2.0 – 1.5 % isoflurane (Forene, Abbot, Wiesbaden, Germany) in a O_2/N_2O mixture (0.3/ 0.7l/min). The animal's body temperature was measured rectally and maintained at 37°C by placing animals on a blanket

with circulating warm water. Ventilation was monitored with a pressure sensitive pad (Small Animal Monitoring & Gating System, SA Instruments, Stony Brook, New York, USA). Morphometric T2-weighted (T2w) and diffusion MRI (dMRI) were performed at a 7 T Bruker BioSpec with a Tx/Rx 1H-cryoprobe and ParaVision 6.0.1 software. T2w images were acquired with a 2D-RARE sequence (TR/TE = 4250 ms/33 ms, RARE factor 8, 2 averages, 40 axial slices with a slice thickness of 0.4 mm, field of view (FOV)=19.2 x 19.2 mm2, matrix of 192 x 192, BW=34.7 kHz and TA=3:24 min). Single shell diffusion Tensor Imaging (DTI) (60 diffusion directions at b=1000 s/mm2, 5 b=0 images) was performed using a 2D segmented spin echo EPI sequence (4 segments) with identical slice thickness and field of view but lower resolution (matrix=160 x 160, TR=3500 ms /TE=30 ms, BW=300 kHz, diffusion gradient duration / separation= 2.7 ms / 8.6 ms, TA=15:10 min).

4.6.2 MRI Data analysis of T2-weighted (T2w) and diffusion MRI (dMRI) axial diffusion matrices

Analysis was performed as previously described in: (Hoffmann et al. 2023). In short, the generated MR images were converted into NIFTI format. T2-weighted (T2w) images were segmented into gray matter, white matter and cerebrospinal fluid. A customly generated brain atlas of 308 anatomical defined regions (154 right / 154 left hemisphere) adapted from Allen mouse brain atlas was used to define anatomical differences in the adult mutant mice T2w image using ANTx2 (https://github.com/ChariteExpMri/antx2). Using T2w eight images and their matching atlas those were coregistered to the b=0 images, allowing for atlas based connectome analysis of the diffusion MRI (dMRI) data. The respective connectomes were generated in mrtrix (https://www.mrtrix.org) the used shell scripts are openly available (https://github.com/ChariteExpMri/rodentDtiConnectomics). In brief, in downstream analysis following steps were performed: denoising, Gibbs ringing removal and the correction of bias field, eddy current and motion. Further the reconstruction of diffusion orientation by spherical deconvolution and connectome by applying streamline tractography and SIFT2 optimization. Connectivity matrices were reconstructed by counting the number of streamlines from atlas region to region. Group statistical comparisons between the mutants and the wildtype adult mice were reconstructed between pairs of regions with t-tests and for the explorative analysis fixed p-values with a threshold of 0.001 were used. Detected differences between the conditions were visualized using custom MATLAB tools (MathWorks, Natick, MA,USA).

4. 7 MMCA Methods

This part of the Methods is taken from (Huang, Henck, Qui *et al.* 2023). Note, that references to publications, figures and tables are not referring to this document.

Embryo collection

Mutants were generated through conventional gene editing tools and breeding or tetraploid aggregation and collected at the embryonic stage E13.5, calculated from the day of vaginal plug (noon = E0.5). Collection and whole embryo dissection was performed as previously described53. The embryos were immediately snap-frozen in liquid nitrogen and shipped to the Shendure Lab (University of Washington) in dry ice. Sets of animals with the same genotype were either all male or half male-half female. All animal procedures were in accordance with institutional, state, and government regulations.

Nuclei isolation and fixation

Snap frozen embryos were processed as previously described3. Briefly, the frozen embryos were cut into small pieces with a blade and further dissected by resuspension in 1 ml ice cold cell lysis buffer (CLB, 10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% IGEPAL CA-630, 1% SUPERase In and 1% BSA) in a 6 cm dish. adding another 3ml CLB, the sample was strained (40 µm) into a 15 ml Falcon tube and centrifuged to a pellet (500g, 5 min). Resuspending the sample with another 1 ml CLB, the isolation of nuclei was ensured. Pelleting the isolated nuclei again (500g, 5 min) was followed by a washing step by fixation in 10 ml 4% Paraformaldehyde (PFA) for 15 minutes on ice. The fixed nuclei were pelleted (500g, 3 min) and washed twice in the nuclei suspension buffer (NSB) (500g, 5 min). The nuclei finally were resuspended in 500µl NSB and split into 2 tubes, each containing 250 µl sample. The tubes were flash frozen in liquid nitrogen and stored in a -80°C freezer, until further use for library preparation. The embryo preparation was preceded randomly for nuclei isolation in order to avoid batch effects.

sci-RNA-seq3 library preparation and sequencing

The library preparation was performed previously described,54. In short, the fixed nuclei were permeabilized, sonicated and washed. Nuclei from each mouse embryo

were then distributed into several individual wells into 4 96-well plates. We split samples into four batches (~25 samples randomly selected in each batch) for sci-RNA-seq3 processing. The ID of the reverse transcription well was linked to the respective embryo for downstream analysis. In a first step the nuclei were then mixed with oligo-dT primers and dNTP mix, denatured and placed on ice, afterwards they were proceeded for reverse transcription including a gradient incubation step. After reverse transcription, the nuclei from all wells were pooled with the nuclei dilution buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2, 1% SUPERase In and 1% BSA), spun down and redistributed into 96-well plates containing the reaction mix for ligation. The ligation proceeded for 10 min at 25°C. Afterwards, nuclei again were pooled with nuclei suspension buffer, spun down and washed and filtered. Next, the nuclei were counted and redistributed for second strand synthesis, which was carried out at 16°C for 3h. Afterwards tagmentation mix was added to each well and tagmentation was carried out for 5 minutes at 55°C. To stop the reaction, DNA binding buffer was added and the sample was incubated for another 5 minutes. Following an elution step using AMPure XP beads and elution mix, the samples were subjected to PCR amplification to generate sequencing libraries.

Finally after PCR amplification, the resulting amplicons were pooled and purified using AMPure XP beads. The library was analysed by electrophoresis and the concentration was calculated using Qubit (Invitrogen). The library was sequenced on the NovaSeq platform (Illumina) (read 1: 34 cycles, read 2: 100 cycles, index 1: 10 cycles, index 2: 10 cycles).

<u>Processing of sequencing reads</u>

Read alignment and cell-x-gene expression count matrix generation was performed based on the pipeline that we developed for sci-RNA-seq33 with the following minor modifications: base calls were converted to fastq format using Illumina's bcl2fastq/v2.20 and demultiplexed based on PCR i5 and i7 barcodes using maximum likelihood demultiplexing package deML55 with default settings. Downstream sequence processing and cell-x-gene expression count matrix generation were similar to sci-RNA-seq56 except that the RT index was combined with hairpin adaptor index, and thus the mapped reads were split into constituent cellular indices by demultiplexing reads using both the the RT index and ligation index (Levenshtein edit distance (ED) < 2, including insertions and deletions). Briefly, demultiplexed reads were filtered based on the RT index and ligation index (ED < 2, including insertions

and deletions) and adaptor-clipped using trim_galore/v0.6.5 with default settings. Trimmed reads were mapped to the mouse reference genome (mm10), using STAR/v2.6.1d57 with default settings and gene annotations (GENCODE VM12 for mouse). Uniquely mapping reads were extracted, and duplicates were removed using the unique molecular identifier (UMI) sequence (ED < 2, including insertions and deletions), reverse transcription (RT) index, hairpin ligation adaptor index and read 2 end-coordinate (i.e. reads with UMI sequence less than 2 edit distance, RT index, ligation adaptor index and tagmentation site were considered duplicates). Finally, mapped reads were split into constituent cellular indices by further demultiplexing reads using the RT index and ligation hairpin (ED < 2, including insertions and deletions). To generate the cell-x-gene expression count matrix, we calculated the number of strand-specific UMIs for each cell mapping to the exonic and intronic regions of each gene with python/v2.7.13 HTseq package58. For multi-mapped reads, reads were assigned to the closest gene, except in cases where another intersected gene fell within 100 bp to the end of the closest gene, in which case the read was discarded. For most analyses, we included both expected-strand intronic and exonic UMIs in the cell-x-gene expression count matrix.

The single cell gene count matrix included 1,941,605 cells after cells with low quality (UMI <= 250 or detected gene <= 100) were filtered out. Each cell was assigned to its original mouse embryo on the basis of the reverse transcription barcode. We applied three strategies to detect potential doublet cells. As the first strategy, we split the dataset into subsets for each individual, and then applied the scrublet/v0.1 pipeline59 to each subset with parameters (min_count = 3, min_cells = 3, vscore_percentile = 85, n_pc = 30, expected_doublet_rate = 0.06, sim_doublet_ratio = 2, n_neighbors = 30, scaling_method = 'log') for doublet score calculation. Cells with doublet scores over 0.2 were annotated as detected doublets (5.5% in the whole data set).

As the second strategy, we used an iterative clustering strategy based on Seurat/v360 to detect the doublet-derived subclusters for cells. Briefly, gene count mapping to sex chromosomes was removed before clustering and dimensionality reduction, and then genes with no count were filtered out and each cell was normalised by the total UMI count per cell. The top 1,000 genes with the highest variance were selected. The data was log transformed after adding a pseudo count, and scaled to unit variance and zero mean. The dimensionality of the data was reduced by PCA (30 components) first and then with UMAP, followed by Louvain clustering performed on the 10 principal components (resolution = 1.2). For Louvain

clustering, we first fitted the top 10 PCs to compute a neighbourhood graph of observations (k.param = 50) followed by clustering the cells into sub-groups using the Louvain algorithm. For UMAP visualisation, we directly fit the PCA matrix with min_distance = 0.1. For subcluster identification, we selected cells in each major cell type and applied PCA, UMAP, Louvain clustering similarly to the major cluster analysis. Subclusters with a detected doublet ratio (by Scrublet) over 15% were annotated as doublet-derived subclusters.

We found the above Scrublet and iterative clustering-based approach is limited in marking cell doublets between abundant cell clusters and rare cell clusters (e.g. less than 1% of the total cell population), thus, we applied a third strategy to further detect such doublet cells. Briefly, cells labeled as doublets (by Scrublet) or from doublet-derived subclusters were filtered out. For each cell, we only retain protein-coding genes, lincRNA genes, and pseudogenes. Genes expressed in less than 10 cells and cells expressing less than 100 genes were further filtered out. The downstream dimension reduction and clustering analysis were done with Monocle/v33. The dimensionality of the data was reduced by PCA (50 components) first on the top 5,000 most highly variable genes and then with UMAP (max_components = 2, n_neighbors = 50, min_dist = 0.1, metric = 'cosine'). Cell clusters were identified using the Leiden algorithm implemented in Monocle/v3 (resolution = 1e-06). Next, we took the cell clusters identified by Monocle/v3 and first computed differentially expressed genes across cell clusters with the top_markers function of Monocle/v3 (reference_cells=1000). We then selected a gene set combining the top ten gene markers for each cell cluster (filtering out genes with fraction_expressing < 0.1 and then ordering by pseudo_R2). Cells from each main cell cluster were selected for dimension reduction by PCA (10 components) first on the selected gene set of top cluster-specific gene markers, and then by UMAP (max_components = 2, n_neighbors = 50, min_dist = 0.1, metric = 'cosine'), followed by clustering identification using the Leiden algorithm implemented in Monocle/v3 (resolution = 1e-04). Subclusters showing low expression of target cell cluster-specific markers and enriched expression of non-target cell cluster-specific markers were annotated as doublets derived subclusters and filtered out in visualisation and downstream analysis. Finally, after removing the potential doublet cells detected by either of the above three strategies, 1,671,270 cells were retained for further analyses.

Whole mouse embryo analysis

As described previously3, each cell could be assigned to the mouse embryo from which it derived on the basis of its reverse transcription barcode. After removing doublet cells and another 25 cells which were poorly assigned to any mouse embryo, 1,671,245 cells from 103 individual mouse embryos were retained (a median of 13,468 cells per embryo). UMI counts mapping to each sample were aggregated to generate a pseudobulk RNA-seq profile for each sample. Each cell's counts were normalised by dividing its estimated size factor, and then the data were log2-transformed after adding a pseudocount followed by performing the PCA. The normalisation and dimension reduction were done in Monocle/v3.

We previously used sci-RNA-seq3 to generate the MOCA dataset, which profiled ~2 million cells derived from 61 wild-type B6 mouse embryos staged between stages E9.5 and E13.5. The cleaned dataset, including 1,331,984 high quality cells, was generated by removing cells with <400 detected UMIs as well as doublets (http://atlas.gs.washington.edu/mouse-rna). UMI counts mapping to each sample were aggregated to generate a pseudobulk RNA-seq profile for each embryo. Each cell's counts were normalised by dividing its estimated size factor, and then the data were log2-transformed after adding a pseudocount, followed by PCA. The PCA space was retained and then the embryos from the MMCA dataset were projected onto it.

Cell clustering and annotation

After removing doublet cells, genes expressed in less than 10 cells and cells expressing less than 100 genes were further filtered out. We also filtered out low-quality cells based on the proportion of reads mapping to the mitochondrial genome (MT%) or ribosomal genome (Ribo%) (specifically, filtering cells with MT% > 10 or Ribo% > 5). We then removed cells from two embryos that were identified as outliers based on the whole-mouse embryo analysis (embryo 41 and embryo 104). This left 1,627,857 cells (median UMI count 845; median genes detected 539) from 101 individual embryos that were retained for all subsequent analyses.

To eliminate the potential heterogeneity between samples due to different mutant types and genotype backgrounds, we sought to perform the dimensionality reduction on a subset of cells from the wildtype mice (including 15 embryos with 215,575 cells, 13.2% of all cells) followed by projecting all remaining cells, derived from the various

mutant embryos, onto this same embedding. These procedures were done using Monocle/v3. In brief, the dimensionality of the subset of data from the wildtype mice was reduced by PCA, retaining 50 components, and all remaining cells were projected onto that PCA embedding space. Next, to mitigate potential technical biases, we combined all cells from wildtype and mutant mice and applied the align_cds function implemented in Monocle/v3, with MT%, Ribo%, and log-transformed total UMI of each cell as covariates. We took the subset of cells from wildtype mice, using their "aligned" PC features to perform UMAP (max_components = 3, n_neighbors = 50, min_dist = 0.01, metric = 'cosine') by uwot/v0.1.8, followed by saving the UMAP space. Cell clusters were identified using the Louvain algorithm implemented in Monocle/v3 on three dimensions of UMAP features, resulting in 13 isolated major trajectories (Fig. 1e). We then projected all of the remaining cells from mutant mouse embryos onto the previously saved UMAP space and predicted their major-trajectory labels using a k-nearest neighbour (k-NN) heuristic. Specifically, for each mutant-derived cell, we identified its 15 nearest neighbour wildtype-derived cells in UMAP space and then assigned the major trajectory with the maximum frequency within that set of 15 neighbours as the annotation of the mutant cell. We calculated the ratio of the maximum frequency to the total as the assigned score. Of note, over 99.9% of the cells from the mutant mice had an assigned score greater than 0.8. The cell-type annotation for each major trajectory was based on expression of the known marker genes (Supplementary Table 2).

Within each major trajectory, we repeated a similar strategy, but with slightly adjusted PCA and UMAP parameters. For the major trajectories with more than 50,000 cells, we reduced the dimensionality by PCA to 50 principal components; for the other major trajectories of more than 1,000 cells, we reduced the dimensionality by PCA to 30 principal components; for the remaining major trajectories, we reduced the dimensionality by PCA to 10 principal components. UMAP was performing with max_components = 3, n_neighbors = 15, min_dist = 0.1, metric = 'cosine'. For the mesenchymal trajectory, we observed a significant separation of cells by their cell-cycle phase in the UMAP embedding. We calculated a g2m index and a s index for individual cells by aggregating the log-transformed normalised expression for marker genes of the G2M phase and the S phase and then included them in align_cds function along with the other factors. Applying these procedures to all of the main trajectories, we identified 64 sub-trajectories in total. Similarly, after assigning each cell from the mutant mice with a sub-trajectory label, we calculated the ratio of the maximum frequency to the total as the assigned score. Of note, over 96.7% of the

cells from the mutant mice had an assigned score greater than 0.8. The cell-type annotation for each sub-trajectory was also based on the expression of known marker genes (Supplementary Table 2).

Identification of significant cell composition changes in mutant mice using beta-binomial regression

A cell number matrix of all 64 developmental sub-trajectories (rows) and 101 embryos (columns) was created and the cell number were then normalised by the size factor of each column which was estimated by estimate_size_factors function in Monocle/v3. 10 sub-trajectories with a mean of cell number across individual embryo < 10 were filtered out. The beta-binomial regression was performed using the VGAM package of R, based on the model "(trajectory specific cell number, total cell number of that embryo - trajectory specific cell number) ~ genotype". Of note, embryos from the four different mouse strain backgrounds were analysed independently.

We hypothesise that the power of our strategy to detect the cell proportion changes between different genotypes is affected by three factors: a) the abundance of a given cell type; b) the number of replicates in each genotype group; and c) the effect size. To evaluate power, we performed a simulation analysis that varied these factors, implemented as follows:

We selected the 20 most abundant cell types in wildtype embryos. Their abundances ranged from ~1% to ~20%. The proportions of these cell types served as the basis for our simulations.
We simulated ten groups of "wildtype" samples with 4, 8, 16, …, 40 replicates in each group, wherein each sample consisted of cells drawn from the 20 cell types. For each replicate, the simulated number of cells of each cell type was calculated as the product of: a) the cell-type proportions, simulated by fitting a dirichlet model based on the real proportions from step 1; and b) the total number of cells recovered for that replicate, simulated based on the mean (n = ~15,000) and standard deviation of the cell numbers across replicates in the real dataset.
We simulated ten groups of "mutant" samples by repeating the above step except adding shifts to the numbers of cells within each cell type. The shifting scales were based on different effect sizes. For instance, effect size = 0.1 represents a 10% reduction in the number of cells.

We performed beta-binomial regression (the same test used in Fig. 2a) to test if the cell type proportions were significantly changed between simulated "wildtype" and "mutant" samples, further checking the results as stratified by cell type (with different abundances), the number of replicates, and the effect size.

The results are in line with our hypothesis that the detection power of our strategy varies among comparisons with different effect sizes, sample sizes, or cell-type abundances (Extended Data Fig. 4). The main "take-home" messages are summarised below:

25% changes are robustly detectable, even for rare cell types like <2%, with modest numbers of replicates.
10% changes are possible to detect, but only for abundant cell types (e.g. >5%). More replicates can help in this zone.
1% changes are almost impossible to detect with a cell proportions approach, even with very large numbers of replicates.

In general, at the level of single cell sampling performed in our study, four samples (corresponding to the number of samples used in the manuscript) would be sufficient to detect a 25% effect size for those cell types that is present at a 1% proportion in wildtype embryos.

<u>Defining and calculating lochNESS</u>

To identify local enrichments or depletions of mutant cells, we aim to define a metric for each single cell to quantify the enrichments or depletions of mutant cells in its surrounding neighbourhood. For these analyses, we consider a mutant and a pooled wildtype combining all 4 background strains in a main trajectory as a dataset. For each dataset, we define "lochNESS" as:

$$lochNESS = \frac{\#\ of\ mutant\ cells\ in\ kNNs}{k} / \frac{\#\ of\ mutant\ cells\ in\ dataset}{N} - 1 ,$$

where N is the total number of cells in the dataset, $k = \frac{\sqrt{N}}{2}$ *scales with N and the cells from the same embryo as the cell are excluded from the k-NNs. Note that this value is equivalent to the fold change of mutant cell percentage in the neighbourhood of a cell relative to in the whole main trajectory. For implementation, we took the aligned PCs in each sub-trajectory as calculated above and for each cell in an embryo we find the k-NNs in the remaining mutant embryo cells and wildtype cells.*

We plot the lochNESS in a red-white-blue scale, where white corresponds to 0 or the median lochNESS, blue corresponds to high lochNESS or enrichments, and red corresponds to low lochNESS or depletions.

Currently we calculate lochNESS using a pooled wildtype combining all 4 background strains to include larger numbers of cells in constructing the k-NN graph. If the numbers of cells are sufficient, a wildtype from the matched background strain can be used. Additionally, if the numbers of cells are sufficient, one set of lochNESS can be calculated for each wildtype sample separately and the variability between samples can be considered.

Examining global distributions of lochNESS

Plotting the global distributions of lochNESS for each mutant across all sub-trajectories, we further observed that some mutants (e.g. most TAD boundary knockouts; Scn11a GOF) exhibit unremarkable distributions (Extended Data Fig. 6a). However, others (e.g. Sox9 regulatory INV; Scn10a/11a DKO) are associated with a marked excess of high lochNESS, consistent with mutant-specific effects on transcriptional state across many developmental systems. For reference, we simultaneously create a null distribution of lochNESS using random permutation of the mutant and wildtype cell labels, simulating datasets in which the cells are randomly mixed. Of note, we confirmed that repeating the calculation of lochNESS after random permutation of mutant and wildtype labels resulted in bell-shaped distributions centred around zero (Extended Data Fig. 6b). As such, the deviance of lochNESS can be summarised as the average euclidean distance between lochNESS versus lochNESS under permutation (Extended Data Fig. 6c). In addition, we computed lochNESS between wildtypes from different background strains and observed minimal variation in cell distribution between wildtype from G4, FVB and BALB/C strains and potential strain-specific distributions in C57BL/6 wildtype mice (Extended Data Fig. 6d).

Identifying lochNESS associated gene expression changes

To identify gene expression changes associated with mutant enriched or depleted areas, we find differentially expressed genes through fitting a regression model for each gene accounting for lochNESS. We use the fit_models() function implemented in monocle/v3 with lochNESS as the model_formula_str. This essentially fits a

generalised linear model for each gene: $log(y_i) = \beta_0 + \beta_n * x_n$, *where* y_i *is the gene expression of* $gene_i$, β_n *captures the effect of the lochNESS* x_n *on expression of* $gene_i$ *and* β_n *is the intercept. For each* $gene_i$, *we test if* β_i *is significantly different from zero using a Wald test and after testing all genes, we adjust the p-values using the Bejamini and Hochberg procedure to account for multiple hypotheses testing. We identify the genes that have adjusted p-value<0.05 and large positive* β_i *values as associated with mutant enriched areas, and those with large negative* β_i *values as associated with mutant depleted areas.*

Calculating mutant and embryo similarity scores

We can extend the lochNESS analysis, which is computed on each mutant and its corresponding wildtype mice, to compute "similarity scores" between all pairs of individual embryos from the same background strain. We consider all embryos in the same background in a main trajectory as a dataset. For each dataset, we take define a "similarity score" between $cell_n$ *and* $embryo_j$ *as:*

$$similarity\ score_{cell_n, embryo_j} = \frac{\#\ of\ cells\ from\ embryo_i\ in\ kNNs\ of\ cell_n}{k} \Big/ \frac{\#\ of\ cells\ from\ embryo_j\ in\ dataset}{N}$$

Where N *is the total number of cells in the dataset and* $k = \frac{\sqrt{N}}{2}$. *We take the mean of the similarity scores across all cells in the same embryo, resulting in an embryo similarity score matrix where entries are:*

$$similarity\ score_{embryo_i, embryo_j} = \frac{1}{n_i} \sum_{n=1}^{n_i} similarity\ score_{cell_n, embryo_i}$$

Where n_i *is the number of cells in* $embryo_i$.

Identifying and quantifying developmental delay

To identify potential mutant related developmental delay, we integrate MMCA with MOCA. We consider a mutant and its corresponding wildtype in a sub trajectory as a dataset. We take the cells from E11.5-E13.5 with similar annotations from MOCA and co-embed with the MMCA cells. We take the raw counts from both datasets, normalise, and process the data together without explicit batch correction as both datasets were generated with sci-RNA-seq3 and were similar in dataset quality. We

visualise the co-embedded data in 3D UMAP space and check for developmental delay in the mutant cells (i.e. mutant cells embedded closer to early MOCA cells compared to wildtype cells). To quantify the amount of developmental delay, we find k-NNs in MOCA for each cell in MMCA and calculate $time\ score = \frac{\sum_{n=1}^{k} T_n}{k}$ *, where* T_n *is the developmental time of MOCA* $cell_n$ *in the k-NNs of the MMCA cell. Afterwards, we test if the average time scores of mutant cells are significantly different from that of wildtype cells using a student's t-test.*

RNAscope in situ Hybridization

For RNAscope, embryos were collected at stage E13.5 and fixed for 4 hours in 4% PFA/PBS at room temperature. The embryos were washed twice in PBS before incubation in a sucrose series (5%, 10% and finally 15% sucrose (Roth) /PBS) each for an hour or until the embryos sank to the bottom of the tube. Finally, the embryos were incubated in 15% sucrose/PBS and O.C.T. (Sakura) in a 1:1 solution before embedding the embryos in O.C.T in a chilled ethanol bath and put into -80°C for sectioning. The embryos were cut into 5 µm thick sections on slides for RNAscope.

Simultaneous RNA in situ hybridization was performed using the RNAscope® technology (Advanced Cell Diagnostics [ACD]) and the following probes specific for Mm-Kcnj2 (Cat. No. 476261, ACD) and Mm-Sox9 (Cat. No. 401051-C2, ACD) on five µm sections of the mouse embryos. RNAscope probes were purchased by ACD and designed as described by Wang et al.63. The RNAscope® assay was run on a HybEZ™II Hybridization System (Cat. No. 321720, ACD) using the RNAscope® Multiplex Fluorescent Reagent Kit v2 (Cat. No. 323100, ACD) and the manufacturer's protocol for fixed-frozen tissue samples with target retrieval on a hotplate for 5 minutes. Fluorescent labelling of the RNAscope® probes was achieved by using OPAL 520 and OPAL 570 dyes (Cat. No. FP1487001KT + Cat. No. FP1488001KT, Akoya Biosciences, Marlborough, MA, USA) and stained sections were scanned at 25x magnification using a LSM 980 with Airyscan 2 (Carl Zeiss AG, Oberkochen, DE).

Image analysis

*For quantitative analysis of the RNAscope images, representative fields of view for each stained section were analysed using the image processing software Fiji64. Each organ of interest mRNA signal was counted in a defined area (1 x 1 mm²) with an n=6 per condition. Statistics were calculated using student t-Test and evaluated (- p > 0,05 = non significant, p < 0,05 - ≥ 0,01 = *, p < 0,01 - ≥ 0,001= ** - p < 0,001= ***).*

Ttc21b and Gli2 Mutant fixation for H&E and Immunofluorescence

Ttc21b homozygous, heterozygous mutants and wild-type E13.5 mouse embryos were fixed overnight in 4% PfA at 4 degrees. To stop fixation, the samples were transferred into 70% ethanol, washed twice and dehydrated. In the following, the embryos were embedded in paraffin, and cut into 2,5 µm thick sections.

Ttc21b Mutant H&E staining

Histochemical staining was performed on the eyes of the embryos using haematoxylin and eosin. Slides were scanned with a digital slide scanner (NanoZoomer 2.0HT, Hamamtsu, Japan) and analysed using NDP.view2 software (Hamamatsu Photonics). Numbers of processed embryos: wild type = 2, Ttc21b heterozygous=2, Ttc21b homozygous = 4.

Gli2 Mutant H&E and Immunofluorescence

For the histological analysis, hematoxylin and eosin staining of E13.5 Gli2 -/- mouse embryos, and respective wild-type littermates (n=4 and n=2, respectively), was performed on 4% paraformaldehyde fixed paraffin-embedded sections (3 µm). Paraffin stained sections were scanned using a digital slide scanner (NanoZoomer 2.0HT, Hamamatsu, Japan) and examined using NDP.view2 software. The cut regions and positions were annotated according to "The Atlas of Mouse Development" by M.H.Kaufman (1992, Academic Press, London) (https://www.emouseatlas.org/emap/eHistology/kaufman/index.php).
The spatial abundance patterns of Prealbumin as a marker for choroid plexus and Pax6 as a marker for neural tube development were analysed by immunofluorescence, using specific antibodies (Rabbit monoclonal [EPR20971] to Prealbumin (1:1000, Abcam) and rabbit polyclonal to Pax6 (1:200, AB2237 Merck

Sigma) in an automated BOND Research Detection system. Antibody binding was detected by Goat Anti-Rabbit Alexa Fluor 488-conjugated secondary antibody (Leica, A-11008). Nuclear counterstaining was achieved using 4',6'-diamino-2-phenylindole (DAPI). In negative control sections, the primary antibodies were omitted and antibody diluent was applied. Stained embryo sections were scanned with an AxioScan 7 digital slide scanner (Zeiss, Jena, Germany).

Fluorescence quantification

Quantification of prealbumin expression cells was performed using image analysis software Definiens Developer XD2 (Definiens AG, Germany). The region of interests (ROI 1-4) within the fourth and lateral ventricle ChP, were annotated manually in serial sections. The calculated parameter was the ratio of the total number of prealbumin-positive cells over the embryo section area (µm).

Density visualisation and RNA velocity analysis

Using Seurat/v4.0.6, the raw counts were log-normalised, and PCA was performed with default parameters on top highly variable genes 2000 genes, selected using the "vst" method. Dimensionality reduction was performed using PCA using default parameters, after which the UMAP embedding was carried out on all computed PC components. Density plots were created using the stat_2d_density_filled function in ggplot2/v3.3.5. For RNA velocity analysis using scVelo/v0.2.4, the total, spliced, and unspliced count matrices, along with the UMAP embeddings were exported as an h5ad file using anndata/v0.7.5.2 for R. The count matrices were filtered and normalised using scv.pp.filter_and_normalize, with min_shared_counts=20 and n_top_genes=2000. Means and variances between 30 nearest neighbours were calculated in the PCA space (n_pcs=50, to be consistent with default value in Seurat). The velocities were calculated using default parameters and projected onto the UMAP embedding exported from Seurat.

5. Results

The results section is divided into two parts. The first section's subject is single cell methods as a tool for phenotyping mouse mutants, a summary of the published single cell mouse mutant cell atlas (MMCA) dataset (Huang, Henck, Qui et al. 2023). The second part is an introduction into an in-depth analysis of two mouse models for the neurodegenerative disorder acute demyelinating leukodystrophy (ADLD).

Statement of contribution for the MMCA project : Embryo collection was carried out by **Malte Spielmann and me**. The sample preparation, fixation and library preparation was performed by **Jun Cao** and me, data analysis was performed by **Chengxiang Qiu**, **Xingfan Huang**, **Varun K. A. Sreenivasan** and me (specifically Quality Control and Cell Composition Analysis was performed by Chengxiang Qiu and me, the lochNESS analysis was developed by Xingfan Huang and me, the RNA velocity analysis of the *Sox9* regulatory INV mutant was performed by Varun K. A. Sreenivasan and me). Moreover, the validation experiments of the single cell experiments were performed by **Oana V Amarie**, **Rose Behncke**, **Patricia da Silva-Buttkus**, **Nils Hansmeier**, **Sascha Ulferts** and me (specifically, I prepared and fixed the samples, the RNAscope experiment was carried out by Rose Behncke, Sascha Ulferts and Nils Hansmeier, while the H&E staining and the immunofluorescence imaging was carried out by Oana V Amarie and Patricia da Silva-Buttkus).
Statement of contribution for the ADLD project: I conceived and designed the outline of the experiments and direction of the project. Further, I generated the mutant cell lines and mouse mutant lines, genotyped and collected all indicated samples. All sample preparations and library generations were performed by me. The single cell analysis was performed by me, using and altering the single cell pipeline generated by **Varun K. A. Sreenivasan** and **Saranya Balachandran.** Capture HiC analysis was performed by **Josh Kim** and me. The MRI experiment was designed, performed and analyzed by me and **Susanne Mueller** and **Professor Dr. Philipp Böhm-Sturm**.
Please note that, for consistency of writing and to avoid confusion, the term “we” is used throughout the following section.

5.1 Mouse mutant cell atlas (MMCA)

In recent years various single cell RNA sequencing (scRNA-seq) methods (a.o. Smart-seq2, Drop-seq, Chromium and sci-RNAseq3) have been published and successfully implemented to dissect gene expression of a biopsy, tissue or whole embryo at single cell resolution (Picelli et al. 2013; Macosko et al. 2015; Zheng et al. 2017; Cao et al. 2017). These methods

have contributed to uncover previously inaccessible biological processes, from the diversification of cellular composition contributing to the embryology of several organisms to the origins of diseases and the question of how and when cell types are disease causing (Cao et al. 2019; Qiu et al. 2022; Calderon et al. 2022; D. E. Wagner et al. 2018; Packer et al. 2019; Smajić et al. 2021; Rozenblatt-Rosen et al. 2020; Zhou et al. 2020; Sreenivasan et al. 2023).

Due to the limitation of cell number input in these methods, research of a disease using scRNA-seq was restrictively applied to tissues or biopsies of interest, excluding the potential detection of pleiotropic effects of the respective disease/disorder on the whole organism.

Performing scRNA-seq to phenotype 22 different whole mouse embryo mutants, ranging in severity of mutation from inversions of regulatory landscapes to single TAD boundary mutants, we aimed to challenge whether this method can be a tool implemented to phenotyping diseases in whole embryonic development. Additionally, we wanted to build computational tools to easily access and explore the complex dataset and allow for the comparison of phenotypes between the mouse models.

In total we collected 4 replicates each per mutant and the 4 background mouse strains (BALB/C, C57BL/6J, FVB and G4) for a total of 104 mouse embryos at the developmental stage E13.5 to subject them to one scRNA-seq experiment (**Fig. 5.1a, Supplementary Table 7.3.2**). We applied sci-RNA-seq3, a single cell method based on a three round split and pool barcoding strategy, enabling us to capture a vast amount of over 1.6 million cells (**Fig. 5.1b**)(Cao et al. 2019). The cells were sequenced and after quality control, we confidentially annotated 13 major trajectories and 59 sub-trajectories in the dataset (**Fig. 5.1c**, **Supplementary Figure 7.3**). To properly analyze the data we developed and applied two computational phenotyping strategies, namely, cell composition analysis and lochNESS, which will be further explored in the following section.

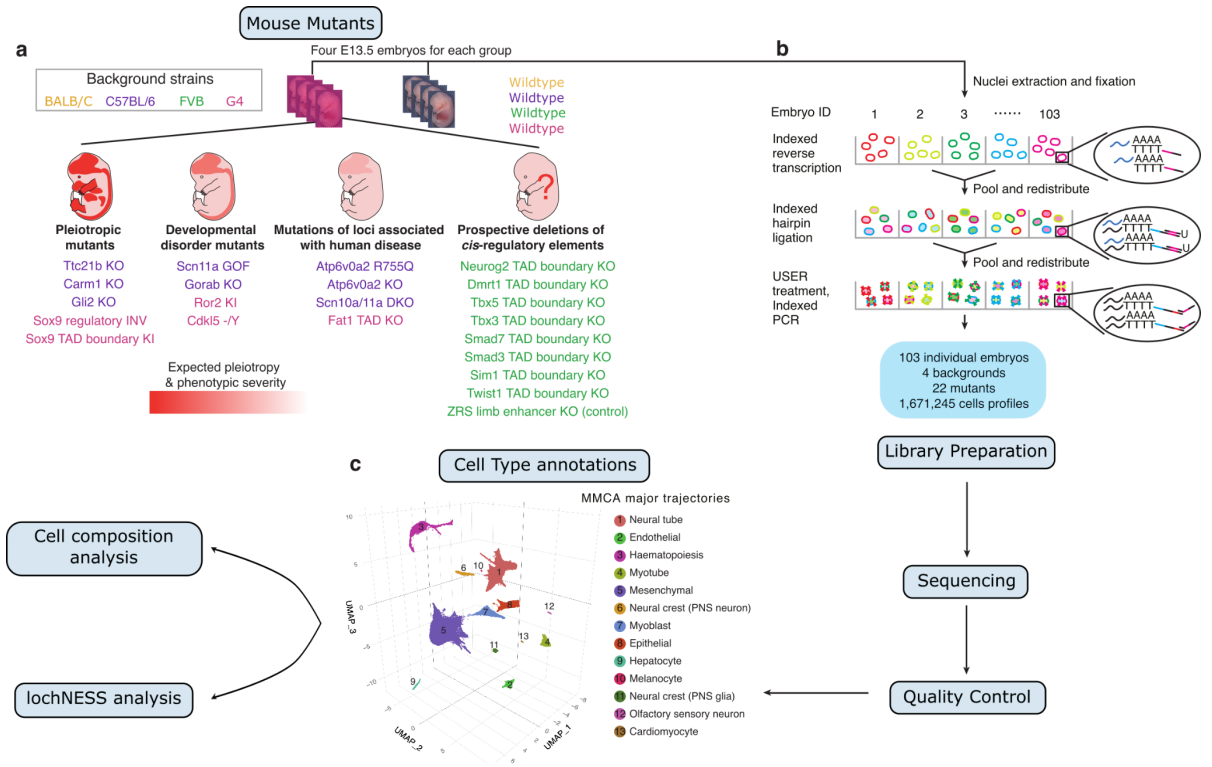


Figure 5.1 Schematic overview of the sci-RNA-seq3 experiment and analysis
a, Distribution of the mouse mutants into different categories of phenotype severities colored by genetic backgrounds **b,** Schematic of split and pool based 3 level sci-RNAseq barcoding and library preparation **c,** 3D-UMAP visualization of the wildtype dataset (215,575 cells from n=15 embryos), coloured by the major trajectories.

5.1.1 Quality of the dataset

First, we investigated the quality of the dataset by principal component analysis (PCA) on 'pseudobulk profiles' of each embryo (**Fig. 5.2a**). PCA revealed pseudobulk profiles of the FVB background embryos clustering separately from the BALB/C, C57BL/6 and G4 embryos. Further, embryo #104 was identified as an outlier from the dataset (**Fig. 5.2a**). For a reference of correct staging we integrated our samples with the published **M**ouse **O**rganogenesis **C**ell **A**tlas (MOCA) wildtype mouse embryos E9.5-E13.5 (Cao et al. 2019). Most pseudobulk profiles clustered around the MOCA stage E13.5 embryos confirming the correct staging of the MMCA collected embryos (**Fig. 5.2b**). Exceptions were embryo #104 which clustered separately again and was therefore excluded in further analysis. Also, 5 pseudobulk profiles clustered with an earlier stage of the MOCA atlas , three embryos from the *Scn11a* GOF mutant, which could be due to earlier sampling of this mutant or a specific developmental delay of this mutant. The other two outliers, #41 and #101 are single replicates of the C57BL/6 wildtype background and the *Carm1* KO mutant. Both were excluded from further analysis, as the other three replicates respectively clustered around the expected stage (**Fig. 5.2b**). Removing the two samples, we were left with 1,627,857

cells, distributed in numbers from a little less than 10.000 to 40.000 cells per sample in 101 embryos (**Fig. 5.2c**).

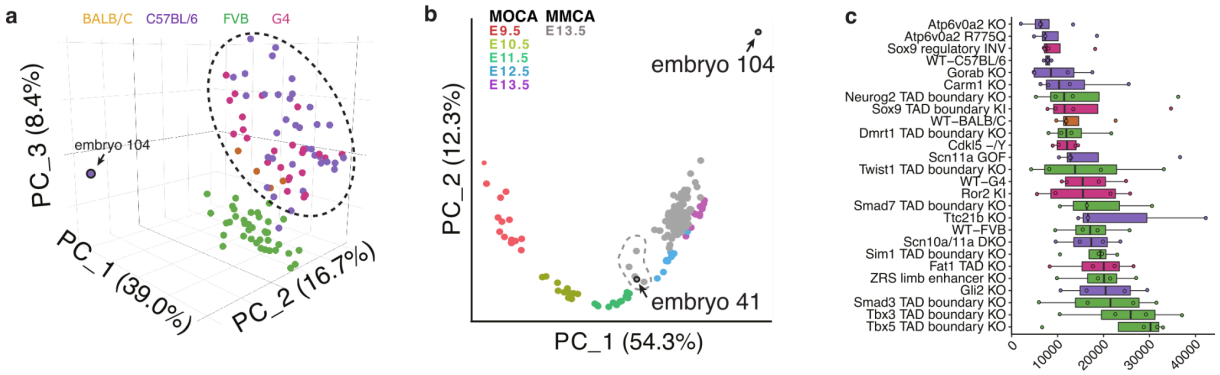


Figure 5.2 Quality control analysis of the dataset
a, PCA space pseudobulk MMCA embryo profiles. Visualization of the top three principal components, embryos are coloured by the respective mouse background strain. The black dotted circle highlights a cluster of two backgrounds (C57BL/6, BALB/c, G4) vs the FVB background. Embryo #104 is marked as an outlier sample.**b,** PCA space of pseudobulk MOCA embryo profiles (Cao et al. 2019), coloured by developmental stage, with MMCA embryos (grey) projected onto it. The top two principal components are visualized. Dashed line highlights five MMCA embryos which map with E11.5 or E12.5 embryos from the MOCA datasets. These outliers consist of three *Scn11a* GOF (#33, #34, #36) and one *Carm1* KO (#101) mutant and a C57BL/6 wildtype (#41). **c,** Boxplots of the number of cells per embryo, coloured by mouse background strain and ordered after median number of cells. The replicates (n=3 for WT-C57BL/6, n=4 for all others) are represented by dots.

5.1.2 Cell composition analysis

First, we aimed to investigate global changes in cell composition of the mutants across the annotated sub-trajectories. We developed the cell composition analysis for which we plotted all mutants (y-axis) against the sub-trajectories (x-axis), removing the 10 sub-trajectories consisting of a mean of under 10 cells across individual embryos (**Fig. 5.3a**). We applied beta-binomial regression to extract the log2-transformed ratio of cell number changes of each mutant against pooled references. These pooled references were generated by combining the other mutants and the respective wildtype background of the strain. The gain and loss of cell numbers are visualized in pie charts which characterize the direction of change (loss or gain) as well as intensity of change through color intensity and fullness of the circle (**Fig. 5.3a**).

A lot of different changes are detectable, within those a few trends are further explored below. First, there is an overall trend of smaller changes in cell type composition within the FVB background, compared to the C57BL/6 and G4 mutants, which possibly can be credited to them being TAD boundary mutants, solely carrying a deletion of a boundary element,

therefore likely less severe (Rajderkar et al. 2023). Another interesting observation is, that mutants that carry mutations targeting the same gene (*Atp6v0a2* KO, *Atp6v0a2* R755Q) are very similar in the depicted changes in cell composition, depicting the same direction of gain and loss of cell numbers within the different trajectories (**Fig. 5.3**) (Fischer et al. 2012).

Two mutants that further confirmed the success of this analytical approach are the ZRS limb enhancer KO and the *Sox9* regulatory INV mutant. The ZRS limb enhancer KO mutant carries the deletion of the ZRS enhancer (ZRS - **Z**one of polarizing activity **R**egulatory **S**equence) which is known to solely regulate the expression of *Shh* in the distal limb and when deleted, knockout mice depict an exclusive loss of the distal limb features (Kvon et al. 2016). The specific phenotype is reflected in the cell composition of this mutant, which shows the strongest change in the limb mesenchyme and minor, non-significant changes in the other trajectories (**Fig. 5.3**).

On the opposite end of cell composition changes is the *Sox9* regulatory INV mutant, which carries the inversion of a significant part of the regulatory landscape at the locus of *Sox9* gene including a TAD boundary and showed strong changes in several sub-trajectories, consistent with the known role of *Sox9* in the development of various tissues (Despang et al. 2019).

In a few mutants, specific sub-trajectories appeared extremely depleted. One example is the retinal neuron trajectory in the *Ttc21b* KO mutant. This mutant carries a mutation in the gene *Ttc21b*, which encodes for a cilial protein (Stottmann et al. 2009). The knockout was previously described to result in multiple malformations, for example in the forebrain, the neural tube, limb and eye (Stottmann et al. 2009).

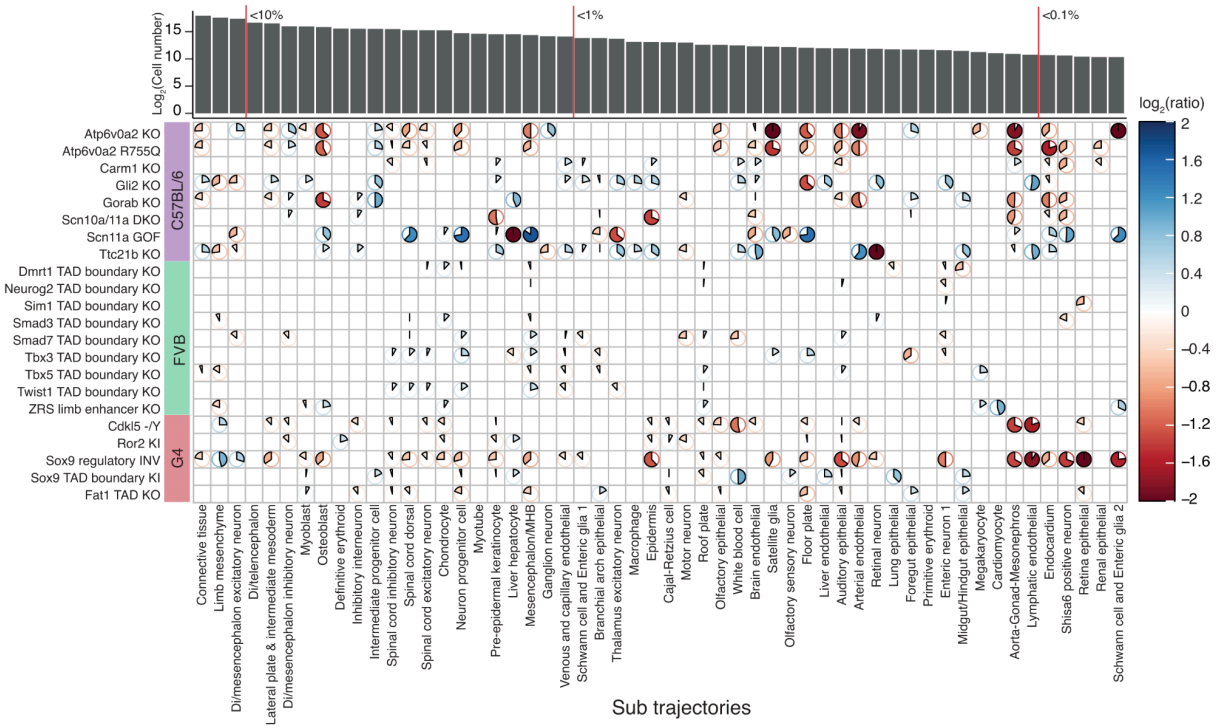


Figure 5.3 Cell composition whole dataset analysis
Pie charts of log2-transformed ratios of cell proportions between each mutant type (*y*-axis) against a pooled reference of wildtype and other mutants from the same strain across individual sub-trajectories (*x*-axis). Color and direction of the pies are representative of the log2-transformed ratios (gain=blue, clockwise; loss=red, anticlockwise). Only the significant values are displayed.

The eye phenotype was never investigated in further detail due to the complexity of the phenotype presented in the mutant. When examining the cell composition changes in UMAP space compare to the wildtype and other mutant of the C57BL/6 background, the retinal neuron trajectory as well as the lens and retinal epithelial trajectory appeared to be dramatically reduced, hinting at a complex multi-trajectory encompassing phenotype (**Fig. 5.4a**). This phenotype was confirmed, when comparing H&E staining of sections of the eye from the homozygous mutant to the heterozygous mutant and the wildtype at the embryonic day E13.5 (**Fig. 5.4b**). In the homozygous mutant, the structure of the lens, the retinal neurons as well as the optic nerve appear completely lost, while the retinal epithelium is collapsed to an aggregation of cells and the cornea is deformed (**Fig. 5.4b**). Taken together, these findings highlight the complexity of the eye phenotype caused by multiple small sub-trajectories as well as establish the capability of scRNA-seq and cell composition analysis to extract and visualize complex information from this type of datasets.

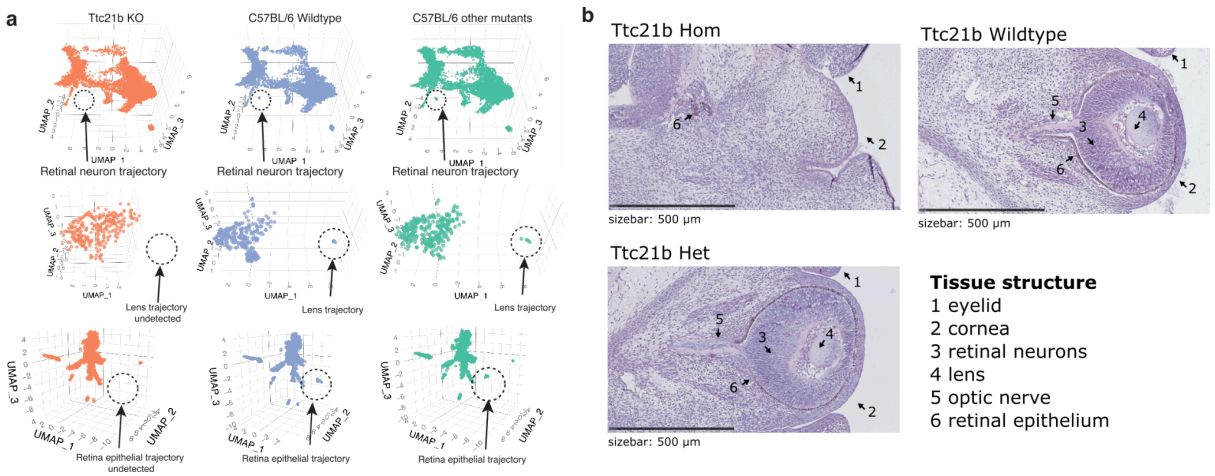


Figure 5.4 *Ttc21b* KO mutant eye development
a,3D-UMAP visualizations of retinal neuron, lens and retinal epithelial trajectories of the *Ttc21b* KO mutant (left), C57BL/6 wildtype (middle) and the other C57BL/6 background mutants (right). Trajectories of interest are highlighted with dotted circles. **b,** Developing eye section H&E stainings of E13.5 *Ttc21b* homozygous (Hom) & heterozygous (Het) mutants and C57BL/6 wildtype embryos. Tissues are indicated by arrows.

5.1.3 LochNESS analysis in the *Gli2* KO mutant

Small, disease related changes such as local de- or increase in cell numbers or deviations in gene expression within the trajectories that balance out across the whole trajectory can not

be detected by cell composition analysis. To capture those changes, we developed lochNESS (local cellular heuristic Neighborhood Enrichment Specificity Score). In essence, for every sub-trajectory, the score considers the aligned principal component features and determines the k-NNs for each cell to calculate the observed over expected number of mutant and wildtype cells in the environment (**Fig. 5.5a**).

For example, when the surrounding cells are a mixture of wildtype and mutant cells at the expected level these cells are given a neutral score whereas a de- or increase of mutant cells will result in a respectively negative or positive lochNESS.

Visualizing the lochNESS as a distribution across a sub-trajectory facilitates an easily accessible overview of potential local de- and increases in a variety of cell populations within the mutant. This way the method enables a fast visualization of a complex dataset. Visualization of the score in UMAP embeddings, and plotting the gene expression signatures associated with the changes in score enables to identify the cellular identity causative of the altered score. As a proof of concept, we looked at the neural tube lochNESS distribution of the previously discussed *Ttc21b* KO mutant, which, in line with the cell composition results, displays a strong depletion in the retinal neuron trajectory (**Fig. 5.3; Fig. 5.4a; Fig. 5.5b, left dashed box**). In comparison, the *Gli2* KO mutant displays a shift towards negative lochNESS values in the floor plate trajectory and two-peak distribution in the roof plate (**Fig. 5.5b, right two dashed boxes**). These two transient structures are emerging on the dorsal and ventral side of the neural tube and are organizing centers integral for the dorso-ventral patterning of the neural tube, development of the spinal cord and the proper migration of neurons by secreting signaling molecules from the ventral (floor plate) and dorsal (roof plate) side, respectively (Brambach et al. 2021; Matise et al. 1998). The gene *Gli2* encodes a transcription factor involved in the *Shh* regulatory network and mutations of the gene are associated with the Culler-Jones syndrome (Valenza et al. 2019). Amongst others microcephaly, craniofacial and skeletal abnormalities have been observed in homozygous mutant embryos, which are consistent with the knowledge about the transcription factors role in the development of these tissues (Mo et al. 1997). To explore the altered lochNESS within these trajectories of the *Gli2* KO we plotted the lochNESS distribution in the floor and roof plate in the mutant next. The floor plate depicts overall negative lochNESS values, whereas the roof plate splits in an area of negative and positive lochNESS values, resulting in the two peaks observed in the distribution (**Fig. 5.5c**).

When further annotating the respective clusters, genes like *Robo1* and *Slit1*, known to be involved in neuronal axon guidance, showed specific expression in the floor plate relative to the roof plate (**Fig. 5.5d**).

We confidentially annotated the two subclusters of the roof plate using genes, predicted by lochNESS. The cluster enriched in the mutant expressed *Ttr* encoding for Transtheryin, the

classical marker of the choroid plexus (ChP) and a major transport protein secreted with the cerebrospinal fluid (CSF) (**Fig. 5.5d**) (Lun, Monuki, and Lehtinen 2015).

The ChP is a derivative tissue of the roof plate in the brain, important for the production and secretion of the CSF (MacAulay, Keep, and Zeuthen 2022; Lun, Monuki, and Lehtinen 2015). During development the ChP develops independently in three distinct areas of the brain along the roof plate, namely the lateral ventricle, the 3rd and the 4th ventricle. The development starts at E11.5, with the 4th ventricle ChP to appear first, followed by the lateral and then the 3rd ventricle (Dani et al. 2021).

Further, the mutant depleted cluster depicted enriched expression of Wnt-related genes and was therefore identified as the lateral ChP associated Cortical hem, a tissue highly expressing Wnt and Bmp signaling molecules and involved in the organization of the hippocampus (**Fig. 5.5d**) (Caronia-Brown et al. 2014).

The proper expression of *Gli2* has long been established essential for the induction of the floor plate as previous reports of *Gli2* knockouts report a loss of floor plate structures and reduction of Shh signaling from the notochord into the neural tube resulting in the "'dorsalization" of the neural tube due to a disturbed dorso-ventral signaling (Mo et al. 1997; Q. Ding et al. 1998; Matise et al. 1998).

Less anticipated were the changes detected in the roof plate,in which *Gli2* expression has not previously been reported to be involved in developmental processes.

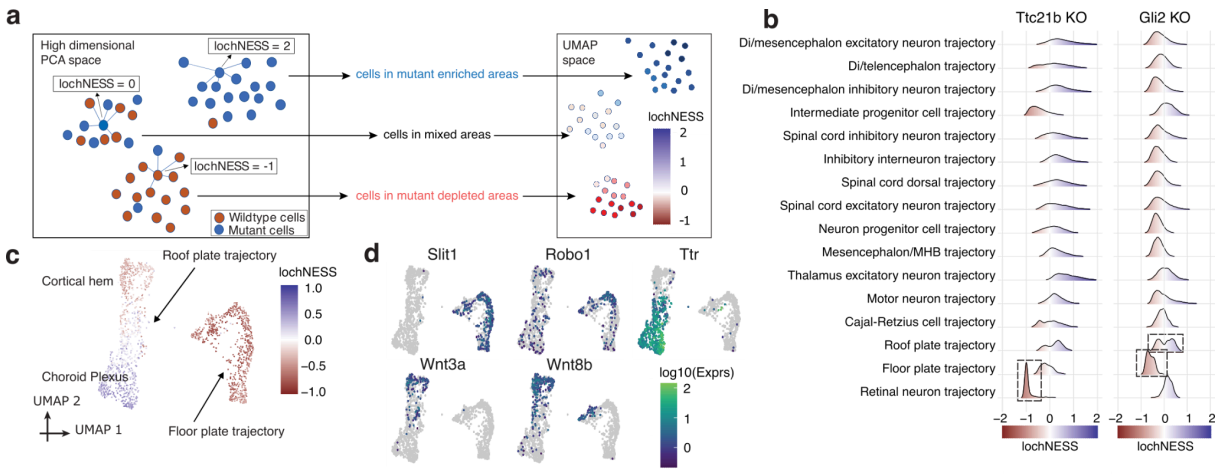


Figure 5.5 lochNESS analysis of the *Gli2* KO in the floor and roof plate
a, Schematic overview of lochNESS calculation per cell and visualization in UMAP space. **b,** LochNESS distribution across *Ttc21b* KO and *Gli2* KO mutant neural tube sub-trajectories. Marked changes in distributions are highlighted in dashed boxes. **c,** Floor plate and roof plate sub-trajectories lochNESS coloured UMAPs of the *Gli2* KO mutant and the pooled wildtype. Further annotations of the trajectories are indicated by arrows. ChP=choroid plexus **d,** Floor plate and roof plate sub-trajectories marker gene expression coloured UMAPs.

5.1.4 *Gli2* KO mutant morphological changes in the neural tube and choroid plexus

To experimentally validate these findings, we first investigated the organization of neural tube and ChP in E13.5 *Gli2* KO embryos compared to wildtype littermates. By performing H&E staining on transverse cranial sections of the mutant and the wildtype we could detect a severe delay in neural tube patterning and frontal lobe formation within the mutant, observable by eye (**Fig. 5.6a**). Taking the neural tube into focus, we performed immunofluorescence staining with the neural tube marker *Pax6* to reveal a strong dorsalization of the mutant neural tube (**Fig. 5.6b**). This finding is consistent with previous finding of *Gli2* misexpression resulting in malformations of the dorsal neural tube and could be explained by the failure to properly propagate *Shh* signaling to the ventral neural tube which, in consequence, collapses (Stamataki et al. 2005; Jacob and Briscoe 2003; Q. Ding et al. 1998; Matise et al. 1998).

Using immunofluorescence capturing *Ttr* expression, we next examined mutant and wildtype choroid plexus morphology within the lateral (LV) and 4th ventricle (4V) (**Fig. 5.6c,d**). The findings of the lochNESS analysis could be validated as there were detectable differences in the structure of the developing ChP between the *Gli2* KO mutant and the wildtype. While in the wildtype, most branches display a single layer of *Ttr*-positive cells, the mutant exhibits a 'double cell' layer in comparable locations within the tissue (**Fig. 5.6c,d;red arrows**). It is known that the ChP has to form a single layer of multiciliated epithelial cells that form the blood- CSF barrier, important for the production and clearance of CSF (Lun, Monuki, and Lehtinen 2015). However, our findings point towards a disordered ChP organization within the *Gli2* KO mutant indicating the expansion of the phenotype to a new tissue. To further connect the observations from immunofluorescence images to our lochNESS findings, the *Ttr*-positive cells were quantified relative to the size of the embryos head within the respective sections (**Fig. 5.6e**). The *Gli2* KO mutants displayed a marked increase in *Ttr*-positive cells compared to the wildtype, supporting the lochNESS results of a local increase in score within the roof plate derivatives of the *Gli2* KO mutant (**Fig. 5.5c**;**Fig. 5.6e**).

In summary, we could establish the lochNESS analysis as a tool to access local changes in mutants that would otherwise remain hidden in the complexity of the dataset. Notably, the changes described in the *Gli2* KO mutant were all detected in small sub-trajectories, demonstrating the granularity at which lochNESS can successfully extract meaningful results.

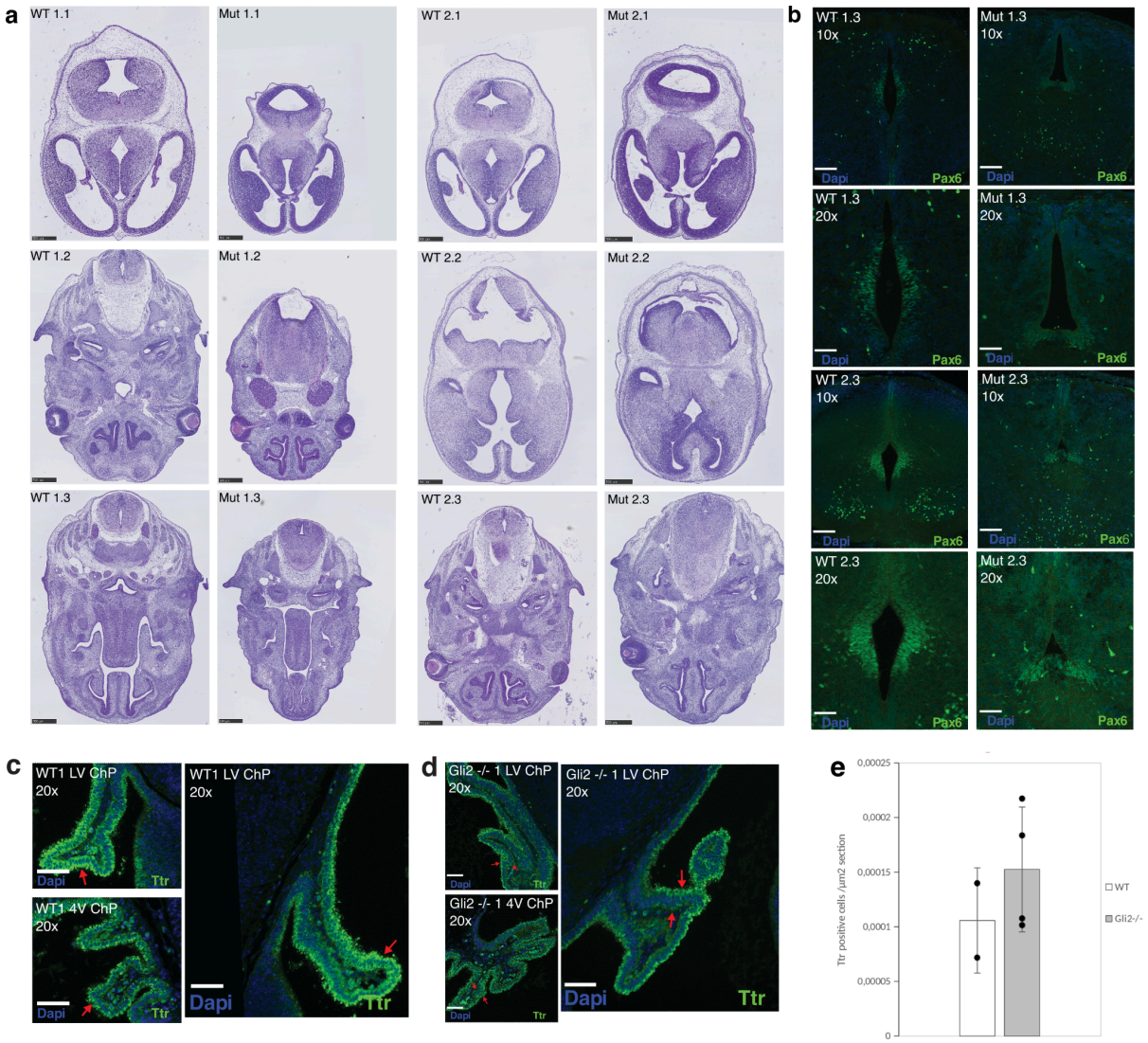


Figure 5.6 *Gli2* KO Mutant neural tube and choroid plexus deformations
a, H&E staining of head sections of two *Gli2* KO mutants (Mut1;2) and two wildtypes (WT1;2) at E13.5 ordered cranial to caudal matched based on tissue landmarks such as the positioning of the eye, tongue and nasal cavities (scale bar: 500 µm) **b,** Immunofluorescence imaging of *Pax6* neural tube marker in *Gli2* KO (Mut) and wildtype (WT) transversal neural tube sections of 10x and 20x magnification (scale bars: 100 µm (10x magnification) and 50 µm (20x magnification)). **c&d,** Immunofluorescence imaging of choroid plexus *Ttr* marker and DAPI within choroid plexus sections of **c,** wildtype (WT) and **d,** *Gli2* KO mutants (*Gli2* -/-) mutants in 20x magnification. (LV = lateral ventricle, 4V = 4th ventricle, ChP = choroid plexus) (scale bar: 50 µm). Red arrows highlight differences in choroid plexus morphological differences between the wildtype and the *Gli2* KO mutant **e,** Barplot fluorescence quantification of *Ttr* immunofluorescence signal within the choroid plexus of the wildtype (white bar) and the *Gli2* KO mutant (grey bar). Dots are the ratio of the total number of prealbumin-positive cells over the embryo section area (µm) per sample wildtype n=2, *Gli2* KO n=4.

5.1.5 Combinatorial in-depth analysis of the *Sox9* regulatory INV mutant

In the previous sections we demonstrated that the approaches cell composition analysis and lochNESS individually allow us to unravel new insights from a wide range of mouse mutants.

Next we sought to combine both approaches to investigate a phenotype in more detail. The *Sox9* regulatory INV mutant is an interesting candidate, as the gene encodes for a pleiotropic transcription factor known to be involved in the development of various tissues such as the nervous system, skeleton, gonads, lung, liver, intestines, skin and many more, orchestrated by a complex regulatory landscape (Despang et al. 2019; Yao et al. 2015; Scott et al. 2010; T. Wagner et al. 1994; Rockich et al. 2013). The mutant carries an inversion of a part of the regulatory landscape of the *Sox9* gene, resulting in the repositioning of a TAD boundary between the gene and some of its associated regulatory elements (**Fig. 5.7a**). Research about this mutant has demonstrated that the inversion caused a 50% decrease in *Sox9* expression. Notably, the inversion further resulted in misexpression of the functionally unrelated neighboring TAD gene *Kcnj2* in a *Sox9*-like pattern in the developing limbs “digit anlagen”, indicative of enhancer hijacking by the gene. The reported expression pattern results in a phenotype including abnormal digit formation as well as a cleft palate, deformed bones and an overall delay in ossification (Despang et al. 2019). Additionally, the *Sox9* regulatory INV mutant is perinatally lethal, suggestive of a more complex phenotype resulting from the inversion. This is supported by our cell composition analysis, which revealed differences in cell numbers in a wide range of trajectories of the *Sox9* regulatory INV mutant (**Fig. 5.3**).

We first examined the lochNESS distribution within the mesenchyme trajectory compared to other mutants from the same background (**Fig. 5.7b**). The *Sox9* INV mutant displays a strong gain in lochNESS compared to the other mutants. When split up into the mesenchymal sub-trajectories, a peak of cells with high lochNESS is revealed in the limb mesenchyme (**Fig. 5.7b**). Exploring the cellular distribution within UMAP embedding of the mutant and the wildtype mesenchymal trajectory, the mutant displays an accumulation of cells in the limb mesenchyme (**Fig. 5.7c**). Comparing the density further, the wiltype appears relatively evenly distributed across the other sub-trajectories, while the mutant is underrepresented. Interestingly, through applying RNA velocity analysis on the mesenchyme, it becomes apparent that all other sub-trajectories develop from the ‘limb mesenchyme’ challenging the correct annotation of this trajectory. Assuming this trajectory is a misannotated overall mesenchymal progenitor cell state which differentiates into chondrocytes, connective tissue, osteoblasts and intermediate mesoderm, this stalling phenotype builds on the findings from the original publication. The delay of ossification and limb phenotype, would expand to an overall delay of mesenchymal differentiation, resulting from developmental stalling of the *Sox9* regulatory INV mutant in a mesenchymal progenitor stage.

To test this hypothesis, we performed RNAscope on sections of the E13.5 *Sox9* regulatory INV mutant and a wildtype littermate (**Fig. 5.7d**). Expanding the in-depth analysis from previous work on the limb to the rib of the mutant, RNAscope analysis reveals a distinct

misexpression of *Kcnj2* in the pattern of *Sox9* as well as a reduction of *Sox9* signal in the developing bone (**Fig. 5.7d**). This phenotype is likely a result of the isolation of *Sox9* from its regulatory elements within the inverted region, known to drive expression in chondrogenesis (previously annotated: E250, E195) (Yao et al. 2015). These and other regulatory elements are now closer in proximity to *Kcnj2*, leading to the misexpression of the gene in a *Sox9*-like pattern. However there are additional regulatory elements positioned in the inverted region, associated with activity in the lung and the brain (previously annotated: E250, E195) (Yao et al. 2015). To examine whether the effect we detect in bones also appears in other tissues we investigated the gene expression in scRNA-seq as well as RNAscope quantification (**Fig. 5.7e,f**). In all tissues analyzed, there was an increase in *Kcnj2* levels, detectable with both approaches. However, while the effects of *Sox9* downregulation coincide in bone in both approaches, the gene was slightly upregulated in scRNA-seq data in all other tissues while showing a reduction of signal in telencephalon and lung in RNAscope (**Fig. 5.7e,f**). These differences are likely due to differences in cluster annotation of scRNA-seq versus the sections stained in the RNA scope data additionally to the small fraction of cells captured for expression analysis. Yet, through combined analysis we were able to add to the previously known phenotype of the *Sox9* regulatory INV mutant, the mechanism of stalling in the mesenchyme as well as other tissues involved through the inversion.

Taken together, scRNA-seq analysis of multiple whole embryo mutants and their backgrounds enabled us to expand on known phenotypes, explore unknown mutants and establish the technology of scRNA-seq as a phenotyping tool of multiple mutants in parallel, including the new analytical tools described.

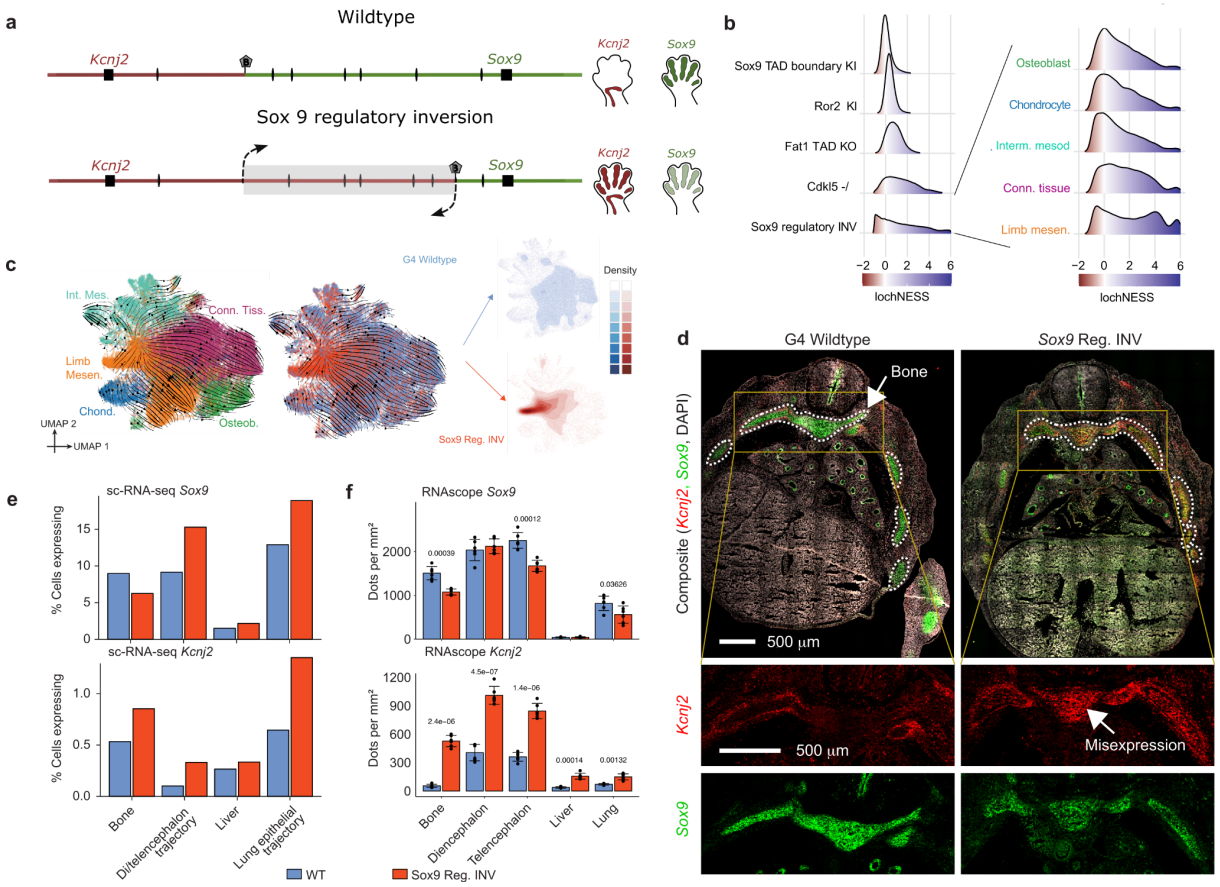


Figure 5.7 *Sox9* regulatory INV mutant changes in regulatory region related tissues
a, Model of the *Sox9* locus and the *Sox9* regulatory inversion structural variation depicting ectopic *Kcnj2* regulation through enhancer adoption (adopted from (Despang et al. 2019). **b,** LochNESS distributions of the G4 mutants in the mesenchymal trajectory and the mesenchymal sub-trajectories of the *Sox9* regulatory INV mutant **c,** RNA velocity plots of the mesenchymal G4 wildtype and *Sox9* regulatory INV mutant coloured by sub-trajectories (left) or genotype (right) and split into the 2D density plots split by genotype G4 (blue) and *Sox9* regulatory INV mutant (red). **d,** RNA scope images (red: *Kcnj2*; green: *Sox9*) of a transversal ribcage section of the E13.5 wildtype and heterozygous *Sox9* regulatory INV littermates, bone development is circled in white and the area of *Kcnj2* misexpression is highlighted (white arrow) **e,** scRNA-seq quantification of *Sox9* and *Kcnj2* expression in the wildtype (blue) and *Sox9* regulatory INV mutant (red) in candidate trajectories **f,** RNAscope image mRNA signal quantification of *Sox9* and *Kcnj2* expression based on RNAscope images of E13.5 wildtype and heterozygous littermate *Sox9* regulatory INV mutant embryos (n=6 embryos for each condition) of comparable tissues to the candidate trajectories (**Fig. 5.7f**.) The signal was quantified in the tissues in defined areas (1 x 1 mm²). Using a two-sided student's t-test, numbers representing mean differences of p-values, non significant values are not displayed. Standard deviation is represented by errors. The model of the locus and the limb was inspired by (Despang et al. 2019)

5.2 Mouse model of acute demyelinating leukodystrophy

As discussed before, single cell methods are enabling us to access the molecular phenotypes of mouse mutants at new sample sizes and granularity within one experiment. Still, there are limitations of the approach regarding the approach of phenotyping disease models. To briefly list a few, capturing only a fraction of cells from the respective tissue limits statements about quantification of expression data as touched on in the *Sox9* regulatory INV expression analysis. Second, the dissociation of tissues from whole embryos is yet to be improved on tougher tissues such as ossified bones, making full analysis of older embryos and aging organisms a challenge. Finally, certain diseases might not be molecularly established at the time of sampling. Single cell RNA-seq of a developmental stage serves as a snapshot of the developmental processes at that time. One extreme example are neurodegenerative diseases in which the tissue of interest only reaches full complexity postnatally. Current state of the art in the field is that neurodegenerative disorders are considered to only develop later in the lifespan of an organism and, in consequence the disease onset might not be detectable in early-stage full embryo analysis. Counterargument to this hypothesis is, if the disease driving mutation is caused in the germline, the molecular phenotype might establish prior to the onset of clinical symptoms (Falcão de Campos et al. 2023; Swaddiwudhipong et al. 2023; Katsuno, Tanaka, and Sobue 2012). This is a critical question to be explored in the field of neurodegenerative diseases in which the onset of degeneration often precedes the clinical diagnosis in years.

With the aim of challenging the state of the art in this field, we chose to investigate autosomal dominant leukodystrophy (ADLD), a slowly progressing disorder of the white matter caused by two different mutations, a duplication of the gene *Lmnb1* and the deletion of an upstream region close to the *Lmnb1* locus resulting in a demyelinating leukodystrophy in human patients with a clinical onset of symptoms at the age of 40-50 years (Giorgio et al. 2015; Nmezi et al. 2019; Quasar Saleem Padiath and Fu 2010; Brussino et al. 2010; Giorgio et al. 2013). On the basis of this disorder, we aim to unravel the time of onset of neurodegenerative disorders and disentangle the mechanism of disease development caused by the ADLD-deletion genotype. Alterations in the non-coding genome such as structural variants have been shown to result in disorders in rare skeletal phenotypes. In the field of neurological research, the focus remains on the coding genome as some of the more understood disorders such as Huntington’s and Parkinson’s disease are known to be caused by coding genome mutations (Spielmann and Mundlos 2016; Kraft et al. 2019; Will et al. 2017; Benko et al. 2009). Nevertheless, as neurological disorders combine to rank second in deaths, there is a huge need to study other potential causes such as mutations in the

non-coding genome to expand our knowledge on this type of disease (GBD 2016 Neurology Collaborators 2019).

Mouse models of a patient derived deletion and a duplication were generated using the CRISPR/Cas9 genome editing system in mouse embryonic stem cells (mESCs) (**Fig. 5.8**). Transgenic mouse lines were created through tetra- and diploid aggregations. The brains from mutant and wildtype mice were extracted from two embryonic stages, E11.5 and E18.5 and the changes in chromatin organization were investigated using capture HiC (cHiC) and the molecular phenotype using 10x based single cell RNAseq. Finally, adult mice were investigated in MRI scans for potential morphological phenotypes resulting from the introduced mutations (**Fig. 5.8**).

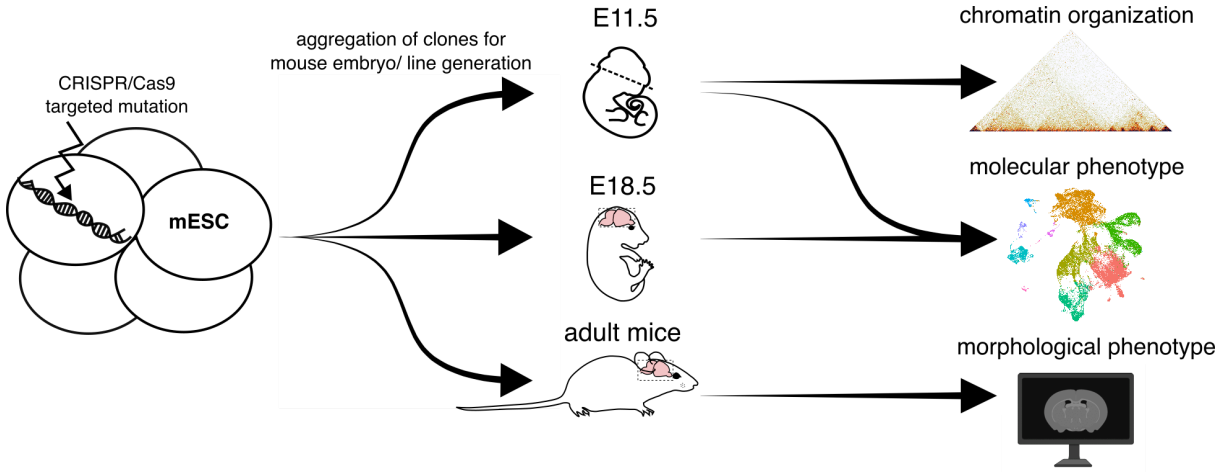


Figure 5.8 Visualization of the experimental strategy of the ADLD project
Introduction of mutations using the CRISPR/Cas9 based genome editing method was followed by embryonic stem cell (ESC) aggregation into mouse lines. Mouse brains were consequently used at different developmental stages in different approaches. Using embryonic stage E11.5 the chromatin organization of the locus (using cHiC) was investigated, the molecular phenotype (using scRNA-seq) was investigated at two embryonic stages (E11.5 and E18.5) and the morphological phenotype was investigated in adult mice (using Magnetic Resonance Imaging (MRI)).

5.2.1 The 3D chromatin changes upon ADLD based mutations

As described briefly in the introduction, *Lmnb1,* the gene encoding for the nuclear envelope protein Lamin-B1, is a nuclear envelope protein involved in many important biological processes (Camps, Erdos, and Ried 2015; Freund et al. 2012; Shimi et al. 2011; Cristofoli et al. 2020; Kaneshiro, Capitanio, and Hetzer 2023). To assess the changes in chromatin structure caused by introducing the mutations, the wildtype locus at the developmental time point E11.5 in the mouse brain was investigated. This stage was chosen, as this is the earliest time point in literature in which the enhancers at the *Zfp608* locus are active and the

gene *Lmnb1* is expressed (Visel et al. 2007; Bonev et al. 2017; Mahajani et al. 2017). The wildtype cHiC demonstrates that *Lmnb1* is located in a gene-dense topologically associating domain (TAD), which at E11.5 displays multiple loops characteristic of frequent chromatin interaction within the region (**Fig. 5.9**). Although only one enhancer is confidentially annotated via VISTA enhancer browser (vista annotation hs2489), these loops are characteristic for activity of multiple genes within this region at the developmental time point of E11.5 (Visel et al. 2007). Within the neighboring TAD, *Zfp608*, a zinc finger protein encoding gene, which is known to be expressed in the brain, is located in the otherwise gene-empty region (**Fig. 5.9**). Using the Fan-C tool build by the Vaquerizas lab we calculated the insulation score of the landscape, identifying two local minima in insulation score, indicative of two potential boundaries in between the genes *Lmnb1* and *Zfp608* (**Fig. 5.9**) (Kruse, Hug, and Vaquerizas 2020). For designing the targetable region in mESCs, the exact position of breakpoints are translated from known human ADLD patients. For the duplication it was important to choose a ‘classic case’ that represents the average duplication patient, without interfering too much with the surrounding coding genome. For the deletion, the selection was more restricted due to only a few cases defined, so choosing one of the big deletions allowed to investigate the effect of removing both insulation scores called boundaries and therefore get a strong contrast to the ‘classic duplication’ genotype (**Fig. 1.7**). Notably, research on the human deletion cases hypothesizes about the existence of a single boundary between the TADs, which would be targeted by the small as well as the big deletion (Nmezi et al. 2019). This is important to note as chromatin conformations can differ between species. Therefore the smaller deletion might not result in comparable effects as the bigger deletion in mice, due to a difference in chromatin conformation.

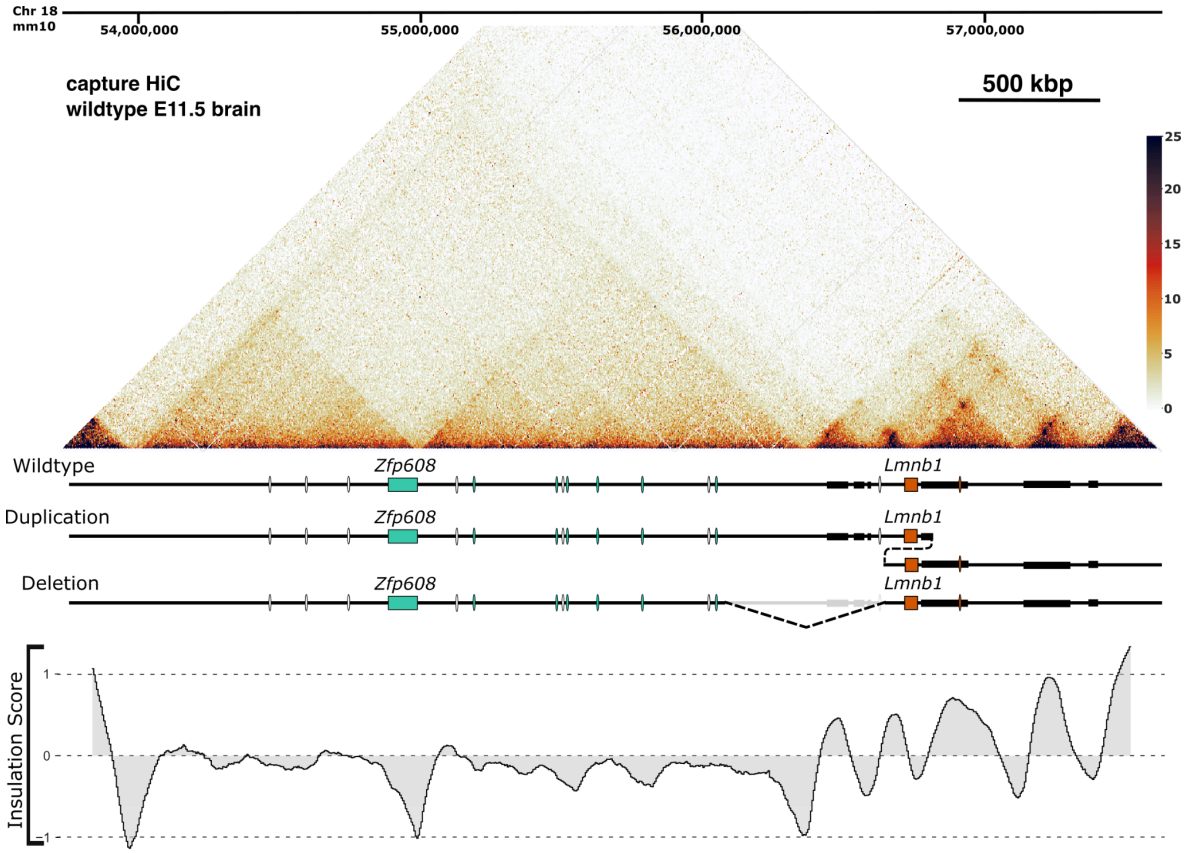


Figure 5.9 Wildtype *Lmnb1* locus with CRISPR mutation targets and wildtype insulation score
cHiC from the E11.5 wildtype mouse brain sample at the *Lmnb1* and *Zfp608* locus. The genes of interest *Lmnb1* (orange) and *Zfp608* (light blue) are highlighted in the genome for reference. The respective potential endogenous enhancers, sourced from VISTA enhancer browser (Visel et al. 2007) are indicated by color code matching the respective gene. Mutations (duplication and deletion) introduced into the genome are schematically visualized. The insulation score, calculated from the wildtype cHiC contact map is visualized below.

To investigate the chromatin changes induced by CRISPR/Cas9 into mESCs, mutants were generated and cHiC was performed on the same developmental stage as the wildtype (**Fig. 5.9**). The capture HiC maps of the heterozygous duplication and deletion depict very distinct changes in chromatin contacts compared to the wildtype (**Fig. 5.10**). Namely, while the duplication shows a local increase at the position of the introduced mutation, the deletion map shows a pattern typical for a deletion mapped to the wildtype genome in which the area of the deleted region is showing a v-shaped cut out, where the contact is removed in both directions. In addition, a stripe-like increase in contact formed in the deletion cHiC map which gets more apparent in the subtraction map of this mutant.

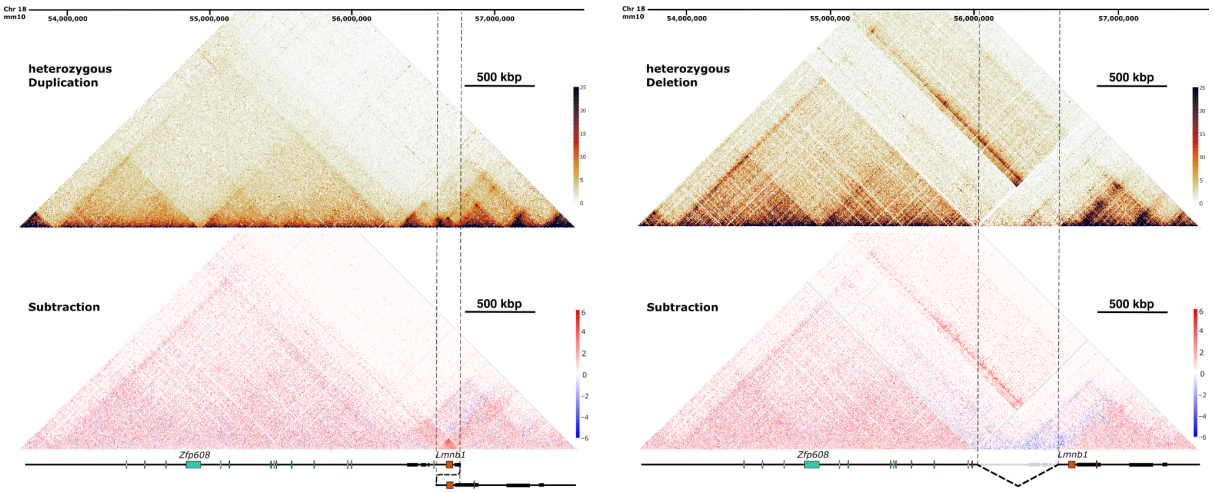


Figure 5.10 Subtraction maps of the duplication and deletion
cHiC of the E11.5 heterozygous duplication (top left panel) and deletion (top right panel) mutant mouse brain samples. Subtraction maps of the duplication (bottom left panel) and the deletion (bottom right panel) show contacts scaled to the wildtype contact map; increase (red) and decrease (blue) in contact is visualized relative to the wildtype contacts. Dotted lines mark the area of the duplicated region (top and bottom left panel) and the area of the deleted region (top and bottom right panel). The genes of interest *Lmnb1* (orange) and *Zfp608* (light blue) are highlighted in the genome for reference.

To inspect the structural changes in the deletion mutant further, we mapped the deletion cHiC contacts onto a custom map, which allows us to investigate the cHiC contact signal solely from the perspective of the allele carrying the deletion. Strikingly, the stripe-like formation detected in the heterozygous deletion mapped to the wildtype allele, is now located from the promoter of *Lmnb1* to the neighboring TAD of *Zfp608* (**Fig. 5.10;5.11**). Local increases of signal suggestive of loop formation become apparent, which notably match the position of known active enhancers in the *Zfp608* TAD at E11.5 in the brain (**Fig. 5.11; dotted circles**). Taken together the map of the deletion mutant shows contact of the *Lmnb1* locus to the neighboring TAD with loops to enhancers known to drive *Zfp608* expression at that developmental stage, further supporting the theory of enhancer hijacking of ectopic regulatory element by *Lmnb1* facilitated by the deletion of the boundary to the neighboring TAD. In contrast, the duplication cHiC map exhibits local changes in chromatin contact, indicating that, although the two mutations result in the same clinical phenotype, the disease mechanism is different regarding the regulatory elements involved in contacting the biomarker of disease development.

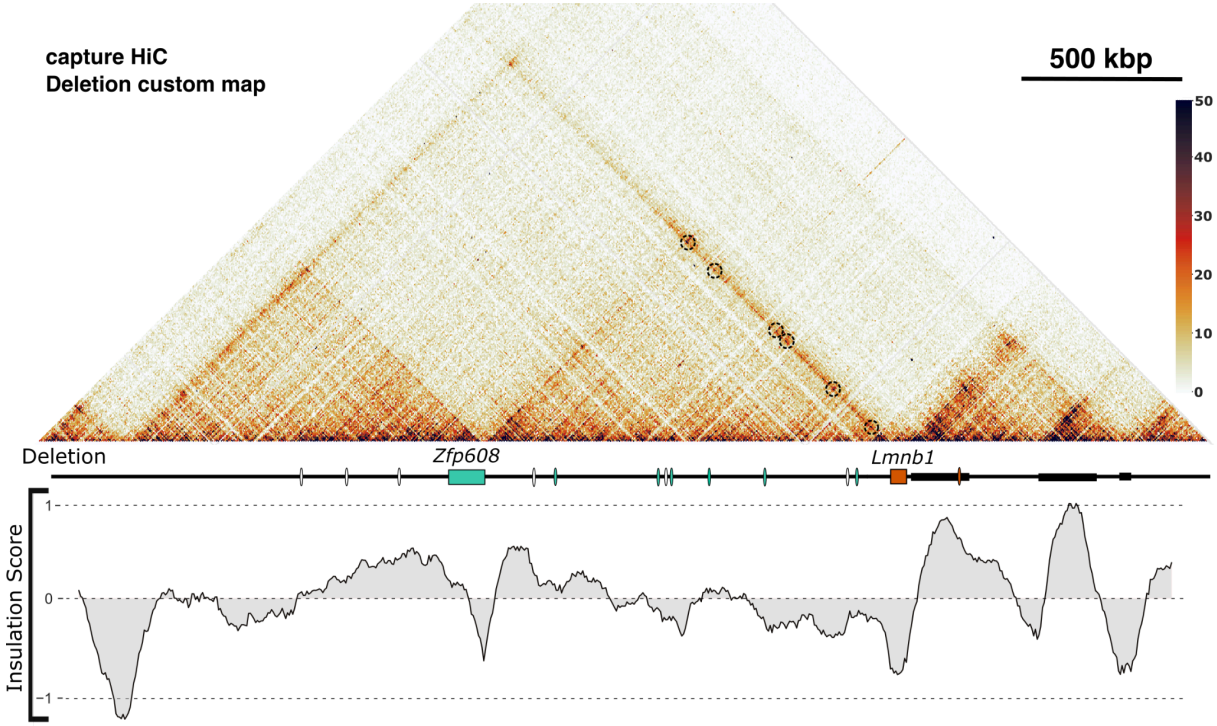


Figure 5.11 Deletion mutant custom map and insulation score
cHiC of the E11.5 deletion mouse mutant brain sample mapped to a custom genome excluding the deleted region with the calculated insulation score of the locus. The genes of interest *Lmnb1* (orange) and *Zfp608* (light blue) are highlighted in the genome for reference. Circles highlight increase in contact between the *Lmnb1* gene and neighboring TAD regulatory elements.

5.2.2 Molecular phenotype of the deletion and duplication mutant

To investigate phenotypic changes caused by aberrant chromatin conformation of the two mutations discussed above, scRNA-seq using the 10x technology was performed on both mutants and the corresponding wildtype. Knowing from the cHiC experiment that chromatin conformation changes are present during embryogenesis and to test whether the potential molecular changes already appear during embryology, the molecular state of E11.5 brains was explored. Additionally a late stage of embryonic development E18.5 was included to capture a cellular diversity representative of multiple cell types that is not present at the earlier time point (E11.5). Overexpression of *Lmnb1* has been studied previously with results stating different cell types to be causative in the ADLD, so the duplication mutant was included as a positive control to the deletion mutant to examine whether the two mutations result in differing phenotypes (Lee et al. 2014; Ratti, Rusciano, Mongiorgi, Owusu Obeng, et al. 2021; Rolyan et al. 2015; Heng et al. 2013; Ratti, Rusciano, Mongiorgi, Neri, et al. 2021).
To properly investigate the two developmental timepoints, integration of the datasets was performed using Harmony and the trajectories were annotated using known markers (**Fig. 5.12a, Supplementary Table 7.3.1**) (Korsunsky et al. 2019). As example, a few typical marker genes are displayed (**Fig. 5.12b**). The trajectories are interneurons 1 and 2 (*Gad2*),

excitatory neurons and their dendritogenesis & layer development (*Neurod2*), radial glia (*Hes5*), meninges (*Postn*), oligodendrocytes (*Olig1*), endothelial cells (*Rgs5*), the erythroid lineage (Hba-x) and microglia & macrophages (*Mrc1*) (**Fig. 5.12b**). The genes of the locus *Lmnb1* and *Zfp608* are expressed in both timepoints across the trajectories with no distinct differences within trajectories detectable in UMAP distribution (**Fig. 5.12c**). When comparing *Lmnb1* and *Zfp608* expression across all trajectories, the duplication shows significant upregulation in *Lmnb1* expression, while the deletion mutant depicts less but still significantly upregulation in *Lmnb1* and, notably also significant downregulation of gene expression in *Zfp608* expression compared to the wildtype (**Fig. 5.12d**). Splitting *Lmnb1* and *Zpf608* expression by the main trajectories, the genes are expressed in all trajectories with differences in respect to the strength of expression per trajectory in the wildtype sample (**Fig. 5.12e**). *Lmnb1* expression is highest in radial glia, while *Zfp608* is expressed at high levels in all the neuronal trajectories and the Endothelial cells (**Fig. 5.12e**).

In summary, at first impression the data reveals two important points 1. The upregulation of *Lmnb1* expression, which is known to be the disease-causing factor in ADLD, is detectable already during embryogenesis in both mutants and 2. The deletion shows a significant downregulation of expression in the neighboring TADs gene *Zfp608*. This strengthens the theory that the deletion is causing the phenotype through enhancer hijacking. In which case, the enhancers which typically regulate the transcription of *Zfp608* are now partially contacting *Lmnb1* causing the gene to be upregulated and, in consequence, reduce the contact to the endogenous gene *Zfp608*, causing the expression to be lower.

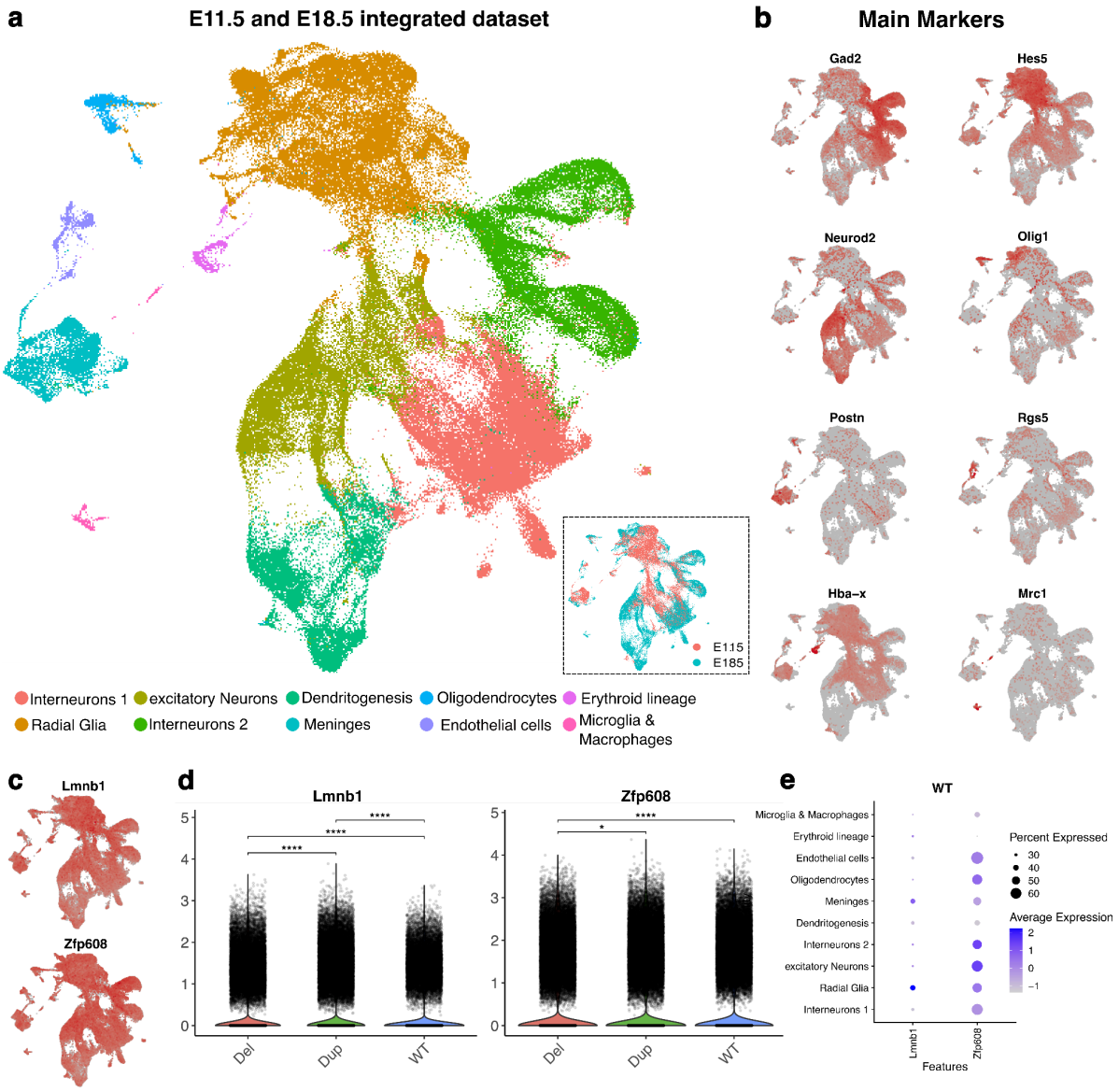


Figure 5.12 The mutant single cell dataset and locus specific gene expression changes
a, Uniform Manifold Approximation and Projection **(**UMAP) dimensional reduction visualization of the two integrated timepoints (E11.5 and E18.5) including both mutants (deletion and duplication) and the wildtype (Total of 128,511 cells from n=12 samples). The cells are coloured by main trajectory annotation. The dotted box highlights the same UMAP embedding colored by developmental time points (E11.5=red; E18.5= blue) **b,** UMAP visualization of gene expression of some marker genes used to annotate the main trajectories. **c,d,** UMAP and violin plot summarizing the gene expression of interesting locus and disease related genes *Lmnb1* and *Zfp608* with violin plot significance in gene expression differences between the samples (Del=deletion, Dup=duplication, WT=wildtype) marked as followed: - p > 0,05= non significant, p < 0,05 - ≥ 0,01 = *, p < 0,01 - ≥ 0,001= ** - p < 0,001= ***. **e,** Dotplot of *Lmnb1* and *Zfp608* gene expression split into the wildtype main trajectories.

To get a first impression of the overall pathways expressed across the main trajectories between the conditions, a gene set enrichment analysis (GSEA) was performed using hallmark gene sets (**Fig. 5.13a**) (Liberzon et al. 2015). Note, that the more cells a trajectory

consists of, the more likely there are stronger changes in gene sets between conditions. Radial glia are an interesting cell type to investigate, because the cell types thought to be involved in the cause of developing ADLD are astrocytes and oligodendrocytes, which are both descendants of the radial glia trajectory and it depicts the highest expression in the wildtype main trajectories indicative of a role of the gene within the cell type (**Fig. 5.12e**) (Rolyan et al. 2015; Heng et al. 2013; Koufi et al. 2023; Ratti, Rusciano, Mongiorgi, Neri, et al. 2021). Looking at the ten most significantly altered gene sets between the conditions (deletion vs wildtype; duplication vs wildtype; deletion vs wildtype) in the radial glia trajectory, cholesterol homeostasis appears downregulated in the duplication compared to the wildtype and the deletion, respectively (**Fig. 5.13b**). This is an interesting geneset to explore further as *Lmnb1* upregulation has previously been linked to cholesterol biosynthesis disruption in aging mice (Rolyan et al. 2015; Yattah et al. 2020). Cholesterol is a crucial building block in cell membranes and involved in various parts of cell survival pathways such as cell cycle checkpoints as well as in the synthesis of myelin. The disruption of cholesterol synthesis could be a primary downstream effect of *Lmnb1* gene misexpression and an important step in the process of disease development (Lasunción et al. 2022; Singh et al. 2013; Rolyan et al. 2015). Looking into cholesterol specific pathway regulation between the deletion and the duplication samples, all pathways are enriched in the deletion, hinting at differences in the regulation of cholesterol biosynthesis between the mutants (**Fig. 5.13c**). Further, investigating genes involved in the biosynthesis of cholesterol through the mevalonate pathway, a vital pathway in the synthesis of cholesterol from Acetyl-CoA, intermediates involved in the pathway such as *Lss*, *Nsdhl*, *Abca1*, *Mvk*, *Fdps* and *Sqle* are downregulated significantly in the duplication compared to the wildtype and partially the deletion (**Fig. 5.13d**) (Goldstein and Brown 1990; Lasunción et al. 2022). Markedly, intermediates involved in the initial stages of Acetyl-CoA processing toward the mevalonate pathway, *Hmgcr* and *Hmgcs1* were significantly upregulated in the deletion, together with the gene *Srsf2*, which encodes for the serine and arginine rich splicing factor 2 and is part of the spliceosome (**Fig. 5.13e**). The disruption of cholesterol pathways in astrocytes has been previously linked to the downregulation of *Hmgcr*, *Hmgcs1* and *Srsf2* inducing an altered crosstalk of astrocytes with neurons resulting in changes in neuron synapse transmission and age related memory defect (Raihan et al. 2018; Korinek et al. 2020). Although opposite expression trends are observed in the deletion mutant, the correlation between gene expression changes in glia cell related pathways and neuronal dysfunction is a promising direction of further data exploration toward the deletion mutant specific disease mechanism.

Taken together, cholesterol pathways are differentially regulated in the two mutants compared to the wildtype. Interestingly the importance of cholesterol synthesis has been previously stated in connection with myelin sheathing of neurons, and involved in

neurodegenerative disorders such as multiple sclerosis (Berghoff, Spieth, and Saher 2022). Studies investigating *Lmnb1* overexpression have demonstrated myelinsheat aberrations as a cause of leukodystrophies, so this pathway might be a promising target to explore further in the development of the disease (Rolyan et al. 2015).

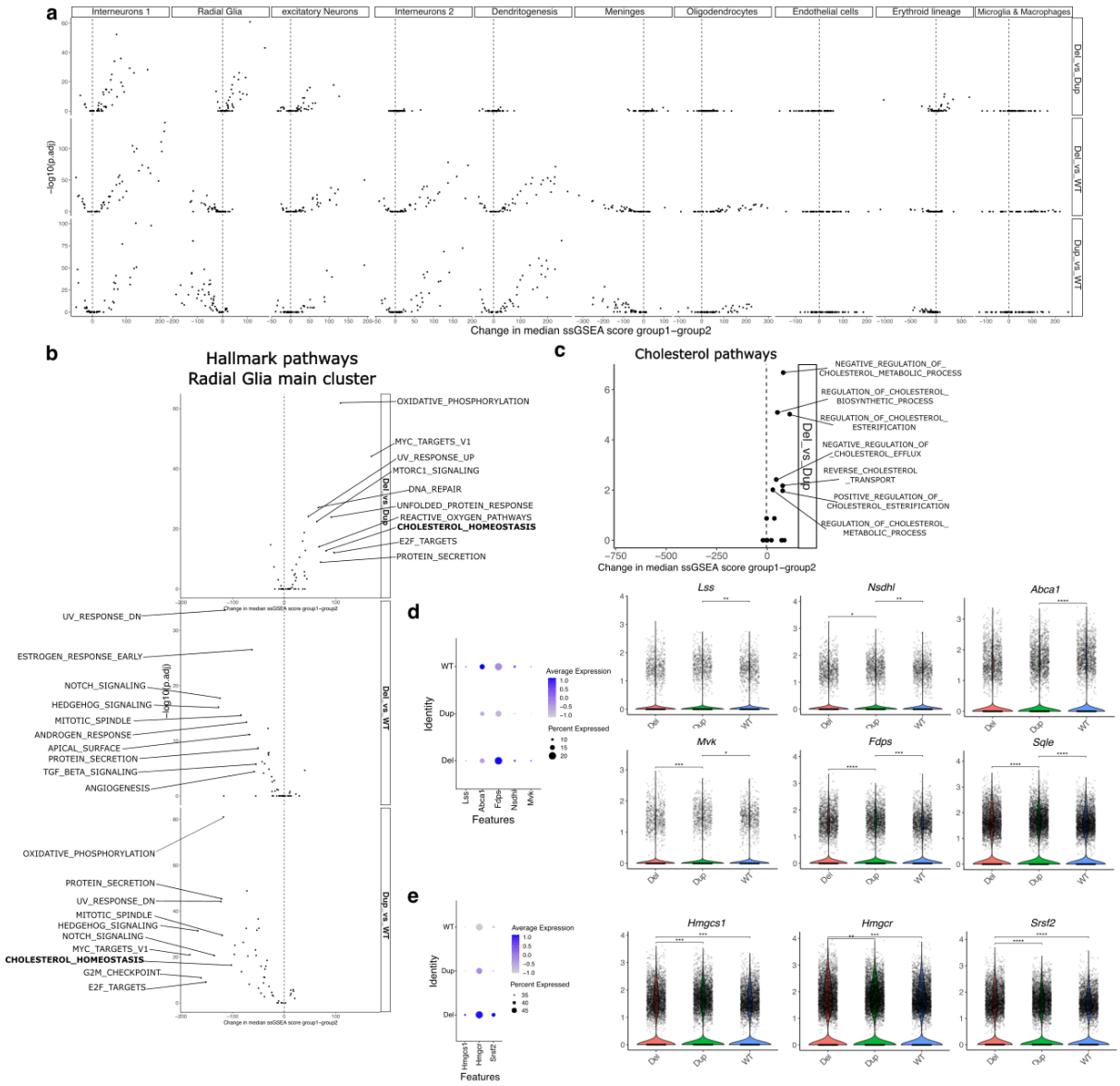


Figure 5.13 GSEA related changes in cholesterol pathways of radial glia
a, gene set enrichment analysis main trajectory for hallmark gene sets (Liberzon et al. 2015). Mean GSEA score comparison of mutant vs wildtype and mutant vs mutant per main trajectory. **b,** Mean GSEA score comparison of hallmark gene sets in the radial glia trajectory with the top ten significantly altered pathways labeled in the comparison of mutant vs wildtype and mutant vs mutant. **c,** Mean GSEA score comparison of Cholesterol pathways of the C5 ontology gene sets extracted by search term "C5,GOBP.+_cholesterol_" between the duplication and the deletion. Significantly altered pathways are labeled. **d, & e,** Dotplot and violinplot summarizing the gene expression of selected marker genes of the cholesterol pathway with violin plot significance in gene expression differences between the samples (Del=deletion, Dup=duplication, WT=wildtype) marked as followed: - p > 0,05= non significant, p < 0,05 - ≥ 0,01 = *, p < 0,01 - ≥ 0,001= ** - p < 0,001= ***.

GSEA demonstrated that the radial glia is an interesting trajectory for further exploration, so cell composition analysis was performed, looking into the changes in cell numbers between the duplication and deletion in the respective developmental time points (E11.5 and E18.5). Interestingly, the radial glia main trajectory is significantly reduced within the duplication mutant and the endothelial (gain) and microglia & macrophage (loss) trajectories are significantly changed in the deletion mutant at embryonic day E11.5 (**Fig. 5.14a**). There are no significant cell composition changes within the main trajectories at embryonic day E18.5 (**Fig. 5.14b**). Further looking into the sub-trajectories of the radial glia at E11.5, the gain of cell proportions in the sub-trajectory radial glia as well as cortical hem & choroid plexus are significant in both the deletion and the duplication mutant (**Fig. 5.14c**). Additionally, the ependymal cells show a significant gain in cell numbers in the duplication mutant (**Fig. 5.14c**). This demonstrates that the changes in other sub-trajectories might also be significantly altered and require further analysis of the dataset.

To understand what happens during glia cell differentiation in the mutants, GSEA analysis of glia differentiation was applied to both developmental time points separately (**Fig. 5.14d**). Interestingly, the significant gene sets downregulated in E11.5 are upregulated in E18.5 in the mutants compared to the wildtype, indicative of a change in glia differentiation pathways between the two developmental timepoints. Moreover, the changes of glia differentiation are stronger downregulated in the duplication at E11.5 compared to the deletion, which could be a reason for the stronger changes in cell composition in radial glia sub-trajectories in the duplication at this developmental time point (**Fig. 5.14d; Del vs. Dup**). With the aim of taking a closer look into the differences in cellular dynamics in the radial glia, RNA velocity was applied on the radial glia trajectory E11.5 cells of the wildtype, duplication and deletion (**Fig. 5.14e**) (La Manno et al. 2018). Detectable by eye in the neuronal progenitor subcluster, the wildtype cells have progressed outwardly, while the duplication and deletion cells are more concentrated within the center of the cluster, suggestive of a slight delay in the mutant's differentiation. Additionally the radial glia subcluster depicts differences in velocity directions across the conditions. While the arrows within the wildtype are distributed across the subcluster, the mutants present a concentration towards the lower edge of the subcluster, indicative of differences in differentiation pathway differences (**Fig. 5.14e, dotted circle**).

Taken together these results show that there are cell composition changes in the mutant radial glia. This trajectory showed the highest expression of *Lmnb1* in the wildtype dataset (**Fig. 5.12e**) and depicted pathway alterations such as in cholesterol biosynthesis which can be linked to *Lmnb1* overexpression, concluding in a promising trajectory for further investigations in disease development (Rolyan et al. 2015).

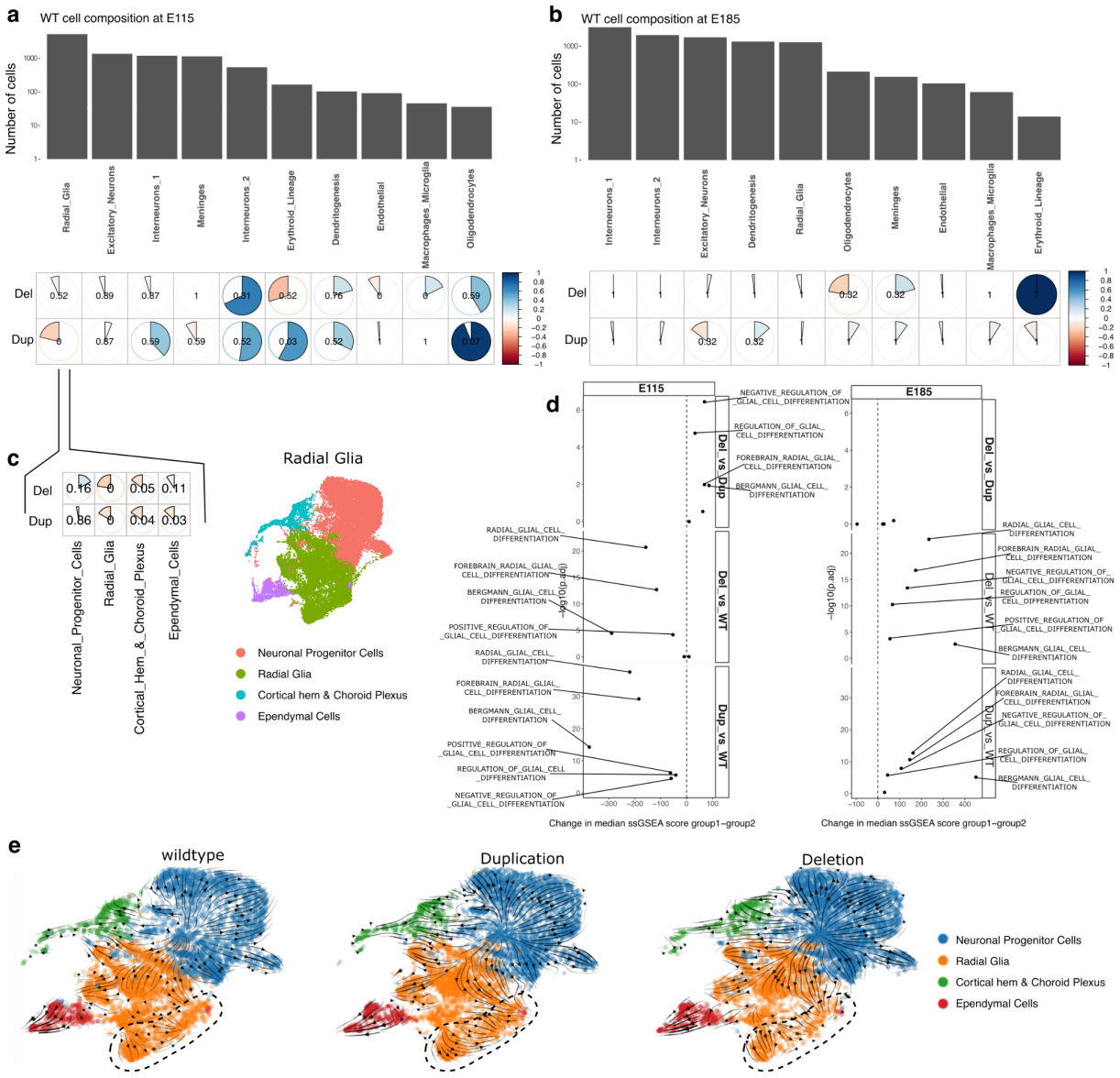


Figure 5.14 Changes in cell composition and difference in stage dependent glia differentiation
a,b, Pie charts of cell proportions changes in the mutants (Dup= duplication, Del= deletion) against the wildtype (**a,** time point E11.5; **b,** timepoint E18.5) per main trajectory ordered in size of trajectory, numbers of cells per trajectory are shown above. The color and direction of each pie correspond to whether log2-transformed ratio is gained above 0 (blue, clockwise)) or lost (below zero (red, anticlockwise)). Log2-transformed ratios that were above or below 0 were manually set to 1 or -1. All p-values are depicted on the respective pie. **c,** Pie chart and UMAP of radial glia sub-trajectories in the E11.5 dataset. **d,** Mean GSEA score comparison of C5 ontology gene sets including Glia Differentiation extracted by applying the search term "C5,GOBP.+_glial_cell_di" **e,** RNA velocity plots of radial glia trajectory E11.5 wildtype, duplication and deletion mutant cells (left to right) coloured by sub-trajectories.

5.2.3 Morphological phenotype of the aging mouse mutants

Since ADLD is defined as a late onset disease we next wanted to investigate potential brain malformations in the adult mutant brains. For that we applied magnetic resonance imaging

(MRI) on aging duplication, deletion and wildtype mice. The ages of the male mice ranged between postnatal day P410-P440 (59-63 weeks) roughly equivalent to human age of 47-52 years according to Jackson Laboratory ("Life Span as a Biomarker" n.d.). The brain MRI included a T2-weighted scan to analyze potential anatomical differences and diffusion tensor imaging (DTI) scanning to evaluate microstructural differences between the different mutants, as well as the wildtype. Note that the sample size per genotype (deletion n=4, duplication n=3, wildtype n=4) does not allow for meaningful statistical conclusions and requires follow up experiments with increased numbers per genotype. The first analysis of the data pointed towards changes between the mutants in brain morphology and diffusion properties (**Fig. 5.15a,b**). We observed a difference in the mutants' lateral ventricle size (1.01 ± 0.18; 0.29 ± 0.09; t value= 6.23; p value= 0.001), with the duplication mutant's being significantly smaller compared to the deletion as displayed in the example T2 scans (area of lateral ventricle marked with red arrows) (**Fig. 5.15a**). To investigate potential changes between the mice in more detail we investigated the white-matter microstructural integrity in axial diffusivity (AD) maps from the dMRI scan (**Fig. 5.15b**). These maps have previously been used to shed light into axonal degeneration and gliosis in leukodystrophy in patients (van Rappard et al. 2018).
Voxel-based analysis of AD maps showed a significant decrease in the duplication mutant compared to the deletion mutant (t-test, cluster-level corrected) and wildtype (t-test, cluster-level corrected, see colored areas within the respective brain maps) (**Fig. 5.15b**)**.** Interestingly, we found similar effects in both comparisons. The lateral ventricle, striatum and the periventricular hypothalamic nucleus preoptic and intermediate part are significantly altered in the duplication compared to the deletion mutant and the wildtype regarding AD map analysis (**Supplementary Table 7.3.3**). These effects partially overlap with the phenotype found in cell composition of the E11.5 single cell data discussed earlier, which might hint towards differences between the mutant phenotypes appearing in cell composition and morphological phenotype.

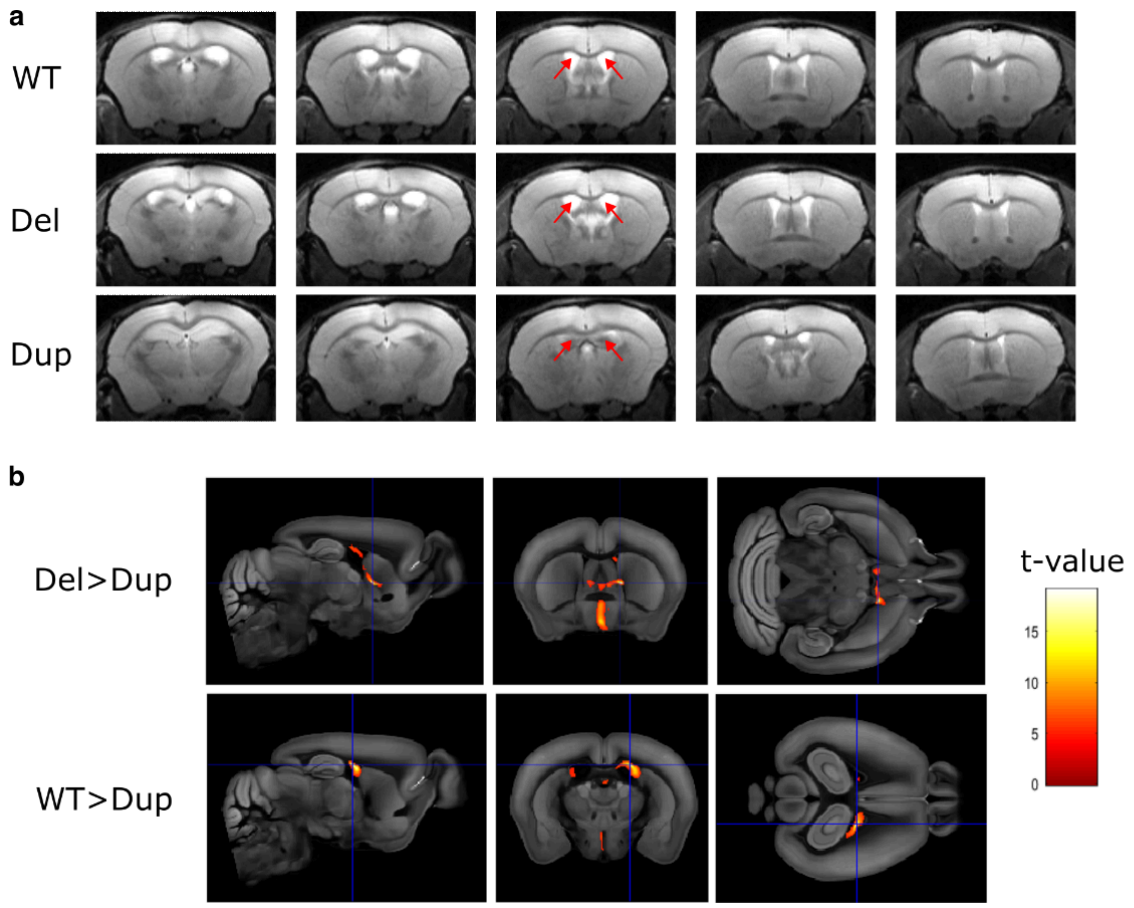


Figure 5.15 MRI based differences in aging mouse brains
a, T2-weighted scan examples of the investigated brains of the different conditions (wildtype, deletion mutant and duplication mutant). The volume of the lateral ventricles (red arrows) was significantly different between the deletion mutant and the duplication mutant. **b,** Voxel-based statistics of diffusion tensor imaging derived axial diffusivity maps. Voxels with significant contrasts of deletion vs. duplication and wildtype vs. duplication are shown color-coded (cluster-level corrected t-test, color marks t value).

6. Discussion

6.1 Summary of the studies

The exploration of phenotype development in models such as the laboratory mouse is a helpful tool to thoroughly investigate processes in disease development. Most of the available phenotyping methods are laborious and time consuming, and a combination of multiple methods is required to arrive at conclusive results. In this work, we present single cell sequencing as a new phenotyping tool to decipher the molecular basis of developmental disorders. For this we applied the tool in two studies: the ADLD project and the MMCA project. In two different settings, we used scRNA-seq to phenotype mouse mutants, aided by the development and application of novel computational tools.

In the MMCA project, we aimed to test the sensitivity of the scRNA-seq technology to detect phenotypes of mouse mutants ranging in variant severity. Additionally, we developed new analytical tools to achieve sufficient granularity in characterizing those phenotypes. In the subsequent section, we will delve deeper into the presented results, explore the limitations of the technology and experimental approach, and discuss the possibilities that they hold for the future.

A central question that I aimed to address in the ADLD project is whether structural variants such as TAD boundary deletions can result in neurodegenerative disorders. Importantly, these disorders are highly prevalent in society and present multiple challenges in the process of diagnosis, treatment, and research on disease progression. This is partly due to the fact that neurological disorders, especially neurodegenerative disorders, are often diagnosed late in disease progression, and clinical phenotypes are preceded by the cellular and molecular changes in the disease progression, making the treatment and dissection of primary disease causing mechanisms difficult.

In the following section, I want to discuss the results gathered from the current research on the neurodegenerative disorder ADLD. Further, I want to examine the general difficulties of the research on neurodegenerative disorders and how the findings are contributing to the definition of neurodegenerative disorders as a category compared to neurodevelopmental disorders.

6.2 Summary of the MMCA scRNA-seq dataset

The mouse mutant cell atlas (MMCA) consists of 22 mouse mutants and their respective wildtype background at embryonic stage E13.5. The mutants were selected based on their

predicted or published severity, ranging from pleiotropic phenotypes to the deletion of non-coding elements near developmental genes. By applying scRNA-seq to whole embryo samples, our aim was to challenge the technology's capability to detect granular changes across trajectories within the mutants for a comparative and scalable analysis. For this dataset, we developed two novel analytical approaches, namely the cell composition analysis (**Fig 5.3**) and the lochNESS analysis (**Fig. 5.5a**). These tools allow for fast and easily accessible detection of phenotypic changes in cell composition and local differences in gene expression between the mutants and their wildtype backgrounds. With this, we wanted to establish the scRNA-seq method as a scalable tool for systematic phenotyping of mutant mouse models of human disease and developmental variants.

Overall, our approach yielded promising results. We were able to expand the knowledge of previously captured phenotypes in greater detail, as for example in the retinal phenotype of the *Ttc21b* KO mutant (**Fig. 5.3; Fig. 5.4 a,b**). By analyzing the whole embryo, we could investigate how multiple trajectories are involved in the phenotypic changes detected in the eye of this mutant.

Moreover, through the unbiased approach of analyzing all trajectories, we could detect previously unknown phenotypes such as morphological differences in roof plate derivative ChP within the *Gli2* KO mutant (**Fig. 5.5c; Fig. 5.6 c,d**). Through combining the computational tools, we were further able to decipher multiple effects within the *Sox9* regulatory INV mutant (**Fig. 5.7**). This mutant has been previously described to carry an inversion of multiple regulatory elements away from the pleiotropic gene *Sox9* to ectopically express the neighboring TAD located gene *Kcnj2*. In addition, by looking at the skeletal preparations of the mutant, it was revealed to carry defects in bone formation (Despang et al. 2019). Our approach allowed us to expand on the previously described phenotype, connecting the skeletal phenotype with a stalling mesenchymal trajectory within this mutant (**Fig. 5.7c,d**). Moreover, because we captured the whole embryo, we were able to extend the analysis to other trajectories and detect upregulation of *Kcnj2* expression in other tissues than those related to the captured ossification phenotype (**Fig. 5.7e**). Interestingly, most of these tissues did not show the expected trend of *Sox9* downregulation. This could either be explained by an insufficient amount of cells captured by this approach which are affected by the expression changes in the mutant or by compensatory effects from regulatory elements not involved in the inversion contacting *Sox9*. As discussed before, in some cases multiple enhancers are known to regulate a gene's expression (**Fig. 1.1**). *Sox9* is an example for such a complex regulatory landscape, as about 52 enhancers have been found within the *Sox9* TAD (Despang et al. 2019). These enhancers hold the potential of regulating *Sox9* expression in an additive, synergistic or hierarchical manner within the different tissues

where this pleiotropic gene is known to play a role in (Yao et al. 2015; Scott et al. 2010; T. Wagner et al. 1994; Rockich et al. 2013).
Overall, by applying one scRNA-seq experiment to multiple mutants and building an analytical framework for this type of data, we could demonstrate that this technology is capable of detecting phenotypic changes within the cell composition and transcriptome within multiple mutants and on a granular level.

6.2.1 Limitations of the MMCA scRNA-seq dataset

This dataset is the first in which multiple whole embryo mutants were compared within one experiment. Thus, there are some notable limitations to our experimental design.
First, capturing 10.000 to 40.000 cells per mutant replicate entails the risk of missing mutant specific changes within smaller trajectories. Although we were able to capture and annotate 59 trajectories and explore the sub-trajectories such as the defects in multiple retinal trajectories of the *Ttc21b* KO mutant (**Fig. 5.3, Fig. 5.4 a,b**), we cannot exclude that other minor cell types are insufficiently represented in the dataset.
Second, we are exclusively capturing the developmental time point of E13.5. This dataset therefore is a snapshot within the development of the mutants and only allows to draw conclusions about the phenotypic differences within those mutants at that particular stage. This is critical to consider when investigating certain mutants such as in the *Cdkl5* -/Y (hemizygous) mutant which is modeled after a human disorder causing neurological dysfunctions in patients (**Fig. 5.3**). *Cdkl5* encodes for a protein associated with synapse outgrowth and dendritic connection (Ricciardi et al. 2012; Zhu and Xiong 2019). The phenotype of patients with *Cdkl5* deficiency disorder is variable, but clinical features are closely related to the Rett syndrome, including early-onset infantile epilepsy, intellectual disability and autism (Ricciardi et al. 2012). In our dataset, the neural tube trajectory contains a diversity of developing neuronal cell types (**Supplementary Figure 7.3**), yet it is not known whether at embryonic stage E13.5 neurons are differentiated to the extent of synaptic maturity and dendrite development to reflect the phenotypic changes of disease development in the mutant *Cdkl5* -/Y. Because the process of brain development extends well beyond birth in mice and in humans, the question arises whether the phenotype can be captured within the selected time point E13.5 (Lemaître et al. 2021; Dubois et al. 2014; Hammelrath et al. 2016).

Relating to the point made above, one might argue that the mutants used in the study could have been selected in a more streamlined way, for example only including mouse mutants that carry knockouts of genes involved in critical developmental steps active at E13.5.

Instead, we included samples carrying a wide range of different mutations, such as knockouts, missense mutations, TAD boundary deletions and enhancer deletions among others (**Fig. 5.1a**). These mutations target or are associated with a great variety of genes with different functions in development. For instance, some genes like *Sox9* have pleiotropic functions across multiple tissues, while others such as *Scn11a* are associated with a very specific sodium ion channel which has a role in pain perception (Leipold et al. 2013; Despang et al. 2019). This made downstream analysis of the dataset more challenging than if we would have chosen a coherent series of knockout mouse models. On the other hand, the selection challenged us to build dynamic analytical tools such as lochNESS, which visualizes granular changes within less severely affected mutants and generates an easily accessible overview via the lochNESS trajectory distribution (**Fig. 5.5b,c**).
As this dataset is the first mouse mutant single cell atlas, we aimed for a broad representation of mouse models, thereby enabling the scientific community to draw conclusions from our data, which might help to decide whether a single cell experiment is the right approach for the mouse mutant of their interest. In addition, we included four different wildtype mouse strains, which are commonly used to generate mouse models. This gave us the opportunity to report differences in wildtype pseudo bulk profile (**Fig. 5.2a**) further reinforcing the need to include a well chosen wildtype control in the scRNA-seq design and additionally providing wildtype controls for other scientists to source from in future experiments.

Another limitation is the fact that we are selectively investigating the transcriptome of the mutants by using scRNA-seq. As discussed before, there are multiple single cell methods developed up to this date. One could argue that the application of a combined analysis of both RNA and ATAC profiles by using a multi-omic approach might have enabled us to gather more information about the mutants. However, for the chosen method, namely sci-RNAseq3, there are no available multi-omic approaches yet. Choosing a commercially available multi-omic approach such as the 10x Genomics approach would have limited the sample size to 10.000-50.000 captured cells within one experiment, compared to the over 1.5 million cells we were able to confidently capture with sci-RNAseq3 (https://www.10xgenomics.com/support/single-cell-multiome-atac-plus-gene-expression).

Further, the dataset is comprised of 4 replicates per mutant, which is quite few to allow statements about the significance of cell composition changes. Sample availability is a focal limitation factor when working with mouse models in comparison to other model organisms such as C. Elegans and Drosophila, as mouse generation time is slower and offspring numbers are lower. In our case, some of the mutant mouse lines were established over 20

years ago and the recent crossings were not generating a lot of offspring, limiting the number of available samples to 4 replicates per mutant.

Based on a simulation analysis, these numbers are sufficient to detect minor changes in bigger cell types, but only stronger changes of cell composition in smaller cell types. For detecting subtle changes in smaller cell types application of lochNESS is advised, which is based on the embedding of cells within the trajectory, rather than on counts as it is the case in the cell composition analysis.

We confirmed the MMCA trajectory annotation through correlation with the mouse organogenesis cell atlas (MOCA) in a non-negative least-squares (NNLS) regression analysis (**Supplementary Figure 7.5**) (Cao et al. 2019). Yet, inconsistencies like in the limb mesenchymal trajectory annotation, which, as indicated in RNA velocity, is the origin of all other mesenchymal trajectories, were emerging in the downstream phenotype analysis (**Fig. 5.2a**). In this case, it is highly likely that the general mesenchymal progenitor was mislabelled as limb mesenchymal trajectory. This has recently been discussed as a risk of reference atlas based annotation (Domcke and Shendure 2023). Further, the authors propose the creation of reference trees for the annotation of developing organisms including information about the molecular state and history of lineage development. These reference trees would help to achieve consistent and transferable annotations for the vastly growing amount of single cell datasets. At the moment, there are an abundance of different annotations for dynamic developmental trajectories within the field that are not coinciding with one another, complicating annotation of new single cell datasets and phenotype comparison between different experiments.

Since single cell technologies are a comparably recent tool, all phenotypic findings, especially the ones that are unexpected, require further experimental validation to verify the results found in the study. All the phenotypic findings in our study were supported by validation experiments. Because this study introduces a new strategy to phenotype mutants, the requirement of further validation is absolutely reasonable, yet the final goal is to be able to apply scRNA-seq as a verified method for mutant analysis.

One reason for the need of follow-up experiments is missing information about the localization of scRNA-seq identified aberrant cells. UMAP embeddings provide information about transcriptional identity of the cells through proximity based visualization of the data points, yet the localization of cells within the sample typically is validated through IF (*Gli2* KO) or RNAscope (*Sox9* regulatory INV) (**Fig. 5.5d; Fig. 5.6b,c,d; Fig. 5.7c,d**). Recent advances in spatial transcriptomics have enabled the first “digital in situ” of wiltype whole mouse embryos (Srivatsan et al. 2021). Application of spatial transcriptomics in developing mouse mutant embryos would enable the visualization of morphogenesis and reconstruction of cell to cell interaction to decipher the local deformation of tissues leading to the phenotypic

manifestation (Veenvliet et al. 2021). However, this method is not yet applicable at atlas-scale.

In summary, this study provides evidence of the successful application of scRNA-seq as a phenotyping tool to gather unbiased and whole organism information about cell composition and gene expression changes. This study further provided the first analytical framework to analyze multiple mutants in parallel and on a granular level.

6.2.2 Perspectives of the MMCA scRNA-seq dataset

With the experiment, we generated a large-scale dataset containing more information than can be dissected within the framework of a single publication. Using the cell composition analysis alone, most mutants displayed interesting changes in cell numbers, worth further exploration (**Fig. 5.3**). Hence, we aim for the scientific community to continuously explore and make use of the generated data. To enable this, we generated an interactive browser for easy access to our data (https://atlas.gs.washington.edu/mmca_v2/). The browser offers the direct exploration of trajectories, gene expression profiles and lochNESS within all mutants from this project. Further, the computational tools built for the analysis and the scRNA-seq dataset itself are accessible on this website for an in depth exploration of the data.

By generating datasets like this, we hope to initiate the generation of a comparative database of as many mouse mutants as possible using scRNA-seq. Ideally, this approach could be applied to generate an atlas of all mouse models. The expansion of scRNA-seq to more mouse mutants would drive the development of new analytical tools, which would generate comparable approaches to analyze the transcriptome of mouse mutants generated by different labs, compare similarities between different genotypes and explore the subtle differences in similar genotypes as in the example of the *Atp6v0a2* KO and *Atp6v0a2* R755Q from our study (**Fig. 5.3**).

Notably, a recent study presented the successful integration of datasets capturing gastrulation and organogenesis of the mouse embryo generated by different scRNA-seq experiments (Qiu et al. 2022). This approach demonstrates that careful integration of datasets generated through different methods is possible. Hence, the usage of different methods to generate whole embryo phenotyping data should not present a barrier in the aim to generate a database of mouse mutants from different whole embryo single cell phenotyping experiments. Another step would be the expansion towards a multi-omic database of mutants exploring accessible chromatin, protein and expression of the mutants in integrated databases.

A limitation to create such a database is that single cell experiments remain very cost extensive and are therefore not an accessible research option for a big part of the scientific

community. One option to overcome this hurdle is to combine efforts by several labs to investigate multiple mutants within one experiment. As calculated for sci-RNAseq3 specifically, the cost per cell is around $0.03-$0.20, which might be even more reduced by upscaling the experiment up to 10 million cells, an option achievable by combining multiple mutants from several research groups (Cao et al. 2017). Further, we hope that by continually making this data available to the scientific community, we can contribute to leverage the playing field, thereby enabling broader access to large datasets and cutting-edge technologies.

6.3 Summary of the findings in the ADLD project

In patients, the neurodegenerative disorder ADLD has been described to result in autonomic, cerebellar, and pyramidal changes and finally cognitive symptoms (Quasar S. Padiath 2019; Terlizzi et al. 2016; Raininko, Gosky, and Padiath 2016; Finnsson et al. 2015). The disorder is caused by a duplication of the gene *Lmnb1*, resulting in *Lmnb1* overexpression (Giorgio et al. 2013; Quasar S. Padiath et al. 2006; Brussino et al. 2009). More recently, some patients were described to carry deletions upstream of the gene (Brussino et al. 2010; Giorgio et al. 2015; Nmezi et al. 2019). A hypothesis on how these deletions could cause similar symptoms postulates that the deletions include a TAD boundary which consecutively results in enhancer hijacking from the neighboring TADs (Giorgio et al. 2015). However, the hypothesis was not further investigated to this date. In this project, I wanted to address the mechanism of disease in the deletion variant. Further, a central aim was to address at what point in time the molecular phenotype appears in the variants, and whether those changes precede the clinical phenotype described in patients.
To investigate this, I generated two mouse lines with CRISPR based mutations, designed after a duplication and a deletion found in patients. I subjected embryonic brain samples to capture HiC (cHiC) and scRNA-seq with 10x technology. Further, I subjected adult mice to magnetic resonance imaging (MRI) to see whether the variants induced any morphological changes in the brain during aging.

6.3.1 Capture HiC results require follow up exploration of the locus

As discussed above, E11.5 brain cHiC maps show chromatin changes in the 3D organization of the *Lmnb1* - *Zfp608* landscape in the deletion mutant. These changes lend support to the proposed mechanism of boundary deletion and enhancer hijacking responsible for the *Lmnb1* misexpression. Nevertheless, this data lacks detailed information about the exact enhancer and boundary elements involved in the deletion-based misregulation.

To identify the specific boundary causative of *Lmnb1* misexpression, one option would be to delete the local minimums of insulation score detected in the wildtype cHiC between *Lmnb1* and *Zfp608* indicative of a boundary (**Fig. 5.9**). The definition of the exact boundaries could be further supported by a CTCF ChiPseq, a readout for localization of TAD boundaries.

However, other publications demonstrate that the deletion of a single boundary element has no effect in TAD configuration and causes none to little changes in gene expression (Despang et al. 2019; Rajderkar et al. 2023). In most cases where deletions of TAD boundaries result in disease, a significant region additional to the boundary was deleted (Lupiáñez et al. 2015; Flöttmann et al. 2015). This could be due to the need for enhancers to get into relative proximity to the target promoter for contact formation which is counterintuitive to the knowledge about long-range contact between enhancers and their target promoters (Lettice et al. 2014; Kvon et al. 2016). Hence, the targeted deletion of a boundary element is not expected to result in major changes within the 3D structure of the locus or gene expression. Also, it is important to note that the cHiC maps display an average contact across all cell types. This means that non-neuronal cell types such as the meninges, which contribute around 1000 cells to the E11.5 scRNA-seq dataset, also influence the boundary profile displayed in our insulation score. As stated in other studies, TAD organization varies significantly in between different cell types (McArthur and Capra 2021; Du et al. 2017; Winick-Ng et al. 2021). In order to identify the boundaries involved in neuronal 3D chromatin architecture of the locus, it would be necessary to select or enrich these cell types as a step prior to the capture HiC experiment. This way we would be able to visualize the precise boundaries targetable in the disease.

Further, the identification of the specific enhancer elements involved in the proposed enhancer hijacking by *Lmnb1* in the deletion mutant would require a series of CRISPR knock outs of enhancers in the *Zfp608* TAD, followed by *Lmnb1* expression analysis. A limiting factor for this experimental set-up is that mutants are heterozygously carrying the variants, designed after heterozygous variants in patients. Thus, it would require extensive CRISPR experiments followed by locus sequencing to ensure deletion-allele specific targeting of regulatory elements.

An alternative strategy to investigate the regulatory elements could be the inversion of a part of the regulatory landscape including a boundary, similar to the *Sox9* regulatory INV mutant. Introducing a balanced variant into the mouse genome at this locus would solve two open questions remaining from the current locus analysis (**Fig. 6.1**). First, we could dissect which of the detected local minimums in the insulation score is involved in the boundary by solely including one of them in the respective inversion. Second, we could perform gene expression analysis to investigate whether the regulatory elements introduced into the neighboring TAD are causing *Lmnb1* misexpression.

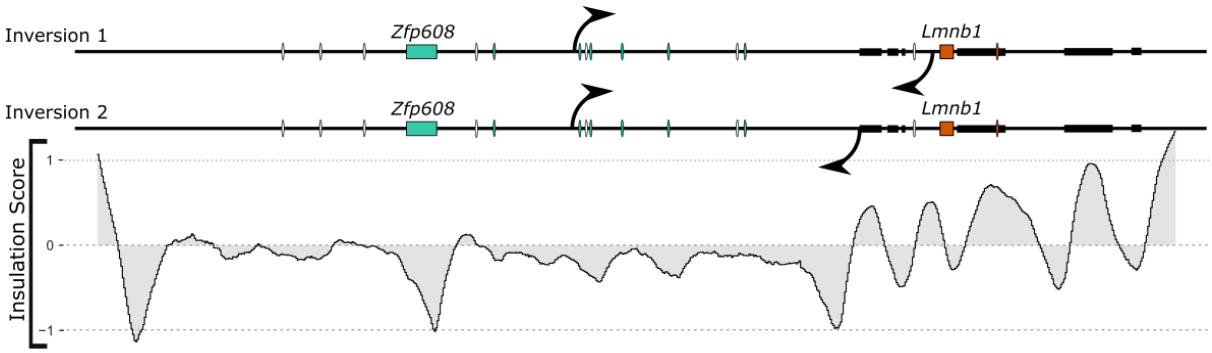


Figure 6.1 Inversion mutants for further analysis of the locus
Two proposed balanced mutations (Inversion 1 and Inversion 2) are schematically visualized by arrows positioned at the potential breakpoints for inversion. Below, the insulation score, calculated from the wildtype cHiC contact map is visualized: Inversion 1 would include both minimums of the insulation score while Inversion 2 would only include the stronger minimum in the inversion.

Altogether, the cHiC found 3D conformation changes are a good indication for the mechanism of disease onset in the deletion mutant, further supported by scRNA-seq data. These findings could be extended to find the exact regulatory elements involved in the misregulation of *Lmnb1* expression by performing follow-up dissection of the locus.

6.3.2 Structural variants result in early molecular changes within both mutants

As described above, unlike the duplication, the deletion mutant causes a distinct alteration chromatin structure of the *Lmnb1* locus. Specifically, the cHiC contact map visualizes the *Lmnb1* gene looping towards the neighboring TADs regulatory elements (**Fig. 5.10; Fig. 5.11; dotted circles**). This finding, together with the new boundary composition of this mutant, feeds the hypothesis of a boundary deletion and enhancer hijacking with experimental evidence (**Fig. 5.9; Fig. 5.11**). In this, the *Lmnb1* gene can now contact enhancers from the neighboring TAD, which are shown to regulate expression of *Zfp608* in different areas of the developing mouse brain as shown in the VISTA enhancer LacZ stainings (**Fig. 1.7**).
By investigating the scRNA-seq data of the duplication and the deletion mutants brain development, we could detect significant upregulation of *Lmnb1* expression in both mutants (**Fig. 5.12d**). Upregulation of *Lmnb1* expression is known to be the molecular marker of disease development in ADLD (Neri et al. 2023), so capturing the overexpression in an early developmental stage is a central finding in our study. Moreover, the deletion mutant displayed significant downregulation of *Zfp608* expression (**Fig. 5.12d**) hinting towards a manipulation of the regulatory elements originally wiring to *Zfp608* in the deletion, thereby supporting the 3D structural changes found in this mutant.

Taken together, these results are a first indication of early onset molecular changes as well as differences between the deletion and the duplication variant in disease progression.

6.3.3 Single cell RNA-seq analysis reveals cholesterol and glia differentiation difference in the mutants

The scRNA-seq dataset was further dissected in downstream analysis. In the radial glia trajectory, we could detect changes in cholesterol biosynthesis in the mutants compared to the wildtype sample (**Fig. 5.13d**). These changes varied between the mutants as some steps in the biosynthesis of cholesterol were downregulated in the duplication mutant and other intermediates were upregulated in the deletion mutant (**Fig. 5.13e,f**). Previous research on *Lmnb1* overexpression connected the molecular phenotype to lipid and cholesterol pathway changes in an oligodendrocyte specific overexpression mouse model (Rolyan et al. 2015). We could expand this phenotype to cholesterol biosynthesis defects in the radial glia to mutants that carry genomic alterations mimicking those found in patients. Our findings expand on the cell types involved in the disease development and emphasizes differences in disease progression in the two variants.

To our surprise, there were only few significant cell composition changes in the main trajectories within the E11.5 and none in the E18.5 dataset (**Fig. 5.14a,b**). One reason for this could be that slight developmental malformations in early development could be corrected for; another might be insufficient sample numbers, which will be discussed later.

One of the trajectories altered within the E11.5 is the radial glia trajectory in the duplication mutant. The sub-trajectories of the radial glia main trajectory showed significant reduction in cell numbers in the radial glia as well as cortical hem & choroid plexus in both mutants and a significant change within the ependymal cells within the duplication (**Fig. 5.14c**). Since astrocytes and oligodendrocytes, both descendants of the radial glia, have been previously investigated to contribute to disease progression in ADLD, we further looked into glia differentiation (**Fig. 5.14d**). While glia differentiation pathways are downregulated in the E11.5 mutants, they are upregulated in the E18.5 mutants, suggestive of a change in trajectory gene set expression between the two time points. This is another lead towards a switch in gene expression between the mutant to compensate for mutation induced alteration in the mouse embryos. This hypothesis would serve as an explanation for the differences detected in cell composition between E11.5 and E18.5 and the reversed profiles of glia differentiation gene sets in the time points.

6.3.4 MRI reveals morphological changes within the duplication mutant

MRI of adult mouse brains revealed a significant reduction of choroid plexus volume in the duplication mutant compared to the deletion mutant (**Fig. 5.15a**). This partially supports the changes found in cell composition of embryonic scRNA-seq analysis, as the trends in scRNA-seq mark a loss of this tissue in both the deletion and the duplication mutant (**Fig. 5.15a; Fig. 5.12d**). Since the time points as well as experimental conditions of MRI and scRNA-seq differ significantly, a direct comparison between the scRNA-seq and MRI results would be highly speculative.

The choroid plexus is involved in adult neurogenesis and gliogenesis, central nervous system immunity and production of cerebrospinal fluids (CSF) which has a crucial role in nutrient supply, protection and removal of waste products generated in the brain (Silva-Vargas et al. 2016; Thompson, Brissette, and Watt 2022; Wichmann, Damkier, and Pedersen 2022). Reduced choroid plexus function and CSF production has already been explored in the context of amyloid-beta peptide accumulation in Alzheimer's disease, another neurodegenerative disorder (Gião et al. 2022; Kant et al. 2018). Based on these findings and our results in the MRI and scRNA-seq of the ADLD mouse models, the choroid plexus presents an interesting tissue to investigate the potential accumulation of inflammatory components in ADLD onset and progression. Further, MRI axial diffusion maps generated by diffusion tensor imaging (DTI) scans provided additional evidence for structural changes within the duplication mutant (**Fig. 5.15b**). The maps show a significant decrease in several tissues of the duplication mutant compared to the wildtype and the deletion mutant. This result further underlines that changes in the duplication mutant are stronger than in the deletion mutant, resulting in morphological changes in this stage of disease progression.

Taken together, our analysis of the ADLD patient derived mouse models provide new evidence of cell type specific composition changes, differences in gene expression and morphological differences, worth further exploration in the context of differential disease development in the two variants.

6.3.5 Directions of exploration in the ADLD scRNA-seq dataset

The scRNA-seq dataset, comprising the brain of the two mutants and its corresponding wildtype at two developmental time points, produced interesting findings that require further exploration to fully exploit the available information.

As an example, the cell composition analysis revealed that the radial glia sub-trajectories showed significant alterations in the deletion mutant which were not discovered in the main trajectory analysis (**Fig. 5.14a,b**). This indicates that there could be interesting changes in

other sub-trajectories, raising the need for cell composition analysis in the sub-trajectory of both time points.

The choice to further investigate radial glia was hypothesis driven, based on previous knowledge about cell types involved in ADLD (Rolyan et al. 2015; Ratti, Rusciano, Mongiorgi, Owusu Obeng, et al. 2021; Heng et al. 2013). One of the protruding advantages of scRNA-seq is the unbiased analysis of all cell types involved in the tissue at hand. This opens the opportunity for the investigation of previously disregarded cell types. Although unlikely to play a role in the brain related disease development of ADLD, the non-neuronal cells should be investigated as well, especially since *Lmnb1* is known to play an ubiquitous role in the establishment of the nuclear lamina across cell types. So there might be unexplored yet noteworthy changes in gene expression unrelated to the primary neuronal phenotype.

Ideally, each condition would include more than two replicates to strengthen statements about the significance of the cell composition changes observed. This, unfortunately, is restricted by the costs of 10x single cell experiments. For detecting smaller and local changes in the trajectories without increasing replicate numbers, application of the MMCA developed tool lochNESS presents a promising alternative.

Another direction which could be further explored is the RNA velocity of the E11.5 radial glia, which revealed the cellular dynamics of the samples marking potential differences in developmental direction (**Fig. 5.14e**). This finding would greatly benefit from further exploring the density plot differences between the samples in the radial glia. If implemented similarly to the MMCA project, the density plots would enable the visualization of potential differences in cell distributions across a trajectory (**Fig. 5.7c**).

In disease development morphological abnormalities are preceded by molecular and cellular changes. Cellular changes include a step in which cells exit the healthy trajectory to develop aberrantly based on the different gene expression program. This moment in disease development is described in a recent review as critical to capture in order to understand and target the cellular origin and cause of disease (Rajewsky et al. 2020). For instance, *Lmnb1* overexpression might not be specific to the disease causing cell types, but could be compensated for in non-disease related cell types by regulations such as post-transcriptional modification. Our conclusions, which, at the moment, are solely based on molecular changes detected, could therefore be a skewed view on the disease mechanism. To capture cellular changes within etiology, the addition of another developmental stage to the dataset might be interesting to consider. However, the difficulty of finding the right samples representative of this stage is specifically tricky in degenerative diseases. Between the embryonic expression changes until slight morphological changes, both described in our study, lies over a year of

disease progression. One landmark worth considering is the coinciding phenotype of aberrant myelination in patients and mouse models described so far (Quasar S. Padiath 2016; Lin, Ptácek, and Fu 2011). This is in line with findings that the demyelination phenotype was stronger than axon loss, indicative of aberrant myelination being a primary feature in the disease (Coffeen et al. 2000).

Studies describing the myelination of the CNS in mice demonstrated that mature oligodendrocytes coat the axons of neurons with a lipid rich layer of myelin allowing for the rapid conduction of signals, starting predominantly postnatally (From et al. 2014; Bechler, Swire, and Ffrench-Constant 2018). This coincides with the relatively low numbers of mature oligodendrocytes cells present in our dataset (**Supplementary Figure 7.4**). Further, most of the primary myelination happening in the mouse brain, also known as intrinsic myelination, occurs within the first three weeks of postnatal development (Morell and Quarles 1999). Based on this and additional evidence that mature, active oligodendrocytes are a disease driving cell type, scRNA-seq samples around the three week window of mouse postnatal development might be a good addition to the dataset. This would enable the construction of a trajectory detailing early disease development using single cell technology.

Opposing this hypothesis, myelination is a continuous process that requires the life long maintenance of myelin sheath (Toyama et al. 2013). Additionally, myelination based on strong neuronal activity has been detected throughout the lifespan of humans and is considered adaptive myelination (Noori et al. 2020). Therefore, it remains impossible to define exactly when oligodendrocytes develop aberrantly between the first onset of molecular changes and the development of a morphological phenotype, leaving it up for further investigation whether ADLD observed demyelination is caused by initial intrinsic or adaptive myelination.

6.3.6 MRI requires replicates for significance statement

While the investigation of phenotypic aberrations in the brain using MRI has been investigated in various mouse models of neurodegenerative disorder, this method has never been applied in mouse models of ADLD (Jullienne, Trinh, and Obenaus 2022; Cowin et al. 2011; Underwood et al. 2011; Nathoo et al. 2013). Because MRI is a great tool to investigate inflammation and morphological changes, we set up a preliminary study with adult mice. We were pleasantly surprised to detect morphologic and diffusion differences in the duplication mutant (**Fig. 5.15**). But as previously discussed, these results have to be interpreted with caution. Typically, exploratory studies should contain a much higher sample size than the n=3 and n=4 reported here to warrant statements about the significance of the observed changes. Our current MRI dataset requires a follow up experiment with more individuals to

evaluate the trends. It is worth noting, however, that, contrary to the duplication mutants, the deletion mutants do not show any notable changes. This is interesting in the context that the primary molecular changes, such as the increase in *Lmnb1* expression as well as the cell composition changes that we explored in the radial glia, are both slightly stronger in the duplication. Taken together, the expansion of replicates in MRI analysis will considerably advance our knowledge of the mutants.

6.3.7 Lamin B1 involvement in global 3D genome architecture in disease development

It is known that neurons are a unique cell type in many regards. Unlike other cell types, they are believed to mostly exit the cell cycle into quiescence after being fully differentiated (Frade and Ovejero-Benito 2015). Although previously stated non-existent, there is adult neurogenesis in restricted areas of the brain, namely the subventricular and the subgranular zone, mainly involved in memory and learning (Ming and Song 2011). Interestingly, the aging brain underlies more dynamic changes than previously believed as demonstrated by a study showing the dynamic modulation of 3D genome architecture of neurons throughout the lifespan of humans and mice (Tan et al. 2023; Akbarian and Won 2023). Moreover, recent work in Alzheimer models proposed a role of Lamin B1, the protein encoded by the *Lmnb1* gene, in 3D chromatin changes involved in the neurodegenerative disease (Dileep et al. 2023).

Building on the known role of Lamin B1 in chromatin organization at the nuclear membrane and the obstruction of DSB repair upon its overexpression, the study reveals Lamin B1 misregulation to play a role in the accumulation of DSB and disruption of 3D genome architecture of neurons in disease progression (Solovei et al. 2013; Etourneaud et al. 2021; Dileep et al. 2023). Interestingly, the meta-organization of the chromatin into A- and B-compartments remained intact, demonstrating no compartment switches upon Lamin B1 misregulation. Further, a related publication described a decrease in Lamin B1 associated with global epigenetic dysregulations in the progression of Alzheimer's disease (Xiong et al. 2023).

Although in opposing trends compared to the observations in our study, these publications describe a global role of Lamin B1 in the progression of a neurodegenerative disorder, highlighting that this gene holds a crucial role in aging and deregulation of processes involved in brain related diseases.

6.4 Conclusion and Outlook

In conclusion, we present two different approaches to apply scRNA-seq as a phenotyping tool in mouse mutants. In the MMCA scRNA-seq experiment, we were able to highlight the strength of the technology to perform large-scale, whole embryo analysis of multiple mutants within one experiment. Using this technology in combination with newly built analytical tools, we could (I) deepen our knowledge about previously known phenotypes, (II) uncover new phenotypes, and (III) expand the cellular dynamics involved in the progression of malformations. In light of the limitations of the approach discussed in the sections above, the results discussed here support an introduction of the technology as a systematic, unbiased approach that enables phenotype discovery at an incomparable scale compared to conventional phenotyping methods.

The ADLD project is, contrary to the MMCA study, an in-depth analysis of one particular disease. In this study, we generated the first mouse models of the neurodegenerative disorder ADLD based on patient variants. Specifically, the deletion mutant reveals 3D genomic alterations that point towards a mechanism of TAD boundary deletion resulting in enhancer adoption from the neighboring TAD (**Fig. 6.2**).

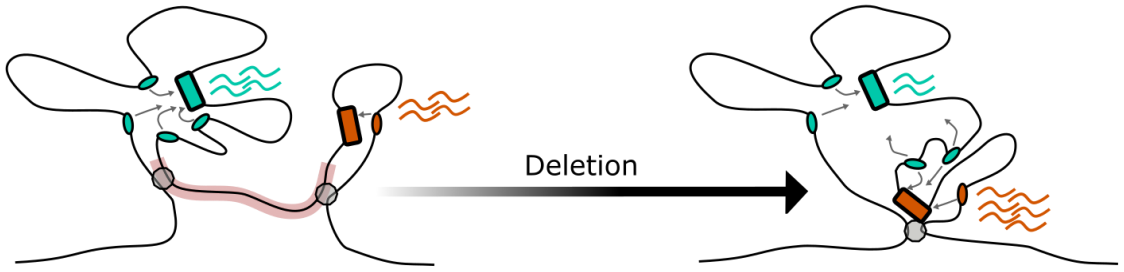


Figure 6.2 Model of the deletion variant

In a 3D model translated from the wildtype cHiC contact map of the *Lmnb1* and *Zfp608* locus, the gene *Zfp608* gets activated by multiple enhancers (light blue) and is separated from *Lmnb1* and its proposed endogenous enhancer (orange). Both genes are similarly expressed as indicated by the blue and orange waves. Our data suggests that, through the deletion of a region including boundary elements, *Lmnb1* gets relocated into proximity of *Zfp608* endogenous enhancers. Our model proposes that the mutation results in enhancer adoption, *Lmnb1* misexpression and *Zfp608* downregulation.

Strikingly, we could show that molecular changes appear already in the mutant's embryology. This finding somewhat challenges the categorization between early-onset neurodevelopmental disorders, such as attention deficit hyperactivity disorder (ADHD) or autism spectrum disorder (ASD), and late-onset neurodegenerative disorders, like Alzheimer's, Huntington's or Parkinson's disease, which are based on the onset of clinical symptoms. Similar to other studies discussing early molecular changes found in research

about Parkinson's and Huntington's (Schwamborn 2018; Ring et al. 2015), our study shows that molecular changes precede clinical symptoms in ADLD by far. With a notable rise in cases and absent treatments for clinically established neurodegenerative disorders, the detection of these diseases in the molecular and cellular stages is of high priority (Van Schependom and D'haeseleer 2023). Early detection involves increased research on biomarkers, development of treatments for molecular and cellular changes and finally, the re-categorization of neurodegenerative disorders into early onset. In order to change the category of neurodegenerative disorders from a fatal diagnosis into early detectable and treatable diseases, there is a strong need for building awareness. We hope to have contributed a small building block towards reaching that goal with this study.

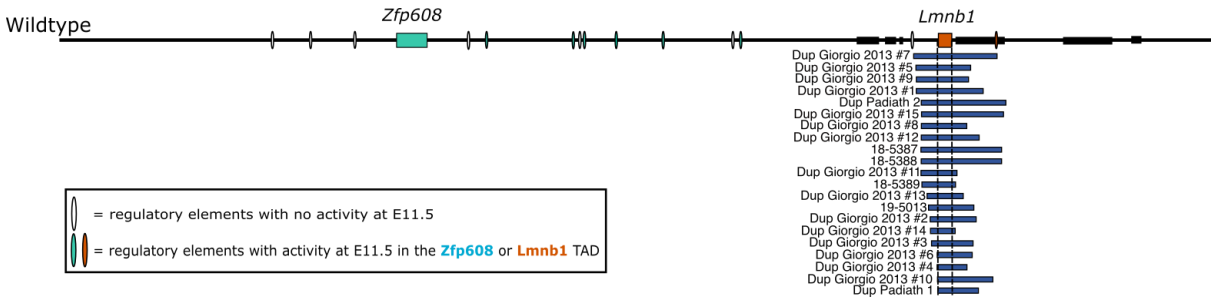
Nevertheless, the project requires further exploration of disease progression to decipher the correlation between the beginning of molecular changes in embryology and the resulting morphological phenotype involved in the mutants.

7. Appendix

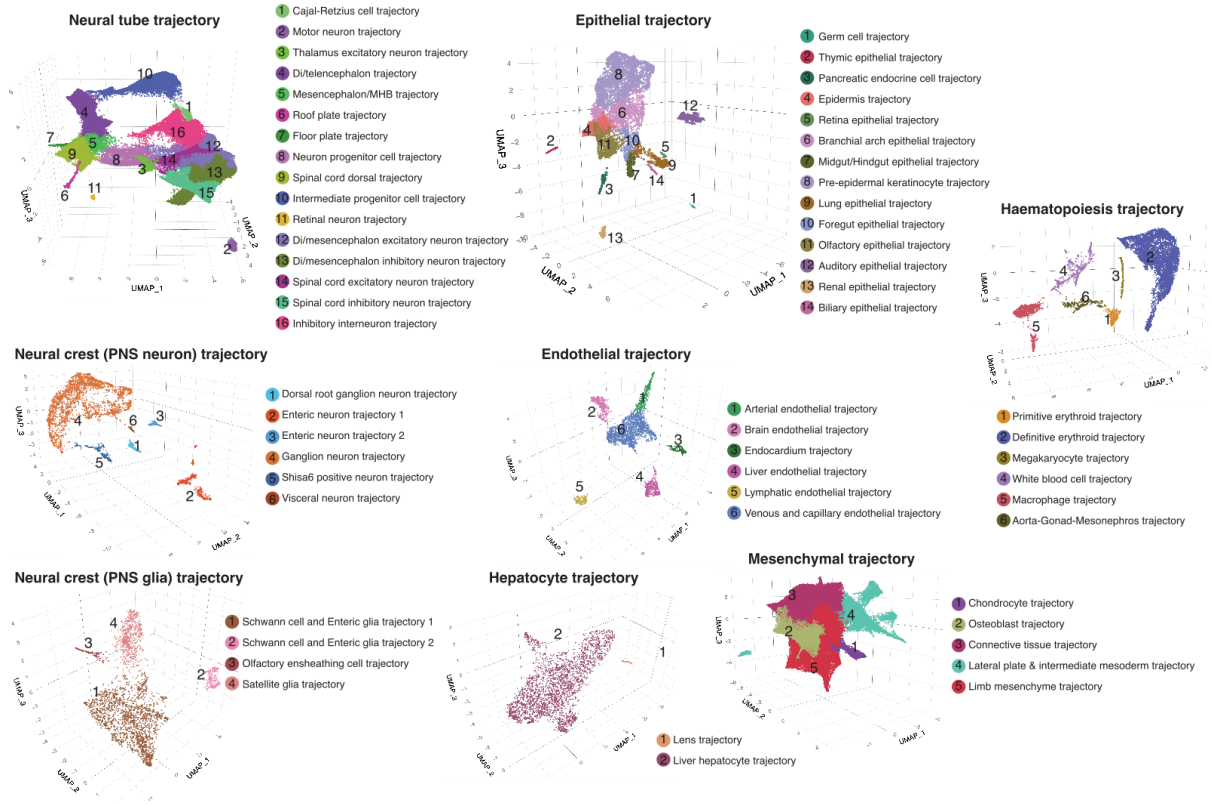
7.1 Plasmid Maps

Supplementary Figure 7.1 Vector Map of the pSpCas9(BB)-2A-Puro (px459) Vector

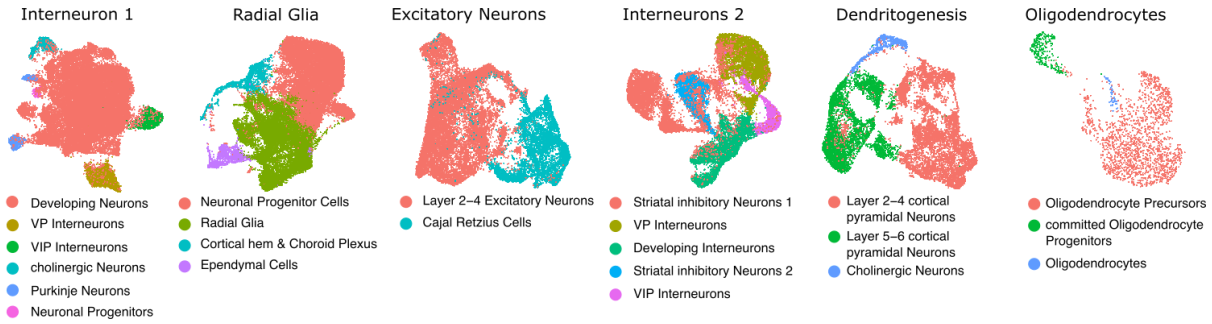
7.2 Supplementary Figures



Supplementary Figure 7.2 Duplication Variants from patients in the *Lmnb1*/*Zfp608* mouse locus Patient found duplication were converted to the mouse genome (mm10) using UCSC LiftOver. The minimal critical region is visualized in the dotted lines. Patient duplications are sourced from following publications: (Borja et al. 2022; Nmezi et al. 2019; Giorgio et al. 2015, 2013; Quasar S. Padiath et al. 2006).

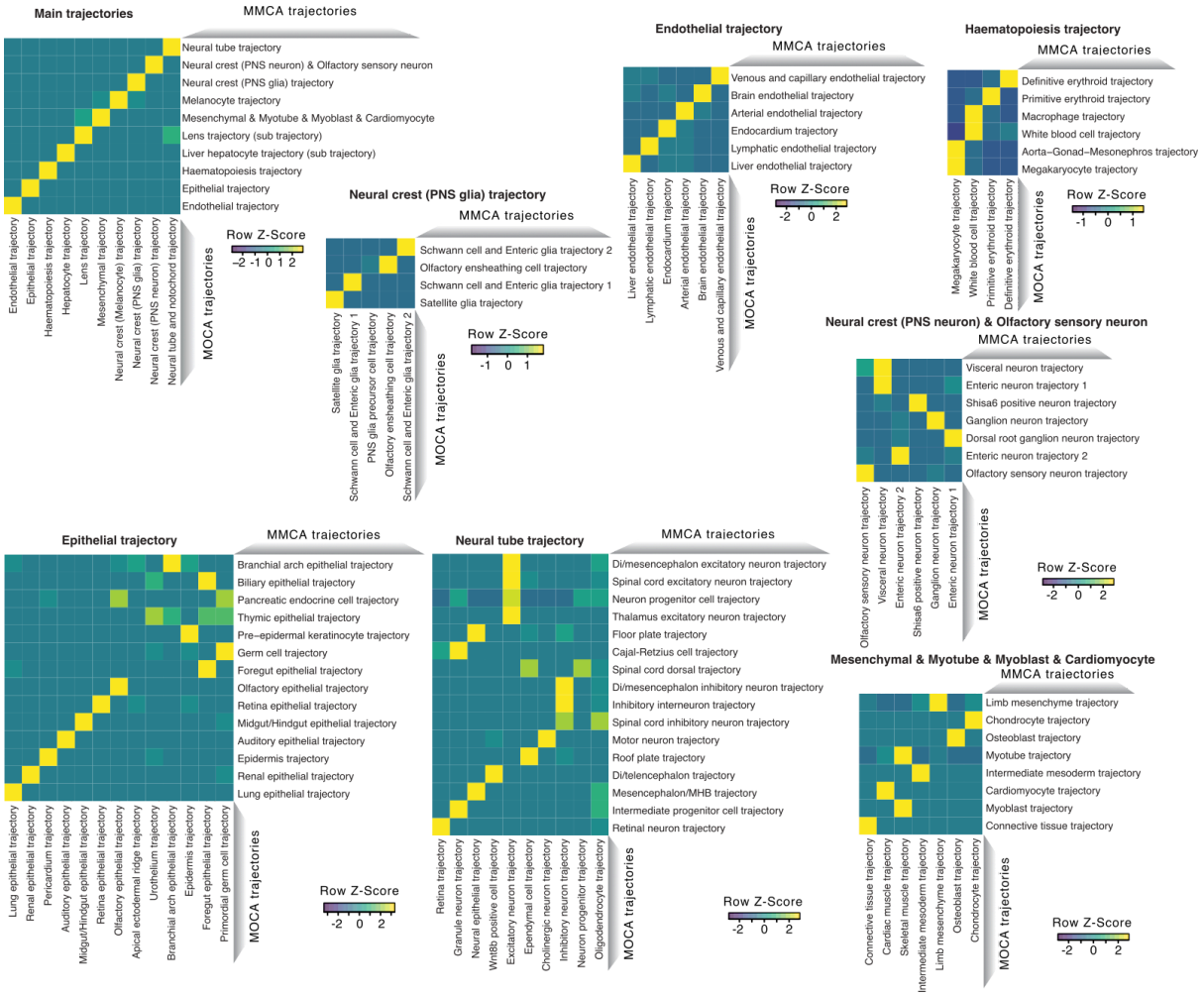


Supplementary Figure 7.3 Annotation of sub-trajectories of the MMCA E13.5 embryos
In the wildtype E13.5 MMCA dataset of 215,517 single cells we annotated 13 major trajectories and for 8 of them we identified further sub-trajectories displayed in UMAP embedding. The clusters are colored by their annotations.



Supplementary Figure 7.4 Annotation of the neuronal subclusters of the ADLD E11.5 and E18.5 integrated brain single cell dataset

The neuronal trajectory associated clusters of the integrated dataset were further annotated for the subclusters. Trajectories are visualized in UMAP embedding and colored by their annotations.



Supplementary Figure 7.5 Non-negative least-squares (*NNLS*) correlation of the major and sub-trajectories between the MOCA dataset and the MMCA dataset

Heat maps displaying the regression coefficients between trajectories or respective sub-trajectories from MMCA dataset (rows) and MOCA dataset (columns).

7.3 Supplementary Tables

Supplementary Table 7.3.1 Literature based marker gene annotation of the ADLD single cell dataset

cluster	literature based marker genes	annotation
0	Galnt16, Klhl1, Sorcs3 ,Adarb2, Ebf1	Interneurons 1
1	Gli3, Top2a, Sox6, Hes5, Ccna2, Pax6, Egfr, Scd1, Dll1	Radial Glia
2	Tiam2, Satb2, Neurod2, Tbr1, Igfbpl1, Dcc, Srrm4, Plxna4	excitatory Neurons
3	Gad2, Sox2ot, Adarb2, Bcl11b, Erbb4, Sp9, Sorcs3, Dlx5	Interneurons 2
4	Arpp21, Rgs6, Tmem108, Dlg2, Satb2	Dendritogenesis (excitatory)
5	Col3a1, Lama2, Tgfbi, Postn, Foxd1, Col1a1, Col1a2, Tbx18, Dcn	Meninges/Fibroblasts
6	Pdgfra, Serpine2, Ednrb, Olig1, Lhfpl3	Oligodendrocyte Progenitors
7	Flt1, Col4a2, Igfbp7, Pecam1, Cldn5, Ptprb,Rgs5	Endothelial Cells/Pericytes
8	Hba-x, Hba-a1, Hba-a2, Hbb-h, Hbb-bt	Erythroid Lineages
9	C1qb, C1qa, C1qc, Apoe, Mrc1, Trem2, Lyvc1, Spp1,Tyrobp	Microglia/Macrophages

Supplementary Table 7.3.2 MMCA Mutants

Mutant Name	Original Paper Name	Background Strain	Source	Reference
ZRS limb enhancer KO	ZRS-/-	FVB Strain	Visel & Pennacchio & Dickels Lab (Berkeley, California)	DOI: 10.1016/j.cell.2016.09.028
Tbx3 TAD boundary KO	B 5	FVB Strain	Visel & Pennacchio & Dickels Lab (Berkeley, California)	https://doi.org/10.1038/s42003-023-04819-w
Dmrt1 TAD boundary KO	B 8	FVB Strain	Visel & Pennacchio & Dickels Lab (Berkeley, California)	https://doi.org/10.1038/s42003-023-04819-w
Smad7 TAD boundary KO	B 3	FVB Strain	Visel & Pennacchio & Dickels Lab (Berkeley, California)	https://doi.org/10.1038/s42003-023-04819-w
Sim1 TAD boundary KO	B 4	FVB Strain	Visel & Pennacchio & Dickels Lab (Berkeley, California)	https://doi.org/10.1038/s42003-023-04819-w
Neurog2 TAD boundary KO	B 7	FVB Strain	Visel & Pennacchio & Dickels Lab (Berkeley, California)	https://doi.org/10.1038/s42003-023-04819-w

Tbx5 TAD boundary KO	B 2	FVB Strain	Visel & Pennacchio & Dickels Lab (Berkeley, California)	https://doi.org/10.1038/s42003-023-04819-w
Twist1 TAD boundary KO	B 6	FVB Strain	Visel & Pennacchio & Dickels Lab (Berkeley, California)	https://doi.org/10.1038/s42003-023-04819-w
Smad3 TAD boundary KO	B 1	FVB Strain	Visel & Pennacchio & Dickels Lab (Berkeley, California)	https://doi.org/10.1038/s42003-023-04819-w
Gorab KO	Gorab Null	C57BL6 Strain	Kornak Lab (Göttingen, Germany)	DOI: 10.1371/journal.pgen.1007242
Scn10a/11a DKO	not published	C57BL6 Strain	Kurth Lab (Aachen, Germany)	The phenotype is unpublished.
Scn11a GOF	Scn11a+/L 799P	C57BL6 Strain	Kurth Lab (Aachen, Germany)	DOI: 10.1038/ng.2767
Atp6v0a2 R755Q	Atp6v0a2 RQ	C57BL6 Strain	Kornak Lab (Göttingen, Germany)	https://doi.org/10.1007/s00439-012-1197-8
Atp6v0a2 KO	Atp6v0a2 -/-	C57BL6 Strain	Kornak Lab (Göttingen, Germany)	The phenotype is unpublished.
Gli2 KO	Gli2^tm1Al j	C57BL6 Strain	Beier Lab (Seattle, Washington)	https://doi.org/10.1242/dev.124.1.113
Carm1 KO	Carm1^tm 1Mtb	C57BL6 Strain	Beier Lab (Seattle, Washington)	DOI: 10.1073/pnas.1232272100
Ttc21b KO	alien	C57BL6 Strain	Beier Lab (Seattle, Washington)	DOI: 10.1016/j.ydbio.2009.08.023
Ror2 KI	not published	G4 Strain	Kalscheuer Lab (Berlin, Germany)	The phenotype is unpublished.
Cdkl5 -/Y	not published	G4 Strain	Kalscheuer Lab (Berlin, Germany)	The phenotype is unpublished.
Fat1 TAD KO	delta D1+2	G4 Strain	Mundlos Lab (Berlin, Germany)	DOI: 10.1016/j.cell.2022.09.006
Sox9 TAD boundary KI	Bor-KnockI n	G4 Strain	Mundlos Lab (Berlin, Germany)	DOI: 10.1038/s41588-019-0466-z
Sox9 regulatory INV	Sox9_InvC	G4 Strain	Mundlos Lab (Berlin, Germany)	DOI: 10.1038/s41588-019-0466-z
G4 Wildtype	-	G4 Strain	Mundlos Lab (Berlin, Germany)	Wildtype strain used for comparison in cell type composition and lockNess score analysis to detect changes in mutants.
C57BL6 Wildtype	-	C57BL6 Strain	Mundlos Lab (Berlin, Germany)	Wildtype strain used for comparison in cell type composition and lockNess score analysis to detect changes in mutants.
BALB/C Wildtype	-	BALB/C Strain	Visel & Pennacchio & Dickels Lab (Berkeley, California)	Wildtype strain used for comparison in cell type composition and lockNess score analysis to detect changes in mutants.
FVB Wildtype	-	FVB Strain	Visel & Pennacchio & Dickels Lab (Berkeley, California)	Wildtype strain used for comparison in cell type composition and lockNess score analysis to detect changes in mutants.

Supplementary Table 7.3.3 Axial diffusion map voxel based statistics

Deletion > Duplication

set	set	cluster	cluster	cluster	cluster	peak	peak	peak	peak	peak	x	y	z	Labels
p	c	p(FWE-corr)	p(FDR-corr)	equivk	p(unc)	p(FWE-corr)	p(FDR-corr)	T	equivZ	p(unc)	mm	mm	mm	-
4,02E-14	4	2,20E-12		3578	3,70E-14	0,197184694	0,060024554	19,11194992	5,145160218	1,34E-07	1,114999115	0,095518768	-3,0603351	Striatum
						0,897537248	0,060024554	12,62574005	4,586037612	2,26E-06	0,064999223	-0,044481218	-5,090334892	Periventricular hypothalamic nucleus preoptic part
						0,998695708	0,06173375	9,46287632	4,167844757	1,54E-05	-0,075000763	-0,814481139	-4,880334914	Periventricular hypothalamic nucleus intermediate part
		2,01E-06		1484	3,39E-08	0,37667123	0,060024554	17,3939476	5,022288347	2,55E-07	1,394999087	-1,094481111	-1,170335293	lateral ventricle
						0,997270408	0,06173375	9,775366783	4,216195674	1,24E-05	0,764999151	-0,674481153	-1,380335271	lateral ventricle
						0,99817348	0,06173375	9,599833488	4,189267876	1,40E-05	1,814999044	-1,234481096	-1,660335243	lateral ventricle
		4,08E-05		1112	6,88E-07	0,880973092	0,060024554	12,85417271	4,611229087	2,00E-06	0,274999201	-2,354480982	-1,660335243	root
						0,972267895	0,060024554	11,19171238	4,414194054	5,07E-06	0,134999216	-1,304481089	-2,0803352	choroid plexus
		5,53E-06		1355	9,32E-08	0,998251211	0,06173375	9,581651688	4,186444988	1,42E-05	2,864998937	-1,864481032	-4,250334978	Striatum
						0,999959084	0,062971557	8,412625313	3,990184347	3,30E-05	3,774998844	-2,004481018	-4,180334985	Lateral amygdalar nucleus
						0,999998113	0,06490534	7,792992115	3,87252324	5,39E-05	3,354998887	-2,004481018	-4,460334957	Lateral amygdalar nucleus

Wildtype > Duplication

set	set	cluster	cluster	cluster	cluster	peak	peak	peak	peak	peak	x	y	z	Labels
p	c	p(FWE-corr)	p(FDR-corr)	equivk	p(unc)	p(FWE-corr)	p(FDR-corr)	T	equivZ	p(unc)	mm	mm	mm	-
0	5	1,15E-09		2549	1,93E-11	0,245695852	0,065005965	18,51053619	5,103713938	1,67E-07	1,394999087	-1,094481111	-1,170335293	lateral ventricle
						0,65235798	0,065005965	15,43015289	4,862549885	5,79E-07	0,694999158	-0,814481139	-1,310335279	lateral ventricle
						0,978784051	0,091429966	10,97720146	4,386208235	5,77E-06	1,744999051	-1,164481103	-1,59033525	lateral ventricle
		0,000597064		812	1,01E-05	0,97143675	0,087377167	11,21656227	4,417394211	4,99E-06	-1,475000062	-0,884481132	-1,870335221	Caudoputamen
						0,999999772	0,123996659	7,460274696	3,804701559	7,10E-05	-0,845000684	-0,464481175	-1,520335257	lateral ventricle
						1	0,133284105	5,963832855	3,449254053	0,000281069	-0,635000706	0,375518739	-1,870335221	lateral ventricle
		0,000538275		823	9,07E-06	0,997675134	0,110239298	9,703546524	4,205248666	1,30E-05	0,274999201	-2,354480982	-1,660335243	root
						0,997892091	0,110239298	9,660835266	4,198692346	1,34E-05	0,064999223	-2,634480953	-2,57033515	Precommissural nucleus
						0,999910311	0,123996659	8,607889175	4,025150126	2,85E-05	0,134999216	-1,514481068	-2,150335193	Medial habenula
		3,59E-05		1127	6,06E-07	0,999495159	0,120744723	9,118253708	4,112254951	1,96E-05	2,864998937	-1,864481032	-4,250334978	Striatum
						0,999996743	0,123996659	7,889411926	3,891543772	4,98E-05	3,354998887	-2,004481018	-4,530334949	Lateral amygdalar nucleus
		0,000299854		886	5,05E-06	0,999969061	0,123996659	8,347458839	3,978300124	3,47E-05	0,064999223	0,025518775	-5,090334892	Periventricular hypothalamic nucleus preoptic part
						0,999999047	0,123996659	7,678975582	3,849671034	5,91E-05	-0,075000763	-0,744481146	-4,880334914	Periventricular hypothalamic nucleus intermediate part
						0,999999998	0,123996659	6,857079983	3,672240984	0,000120216	-0,005000077	0,095518768	-4,320334971	third ventricle

7.4 List of Abbreviations

°C	degrees celsius
4V	fourth ventricle
AD	axial diffusion
ADHD	attention deficit hyperactivity disorder
ADLD	acute demyelinating leukodystrophy
ASD	autism spectrum disorder
ATAC	Assay for Transposase-Accessible Chromatin
bp	basepairs
Cdkl5	Cyclin Dependent Kinase Like 5
cDNA	complementary DNA
cHiC	capture HiC
ChIP	chromatin immunoprecipitation
ChP	choroid plexus
Chr	Chromosome
CLB	cellular lysis buffer
CREs	cis-regulatory elements
CRISPR	clustered regularly interspaced short palindromic repeats
CSF	cerebrospinal fluids
CTCF	CCCTC-binding factor
DEG	differentially expressed genes
Del	deletion
dMRI	diffusion MRI
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide
DSB	double strand break
DTI	diffusion tensor imaging
Dup	duplication
g	gram
gDNA	genomic DNA
Gli2	GLI Family Zinc Finger 2
GSEA	Gene Set Enrichment Analysis
h	hour
H&E	hematoxylin-eosin
Het	heterozygous
Hom	homozygous

INV	inversion
kb	kilobase
Kcnj2	potassium inwardly rectifying channel subfamily J member 2
KO	knock-out
L	liters
LaGeSo	Landesamt für Gesundheit und Soziales
LIF	leukemia inhibiting factor
Lmnb1	Lamin B1
LN	liquid nitrogen
LV	lateral ventricle
M	Molar
Mb	Megabase
Med	mediator complex
MEF	mouse embryonic fibroblasts
mESC's	mouse embryonic stem cells
min	minute
ml	millilitre
mm	*Mus Musculus*
MMCA	mouse mutant cell atlas
MOCA	mouse organogenesis cell atlas
MRI	magnetic resonance imaging
ng	nanogram
nhvg	number of highly variable genes
NNLS	non-negative least-squares
npcs	number of principal components
NSB	nuclear suspension buffer
ON	over night
PBS	phosphate-buffered saline
PCA	principal component analysis
PFA	paraformaldehyde
pH	potential of hydrogen
PIC	preinitiation complex
rcf	relative centrifugal force
RNA	ribonucleic acid
rpm	revolutions per minute
RT	room temperature
sc	single cell

scRNA-seq	single cell RNA sequencing
seq	sequencing
sgRNA	single guide RNA
Sox9	SRY-Box Transcription Factor 9
TAD	topologically associating domain
TSS	transcriptional start site
Ttc21b	Tetratricopeptide Repeat Domain 21B
U	units
UCSC	University of California, Santa Cruz
UMAP	uniform manifold approximation and projection
WT	wildtype
Y/-	hemizygous
Zfp608	Zinc Finger Protein 608
ZPA	zone of polarizing activity
ZRS	ZPA regulating sequence
µg	microgram
µl	microlitres

7.5 List of Figures

Figure 1.1 The activity of a gene can be regulated by a complex landscape in different tissues **3**
Figure 1.2 Schematic visualization of chromatin organization in the nucleus **7**
Figure 1.3 Schematic overview of cHiC library generation **8**
Figure 1.4 Intra-TAD structural variants **10**
Figure 1.5 Inter-TAD structural variants **11**
Figure 1.6 CRISPR/Cas9 technology in genome editing **13**
Figure 1.7 Structural Variants from patients in the Lmnb1/Zfp608 mouse brain cHiC **16**
Figure 1.8 ADLD patient MRI and phenotype of mouse Lmnb1 overexpression **18**
Figure 1.9 Overview of single cell dissection and barcoding methods of the mouse brain **21**
Figure 5.1 Schematic overview of the sci-RNA-seq3 experiment and analysis **57**
Figure 5.2 Quality control analysis of the dataset **58**
Figure 5.3 Cell composition whole dataset analysis **59**
Figure 5.4 Ttc21b KO mutant eye development **60**
Figure 5.5 lochNESS analysis of the Gli2 KO in the floor and roof plate **62**
Figure 5.6 Gli2 KO Mutant neural tube and choroid plexus deformations **64**
Figure 5.7 Sox9 regulatory INV mutant changes in regulatory region related tissues **67**
Figure 5.8 Visualization of the experimental strategy of the ADLD project **69**
Figure 5.9 Wildtype Lmnb1 locus with CRISPR mutation targets and wildtype insulation score **71**

Figure 5.10 Subtraction maps of the duplication and deletion **72**
Figure 5.11 Deletion mutant custom map and insulation score **73**
Figure 5.12 The mutant single cell dataset and locus specific gene expression changes **75**
Figure 5.13 GSEA related changes in cholesterol pathways of radial glia **77**
Figure 5.14 Changes in cell composition and difference in stage dependent glia differentiation **79**
Figure 5.15 MRI based differences in aging mouse brains **81**
Figure 6.1 Inversion mutants for further analysis of the locus **90**
Figure 6.2 Model of the deletion variant **96**
Supplementary Figure 7.1 Vector Map of the pSpCas9(BB)-2A-Puro (px459) Vector **98**
Supplementary Figure 7.2 Duplication Variants from patients in the Lmnb1/Zfp608 mouse locus **99**
Supplementary Figure 7.3 Annotation of sub-trajectories of the MMCA E13.5 embryos **99**
Supplementary Figure 7.4 Annotation of the neuronal subclusters of the ADLD E11.5 and E18.5 integrated brain single cell dataset **100**
Supplementary Figure 7.5 Non-negative least-squares (NNLS) correlation of the major and sub-trajectories between the MOCA dataset and the MMCA dataset **100**

7.6 List of Tables

Table 3.1 Antibodies, Enzymes and Probes used in the ADLD and MMCA Project **23**
Table 3.2 standard solutions used in this study **23**
Table 3.3 Buffers and Solutions for 10x based single cell nuclei preparation **24**
Table 3.4 Buffers and Solutions for sci RNA seq based single cell nuclei preparation **24**
Table 3.5 Buffers and Solutions for cHiC sample and library preparation **24**
Table 3.6 Instruments used in this study **25**
Table 3.7 Kits used in this study **25**
Table 3.8 Media used in this study **26**
Table 3.9 Oligonucleotides used in the ADLD project **26**
Table 3.10 Transgenic mouse lines generated in the ADLD project **27**
Table 3.11 Software, Tools & Programming Languages used in the ADLD & MMCA project **27**
Table 4.1 & 4.2 Chemicals and cycle numbers of the standard genotyping PCR **32**
Supplementary Table 7.3.1 Literature based marker gene annotation of the ADLD single cell dataset **101**
Supplementary Table 7.3.2 MMCA Mutants **101**
Supplementary Table 7.3.3 Axial diffusion map voxel based statistics **103**

8. References

Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, P. G. Amanatides, S. E. Scherer, et al. 2000. “The Genome Sequence of Drosophila Melanogaster.” *Science* 287 (5461): 2185–95.

Aganezov, Sergey, Stephanie M. Yan, Daniela C. Soto, Melanie Kirsche, Samantha Zarate, Pavel Avdeyev, Dylan J. Taylor, et al. 2022. “A Complete Reference Genome Improves Analysis of Human Genetic Variation.” *Science* 376 (6588): eabl3533.

Aires, Ana, Ariana Barros, Célia Machado, Diogo Fitas, Gonçalo Cação, Rui Pedrosa, João Cerqueira, et al. 2019. “Diagnostic Delay of Multiple Sclerosis in a Portuguese Population.” *Acta Medica Portuguesa* 32 (4): 289–94.

Akbarian, Schahram, and Hyejung Won. 2023. “Chromosomal Contacts Change with Age.” *Science* 381 (6662): 1049–50.

Allen, Benjamin L., and Dylan J. Taatjes. 2015. “The Mediator Complex: A Central Integrator of Transcription.” *Nature Reviews. Molecular Cell Biology* 16 (3): 155–66.

Allis, C. David, and Thomas Jenuwein. 2016. “The Molecular Hallmarks of Epigenetic Control.” *Nature Reviews. Genetics* 17 (8): 487–500.

“A Physical Map of the Human Genome.” 2001. *Nature* 409 (6822): 934–41.

Ayoub, Albert E., Sunghee Oh, Yanhua Xie, Jing Leng, Justin Cotney, Martin H. Dominguez, James P. Noonan, and Pasko Rakic. 2011. “Transcriptional Programs in Transient Embryonic Zones of the Cerebral Cortex Defined by High-Resolution mRNA Sequencing.” *Proceedings of the National Academy of Sciences of the United States of America* 108 (36): 14950–55.

Bageritz, Josephine, and Gianmarco Raddi. 2019. “Single-Cell RNA Sequencing with Drop-Seq.” *Methods in Molecular Biology* 1979: 73–85.

Banigan, Edward J., and Leonid A. Mirny. 2020. “Loop Extrusion: Theory Meets Single-Molecule Experiments.” *Current Opinion in Cell Biology* 64 (June): 124–38.

Bannister, Andrew J., and Tony Kouzarides. 2011. “Regulation of Chromatin by Histone Modifications.” *Cell Research* 21 (3): 381–95.

Bechler, Marie E., Matthew Swire, and Charles Ffrench-Constant. 2018. “Intrinsic and Adaptive Myelination-A Sequential Mechanism for Smart Wiring in the Brain.” *Developmental Neurobiology* 78 (2): 68–79.

Bekris, Lynn M., Chang-En Yu, Thomas D. Bird, and Debby W. Tsuang. 2010. “Genetics of Alzheimer Disease.” *Journal of Geriatric Psychiatry and Neurology* 23 (4): 213–27.

Benko, Sabina, Judy A. Fantes, Jeanne Amiel, Dirk-Jan Kleinjan, Sophie Thomas, Jacqueline Ramsay, Negar Jamshidi, et al. 2009. “Highly Conserved Non-Coding Elements on Either Side of SOX9 Associated with Pierre Robin Sequence.” *Nature Genetics* 41 (3): 359–64.

Berg, Debbie L. C. van den, Roberta Azzarelli, Koji Oishi, Ben Martynoga, Noelia Urbán, Dick H. W. Dekkers, Jeroen A. Demmers, and François Guillemot. 2017. “Nipbl Interacts with Zfp609 and the Integrator Complex to Regulate Cortical Neuron Migration.” *Neuron* 93 (2): 348–61.

Berghoff, Stefan A., Lena Spieth, and Gesine Saher. 2022. “Local Cholesterol Metabolism Orchestrates Remyelination.” *Trends in Neurosciences* 45 (4): 272–83.

Bock, Christoph, Paul Datlinger, Florence Chardon, Matthew A. Coelho, Matthew B. Dong, Keith A. Lawson, Tian Lu, et al. 2022. “High-Content CRISPR Screening.” *Nature Reviews Methods Primers* 2 (1): 1–23.

Boija, Ann, Isaac A. Klein, Benjamin R. Sabari, Alessandra Dall’Agnese, Eliot L. Coffey, Alicia V. Zamudio, Charles H. Li, et al. 2018. “Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains.” *Cell* 175 (7): 1842–55.e16.

Bolt, Christopher Chase, Lucille Lopez-Delisle, Aurélie Hintermann, Bénédicte Mascrez, Antonella Rauseo, Guillaume Andrey, and Denis Duboule. 2022. “Context-Dependent Enhancer Function Revealed by Targeted Inter-TAD Relocation.” *Nature Communications* 13 (1): 1–15.

Bonev, B., N. M. Cohen, Q. Szabo, L. Fritsch, G. L. Papadopoulos, Y. Lubling, X. L. Xu, X. D. Lv, J. P. Hugnot, A. Tanay, and G. Cavalli. 2017. Multiscale 3D Genome Rewiring during Mouse Neural Development.Cell171 (3):557-+.

Bonev, Boyan, and Giacomo Cavalli. 2016. “Organization and Function of the 3D Genome.” *Nature Reviews. Genetics* 17 (11): 661–78.

Borja, Nicholas, Stephanie Bivona, Lé Shon Peart, Brittany Johnson, Joanna Gonzalez, Deborah Barbouth, Henry Moore, et al. 2022. “Genome Sequencing Reveals Novel Noncoding Variants in PLA2G6 and LMNB1 Causing Progressive Neurologic Disease.” *Molecular Genetics & Genomic Medicine* 10 (4): e1892.

Botten, Giovanni A., Yuannyu Zhang, Kseniia Dudnyk, Yoon Jung Kim, Xin Liu, Jacob T. Sanders, Aygun Imanci, et al. 2023. “Structural Variation Cooperates with Permissive Chromatin to Control Enhancer Hijacking-Mediated Oncogenic Transcription.” *Blood* 142 (4): 336–51.

Brambach, Max, Ariane Ernst, Sara Nolbrant, Janelle Drouin-Ouellet, Agnete Kirkeby, Malin Parmar, and Victor Olariu. 2021. “Neural Tube Patterning: From a Minimal Model for Rostrocaudal Patterning toward an Integrated 3D Model.” *iScience* 24 (6): 102559.

Brown, S. W. 1966. “Heterochromatin.” *Science* 151 (3709): 417–25.

Brussino, A., G. Vaula, C. Cagnoli, A. Mauro, L. Pradotto, D. Daniele, E. Di Gregorio, et al. 2009. “A Novel Family with Lamin B1 Duplication Associated with Adult-Onset Leucoencephalopathy.” *Journal of Neurology, Neurosurgery, and Psychiatry* 80 (2): 237–40.

Brussino, A., G. Vaula, C. Cagnoli, E. Panza, M. Seri, E. Di Gregorio, S. Scappaticci, et al. 2010. “A Family with Autosomal Dominant Leukodystrophy Linked to 5q23.2–q23.3 without Lamin B1 Mutations.” *European Journal of Neurology: The Official Journal of the European Federation of Neurological Societies* 17 (4): 541–49.

Bryda, Elizabeth C. 2013. “The Mighty Mouse: The Impact of Rodents on Advances in Biomedical Research.” *Missouri Medicine* 110 (3): 207–11.

Buenrostro, Jason D., Paul G. Giresi, Lisa C. Zaba, Howard Y. Chang, and William J. Greenleaf. 2013. “Transposition of Native Chromatin for Fast and Sensitive Epigenomic Profiling of Open Chromatin, DNA-Binding Proteins and Nucleosome Position.” *Nature Methods* 10 (12): 1213–18.

Butler, Andrew, Paul Hoffman, Peter Smibert, Efthymia Papalexi, and Rahul Satija. 2018. “Integrating Single-Cell Transcriptomic Data across Different Conditions, Technologies, and Species.” *Nature Biotechnology* 36 (5): 411–20.

Calderon, Diego, Ronnie Blecher-Gonen, Xingfan Huang, Stefano Secchia, James Kentro, Riza M. Daza, Beth Martin, et al. 2022. “The Continuum of Embryonic Development at Single-Cell Resolution.” *Science* 377 (6606): eabn5800.

Camps, Jordi, Michael R. Erdos, and Thomas Ried. 2015. “The Role of Lamin B1 for the Maintenance of Nuclear Structure and Function.” *Nucleus* 6 (1): 8–14.

Cannavò, Enrico, Pierre Khoueiry, David A. Garfield, Paul Geeleher, Thomas Zichner, E. Hilary Gustafson, Lucia Ciglar, Jan O. Korbel, and Eileen E. M. Furlong. 2016. “Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks.” *Current Biology: CB* 26 (1): 38–51.

Cao, Junyue, Jonathan S. Packer, Vijay Ramani, Darren A. Cusanovich, Chau Huynh, Riza Daza, Xiaojie Qiu, et al. 2017. “Comprehensive Single-Cell Transcriptional Profiling of a Multicellular Organism.” *Science* 357 (6352): 661–67.

Cao, Junyue, Malte Spielmann, Xiaojie Qiu, Xingfan Huang, Daniel M. Ibrahim, Andrew J. Hill, Fan Zhang, et al. 2019. “The Single-Cell Transcriptional Landscape of Mammalian Organogenesis.” *Nature* 566 (7745): 496–502.

Caronia-Brown, Giuliana, Michio Yoshida, Forrest Gulden, Stavroula Assimacopoulos, and Elizabeth A. Grove. 2014. “The Cortical Hem Regulates the Size and Patterning of Neocortex.” *Development* 141 (14): 2855–65.

C. elegans Sequencing Consortium. 1998. “Genome Sequence of the Nematode C. Elegans: A Platform for Investigating Biology.” *Science* 282 (5396): 2012–18.

Chen, Song, Blue B. Lake, and Kun Zhang. 2019. “High-Throughput Sequencing of the

Transcriptome and Chromatin Accessibility in the Same Cell." *Nature Biotechnology* 37 (12): 1452–57.
Chen, Xizi, Yilun Qi, Zihan Wu, Xinxin Wang, Jiabei Li, Dan Zhao, Haifeng Hou, et al. 2021. "Structural Insights into Preinitiation Complex Assembly on Core Promoters." *Science* 372 (6541). https://doi.org/10.1126/science.aba8490.
Christian, Susan L., Judy A. Fantes, Stephanie K. Mewborn, Bing Huang, and David H. Ledbetter. 1999. "Large Genomic Duplicons Map to Sites of Instability in the Prader-Willi/Angelman Syndrome Chromosome Region (15q11–q13)." *Human Molecular Genetics* 8 (6): 1025–37.
Claringbould, Annique, and Judith B. Zaugg. 2021. "Enhancers in Disease: Molecular Basis and Emerging Treatment Strategies." *Trends in Molecular Medicine* 27 (11): 1060–73.
Coffeen, Christin M., Catherine E. McKenna, Arnulf H. Koeppen, Nikki M. Plaster, Nicholas Maragakis, Jason Mihalopoulos, John D. Schwankhaus, et al. 2000. "Genetic Localization of an Autosomal Dominant Leukodystrophy Mimicking Chronic Progressive Multiple Sclerosis to Chromosome 5q31." *Human Molecular Genetics* 9 (5): 787–93.
Coffinier, Catherine, Hea-Jin Jung, Chika Nobumori, Sandy Chang, Yiping Tu, Richard H. Barnes II, Yuko Yoshinaga, et al. 2011. "Deficiencies in Lamin B1 and Lamin B2 Cause Neurodevelopmental Defects and Distinct Nuclear Shape Abnormalities in Neurons." *Molecular Biology of the Cell*, October. https://doi.org/10.1091/mbc.e11-06-0504.
Cong, Le, F. Ann Ran, David Cox, Shuailiang Lin, Robert Barretto, Naomi Habib, Patrick D. Hsu, et al. 2013. "Multiplex Genome Engineering Using CRISPR/Cas Systems." *Science* 339 (6121): 819–23.
Cowin, Gary J., Tim J. Butler, Nyoman D. Kurniawan, Charles Watson, and Robyn H. Wallace. 2011. "Magnetic Resonance Microimaging of the Spinal Cord in the SOD1 Mouse Model of Amyotrophic Lateral Sclerosis Detects Motor Nerve Root Degeneration." *NeuroImage* 58 (1): 69–74.
Cramer, Patrick. 2019. "Organization and Regulation of Gene Transcription." *Nature* 573 (7772): 45–54.
Crane, Emily, Qian Bian, Rachel Patton McCord, Bryan R. Lajoie, Bayly S. Wheeler, Edward J. Ralston, Satoru Uzawa, Job Dekker, and Barbara J. Meyer. 2015. "Condensin-Driven Remodelling of X Chromosome Topology during Dosage Compensation." *Nature* 523 (7559): 240–44.
Cremer, T., C. Cremer, H. Baumann, E. K. Luedtke, K. Sperling, V. Teuber, and C. Zorn. 1982. "Rabl's Model of the Interphase Chromosome Arrangement Tested in Chinise Hamster Cells by Premature Chromosome Condensation and Laser-UV-Microbeam Experiments." *Human Genetics* 60 (1): 46–56.
Cristofoli, Francesca, Tonya Moss, Hannah W. Moore, Koen Devriendt, Heather Flanagan-Steet, Melanie May, Julie Jones, et al. 2020. "De Novo Variants in LMNB1 Cause Pronounced Syndromic Microcephaly and Disruption of Nuclear Envelope Integrity." *American Journal of Human Genetics* 107 (4): 753–62.
Cusanovich, Darren A., Riza Daza, Andrew Adey, Hannah A. Pliner, Lena Christiansen, Kevin L. Gunderson, Frank J. Steemers, Cole Trapnell, and Jay Shendure. 2015. "Multiplex Single Cell Profiling of Chromatin Accessibility by Combinatorial Cellular Indexing." *Science* 348 (6237): 910–14.
Cusanovich, Darren A., Andrew J. Hill, Delasa Aghamirzaie, Riza M. Daza, Hannah A. Pliner, Joel B. Berletch, Galina N. Filippova, et al. 2018. "A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility." *Cell* 174 (5): 1309–24.e18.
Dani, R.H. Herbst, C. McCabe, G.S. Green, K. Kaiser, J.P. Head, J. Cui, F.B. Shipley, A. Jang, D. Dionne, L. Nguyen, C. Rodman, S.J. Riesenfeld, J. Prochazka, M. Prochazkova, R. Sedlacek, F. Zhang, V. Bryja, O. Rozenblatt-Rosen, N. Habib, A. Regev, M.K. Lehtinen"A Cellular and Spatial Map of the Choroid Plexus across Brain Ventricles and Ages." 2021. *Cell* 184 (11): 3056–74.e21.
Dekker, Job, and Edith Heard. 2015. "Structural and Functional Diversity of Topologically Associating Domains." *FEBS Letters* 589 (20 Pt A): 2877–84.
Dekker, Job, Karsten Rippe, Martijn Dekker, and Nancy Kleckner. 2002. "Capturing

Chromosome Conformation." *Science* 295 (5558): 1306–11.
Demirci, Selami, Alexis Leonard, Juan J. Haro-Mora, Naoya Uchida, and John F. Tisdale. 2019. "CRISPR/Cas9 for Sickle Cell Disease: Applications, Future Possibilities, and Challenges." *Advances in Experimental Medicine and Biology* 1144: 37–52.
Der, Evan, Saritha Ranabothu, Hemant Suryawanshi, Kemal M. Akat, Robert Clancy, Pavel Morozov, Manjunath Kustagi, et al. 2017. "Single Cell RNA Sequencing to Dissect the Molecular Heterogeneity in Lupus Nephritis." *JCI Insight* 2 (9). https://doi.org/10.1172/jci.insight.93009.
Despang, Alexandra, Robert Schöpflin, Martin Franke, Salaheddine Ali, Ivana Jerković, Christina Paliou, Wing-Lee Chan, et al. 2019. "Functional Dissection of the Sox9-Kcnj2 Locus Identifies Nonessential and Instructive Roles of TAD Architecture." *Nature Genetics* 51 (8): 1263–71.
Dileep, Vishnu, Carles A. Boix, Hansruedi Mathys, Asaf Marco, Gwyneth M. Welch, Hiruy S. Meharena, Anjanet Loon, et al. 2023. "Neuronal DNA Double-Strand Breaks Lead to Genome Structural Variations and 3D Genome Disruption in Neurodegeneration." *Cell* 186 (20): 4404–21.e20.
Ding, Chenyu, Yuying Wu, Xiaoyong Chen, Yue Chen, Zanyi Wu, Zhangya Lin, Dezhi Kang, Wenhua Fang, and Fa Chen. 2022. "Global, Regional, and National Burden and Attributable Risk Factors of Neurological Disorders: The Global Burden of Disease Study 1990-2019." *Frontiers in Public Health* 10 (November): 952161.
Ding, Q., J. Motoyama, S. Gasca, R. Mo, H. Sasaki, J. Rossant, and C. C. Hui. 1998. "Diminished Sonic Hedgehog Signaling and Lack of Floor Plate Differentiation in Gli2 Mutant Mice." *Development* 125 (14): 2533–43.
Dixon, Jesse R., Siddarth Selvaraj, Feng Yue, Audrey Kim, Yan Li, Yin Shen, Ming Hu, Jun S. Liu, and Bing Ren. 2012. "Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions." *Nature* 485 (7398): 376–80.
Dixon, Jesse R., Jie Xu, Vishnu Dileep, Ye Zhan, Fan Song, Victoria T. Le, Galip Gürkan Yardımcı, et al. 2018. "Integrative Detection and Analysis of Structural Variation in Cancer Genomes." *Nature Genetics* 50 (10): 1388–98.
Dobson, Philip F., Mariana C. Rocha, John P. Grady, Alexia Chrysostomou, Daniel Hipps, Sharon Watson, Laura C. Greaves, David J. Deehan, and Doug M. Turnbull. 2016. "Unique Quadruple Immunofluorescence Assay Demonstrates Mitochondrial Respiratory Chain Dysfunction in Osteoblasts of Aged and PolgA−/− Mice." *Scientific Reports* 6 (1): 1–10.
Domcke, Silvia, and Jay Shendure. 2023. "A Reference Cell Tree Will Serve Science Better than a Reference Cell Atlas." *Cell* 186 (6): 1103–14.
Doudna, Jennifer A., and Emmanuelle Charpentier. 2014. "Genome Editing. The New Frontier of Genome Engineering with CRISPR-Cas9." *Science* 346 (6213): 1258096.
Dubois, J., G. Dehaene-Lambertz, S. Kulikova, C. Poupon, P. S. Hüppi, and L. Hertz-Pannier. 2014. "The Early Development of Brain White Matter: A Review of Imaging Studies in Fetuses, Newborns and Infants." *Neuroscience* 276 (September): 48–71.
Du, Zhenhai, Hui Zheng, Bo Huang, Rui Ma, Jingyi Wu, Xianglin Zhang, Jing He, et al. 2017. "Allelic Reprogramming of 3D Chromatin Architecture during Early Mammalian Development." *Nature* 547 (7662): 232–35.
Eichmüller, Oliver L., and Juergen A. Knoblich. 2022. "Human Cerebral Organoids - a New Tool for Clinical Neurology Research." *Nature Reviews. Neurology* 18 (11): 661–80.
Etourneaud, Laure, Angela Moussa, Emilie Rass, Diane Genet, Simon Willaume, Caroline Chabance-Okumura, Paul Wanschoor, et al. 2021. "Lamin B1 Sequesters 53BP1 to Control Its Recruitment to DNA Damage." *Science Advances* 7 (35). https://doi.org/10.1126/sciadv.abb3799.
Falcão de Campos, Catarina, Marta Gromicho, Hilmi Uysal, Julian Grosskreutz, Magdalena Kuzma-Kozakiewicz, Miguel Oliveira Santos, Susana Pinto, Susanne Petri, Michael Swash, and Mamede de Carvalho. 2023. "Trends in the Diagnostic Delay and Pathway for Amyotrophic Lateral Sclerosis Patients across Different Countries." *Frontiers in Neurology* 13 (January): 1064619.

Feigin, Valery L., Theo Vos, Emma Nichols, Mayowa O. Owolabi, William M. Carroll, Martin Dichgans, Günther Deuschl, Priya Parmar, Michael Brainin, and Christopher Murray. 2020. “The Global Burden of Neurological Disorders: Translating Evidence into Policy.” *Lancet Neurology* 19 (3): 255–65.
“Finishing the Euchromatic Sequence of the Human Genome.” 2004. *Nature* 431 (7011): 931–45.
Finnsson, Johannes, Jimmy Sundblom, Niklas Dahl, Atle Melberg, and Raili Raininko. 2015. “LMNB1-Related Autosomal-Dominant Leukodystrophy: Clinical and Radiological Course.” *Annals of Neurology* 78 (3): 412–25.
Fischer, Björn, Aikaterini Dimopoulou, Johannes Egerer, Thatjana Gardeitchik, Alexa Kidd, Dominik Jost, Hülya Kayserili, et al. 2012. “Further Characterization of ATP6V0A2-Related Autosomal Recessive Cutis Laxa.” *Human Genetics* 131 (11): 1761–73.
Flöttmann, Ricarda, Johannes Wagner, Karolina Kobus, Cynthia J. Curry, Ravi Savarirayan, Gen Nishimura, Natsuo Yasui, et al. 2015. “Microdeletions on 6p22.3 Are Associated with Mesomelic Dysplasia Savarirayan Type.” *Journal of Medical Genetics* 52 (7): 476–83.
Frade, José M., and María C. Ovejero-Benito. 2015. “Neuronal Cell Cycle: The Neuron Itself and Its Circumstances.” *Cell Cycle* 14 (5): 712–20.
Franke, Martin, Daniel M. Ibrahim, Guillaume Andrey, Wibke Schwarzer, Verena Heinrich, Robert Schöpflin, Katerina Kraft, et al. 2016. “Formation of New Chromatin Domains Determines Pathogenicity of Genomic Duplications.” *Nature* 538 (7624): 265–69.
Freund, Adam, Remi-Martin Laberge, Marco Demaria, and Judith Campisi. 2012. “Lamin B1 Loss Is a Senescence-Associated Biomarker.” *Molecular Biology of the Cell* 23 (11): 2066–75.
From, Renana, Raya Eilam, Dekel D. Bar-Lev, Smadar Levin-Zaidman, Michael Tsoory, Patrizia LoPresti, Michael Sela, Ruth Arnon, and Rina Aharoni. 2014. “Oligodendrogenesis and Myelinogenesis during Postnatal Development Effect of Glatiramer Acetate.” *Glia* 62 (4): 649–65.
Furlong, Eileen E. M., and Michael Levine. 2018. “Developmental Enhancers and Chromosome Topology.” *Science* 361 (6409): 1341–45.
Gailus-Durner, Valérie, Helmut Fuchs, Lore Becker, Ines Bolle, Markus Brielmeier, Julia Calzada-Wack, Ralf Elvert, et al. 2005. “Introducing the German Mouse Clinic: Open Access Platform for Standardized Phenotyping.” *Nature Methods* 2 (6): 403–4.
Gaj, Thomas, Charles A. Gersbach, and Carlos F. Barbas 3rd. 2013. “ZFN, TALEN, and CRISPR/Cas-Based Methods for Genome Engineering.” *Trends in Biotechnology* 31 (7): 397–405.
GBD 2016 Neurology Collaborators. 2019. “Global, Regional, and National Burden of Neurological Disorders, 1990-2016: A Systematic Analysis for the Global Burden of Disease Study 2016.” *Lancet Neurology* 18 (5): 459–80.
Giacomini, Caterina, Sameehan Mahajani, Roberta Ruffilli, Roberto Marotta, and Laura Gasparini. 2016. “Lamin B1 Protein Is Required for Dendrite Development in Primary Mouse Cortical Neurons.” *Molecular Biology of the Cell* 27 (1): 35–47.
Gião, Tiago, Tiago Teixeira, Maria Rosário Almeida, and Isabel Cardoso. 2022. “Choroid Plexus in Alzheimer’s Disease-The Current State of Knowledge.” *Biomedicines* 10 (2). https://doi.org/10.3390/biomedicines10020224.
Giorgio, Elisa, Daniel Robyr, Malte Spielmann, Enza Ferrero, Eleonora Di Gregorio, Daniele Imperiale, Giovanna Vaula, et al. 2015. “A Large Genomic Deletion Leads to Enhancer Adoption by the Lamin B1 Gene: A Second Path to Autosomal Dominant Adult-Onset Demyelinating Leukodystrophy (ADLD).” *Human Molecular Genetics* 24 (11): 3143–54.
Giorgio, Elisa, Harshvardhan Rolyan, Laura Kropp, Anish Baswanth Chakka, Svetlana Yatsenko, Eleonora Di Gregorio, Daniela Lacerenza, et al. 2013. “Analysis of LMNB1 Duplications in Autosomal Dominant Leukodystrophy Provides Insights into Duplication Mechanisms and Allele-Specific Expression.” *Human Mutation* 34 (8): 1160–71.
Goldstein, J. L., and M. S. Brown. 1990. “Regulation of the Mevalonate Pathway.” *Nature* 343

(6257): 425–30.
Gregory, Simon G., Mandeep Sekhon, Jacqueline Schein, Shaying Zhao, Kazutoyo Osoegawa, Carol E. Scott, Richard S. Evans, et al. 2002. “A Physical Map of the Mouse Genome.” *Nature* 418 (6899): 743–50.
Gwathmey, Kelly G., Philippe Corcia, Chris J. McDermott, Angela Genge, Stefan Sennfält, Mamede de Carvalho, and Caroline Ingre. 2023. “Diagnostic Delay in Amyotrophic Lateral Sclerosis.” *European Journal of Neurology: The Official Journal of the European Federation of Neurological Societies* 30 (9): 2595–2601.
Hammelrath, Luam, Siniša Škokić, Artem Khmelinskii, Andreas Hess, Noortje van der Knaap, Marius Staring, Boudewijn P. F. Lelieveldt, Dirk Wiedermann, and Mathias Hoehn. 2016. “Morphological Maturation of the Mouse Brain: An in Vivo MRI and Histology Investigation.” *NeuroImage* 125 (January): 144–52.
Han, Lin, Xiaoyuan Zi, Lana X. Garmire, Yu Wu, Sherman M. Weissman, Xinghua Pan, and Rong Fan. 2014. “Co-Detection and Sequencing of Genes and Transcripts from the Same Single Cells Facilitated by a Microfluidics Platform.” *Scientific Reports* 4 (September): 6485.
Hao, Yuhan, Stephanie Hao, Erica Andersen-Nissen, William M. Mauck 3rd, Shiwei Zheng, Andrew Butler, Maddie J. Lee, et al. 2021. “Integrated Analysis of Multimodal Single-Cell Data.” *Cell* 184 (13): 3573–87.e29.
Hartlerode, Andrea J., and Ralph Scully. 2009. “Mechanisms of Double-Strand Break Repair in Somatic Mammalian Cells.” *Biochemical Journal* 423 (2): 157–68.
Heintzman, Nathaniel D., Rhona K. Stuart, Gary Hon, Yutao Fu, Christina W. Ching, R. David Hawkins, Leah O. Barrera, et al. 2007. “Distinct and Predictive Chromatin Signatures of Transcriptional Promoters and Enhancers in the Human Genome.” *Nature Genetics* 39 (3): 311–18.
Heitz, E. 1933. “Die somatische Heteropyknose bei Drosophila melanogaster und ihre genetische Bedeutung.” *Zeitschrift für Zellforschung und Mikroskopische Anatomie* 20 (1): 237–87.
Heng, Mary Y., Shu-Ting Lin, Laure Verret, Yong Huang, Sherry Kamiya, Quasar S. Padiath, Ying Tong, et al. 2013. “Lamin B1 Mediates Cell-Autonomous Neuropathology in a Leukodystrophy Mouse Model.” *The Journal of Clinical Investigation* 123 (6): 2719–29.
Hnisz, Denes, Brian J. Abraham, Tong Ihn Lee, Ashley Lau, Violaine Saint-André, Alla A. Sigova, Heather A. Hoke, and Richard A. Young. 2013. “Super-Enhancers in the Control of Cell Identity and Disease.” *Cell* 155 (4): 934–47.
Hnisz, Denes, Krishna Shrinivas, Richard A. Young, Arup K. Chakraborty, and Phillip A. Sharp. 2017. “A Phase Separation Model for Transcriptional Control.” *Cell* 169 (1): 13–23.
Hoffmann, Christian J., Melanie Tc Kuffner, Leif Koschützke, Stephanie Lorenz, Janet Lips, Philipp Boehm-Sturm, Susanne Mueller, et al. 2023. “Zfp580 Inactivation as a New Therapeutic Target to Enhance Recovery after Stroke in Mice.” *Journal of Cerebral Blood Flow and Metabolism: Official Journal of the International Society of Cerebral Blood Flow and Metabolism* 43 (8): 1400–1418.
Huang, Hui, Quan Zhu, Adam Jussila, Yuanyuan Han, Bogdan Bintu, Colin Kern, Mattia Conte, et al. 2021. “CTCF Mediates Dosage- and Sequence-Context-Dependent Transcriptional Insulation by Forming Local Chromatin Domains.” *Nature Genetics* 53 (7): 1064–74.
Huang, Jialiang, Kailong Li, Wenqing Cai, Xin Liu, Yuannyu Zhang, Stuart H. Orkin, Jian Xu, and Guo-Cheng Yuan. 2018. “Dissecting Super-Enhancer Hierarchy Based on Chromatin Interactions.” *Nature Communications* 9 (1): 943.
Huang, Jialiang, Xin Liu, Dan Li, Zhen Shao, Hui Cao, Yuannyu Zhang, Eirini Trompouki, et al. 2016. “Dynamic Control of Enhancer Repertoires Drives Lineage and Stage-Specific Transcription during Hematopoiesis.” *Developmental Cell* 36 (1): 9–23.
Huang, Xingfan, Jana Henck, Chengxiang Qiu, Varun KA Sreenivasan, Saranya Balachandran, Rose Behncke, Wing-Lee Chan et al. "Single cell, whole embryo phenotyping of pleiotropic disorders of mammalian development." *Nature* 2023.

DOI:10.1038/s41586-023-06548-w
Inoue, Fumitaka, and Nadav Ahituv. 2015. “Decoding Enhancers Using Massively Parallel Reporter Assays.” *Genomics* 106 (3): 159–64.
Ishino, Y., H. Shinagawa, K. Makino, M. Amemura, and A. Nakata. 1987. “Nucleotide Sequence of the Iap Gene, Responsible for Alkaline Phosphatase Isozyme Conversion in Escherichia Coli, and Identification of the Gene Product.” *Journal of Bacteriology* 169 (12): 5429–33.
Jacob, John, and James Briscoe. 2003. “Gli Proteins and the Control of Spinal-Cord Patterning.” *EMBO Reports* 4 (8): 761–65.
Jasin, Maria, and Rodney Rothstein. 2013. “Repair of Strand Breaks by Homologous Recombination.” *Cold Spring Harbor Perspectives in Biology* 5 (11): a012740.
Jerkovic´, Ivana, and Giacomo Cavalli. 2021. “Understanding 3D Genome Organization by Multidisciplinary Methods.” *Nature Reviews. Molecular Cell Biology* 22 (8): 511–28.
Jiang, Jing, Cankun Wang, Ren Qi, Hongjun Fu, and Qin Ma. 2020. “scREAD: A Single-Cell RNA-Seq Database for Alzheimer’s Disease.” *iScience* 23 (11): 101769.
Jindal, G. and Farley E. “Enhancer Grammar in Development, Evolution, and Disease: Dependencies and Interplay.” 2021. Developmental Cell 56 (5): 575–87.
Jinek, Martin, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A. Doudna, and Emmanuelle Charpentier. 2012. “A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity.” *Science* 337 (6096): 816–21.
Ji, Y., E. E. Eichler, S. Schwartz, and R. D. Nicholls. 2000. “Structure of Chromosomal Duplicons and Their Role in Mediating Human Genomic Disorders.” *Genome Research* 10 (5): 597–610.
Jullienne, Amandine, Michelle V. Trinh, and Andre Obenaus. 2022. “Neuroimaging of Mouse Models of Alzheimer’s Disease.” *Biomedicines* 10 (2). https://doi.org/10.3390/biomedicines10020305.
“June 2000 White House Event.” n.d. Genome.gov. Accessed September 9, 2023. https://www.genome.gov/10001356/june-2000-white-house-event.
Kaneshiro, Jeanae M., Juliana S. Capitanio, and Martin W. Hetzer. 2023. “Lamin B1 Overexpression Alters Chromatin Organization and Gene Expression.” *Nucleus* 14 (1): 2202548.
Kant, Shawn, Edward G. Stopa, Conrad E. Johanson, Andrew Baird, and Gerald D. Silverberg. 2018. “Choroid Plexus Genes for CSF Production and Brain Homeostasis Are Altered in Alzheimer’s Disease.” *Fluids and Barriers of the CNS* 15 (1): 1–10.
Karnuta, Jaret M., and Peter C. Scacheri. 2018. “Enhancers: Bridging the Gap between Gene Control and Human Disease.” *Human Molecular Genetics* 27 (R2): R219–27.
Katsuno, Masahisa, Fumiaki Tanaka, and Gen Sobue. 2012. “Perspectives on Molecular Targeted Therapies and Clinical Trials for Neurodegenerative Diseases.” *Journal of Neurology, Neurosurgery, and Psychiatry* 83 (3): 329–35.
Kim, Youngjo, Alexei A. Sharov, Katie McDole, Melody Cheng, Haiping Hao, Chen-Ming Fan, Nicholas Gaiano, Minoru S. H. Ko, and Yixian Zheng. 2011. “Mouse B-Type Lamins Are Required for Proper Organogenesis but Not by Embryonic Stem Cells.” *Science* 334 (6063): 1706–10.
Klein, Allon M., Linas Mazutis, Ilke Akartuna, Naren Tallapragada, Adrian Veres, Victor Li, Leonid Peshkin, David A. Weitz, and Marc W. Kirschner. 2015. “Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells.” *Cell* 161 (5): 1187–1201.
Klein, Christine, and Ana Westenberger. 2012. “Genetics of Parkinson’s Disease.” *Cold Spring Harbor Perspectives in Medicine* 2 (1): a008888.
Knight, Philip A., and Daniel Ruiz. 2012. “A Fast Algorithm for Matrix Balancing.” *IMA Journal of Numerical Analysis* 33 (3): 1029–47.
Konermann, Silvana, Mark D. Brigham, Alexandro E. Trevino, Julia Joung, Omar O. Abudayyeh, Clea Barcena, Patrick D. Hsu, et al. 2014. “Genome-Scale Transcriptional Activation by an Engineered CRISPR-Cas9 Complex.” *Nature* 517 (7536): 583–88.
Korinek, Miloslav, Inmaculada M. Gonzalez-Gonzalez, Tereza Smejkalova, Dragana Hajdukovic, Kristyna Skrenkova, Jan Krusek, Martin Horak, and Ladislav Vyklicky. 2020.

"Cholesterol Modulates Presynaptic and Postsynaptic Properties of Excitatory Synaptic Transmission." *Scientific Reports* 10 (1): 1–18.
Korsunsky, Ilya, Nghia Millard, Jean Fan, Kamil Slowikowski, Fan Zhang, Kevin Wei, Yuriy Baglaenko, Michael Brenner, Po-Ru Loh, and Soumya Raychaudhuri. 2019. "Fast, Sensitive and Accurate Integration of Single-Cell Data with Harmony." *Nature Methods* 16 (12): 1289–96.
Kothary, R., S. Clapoff, S. Darling, M. D. Perry, L. A. Moran, and J. Rossant. 1989. "Inducible Expression of an hsp68-lacZ Hybrid Gene in Transgenic Mice." *Development* 105 (4): 707–14.
Koufi, Foteini-Dionysia, Irene Neri, Giulia Ramazzotti, Isabella Rusciano, Sara Mongiorgi, Maria Vittoria Marvi, Antonietta Fazio, et al. 2023. "Lamin B1 as a Key Modulator of the Developing and Aging Brain." *Frontiers in Cellular Neuroscience* 17 (August): 1263310.
Kraft, Katerina, Sinje Geuer, Anja J. Will, Wing Lee Chan, Christina Paliou, Marina Borschiwer, Izabela Harabula, et al. 2015. "Deletions, Inversions, Duplications: Engineering of Structural Variants Using CRISPR/Cas in Mice." *Cell Reports* 10 (5): 833–39.
Kraft, Katerina, Andreas Magg, Verena Heinrich, Christina Riemenschneider, Robert Schöpflin, Julia Markowski, Daniel M. Ibrahim, et al. 2019. "Serial Genomic Inversions Induce Tissue-Specific Architectural Stripes, Gene Misexpression and Congenital Malformations." *Nature Cell Biology* 21 (3): 305–10.
Kruse, Kai, Clemens B. Hug, and Juan M. Vaquerizas. 2020. "FAN-C: A Feature-Rich Framework for the Analysis and Visualisation of Chromosome Conformation Capture Data." *Genome Biology* 21 (1): 303.
Kurniawan, Nyoman D. 2018. "MRI in the Study of Animal Models of Neurodegenerative Diseases." *Methods in Molecular Biology* 1718: 347–75.
Kvon, Evgeny Z., Olga K. Kamneva, Uirá S. Melo, Iros Barozzi, Marco Osterwalder, Brandon J. Mannion, Virginie Tissières, et al. 2016. "Progressive Loss of Function in a Limb Enhancer during Snake Evolution." *Cell* 167 (3): 633–42.e11.
La Manno, Gioele, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, et al. 2018. "RNA Velocity of Single Cells." *Nature* 560 (7719): 494–98.
Larson, Matthew H., Luke A. Gilbert, Xiaowo Wang, Wendell A. Lim, Jonathan S. Weissman, and Lei S. Qi. 2013. "CRISPR Interference (CRISPRi) for Sequence-Specific Control of Gene Expression." *Nature Protocols* 8 (11): 2180–96.
Lasunción, Miguel A., Javier Martínez-Botas, Covadonga Martín-Sánchez, Rebeca Busto, and Diego Gómez-Coronado. 2022. "Cell Cycle Dependence on the Mevalonate Pathway: Role of Cholesterol and Non-Sterol Isoprenoids." *Biochemical Pharmacology* 196 (February): 114623.
Lee, John M., Yiping Tu, Angelica Tatar, Daniel Wu, Chika Nobumori, Hea-Jin Jung, Yuko Yoshinaga, et al. 2014. "Reciprocal Knock-in Mice to Investigate the Functional Redundancy of Lamin B1 and Lamin B2." *Molecular Biology of the Cell* 25 (10): 1666–75.
Leipold, Enrico, Lutz Liebmann, G. Christoph Korenke, Theresa Heinrich, Sebastian Giesselmann, Jonathan Baets, Matthias Ebbinghaus, et al. 2013. "A de Novo Gain-of-Function Mutation in SCN11A Causes Loss of Pain Perception." *Nature Genetics* 45 (11): 1399–1404.
Lemaître, Hervé, Pierre Augé, Ana Saitovitch, Alice Vinçon-Leite, Jean-Marc Tacchella, Ludovic Fillon, Raphael Calmon, et al. 2021. "Rest Functional Brain Maturation during the First Year of Life." *Cerebral Cortex* 31 (3): 1776–85.
Lettice, Laura A., Sarah Daniels, Elizabeth Sweeney, Shanmugasundaram Venkataraman, Paul S. Devenney, Philippe Gautier, Harris Morrison, Judy Fantes, Robert E. Hill, and David R. FitzPatrick. 2011. "Enhancer-Adoption as a Mechanism of Human Developmental Disease." *Human Mutation* 32 (12): 1492–99.
Lettice, Laura A., Iain Williamson, Paul S. Devenney, Fiona Kilanowski, Julia Dorin, and Robert E. Hill. 2014. "Development of Five Digits Is Controlled by a Bipartite

Long-Range Cis-Regulator." *Development* 141 (8): 1715–25.
Levine, Mike. 2010. "Transcriptional Enhancers in Animal Development and Evolution." *Current Biology: CB* 20 (17): R754–63.
Liberzon, Arthur, Chet Birger, Helga Thorvaldsdóttir, Mahmoud Ghandi, Jill P. Mesirov, and Pablo Tamayo. 2015. "The Molecular Signatures Database (MSigDB) Hallmark Gene Set Collection." *Cell Systems* 1 (6): 417–25.
Lieberman-Aiden, Erez, Nynke L. van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, et al. 2009. "Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome." *Science* 326 (5950): 289–93.
"Life Span as a Biomarker." n.d. The Jackson Laboratory. Accessed September 15, 2023. https://www.jax.org/research-and-faculty/research-labs/the-harrison-lab/gerontology/life-span-as-a-biomarker.
Lin, Shu-Ting, Louis J. Ptácek, and Ying-Hui Fu. 2011. "Adult-Onset Autosomal Dominant Leukodystrophy: Linking Nuclear Envelope to Myelin." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 31 (4): 1163–66.
Li, Shengnan, Dexing Lin, Yunwei Zhang, Min Deng, Yongxing Chen, Bin Lv, Boshu Li, et al. 2022. "Genome-Edited Powdery Mildew Resistance in Wheat without Growth Penalties." *Nature* 602 (7897): 455–60.
Liu, Guanwen, Qiupeng Lin, Shuai Jin, and Caixia Gao. 2022. "The CRISPR-Cas Toolbox and Gene Editing Technologies." *Molecular Cell* 82 (2): 333–47.
Liu, Jing, Mujahid Ali, and Qi Zhou. 2020. "Establishment and Evolution of Heterochromatin." *Annals of the New York Academy of Sciences* 1476 (1): 59–77.
Lohan, S., M. Spielmann, S. C. Doelken, R. Flöttmann, F. Muhammad, S. M. Baig, M. Wajid, et al. 2014. "Microduplications Encompassing the Sonic Hedgehog Limb Enhancer ZRS Are Associated with Haas-Type Polysyndactyly and Laurin-Sandrow Syndrome." *Clinical Genetics* 86 (4): 318–25.
Lo Martire, Viviana, Sara Alvente, Stefano Bastianini, Chiara Berteotti, Cristiano Bombardi, Giovanna Calandra-Buonaura, Sabina Capellari, et al. 2018. "Mice Overexpressing Lamin B1 in Oligodendrocytes Recapitulate the Age-Dependent Motor Signs, but Not the Early Autonomic Cardiovascular Dysfunction of Autosomal-Dominant Leukodystrophy (ADLD)." *Experimental Neurology* 301 (Pt A): 1–12.
Lun, Melody P., Edwin S. Monuki, and Maria K. Lehtinen. 2015. "Development and Functions of the Choroid Plexus-Cerebrospinal Fluid System." *Nature Reviews. Neuroscience* 16 (8): 445–57.
Lupiáñez, Darío G., Katerina Kraft, Verena Heinrich, Peter Krawitz, Francesco Brancati, Eva Klopocki, Denise Horn, et al. 2015. "Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions." *Cell* 161 (5): 1012–25.
Lupiáñez, Darío G., Malte Spielmann, and Stefan Mundlos. 2016. "Breaking TADs: How Alterations of Chromatin Domains Result in Disease." *Trends in Genetics: TIG* 32 (4): 225–37.
Lupski, James R., and Pawel Stankiewicz. 2005. "Genomic Disorders: Molecular Mechanisms for Rearrangements and Conveyed Phenotypes." *PLoS Genetics* 1 (6): e49.
MacAulay, Nanna, Richard F. Keep, and Thomas Zeuthen. 2022. "Cerebrospinal Fluid Production by the Choroid Plexus: A Century of Barrier Research Revisited." *Fluids and Barriers of the CNS* 19 (1): 26.
MacDonald, B. K., O. C. Cockerell, J. W. A. S. Sander, and S. D. Shorvon. 2000. "The Incidence and Lifetime Prevalence of Neurological Disorders in a Prospective Community-Based Study in the UK." *Brain: A Journal of Neurology* 123 (4): 665–76.
Mahajani, Sameehan, Caterina Giacomini, Federica Marinaro, Davide De Pietri Tonelli, Andrea Contestabile, and Laura Gasparini. 2017. "Lamin B1 Levels Modulate Differentiation into Neurons during Embryonic Corticogenesis." *Scientific Reports* 7 (1): 1–11.
Macosko E Z, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas A R,

Kamitaki N, Martersteck E M, Trombetta J J, Weitz D A, Sanes J R, Shalek A K, Regev A, McCarroll S A.“Highly Parallel Genome-Wide Expression Profiling of Individual Cells Using Nanoliter Droplets.” 2015. *Cell* 161 (5): 1202–14.
Malinverno, Matteo, Claudio Maderna, Abdallah Abu Taha, Monica Corada, Fabrizio Orsenigo, Mariaelena Valentino, Federica Pisati, et al. 2019. “Endothelial Cell Clonal Expansion in the Development of Cerebral Cavernous Malformations.” *Nature Communications* 10 (1): 2761.
Mali, Prashant, Luhan Yang, Kevin M. Esvelt, John Aach, Marc Guell, James E. DiCarlo, Julie E. Norville, and George M. Church. 2013. “RNA-Guided Human Genome Engineering via Cas9.” *Science* 339 (6121): 823–26.
Martin, Beth K., Chengxiang Qiu, Eva Nichols, Melissa Phung, Rula Green-Gladden, Sanjay Srivatsan, Ronnie Blecher-Gonen, et al. 2022. “Optimized Single-Nucleus Transcriptional Profiling by Combinatorial Indexing.” *Nature Protocols* 18 (1): 188–207.
Matise, M. P., D. J. Epstein, H. L. Park, K. A. Platt, and A. L. Joyner. 1998. “Gli2 Is Required for Induction of Floor Plate and Adjacent Cells, but Not Most Ventral Neurons in the Mouse Central Nervous System.” *Development* 125 (15): 2759–70.
Mazzarella, R., and D. Schlessinger. 1998. “Pathological Consequences of Sequence Duplications in the Human Genome.” *Genome Research* 8 (10): 1007–21.
McArthur, Evonne, and John A. Capra. 2021. “Topologically Associating Domain Boundaries That Are Stable across Diverse Cell Types Are Evolutionarily Constrained and Enriched for Heritability.” *American Journal of Human Genetics* 108 (2): 269–83.
McDonald, David A., Robert Shenkar, Changbin Shi, Rebecca A. Stockton, Amy L. Akers, Melanie H. Kucherlapati, Raju Kucherlapati, et al. 2011. “A Novel Mouse Model of Cerebral Cavernous Malformations Based on the Two-Hit Mutation Hypothesis Recapitulates the Human Disease.” *Human Molecular Genetics* 20 (2): 211–22.
Meijer, Inge A., Ana A. Simoes-Lopes, Sandra Laurent, Tanya Katz, Judith St-Onge, Dominique J. Verlaan, Nicolas Dupré, et al. 2008. “A Novel Duplication Confirms the Involvement of 5q23.2 in Autosomal Dominant Leukodystrophy.” *Archives of Neurology* 65 (11): 1496–1501.
Melberg, A., L. Hallberg, H. Kalimo, and R. Raininko. 2006. “MR Characteristics and Neuropathology in Adult-Onset Autosomal Dominant Leukodystrophy with Autonomic Symptoms.” *AJNR. American Journal of Neuroradiology* 27 (4): 904–11.
Merkenschlager, Matthias, and Elphège P. Nora. 2016. “CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation.” *Annual Review of Genomics and Human Genetics* 17 (August): 17–43.
Mezaki, Naomi, Takeshi Miura, Kotaro Ogaki, Makoto Eriguchi, Yuri Mizuno, Kenichi Komatsu, Hiroki Yamazaki, et al. 2018. “Duplication and Deletion Upstream of in Autosomal Dominant Adult-Onset Leukodystrophy.” *Neurology. Genetics* 4 (6): e292.
Mifsud, Borbala, Filipe Tavares-Cadete, Alice N. Young, Robert Sugar, Stefan Schoenfelder, Lauren Ferreira, Steven W. Wingett, et al. 2015. “Mapping Long-Range Promoter Contacts in Human Cells with High-Resolution Capture Hi-C.” *Nature Genetics* 47 (6): 598–606.
Mikkelsen, Tarjei S., Manching Ku, David B. Jaffe, Biju Issac, Erez Lieberman, Georgia Giannoukos, Pablo Alvarez, et al. 2007. “Genome-Wide Maps of Chromatin State in Pluripotent and Lineage-Committed Cells.” *Nature* 448 (7153): 553–60.
Ming, Guo-Li, and Hongjun Song. 2011. “Adult Neurogenesis in the Mammalian Brain: Significant Answers and Significant Questions.” *Neuron* 70 (4): 687–702.
Mittnenzweig, Markus, Yoav Mayshar, Saifeng Cheng, Raz Ben-Yair, Ron Hadas, Yoach Rais, Elad Chomsky, et al. 2021. “A Single-Embryo, Single-Cell Time-Resolved Model for Mouse Gastrulation.” *Cell* 184 (11): 2825–42.e22.
Mohammed, Hisham, Irene Hernando-Herraez, Aurora Savino, Antonio Scialdone, Iain Macaulay, Carla Mulas, Tamir Chandra, et al. 2017. “Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation.” *Cell Reports* 20 (5): 1215–28.
Molloy, A., O. Cotter, R. van Spaendonk, E. Sistermans, and B. Sweeney. 2012. “A Patient

with a Rare Leukodystrophy Related to Lamin B1 Duplication." *Irish Medical Journal* 105 (6): 186–87.
Moore, Jill E., Michael J. Purcaro, Henry E. Pratt, Charles B. Epstein, Noam Shoresh, Jessika Adrian, Trupti Kawli, et al. 2020. "Expanded Encyclopaedias of DNA Elements in the Human and Mouse Genomes." *Nature* 583 (7818): 699–710.
Morell, Pierre, and Richard H. Quarles. 1999. "Myelin Formation, Structure and Biochemistry." In *Basic Neurochemistry: Molecular, Cellular and Medical Aspects. 6th Edition*. Lippincott-Raven.
Mo, R., A. M. Freer, D. L. Zinyk, M. A. Crackower, J. Michaud, H. H. Heng, K. W. Chik, et al. 1997. "Specific and Redundant Functions of Gli2 and Gli3 Zinc Finger Genes in Skeletal Patterning and Development." *Development* 124 (1): 113–23.
Mossbrugger, I., G. Hoelzlwimmer, J. Calzada-Wack, and L. Quintanilla-Martinez. 2007. "Standardized Morphological Phenotyping of Mouse Models of Human Diseases within the German Mouse Clinic." *Verhandlungen Der Deutschen Gesellschaft Fur Pathologie* 91: 98–103.
"Multiscale 3D Genome Rewiring during Mouse Neural Development." 2017. *Cell* 171 (3): 557–72.e24.
Nanni, Luca, Stefano Ceri, and Colin Logie. 2020. "Spatial Patterns of CTCF Sites Define the Anatomy of TADs and Their Boundaries." *Genome Biology* 21 (1): 1–25.
Nathoo, Nabeela, Smriti Agrawal, Ying Wu, Sarah Haylock-Jacobs, V. Wee Yong, Tad Foniok, Samuel Barnes, Andre Obenaus, and Jeff F. Dunn. 2013. "Susceptibility-Weighted Imaging in the Experimental Autoimmune Encephalomyelitis Model of Multiple Sclerosis Indicates Elevated Deoxyhemoglobin, Iron Deposition and Demyelination." *Multiple Sclerosis* 19 (6): 721–31.
Neri, Irene, Giulia Ramazzotti, Sara Mongiorgi, Isabella Rusciano, Marianna Bugiani, Luciano Conti, Margot Cousin, et al. 2023. "Understanding the Ultra-Rare Disease Autosomal Dominant Leukodystrophy: An Updated Review on Morpho-Functional Alterations Found in Experimental Models." *Molecular Neurobiology*, July. https://doi.org/10.1007/s12035-023-03461-1.
Nmezi, Bruce, Elisa Giorgio, Raili Raininko, Anna Lehman, Malte Spielmann, Mary Kay Koenig, Rahmat Adejumo, et al. 2019. "Genomic Deletions Upstream of Lamin B1 Lead to Atypical Autosomal Dominant Leukodystrophy." *Neurology. Genetics* 5 (1): e305.
Nmezi, Bruce, Laura L. Vollmer, Tong Ying Shun, Albert Gough, Harshvardhan Rolyan, Fang Liu, Yumeng Jia, Quasar S. Padiath, and Andreas Vogt. 2020. "Development and Optimization of a High-Content Analysis Platform to Identify Suppressors of Lamin B1 Overexpression as a Therapeutic Strategy for Autosomal Dominant Leukodystrophy." *SLAS Discovery : Advancing Life Sciences R & D* 25 (8): 939–49.
Noori, Rabiya, Daniel Park, John D. Griffiths, Sonya Bells, Paul W. Frankland, Donald Mabbott, and Jeremie Lefebvre. 2020. "Activity-Dependent Myelination: A Glial Mechanism of Oscillatory Self-Organization in Large-Scale Brain Networks." *Proceedings of the National Academy of Sciences of the United States of America* 117 (24): 13227–37.
Northcott, Paul A., Catherine Lee, Thomas Zichner, Adrian M. Stütz, Serap Erkek, Daisuke Kawauchi, David J. H. Shih, et al. 2014. "Enhancer Hijacking Activates GFI1 Family Oncogenes in Medulloblastoma." *Nature* 511 (7510): 428–34.
Nurk, Sergey, Sergey Koren, Arang Rhie, Mikko Rautiainen, Andrey V. Bzikadze, Alla Mikheenko, Mitchell R. Vollger, et al. 2022. "The Complete Sequence of a Human Genome." *Science* 376 (6588): 44–53.
O'Kane, C. J., and W. J. Gehring. 1987. "Detection in Situ of Genomic Regulatory Elements in Drosophila." *Proceedings of the National Academy of Sciences of the United States of America* 84 (24): 9123–27.
Oudelaar, A. Marieke, Robert A. Beagrie, Matthew Gosden, Sara de Ornellas, Emily Georgiades, Jon Kerry, Daniel Hidalgo, et al. 2020. "Dynamics of the 4D Genome during in Vivo Lineage Specification and Differentiation." *Nature Communications* 11 (1): 1–12.
Oudelaar, A. Marieke, Damien J. Downes, James O. J. Davies, and Jim R. Hughes. 2017.

"Low-Input Capture-C: A Chromosome Conformation Capture Assay to Analyze Chromatin Architecture in Small Numbers of Cells." *Bio-Protocol* 7 (23). https://doi.org/10.21769/BioProtoc.2645.
Packer, Jonathan S., Qin Zhu, Chau Huynh, Priya Sivaramakrishnan, Elicia Preston, Hannah Dueck, Derek Stefanik, et al. 2019. "A Lineage-Resolved Molecular Atlas of Embryogenesis at Single-Cell Resolution." *Science* 365 (6459). https://doi.org/10.1126/science.aax1971.
Padiath, Quasar S. 2016. "Lamin B1 Mediated Demyelination: Linking Lamins, Lipids and Leukodystrophies." *Nucleus* 7 (6): 547–53.
———. 2019. "Autosomal Dominant Leukodystrophy: A Disease of the Nuclear Lamina." *Frontiers in Cell and Developmental Biology* 7 (March): 41.
Padiath, Quasar Saleem, and Ying-Hui Fu. 2010. "Autosomal Dominant Leukodystrophy Caused by Lamin B1 Duplications a Clinical and Molecular Case Study of Altered Nuclear Function and Disease." *Methods in Cell Biology* 98: 337–57.
Padiath, Quasar S., Kazumasa Saigoh, Raphael Schiffmann, Hideaki Asahara, Takeshi Yamada, Anulf Koeppen, Kirk Hogan, Louis J. Ptácek, and Ying-Hui Fu. 2006. "Lamin B1 Duplications Cause Autosomal Dominant Leukodystrophy." *Nature Genetics* 38 (10): 1114–23.
Peng, Yanling, and Yubo Zhang. 2018. "Enhancer and Super-Enhancer: Positive Regulators in Gene Transcription." *Animal Models and Experimental Medicine* 1 (3): 169–79.
Pennacchio, Len A., Wendy Bickmore, Ann Dean, Marcelo A. Nobrega, and Gill Bejerano. 2013. "Enhancers: Five Essential Questions." *Nature Reviews. Genetics* 14 (4): 288–95.
Petrenko, Natalia, Yi Jin, Koon Ho Wong, and Kevin Struhl. 2016. "Mediator Undergoes a Compositional Change during Transcriptional Activation." *Molecular Cell* 64 (3): 443–54.
Picelli, Simone, Åsa K. Björklund, Omid R. Faridani, Sven Sagasser, Gösta Winberg, and Rickard Sandberg. 2013. "Smart-seq2 for Sensitive Full-Length Transcriptome Profiling in Single Cells." *Nature Methods* 10 (11): 1096–98.
Pijuan-Sala, Blanca, Jonathan A. Griffiths, Carolina Guibentif, Tom W. Hiscock, Wajid Jawaid, Fernando J. Calero-Nieto, Carla Mulas, et al. 2019. "A Single-Cell Molecular Map of Mouse Gastrulation and Early Organogenesis." *Nature* 566 (7745): 490–95.
Qiu, Chengxiang, Junyue Cao, Beth K. Martin, Tony Li, Ian C. Welsh, Sanjay Srivatsan, Xingfan Huang, et al. 2022. "Systematic Reconstruction of Cellular Trajectories across Mouse Embryogenesis." *Nature Genetics* 54 (3): 328–41.
Raihan, O.; Brishti, A.; Molla, R.; Li, W.; Zhang, Q.; Xu, P.; Khan, M.I.; Zhang, J.; Liu, Q. "The Age-Dependent Elevation of miR-335-3p Leads to Reduced Cholesterol and Impaired Memory in Brain." 2018. *Neuroscience* 390 (October): 160–73.
Raininko, Raili, Michael Gosky, and Quasar S. Padiath. 2016. "-Related Autosomal Dominant Leukodystrophy." In *GeneReviews*, edited by Margaret P. Adam, Ghayda M. Mirzaa, Roberta A. Pagon, Stephanie E. Wallace, Lora J. H. Bean, Karen W. Gripp, and Anne Amemiya. Seattle (WA): University of Washington, Seattle.
Rajderkar, Sudha, Iros Barozzi, Yiwen Zhu, Rong Hu, Yanxiao Zhang, Bin Li, Ana Alcaina Caro, et al. 2023. "Topologically Associating Domain Boundaries Are Required for Normal Genome Function." *Communications Biology* 6 (1): 1–10.
Rajewsky, Nikolaus, Geneviève Almouzni, Stanislaw A. Gorski, Stein Aerts, Ido Amit, Michela G. Bertero, Christoph Bock, et al. 2020. "LifeTime and Improving European Healthcare through Cell-Based Interceptive Medicine." *Nature* 587 (7834): 377–86.
Ran, F. Ann, Patrick D. Hsu, Jason Wright, Vineeta Agarwala, David A. Scott, and Feng Zhang. 2013. "Genome Engineering Using the CRISPR-Cas9 System." *Nature Protocols* 8 (11): 2281–2308.
Rao, Suhas S. P., Miriam H. Huntley, Neva C. Durand, Elena K. Stamenova, Ivan D. Bochkov, James T. Robinson, Adrian L. Sanborn, et al. 2014. "A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping." *Cell* 159 (7): 1665–80.
Rappard, Diane F. van, Marsh Königs, Marjan E. Steenweg, Jaap Jan Boelens, Jaap Oosterlaan, Marjo S. van der Knaap, Nicole I. Wolf, and Petra J. W. Pouwels. 2018.

"Diffusion Tensor Imaging in Metachromatic Leukodystrophy." *Journal of Neurology* 265 (3): 659–68.
Ratti, Stefano, Isabella Rusciano, Sara Mongiorgi, Irene Neri, Alessandra Cappellini, Pietro Cortelli, Pann-Ghill Suh, et al. 2021. "Lamin B1 Accumulation's Effects on Autosomal Dominant Leukodystrophy (ADLD): Induction of Reactivity in the Astrocytes." *Cells* 10 (10). https://doi.org/10.3390/cells10102566.
Ratti, Stefano, Isabella Rusciano, Sara Mongiorgi, Eric Owusu Obeng, Alessandra Cappellini, Gabriella Teti, Mirella Falconi, et al. 2021. "Cell Signaling Pathways in Autosomal-Dominant Leukodystrophy (ADLD): The Intriguing Role of the Astrocytes." *Cellular and Molecular Life Sciences: CMLS* 78 (6): 2781–95.
Ricciardi, Sara, Federica Ungaro, Melanie Hambrock, Nils Rademacher, Gilda Stefanelli, Dario Brambilla, Alessandro Sessa, et al. 2012. "CDKL5 Ensures Excitatory Synapse Stability by Reinforcing NGL-1-PSD95 Interaction in the Postsynaptic Compartment and Is Impaired in Patient iPSC-Derived Neurons." *Nature Cell Biology* 14 (9): 911–23.
Ring, Karen L., Mahru C. An, Ningzhe Zhang, Robert N. O'Brien, Eliana Marisa Ramos, Fuying Gao, Robert Atwood, et al. 2015. "Genomic Analysis Reveals Disruption of Striatal Neuronal Development and Therapeutic Targets in Human Huntington's Disease Neural Stem Cells." *Stem Cell Reports* 5 (6): 1023–38.
Ringel AR, Szabo Q, Chiariello AM, Chudzik K, Schöpflin R, Rothe P, Mattei AL, Zehnder T, Harnett D, Laupert V, Bianco S, Hetzel S, Glaser J, Phan MHQ, Schindler M, IbrahimDM, Paliou C, Esposito A, PradaMedina CA, Haas SA, Giere P, Vingron M, Wittler L, Meissner A, Nicodemi M, Cavalli G, Bantignies F, Mundlos S, Robson MI."Repression and 3D-Restructuring Resolves Regulatory Conflicts in Evolutionarily Rearranged Genomes." 2022. *Cell* 185 (20): 3689–3704.e21.
Robson, Michael I., Alessa R. Ringel, and Stefan Mundlos. 2019. "Regulatory Landscaping: How Enhancer-Promoter Communication Is Sculpted in 3D." *Molecular Cell* 74 (6): 1110–22.
Rockich, Briana E., Steven M. Hrycaj, Hung Ping Shih, Melinda S. Nagy, Michael A. H. Ferguson, Janel L. Kopp, Maike Sander, Deneen M. Wellik, and Jason R. Spence. 2013. "Sox9 Plays Multiple Roles in the Lung Epithelium during Branching Morphogenesis." *Proceedings of the National Academy of Sciences of the United States of America* 110 (47): E4456–64.
Rodriques, Samuel G., Robert R. Stickels, Aleksandrina Goeva, Carly A. Martin, Evan Murray, Charles R. Vanderburg, Joshua Welch, Linlin M. Chen, Fei Chen, and Evan Z. Macosko. 2019. "Slide-Seq: A Scalable Technology for Measuring Genome-Wide Expression at High Spatial Resolution." *Science* 363 (6434): 1463–67.
Rolyan, Harshvardhan, Yulia Y. Tyurina, Marylens Hernandez, Andrew A. Amoscato, Louis J. Sparvero, Bruce C. Nmezi, Yue Lu, et al. 2015. "Defects of Lipid Synthesis Are Linked to the Age-Dependent Demyelination Caused by Lamin B1 Overexpression." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 35 (34): 12002–17.
Rosenthal, Nadia, and Steve Brown. 2007. "The Mouse Ascending: Perspectives for Human-Disease Models." *Nature Cell Biology* 9 (9): 993–99.
Roy, Ananda L., and Dinah S. Singer. 2015. "Core Promoters in Transcription: Old Problem, New Insights." *Trends in Biochemical Sciences* 40 (3): 165–71.
Rozenblatt-Rosen, Orit, Aviv Regev, Philipp Oberdoerffer, Tal Nawy, Anna Hupalowska, Jennifer E. Rood, Orr Ashenberg, et al. 2020. "The Human Tumor Atlas Network: Charting Tumor Transitions across Space and Time at Single-Cell Resolution." *Cell* 181 (2): 236–49.
Sabari, Benjamin R., Alessandra Dall'Agnese, Ann Boija, Isaac A. Klein, Eliot L. Coffey, Krishna Shrinivas, Brian J. Abraham, et al. 2018. "Coactivator Condensation at Super-Enhancers Links Phase Separation and Gene Control." *Science* 361 (6400). https://doi.org/10.1126/science.aar3958.
Samad, Tahmina, and Sean M. Wu. 2021. "Single Cell RNA Sequencing Approaches to Cardiac Development and Congenital Heart Disease." *Seminars in Cell &*

Developmental Biology 118 (October): 129–35.
Schafflick, David, Chenling A. Xu, Maike Hartlehnert, Michael Cole, Andreas Schulte-Mecklenbeck, Tobias Lautwein, Jolien Wolbert, et al. 2020. “Integrated Single Cell Analysis of Blood and Cerebrospinal Fluid Leukocytes in Multiple Sclerosis.” *Nature Communications* 11 (1): 247.
Schuster, Jens, Jimmy Sundblom, Ann-Charlotte Thuresson, Sharon Hassin-Baer, Thomas Klopstock, Martin Dichgans, Oren S. Cohen, Raili Raininko, Atle Melberg, and Niklas Dahl. 2011. “Genomic Duplications Mediate Overexpression of Lamin B1 in Adult-Onset Autosomal Dominant Leukodystrophy (ADLD) with Autonomic Symptoms.” *Neurogenetics* 12 (1): 65–72.
Schwamborn, Jens C. 2018. “Is Parkinson’s Disease a Neurodevelopmental Disorder and Will Brain Organoids Help Us to Understand It?” *Stem Cells and Development* 27 (14): 968–75.
Scott, Charlotte E., Sarah L. Wynn, Abdul Sesay, Catarina Cruz, Martin Cheung, Maria-Victoria Gomez Gaviro, Sarah Booth, et al. 2010. “SOX9 Induces and Maintains Neural Stem Cells.” *Nature Neuroscience* 13 (10): 1181–89.
Selvaraj, Sridhar, Neha R. Dhoke, James Kiley, Alba Judith Mateos-Aierdi, Sudheer Tungtur, Ricardo Mondragon-Gonzalez, Grace Killeen, Vanessa K. P. Oliveira, Adolfo López de Munain, and Rita C. R. Perlingeiro. 2019. “Gene Correction of LGMD2A Patient-Specific iPSCs for the Development of Targeted Autologous Cell Therapy.” *Molecular Therapy: The Journal of the American Society of Gene Therapy* 27 (12): 2147–57.
Sexton, Tom, Eitan Yaffe, Ephraim Kenigsberg, Frédéric Bantignies, Benjamin Leblanc, Michael Hoichman, Hugues Parrinello, Amos Tanay, and Giacomo Cavalli. 2012. “Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome.” *Cell* 148 (3): 458–72.
Shalem, Ophir, Neville E. Sanjana, and Feng Zhang. 2015. “High-Throughput Functional Genomics Using CRISPR–Cas9.” *Nature Reviews. Genetics* 16 (5): 299–311.
Shen, Yin, Feng Yue, David F. McCleary, Zhen Ye, Lee Edsall, Samantha Kuan, Ulrich Wagner, et al. 2012. “A Map of the Cis-Regulatory Sequences in the Mouse Genome.” *Nature* 488 (7409): 116–20.
Shimi, Takeshi, Veronika Butin-Israeli, Stephen A. Adam, Robert B. Hamanaka, Anne E. Goldman, Catherine A. Lucas, Dale K. Shumaker, Steven T. Kosak, Navdeep S. Chandel, and Robert D. Goldman. 2011. “The Role of Nuclear Lamin B1 in Cell Proliferation and Senescence.” *Genes & Development* 25 (24): 2579–93.
Silva-Vargas, Violeta, Angel R. Maldonado-Soto, Dogukan Mizrak, Paolo Codega, and Fiona Doetsch. 2016. “Age-Dependent Niche Signals from the Choroid Plexus Regulate Adult Neural Stem Cells.” *Cell Stem Cell* 19 (5): 643–52.
Singh, Pushpendra, Roopali Saxena, Gunda Srinivas, Gopal Pande, and Amitabha Chattopadhyay. 2013. “Cholesterol Biosynthesis and Homeostasis in Regulation of the Cell Cycle.” *PloS One* 8 (3): e58833.
Smajić, Semra, Cesar A. Prada-Medina, Zied Landoulsi, Jenny Ghelfi, Sylvie Delcambre, Carola Dietrich, Javier Jarazo, et al. 2021. “Single-Cell Sequencing of Human Midbrain Reveals Glial Activation and a Parkinson-Specific Neuronal State.” *Brain: A Journal of Neurology* 145 (3): 964–78.
Solovei, Irina, Audrey S. Wang, Katharina Thanisch, Christine S. Schmidt, Stefan Krebs, Monika Zwerger, Tatiana V. Cohen, et al. 2013. “LBR and Lamin A/C Sequentially Tether Peripheral Heterochromatin and Inversely Regulate Differentiation.” *Cell* 152 (3): 584–98.
Spielmann, Malte, Darío G. Lupiáñez, and Stefan Mundlos. 2018. “Structural Variation in the 3D Genome.” *Nature Reviews. Genetics* 19 (7): 453–67.
Spielmann, Malte, and Stefan Mundlos. 2016. “Looking beyond the Genes: The Role of Non-Coding Variants in Human Disease.” *Human Molecular Genetics* 25 (R2): R157–65.
Sreenivasan, Varun K. A., Riccardo Dore, Julia Resch, Julia Maier, Carola Dietrich, Jana Henck, Saranya Balachandran, Jens Mittag, and Malte Spielmann. 2023. “Single-Cell RNA-Based Phenotyping Reveals a Pivotal Role of Thyroid Hormone Receptor Alpha for

Hypothalamic Development." *Development* 150 (3): dev201228.
Srivatsan, M.C. Regier, E. Barkan, J.M. Franks, J.S. Packer, P. Grosjean, M. Duran, S. Saxton, J.J. Ladd, M. Spielmann, et al. ,Embryo-scale, single-cell spatial transcriptomics. Science373,111-117(2021).DOI:10.1126/science.abb9536
Ståhl, Patrik L., Fredrik Salmén, Sanja Vickovic, Anna Lundmark, José Fernández Navarro, Jens Magnusson, Stefania Giacomello, et al. 2016. "Visualization and Analysis of Gene Expression in Tissue Sections by Spatial Transcriptomics." *Science* 353 (6294): 78–82.
Stamataki, Despina, Fausto Ulloa, Stavroula V. Tsoni, Anita Mynett, and James Briscoe. 2005. "A Gradient of Gli Activity Mediates Graded Sonic Hedgehog Signaling in the Neural Tube." *Genes & Development* 19 (5): 626–41.
Stottmann, R. W., P. V. Tran, A. Turbe-Doan, and D. R. Beier. 2009. "Ttc21b Is Required to Restrict Sonic Hedgehog Activity in the Developing Mouse Forebrain." *Developmental Biology* 335 (1): 166–78.
Sundblom, J., A. Melberg, H. Kalimo, A. Smits, and R. Raininko. 2009. "MR Imaging Characteristics and Neuropathology of the Spinal Cord in Adult-Onset Autosomal Dominant Leukodystrophy with Autonomic Symptoms." *AJNR. American Journal of Neuroradiology* 30 (2): 328–35.
Swaddiwudhipong, Nol, David J. Whiteside, Frank H. Hezemans, Duncan Street, James B. Rowe, and Timothy Rittman. 2023. "Pre-Diagnostic Cognitive and Functional Impairment in Multiple Sporadic Neurodegenerative Diseases." *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 19 (5): 1752–63.
Szabo, Quentin, Frédéric Bantignies, and Giacomo Cavalli. 2019. "Principles of Genome Folding into Topologically Associating Domains." *Science Advances* 5 (4): eaaw1668.
Tan, Longzhi, Jenny Shi, Siavash Moghadami, Bibudha Parasar, Cydney P. Wright, Yunji Seo, Kristen Vallejo, et al. 2023. "Lifelong Restructuring of 3D Genome Architecture in Cerebellar Granule Cells." *Science* 381 (6662): 1112–19.
Terlizzi, Rossana, Giovanna Calandra-Buonaura, Stefano Zanigni, Giorgio Barletta, Sabina Capellari, Pietro Guaraldi, Vincenzo Donadio, et al. 2016. "A Longitudinal Study of a Family with Adult-Onset Autosomal Dominant Leukodystrophy: Clinical, Autonomic and Neuropsychological Findings." *Autonomic Neuroscience: Basic & Clinical* 195 (February): 20–26.
Thompson, Derick, Catherine A. Brissette, and John A. Watt. 2022. "The Choroid Plexus and Its Role in the Pathogenesis of Neurological Infections." *Fluids and Barriers of the CNS* 19 (1): 1–20.
Thurman, Robert E., Eric Rynes, Richard Humbert, Jeff Vierstra, Matthew T. Maurano, Eric Haugen, Nathan C. Sheffield, et al. 2012. "The Accessible Chromatin Landscape of the Human Genome." *Nature* 489 (7414): 75–82.
Tirosh, Itay, Benjamin Izar, Sanjay M. Prakadan, Marc H. Wadsworth 2nd, Daniel Treacy, John J. Trombetta, Asaf Rotem, et al. 2016. "Dissecting the Multicellular Ecosystem of Metastatic Melanoma by Single-Cell RNA-Seq." *Science* 352 (6282): 189–96.
Toyama, Brandon H., Jeffrey N. Savas, Sung Kyu Park, Michael S. Harris, Nicholas T. Ingolia, John R. Yates 3rd, and Martin W. Hetzer. 2013. "Identification of Long-Lived Proteins Reveals Exceptional Stability of Essential Cellular Structures." *Cell* 154 (5): 971–82.
Underwood, Clare K., Nyoman D. Kurniawan, Tim J. Butler, Gary J. Cowin, and Robyn H. Wallace. 2011. "Non-Invasive Diffusion Tensor Imaging Detects White Matter Degeneration in the Spinal Cord of a Mouse Model of Amyotrophic Lateral Sclerosis." *NeuroImage* 55 (2): 455–61.
Urnov, Fyodor D. 2021. "The Cas9 Hammer-and Sickle: A Challenge for Genome Editors." *The CRISPR Journal* 4 (1): 6–13.
Uyehara, Christopher M., and Effie Apostolou. 2023. "3D Enhancer-Promoter Interactions and Multi-Connected Hubs: Organizational Principles and Functional Roles." *Cell Reports*, April, 112068.
Valenza, Fabiola, Davide Cittaro, Elia Stupka, Donatella Biancolini, Maria Grazia Patricelli, Dario Bonanomi, and Dejan Lazarević. 2019. "A Novel Truncating Variant of GLI2

Associated with Culler-Jones Syndrome Impairs Hedgehog Signalling." *PloS One* 14 (1): e0210097.
Vandamme, Thierry F. 2014. "Use of Rodents as Models of Human Diseases." *Journal of Pharmacy & Bioallied Sciences* 6 (1): 2–9.
Van Schependom, Jeroen, and Miguel D'haeseleer. 2023. "Advances in Neurodegenerative Diseases." *Journal of Clinical Medicine Research* 12 (5). https://doi.org/10.3390/jcm12051709.
Veenvliet, J. V., Lenne, P.-F., Turner, D. A., Nachman, I. & Trivedi, V. Sculpting with stem cells: how models of embryo development take shape. Development 148, dev192914 (2021).
Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, et al. 2001. "The Sequence of the Human Genome." *Science* 291 (5507): 1304–51.
Vergnes, Laurent, Miklós Péterfy, Martin O. Bergo, Stephen G. Young, and Karen Reue. 2004. "Lamin B1 Is Required for Mouse Development and Nuclear Integrity." *Proceedings of the National Academy of Sciences of the United States of America* 101 (28): 10428–33.
Visel, Axel, Simon Minovitsky, Inna Dubchak, and Len A. Pennacchio. 2007. "VISTA Enhancer Browser--a Database of Tissue-Specific Human Enhancers." *Nucleic Acids Research* 35 (Database issue): D88–92.
Wagner, Daniel E., Caleb Weinreb, Zach M. Collins, James A. Briggs, Sean G. Megason, and Allon M. Klein. 2018. "Single-Cell Mapping of Gene Expression Landscapes and Lineage in the Zebrafish Embryo." *Science* 360 (6392): 981–87.
Wagner, T., J. Wirth, J. Meyer, B. Zabel, M. Held, J. Zimmer, J. Pasantes, et al. 1994. "Autosomal Sex Reversal and Campomelic Dysplasia Are Caused by Mutations in and around the SRY-Related Gene SOX9." *Cell* 79 (6): 1111–20.
Waltz, Emily. 2022. "GABA-Enriched Tomato Is First CRISPR-Edited Food to Enter Market." *Nature Biotechnology* 40 (1): 9–11.
Wichmann, Thea Overgaard, Helle Hasager Damkier, and Michael Pedersen. 2022. "A Brief Overview of the Cerebrospinal Fluid System and Its Implications for Brain and Spinal Cord Diseases." *Frontiers in Human Neuroscience* 15 (January): 737217.
Will, Anja J., Giulia Cova, Marco Osterwalder, Wing-Lee Chan, Lars Wittler, Norbert Brieske, Verena Heinrich, et al. 2017. "Composition and Dosage of a Multipartite Enhancer Cluster Control Developmental Expression of Ihh (Indian Hedgehog)." *Nature Genetics* 49 (10): 1539–45.
Winick-Ng, Warren, Alexander Kukalev, Izabela Harabula, Luna Zea-Redondo, Dominik Szabó, Mandy Meijer, Leonid Serebreni, et al. 2021. "Cell-Type Specialization Is Encoded by Specific Chromatin Topologies." *Nature* 599 (7886): 684–91.
Wolock, Samuel L., Romain Lopez, and Allon M. Klein. 2019. "Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data." *Cell Systems* 8 (4): 281–91.e9.
Xiong, Xushen, Benjamin T. James, Carles A. Boix, Yongjin P. Park, Kyriaki Galani, Matheus B. Victor, Na Sun, et al. 2023. "Epigenomic Dissection of Alzheimer's Disease Pinpoints Causal Variants and Reveals Epigenome Erosion." *Cell* 186 (20): 4422–37.e21.
Xu, Jie, Fan Song, Huijue Lyu, Mikoto Kobayashi, Baozhen Zhang, Ziyu Zhao, Ye Hou, et al. 2022. "Subtype-Specific 3D Genome Alteration in Acute Myeloid Leukaemia." *Nature* 611 (7935): 387–98.
Yao, Baojin, Qiuqing Wang, Chia-Feng Liu, Pallavi Bhattaram, Wei Li, Timothy J. Mead, James F. Crish, and Véronique Lefebvre. 2015. "The SOX9 Upstream Region Prone to Chromosomal Aberrations Causing Campomelic Dysplasia Contains Multiple Cartilage Enhancers." *Nucleic Acids Research* 43 (11): 5394–5408.
Yatskevich, Stanislau, James Rhodes, and Kim Nasmyth. 2019. "Organization of Chromosomal DNA by SMC Complexes." *Annual Review of Genetics* 53 (December): 445–82.
Yattah, Camila, Marylens Hernandez, Dennis Huang, Hyejin Park, Will Liao, and Patrizia Casaccia. 2020. "Dynamic Lamin B1-Gene Association During Oligodendrocyte

Progenitor Differentiation." *Neurochemical Research* 45 (3): 606–19.
Zappia, Luke, and Fabian J. Theis. 2021. "Over 1000 Tools Reveal Trends in the Single-Cell RNA-Seq Analysis Landscape." *Genome Biology* 22 (1): 1–18.
Zarghamian, Parinaz, Julia Klermund, and Toni Cathomen. 2022. "Clinical Genome Editing to Treat Sickle Cell Disease-A Brief Update." *Frontiers of Medicine* 9: 1065377.
Zerbin-Rüdin, E., and J. Peiffer. 1964. "[Genetic contribution to the problem of the late form of Pelizaeus-Merzbacher disease]." *Humangenetik* 1 (2): 107–22.
Zhang, Jian-Ping, Xiao-Lan Li, Guo-Hua Li, Wanqiu Chen, Cameron Arakaki, Gary D. Botimer, David Baylink, et al. 2017. "Efficient Precise Knockin with a Double Cut HDR Donor after CRISPR/Cas9-Mediated Double-Stranded DNA Cleavage." *Genome Biology* 18 (1): 1–18.
Zheng, Grace X. Y., Jessica M. Terry, Phillip Belgrader, Paul Ryvkin, Zachary W. Bent, Ryan Wilson, Solongo B. Ziraldo, et al. 2017. "Massively Parallel Digital Transcriptional Profiling of Single Cells." *Nature Communications* 8 (1): 1–12.
Zhou, Yingyue, Wilbur M. Song, Prabhakar S. Andhey, Amanda Swain, Tyler Levy, Kelly R. Miller, Pietro L. Poliani, et al. 2020. "Author Correction: Human and Mouse Single-Nucleus Transcriptomics Reveal TREM2-Dependent and TREM2-Independent Cellular Responses in Alzheimer's Disease." *Nature Medicine* 26 (6): 981.
Zhu, Yong-Chuan, and Zhi-Qi Xiong. 2019. "Molecular and Synaptic Bases of CDKL5 Disorder." *Developmental Neurobiology* 79 (1): 8–19.
Zuin, Jessica, Gregory Roth, Yinxiu Zhan, Julie Cramard, Josef Redolfi, Ewa Piskadlo, Pia Mach, et al. 2022. "Nonlinear Control of Transcription through Enhancer–promoter Interactions." *Nature* 604 (7906): 571–77.

9. Acknowledgements

Ich möchte mich allen voran bei meinem Supervisor Malte Spielmann bedanken. Danke, dass du mir die Chance gegeben hast, an diesen Projekten zu arbeiten, selbst zu entscheiden, in welche Richtung die Projekte gehen und mich darin unterstützt hast, ein Selbstbewusstsein für meine Arbeit zu entwickeln. Alle Möglichkeiten, Herausforderungen und Hindernisse, haben wir meiner Meinung nach gut bewerkstelligt, besonders für eine Long-Distance-Supervision.

Danke an mein Thesis Advisory Committee: Edda Schulz, Jan-Philipp Junker und Sigmar Stricker für die hilfreichen Kommentare und die Unterstützung der Projekte mit Ideen. Besonderer Dank geht an Sigmar für die Übernahme der Rolle des Zweitgutachters.

Thank you to Jay Shendure, for the opportunity to work in your lab. It was an amazing experience to get to know the team, the projects and Seattle. I was very warmly welcomed by everyone and it was a pleasure to extend the work with Jun, Chengxiang and Xingfan to a long standing collaboration and the weekly meetings from which I learned tremendously.

Danke an Stefan Mundlos und das gesamte Mundlos-Labor, past and present für die Unterstützung, die familiäre Atmosphäre und die Aufnahme als eines eurer Gruppenmitglieder. Die Meetings, der Alltag im Labor und die Zeit mit euch haben meinen PhD ausgemacht. Ich habe besonders Asita und Ute zu danken, die mir immer geholfen haben, auch wenn ich längst wissen sollte, wo etwas steht und wie etwas funktioniert und immer Zeit haben, für eine nette Unterhaltung, ihr haltet uns zusammen! Danke auch an Carola, du hast mich in der Anfangszeit so wunderbar unterstützt, wir waren ein tolles Team. Ein lieber Dank geht an Alex, Alessa und insbesondere Andreas, welche mich am Anfang an die Hand genommen und mir alles gezeigt haben, meine erste Office Crew waren und den Grundstein für alle Laborkenntnisse gesetzt haben. Thank you to Cesar who set up what can only be defined as a sisyphus task for me to learn to analyze my own data. Thank you for motivating me to do this on my own. Thank you to Varun and Saranya for continuing this journey with me. I know I will never be a bioinformatic genius like you three, but you made me feel empowered to analyze my own data and that was a central experience of my PhD. Thank you to everyone else who got involved in the projects directly: Mikie and Josh for the support on the capture HiC front, Susanne und Philipp für eure Unterstützung bei dem MRT Abenteuer, Norbert für die Hilfe bei den single cell Experimenten und Corinna für die wundervolle Arbeit an den Mauslinien.

Abseits der Arbeit, aber ganz zentral für meine persönliche Entwicklung, möchte ich an allererster Stelle meiner Familie danken. Dank geht an meine Eltern, dass ihr immer an mich und die Entscheidung, diesen Weg einzuschlagen, geglaubt habt und mich

unterstützt habt, wo ihr konntet. Nina, du bist so unglaublich wichtig in allem, was ich tue, danke, du bist das Beste an Allem. An Helmut, Gina, Heike, Robin und alle anderen Familienmitglieder, danke für eure Unterstützung über die Jahre und die Fürsorge, mit der ihr für mich da seid.

An alle meine Freunde, danke, dass ihr mir den Rücken stärkt, mich aus dem Alltag reißt, mich inspiriert, zum Lachen bringt und egal über welche Distanz, immer ganz nah bei mir seid. Besonderer Dank für diese Arbeit geht an Johannes, der diese Arbeit Korrektur gelesen hat und meine Projekte mittlerweile am zweitbesten kennt und an Lisa-Kristin für den Feinschliff, danke euch beiden.

An Simon, meinen Partner, möchte ich einen besonderen Dank aussprechen. Für deine jahrelange Unterstützung, dein Verständnis und für all die Zeit, in der ich an dieser Arbeit saß und in der ich plötzlich andere Prioritäten hatte als geplant. Ich schätze alles, was wir uns gemeinsam erarbeitet und aufgebaut haben.

And in the words of the great Snoop Dogg:

“I want to thank me for believing in me. I want to thank me for doing all this hard work.”

10. Scientific Publications

Huang X.*,**Henck J.***, Qiu C.*, Sreenivasan V.K.A., Balachandran S., Amarie O.V., Hrabe de Angelis M., Behncke R., Chan W.L., Despang A., Dickel D.E., Duran M., Feuchtinger A., Fuchs H., Gailus-Durner V., Haag N., Hägerling R., Hansmeier N., Hennig F., Marshall C., Rajderkar S., Ringel A., Robson M., Saunders L., Silva-Buttkus P. d., Spielmann N., Srivatsan S.R., Ulferts S., Wittler L., Zhu Y., Kalscheuer V.M., Ibrahim D., Kurth I., Kornak U., Visel A., Pennacchio L.A., Beier D. R., Trapnell C., Cao J., Shendure J., Spielmann M. *equal contribution
Single cell, whole embryo phenotyping of pleiotropic disorders of mammalian development *(Nature, 2023)* DOI: 10.1038/s41586-023-06548-w

Sreenivasan V. K. A, Dore R., Resch J., Maier J., Dietrich C., Balachandran S., **Henck J.**, Mittag J., Spielmann M.
Single-cell RNA-based phenotyping reveals a pivotal role of thyroid hormone receptor alpha for hypothalamic development *(Development, 2023)* https://doi.org/10.1242/dev.201228

Sreenivasan V. K. A, **Henck J.**, Spielmann M.
Single-cell sequencing: promises and challenges for human genetics *(Medizinische Genetik, 2022)* https://doi.org/10.1515/medgen-2022-2156

Smaji S.,Prada-Medina C.S.,Landoulsi Z., Ghel J., Delcambre S., Dietrich C., Jarazo J., **Henck J.**, Balachandran S., Pachchek S.,Morris C.M., Antony P., Timmermann B., Sauer S., Pereira S. L., Schwamborn J.C., May P., Grünewald A., Spielmann M.
Single-cell sequencing of human midbrain reveals glial activation and a Parkinson-specific neuronal state *(Brain, 2021)* https://doi.org/10.1093/brain/awab446

Kari V.,Raul S.K., **Henck J.**, Kitz J., Kramer F., Kosinsky R.K., Übelmesser N., Mansour W.Y., Eggert J., Spitzner M., Najafova Z. ,Bastians H., Grade M., Gaedcke J., Florian Wegwitz, Johnsen S.A.
The histone methyltransferase DOT1L is required for proper DNA damage response, DNA repair, and modulates chemotherapy responsiveness *(Clinical Epigenetics,2019)* DOI: 10.1186/s13148-018-0601-1

11. Scientific Conferences

Selected Talk & Invited Session Chair, European Society of Human Genetics -06/2023
Henck J. et al. “Position effects in the 3D genome as the cause of demyelinating leukodystrophy”

Invited Talk, Belgian Society of Human Genetics - 03/2023
Huang X.*,**Henck J.***, Qiu C.* et al. “Single cell, whole embryo phenotyping of pleiotropic disorders of mammalian development”

Selected Talk, European Society of Human Genetics - 06/2022
Huang X.*,**Henck J.***, Qiu C.* et al. “Single cell, whole embryo phenotyping of pleiotropic disorders of mammalian development”

Selected Talk, German Society of Human Genetics -03/2022
Huang X.*,**Henck J.***, Qiu C.* et al. “Single cell, whole embryo phenotyping of pleiotropic disorders of mammalian development”

Selected Talk, American Society of Human Genetics- 10/2021
Huang X.*,**Henck J.***, Qiu C.* et al. “Single cell, whole embryo phenotyping of pleiotropic disorders of mammalian development”