DISS. ETH NO. 28460

Discovery of functional enhancer-promoter pairs using brain organoids

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

Yan Jun Lan

MSc Pharmaceutical Sciences
University of Basel, Switzerland

born on 07.12.1991

citizen of
Vienna, Austria

accepted on the recommendation of

Prof. em. Dr. med. Michael Detmar, examiner
Prof. Dr. Gerhard Schratt, co-examiner
Dr. Jay W. Shin, co-examiner

2022

Find your purpose.

Table of Contents

Summary

The human genome contains an estimated one million enhancers. With this significant number of enhancers, they make up an essential part of our genome. Enhancers are regulatory elements that interact with their target promoters through transcription factors that regulate the transcription process and thus have a significant impact on gene expression. Active enhancers, which are under strict spatial and temporal control, can be characterized by enhancer RNAs (eRNAs). eRNAs are 200-500 base pairs long RNA molecules and can be identified with 5'-based RNA sequencing. The activity of enhancers contributes significantly to cell type differentiation since a cell's function and identity are determined by its gene expression. Because the brain is one of the most complex organs with remarkable cellular diversity, elucidating the functions of enhancers is essential to understanding brain development. Because most cell types of a mature brain arise from the same precursor cells, it exhibits surprising cellular diversity and complexity.

The objective of this thesis was to investigate enhancers that contribute to the differentiation of brain cells. Since enhancers interact with promoters, this study aimed to elucidate the biological function of the enhancers based on their interaction with their target promoters. We used brain organoids, which are 3D cellular structures that can recapitulate regions of a brain. We analyzed brain organoids at 40, 80, and 120 days using 5'-based single-cell RNA sequencing. Our data showed that the cell type diversity increased with increasing cultivation time, and the cells also differentiated into more mature cells. Initially, radial glia progenitor cells accounted for 20% of the cells, while after 120 days of culture, this population decreased to less than 5%.

Furthermore, the organoids from day 40 to day 120 had gene expression profiles similar to those of human embryonic brains at post-conception week 12. Since the different stages of cellular differentiation processes are precisely regulated by specific sets of transcription factors, we next investigated the transcription factor signatures during brain development. The transcription factor motifs we detected supported our

observation that the cells in the brain organoids had undergone a similar differentiation process from progenitor cells to more mature cells to that of a human brain.

In the second part of the thesis, the purpose was to determine biologically and functionally relevant enhancer-promoter pairs. Enhancers interact with promoters to regulate biological processes by forming chromatin loops that bring them to close proximity. Using the 5'-based sequencing method combined with the SCAFE tool, we aimed to identify active enhancers with increased sensitivity compared to previous studies. SCAFE detects genuine enhancers utilizing logistic regression. We initially discovered 15,306 potential enhancer elements with SCAFE, of which about 50% were novel. Next, we identified approximately 50,000 robust enhancer-promoter pairs, from which 1627 showed a cell-type-specific expression. Finally, after integrating public HiC data, we identified 52 pairs with known genomic interaction, supporting the validity of these interactions.

Our data show that brain organoids represent an almost ideal model for studying brain development. With the increasing need for better characterization and identification of enhancers, our study contributes to the understanding of gene regulation. This work extends current resources with biologically relevant and functional enhancer-promoter pairs.

Zusammenfassung

Das menschliche Genom enthält schätzungsweise eine Million Enhancer. Mit dieser Anzahl machen Enhancer einen wesentlichen Bestandteil unseres Genoms aus. Enhancer fungieren mit Transkriptionsfaktoren als regulatorische Elemente, die den Transkriptionsprozess beeinflussen, und steuern damit die Genexpression. Die Aktivität von Enhancern steht unter strikter räumlicher und zeitlicher Kontrolle. Charakteristisch für aktive Enhancer sind Enhancer-RNAs (eRNAs). eRNAs sind 200-500 Basenpaare lange RNA-Moleküle und können mit einer 5‘-basierten RNA-Sequenzierung identifiziert werden. Die Aktivität von Enhancern trägt massgeblich zur Entstehung der Vielfalt von Zellarten bei. Da die Funktion und Identität einer Zelle durch ihre Genexpression festgelegt wird, haben Enhancer einen großen Einfluss auf die Eigenschaften einer Zelle. Das Gehirn ist eines der komplexesten Organe mit der größten Zellvielfalt. Obwohl die meisten Gehirnzellarten eines Erwachsenen aus ein und derselben Gehirnvorläuferzelle entsteht, weist das Gehirn eine erstaunliche zelluläre Vielfalt und Komplexität auf. Diese Studie zielt darauf ab, Enhancer, die zur Differenzierung von Gehirnzellen beitragen, zu untersuchen. Da Enhancer mit Promotoren wechselwirken, sollte im Rahmen dieser Studie die biologische Funktion der Enhancer anhand der Wechselwirkung mit Promotern herausgefunden werden. Hierfür wurden Gehirn-Organoide, 3D-Modelle, die verschiedene Regionen des Gehirns nachbilden können etabliert und die Genexpression dieser Gehirn-Organoide nach 40, 80 und 120 Tagen mithilfe von 5‘-basierter Einzelzell-RNA-Sequenzierung aufgeschlüsselt. Unsere Daten zeigen, dass mit zunehmender Kultivierungszeit sowohl eine Zunahme von Zellarten entstand, und dass die Zellen auch eine höher entwickelte Form von Gehirnzellen annahmen. Zunächst machten radiale Gila Vorläuferzellen 20% der Zellen aus, während nach 120 Tagen Kultivierung der Anteil auf 5% sank. Weiterhin wiesen die Organoide von 40 bis 120 Tagen, ein ähnliches Genexpressionsprofil wie menschliche Gehirne im embryonalen Entwicklungsstand der 12. Embryonalwoche. Da die Teilnahme von Transkriptionsfaktoren zur Differenzierung von Zellen beiträgt, untersuchten wir in unserem Modell als nächstes die Aktivität von Transkriptionsfaktoren anhand von ihren Motiven. Die Transkriptionsfaktor-Motivaktivitätsanalyse zeigte, dass die Zellen einen

Entwicklungsprozess von Vorläuferzellen zu mehr ausgereiften Zellen durchlaufen hatten.

Im Rahmen des zweiten Projektes wurden die funktionellen Enhancer mit Promotoren verknüpft. Enhancer interagieren mit Promotoren, um biologische Prozesse zu regulieren. Die Verwendung der 5‘-basierten Sequenzierungsmethode in Kombination mit dem aktuellen SCAFE-Algorithmus bietet eine empfindlichere Möglichkeit als bisherige Technologien, um funktionelle Enhancer zu identifizieren. SCAFE bestimmt „echte“ eRNAs mit Hilfe von logistischen Regressionen. Wir entdeckten zunächst 15‘306 potenzielle Enhancer-Elemente. Ca. 50% der Enhancer wurden in unserer Studie erstmals beschrieben. Als nächstes identifizierten wir ca. 50‘000 robuste Enhancer-Promoter-Paare, von denen 1627 einen Promoter, der Zelltyp-spezifisch exprimiert war, aufwiesen. Nach Einbindung von öffentlichen HiC Daten fanden wir 52 Paare mit genomischen Interaktionen, welche auf eine Enhancer-Promoter Wechselwirkung hindeuten.

Zusammengefasst zeigen unsere Daten, dass Gehirn-Organoide ein nahezu ideales Modell für die Untersuchung der Gehirnentwicklung darstellen. Mit dem steigenden Bedarf an besserer Charakterisierung und Identifizierung von Enhancern trägt unsere Studie zum Verständnis der Genregulation bei. Diese Arbeit erweitert aktuelle Ressourcen mit biologisch relevanten und funktionalen Enhancer-Promoter-Paaren.

List of Abbreviations

AP	Astrocytes
ARX	Aristaless-related homeobox
ATAC	Assay for transposase-accessible chromatin
BMP	Bone morphogenetic protein
BRD4	Bromodomain Containing 4
CAGE	Cap Analysis of Gene Expression
CBP	CREB-binding protein
ChIP seq	Chromatin immunoprecipitation sequencing
CO	Cerebral organoids
CO2	Carbon dioxide
CP	Choroid Plexus
CREID	Cis regulatory element ID
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CYP19	Cytochrome P450 aromatase
DMEM-F12	Dulbecco's Modified Eagle Medium F-12
eQTL	Expression Quantitative trait loci
ERK	Extracellular signal regulated kinase
FANTOM5	Functional annotation of the mammalian genome
IN	Interneurons
IP	intermediate progenitor Cells
LASSO	Least Absolute Shrinkage and selection operator
MAPK	Mitogen-activated protein kinase
MARA	Motif activity response analysis
MEM-NEAA	Minimum essential medium non-essential amino acid
MGE	Medial Ganglionic Eminence
MH	Midbrain Hindbrain
NGS	Next Generation Sequencing
NPC	Neuronal Progenitor Cells

OP	Oligodendrocytes
ORF	Open reading frame
P-TEFb	Positive transcription elongation factor b
PBS	Phosphate-buffered saline
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
RGC	Radial Glial Cells
RNA	Ribonucleic acid
ROCK	Rho-associated protein kinase inhibitor
SCAFE	Single Cell Analysis of Five-prime Ends
SHH	Sonic Hedgehog
SMAD4	Mothers against decapentaplegic homolog 4
SMAD9	Mothers against decapentaplegic homolog 9
SNCA	Synuclein Alpha
SNP	Single Nucleotide Polymorphism
TF	Transcription Factor
TGFb	Transforming Growth Factor beta
TSO	Template Switching Oligo
UMI	Unique Molecule Identifier
UTR	Untranslated Region
ZNF143	Zinc finger 143

1 Introduction

1.1 History of enhancers

1.1.1 The first discovery of an enhancer

Starting from the first discovery of enhancers as transcriptional regulators in 1983 (Mercola *et al.* 1983), the definition of an enhancer has varied over time. The first reported enhancer sequence was described as a cis-regulatory and non-coding sequence that could enhance the expression of a gene. This finding soon became the object of interest in many biological fields. However, unlike genes, enhancers could not solely be identified by transcriptomic sequencing, and they lacked a defined grammar that supported their assignment as being functional.

1.1.2 Current understanding of enhancers and enhancer RNAs

The human genome carries around one million enhancers (Dunham *et al.*, 2012). An enhancer is a regulatory DNA segment of about 200 base pairs that can strongly stimulate or inhibit the transcription of a promoter (Figure 1). Enhancers govern the transcription from upstream or downstream of the site of transcription initiation. Most functionally validated enhancers act in *cis* (Walters *et al.* 1996) and are composed of multiple binding sites for transcriptional activator proteins (Spitz and Furlong 2012; Visel *et al.* 2009). However, some studies support the hypothesis that enhancers can function in *trans* (Müller and Schaffner 1990; Mattioli *et al.* 2020; Bateman, Johnson, and Locke 2012). They are critical gene-regulatory elements that govern cell-type-specific spatiotemporal gene expression by physically contacting genes through long-range chromosomal interactions. Enhancers can be distal to their interacting promoter and regulate more than one promoter. They generally function independently of orientation and at various distances from their target promoters. Enhancers can be very far away from their regulating promoter and often have several modes of action. Despite our long history of enhancer research, the current definition of enhancers is not definite and is the subject of ongoing research.

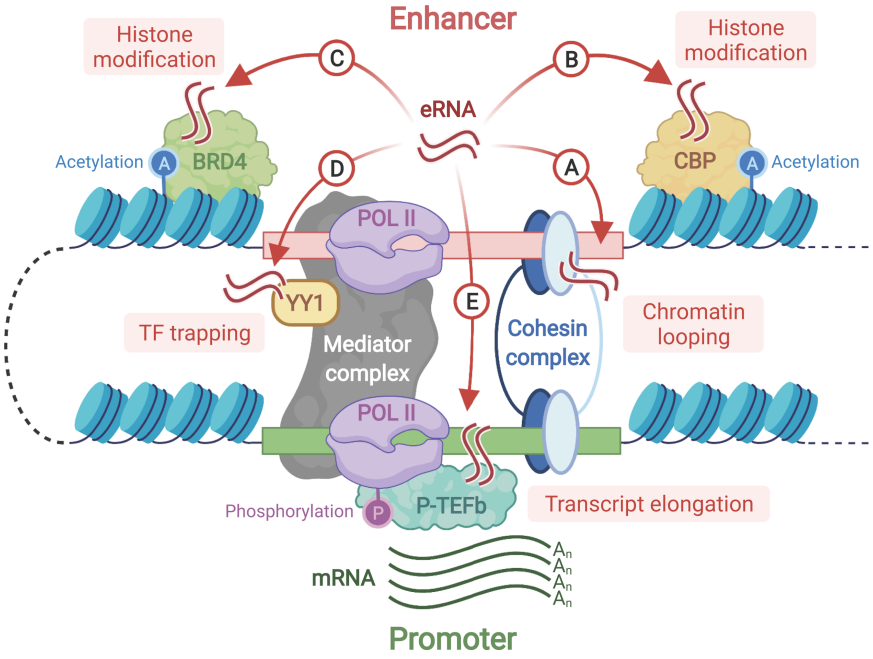

Figure 1. Functional role of enhancer RNAs (eRNAs) during the transcriptional process. eRNAs emerge from the translation of the enhancer. (A) Interaction of eRNA with cohesion leads to chromatin looping. (B and C) Interaction of eRNA with CBP and BRD4 facilitates histone acetylations. (D) Interaction of eRNA with the transcription factor YY1 leads to transcription factor trapping. (E) Interaction of the eRNA with P-TEFb leads to phosphorylation of POL II. Illustration from Syed and Hon (2021).

In two independent discoveries in 2010, De Santa *et al.* and Kim *et al.* could show that enhancer regions are actively transcribed into enhancer RNAs (De Santa *et al.* 2010; Kim *et al.* 2010). The surprising product of this transcription, termed "enhancer RNA" or "eRNA," has subsequently been a subject of controversy, and not every enhancer was shown to produce eRNAs. De Santa *et al.* suggested that the transcription of enhancers could be a byproduct of a leaky transcription process. However, with the rise of perturbation studies, eRNAs have proven to lead the functional role of the enhancer-promoter interaction in the gene regulation process (Li *et al.* 2013; Aguilo *et al.* 2016; Hsieh *et al.* 2014). Furthermore, these studies have shown that eRNAs are involved in diverse aspects of cell functions and are therefore unlikely to be a random byproduct of transcription.

Pnueli *et al.* also suggested that eRNAs are necessary for stabilizing the enhancer-promoter loop (Pnueli *et al.* 2015) and therefore have a supporting function in the enhancer activity. Additional studies have shown that overexpression of eRNAs leads to

an increase in their promoter targets (Shii *et al.* 2017; Jiao *et al.* 2018; Alvarez-Dominguez *et al.* 2017). These studies have further proven that eRNAs play a functional role in the enhancer to promoter interaction. However, it remains unclear how generalizable these characteristics are on a global view of enhancers. Thus, more important than if eRNAs have a biological role is the question of which eRNAs are functional and how we can infer their activity. The field of enhancer-promoter relationship needs further studies to obtain a more comprehensive profile of how and when enhancers interact with promoters.

1.1.3 Accessible versus transcribed enhancers

Enhancers can be identified using different technologies (Table 1), each technology capturing a distinct feature of our current enhancer definition.

Enhancers can be identified by open-chromatin regions (Creyghton *et al.* 2010; Thurman *et al.* 2012; Rivera and Ren 2013). Active enhancers are identified by open-chromatin regions that carry an epigenetic mark, such as methylation or acetylation. Epigenetic marks contribute to chromatin formation changes and the function of enhancers.

Another approach utilized enhancer RNAs to identify active enhancers (Djebali *et al.* 2012; Andersson *et al.* 2014; Arner *et al.* 2015; Li, Notani, and Rosenfeld 2016; Wang *et al*. 2011). Enhancer RNA is produced when the enhancer is looping to its target gene. The enhancer region gets into proximity to the polymerase and leads to the transcription of the enhancer region, thus characterizing an active enhancer. Several studies show that transcribed enhancer regions are more likely to be involved in looping and DNA-DNA interactions than enhancer regions that are not transcribed.

Table 1. Technologies to identify the most common features of enhancers. ChIP-seq can detect a wide range of enhancer features, such as histone modification H3K4me1 and H3K27Ac, as well as binding of transcription actors. Adapted from Lewis, Li, and Franco (2019).

Feature	Commonly used methods for detection
Chromatin accessibility	DNase-seq, ATAC-seq, and FAIRE-seq
Histone modification H3K4me1	ChIP-qPCR, ChIP-chip, ChIP-seq
Histone modification H3K27Ac	ChIP-qPCR, ChIP-chip, ChIP-seq
Enhancer-promoter looping	3C, 4C, HiC, and Hi-ChIP
Binding of transcription factors	ChIP-qPCR, ChIP-chip, and ChIP-seq
Production of eRNAs	GRO-seq, CAGE, PRO-seq

Both methods are currently used to identify active enhancers. A study in 2013 found that eRNAs are more accurate than chromatin accessibility in identifying enhancers (Zhu *et al.* 2013). This study created a regression model to understand the relationship between enhancer RNAs and chromatin modifications. Interestingly, the authors concluded that enhancer RNAs are more indicative of enhancer activity than chromatin modifications. Similar conclusions were made from a study looking at an enhancer regulating the MAPK/ERK pathway (Tyssowski *et al.* 2018). The authors found that the eRNA was repressed as expected upon inhibition of the MAPK/ERK pathway, but acetylation was still accumulating at the enhancer region. These studies suggest that eRNA transcription is a more sensitive marker for enhancer activation than histone modifications.

1.1.4 Cell type specificity of enhancer RNAs

While the human body consists of hundreds of cell types, the genome of each cell is the same. However, every cell type is characterized by a distinct gene expression profile. Coordinating this variety of gene expression for each cell type needs precise control in a time- and environment-specific manner. Previous studies have shown that enhancers in combination with transcription factors play a vital role in this regulation. While

enhancers are critical to all cell types, enhancer activity varies dramatically across tissues (Nord *et al.* 2013). The FANTOM5 project identified immune cells, neural tissues and hepatocytes among those with the highest abundance of cell-specific enhancers, a higher ratio of enhancers to genes, and high enhancer transcription (Andersson *et al.* 2014). In this study, eRNAs were used to identify active enhancers. Indeed, studies in neurons, macrophages, and other immune cells are prominent among those contributing to our understanding of enhancers and active enhancers. In contrast, smooth muscle cells, fibroblasts, and epithelial cells utilize enhancers with less cell-specificity and lower enhancer to gene ratio (Andersson *et al.* 2014). This diversity of enhancer utilization likely reflects each cell type's unique roles and needs to respond dynamically to its environment.

Previous studies identified millions of enhancers in the human genome, but only a subset of enhancers are activated at a time. Enhancers can interact with transcription factors to regulate gene expression (Arnold, Wells, and Li 2020; Michida *et al.* 2020; Palstra 2012; Spitz and Furlong 2012). Transcription factors are proteins that bind to specific DNA sequences and motifs to activate or repress gene transcription. Transcription factors can operate alone or with other proteins in a complex. Studies have shown that enhancers distal to their regulating promoter can interact with their promoter by recruiting transcription factors that initiate the looping of the DNA, bringing the enhancer close to the interacting promoter (Rhee *et al.* 2016). As enhancer activities undergo progressive changes during cell type differentiation, the interaction with transcription factors is dynamic. Transcription factors function in complex combinatorial and regulatory patterns to secure precise gene regulation.

1.1.5 Identification of enhancers and their regulating gene

From previous studies, we know that enhancers regulate the transcriptional process by interaction with the promoter of that gene (Figure 2). However, there is no gold standard yet to identify this enhancer-promoter interaction. There are many obstacles to overcome in identifying and characterizing enhancer-promoter pairs. First of all, enhancers are distributed across 99% of the non-coding region of the genome, and

enhancer sequence codes are poorly understood if we compare the well-annotated sequence code of protein-coding genes. Thus, enhancers cannot be identified by sequence only. Instead, we rely on epigenetic marks or eRNAs to elucidate the function of an enhancer.

Because the location of enhancers to their target gene is variable, the distance from enhancer to promoter cannot be used as a characteristic to define an enhancer. In addition, there is no universal rule that enhancers must be upstream, downstream, or within a gene body. Having multiple variables makes identifying the target gene of enhancers challenging.

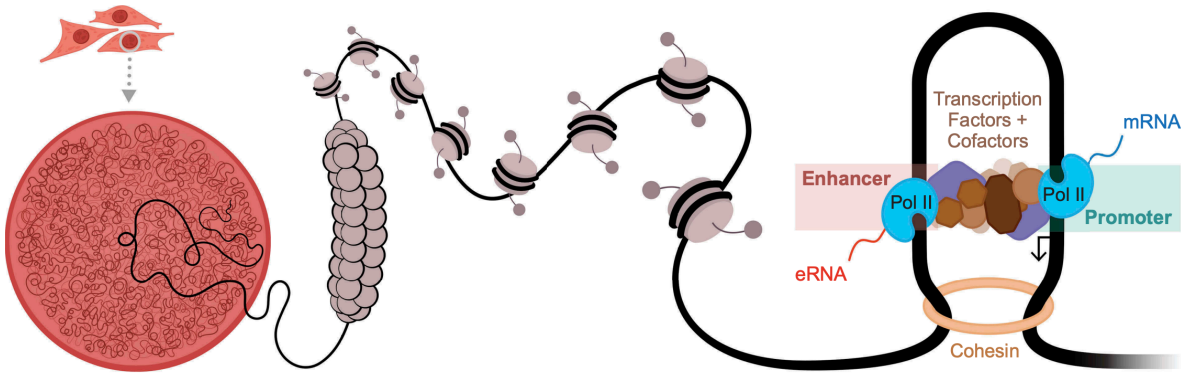

Figure 2. Current understanding of the enhancer-promoter interaction. The target gene can be located far from the enhancer. Changes in the chromatin loop and higher-order chromatin structure can facilitate the enhancer to come to proximity to the target promoter. The changes in chromatin structure are initiated by transcription factors, which recruit chromatin remodeling enzymes, histone enzymes, and the RNA polymerase II complex. Transcription of enhancers results in enhancer RNAs (eRNAs).

However, continuous research has shown that different methods can overcome these obstacles. There are several ways of inferring enhancers to target gene interactions (Hariprakash and Ferrari 2019; Wang, Hu, and Li 2020; Jing, Zhang, and Zhang 2020; Yang Yang *et al.* 2017; Zhao, Li, and Hu 2016; S. Liu *et al.* 2020).

One method of inferring enhancer-promoter pairs uses HiC data, representing a whole-genome pull-down of all genomic regions interacting with each other. Therefore, HiC is widely accepted for identifying available enhancer-promoter pairs since this method proves physical contact. Common approaches to identify enhancer-promoter pairs

include taking the nearest protein-coding gene to the enhancer or setting a defined range. However, with the emergence of eRNAs, correlating the eRNA expression pattern with promoters has become an alternative way of predicting enhancer-promoter pairs. Even though the low eRNA expression levels complicate this method, recent studies showed that eRNA transcription and induction of mRNA transcription at neighboring genes are correlated (Arner *et al.* 2015; Andersson *et al.* 2014; Kouno *et al.* 2019; Li *et al.* 2013; Kim *et al.* 2010).

Other approaches include transcription factor binding site motif analysis, as transcription factors are known to be involved in the enhancer-promoter interaction. Transcription factors are specialized proteins that can facilitate the binding of the enhancer or recruit other proteins to aid the enhancer to promoter interaction. Transcription factors recognize specific DNA sequences called motifs. Therefore, transcription factors can be used to identify enhancers and promoters that carry the same motif.

Recent publications show that the successful identification of enhancer-promoter pairs remains a challenging and ongoing process but have demonstrated that bioinformatics approaches are powerful in predicting enhancer-promoter pairs on a genome-wide scale (Gao and Qian 2019; Hait *et al.* 2018; Hwang *et al.* 2015). Using *in silico* methods enables inferring enhancer-promoter pairs on a large scale and reduces the experimental workload while saving resources. Therefore, having a reliable way to infer enhancer-promoter pairs is the first step toward finding functional pairs with biological relevance.

1.1.6 Further characteristics of enhancers

As we are uncovering the true nature of enhancers, resources and individual studies add to our growing understanding of enhancers.

Although most described enhancers have one promoter they interact with, there have been studies that prove that enhancers can also act in groups of super-enhancers (Hnisz *et al.* 2013; Su *et al.* 2015; Luo *et al.* 2016). One enhancer can interact with multiple

genes or vice versa. Super-enhancer regions are usually 600-1400 base pairs long (Whyte *et al.*, 2013). Complex interactions are challenging to study because the contribution of every single enhancer is unclear. Several studies have investigated the effect of individual enhancers in a super-enhancer hub (Xie *et al.* 2017; Kai *et al.* 2021). However, there is no consensus in predicting the enhancer with the most functional role in a super-enhancer hub. Predicting the active enhancer in a super-enhancer hub is beneficial when we want to understand the function of the super-enhancer. Bioinformatic approaches can reduce the experimental time to test every single enhancer from a super-enhancer hub by predicting the importance of the enhancer based on previous studies or characteristics of the enhancer.

1.2 Human brain organoids

1.2.1 Early model systems to study brain development

The brain is a complex tissue with many cell types and each performing a distinct function in space and time (Stiles and Jernigan 2010). Unfortunately, there has been a shortage of possibilities in human brain developmental model systems due to technical restrictions (Farahany *et al.* 2018). Limitations include accessibility to samples and the lack of cellular heterogeneity in model systems.

A significant amount of our current knowledge is based on mice and other non-human species (Y. E. Li *et al.* 2021; Gorkin *et al.* 2020). Animal models serve as an essential source for studying the function of the brain. Mice studies are highly valued due to the availability of the model system and flexibility. However, findings in mice might be limited to mice due to species differences and mice-only features (Hodge *et al.* 2019). Given the limitation of the mouse system, researchers have been working on alternative model systems to study the brain.

With the rapid progress in stem cell technologies (Takahashi *et al.* 2007; Takahashi and Yamanaka 2006), stem cells have become a widely used model system for generating neuronal cells in a dish (Shimojo *et al.* 2015; Pistollato *et al.* 2017). Studying neuronal cells became available on a large scale with the development of differentiation protocols into key neuronal cell types. However, commonly used 2D cultures are designed for culturing a few cell types at a time. This results in morphological and cell-to-cell interaction constraints because neurons function in neurological networks. With the discovery of new cell types, more and more sophisticated model systems are required to study the biological functions of these cell types (Masland 2004; Zeng and Sanes 2017; Sanes and Masland 2015). The number of currently known radial glia cells increased from appr. 12 to appr. 30 as more advanced technologies became available. 2D model systems are limited in their capabilities to mimic the microenvironment of the brain, and the lack of a multi-cell type model system is a significant bottleneck in advancing human brain developmental studies. However, studying the interplay of cell types at early developmental stages on a large scale became possible with the rise of brain organoids.

In 2014, Lancaster and Knoblich introduced the first human brain organoid platform that overcomes parts of previously mentioned challenges.

1.2.2 Current human brain organoid protocols

In brief, brain organoid protocols depend on the self-assembling power of stem cells. These cells can be embryonic pluripotent stem cells or induced pluripotent stem cells (Figure 3). Putting these cells in close proximity allows them to self-assemble into a tight aggregation of cells, often referred to as embryonic bodies or spheres. There are various culturing methods, but they share the common goal of mimicking certain brain regions or specific functions of the brain. Specifically, the priming medium in the early stages of brain organoid development initiates the lineage specification of the brain organoids. Recent advances have led to protocols that allow brain organoids' efficient and rapid formation.

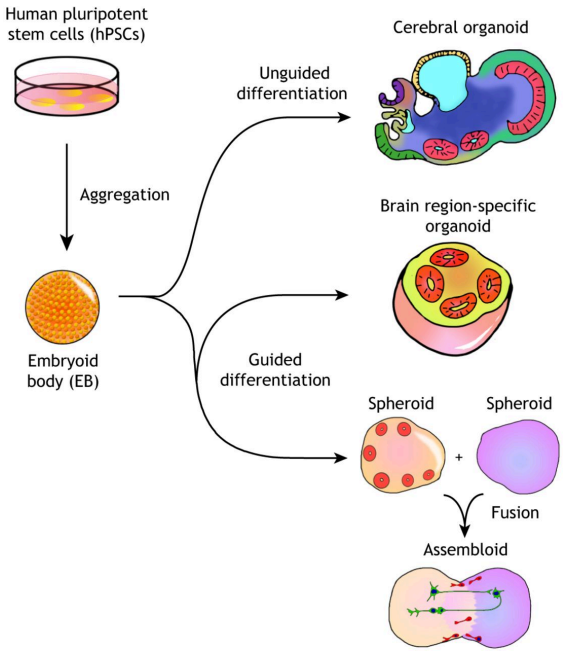

Figure 3. Guided and unguided protocol to generate brain organoids. Unguided approaches (top) depend on the intrinsic self-assembly of stem cells and spontaneous differentiation into neuronal cells. The resulting cerebral organoid carries cell types from various regions of the brain. The guided approach utilizes small molecules and growth factors to generate brain region-specific spheroids. By patterning the stem cell from an early time point, the cells are primed towards a specific progenitor fate leading to a particular cell tissue. Two spheroids can be fused together to mimic interactions between different brain regions, resulting in an assembloid. Adapted from Sozen *et al.* (2018).

Brain organoid protocols are broadly classified into either a guided or non-guided approach. The first brain organoid protocol from Lancaster et al. used the intrinsic property of human-induced pluripotent stem cells to differentiate the cells toward the neuroectodermal lineage under serum-free conditions (Lancaster *et al.* 2013). This protocol did not display a specific regional identity but broadly contained clusters of cells representing the regions of diverse brain regions. Non-guided protocols result in brain organoids giving rise to various rostral forebrain cell types. Large-scale single-cell RNA sequencing revealed cellular diversity, maturity, and heterogeneous neural networks in the Lancaster protocol (Quadrato *et al.* 2017). In addition, optimization of maintaining the brain organoids has improved neuronal survival and outgrowth in these brain organoids (Giandomenico *et al.* 2019).

After the cerebellar brain organoid protocol, several brain region-specific protocols and multi-brain region co-culturing protocols provide promising results for further translational applications (Qian *et al.* 2016; Ozone *et al.* 2016; Sakaguchi *et al.* 2015; Muguruma *et al.* 2015; Monzel *et al.* 2017). For improving the survival of neuronal progenitors, lipids, serum, heparin, and Matrigel were included in the media, supporting the survival of neuronal progenitors (Kadoshima *et al.* 2013). The Pasca group used growth factors during neuronal differentiation to improve cortical lamination and functional synaptogenesis. Growth factors support lineage commitment and organoid morphogenesis (Birey *et al*. 2017; Pasca *et al*. 2015). This method was also used to generate cortical organoids (Madhavan *et al.* 2018) and to establish oscillatory networks in these organoids under long-term culture with additional nutrients (Trujillo *et al*. 2019). Xiang *et al.* (2017) used a guided approach to generate the dorsal forebrain using a cocktail of WNT, BMP, and TGFb inhibitors without Matrigel, while inhibiting the sonic hedgehog (SHH) pathway (Fiddes *et al.* 2018) generated dorsal cortical organoids.

Taken together, currently, two main brain organoid protocols are being used, leading to brain organoids with a region-specific brain organoid or general cerebral structure. Despite the differences in media formulation, both protocols use the intrinsic self-assembling attribute of stem cells (Figure 4).

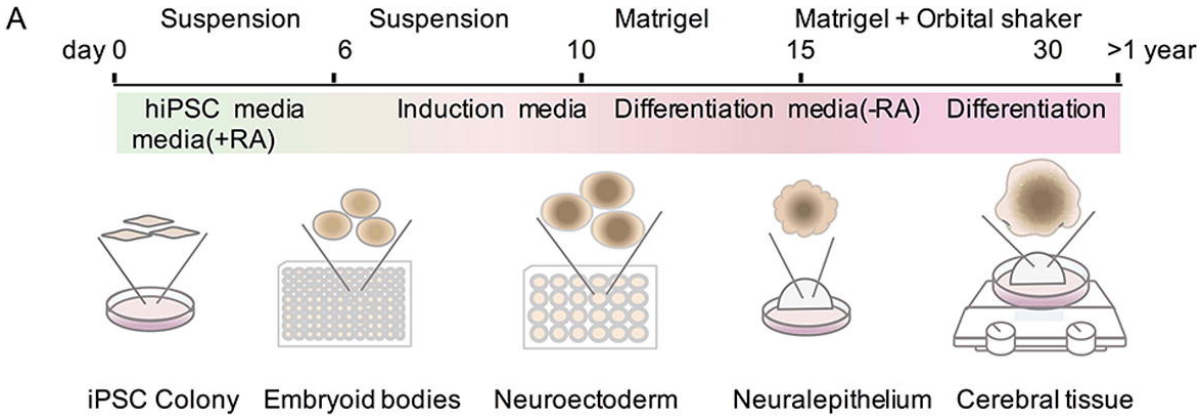

Figure 4. General current protocol of brain organoids. After the 2D culture of the stem cell, cells are transferred to a round ultra-low-attachment plate where the cells self-assemble and form embryonic bodies. After priming cells towards a general neuronal progenitor identity, the embryonic body is encapsulated in Matrigel and cultured on an orbital shaker to improve oxygen flow. Matrigel provides support for the tissue in the initial stages. The brain organoid eventually will outgrow the Matrigel. The figure was adapted from Liu *et al.* (2019).

1.2.3 Brain organoids as a model system to study brain development and brain diseases

Brain organoid model systems are based on previous work on the culture of spheres and embryonic bodies. Next to the extraordinary accomplishment of the generation of organoids, the therapeutic purpose of organoids has tremendous potential. Brain organoids hold the unique possibility of understanding how viruses affect the development of the human brain (Zhang *et al*. 2018) or as an oncology model system (da Silva *et al.* 2018).

Especially in combination with the technology of induced pluripotent stem cells, brain organoids can be used to study genetic disorders from patient-derived tissue (Allende *et al.* 2018). Allende *et al.* generated brain organoids from patients suffering from Sandhoff disease. Sandhoff disease is characterized by a progressive loss of nerve cells in the brain and spinal cord. It belongs to the family of rare inherited disorders. Because it affects early children, access to samples is challenging. Therefore, brain organoids are an excellent model system to study diseases with early-onset or biological processes in early developmental stages.

1.2.4 Importance of enhancers and eRNAs in the human brain

Enhancers play a crucial role in regulating cell type-specific gene expression patterns, and several enhancer features are associated with the brain's developmental process (Nord *et al.* 2013). Furthermore, brain development is a dynamic process in which gene expression needs to adapt fast to the changing environment of the cells (Stiles and Jernigan 2010b; Tierney and Nelson 2009). Therefore, disturbing the enhancer activity during the cell differentiation process affects the lineage specification leading to impaired brain development that can manifest into severe brain diseases (Yousefi *et al.* 2021).

Nord *et al.* (2013) demonstrated that the enhancer activity is a stage-specific process during mouse forebrain development. Active enhancers were identified using H3K27 acetylation by performing Chromatin Immunoprecipitation sequencing (ChIP-Seq). 85% of their identified enhancers showed temporally specific enhancer activity depending on the time point they sampled the mouse forebrain. Furthermore, they confirmed that the specific spatiotemporal activity of the enhancers is reflected in the tissue-specific gene expression of the interacting gene. Nord *et al.* identified enrichment of neurodevelopmental transcription factors in early active enhancers, whereas transcription factors associated with neuron-specific functions were enriched in delayed active enhancers. Dickel *et al.* (2018) showed that impairment of an enhancer regulating the ARX gene results in a decrease of neurons. ARX was first described as a gene involved in non-syndromic X-linked mental retardation disease (Bienvenu *et al.* 2002) and X-linked West syndrome (Strømme *et al.* 2002). Since then, ARX has been implicated in multiple phenotypes associated with brain-related malformations (Kitamura *et al.* 2009; Kato *et al.* 2004; Guerrini *et al.* 2007). These studies highlight the importance of highly controlled enhancer activity during development and the requisite of understanding the regulatory role of enhancers during brain development to prevent malformations.

Over the past decade, genomic sequencing efforts confirmed that single nucleotide polymorphisms (SNPs) in the brain could lead to neuropsychiatric diseases (Dong *et al.* 2018; Hnisz *et al.* 2013; Li *et al.* 2019; Kim *et al.* 2012). These predictions afforded a better understanding of the pervasiveness of mutations in distant cis-regulatory

elements underlying brain diseases. In an early demonstration of this, Soldner *et al.* (2016) found one single nucleotide polymorphism (SNP) in a distal enhancer region of SNCA, a marker gene for Parkinson's disease. This study highlights the impact of one SNP on the function of the enhancer activity. Although it is unknown how many of the enhancers which carry one SNP are functionally affected, the functional characterization of putative disease-causing regulatory mutations is essential to understand the underlying cause of the disease. Experimental strategies involve reporter assays and knock-out studies. Genetic variation in distant enhancers has been linked to several human disorders. Given the broad abundance of enhancers in the genome, further studies focusing on the functional relevance of SNPs in enhancer regions are crucial.

1.3 5’ single-cell RNA sequencing

1.3.1 First single-cell RNA seq platforms

RNA sequencing is a powerful technology to capture the transcriptome of the cell. RNA sequencing is widely used due to its reliability, reproducibility, and easiness of use. However, bulk RNA sequencing leads to averaged expression of highly heterogeneous cell types and states. Our current understanding of one cell type is restricted by technology, and the rise of advanced technologies results in a progressive definition of cell types. For example, studies have shown that cancerous cells are a mixture of cells in different states, although commonly referred to as one cell population (Meacham and Morrison 2013). Another study revealed the heterogeneous effect of drug treatment on breast cancer cells (Hoffman *et al*. 2020). Recognizing cell to cell differences is crucial to understanding a drug's effectiveness and preventing side effects. Therefore, developing a technique to study the transcriptome of a single cell was essential.

With the advancement of droplet technologies and next-generation sequencing (NGS), measuring the gene expression profiles of thousands of single cells became possible, which significantly increased the sensitivity of our transcriptomic understanding in a single cell. Since the first demonstration of single-cell RNA seq in the human oocytes (Tang *et al.* 2009), the field of single-cell RNA sequencing applications has been increasing. Although the technology was once considered exclusive and expensive, it soon became the technology of choice for many areas of biology (Svensson, da Veiga Beltrame, and Pachter 2020). Capturing cellular differences on a single cell level pushed the boundaries and enhanced our understanding of cell populations.

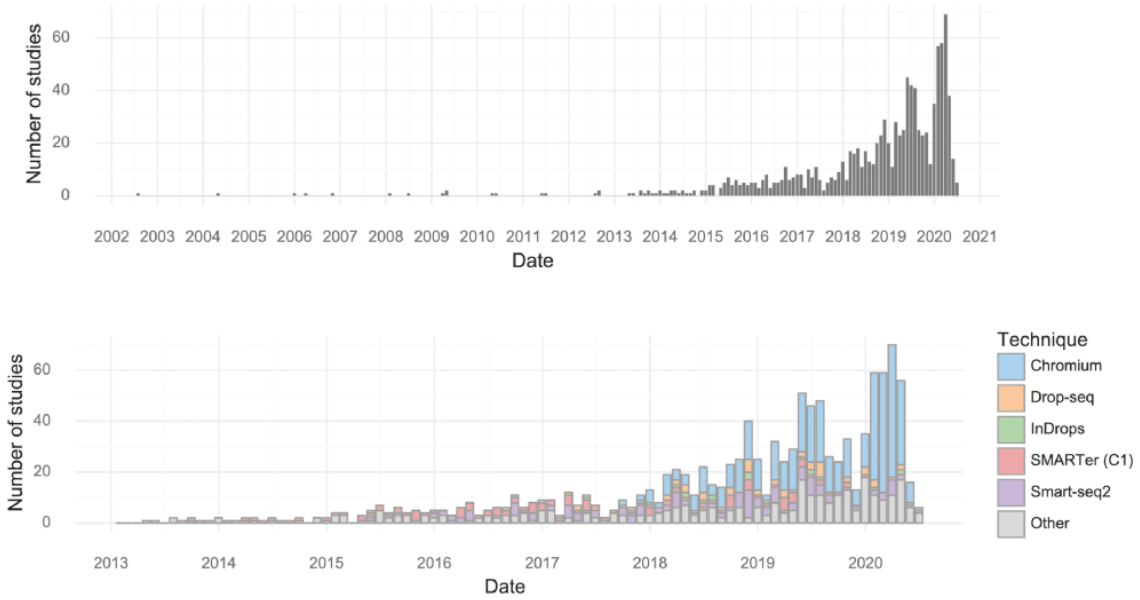

Figure 5. Trends in single-cell RNA sequencing publications and platforms. The upper panel shows the number of single-cell transcriptomics studies published per month. The lower panel shows the number of single-cell RNA seq studies published per month color-coded by the used method. Adapted from Svensson, da Veiga Beltrame, and Pachter (2020).

1.3.2 3’ vs. 5’ RNA sequencing

In recent years, single-cell studies of the brain have been rapidly growing (Figure 5). These studies were conducted using ex vivo samples of brain organoids. However, most studies are conducted based on 3’ based sequencing methods, leaving behind the advantage of using 5’ end sequencing to more sensitively detect enhancer RNAs. Most studies are conducted using 3’ sequencing methods, as this was the more sensitive method regarding gene detection. This drawback has changed with advanced technologies, and now, 3’ and 5’ detection sensitivities are similar. However, although both technologies are identical, when using the 3’ sequencing, we capture the read from the 3’ end of the transcript. Therefore, we do not capture the genuine transcription start site of the transcript. This is especially important as our annotation of enhancers is still incomplete.

Consequently, we could miss out on these elements when we have to infer the transcription start site from transcripts based on current annotations. There have been efforts to provide a more comprehensive annotation of enhancers using CAGE

sequencing (Shiraki *et al.* 2003; Andersson *et al.* 2014; Arner *et al.* 2015). CAGE captures the 5’ end and has thus a significant advantage over the 3’ back in detecting enhancer RNAs. Capturing the precise transcription start site makes the detection of novel eRNAs more robust than using the 3’ end of the transcript. This means that if a promoter and enhancer have the same 3’ end, we would not be able to distinguish if this transcript resulted from a promoter or enhancer. Many enhancers are also considered alternative promoters (Kowalczyk *et al.* 2012) and therefore, this is one of the primary bottlenecks when we want to identify enhancers using enhancer RNAs.

We utilized a 5’ based single-cell RNA sequencing method to identify cell-type-specific enhancers combined with a template-switching oligonucleotide (TSO) technology (Zheng *et al.* 2017). TSO plays a vital role in completing the cDNA amplification step (Figure 6). It binds to the C overhang the reverse transcriptase created and adds a common 5’ sequence to the final cDNA.

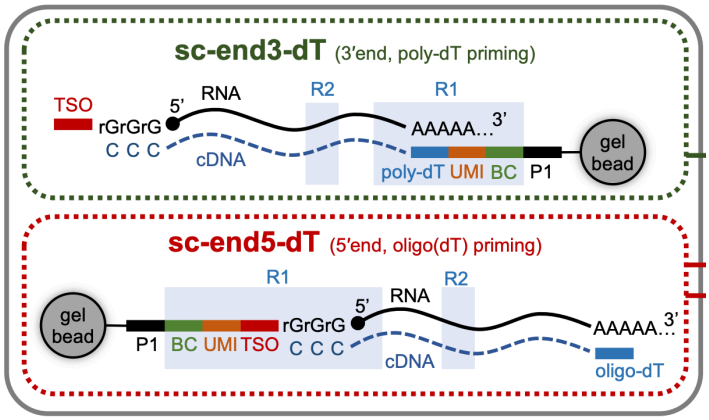

Figure 6. Difference between 3’ and 5’ based RNA sequencing methods. 5’ and 3’ based single-cell RNA sequencing methods capture different ends of the transcript. Although both methods use the Template Switching Oligonucleotide technology, in the 3’ technology, Read 1 captures the poly-A tail of the transcript. 5’ based technologies capture the Transcription Start Site of the transcript. BC: cell barcode, UMI: unique molecular identifier, cDNA: complementary DNA, TSO: template-switching oligonucleotide, R1: read 1, R2: Read 2. This figure was adapted from Moody *et al.* (2021).

1.3.3 Technologies to capture enhancers

Although there are multiple described technologies to capture distinct features of enhancers (Table 1), technologies for capturing cell-type-specific enhancers on a single cell level are still unfolding. Since the first scRNA-seq method was published in 2009,

many other single-cell RNA-seq approaches have been developed. These scRNA-seq are similar at the core. However, their technologies differ in one of the following aspects: cell isolation, cell lysis, reverse transcription, amplification, transcript coverage, strand specificity, or unique molecular identifiers (UMI), which are molecular tags. UMIs are used to detect and quantify individual transcripts.

Table 2. Current technologies to capture enhancers. Each technique highlights another studied aspect, capturing a distinct feature of enhancers. While each technology represents one aspect of enhancers, a technology combining multiple characteristics would maximize the use of each technology. Adapted from Weber *et al.* (2016).

Techniques	Studied Aspect	Advantages	Disadvantages
TF Binding Motif Scan	TF binding motifs	Identifies TF binding sites	High false discovery rate; prior knowledge on TF binding motif required
DNase-seq	Open chromatin	TF binding motifs can be detected	DNase I can introduce cleavage bias, affecting TF footprint detection
ChIP-seq	Histone modifications, TFs, chromatin-associated proteins	A wide range of targets can be studied	Relies on the availability of high-quality antibodies or tagged proteins
RNA-seq	Transcript levels	eRNA levels implicate enhancer activity, detects directionality of transcription	eRNA expression is low, high sequencing depth required
CAGE	Transcript levels	eRNA levels implicate enhancer activity, detects directionality of transcription	Only detects capped eRNAs
GRO-seq	Nascent transcript levels	eRNA transcription implicates enhancer activity	Challenging technique
STARR-seq	Enhancer mapping (and activity)	High-throughput identification and validation of enhancers in parallel	Minimal promoter used influences the set of identified enhancers
BS-seq	DNA methylation	Single bp resolution	High sequence depth needed; incomplete BS conversion affects data interpretation
Enhancer trapping	Enhancer activity	Visualizes tissue-specific pattern mediated by endogenous *cis*-regulatory sequences	Difficult to locate trapped enhancers
3C technology	Chromatin interactions	Identifies promoter-enhancer interactions	Challenging technique; trade-off between number of observed interactions and resolution
Reporter assay	Transcriptional activity	Confirms activity and tissue specificity of enhancer candidates	Potential expression bias arising from test conditions and the minimal promoter used

Compared to 3’ or 5’ based protocols, full-length scRNA-seq methods have unique advantages in isoform and allelic expression detection. This is due to their superiority in transcript coverage. For example, one study showed that full-length scRNA-sequencing approaches are more sensitive than 3’ sequencing methods (Ziegenhain *et al.* 2017) in detecting lowly expressed transcripts. Notably, droplet-based technologies (Macosko *et al.* 2015), InDrop (Klein *et al.* 2015), and Chromium (Zheng *et al.* 2017) can generally provide a more significant throughput of cells and a lower sequencing cost per cell compared to whole transcript scRNA-seq. Thus, droplet-based protocols are suitable for processing large amounts of cells to identify the cell subpopulations of complex tissues.

1.3.4 Enhancers as alternative promoters

As previously discussed, enhancers can be at various genome locations, including close to a transcription start site. The proximity to a transcription start site makes it challenging to distinguish alternative promoters from enhancer regions (Kowalczyk *et al.* 2012; Dao and Spicuglia 2018). Human protein-coding genes are estimated around 20,000 (Salzberg 2018; Piovesan *et al.* 2019). In contrast, simpler organisms such as *Drosophila melanogaster* are fewer with around 13,000 genes (Misra *et al.* 2002). Given that mammalian genomes have roughly 40% more genes than *Drosophila melanogaster*, it is generally believed that the phenotypic complexity of higher organisms is achieved not only by higher gene numbers but also by multiple proteins encoded by a single gene and by the number of protein-protein interactions. Alternative splicing is the most well-described mechanism that produces multiple protein isoforms from a single gene locus. It has been estimated that 35-50% of all human genes give rise to alternatively spliced mRNAs (Jiang and Chen 2020). An increase in the number of recent studies reporting the existence of alternative promoters for genes demonstrated that this phenomenon is another important source for generating protein and regulatory diversity.

No variation in the resulting proteins has been reported for many genes for which multiple promoters have been documented. However, even in the absence of protein isoforms, the mRNA variants differ in their transcriptional patterns and translational

efficiencies. These alternative promoters have different tissue specify, developmental activity, expression levels, or variant 5' untranslated regions (UTRs), which might differ in their secondary structure and affect translation (Figure 7). A well-documented example of a human gene with tissue-specific expression governed by the usage of alternative promoters in the CYP19 gene *(Singer et al. 2008; Demircioğlu et al.* 2019). This gene encodes the aromatase P450 protein, which converts C19 steroids to C18 estrogens, and gonadal and brain-specific promoters in humans direct the tissue-restricted expression of CYP19. Alternative promoters display tissue-specific regulation and impact isoform diversity. These characteristics of alternative promoters are similar to enhancer RNAs and can cause misinterpretation of an enhancer region as an alternative promoter or vice versa. Therefore, we need technologies to precisely detect the transcription start site to prevent miss annotation of an alternative promoter as an enhancer. However, there have also been cases where an intragenic enhancer can function as an alternative promoter (Kowalczyk *et al.* 2012). As the number of such cases is low, it remains unclear if this is a rare case or if more enhancers serve as an alternative promoter.

1.3.5 Analysis tools for single-cell RNA sequencing

With the rise of single-cell RNA sequencing data, there was a need for more sophisticated bioinformatics tools to discover new biological insights from single-cell data. Monocle is a highly used analysis tool to understand the differentiation process and enhancer-promoter pairs (Trapnell *et al.* 2014).

Single-cell trajectory analysis is a computational approach that orders cells along a pseudo time axis. This approach models transitional processes such as cell type differentiation during development. Pseudo time positions the cells along a trajectory that quantifies the relative progression of the underlying differentiation process, and it measures the cells' progress through the transition phase. Pseudo time is considered a unit of progress along the trajectory and represents the distance between a cell and the start of the trajectory. Cell types emerge from progenitor cells and transition from one state into another to differentiate into their mature form.

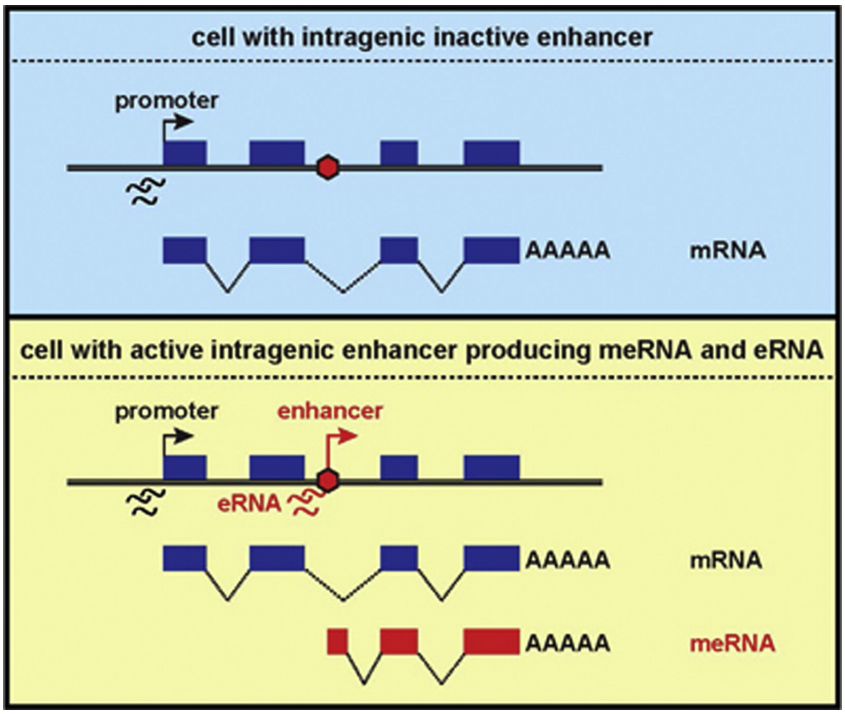

Figure 7. Enhancers as alternative promoters. Promoters are regions that initiate gene transcription and enhancers are regulatory elements that can regulate the expression of the promoter. Despite our common definition, intragenic enhancers can behave as alternative promoters resulting in messenger RNAs (meRNAs). During transcription, the promoter results in mRNAs, while enhancers produce enhancer RNAs (eRNAs). Adapted from Kowalczyk *et al*. (2012).

Pseudo time catches single cells in that transitional state and aligns them on a trajectory (Trapnell *et al.* 2014; Traag, Waltman, and van Eck 2019; Qiu *et al.* 2017). Current studies use trajectory analysis to identify marker genes that drive the lineage specification. This method is used to quantify the relative gene expression changes along the maturation process of the cells. In addition, this approach has been used to identify cell type-specific transcription factors or key gene regulators for various biological processes. Monocle is a powerful tool to unravel the changes every single cell undergoes on a pseudo time scale.

2 Aims of the Thesis

With the rise of brain organoids, studying brain development and the brain microenvironment has become possible on a large scale which was restricted by accessibility previously. Because brain organoids can facilitate an environment to mimic the microenvironment of the human brain, they represent a unique model system to study the distinct cell types in the brain during brain development. However, most of the work to date has focused on the protein-coding genes using the 3' based sequencing methods, and only a subset of the studies has concentrated on gene regulatory elements such as enhancers. Therefore, this study aims to identify the cell type-specific enhancer-promoter pairs essential in brain development and disease.

In the first phase of this project, we modify the Lancaster brain organoid protocol and successfully grow brain organoids, profiling them at three different developmental stages to identify the emergence and maturation of different brain cell types. At each stage, we generate the sequencing libraries from the organoids and perform 5' single-cell RNA sequencing to reveal the formation of more complex cell type combinations as development progresses. To this end, we annotate the cells using marker genes for the corresponding developmental stages from the literature. Next, we investigate the molecular cell type differentiation process in brain organoids over time by aligning the cells on a pseudo time using the R package monocle. Based on the detected cell types in our data, we choose seven main cell types which emerge during human brain development, such as glial cells and cortical neurons. We then perform differential expression analysis on a subset of cell types to identify cell type-specific genes. As the brain functions in circuits as a network, we assess whether the brain organoids recapitulate a cellular network by generating the transcription factor activity networks driving the cellular activities between cell types. Finally, to understand which human developmental stage the brain organoids represent, we compare the human brain organoid transcriptome to publicly available human fetal brain developmental transcriptome data. These findings will reveal what cell types the brain organoids consist of and which developmental stage our brain organoids represent.

The second part of the thesis aims to identify cell-type-specific enhancer-promoter pairs using 5’ based scRNA sequencing. Current approaches in transcriptomic profiling utilize 3’ based methods due to the robustness of the 3’ based methods. However, as enhancer RNAs can be close to promoter regions, using the 5’ based detection method serves as a more sensitive tool to identify the transcription start site of enhancers than 3’ based sequencing methods. Therefore, we first determine how many enhancers RNAs the 5’ technology can capture compared to previously used methods and integrated public datasets. Next, we link the enhancers to their target promoters and elucidate their biological functions using Cicero, an R package that infers enhancer-promoter pairs from transcriptomics data. We also integrate public HiC data to assess whether the identified enhancer-promoter pairs have genomic interaction. Finally, we search for single nucleotide polymorphisms (SNPs) and expression quantitative trait loci (eQTL) in our enhancer-promoter pairs to further validate their functional linkage.

3 Materials and Methods

3.1 Experimental Procedures

Human-induced pluripotent stem cell culture and passaging

All human induced pluripotent stem cells (hiPSCs) were maintained in a 5% CO_2 incubator at 37℃. Standard procedures were used to culture and split hiPSCs. In brief, hiPSCs were cultured on iMatrix-511™ (Nippi®) coated six-well plates with 1.5 ml Stemfit media (Ajinomoto®). Six-well plates were coated with 0.5 µg/cm^2 iMatrix-511™ (Nippi®) in 1.5 ml of PBS for an hour at 37℃. The coating was replaced with culture media before usage. Media were changed every day. Cells were split when reaching 80% confluence. For the passaging, cells were washed with 1.5 ml PBS, and 1.5 ml Accutase™ (Sigma) was added. Cells were incubated at 37℃ for 5 min. Accutase was inactivated by adding an equal amount of StemFit™ plus ten µM ROCK inhibitor (FUJIFILM Wako). Cells were spun down at 800 x g, 3 min, and resuspended in StemFit™ plus ten µM ROCK inhibitor (FUJIFILM Wako). We maintained a seeding density of 4,000 cells/cm^2.

Brain organoid generation

To generate human cerebral organoids (COs), we modified a previously published method (Lancaster and Knoblich 2014). Briefly, when hiPSCs reached 80%, confluent cells were washed with 1.5 ml PBS and dissociated into single cells with 1.5 ml Accutase™ at 37℃ for 5 min. Accutase was inactivated by adding an equal amount of StemFit™ plus ten µM ROCK inhibitor (FUJIFILM Wako). Cells were spun down at 800 x g, 3 min, and resuspended in StemFit™ plus ten µM ROCK inhibitor (FUJIFILM Wako). Cells were counted using the Countess™ II Automated Cell Counter (Thermo Fisher). Nine thousand hiPS cells were seeded per 96 wells in an ultra-low attachment plate (Corning) in Stemfit™ media plus 10 µM ROCK inhibitor. 100 µl media was replaced every other day with new media. On day 6, the embryonic bodies were transferred to a 24-well ultra-low attachment plate (Corning) using a cut 200 µl tip. The embryonic bodies were cultured in 0.5 ml neural induction media (DMEM-F12 with 1% (vol/vol) N2 supplement, 1% (vol/vol) GlutaMAX supplement and 1% (vol/vol) MEM-NEAA). On day

8, 0.5 ml neural induction medium was added. On day 10, the embryonic bodies were embedded in 35 μl growth factor-reduced Matrigel (Corning) and transferred to an ultra-low attachment six-well plate containing 4 ml cerebral media without vitamin A (for 125 ml of DMEM-F12 add 100% (vol/vol) Neurobasal medium, 1% (vol/vol) of N2 supplement, 62.5 μl of insulin, 2% (vol/vol) of GlutaMAX supplement, 1% (vol/vol) of MEM-NEAA, 2% (vol/vol) of penicillin-streptomycin, 8 μl of 2-mercaptoethanol, and 2% (vol/vol) of B27 without vitamin A supplement). A sterile parafilm was put on top of a tip box for the embedding, and gentle dimples were created. This parafilm was put into a 10 cm dish, and embryonic bodies were put into those dimples using a cut 200 μl tip. Excess media was removed before Matrigel was added. The dish was incubated for 30 min at 37℃, 5% CO_2 before transferring it to media without vitamin A using a sterile spoon. On day 10, the media was changed to 4 ml cerebral media containing vitamin A (for 125 ml of DMEM-F12, add 100% (vol/vol) Neurobasal medium, 1% (vol/vol) of N2 supplement, 62.5 μl of insulin, 2% (vol/vol) of GlutaMAX supplement, 1% (vol/vol) of MEM-NEAA, 2% (vol/vol) of penicillin-streptomycin, eight μl of 2-mercaptoethanol, and 2% (vol/vol) of B27 with vitamin A supplement) and transferred to an orbital shaker. COs were maintained in cerebral media from day ten onwards with medium changes every four days.

Single-cell dissociation

COs were transferred to an Eppendorf tube and washed twice in 1.5 ml PBS, followed by incubation in 1.5 ml Accutase™ solution at 37℃ for 45 min. COs were gently pipetted using a wide bore tip every 10 minutes. After 45 min, COs were pipetted using a wide bore tip until completely dissociated into a single cell solution. The single-cell suspension was filtered through a 35 μm filter and washed with 1.5 ml of 0.1% bovine serum albumin (Invitrogen) solution in PBS solution. Cells were washed twice, and cell number and viability were measured using the Countess™ II Automated Cell Counter (Thermo Fisher). Only samples with cell viability > 80% were used for library preparation. Samples were kept on ice until processed on the 10x Chromium machine (10x Genomics). GEM emulsion was generated following the standard protocol of 10x Genomics.

Library preparation

For all samples, 1 million cells were taken from the single-cell suspension, and volume was adjusted to a final of 1 ml with a final cell concentration of 1000 cells/μl. The target cell number was 4000 cells for the 40 day time point. For days 80 and 120, we set 6000 as the target cell number. In brief, the single-cell suspension was mixed with the Single Cell Master Mix using the oligo (dt) Reverse transcription primer (AAGCAGTGGTATCAACGCAGAGTACGAGAC–T(30)–VN) (10x Genomics, PN – 10000151) and loaded on the Chromium Controller (10x Genomics) to generate single-cell gel beads in emulsion according to manufacturer's instructions. Following PCR reactions were conducted using a Veriti™ Thermal Cycler (Applied Biosystems). Libraries were generated using the Chromium Single Cell 5' Library & Gel Bead Kit (10x Genomics, PN-1000014) by following the manufacturer's instructions for the library preparation. We applied the following changes to the recommended protocol. cDNA amplification PCR was set to 16 cycles. All libraries were indexed using the Chromium i7 Multiplex Kit (10x Genomics PN-120262). The size of the libraries was confirmed using the Bioanalyzer™ (Agilent) followed by a quantification step using KAPA™ Library Quantification kit (Kapa Biosystems). For each timepoint, libraries were pooled based on their molar concentration and sequenced on HiSeq™ 2500 (Illumina) in rapid mode with the 50 bp paired-end configuration.

3.2 Bioinformatic pipeline

All following analyses were performed using R studio 4.0.3.

Genome version

Human genome assembly version hg38 was used for all analyses of this study. Integrated datasets were first converted to hg38 using liftover (http://genome.ucsc.edu) if the initial study was conducted on hg19.

De novo peak calling and peak annotation

Raw base call files were preprocessed (alignment and quantification) using the CellRanger v2.1.1. Pipeline (10x Genomics) functions "cellranger mkfastq" and "cellranger count" with default options. We used the GRCh38 reference genome. We used SCAFE (Moody *et al.* 2021) to identify novel enhancer peaks from the "cellranger count" output folder on the bam file. We generate a custom genome for the annotation using the "scafe.tool.cm.prep_genome" function from SCAFE. We used the hg38 fasta file from UCSC (https://hgdownload.cse.ucsc.edu/goldenpath/hg38/bigZips/hg38.fa.gz) and the FANTOM6 transcript gtf file (Hon *et al.* 2017). "scafe.workflow.sc.solo" was performed for each sample. The output count table format from "scafe.workflow.sc.solo" uses cis-regulatory-element ID (CREID) as an annotation instead of gene names. To unify the detected CREID across all samples, we used the "scafe.workflow.sc.pool" function with default options on all count matrix folders from "scafe.workflow.sc.solo". "scafe.workflow.sc.pool" generates a count matrix for each input sample with unified CREIDs which was used for downstream analysis. Metadata annotation of CREID was added based on the FANTOM6 annotation files (Hon *et al.* 2017).

Cluster annotation

The "scafe.workflow.sc.pool" (Moody *et al.* 2021) output count matrix was used as input for the Seurat package 4.0.4 (Hao *et al.* 2021). We used Seurat to cluster the data,

followed by Harmony 0.1.0 (Korsunsky *et al.* 2018) for batch correction. We generated a cell data set for each time point and merged the three timepoints. We next filtered the data set, removing any cells with less than 500 CREIDs and any CREID expressed in less than three cells. We then identified the group of most variable genes across the entire dataset. We used them to estimate the most significant principal components (PCs) after batch correction and unique molecule identifier (UMI) correction. Unbiased uniform manifold approximation and projection (UMAP) (McInnes, Healy, and Melville 2020) clustering identified 34 clusters originating from all the samples. Differential expression analysis between individual clusters and the other clusters highlighted a set of marker genes for each cluster. We used published datasets of cell markers from single-cell RNA-seq studies of fetal brain samples to annotate our clusters.

Integration of public human fetal brain tissue data

We used publicly available human fetal brain transcriptomic data (Camp *et al.* 2015) to elucidate which human brain development stage the brain organoids represent. We downloaded the fetal neocortex data from 12 weeks post-conception and generated a count matrix. Using cor (R package stats, version 4.3.0), we tested the correlation of marker gene expression among the same cell types from our brain organoid data and the available human fetal brain data.

Cell trajectory analysis.

We used Monocle 3 package 1.0.0 (Cao *et al.* 2019) to predict the pseudo time. In brief, we extracted the count matrix for the Radial Glia Cell cluster and the cell cluster of interest from the Seurat object. We created a new object using Monocle3. We used principal component analysis (PCA) as our preprocessing method and chose UMAP as the reduction method. We next performed cluster_cells and learn_graph on default settings. In the oder_cells, we chose Radial Glia Cells as our starting cell cluster. We used the graph_test to identify differentially expressed genes across the pseudo time. To determine five prominent expression profiles across the pseudo time, we binned the

modules of genes. The dynamic gene expression patterns of each trajectory were plotted using pheatmap.

Motif enrichment and motif discovery analysis

We used MARA to identify the motif activity (Suzuki *et al.* 2009; Balwierz *et al.* 2014). This was performed in three steps. All steps were performed for each lineage. First, we generated a profile using "makeProfile" on the default option. Second, we used "assoTfbs" to associate the transcription factors. Finally, we used "CalcMotifAct" to calculate the motif activity.

Network construction

Using the transcription factors from the MARA analysis, we constructed a network using Cytoscape v.3.8.1 (Shannon *et al.* 2003; Maere, Heymans, and Kuiper 2005; Bindea *et al.* 2009; Smoot *et al.* 2011). Using the genes of the TFs, we first correlated the mRNA expression of the TFs with the motif activity, resulting in a weighted expression matrix. Then, this matrix was used as input to create the network in Cytoscape.

Enhancer and promoter co-expression during lineage commitment

We used the object generated from the cell trajectory analysis for this analysis. First, we used the expression profile of the five bins and calculated the average expression level for each compartment. This was performed for each enhancer and promoter in each lineage. Next, we correlated enhancer expression and promoter mRNA expression along the pseudo time using Pearson correlation for each lineage. Finally, we used the R package pheatmap to generate the heatmap visualization.

Enhancer–promoter pairs

To identify enhancer-promoter pairs, we used the Cicero version 1.2.4.11 package (Pliner *et al.* 2018), which aims to identify all co-accessible sites using the graphical

LASSO method (Friedman, Hastie, and Tibshirani 2008). Cicero estimates the correlation matrix and generates a score that penalizes pairs of distant sites more than proximal sites. As Cicero predicts all combinations of input CREIDs, we selected enhancer-promoter co-accessible pairs for downstream analysis. We applied a co-accessibility > 0.3 cut-off for pairs with a distance < 10-kilo base pairs and a co-accessibility > 0.1 cut-off for distances> 10 kilobase pairs to remain distant pairs.

Integration of enhancer atlases

We downloaded data from EnhancerAtlas (Gao and Qian 2020), SuperEnhancers (Khan and Zhang 2016), and UCSC (Fishilevich et al. 2017; Stelzer et al. 2016) to validate our enahncers. Data was first liftover to hg38 and intersected with our enhancer set. We also downloaded ChIP-seq data from Encode (Dunham *et al.* 2012; Davis *et al.* 2018) to ascertain if our enhancer set showed epigenetic enhancer marks.

Integration of HiC data

For the functional annotation of enhancer-promoter pairs, we integrated the public HiC data (Kim *et al.* 2021). In brief, this dataset was first lifted over to hg38 using LiftOver (Kent *et al.* 2002). We next used bedtools v2.30.0 (Quinlan and Hall 2010) to identify enhancer-promoter pairs overlapping with HiC interaction pairs.

Integration of dataset containing human-specific regions

Human-specific genomic regions were obtained from Reilly *et al.* (2015). Datasets were converted to hg38 using liftover (http://genome.ucsc.edu). Regions were ranked based on fold change and p-value. To be considered as a human gained region, that region has to show an increase in signal in the human sample (Benjamini-Hochberg (Benjamini and Hochberg 1995) method p-value ≤ 0.001 and fold increase ≥ 1.5) compared to mice and rhesus macaques.

Integration of Genome-Wide Association Studies data

To obtain a set of trait-associated variants, we followed the processing methods from Moody *et al.* (2021). Genome-wide association study (GWAS) summary statistics were obtained from UK biobank (Bulik-Sullivan *et al.* 2015), the Price group (https://alkesgroup.broadinstitute.org/), and the Japanese encyclopedia of genetic associations (JENGER, http://jenger.riken.jp/). JENGER was pre-processed using "munge_sumstats.py" scripts from the LDSC software (Bulik-Sullivan *et al.* 2015). To obtain the candidate gene for the trait-associated variants, lead variants with ($p < 5x10-8$) were taken from GWASdb (M. J. Li *et al.* 2016) and NHGRI-EBI GWAS Catalog (Buniello *et al.* 2019). Lead variants were searched within their linkage disequilibrium block of variants using PLINK v1.9 (Purcell *et al.* 2007) in a matched population of the 1000 Genomes Project (Leeuw *et al.* 2015).

Integration of eQTL data

We integrated expression quantitative trait locus (eQTL) data from the GTEx database (Keen and Moore 2015). Data were pre-processed as described in Yip *et al.* (2022). Variant-gene associations and the interacting eGene were taken from 49 tissues with a q-value < 0.25. The genomic location of the enhancer-promoter pairs was intersected with the GTEx data.

Integration of disease-associated genes

To understand the relevance of our enhancers in disease, we integrated data from the gene-disease association database https://www.disgenet.org/downloads (Bauer-Mehren *et al.* 2010; Piñero *et al.* 2020; Bauer-Mehren *et al.* 2011). Association of the gene to a disease infers the regulatory role of the enhancer in that disease. DisGeNET is a contains gene-disease associations, that are collected from different data sources. We included gene-disease association which are seen in at least two independent studies.

We used the R package ggplot2 and R studio 4.0.3. for visualizations if not otherwise stated.

4 Results

4.1 Characterization of human brain organoids

4.1.1 Brain organoids show brain sub-type-specific cell types

We performed 5’ based single-cell RNA sequencing (scRNA-seq) during brain organoid development at 40, 80, and 120 days (Figure 8 A-C). We harvested RNA from two organoids for each time point and generated two scRNA-seq libraries, except for day 40, represented by one library. A total of 9778 cells passed the initial quality control, and after preprocessing (see methods), we identified an increase in the number of cell clusters from 8 (40 days) to 14 cell clusters (120 days) (Figure 8 A-C). We used K-nearest neighbor graph-based clustering approach (Dong, Moses, and Li 2011). We reason that the maturation of brain organoids gave rise to more specialized cell types resulting in more cell clusters. For downstream analysis, we integrated the timepoints into one UMAP (see methods) to identify cell-type-specific differences (Figure 8 D).

Next, we assigned cell type identities to clusters by identifying cluster-specific markers across all timepoints. One cluster is a group of cells with a similar transcriptomic profile. We first looked at the 20 most differentially expressed genes obtained for each cluster by comparing a given cluster to all other clusters. For example, neuronal Progenitor Cells show high expression of proliferating genes, whereas astrocytes and oligodendrocytes show more glia-specific genes. This indicates that induced pluripotent stem cells-derived brain organoids recapitulate crucial marker genes' expression for brain subtype-specific cell types.

Comparison of the three time points (40, 80, and 120 days) indicateed time-dependent transcriptomic changes. For example, next to the Radial Glia Cells (in yellow) are the proliferating cell types such as neuronal progenitor cells. In contrast, we found more differentiated cell types as Inhibitory Interneurons further away from Radial Glial Cells. This also highlights cell type heterogeneity of brain organoids from early rising neuronal cells such as neuronal progenitor cells to specialized cell types such as astrocytes, which occur at later stages of neurodevelopment. Interestingly, in our 120 days organoids, 50% are Inhibitory Interneurons. Inhibitory Interneurons are a heterogeneous population of

neurons essential for firing properties and synaptic connectivity (Swanson and Maffei 2019). We suggest that our organoids begin expressing marker genes indicating their functional phenotype starting from day 80.

As some clusters showed similar expression patterns for the marker genes (Figure 9 C), we combined these clusters into more broad clusters, yielding seven main cell clusters (Figure 9 A-B). The high number of subclusters in the Interneuron cluster shows that these cells are in a wide range of developmental stages, indicating dynamic differences within the cell type. In addition, the Interneuron cluster shows the highest percentage increase compared to all other cell types by making up to 60% of the cell number of the 120 days organoids.

4.1.2 Brain organoids and human fetal brain show similar transcriptomic profile

Next, we wanted to understand which human brain developmental stage the organoids resemble. Therefore, we integrated a publicly available human fetal brain transcriptome dataset (Polioudakis *et al.* 2019) and correlated their expression profiles with the identified cell clusters (Figure 9 D). As a result, we observed that the brain organoid Inhibitory Interneuron Cluster shows the same marker genes as the MGE Inhibitory Neuron cells from the fetal brain tissue. Similarly, the Neuronal Progenitor Cells in brain organoids shared expression profile with dividing cells and MGE progenitor cells in the human fetal brain data set. This suggests that although the organoids were only cultured for a short period, the organoids can capture transcriptomic profiles similar to those of the human fetal brain.

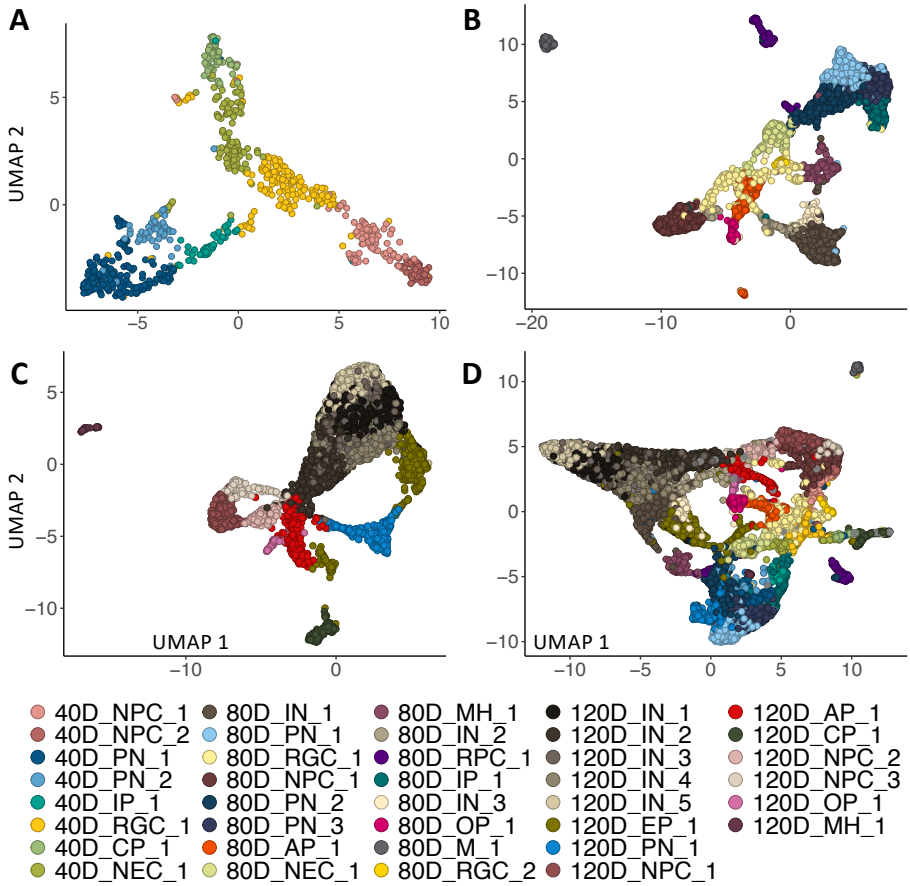

Figure 8. Cell type clustering of iPSCs derived from human brain organoids. Cell clusters were based on a graph-based clustering using the K-nearest neighbor graph. With increasing cultivation time, the number of cell types increased. The cell types also transitioned into more mature cell types and depleted progenitor cells. Cell type clustering is shown at: (A) 40 days, (B) 80 days, and (C) 120 days. (D) Combined cell clustering color-coded by each cluster of each time point. 40D: 40 days, 80D: 80 days, 120D: 120 days. NPC: Neuronal Progenitor Cells, RGC: Radial Glial Cells, AP/A: Astrocyte Progenitors/Astrocytes, OP/O: Oligodendrocyte Progenitors/Oligodendrocytes. IN: Inhibitory Interneurons, PN: Projection Neurons, IP: Intermediate Progenitors, NEC: Neuroepithelial Cells, CP: Choroid Plexus, MH: Midbrain Hindbrain, RPC: Retinal Progenitor Cells, M: Mesenchymal Cells.

4.1.3 Neuronal maturation in brain organoids by motif analysis

Next, we checked if our brain organoid model can recapitulate the maturation of human brain development in brain organoids using monocle software. Cells transition from a progenitor-like state to a more differentiated cell type during development. Monocle is a tool that quantifies at which cell transition state the cells are and projects them onto a pseudo time. This concept can help identify changes in gene expression while the cells

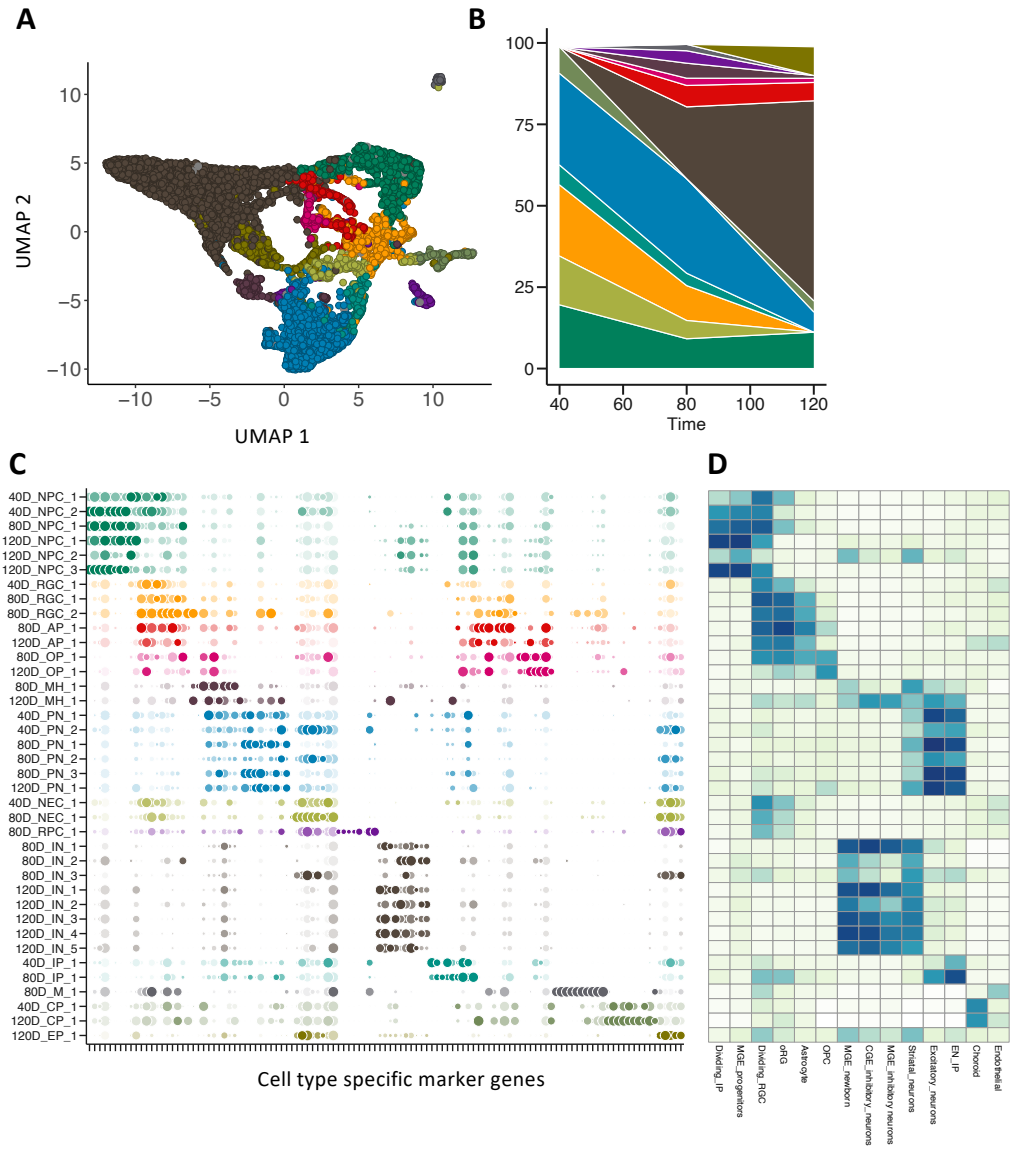

Figure 9. Comparing human iPSC-derived brain organoids with human fetal brain transcriptomic data. (A) UMAP shows broad clustering of the fine clustering from Figure 8. We used broad clustering to further analyze the main cell types from our organoids. (B) Changes in cell-type proportions over time. (C) Dot plot showing the top 20 differentially expressed genes for each lineage by broad clustering. (D) Correlation of gene expression of human iPSC-derived brain organoids to human fetal brain tissue (Polioudakis et al. 2019). 40D: 40 days, 80D: 80 days, 120D: 120 days. NPC: Neuronal Progenitor Cells, RGC: Radial Glial Cells, AP/A: Astrocyte Progenitors/Astrocytes, OP/O: Oligodendrocyte Progenitors/Oligodendrocytes. IN: Inhibitory Interneurons, PN: Projection Neurons, IP: Intermediate Progenitors, NEC: Neuroepithelial Cells, CP: Choroid Plexus, MH: Midbrain Hindbrain, RPC: Retinal Progenitor Cells, M: Mesenchymal Cells, MGE: Medial Ganglionic Eminence, oRG: outer Radial Glia Cells.

transition from stem cells into specialized neuronal cells. After selecting the start point of maturation to be Radial Glia Cells, we explored transcriptional changes of selected

genes (Figure 10 A) along the pseudo time towards more specialized cell types for seven cell types of interest (Figure 10 B). We defined Radial Glia Cells as the starting point of the transition. Then, we identified the differentially expressed genes along the pseudo time using monocle. To classify the dynamic patterns, we binned the differentially expressed genes into five distinct active groups visualizing the change of gene expression along the pseudo time in each lineage (Figure 10 C).

Using the upregulated genes of each lineage, we thus analyzed if these genes show a common transcription factor motif associated with that cell type. Transcription factors (TFs) are sequence-specific DNA binding proteins controlling the process of transcription. TFs often form a complex gene regulatory network to guarantee cell type-specific and time-dependent regulation of their target genes while becoming essential gatekeepers for cell fate decisions and lineage commitment (Kim *et al.* 2021). By identifying motifs among the upregulated genes, we aim to support the hypothesis that brain organoids recapitulate part of the human brain's cellular heterogeneity and cell maturation. We identified a list of enriched motifs, which we then used to estimate the motif activity and overlay the motif with the actual TF expression for each lineage (Figure 10 D). Our analysis showed that the TF activity overlaps with the expression level of the TF. This analysis underlines the power of transcriptomic data to infer motif activity and, therefore, dynamic changes in a cell population. In addition, we identified cell-type-specific TFs for each lineage that showed the same gene expression and motif activity dynamics.

4.1.4 Building a neuronal network using lineage-specific transcription factors

Following the motif activity analysis, we next wanted to know if our human brain organoids recapitulate a transcription factor gene regulatory network. We used the Cytoscape software (Shannon *et al.* 2003; Smoot *et al.* 2011) to visualize the neuronal gene regulatory network in brain organoids for each lineage based on the gene activity analysis from MARA (Figure 11). Motif activity response analysis (MARA) profiles (Suzuki *et al.* 2009) revealed potential regulatory functions of SMAD4, SMAD9, and ZNF143.

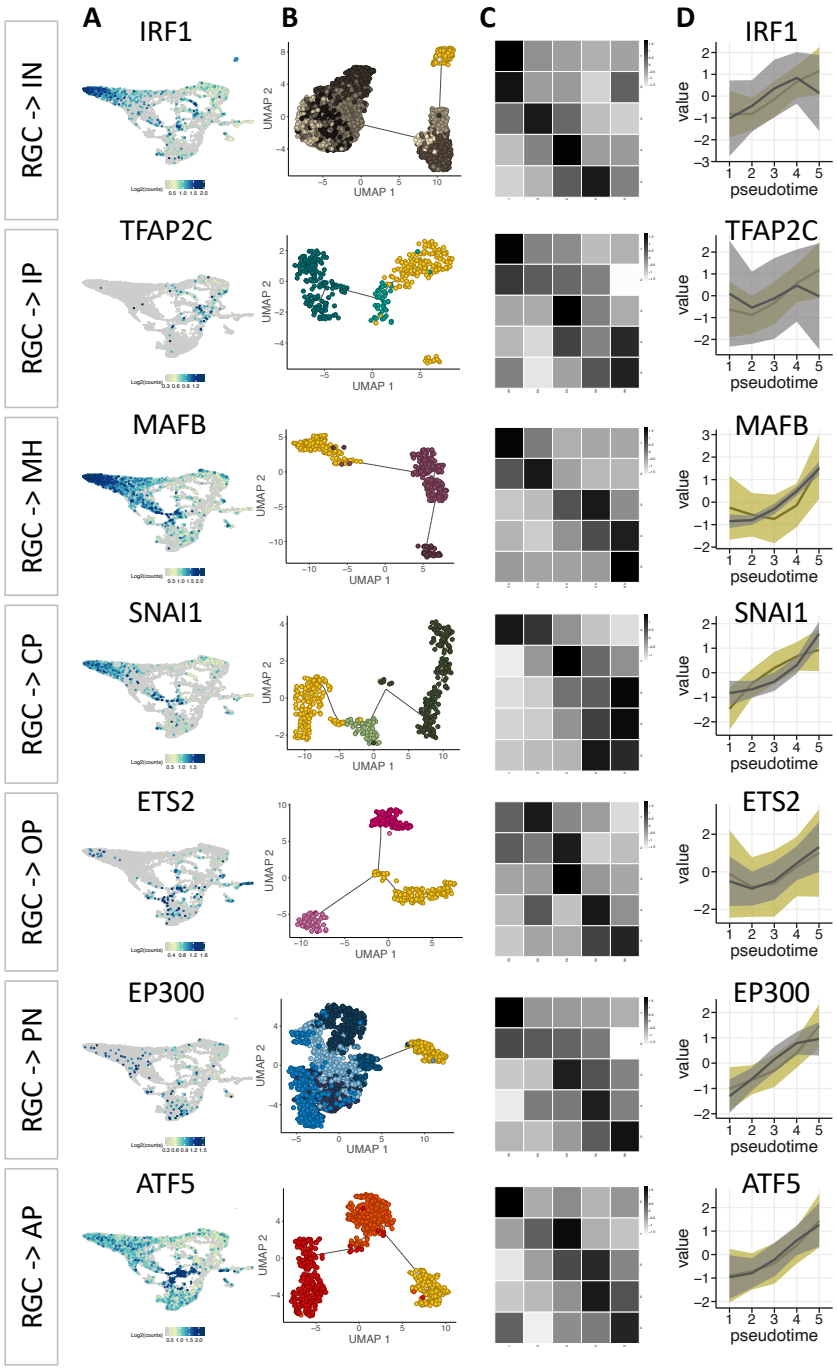

Figure 10. Motif activity and gene expression overlap in neuronal lineages. Each row represents one lineage. (A) Selected marker genes for each lineage. (B) Radial Glia Cells in yellow were chosen as starting point for each lineage. (C) Binned gene expression heatmap for each lineage. White means no expression, and black means high expression. Genes were binned into five dynamic classes. Each bin contains gene with a similar dynamic expression pattern. (D) Correlation of gene expression with motif activity of selected cluster-specific transcription markers over pseudotime. RGC: Radial Glial Cells, AP: Astrocytes, OP: Oligodendrocytes. IN: Inhibitory Interneurons, PN: Projection Neurons, IP: Intermediate Progenitors, CP: Choroid Plexus, MH: Midbrain Hindbrain.

The SMAD signaling pathway was functional in developmental processes such as proliferation and differentiation (Blank and Karlsson 2011; Fernandes, Antoine, and Hébert 2012). A recent study showed that the SMAD signaling pathway is also associated with improved neurogenesis in an Alzheimer's disease model (Wu *et al.* 2021). SMAD4, in particular, is essential for cerebellar development by driving subtypes of neurons (Fernandes, Antoine, and Hébert 2012).

ZNF143 is a mediator for a promoter-enhancer pair for murine hematopoietic stem and progenitor cell function (Bailey *et al.* 2015; Zhou *et al.* 2021). This TF is known to connect enhancers with its target gene by modulating chromatin interactions. Transcription factors can facilitate the looping of chromatin by recruiting proteins that induce or stabilize the folding of the DNA and thus make the interaction of the enhancer with its distal promoter possible. Moreover, it was shown that the modulating chromatin interaction alone was insufficient to mediate the promoter-enhancer interaction, yet ZNF143 TF was crucial for this interaction.

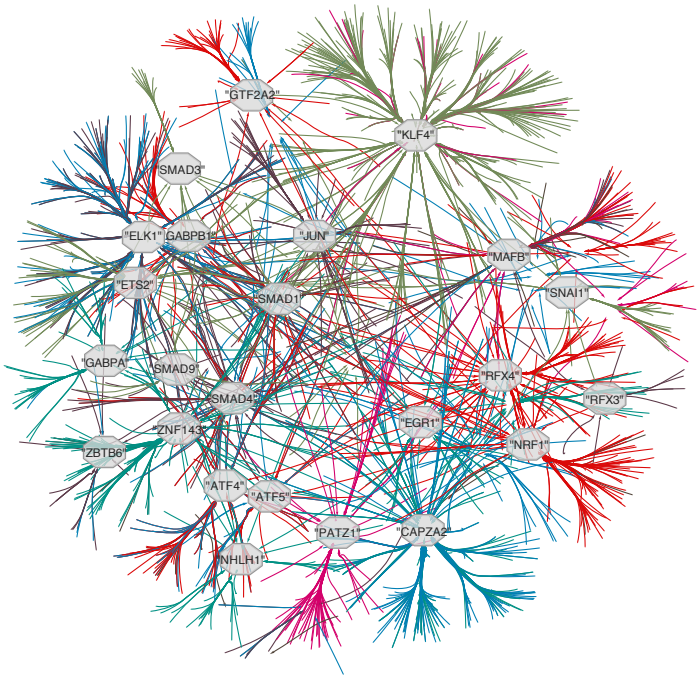

Figure 11. Transcription factor network in human brain organoids. Colors represent the lineages. Green: Choroid plexus. Blue: Progenitor cells. Red: Astrocytes. Pink: Oligodendrocytes. Brown: Inhibitory Interneurons. Blue: Projection Neurons. Branches of certain colors stemming from a given transcription factor mean that this transcription factor regulates the development of the corresponding cell lineage. SMAD1 is an important transcription factor for choroid plexus. However, our data suggest it also regulates astrocytes.

4.2 Identification of enhancer-promoter pairs

4.2.1 Enhancer definition and detection are highly platform-dependent.

eRNAs are relatively lowly expressed and display poor sequence conservation than protein-coding genes. These features hamper the discovery of eRNAs transcription start sites, ultimately leading to their inadequate genomic and functional annotations. To increase the sensitivity of eRNAs identification, we used a recently published tool called SCAFE (Moody *et al.* 2021), which is especially suitable for identifying putative enhancer peaks. SCAFE outperforms previous tools by utilizing a logistic regression to identify genuine enhancers. We detected 15,306 non-coding transcripts using SCAFE compared to 20,904 for coding regions. Such detection rates align with expectations of the roughly equal number of coding and non-coding transcripts originating from the human genome. We then investigated how many of the detected potential enhancer elements overlap with previously identified enhancers from the literature and how many of our enhancers show epigenetic marks. However, most of our enhancers (9146, 59%) are not defined by other resources (Figure 12 A) nor show a DNA hypersensitive site (Figure 12 B). We suggest that the enhancers we detected are unique to the 5' single-cell RNA seq technology. This could infer that RNA sequencing can capture enhancers without DNA hypersensitive sites.

Next, we used the FANTOM collection of enhancer annotations (Forrest *et al.* 2014; Andersson *et al.* 2014; Arner *et al.* 2015). Surprisingly, we found many eRNAs which are absent from the FANTOM set (Figure 12 C). The second largest group overlapped with long non-coding RNAs. This can be explained by the fact that enhancers and long non-coding RNA both belong to the category of non-coding RNAs (Hou, Zhang, and Sun 2019).

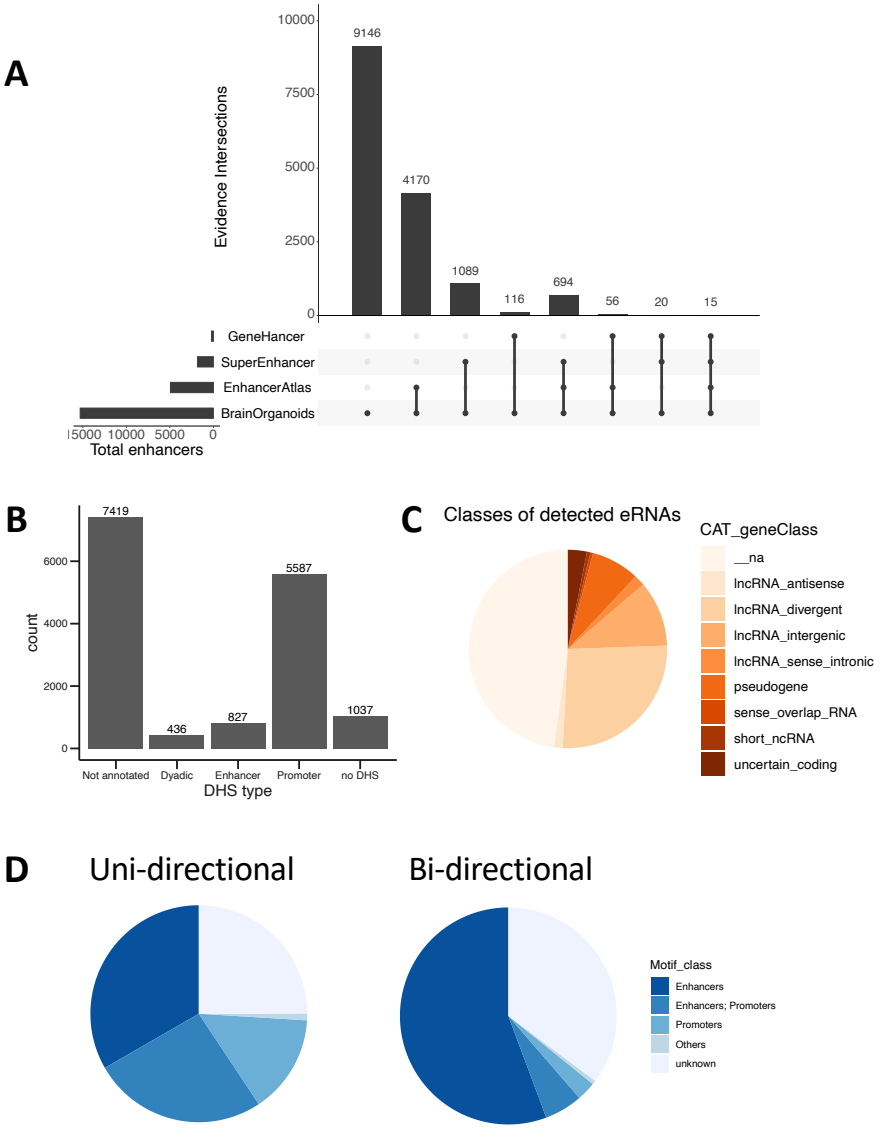

Figure 12. Detection of and characteristics of enhancers in human brain organoids. (A) Validation of enhancers using public datasets. The majority of our enhancers (9146) are only detected in our dataset (a total of 15,306). Next, our dataset showed the most considerable overlap with EnhancerAtlas (4170), followed by SuperEnhancer (1089). From the three integrated databases, our data shows the least overlap with GeneHancer enhancers (116). (B) Gene classification is based on DNA hypersensitive sites (DHS). DHS is a characteristic of active enhancers and we conclude that enhancers from transcriptomic data do not show DHS with our experimental conditions. (C) Using the FANTOM CAT annotation, we aimed to elucidate if our enhancers showed overlapping annotation with long-non-coding RNAs (lncRNAs). (D) eRNAs with bi-directional expression show more enhancer motifs than uni-directional expressed eRNAs.

Next, we wanted to know if eRNAs in our data show distinctive characteristics of enhancers with bi-directional expression profiles (Figure 12 D). Interestingly, when we compared the motif signature of unidirectional compared to bi-directional eRNAs, we

observed that > 50% of the bi-directional eRNAs show an enhancer signature in the ChIP-seq data. In contrast, only 35% of the unidirectional enhancers show an enhancer like a signature. Our results support that bi-directional expression of eRNAs can be used to infer functional enhancers more likely than unidirectional eRNAs. It also suggests that the overlap of enhancer identification across different approaches is low, as previously reported (Benton *et al.* 2019).

4.2.2 Enhancer-promoter pairs are co-expressed in early brain development

Predicting the target genes of enhancers remains a significant challenge. Enhancer-promoter interactions are mainly inferred from Assay for Transposase-Accessible Chromatin (ATAC) or RNA sequencing data, where enhancers are defined based on the open chromatin regions or their eRNA expression. However, some studies have reported the inability of open chromatin regions to infer the activity of enhancers (Carullo *et al.* 2020). Upon perturbation of the enhancer, the chromatin remained open while the eRNA expression was decreased.

Here we hypothesize that enhancer-promoter pairs can be inferred from the 5'-end single-cell RNA sequencing data by co-expression and co-accessibility of the eRNAs and their cognate RNA promoters. Assuming enhancer-promoter pairs from transcriptomic data harbors unique advantages, as the enhancer's transcription as eRNA can prove enhancer activity. Using the pseudo time of monocle, we correlated the eRNA clusters' expression with the promoter clusters' expression along the pseudo time. Each cluster represents a set of enhancers or promoters with similar dynamics along the pseudo time (Figure 13). This result highlights enhancer clusters in the early phase of the pseudo time cluster with early promoter clusters. This could infer that enhancer and promoter expression levels are especially co-expressed in early biological stages like Radial Glial Cells. On the other hand, enhancer-promoter pairs show less co-expressed correlation toward more differentiated cells.

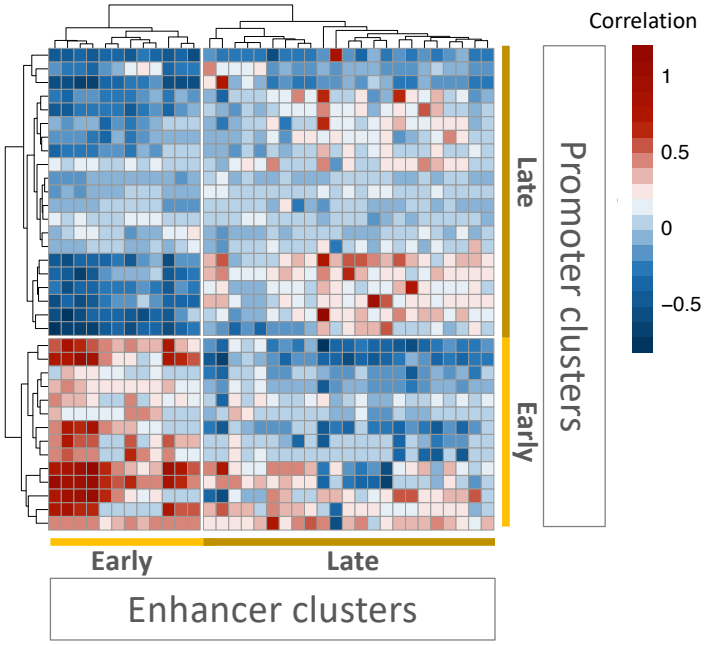

Figure 13. Identification of enhancer-promoter pairs by correlating the expression level of enhancer RNA with promoters by pseudo time. Expression of early enhancer RNA and early promoter showed the highest correlation, indicating that enhancers are more active in early developmental stages. Enhancer clusters and promoter clusters are derived from the trajectory analysis (Figure 10 C).

4.2.3 Co-accessible enhancer-promoter pairs

Next, we used cicero software to calculate the co-accessibility of enhancer-promoter pairs. Cicero is based on the probability of two elements being nearby based on the transcriptomic similarity of the cell and generates a co-accessible score. After calculating the co-accessibility scores of all pair combinations, we selected the 49,403 enhancer-promoter pairs for downstream analysis. As cicero imposes a penalty on scoring distant enhancer-promoter pairs, we corrected that by setting different co-accessibility cut-offs depending on the enhancer-promoter distance. We set a co-accessibility cut-off of 0.3. However, for pairs where the enhancer is further away than 10,000 base pairs from the promoter, we set a 0.1 co-accessibility cut-off. By including this distance correction, we retain distant pairs. This resulted in 4417 pairs filtering out 91% of the initial pairs (Figure 14). Additionally, we selected pairs with the enhancer and promoter expressed in the same pseudo time trajectory, resulting in over 3000 of the 4417 pairs being expressed in only one trajectory, indicating a cell-type-specific expression (Figure 15 B).

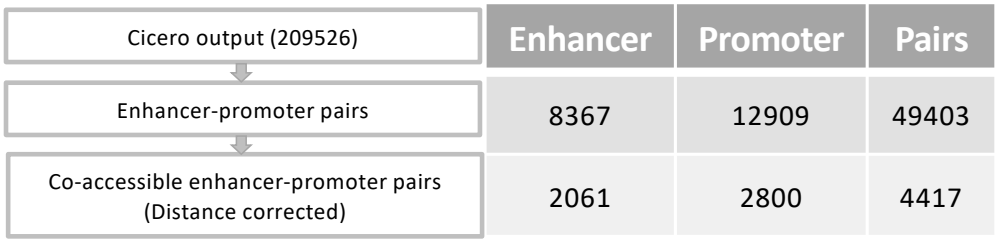

Figure 14. Selection flowchart for a robust set of enhancer-promoter pairs. Cicero output generated 209526 pairs which we filtered for enhancer-promoter pairs. Cicero calculates all possible interactions, including promoter-promoter and enhancer-enhancer interactions. The filtering showed that appr. 25 % of the detected pairs were enhancer-promoter pairs. Distance correlated filtering identified 4417 robust enhancer-promoter pairs (appr. 10%) with a high co-accessibility score. Cicero penalizes distant pairs with low scores. To include distant enhancer-promoter pairs we set a lower cut-off for pairs where the enhancer and promoter are > 10000 base pairs apart.

4.2.4 Cell type-specific enhancer-promoter pairs during brain development

Transcribed enhancers have cell-type-specific expression profiles (Nott *et al.* 2019). We thus hypothesized that 5' based enhancers would contain cell-specific enhancers during brain organoid development. However, as enhancers are generally very lowly expressed, methods utilizing the differences in expression level between cell types to calculate cell-type specificity are not suitable.

We, therefore, explored alternative ways of defining enhancer specificity. First, we described an enhancer as cell type-specific if its target protein-coding gene was cell type-specific and differentially expressed (average log2 fold change > 0.25). Collectively we identified 1627 cell-type-specific enhancer-promoter pairs (Figure 15 A). Next, we investigated in which trajectory the pairs are expressed. We observed that most pairs are expressed in the Projection Neuron Cluster, which has the highest number of sub-clusters (Figure 15 C). This could indicate that Projection Neuron Cells are intrinsically more heterogeneous and potentially undergo a longer differentiation process than other cell types, resulting in more intermediate stages represented by a higher number of clusters.

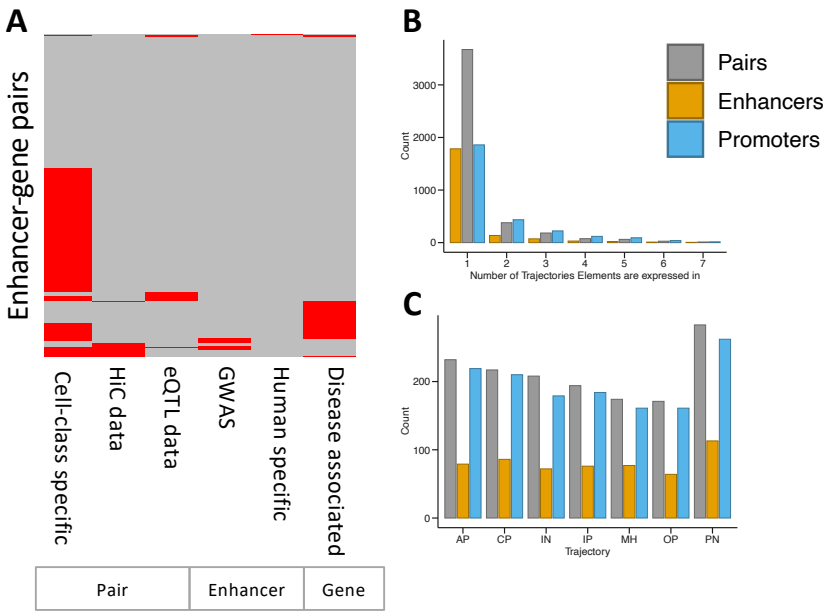

Figure 15. Enhancer-gene and enhancer-promoter characteristics. (A) Heatmap with integrated metadata for pairs, enhancers, and genes. Red means positive, and grey means negative. Each column represents one feature. Cell-class specific, HiC data, and eQTL data represent pairs, affecting the enhancer and the promoter. These features support the discovered enhancer-promoter pairs. GWAS and human-specific affect the enhancer. Enhancers have been associated with a disease or carry a human-specific region. Disease-associated means that the promoter of that pair is linked to a gene associated with a disease. (B) The number of trajectories enhancers, promoters, or pairs is expressed in. (C) In which trajectories are enhancers, promoters, and pairs are expressed. AP: Astrocytes, CP: Choroid Plexus, IN: Inhibitory Interneurons, IP: Intermediate Progenitors, MH: Midbrain/Hindbrain, OP: Oligodendrocytes, PN: Projection Neurons.

Our analysis captured two previously described brain-specific enhancer-promoter pairs (Figure 16). ARX is a gene known to regulate synaptic strength and plasticity (Dickel *et al.* 2018). Dickel *et al.* have identified two enhancers of ARX, which upon knock-out in mice result in depletion of neurons. ARX and its enhancer are mainly expressed in the Interneuron cell cluster indicating the cell-type-specific function of this gene and its enhancer.

Furthermore, we detected a novel enhancer-promoter pair which showed a cell-type-specific expression pattern (Figure 16). PMP2 is a gene that has been associated with myelination and astrocytes before (Matejuk and Ransohoff 2020; Hong *et al.* 2016). In a study from Pantera *et al.* (2018), they detected a super-enhancer region 90-120 kilobase pair upstream of PMP2, whereas the enhancer we identified is 60 kilobase pair upstream

of PMP2. Although enhancers of PMP2 have been described before, we identified an additional enhancer of PMP2.

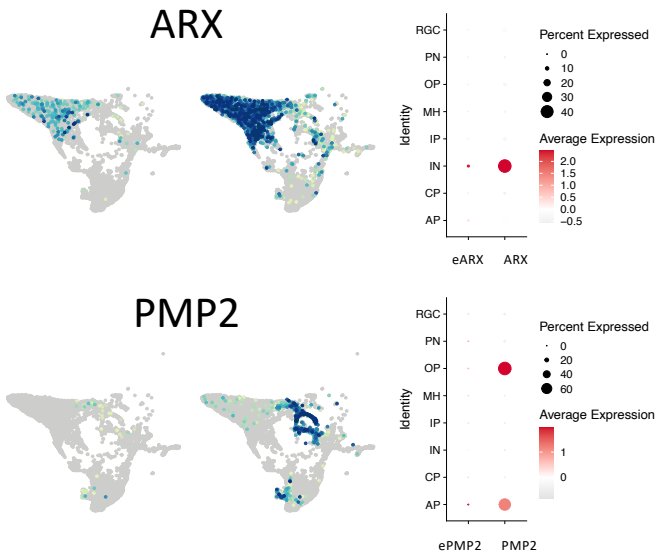

Figure 16. Selected examples for cell-type-specific enhancer-promoter pairs. The left panel showed the enhancer-gene pair on a UMAP plot, indicating the expression of enhancer and gene in the same cell cluster. The color scale indicates the log2 count values. On the right panel, the color of the dot plot depicts the scaled average expression level of the enhancer or promoter. The size of the circle represents the percentage of cells expressing the enhancer or promoter.

4.2.1 Functional enhancer-promoter pairs

Next, we aimed to identify functional enhancer-promoter pairs. Distal enhancers with their target gene by looping to their interacting gene. This looping is facilitated by transcription factors and structural proteins leading to a genomic interaction with the promoter. HiC can capture this interaction. Genomic interaction of enhancers with promoters is used to identify which promoter the enhancer is regulating. We explored whether our enhancer-promoter pairs show genomic interactions by integrating Hi-C data. The intersection of transcriptomic-based enhancer-promoter pairs with HiC-based enhancer-promoter pairs identified 53 pairs from 1627 pairs to show a genomic interaction. Although enhancers interact with promoters through looping and genomic interaction, our data captured more diverse cell types compared to the integrated HiC data. Thus, we suggest that our of enhancer-promoter pairs are robust but not active in

the integrated HiC data set, resulting in only a subset of our enhancer-promoter pairs showing genomic interactions.

Besides genomic interaction, we aimed to identify enhancer-promoter pairs associated with an expression quantitative trait loci (eQTL). eQTLs link single nucleotide polymorphisms (SNP) to a transcript and enhancers are known to harbor SNPs. By intersecting our enhancer-promoter pairs, we identified two pairs that carry an eQTL. WDR47, one of the pairing genes, has been associated with brain development before (Kannan *et al.* 2017). The eQTLs that overlap with our data were identified in brain samples.

Our data identified 53 cell-type-specific enhancer-promoter pairs with genomic interaction support. Furthermore, two of these pairs carry an eQTL. These examples indicate that we can use brain organoids to study cell-type-specific genetic variations.

5 Discussion

Most currently available brain organoid sequencing data is based on 3’ sequencing. Despite 3’ based sequencing being a powerful tool to study gene expression, it does not capture the transcription start site and therefore can miss enhancer RNAs. To precisely and sensitively capture enhancer RNA, we utilized the strength of 5’ based RNA sequencing in combination with the novel SCAFE tool. Using logistic regression, SCAFE outperforms previous tools to annotate genuine enhancers (Moody *et al.* 2021). Enhancers are non-coding elements involved in gene regulation. They interact with transcription factors to activate gene expression through interaction with a promoter. We aimed to identify cell-type-specific enhancers and unrevealed their biological importance by identifying their regulating promoter. With 98% of our genome being non-coding, studies have suggested that they might contribute to regulatory functions (Perenthaler *et al.* 2019).

We successfully established the brain organoid protocol in our lab. Following 5’ based single-cell sequencing captured a transcriptomic profile similar to fetal human brain development. Based on the molecular footprint, we suggest that our brain organoids partially recapitulate human brain development at post-conception week 12. Differential expression analysis revealed that brain organoids capture marker genes of major neuronal cell types such as astrocytes. Furthermore, we detected upregulation of cell type-specific transcription factors indicating the lineage commitment of the main neuronal cell types. Finally, network analysis showed an interplay of cell type-specific transcription factors inferring transcription factors to gene relations during brain development.

Although brain organoids provide an excellent model system to study gene regulation at an early developmental stage, brain organoids show limitations in the growth and complexity of the growing complexity of the human brain (Qian, Song, and Ming 2019; Benito-Kwiecinski and Lancaster 2020). Our findings show that brain organoids with 120 days of culture can recapitulate human brain development of post-conception week 12. We captured progenitor cells and distinct mature cells of the human brain. Despite this,

the lack of intrinsic vascularization and spontaneous differentiation of microglial cells remains a significant challenge in the field of brain organoids. Vasculature supports the continuous growth of tissue and, therefore, the emergence of more mature cells of the human brain. Studies have shown the incorporation of vasculature into brain organoids (Cakir *et al.* 2019; Matsui *et al.* 2021). However, these studies are sparse, indicating technical challenges and a need for better-defined protocols. An alternative approach to overcome this shortage is by restricting the growth of the brain organoid to a smaller size and using directed brain organoid protocols which result in one central area of the brain (Monzel *et al.* 2017). However, gaining more control and reproducibility by using a directed differentiation protocol means loss of molecular heterogeneity and, therefore, the neuronal network, which is one of the compelling advantages of organoids. In addition, providing a 3D structure enables different cell types to interact in a unique environment leading to a more complex transcriptome. Finding the balance between heterogeneity and reproducibility will remain a challenge.

Brain networks represent the interaction of brain regions by showing their functional connectivity (Summers *et al.* 2022). Especially transcription factors are essential players, as they drive cell differentiation and have unique functions for cell fate decisions. Modeling this network of transcription factors in-silico enables us to study regulatory elements in a developing system on a large scale. MARA stands for motive activity response analysis and is a powerful tool for elucidating the role of transcription factors for a cell type. This tool deduces which transcription factors have a role in certain expression profiles by linking the motifs and their activities to the transcription factor that binds to them. Our analysis showed a great correlation between the motif activity and transcription factor expression levels across all trajectories. Because we used cell type-specific transcription factors, the network analysis reflects the interaction of cell types, which can be used to study the differences between organoids based on their transcription factor activity. Brain organoids could serve as a model system to identify novel motifs during early brain development.

In the second part of the thesis, we identified novel enhancers and cell-type-specific enhancer-promoter pairs using 5’ based single-cell RNA sequencing. Many of the

enhancers from our study have not been described before. Although we included various quality control metrics to obtain a robust set of enhancers, this difference could result from the different technologies used. Previous resources and atlas papers have concluded that enhancer detection varies across different platforms due to protocol differences and the unique characteristic each platform uses to identify enhancers (Benton *et al.* 2019; Inoue and Ahituv 2015). Although our results align with previous studies, there is room for improvement in unifying the gap between technologies. Each platform has its unique features, and finding a way to combine each feature from each platform would provide an advanced resource and high-quality enhancers. Several resources have facilitated the integration of different characteristics to define a complete annotation of identified enhancers (Pennacchio *et al.* 2006; Gao and Qian 2020).

On the flip side, the reason why we detect more enhancers than other platforms could mean that RNA sequencing is more sensitive to capturing active enhancers using eRNAs as a surrogate marker compared to other platforms. Although they might not all be functional due to environmental and dynamic reasons, this data provides an excellent resource for in silico detection of enhancers using eRNAs.

By the definition of being cell-type specific, the same enhancer is not active in every cell type. Investigating functional enhancers requires experimental validations or data retrieved from various cell types to investigate the true functional nature of enhancers. A state-of-the-art method to identify eRNA would be the first step toward a full enhancer annotation set.

Going forward, the reliable detection of functional enhancers is crucial. However, the bottleneck is a reliable characteristic of enhancers by which enhancers can be defined. Our current knowledge of enhancers is based on studies highlighting distinct enhancer characteristics. However, these previous investigations contradict and highlight opposing features of enhancers leading to an incomplete understanding of enhancers (Zhu *et al.* 2013; Thurman *et al.* 2012). One way of addressing this challenge is to introduce a better catalog of enhancers by deeply characterizing enhancers. We might use the term enhancers in a broad context and therefore miss the distinct differences

between different enhancer categories. To accelerate ongoing experimental enhancer validations, recent studies have employed machine learning approaches (Wolfe *et al.* 2021; Rajpurkar *et al.* 2021). Wolfe et al. have used deep learning approaches to generate a model that can accurately predict enhancers in Drosophila. Deep learning uses algorithms and neural networks to train a model. Generating high-quality input data to generate an advanced model could expedite current enhancer validations.

Another critically discussed feature of eRNAs is their polyadenylation. There have been studies supporting the idea of polyadenylated eRNAs (Andersen *et al.* 2014; Santa *et al.* 2010), whereas other studies have also found non-polyadenylated eRNAs (Flynn *et al.* 2011; Lubas *et al.* 2015). Based on our experimental design, our study only captured enhancers that are characterized by polyadenylated eRNA. Despite this, polyadenylation might be a functional attribute to the eRNA, inferring that these eRNAs need to be active for a prolonged time. In contrast, non-polyadenylated eRNAs are only required for a short period.

The current trend in the medical field shows a considerable interest in understanding the molecular footprint of transcriptomics on a single cell level since our understanding of tissue grows, and literature has demonstrated that bulk expression can often overshadow lowly expressed genes that are otherwise lost. Although single-cell RNA sequencing might be nosier, it also provides the excellent potential to discover new regions that bulk studies could not access so far. Single-cell data enable us to study the concept of single-cell stage vs. cell type. When cells undergo a developmental process they differentiate from one cell type into a more developed cell type during which the cell is in a transient state. Single-cell gene expression studies facilitate profiling transcriptional regulation in complex biological processes and highly heterogeneous cell populations. These studies simplify the discovery of cell-type-specific genes of a particular cell type from a mix of heterogeneous cell types, mark intermediate states during a natural process, and bifurcate between two alternative cellular fates. The current study provided insight into novel cell-type-specific enhancers and cell-type-specific enhancer-promoter pairs, underlying the importance of single-cell studies.

6 Conclusion and Outlook

Our understanding of enhancers is a dynamic process. With the rise of more sensitive approaches to capture enhancers, the identification and characterization of enhancers have been under massive studies.

This thesis aimed to identify cell type-specific enhancer-promoter pairs essential in brain development. Our analysis showed that brain organoids have an excellent potential for modeling human brain development. The diversity of cell types we obtained from our experiments confirmed that organoids undergo a maturation process recapitulating human fetal brain development. The transcriptomic profile showed that our model system represents the human fetal brain of post-conception week 12. Furthermore, we identified a robust set of active and cell type-specific enhancers. By linking the enhancers to promoters, we elucidated their biological function. Finally, by integrating HiC and eQTL data, we further assessed the functional role of enhancers during brain development.

In conclusion, this work identified cell type-specific enhancers from human-induced stem cell-derived brain organoids. We highlighted the purpose and usage of 5' based transcriptomic approaches. 5' based sequencing approaches are not as common as 3' based approaches. However, enhancers and alternative promoters can be distinguished by capturing the 5' end of the transcript, emphasizing the value of 5' based sequencing approaches. Our analysis extends the current understanding of enhancer-promoter pairs in the brain. We identified a set of novel cell types specific pairs that have not been described yet. We suggest that these pairs are important during brain development, affecting cell type differentiation.

Many of the identified enhancers are not seen in other studies, providing a unique platform to identify novel regulatory regions. However, with this uncertainty comes the lack of support for the novel enhancers, and only functional analysis or experimental validation will prove these as genuine enhancers. Clustered regularly interspaced short palindromic repeats (CRISPR) applications are widely used to study the function of enhancers (Fulco *et al.* 2019; Li *et al.* 2020). Thus, capturing the transcription start site

is a significant advantage when considering perturbation studies using CRISPR interference. Furthermore, it is known that CRISPR applications are most efficient when targeting the transcription start of their element of interest (Yang *et al.* 2021). Therefore, our data aids to design better gRNAs compared to common 3' based sequencing data.

Current studies attempt to integrate multiple technologies to identify the most robust set of enhancers. By integrating multiple technologies, one decreases the chances of falsely annotated enhancers. Whereas single studies lack the consensus on enhancers, integrating atlases into one resource can unite our expertise to interpret the function of enhancers during brain development. The definition of an active enhancer on a global scale is yet to be defined. However, with currently ongoing approaches and integrating various datasets, resources are becoming more defined and taking our understanding of enhancers one step further.

7 References

Aguilo, Francesca, SiDe Li, Natarajan Balasubramaniyan, Ana Sancho, Sabina Benko, Fan Zhang, Ajay Vashisht, et al. 2016. “Deposition of 5-Methylcytosine on Enhancer RNAs Enables the Coactivator Function of PGC-1α.” *Cell Reports* 14 (3): 479–92. https://doi.org/10.1016/j.celrep.2015.12.043.

Allende, Maria L., Emily K. Cook, Bridget C. Larman, Adrienne Nugent, Jacqueline M. Brady, Diane Golebiowski, Miguel Sena-Esteves, Cynthia J. Tifft, and Richard L. Proia. 2018. “Cerebral Organoids Derived from Sandhoff Disease-Induced Pluripotent Stem Cells Exhibit Impaired Neurodifferentiation[S].” *Journal of Lipid Research* 59 (3): 550–63. https://doi.org/10.1194/jlr.M081323.

Andersson, Robin, Claudia Gebhard, Irene Miguel-Escalada, Ilka Hoof, Jette Bornholdt, Mette Boyd, Yun Chen, et al. 2014. “An Atlas of Active Enhancers across Human Cell Types and Tissues.” *Nature* 507 (7493): 455–61. https://doi.org/10.1038/nature12787.

Andersson, Robin, Peter Refsing Andersen, Eivind Valen, Leighton J. Core, Jette Bornholdt, Mette Boyd, Torben Heick Jensen, and Albin Sandelin. 2014. “Nuclear Stability and Transcriptional Directionality Separate Functionally Distinct RNA Species.” *Nature Communications* 5 (1): 5336. https://doi.org/10.1038/ncomms6336.

Arner, Erik, Carsten O. Daub, Kristoffer Vitting-Seerup, Robin Andersson, Berit Lilje, Finn Drabløs, Andreas Lennartsson, et al. 2015. “Transcribed Enhancers Lead Waves of Coordinated Transcription in Transitioning Mammalian Cells.” *Science* 347 (6225): 1010–14. https://doi.org/10.1126/science.1259418.

Arnold, Preston R., Andrew D. Wells, and Xian C. Li. 2020. “Diversity and Emerging Roles of Enhancer RNA in Regulation of Gene Expression and Cell Fate.” *Frontiers in Cell and Developmental Biology* 7. https://www.frontiersin.org/article/10.3389/fcell.2019.00377.

Bailey, Swneke D., Xiaoyang Zhang, Kinjal Desai, Malika Aid, Olivia Corradin, Richard Cowper-Sal·lari, Batool Akhtar-Zaidi, Peter C. Scacheri, Benjamin Haibe-Kains, and Mathieu Lupien. 2015. “ZNF143 Provides Sequence Specificity to Secure

Chromatin Interactions at Gene Promoters." *Nature Communications* 6 (1): 6186. https://doi.org/10.1038/ncomms7186.

Balwierz, Piotr J., Mikhail Pachkov, Phil Arnold, Andreas J. Gruber, Mihaela Zavolan, and Erik van Nimwegen. 2014. "ISMARA: Automated Modeling of Genomic Signals as a Democracy of Regulatory Motifs." *Genome Research*, February. https://doi.org/10.1101/gr.169508.113.

Bateman, Jack R., Justine E. Johnson, and Melissa N. Locke. 2012. "Comparing Enhancer Action in Cis and in Trans." *Genetics* 191 (4): 1143–55. https://doi.org/10.1534/genetics.112.140954.

Bauer-Mehren, Anna, Markus Bundschus, Michael Rautschka, Miguel A. Mayer, Ferran Sanz, and Laura I. Furlong. 2011. "Gene-Disease Network Analysis Reveals Functional Modules in Mendelian, Complex and Environmental Diseases." *PLOS ONE* 6 (6): e20284. https://doi.org/10.1371/journal.pone.0020284.

Bauer-Mehren, Anna, Michael Rautschka, Ferran Sanz, and Laura I. Furlong. 2010. "DisGeNET: A Cytoscape Plugin to Visualize, Integrate, Search and Analyze Gene–Disease Networks." *Bioinformatics* 26 (22): 2924–26. https://doi.org/10.1093/bioinformatics/btq538.

Benito-Kwiecinski, Silvia, and Madeline A. Lancaster. 2020. "Brain Organoids: Human Neurodevelopment in a Dish." *Cold Spring Harbor Perspectives in Biology* 12 (8): a035709. https://doi.org/10.1101/cshperspect.a035709.

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B (Methodological)* 57 (1): 289–300.

Benton, Mary Lauren, Sai Charan Talipineni, Dennis Kostka, and John A. Capra. 2019. "Genome-Wide Enhancer Annotations Differ Significantly in Genomic Distribution, Evolution, and Function." *BMC Genomics* 20 (1): 511. https://doi.org/10.1186/s12864-019-5779-x.

Bienvenu, Thierry, Karine Poirier, Gaelle Friocourt, Nadia Bahi, Delphine Beaumont, Fabien Fauchereau, Lamia Ben Jeema, et al. 2002. "ARX, a Novel Prd-Class-Homeobox Gene Highly Expressed in the Telencephalon, Is Mutated in X-Linked

Mental Retardation." *Human Molecular Genetics* 11 (8): 981–91. https://doi.org/10.1093/hmg/11.8.981.

Bindea, Gabriela, Bernhard Mlecnik, Hubert Hackl, Pornpimol Charoentong, Marie Tosolini, Amos Kirilovsky, Wolf-Herman Fridman, Franck Pagès, Zlatko Trajanoski, and Jérôme Galon. 2009. "ClueGO: A Cytoscape Plug-in to Decipher Functionally Grouped Gene Ontology and Pathway Annotation Networks." *Bioinformatics* 25 (8): 1091–93. https://doi.org/10.1093/bioinformatics/btp101.

Blank, U., and S. Karlsson. 2011. "The Role of Smad Signaling in Hematopoiesis and Translational Hematology." *Leukemia* 25 (9): 1379–88. https://doi.org/10.1038/leu.2011.95.

Bulik-Sullivan, Brendan K., Po-Ru Loh, Hilary K. Finucane, Stephan Ripke, Jian Yang, Nick Patterson, Mark J. Daly, Alkes L. Price, and Benjamin M. Neale. 2015. "LD Score Regression Distinguishes Confounding from Polygenicity in Genome-Wide Association Studies." *Nature Genetics* 47 (3): 291–95. https://doi.org/10.1038/ng.3211.

Buniello, Annalisa, Jacqueline A L MacArthur, Maria Cerezo, Laura W Harris, James Hayhurst, Cinzia Malangone, Aoife McMahon, et al. 2019. "The NHGRI-EBI GWAS Catalog of Published Genome-Wide Association Studies, Targeted Arrays and Summary Statistics 2019." *Nucleic Acids Research* 47 (D1): D1005–12. https://doi.org/10.1093/nar/gky1120.

Cakir, Bilal, Yangfei Xiang, Yoshiaki Tanaka, Mehmet H. Kural, Maxime Parent, Young-Jin Kang, Kayley Chapeton, et al. 2019. "Engineering of Human Brain Organoids with a Functional Vascular-like System." *Nature Methods* 16 (11): 1169–75. https://doi.org/10.1038/s41592-019-0586-5.

Camp, J. Gray, Farhath Badsha, Marta Florio, Sabina Kanton, Tobias Gerber, Michaela Wilsch-Bräuninger, Eric Lewitus, et al. 2015. "Human Cerebral Organoids Recapitulate Gene Expression Programs of Fetal Neocortex Development." *Proceedings of the National Academy of Sciences* 112 (51): 15672–77. https://doi.org/10.1073/pnas.1520760112.

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. https://doi.org/10.1038/s41586-019-0969-x.

Carullo, Nancy V N, Robert A Phillips III, Rhiana C Simon, Salomon A Roman Soto, Jenna E Hinds, Aaron J Salisbury, Jasmin S Revanna, et al. 2020. “Enhancer RNAs Predict Enhancer–Gene Regulatory Links and Are Critical for Enhancer Function in Neuronal Systems.” *Nucleic Acids Research* 48 (17): 9550–70. https://doi.org/10.1093/nar/gkaa671.

Creyghton, Menno P., Albert W. Cheng, G. Grant Welstead, Tristan Kooistra, Bryce W. Carey, Eveline J. Steine, Jacob Hanna, et al. 2010. “Histone H3K27ac Separates Active from Poised Enhancers and Predicts Developmental State.” *Proceedings of the National Academy of Sciences* 107 (50): 21931–36. https://doi.org/10.1073/pnas.1016071107.

Dao, Lan T. M., and Salvatore Spicuglia. 2018. “Transcriptional Regulation by Promoters with Enhancer Function.” *Transcription* 9 (5): 307–14. https://doi.org/10.1080/21541264.2018.1486150.

Davis, Carrie A, Benjamin C Hitz, Cricket A Sloan, Esther T Chan, Jean M Davidson, Idan Gabdank, Jason A Hilton, et al. 2018. “The Encyclopedia of DNA Elements (ENCODE): Data Portal Update.” *Nucleic Acids Research* 46 (D1): D794–801. https://doi.org/10.1093/nar/gkx1081.

De Santa, Francesca, Iros Barozzi, Flore Mietton, Serena Ghisletti, Sara Polletti, Betsabeh Khoramian Tusi, Heiko Muller, Jiannis Ragoussis, Chia-Lin Wei, and Gioacchino Natoli. 2010. “A Large Fraction of Extragenic RNA Pol II Transcription Sites Overlap Enhancers.” *PLoS Biology* 8 (5): e1000384. https://doi.org/10.1371/journal.pbio.1000384.

Demircioğlu, Deniz, Engin Cukuroglu, Martin Kindermans, Tannistha Nandi, Claudia Calabrese, Nuno A. Fonseca, André Kahles, et al. 2019. “A Pan-Cancer Transcriptome Analysis Reveals Pervasive Regulation through Alternative

Promoters." *Cell* 178 (6): 1465-1477.e17. https://doi.org/10.1016/j.cell.2019.08.018.

Dickel, Diane E., Athena R. Ypsilanti, Ramón Pla, Yiwen Zhu, Iros Barozzi, Brandon J. Mannion, Yupar S. Khin, et al. 2018. "Ultraconserved Enhancers Are Required for Normal Development." *Cell* 172 (3): 491-499.e15. https://doi.org/10.1016/j.cell.2017.12.017.

Djebali, Sarah, Carrie A. Davis, Angelika Merkel, Alex Dobin, Timo Lassmann, Ali Mortazavi, Andrea Tanzer, et al. 2012. "Landscape of Transcription in Human Cells." *Nature* 489 (7414): 101–8. https://doi.org/10.1038/nature11233.

Dong, Wei, Charikar Moses, and Kai Li. 2011. "Efficient K-Nearest Neighbor Graph Construction for Generic Similarity Measures." In *Proceedings of the 20th International Conference on World Wide Web - WWW '11*, 577. Hyderabad, India: ACM Press. https://doi.org/10.1145/1963405.1963487.

Dong, Xianjun, Zhixiang Liao, David Gritsch, Yavor Hadzhiev, Yunfei Bai, Joseph J. Locascio, Boris Guennewig, et al. 2018. "Enhancers Active in Dopamine Neurons Are a Primary Link between Genetic Variation and Neuropsychiatric Disease." *Nature Neuroscience* 21 (10): 1482–92. https://doi.org/10.1038/s41593-018-0223-0.

Dunham, Ian, Anshul Kundaje, Shelley F. Aldred, Patrick J. Collins, Carrie A. Davis, Francis Doyle, Charles B. Epstein, et al. 2012. "An Integrated Encyclopedia of DNA Elements in the Human Genome." *Nature* 489 (7414): 57–74. https://doi.org/10.1038/nature11247.

Farahany, Nita A., Henry T. Greely, Steven Hyman, Christof Koch, Christine Grady, Sergiu P. Pașca, Nenad Sestan, et al. 2018. "The Ethics of Experimenting with Human Brain Tissue." *Nature* 556 (7702): 429–32. https://doi.org/10.1038/d41586-018-04813-x.

Fernandes, Marie, Michelle Antoine, and Jean M. Hébert. 2012. "SMAD4 Is Essential for Generating Subtypes of Neurons during Cerebellar Development." *Developmental Biology* 365 (1): 82–90. https://doi.org/10.1016/j.ydbio.2012.02.017.

Fishilevich, Simon, Ron Nudel, Noa Rappaport, Rotem Hadar, Inbar Plaschkes, Tsippi Iny Stein, Naomi Rosen, et al. 2017. “GeneHancer: Genome-Wide Integration of Enhancers and Target Genes in GeneCards.” *Database* 2017 (January): bax028. https://doi.org/10.1093/database/bax028.

Flynn, Ryan A., Albert E. Almada, Jesse R. Zamudio, and Phillip A. Sharp. 2011. “Antisense RNA Polymerase II Divergent Transcripts Are P-TEFb Dependent and Substrates for the RNA Exosome.” *Proceedings of the National Academy of Sciences* 108 (26): 10460–65. https://doi.org/10.1073/pnas.1106630108.

Forrest, Alistair R. R., Hideya Kawaji, Michael Rehli, J. Kenneth Baillie, Michiel J. L. de Hoon, Vanja Haberle, Timo Lassmann, et al. 2014. “A Promoter-Level Mammalian Expression Atlas.” *Nature* 507 (7493): 462–70. https://doi.org/10.1038/nature13182.

Friedman, J., T. Hastie, and R. Tibshirani. 2008. “Sparse Inverse Covariance Estimation with the Graphical Lasso.” *Biostatistics* 9 (3): 432–41. https://doi.org/10.1093/biostatistics/kxm045.

Fulco, Charles P., Joseph Nasser, Thouis R. Jones, Glen Munson, Drew T. Bergman, Vidya Subramanian, Sharon R. Grossman, et al. 2019. “Activity-by-Contact Model of Enhancer–Promoter Regulation from Thousands of CRISPR Perturbations.” *Nature Genetics* 51 (12): 1664–69. https://doi.org/10.1038/s41588-019-0538-0.

Gao, Tianshun, and Jiang Qian. 2019. “EAGLE: An Algorithm That Utilizes a Small Number of Genomic Features to Predict Tissue/Cell Type-Specific Enhancer-Gene Interactions.” *PLOS Computational Biology* 15 (10): e1007436. https://doi.org/10.1371/journal.pcbi.1007436.

Nucleic Acids Research 48 (D1): D58–64. https://doi.org/10.1093/nar/gkz980.

Gorkin, David U., Iros Barozzi, Yuan Zhao, Yanxiao Zhang, Hui Huang, Ah Young Lee, Bin Li, et al. 2020. “An Atlas of Dynamic Chromatin Landscapes in Mouse Fetal Development.” *Nature* 583 (7818): 744–51. https://doi.org/10.1038/s41586-020-2093-3.

Guerrini, R., F. Moro, M. Kato, A. J. Barkovich, T. Shiihara, M. A. McShane, J. Hurst, et al. 2007. “Expansion of the First PolyA Tract of ARX Causes Infantile Spasms and

Status Dystonicus." *Neurology* 69 (5): 427–33. https://doi.org/10.1212/01.wnl.0000266594.16202.c1.

Hait, Tom Aharon, David Amar, Ron Shamir, and Ran Elkon. 2018. "FOCS: A Novel Method for Analyzing Enhancer and Gene Activity Patterns Infers an Extensive Enhancer–Promoter Map." *Genome Biology* 19 (1): 56. https://doi.org/10.1186/s13059-018-1432-2.

Hao, Yuhan, Stephanie Hao, Erica Andersen-Nissen, William M. Mauck, Shiwei Zheng, Andrew Butler, Maddie J. Lee, et al. 2021. "Integrated Analysis of Multimodal Single-Cell Data." *Cell* 184 (13): 3573-3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

Hariprakash, Judith Mary, and Francesco Ferrari. 2019. "Computational Biology Solutions to Identify Enhancers-Target Gene Pairs." *Computational and Structural Biotechnology Journal* 17 (June): 821–31. https://doi.org/10.1016/j.csbj.2019.06.012.

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. https://doi.org/10.1016/j.cell.2013.09.053.

Hodge, Rebecca D., Trygve E. Bakken, Jeremy A. Miller, Kimberly A. Smith, Eliza R. Barkan, Lucas T. Graybuck, Jennie L. Close, et al. 2019. "Conserved Cell Types with Divergent Features in Human versus Mouse Cortex." *Nature* 573 (7772): 61–68. https://doi.org/10.1038/s41586-019-1506-7.

Hoffman, Jackson A., Brian N. Papas, Kevin W. Trotter, and Trevor K. Archer. 2020. "Single-Cell RNA Sequencing Reveals a Heterogeneous Response to Glucocorticoids in Breast Cancer Cells." *Communications Biology* 3 (1): 1–11. https://doi.org/10.1038/s42003-020-0837-0.

Hon, Chung-Chau, Jordan A. Ramilowski, Jayson Harshbarger, Nicolas Bertin, Owen J. L. Rackham, Julian Gough, Elena Denisenko, et al. 2017. "An Atlas of Human Long Non-Coding RNAs with Accurate 5′ Ends." *Nature* 543 (7644): 199–204. https://doi.org/10.1038/nature21374.

Hong, Young Bin, Jaesoon Joo, Young Se Hyun, Geon Kwak, Yu-Ri Choi, Ha Kyung Yeo, Dong Hwan Jwa, et al. 2016. “A Mutation in PMP2 Causes Dominant Demyelinating Charcot-Marie-Tooth Neuropathy.” *PLoS Genetics* 12 (2): e1005829. https://doi.org/10.1371/journal.pgen.1005829.

Hou, Yue, Rongxin Zhang, and Xiao Sun. 2019. “Enhancer LncRNAs Influence Chromatin Interactions in Different Ways.” *Frontiers in Genetics* 10. https://www.frontiersin.org/article/10.3389/fgene.2019.00936.

Hsieh, Chen-Lin, Teng Fei, Yiwen Chen, Tiantian Li, Yanfei Gao, Xiaodong Wang, Tong Sun, et al. 2014. “Enhancer RNAs Participate in Androgen Receptor-Driven Looping That Selectively Enhances Gene Activation.” *Proceedings of the National Academy of Sciences* 111 (20): 7319–24. https://doi.org/10.1073/pnas.1324151111.

Hwang, Yih-Chii, Chiao-Feng Lin, Otto Valladares, John Malamon, Pavel P. Kuksa, Qi Zheng, Brian D. Gregory, and Li-San Wang. 2015. “HIPPIE: A High-Throughput Identification Pipeline for Promoter Interacting Enhancer Elements.” *Bioinformatics* 31 (8): 1290–92. https://doi.org/10.1093/bioinformatics/btu801.

Inoue, Fumitaka, and Nadav Ahituv. 2015. “Decoding Enhancers Using Massively Parallel Reporter Assays.” *Genomics*, Recent advances in functional assays of transcriptional enhancers, 106 (3): 159–64. https://doi.org/10.1016/j.ygeno.2015.06.005.

Jiang, Wei, and Liang Chen. 2020. “Alternative Splicing: Human Disease and Quantitative Analysis from High-Throughput Sequencing.” *Computational and Structural Biotechnology Journal* 19 (December): 183–95. https://doi.org/10.1016/j.csbj.2020.12.009.

Jing, Fang, Shao-Wu Zhang, and Shihua Zhang. 2020. “Prediction of Enhancer–Promoter Interactions Using the Cross-Cell Type Information and Domain Adversarial Neural Network.” *BMC Bioinformatics* 21 (1): 507. https://doi.org/10.1186/s12859-020-03844-4.

Kai, Yan, Bin E. Li, Ming Zhu, Grace Y. Li, Fei Chen, Yingli Han, Hye Ji Cha, et al. 2021. “Mapping the Evolving Landscape of Super-Enhancers during Cell

Differentiation." *Genome Biology* 22 (1): 269. https://doi.org/10.1186/s13059-021-02485-x.

Kannan, Meghna, Efil Bayam, Christel Wagner, Bruno Rinaldi, Perrine F. Kretz, Peggy Tilly, Marna Roos, et al. 2017. "WD40-Repeat 47, a Microtubule-Associated Protein, Is Essential for Brain Development and Autophagy." *Proceedings of the National Academy of Sciences* 114 (44): E9308–17. https://doi.org/10.1073/pnas.1713625114.

Kato, Mitsuhiro, Soma Das, Kristin Petras, Kunio Kitamura, Ken-ichirou Morohashi, Diane N. Abuelo, Mason Barr, et al. 2004. "Mutations of ARX Are Associated with Striking Pleiotropy and Consistent Genotype–Phenotype Correlation." *Human Mutation* 23 (2): 147–59. https://doi.org/10.1002/humu.10310.

Keen, Judy C., and Helen M. Moore. 2015. "The Genotype-Tissue Expression (GTEx) Project: Linking Clinical Data with Molecular Analysis to Advance Personalized Medicine." *Journal of Personalized Medicine* 5 (1): 22–29. https://doi.org/10.3390/jpm5010022.

Kent, W. James, Charles W. Sugnet, Terrence S. Furey, Krishna M. Roskin, Tom H. Pringle, Alan M. Zahler, and and David Haussler. 2002. "The Human Genome Browser at UCSC." *Genome Research* 12 (6): 996–1006. https://doi.org/10.1101/gr.229102.

Khan, Aziz, and Xuegong Zhang. 2016. "DbSUPER: A Database of Super-Enhancers in Mouse and Human Genome." *Nucleic Acids Research* 44 (D1): D164–71. https://doi.org/10.1093/nar/gkv1002.

Kim, Dong Won, Kai Liu, Zoe Qianyi Wang, Yi Stephanie Zhang, Abhijith Bathini, Matthew P. Brown, Sonia Hao Lin, et al. 2021. "Gene Regulatory Networks Controlling Differentiation, Survival, and Diversification of Hypothalamic Lhx6-Expressing GABAergic Neurons." *Communications Biology* 4 (1): 1–16. https://doi.org/10.1038/s42003-020-01616-7.

Kim, Kyukwang, Insu Jang, Mooyoung Kim, Jinhyuk Choi, Min-Seo Kim, Byungwook Lee, and Inkyung Jung. 2021. "3DIV Update for 2021: A Comprehensive Resource of 3D Genome and 3D Cancer Genome." *Nucleic Acids Research* 49 (D1): D38–46. https://doi.org/10.1093/nar/gkaa1078.

Kim, S., H. Cho, D. Lee, and M. J. Webster. 2012. “Association between SNPs and Gene Expression in Multiple Regions of the Human Brain.” *Translational Psychiatry* 2 (5): e113–e113. https://doi.org/10.1038/tp.2012.42.

Kim, Tae-Kyung, Martin Hemberg, Jesse M. Gray, Allen M. Costa, Daniel M. Bear, Jing Wu, David A. Harmin, et al. 2010. “Widespread Transcription at Neuronal Activity-Regulated Enhancers.” *Nature* 465 (7295): 182–87. https://doi.org/10.1038/nature09033.

Kitamura, Kunio, Yukiko Itou, Masako Yanazawa, Maki Ohsawa, Rika Suzuki-Migishima, Yuko Umeki, Hirohiko Hohjoh, et al. 2009. “Three Human ARX Mutations Cause the Lissencephaly-like and Mental Retardation with Epilepsy-like Pleiotropic Phenotypes in Mice.” *Human Molecular Genetics* 18 (19): 3708–24. https://doi.org/10.1093/hmg/ddp318.

Korsunsky, Ilya, Jean Fan, Kamil Slowikowski, Fan Zhang, Kevin Wei, Yuriy Baglaenko, Michael Brenner, Po-Ru Loh, and Soumya Raychaudhuri. 2018. “Fast, Sensitive, and Accurate Integration of Single Cell Data with Harmony.” bioRxiv. https://doi.org/10.1101/461954.

Kouno, Tsukasa, Jonathan Moody, Andrew Tae-Jun Kwon, Youtaro Shibayama, Sachi Kato, Yi Huang, Michael Böttcher, et al. 2019. “C1 CAGE Detects Transcription Start Sites and Enhancer Activity at Single-Cell Resolution.” *Nature Communications* 10 (1): 360. https://doi.org/10.1038/s41467-018-08126-5.

Kowalczyk, Monika S., Jim R. Hughes, David Garrick, Magnus D. Lynch, Jacqueline A. Sharpe, Jacqueline A. Sloane-Stanley, Simon J. McGowan, et al. 2012a. “Intragenic Enhancers Act as Alternative Promoters.” *Molecular Cell* 45 (4): 447–58. https://doi.org/10.1016/j.molcel.2011.12.021.

Kowalczyk, Monika S., Jim R. Hughes, David Garrick, Magnus D. Lynch, Jacqueline A. Sharpe, Jacqueline A. Sloane-Stanley, Simon J. McGowan, et al. 2012b. “Intragenic Enhancers Act as Alternative Promoters.” *Molecular Cell* 45 (4): 447–58. https://doi.org/10.1016/j.molcel.2011.12.021.

Lancaster, Madeline A., and Juergen A. Knoblich. 2014. “Generation of Cerebral Organoids from Human Pluripotent Stem Cells.” *Nature Protocols* 9 (10): 2329–40. https://doi.org/10.1038/nprot.2014.158.

Leeuw, Christiaan A. de, Joris M. Mooij, Tom Heskes, and Danielle Posthuma. 2015. “MAGMA: Generalized Gene-Set Analysis of GWAS Data.” *PLOS Computational Biology* 11 (4): e1004219. https://doi.org/10.1371/journal.pcbi.1004219.

Lewis, Michael W., Shen Li, and Hector L. Franco. 2019. “Transcriptional Control by Enhancers and Enhancer RNAs.” *Transcription* 10 (4–5): 171–86. https://doi.org/10.1080/21541264.2019.1695492.

Li, Kailong, Yuxuan Liu, Hui Cao, Yuannyu Zhang, Zhimin Gu, Xin Liu, Andy Yu, et al. 2020. “Interrogation of Enhancer Function by Enhancer-Targeting CRISPR Epigenetic Editing.” *Nature Communications* 11 (1): 485. https://doi.org/10.1038/s41467-020-14362-5.

Li, Mulin Jun, Zipeng Liu, Panwen Wang, Maria P. Wong, Matthew R. Nelson, Jean-Pierre A. Kocher, Meredith Yeager, et al. 2016. “GWASdb v2: An Update Database for Human Genetic Variants Identified by Genome-Wide Association Studies.” *Nucleic Acids Research* 44 (D1): D869–76. https://doi.org/10.1093/nar/gkv1317.

Li, Peipei, Lee Marshall, Gabriel Oh, Jennifer L. Jakubowski, Daniel Groot, Yu He, Ting Wang, Arturas Petronis, and Viviane Labrie. 2019. “Epigenetic Dysregulation of Enhancers in Neurons Is Associated with Alzheimer’s Disease Pathology and Cognitive Symptoms.” *Nature Communications* 10 (1): 2246. https://doi.org/10.1038/s41467-019-10101-7.

Li, Wenbo, Dimple Notani, Qi Ma, Bogdan Tanasa, Esperanza Nunez, Aaron Yun Chen, Daria Merkurjev, et al. 2013. “Functional Roles of Enhancer RNAs for Oestrogen-Dependent Transcriptional Activation.” *Nature* 498 (7455): 516–20. https://doi.org/10.1038/nature12210.

Li, Wenbo, Dimple Notani, and Michael G. Rosenfeld. 2016. “Enhancers as Non-Coding RNA Transcription Units: Recent Insights and Future Perspectives.” *Nature Reviews Genetics* 17 (4): 207–23. https://doi.org/10.1038/nrg.2016.4.

Li, Yang Eric, Sebastian Preissl, Xiaomeng Hou, Ziyang Zhang, Kai Zhang, Yunjiang Qiu, Olivier B. Poirion, et al. 2021. “An Atlas of Gene Regulatory Elements in Adult Mouse Cerebrum.” *Nature* 598 (7879): 129–36. https://doi.org/10.1038/s41586-021-03604-1.

Liu, Fangkun, Jing Huang, Liyang Zhang, Jindong Chen, Yu Zeng, Yongjian Tang, and Zhixiong Liu. 2019. “Advances in Cerebral Organoid Systems and Their Application in Disease Modeling.” *Neuroscience* 399 (February): 28–38. https://doi.org/10.1016/j.neuroscience.2018.12.013.

Liu, Shuai, Xinran Xu, Zhihao Yang, Xiaohan Zhao, and Wen Zhang. 2020. “EPIHC: Improving Enhancer-Promoter Interaction Prediction by Using Hybrid Features and Communicative Learning.” *ArXiv:2012.15418 [q-Bio]*, December. http://arxiv.org/abs/2012.15418.

Lubas, Michal, Peter Refsing Andersen, Aleks Schein, Andrzej Dziembowski, Grzegorz Kudla, and Torben Heick Jensen. 2015. “The Human Nuclear Exosome Targeting Complex Is Loaded onto Newly Synthesized RNA to Direct Early Ribonucleolysis.” *Cell Reports* 10 (2): 178–92. https://doi.org/10.1016/j.celrep.2014.12.026.

Luo, Chongyuan, Madeline A. Lancaster, Rosa Castanon, Joseph R. Nery, Juergen A. Knoblich, and Joseph R. Ecker. 2016. “Cerebral Organoids Recapitulate Epigenomic Signatures of the Human Fetal Brain.” *Cell Reports* 17 (12): 3369–84. https://doi.org/10.1016/j.celrep.2016.12.001.

Maere, Steven, Karel Heymans, and Martin Kuiper. 2005. “BiNGO: A Cytoscape Plugin to Assess Overrepresentation of Gene Ontology Categories in Biological Networks.” *Bioinformatics* 21 (16): 3448–49. https://doi.org/10.1093/bioinformatics/bti551.

Masland, Richard H. 2004. “Neuronal Cell Types.” *Current Biology* 14 (13): R497–500. https://doi.org/10.1016/j.cub.2004.06.035.

Matejuk, Agata, and Richard M. Ransohoff. 2020. “Crosstalk Between Astrocytes and Microglia: An Overview.” *Frontiers in Immunology* 11. https://www.frontiersin.org/article/10.3389/fimmu.2020.01416.

Matsui, Takeshi K., Yuichiro Tsuru, Koichi Hasegawa, and Ken-ichiro Kuwako. 2021. "Vascularization of Human Brain Organoids." *STEM CELLS* 39 (8): 1017–24. https://doi.org/10.1002/stem.3368.

Mattioli, Kaia, Winona Oliveros, Chiara Gerhardinger, Daniel Andergassen, Philipp G. Maass, John L. Rinn, and Marta Melé. 2020. "Cis and Trans Effects Differentially Contribute to the Evolution of Promoters and Enhancers." *Genome Biology* 21 (1): 210. https://doi.org/10.1186/s13059-020-02110-3.

McInnes, Leland, John Healy, and James Melville. 2020. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction." *ArXiv:1802.03426 [Cs, Stat]*, September. http://arxiv.org/abs/1802.03426.

Meacham, Corbin E., and Sean J. Morrison. 2013. "Tumor Heterogeneity and Cancer Cell Plasticity." *Nature* 501 (7467): 328–37. https://doi.org/10.1038/nature12624.

Mercola, M., X. F. Wang, J. Olsen, and K. Calame. 1983. "Transcriptional Enhancer Elements in the Mouse Immunoglobulin Heavy Chain Locus." *Science (New York, N.Y.)* 221 (4611): 663–65. https://doi.org/10.1126/science.6306772.

Michida, Hiroki, Hiroaki Imoto, Hisaaki Shinohara, Noriko Yumoto, Masahide Seki, Mana Umeda, Tetsutaro Hayashi, et al. 2020. "The Number of Transcription Factors at an Enhancer Determines Switch-like Gene Expression." *Cell Reports* 31 (9): 107724. https://doi.org/10.1016/j.celrep.2020.107724.

Misra, Sima, Madeline A. Crosby, Christopher J. Mungall, Beverley B. Matthews, Kathryn S. Campbell, Pavel Hradecky, Yanmei Huang, et al. 2002. "Annotation of the Drosophila Melanogastereuchromatic Genome: A Systematic Review." *Genome Biology* 3 (12): research0083.1. https://doi.org/10.1186/gb-2002-3-12-research0083.

Monzel, Anna S., Lisa M. Smits, Kathrin Hemmer, Siham Hachi, Edinson Lucumi Moreno, Thea van Wuellen, Javier Jarazo, et al. 2017. "Derivation of Human Midbrain-Specific Organoids from Neuroepithelial Stem Cells." *Stem Cell Reports* 8 (5): 1144–54. https://doi.org/10.1016/j.stemcr.2017.03.010.

Moody, Jonathan, Tsukasa Kouno, Akari Suzuki, Youtaro Shibayama, Chikashi Terao, Jen-Chien Chang, Fernando López-Redondo, et al. 2021. "Profiling of Transcribed Cis-

Regulatory Elements in Single Cells." bioRxiv. https://doi.org/10.1101/2021.04.04.438388.

Muguruma, Keiko, Ayaka Nishiyama, Hideshi Kawakami, Kouichi Hashimoto, and Yoshiki Sasai. 2015. "Self-Organization of Polarized Cerebellar Tissue in 3D Culture of Human Pluripotent Stem Cells." *Cell Reports* 10 (4): 537–50. https://doi.org/10.1016/j.celrep.2014.12.051.

Müller, Hans-Peter, and Walter Schaffner. 1990. "Transcriptional Enhancers Can Act in Trans." *Trends in Genetics* 6 (January): 300–304. https://doi.org/10.1016/0168-9525(90)90236-Y.

Nord, Alex S., Matthew J. Blow, Catia Attanasio, Jennifer A. Akiyama, Amy Holt, Roya Hosseini, Sengthavy Phouanenavong, et al. 2013. "Rapid and Pervasive Changes in Genome-Wide Enhancer Usage during Mammalian Development." *Cell* 155 (7): 1521–31. https://doi.org/10.1016/j.cell.2013.11.033.

Nott, Alexi, Inge R. Holtman, Nicole G. Coufal, Johannes C. M. Schlachetzki, Miao Yu, Rong Hu, Claudia Z. Han, et al. 2019. "Brain Cell Type–Specific Enhancer–Promoter Interactome Maps and Disease-Risk Association." *Science* 366 (6469): 1134–39. https://doi.org/10.1126/science.aay0793.

Ozone, Chikafumi, Hidetaka Suga, Mototsugu Eiraku, Taisuke Kadoshima, Shigenobu Yonemura, Nozomu Takata, Yutaka Oiso, Takashi Tsuji, and Yoshiki Sasai. 2016. "Functional Anterior Pituitary Generated in Self-Organizing Culture of Human Embryonic Stem Cells." *Nature Communications* 7 (1): 10351. https://doi.org/10.1038/ncomms10351.

Palstra, Robert-Jan. 2012. "Transcription Factor Binding at Enhancers: Shaping a Genomic Regulatory Landscape in Flux." *Frontiers in Genetics* 3. https://www.frontiersin.org/article/10.3389/fgene.2012.00195.

Pantera, Harrison, John J Moran, Holly A Hung, Evgenia Pak, Amalia Dutra, and John Svaren. 2018. "Regulation of the Neuropathy-Associated Pmp22 Gene by a Distal Super-Enhancer." *Human Molecular Genetics* 27 (16): 2830–39. https://doi.org/10.1093/hmg/ddy191.

Pennacchio, Len A., Nadav Ahituv, Alan M. Moses, Shyam Prabhakar, Marcelo A. Nobrega, Malak Shoukry, Simon Minovitsky, et al. 2006. “In Vivo Enhancer Analysis of Human Conserved Non-Coding Sequences.” *Nature* 444 (7118): 499–502. https://doi.org/10.1038/nature05295.

Perenthaler, Elena, Soheil Yousefi, Eva Niggl, and Tahsin Stefan Barakat. 2019. “Beyond the Exome: The Non-Coding Genome and Enhancers in Neurodevelopmental Disorders and Malformations of Cortical Development.” *Frontiers in Cellular Neuroscience* 13. https://www.frontiersin.org/article/10.3389/fncel.2019.00352.

Piñero, Janet, Juan Manuel Ramírez-Anguita, Josep Saüch-Pitarch, Francesco Ronzano, Emilio Centeno, Ferran Sanz, and Laura I Furlong. 2020. “The DisGeNET Knowledge Platform for Disease Genomics: 2019 Update.” *Nucleic Acids Research* 48 (D1): D845–55. https://doi.org/10.1093/nar/gkz1021.

Piovesan, Allison, Francesca Antonaros, Lorenza Vitale, Pierluigi Strippoli, Maria Chiara Pelleri, and Maria Caracausi. 2019. “Human Protein-Coding Genes and Gene Feature Statistics in 2019.” *BMC Research Notes* 12 (1): 315. https://doi.org/10.1186/s13104-019-4343-8.

Pistollato, Francesca, David Canovas-Jorda, Dimitra Zagoura, and Anna Price. 2017. “Protocol for the Differentiation of Human Induced Pluripotent Stem Cells into Mixed Cultures of Neurons and Glia for Neurotoxicity Testing.” *Journal of Visualized Experiments : JoVE*, no. 124 (June): 55702. https://doi.org/10.3791/55702.

Pliner, Hannah A., Jonathan S. Packer, José L. McFaline-Figueroa, Darren A. Cusanovich, Riza M. Daza, Delasa Aghamirzaie, Sanjay Srivatsan, et al. 2018. “Cicero Predicts Cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data.” *Molecular Cell* 71 (5): 858-871.e8. https://doi.org/10.1016/j.molcel.2018.06.044.

Pnueli, Lilach, Sergei Rudnizky, Yahav Yosefzon, and Philippa Melamed. 2015. “RNA Transcribed from a Distal Enhancer Is Required for Activating the Chromatin at the Promoter of the Gonadotropin α-Subunit Gene.” *Proceedings of the National*

Academy of Sciences 112 (14): 4369–74. https://doi.org/10.1073/pnas.1414841112.

Polioudakis, Damon, Luis de la Torre-Ubieta, Justin Langerman, Andrew G. Elkins, Xu Shi, Jason L. Stein, Celine K. Vuong, et al. 2019. "A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-Gestation." *Neuron* 103 (5): 785-801.e8. https://doi.org/10.1016/j.neuron.2019.06.011.

Purcell, Shaun, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A. R. Ferreira, David Bender, Julian Maller, et al. 2007. "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses." *The American Journal of Human Genetics* 81 (3): 559–75. https://doi.org/10.1086/519795.

Qian, Xuyu, Ha Nam Nguyen, Mingxi M. Song, Christopher Hadiono, Sarah C. Ogden, Christy Hammack, Bing Yao, et al. 2016. "Brain-Region-Specific Organoids Using Mini-Bioreactors for Modeling ZIKV Exposure." *Cell* 165 (5): 1238–54. https://doi.org/10.1016/j.cell.2016.04.032.

Qian, Xuyu, Hongjun Song, and Guo-li Ming. 2019. "Brain Organoids: Advances, Applications and Challenges." *Development* 146 (8): dev166074. https://doi.org/10.1242/dev.166074.

Qiu, Xiaojie, Qi Mao, Ying Tang, Li Wang, Raghav Chawla, Hannah A. Pliner, and Cole Trapnell. 2017. "Reversed Graph Embedding Resolves Complex Single-Cell Trajectories." *Nature Methods* 14 (10): 979–82. https://doi.org/10.1038/nmeth.4402.

Quinlan, Aaron R., and Ira M. Hall. 2010. "BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features." *Bioinformatics* 26 (6): 841–42. https://doi.org/10.1093/bioinformatics/btq033.

Rajpurkar, Aparna R., Leslie J. Mateo, Sedona E. Murphy, and Alistair N. Boettiger. 2021. "Deep Learning Connects DNA Traces to Transcription to Reveal Predictive Features beyond Enhancer–Promoter Contact." *Nature Communications* 12 (1): 3423. https://doi.org/10.1038/s41467-021-23831-4.

Reilly, Steven K., Jun Yin, Albert E. Ayoub, Deena Emera, Jing Leng, Justin Cotney, Richard Sarro, Pasko Rakic, and James P. Noonan. 2015. "Evolutionary Changes in

Promoter and Enhancer Activity during Human Corticogenesis." *Science*, March. https://doi.org/10.1126/science.1260943.

Rhee, Ho Sung, Michael Closser, Yuchun Guo, Elizaveta V. Bashkirova, G. Christopher Tan, David K. Gifford, and Hynek Wichterle. 2016. "Expression of Terminal Effector Genes in Mammalian Neurons Is Maintained by a Dynamic Relay of Transient Enhancers." *Neuron* 92 (6): 1252–65. https://doi.org/10.1016/j.neuron.2016.11.037.

Rivera, Chloe M., and Bing Ren. 2013. "Mapping Human Epigenomes." *Cell* 155 (1): 39–55. https://doi.org/10.1016/j.cell.2013.09.011.

Sakaguchi, Hideya, Taisuke Kadoshima, Mika Soen, Nobuhiro Narii, Yoshihito Ishida, Masatoshi Ohgushi, Jun Takahashi, Mototsugu Eiraku, and Yoshiki Sasai. 2015. "Generation of Functional Hippocampal Neurons from Self-Organizing Human Embryonic Stem Cell-Derived Dorsomedial Telencephalic Tissue." *Nature Communications* 6 (November): 8896. https://doi.org/10.1038/ncomms9896.

Salzberg, Steven L. 2018. "Open Questions: How Many Genes Do We Have?" *BMC Biology* 16 (August): 94. https://doi.org/10.1186/s12915-018-0564-x.

Sanes, Joshua R., and Richard H. Masland. 2015. "The Types of Retinal Ganglion Cells: Current Status and Implications for Neuronal Classification." *Annual Review of Neuroscience* 38 (1): 221–46. https://doi.org/10.1146/annurev-neuro-071714-034120.

Santa, Francesca De, Iros Barozzi, Flore Mietton, Serena Ghisletti, Sara Polletti, Betsabeh Khoramian Tusi, Heiko Muller, Jiannis Ragoussis, Chia-Lin Wei, and Gioacchino Natoli. 2010. "A Large Fraction of Extragenic RNA Pol II Transcription Sites Overlap Enhancers." *PLOS Biology* 8 (5): e1000384. https://doi.org/10.1371/journal.pbio.1000384.

Shannon, Paul, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. 2003. "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks." *Genome Research* 13 (11): 2498–2504. https://doi.org/10.1101/gr.1239303.

Shimojo, Daisuke, Kazunari Onodera, Yukiko Doi-Torii, Yasuharu Ishihara, Chinatsu Hattori, Yukino Miwa, Satoshi Tanaka, et al. 2015. “Rapid, Efficient, and Simple Motor Neuron Differentiation from Human Pluripotent Stem Cells.” *Molecular Brain* 8 (1): 79. https://doi.org/10.1186/s13041-015-0172-4.

Shiraki, Toshiyuki, Shinji Kondo, Shintaro Katayama, Kazunori Waki, Takeya Kasukawa, Hideya Kawaji, Rimantas Kodzius, et al. 2003. “Cap Analysis Gene Expression for High-Throughput Analysis of Transcriptional Starting Point and Identification of Promoter Usage.” *Proceedings of the National Academy of Sciences* 100 (26): 15776–81. https://doi.org/10.1073/pnas.2136655100.

Silva, Bárbara da, Ryan K. Mathew, Euan S. Polson, Jennifer Williams, and Heiko Wurdak. 2018. “Spontaneous Glioblastoma Spheroid Infiltration of Early-Stage Cerebral Organoids Models Brain Tumor Invasion.” *SLAS DISCOVERY: Advancing the Science of Drug Discovery* 23 (8): 862–68. https://doi.org/10.1177/2472555218764623.

Singer, Gregory AC, Jiejun Wu, Pearlly Yan, Christoph Plass, Tim HM Huang, and Ramana V. Davuluri. 2008. “Genome-Wide Analysis of Alternative Promoters of Human Genes Using a Custom Promoter Tiling Array.” *BMC Genomics* 9 (1): 349. https://doi.org/10.1186/1471-2164-9-349.

Smoot, Michael E., Keiichiro Ono, Johannes Ruscheinski, Peng-Liang Wang, and Trey Ideker. 2011. “Cytoscape 2.8: New Features for Data Integration and Network Visualization.” *Bioinformatics* 27 (3): 431–32. https://doi.org/10.1093/bioinformatics/btq675.

Soldner, Frank, Yonatan Stelzer, Chikdu S. Shivalila, Brian J. Abraham, Jeanne C. Latourelle, M. Inmaculada Barrasa, Johanna Goldmann, Richard H. Myers, Richard A. Young, and Rudolf Jaenisch. 2016. “Parkinson-Associated Risk Variant in Distal Enhancer of α-Synuclein Modulates Target Gene Expression.” *Nature* 533 (7601): 95–99. https://doi.org/10.1038/nature17939.

Sozen, Berna, Gianluca Amadei, Andy Cox, Ran Wang, Ellen Na, Sylwia Czukiewska, Lia Chappell, et al. 2018. “Self-Assembly of Embryonic and Two Extra-Embryonic

Stem Cell Types into Gastrulating Embryo-like Structures." *Nature Cell Biology* 20 (8): 979–89. https://doi.org/10.1038/s41556-018-0147-7.

Spitz, François, and Eileen E. M. Furlong. 2012. "Transcription Factors: From Enhancer Binding to Developmental Control." *Nature Reviews Genetics* 13 (9): 613–26. https://doi.org/10.1038/nrg3207.

Stelzer, Gil, Naomi Rosen, Inbar Plaschkes, Shahar Zimmerman, Michal Twik, Simon Fishilevich, Tsippi Iny Stein, et al. 2016. "The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses." *Current Protocols in Bioinformatics* 54 (1): 1.30.1-1.30.33. https://doi.org/10.1002/cpbi.5.

Stiles, Joan, and Terry L. Jernigan. 2010a. "The Basics of Brain Development." *Neuropsychology Review* 20 (4): 327–48. https://doi.org/10.1007/s11065-010-9148-4.

Strømme, Petter, Marie E. Mangelsdorf, Ingrid E. Scheffer, and Jozef Gécz. 2002. "Infantile Spasms, Dystonia, and Other X-Linked Phenotypes Caused by Mutations in Aristaless Related Homeobox Gene, ARX." *Brain and Development* 24 (5): 266–68. https://doi.org/10.1016/S0387-7604(02)00079-7.

Su, Ying, Ashim Subedee, Noga Bloushtain-Qimron, Virginia Savova, Marcin Krzystanek, Lewyn Li, Andriy Marusyk, et al. 2015. "Somatic Cell Fusions Reveal Extensive Heterogeneity in Basal-like Breast Cancer." *Cell Reports* 11 (10): 1549–63. https://doi.org/10.1016/j.celrep.2015.05.011.

Summers, Kim M, Stephen J Bush, Chunlei Wu, and David A Hume. 2022. "Generation and Network Analysis of an RNA-Seq Transcriptional Atlas for the Rat." *NAR Genomics and Bioinformatics* 4 (1): lqac017. https://doi.org/10.1093/nargab/lqac017.

Suzuki, Harukazu, Alistair R R Forrest, Erik van Nimwegen, Carsten O Daub, Piotr J Balwierz, Katharine M Irvine, Timo Lassmann, et al. 2009. "The Transcriptional Network That Controls Growth Arrest and Differentiation in a Human Myeloid Leukemia Cell Line." *Nature Genetics* 41 (5): 553–62. https://doi.org/10.1038/ng.375.

Svensson, Valentine, Eduardo da Veiga Beltrame, and Lior Pachter. 2020. “A Curated Database Reveals Trends in Single-Cell Transcriptomics.” *Database: The Journal of Biological Databases and Curation* 2020 (November): baaa073. https://doi.org/10.1093/database/baaa073.

Swanson, Olivia K., and Arianna Maffei. 2019. “From Hiring to Firing: Activation of Inhibitory Neurons and Their Recruitment in Behavior.” *Frontiers in Molecular Neuroscience* 12. https://www.frontiersin.org/article/10.3389/fnmol.2019.00168.

Syed, Khaja Mohieddin, and Chung-Chau Hon. 2021. “Heterogeneity among Enhancer RNAs: Origins, Consequences and Perspectives.” *Essays in Biochemistry* 65 (4): 709–21. https://doi.org/10.1042/EBC20200064.

Takahashi, Kazutoshi, Koji Tanabe, Mari Ohnuki, Megumi Narita, Tomoko Ichisaka, Kiichiro Tomoda, and Shinya Yamanaka. 2007. “Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors.” *Cell* 131 (5): 861–72. https://doi.org/10.1016/j.cell.2007.11.019.

Takahashi, Kazutoshi, and Shinya Yamanaka. 2006. “Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors.” *Cell* 126 (4): 663–76. https://doi.org/10.1016/j.cell.2006.07.024.

Tang, Fuchou, Catalin Barbacioru, Yangzhou Wang, Ellen Nordman, Clarence Lee, Nanlan Xu, Xiaohui Wang, et al. 2009. “MRNA-Seq Whole-Transcriptome Analysis of a Single Cell.” *Nature Methods* 6 (5): 377–82. https://doi.org/10.1038/nmeth.1315.

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. https://doi.org/10.1038/nature11232.

Tierney, Adrienne L., and Charles A. Nelson. 2009. “Brain Development and the Role of Experience in the Early Years.” *Zero to Three* 30 (2): 9–13.

Traag, V. A., L. Waltman, and N. J. van Eck. 2019. “From Louvain to Leiden: Guaranteeing Well-Connected Communities.” *Scientific Reports* 9 (1): 5233. https://doi.org/10.1038/s41598-019-41695-z.

Trapnell, Cole, Davide Cacchiarelli, Jonna Grimsby, Prapti Pokharel, Shuqiang Li, Michael Morse, Niall J. Lennon, Kenneth J. Livak, Tarjei S. Mikkelsen, and John L. Rinn. 2014. “The Dynamics and Regulators of Cell Fate Decisions Are Revealed by Pseudotemporal Ordering of Single Cells.” *Nature Biotechnology* 32 (4): 381–86. https://doi.org/10.1038/nbt.2859.

Tyssowski, Kelsey M., Nicholas R. DeStefino, Jin-Hyung Cho, Carissa J. Dunn, Robert G. Poston, Crista E. Carty, Richard D. Jones, et al. 2018. “Different Neuronal Activity Patterns Induce Different Gene Expression Programs.” *Neuron* 98 (3): 530-546.e11. https://doi.org/10.1016/j.neuron.2018.04.001.

Visel, Axel, Matthew J. Blow, Zirong Li, Tao Zhang, Jennifer A. Akiyama, Amy Holt, Ingrid Plajzer-Frick, et al. 2009. “ChIP-Seq Accurately Predicts Tissue-Specific Activity of Enhancers.” *Nature* 457 (7231): 854–58. https://doi.org/10.1038/nature07730.

Walters, M C, W Magis, S Fiering, J Eidemiller, D Scalzo, M Groudine, and D I Martin. 1996. “Transcriptional Enhancers Act in Cis to Suppress Position-Effect Variegation.” *Genes & Development* 10 (2): 185–95. https://doi.org/10.1101/gad.10.2.185.

Wang, Dong, Ivan Garcia-Bassets, Chris Benner, Wenbo Li, Xue Su, Yiming Zhou, Jinsong Qiu, et al. 2011. “Reprogramming Transcription by Distinct Classes of Enhancers Functionally Defined by ERNA.” *Nature* 474 (7351): 390–94. https://doi.org/10.1038/nature10006.

Wang, Saidi, Haiyan Hu, and Xiaoman Li. 2020. “Hundreds of Motif Pairs May Facilitate Enhancer-Promoter Interactions.” bioRxiv. https://doi.org/10.1101/2020.12.29.424656.

Weber, Blaise, Johan Zicola, Rurika Oka, and Maike Stam. 2016. “Plant Enhancers: A Call for Discovery.” *Trends in Plant Science* 21 (11): 974–87. https://doi.org/10.1016/j.tplants.2016.07.013.

Whyte, Warren A., David A. Orlando, Denes Hnisz, Brian J. Abraham, Charles Y. Lin, Michael H. Kagey, Peter B. Rahl, Tong Ihn Lee, and Richard A. Young. 2013. "Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes." *Cell* 153 (2): 307–19. https://doi.org/10.1016/j.cell.2013.03.035.

Wolfe, Jareth C., Liudmila A. Mikheeva, Hani Hagras, and Nicolae Radu Zabet. 2021. "An Explainable Artificial Intelligence Approach for Decoding the Enhancer Histone Modifications Code and Identification of Novel Enhancers in Drosophila." *Genome Biology* 22 (1): 308. https://doi.org/10.1186/s13059-021-02532-7.

Wu, Xiaolei, Qi Shen, Zhan Zhang, Di Zhang, Ying Gu, and Da Xing. 2021. "Photoactivation of TGFβ/SMAD Signaling Pathway Ameliorates Adult Hippocampal Neurogenesis in Alzheimer's Disease Model." *Stem Cell Research & Therapy* 12 (1): 345. https://doi.org/10.1186/s13287-021-02399-2.

Xie, Shiqi, Jialei Duan, Boxun Li, Pei Zhou, and Gary C. Hon. 2017. "Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells." *Molecular Cell* 66 (2): 285-299.e5. https://doi.org/10.1016/j.molcel.2017.03.007.

Yang, Yang, Ruochi Zhang, Shashank Singh, and Jian Ma. 2017. "Exploiting Sequence-Based Features for Predicting Enhancer–Promoter Interactions." *Bioinformatics* 33 (14): i252–60. https://doi.org/10.1093/bioinformatics/btx257.

Yang, Yue, Jin Xu, Shuyu Ge, and Liqin Lai. 2021. "CRISPR/Cas: Advances, Limitations, and Applications for Precision Cancer Research." *Frontiers in Medicine* 8. https://www.frontiersin.org/article/10.3389/fmed.2021.649896.

Yip, Chi Wai, Chung-Chau Hon, Kayoko Yasuzawa, Divya M. Sivaraman, Jordan A. Ramilowski, Youtaro Shibayama, Saumya Agrawal, et al. 2022. "The High-Throughput Perturbation of Long Non-Coding RNA Reveals Functional Features in Stem Cells and across Cell-Types." bioRxiv. https://doi.org/10.1101/2022.02.16.480297.

Yousefi, Soheil, Ruizhi Deng, Kristina Lanko, Eva Medico Salsench, Anita Nikoncuk, Herma C. van der Linde, Elena Perenthaler, Tjakko J. van Ham, Eskeatnaf Mulugeta, and Tahsin Stefan Barakat. 2021. "Comprehensive Multi-Omics

Integration Identifies Differentially Active Enhancers during Human Brain Development with Clinical Relevance." *Genome Medicine* 13 (1): 162. https://doi.org/10.1186/s13073-021-00980-1.

Zeng, Hongkui, and Joshua R. Sanes. 2017. "Neuronal Cell-Type Classification: Challenges, Opportunities and the Path Forward." *Nature Reviews Neuroscience* 18 (9): 530–46. https://doi.org/10.1038/nrn.2017.85.

Zhang, Boya, Yangzhige He, Yanpeng Xu, Fan Mo, Tingwei Mi, Qing Sunny Shen, Chunfeng Li, et al. 2018. "Differential Antiviral Immunity to Japanese Encephalitis Virus in Developing Cortical Organoids." *Cell Death & Disease* 9 (7): 1–10. https://doi.org/10.1038/s41419-018-0763-y.

Zhao, Changyong, Xiaoman Li, and Haiyan Hu. 2016. "PETModule: A Motif Module Based Approach for Enhancer Target Gene Prediction." *Scientific Reports* 6 (1): 30043. https://doi.org/10.1038/srep30043.

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): 14049. https://doi.org/10.1038/ncomms14049.

Zhou, Qiling, Miao Yu, Roberto Tirado-Magallanes, Bin Li, Lingshi Kong, Mingrui Guo, Zi Hui Tan, et al. 2021. "ZNF143 Mediates CTCF-Bound Promoter–Enhancer Loops Required for Murine Hematopoietic Stem and Progenitor Cell Function." *Nature Communications* 12 (1): 43. https://doi.org/10.1038/s41467-020-20282-1.

Zhu, Yun, Lin Sun, Zhao Chen, John W. Whitaker, Tao Wang, and Wei Wang. 2013. "Predicting Enhancer Transcription and Activity from Chromatin Modifications." *Nucleic Acids Research* 41 (22): 10032–43. https://doi.org/10.1093/nar/gkt826.

8 Acknowledgments

First, I would like to thank Professor Michael Detmar for giving me the chance to pursue a Ph.D. in his group. Without him, this work would never have been possible. His support and advice led my scientific path until I learned to follow my ideas.

Next, I would like to thank Dr. Jay W. Shin. His patience and encouragement were key success factors for my Ph.D. He supported me throughout the whole Ph.D. process and made this journey unforgettable. I am very grateful for the hours of discussions, exploring new ideas, and helping me to see the bigger picture of the project.

My special thanks go to Dr. Joachim Luginbühl and Dr. Chi Wai Yip for their guidance and thoughtful pieces of advice. Thank you for embracing all my mistakes with patience and helping me design better experiments. They showed me the fun side of science.

Whenever work was challenging, my lab members gave me the environment and support to learn. My lab members taught me some of my most valuable experiences. My deepest thank for their continuous support in the lab. I am incredibly thankful to Dr. Chung Chau Hon, Dr. Juan Ortiz Quinonez, Dr. Jordan Ramilowski, and Dr. Jonathan Moody for their advice on operating R and the long mails of troubleshooting error messages. I would also like to thank Dr. Julio Jesus Leon Incio, Dr. Anika Prabhu, and Dr. Pauline Robbe for their encouragement and advice on how to nail your Ph.D. Finally, I am very grateful to Kouno-san, who always supported me in the lab with technical advice. Without you guys, this incredible journey of challenges would not have been as exciting as it was. Thank you for all the discussions and fun hours in the FANTOM bar.

Next, I would like to thank my thesis committee for their advice and scientific guidance.

I arrived in Japan with Nicola Hetherington and Wing Hin Yip at the same time. These two became my buddies in crime. I want to thank them for their continuous belief in my skills and cheering me up.

I also want to thank Andrew Kwon, who calmed me down during challenging times and was always there to provide feedback on my problems.

And last, I want to thank my family for always believing in the best version of me. Especially my siblings kept reminding me to look at the bright side of life and challenge yesterday to become a better person today.

9 Curriculum Vitae

Yan Jun LAN, M. Sc.
Chome-7-22 Suehirocho
Tsurumi Ward, Yokohama
Kanagawa 230-0045
Japan

+ 81-70-4431-5689
yanjun.lan@riken.jp

Education

2018 – 2022	**Ph.D. candidate in Pharmaceutical Science**, ETH, Switzerland "Discovery of functional enhancer-promoter pairs using brain organoids"
2015 – 2017	**M. Sc. in Pharmaceutical Science**, University of Basel, Switzerland "Generation of a CLDN5-GFP reporter in Human Embryonic Stem Cells for Modeling of the Blood-Retinal-Barrier" (best grade)
2011 – 2015	**B. Sc. in Pharmaceutical Science,** University of Freiburg, Germany "The Influence of Chloroquine on the Transfection Efficiency of DC30® Lipoplexes in MDCK II Cells" (best grade)

Professional experience

2018 – 2022	**International Program Associate**, RIKEN, Yokohama, Japan
2017	**Research Assistant,** University of São Paulo, Brazil
2017	**Summer Intern,** CiRA, Kyoto, Japan
2016 – 2017	**Intern,** Roche, Basel, Switzerland
2013 – 2014	**Research Assistant,** University of Freiburg, Freiburg, Germany
2011	**Research Assistant,** University of Zurich, Zurich, Switzerland

Scholarships

2018 – 2022	International Program Associate Scholarship (4 years, total 65000 USD)
2019	RIKEN - HUGO scholarship award "Best Poster" Travel Grant (800 USD)

Conferences

2020	**IMS-Stanford ICSBRM Symposium, online** Oral presentation "Decoding the Regulatory Logic behind Brain Development and Neural Diversification using cerebral organoids."

2020	**Human Cell Atlas, online** Oral presentation "Gene-regulatory network during brain development using brain organoids"
2019	**Human Genome Meeting, Seoul, Korea** Oral presentation "5'-Single Cell RNA-seq Reveals Non-coding Regulatory Elements Associated in Human Brain Disorders."
2018	**Human Cell Atlas, Jeju Island, Korea** Oral presentation "Non-coding regulatory elements in human brain organoids"
2017	**StemBANCC Science & 10th SC MEETING, Basel, Switzerland** Oral presentation "Modeling of severe metabolic, genetic diseases in vitro using genome editing and human pluripotent stem cell differentiation to vascular endothelial cells"
2017	**Translational Opportunities in Stem Cell Research ISSCR, Basel, Switzerland** Poster presentation "Development of retinal endothelial cell in vitro model with high-resistance barrier using endothelial cells derived from human pluripotent stem cells"

Publications

Lan, Yan Jun, Joachim Luginbühl, Kouno Tsukasa , Jonathan Moody, Chung-Chau Hon, Michael Detmar, Jay W. Shin. "Discovery of functional enhancer-promoter pairs in brain organoids." *in progress*

Yip, Chi Wai, Chung-Chau Hon, Kayoko Yasuzawa, Divya M. Sivaraman, Jordan A. Ramilowski, Youtaro Shibayama, Saumya Agrawal, Anika V.Prabhu, Callum Parr, Jessica Severin, **Yan Jun Lan**, Josée Dostie, Hiromi Nishiyori-Sueki, Michihira Tagami, Masayoshi Itoh, Fernando López-Redondo, Tsukasa Kouno, Jen-Chien Chang, Joachim Luginbühl, Masaki Kato, Mitsuyoshi Murata, Wing Hin Yip, Xufeng Shu, Imad Abugessaisa, Akira Hasegawa, Harukazu Suzuki, Ken Yagi, Takeya Kasukawa, Michiel de Hoon, Piero Carninci, Jay W. Shin. 2022. "The High-Throughput Perturbation of Long Non-Coding RNA Reveals Functional Features in Stem Cells and across Cell-Types." bioRxiv. https://doi.org/10.1101/2022.02.16.480297.

Sabate-Soler, Sonia, Sarah Louise Nickels, Cláudia Saraiva, Emanuel Berger, Ugne Dubonyte, Kyriaki Barmpa, **Yan Jun Lan**, Tsukasa Kouno, Javier Jarazo, Graham Robertson, Jafar Sharif, Haruhiko Koseki, Christian Thome, Jay W. Shin, Sally A. Cowley, Jens C. Schwamborn. 2022. "Microglia Integration into Human Midbrain Organoids Leads to Increased Neuronal Maturation and Functionality." *Glia* 70 (7): 1267–88. https://doi.org/10.1002/glia.24167.

Eura, Nobuyuki, Takeshi K. Matsui, Joachim Luginbühl, Masaya Matsubayashi, Hitoki Nanaura, Tomo Shiota, Kaoru Kinugawa, Naohiko Iguchi, Takao Kiriyama, Canbin Zheng, Tsukasa Kouno, **Yan Jun Lan**, Pornparn Kongpracha, Pattama Wiriyasermkul, Yoshihiko M. Sakaguchi, Riko Nagata, Tomoya Komeda, Naritaka Morikawa, Fumika Kitayoshi, Miyong Jong, Shinko Kobashigawa, Mari Nakanishi, Masatoshi Hasegawa, Yasuhiko Saito, Takashi Shiromizu, Yuhei Nishimura, Takahiko Kasai, Maiko Takeda, Hiroshi Kobayashi, Yusuke Inagaki, Yasuhito Tanaka, Manabu Makinodan, Toshifumi Kishimoto, Hiroki Kuniyasu, Shushi Nagamori, Alysson R. Muotri, Jay W. Shin, Kazuma Sugie, Eiichiro Mori. 2020. "Brainstem Organoids from Human Pluripotent Stem Cells." *Frontiers in Neuroscience* 14. https://www.frontiersin.org/article/10.3389/fnins.2020.00538.

Roudnicky, Filip, Bo Kyoung Kim, **Yanjun Lan**, Roland Schmucki, Verena Küppers, Klaus Christensen, Martin Graf, Christoph Patsch, Mark Burcin, Claas Aiko Meyer, Peter D Westenskow, Chad A. Cowan. 2020. "Identification of a Combination of Transcription Factors That Synergistically Increases Endothelial Cell Barrier Resistance." *Scientific Reports* 10 (1): 3886. https://doi.org/10.1038/s41598-020-60688-x.

Roudnicky, Filip, **Yanjun Lan**, Max Friesen, Gregor Dernick, Jitao David Zhang, Andreas Staempfli, Natalie Bordag, Antje Wagner-Golbs, Klaus Christensen, Martin Ebeling, Martin Graf, Mark Burcin, Claas Aiko Meyer, Chad A. Cowan, Christoph Patsch. 2019. "Modeling the Effects of Severe Metabolic Disease by Genome Editing of HPSC-Derived Endothelial Cells Reveals an Inflammatory Phenotype." *International Journal of Molecular Sciences* 20 (24): 6201. https://doi.org/10.3390/ijms20246201.

Roudnicky, Filip, Jitao David Zhang, Bo Kyoung Kim, Nikhil J. Pandya, **Yanjun Lan,** Lisa Sach-Peltason, Heloise Ragelle, Pamela Strassburger, Sabine Gruener, Mirjana Lazendic, Sabine Uhles, Franco Revelant, Oliv Eidam, Gregor Sturm, Verena Kueppers, Klaus Christensen, Leonard D. Goldstein, Manuel Tzouros, Balazs Banfai, Zora Modrusan, Martin Graf, Christoph Patsch, Mark Burcin, Claas A. Meyer, Peter D. Westenskow, and Chad A. Cowan. 2020. "Inducers of the Endothelial Cell Barrier Identified through Chemogenomic Screening in Genome-Edited HPSC-Endothelial Cells." *Proceedings of the National Academy of Sciences* 117 (33): 19854–65. https://doi.org/10.1073/pnas.1911532117.

Leonidova, Anna, Vanessa Pierroz, Riccardo Rubbiani, **Yanjun Lan**, Anita G. Schmitz, Andres Kaech, Roland K. O. Sigel, Stefano Ferrari, and Gilles Gasser. 2014. "Photo-Induced Uncaging of a Specific Re(I) Organometallic Complex in Living Cells." *Chemical Science* 5 (10): 4044–56. https://doi.org/10.1039/C3SC53550A.