Aus dem Helmholtz Pioneer Campus
Helmholtz-Zentrum München

Dissertation

zum Erwerb des Doctor of Philosophy (Ph.D.)

an der Medizinischen Fakultät der

Ludwig-Maximilians-Universität München

Development and use of multiomics tools to dissect drivers of neuronal identity

vorgelegt von:

Jeisimhan Diwakar Shunmugapriya

aus:

Palanichettipatti / India

Jahr:

2025

Mit Genehmigung der Medizinischen Fakultät der
Ludwig-Maximilians-Universität München

Erstes Gutachten von:	Dr. Boyan Bonev
Zweites Gutachten von:	Prof. Dr. Gunnar Schotta
Drittes Gutachten von:	Prof. Dr. Axel Imhof
Viertes Gutachtes:	Priv. Doz. Dr. Lisa Ann Gerdes

Dekan:	**Prof. Dr. med. Thomas Gudermann**

Datum der Verteidigung:

24.06.2025

Table of Contents

Summary

Cell-type identities and development are guided by epigenetic mechanisms that modify chromatin to regulate gene expression without altering the DNA sequence. This process depends on transcription factors (TFs) to regulate RNA expression programs by affecting cis-regulatory elements and the chromatin landscape, all within the context of 3D nuclear architecture. Until recently, studies focused mainly on a few epigenetic layers, often neglecting the 3D genome architecture. Yet, high-throughput technologies like Hi-C have provided valuable insights into how 3D chromatin folding influences development. Therefore, in this PhD, we apply and develop multiomics assays that include chromosome conformation technologies to mechanistically understand how lineage-specifying factors dynamically rewire multiple epigenetic layers.

We initiated our study by examining the epigenome remodelling involved in the direct neuronal reprogramming of astrocytes into induced neurons, facilitated by the overexpression of the proneural transcription factor Neurogenin2 (Ngn2) or its phosphorylation-resistant variant (PmutNgn2). Through the integration of single-cell multiomics and Methyl-HiC, we revealed that Ngn2 drives extensive multilayered epigenetic rewiring. Induction with PmutNgn2 resulted in the faster generation of more mature neurons, accompanied by enhanced chromatin remodelling. Interestingly, this effect was not due to superior pioneering activity but rather to the activation of downstream genes that act as co-factors. Among these, we identified Yy1 as a critical Ngn2-recruited co-factor, whose depletion impaired reprogramming efficiency.

To explore how cellular and epigenetic contexts influence TF mediated rewiring, we overexpressed Ngn2 in mouse embryonic stem cells and neural progenitor cells. By integrating 3DRAM-seq with ChIP-mass spectrometry, we found that whilst Ngn2's direct activity at bound sites was largely consistent across both cell types, there were pronounced cell-type-specific indirect effects on the global epigenome, most notably in embryonic stem cells. These distinct effects appear to be modulated by Ngn2 interactors, which included subunits of the SWI/SNF and NuRD chromatin remodelling complexes.

Recognising the intricate interplay between multiple epigenetic layers, we initially aimed to create a method for simultaneously profiling the 3D genome and transcription at single-cell resolution with high throughput. However, given the emergence of several similar methods in recent studies, we expanded upon existing techniques to develop sc-3DRAM-seq, capable of additionally measuring DNA methylation and chromatin accessibility. Bulk quality control experiments have shown promising results, and we are now conducting single-cell tests. The primary goal of this work is to uncover epigenetic variation within the heterogeneous tissues of mouse embryonic brains and human fetal brains.

List of Figures

List of abbreviations

- TF – Transcription Factor
- TSS – Transcription Start Site
- DBD – DNA-Binding Domain
- bHLH – basic-helix-loop-helix
- Ngn2/Neurog2 – Neurogenin2
- PmutNgn2 – Phospmutant Neurogenein2
- iN – induced Neurons
- DEGs – Differentially expressed genes
- EB – Embryoid Bodies
- ES – mouse embryonic stem cells
- NPC – Neural Progenitor cells
- HiT-seq – Hi-C + Transcription sequencing
- Sc-3DRAM-seq – single cell 3D genome, RNA, Accessibility and Methylation sequencing
- NGS – Next generation sequencing

1. Introduction

1.1 Transcription Factors and Epigenetics

Eukaryotic DNA is chromatinised with several layers of epigenetic control. Features include DNA methylation[1], chromatin accessibility[2], histone modifications and 3D genome organisation[3,4]. A synergistic modulation of these multiple layers play a key role in regulating the ability of promoters[5] and cis-regulatory elements located further away from the transcriptional start site (TSS) to initiate and maintain transcriptional profiles.

TFs are a class of proteins that recognise and bind a distinct DNA sequence[6] and regulate gene expression. They achieve this by recruiting co-factors that may alter chromatin states, which facilitates the recruitment of the preinitiation complex (PIC), the Mediator complex, and RNA polymerase II (RNA Pol II) to gene promoters[7,8]. TFs typically contain a DNA-binding domain (DBD), one or more transcriptional activation or repression domains, and often a dimerization domain along with protein-protein interaction domains[9]. The structural features of TFs, co-factors, and the surrounding epigenetic landscape enable the protein to exhibit a highly selective, often 1,000-fold greater, preference for specific binding sites over other sequences[10,11].

Given the complexity required for accurate regulation genome-wide, eukaryotes host a diverse array of transcription factor families, each defined by distinct structural motifs, including C2H2 zinc finger (ZF), homeodomain (HD), basic helix-loop-helix (bHLH), basic leucine zipper (bZIP), and nuclear hormone receptor (NHR)[12]. The mechanism of these TFs can vary significantly depending on their specific family and function.

1.1.1 Pioneer Transcription Factors

One important mechanistic distinction of TFs is the ability or inability to facilitate the opening of inaccessible nucleosomal DNA.

The concept of pioneer TFs first emerged in the late 1970s following the identification of differentially accessible chromatinised regions through techniques such as DNase hyper-sensitivity[13]. Direct experimental evidence then followed with *in vitro* studies showing that FOXA and GATA4 can bind nucleosome arrays compacted with linker histones[14], unlike other TFs such as NF1, C/EBP, and GAL4-AH. The binding of FOXA and GATA4 led to the creation of local nuclease-sensitive sites in the middle of the nucleosome array. This new family of TFs were termed ‘pioneer’ factors, indicative of their initial binding occurring prior to other downstream effects[15].

Unlike most TFs, which can only bind to their target sequences on exposed DNA, the ability to bind nucleosomes stood out. Given that active regulatory regions are already known to be more accessible in chromatin than other sequences[16,17], cell fate changes requires pioneer factors that can bind even when the DNA is shielded in closed chromatin.

The contextual importance of pioneer TFs was identified using DNA footprinting experiments. *In vivo* footprinting of the albumin enhancer in precursor gut endoderm revealed that the transcriptionally inactive gene is marked by the initial binding of FOXA leading to subsequent gene expression and lineage specification[18]. Other TFs with similar properties have since been identified in diverse contexts, including zygotic gene activation[19,20], direct cellular reprogramming[21,22] and neuronal development[23,24].

Although it was previously argued that the distinction between canonical TFs and pioneer TFs is a continuum[25], more recent biophysical and structural studies have clearly shown that pioneer TFs can differentially interact with nucleosome core particles, whereas canonical TFs do it poorly[26]. For example, using cryo-EM, it was shown that the helix-loop-helix domain of a budding yeast pioneer TF Cbf1 interacts with histones H2A and H2B, with the acidic part of the Cbf1 HLH region likely positioned to interact with the highly basic N-terminal H3 tail. These interactions with histones seem to stabilize Cbf1 binding within the nucleosome, leading to significantly slower dissociation rates, subsequent nucleosome eviction and creation of a nucleosome depleted region[26]. Additionally, single-molecule tracking microscopy assays revealed that pioneer TFs exhibit short residence times on closed chromatin during their search for target sequences, whereas non-pioneer TFs tend to avoid closed chromatin entirely[27,28].

To better understand how different pioneer transcription factors (TFs) interact with nucleosomes, a technique called Nucleosome CAP–SELEX was developed. In this method, DNA is reconstituted into nucleosomes, which are then incubated with TFs. After incubation, the bound DNA is isolated using PCR. This approach was used to investigate the binding preferences of 220 TFs on nucleosomal DNA, uncovering five main interaction patterns: 1) binding across both gyres of the nucleosomal DNA, 2) preference for orientation, 3) preference for binding ends, 4) periodic binding, and 5) higher affinity for the dyad region. Most pioneer TFs were found to induce nucleosome dissociation, likely forming a ternary complex upon binding. This complex is thought to be unstable because TFs have a stronger binding affinity for free DNA than for nucleosomal DNA, which generates the free energy needed to drive nucleosome dissociation and regulate gene expression [29].

Interestingly, pioneer TFs are also characterised by their ability to alter the nucleosome structure *in vitro* in an ATPase-independent manner, such as in the case of Foxa which displaces the linker histone H1[30]. However, other studies have shown that using *in vitro* chromatin assembled with *Drosophila* embryo extracts, nucleosome disruption by the pioneer factor GAF depends on ATP hydrolysis via the Nucleosome-remodeling factor (NURF) complex for complete remodelling[31]. Similarly *in vivo*, complete opening of the chromatin does still require ATP-dependent remodellers[32,33], such as through the direct interaction with different subunits of SWI/SNF complexes[34]. The open domains subsequently created can further stabilise the binding of the pioneer TFs to the chromatin due to the greater binding affinity for the remodelled DNA[29] and support the recruitment of additional TFs for regulation.

1.2 Gene regulation and pioneer TFs

Spatiotemporal control of gene expression requires the co-ordination and interplay amongst various gene regulatory elements, with focus on promoters and enhancers as activating regulatory elements. While promoters are located proximal to the transcription start site (TSS) and serve as the assembly site for the transcriptional machinery, cis-regulatory elements such as enhancers can be located up to 1Mb away [35] from the core promoter and play a crucial role in modulating when and how transcription is initiated[36]. Enhancers can function both synergistically and uniquely across various cell types to regulate gene expression specificity. Investigating how cell-type-specific TFs activate enhancers and can transmit key signals over long distances remains a critical area of research in uncovering the mechanisms underlying gene regulation.

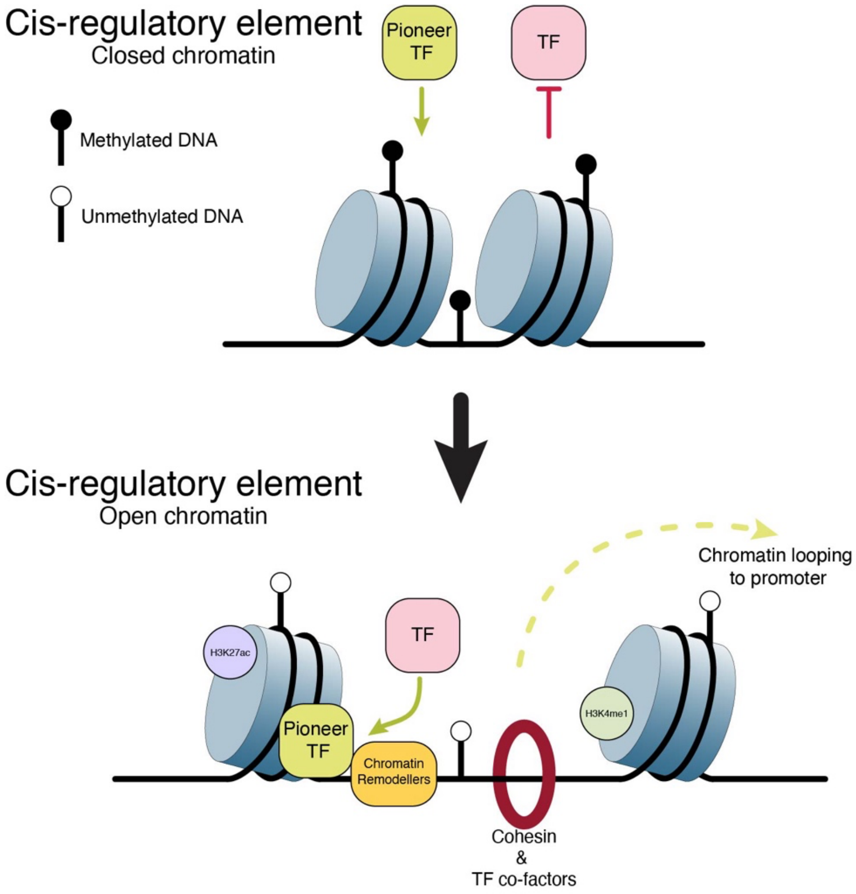

Figure 1 Pioneer factor action at cis-regulatory elements

Pioneer factors can bind closed chromatin by directly associating with nucleosomes, inducing chromatin opening, recruitment of remodellers and other TFs, histone modifications, DNA demethylation and initiating chromatin looping. Adapted from Barral and Zaret, Trends in Genetics, 2024

1.2.1 DNA accessibility

Enhancers contain multiple DNA binding sites that can be recognised in a sequence-specific manner by pioneer TFs. The subsequent nucleosome depletion through the recruitment of ATPase-dependent remodellers creates an open chromatin conformation (Fig. 1)– a common signature of active enhancers, initially identified through techniques such as DNase-seq, FAIRE-seq[37] and now more commonly ATAC-seq[38]. This permits the recruitment of additional machinery and TFs, that further support enhancer function.

1.2.2 DNA Methylation

In cell types where enhancers are active, the underlying DNA sequences typically have reduced levels of DNA methylation. Whereas when inactive, the sequence is often methylated[39,40]. DNA methylation in mammals is an epigenetic mark whereby a methyl group is transferred from metabolites (SAM) onto the C5 (5mC) position on the DNA base cytosine. This mark is often associated with gene repression through recruitment of methyl-binding domain (MBD) proteins, many of which are part of larger repressive chromatin complexes (indirect model), or by preventing methylation sensitive TFs from binding the DNA (direct model)[41,42]. The most common context for cytosine methylation is found within CpG dinucleotides[43,44], which are often clustered in regions known as CpG islands and are present in 70% of annotated mammalian promoters[45], found often unmethylated. Enhancers can also contain CpG islands, often termed orphan CpG islands[46].

Binding of DNA sequences by pioneers is often accompanied by DNA demethylation (Fig. 1)[47]. This can be achieved through the direct recruitment of Ten-eleven translocation (TET) enzymes to enhancers[48] that partially demethylate enhancers to a characteristic intermediate 10%–50% methylation[49]. However, if this demethylation is necessary for subsequent enhancer activation across the genome remains contentious, given that the *in vivo* response of methyl-sensitive TFs upon 5mC perturbation can drastically vary from one context to the other and the redundancy amongst some enhancers in regulating the promoter. For example, CTCF, considered a methylation sensitive TF, upon DNA methylation perturbation >98.5%, the many unoccupied and previously methylated CTCF recognition sequences remain unbound[50], with the differential sensitivity unable to be explained by differences in their recognition motifs[51]. Therefore, 5mC is unlikely to function as a simple on/off switch for enhancer activity; but rather modulating the degree of enhancer activity by influencing TF binding dependent on genomic context. This subtle modulation is likely sufficient to instigate the necessary activity to direct developmental trajectories[49]. Evidence for this can be seen in disease states where alteration of enhancer DNA methylation levels can lead to disruption in development, as with hematopoietic differentiation[52].

1.2.3 Histone Modifications

The remodelled nucleosomes in enhancers typically have two main histone marks that characterise their active nature: H3K4me1 and H3K27ac (Fig. 1). H3K4me1 is placed by the histone methyltransferase MLL3/4[53]. Both the histone mark and the methyltransferase subsequently recruit BRG1/BAF which function to maintain the chromatin open[54]. H3K4me1 also recruits p300 and CBP acetyltransferases which then establishes H3K27ac[55].

H3K4me1 marked enhancers when tested through reporter assays, displayed some activity, however, a significant proportion were not active[56]. It has been hypothesised that this histone mark is more associated with a poised state, as seen with many inactive developmental enhancers enriched with H3K4me1[57].

H3K27ac at enhancers often correlates well with target gene expression[58], however varying evidence exists for if the mark is essential for enhancer activation. In mESCs, where the lysine on the histone variant H3.3 is mutated to an arginine, the transcriptome is largely unaltered. This could be likely due to the lack of change in chromatin accessibility, suggesting H3K27ac alone is not sufficient for activation[59]. An alternative study that chemically blocked CBP/P300-mediated H3K27ac in in MOLM-16 acute myeloid leukemia (AML) cells showed a drastic reduction in enhancer activity due to decrease of p-TEFb and Pol II occupancy and supressed production of eRNAs, leading to loss of gene expression[60]. Due to current discrepancies, the field is also in search for alternative histone marks that are more tightly linked to activity of the enhancer[61].

Nevertheless, it has been suggested that these histone modifications may contribute to the recruitment of other TFs and contribute to the co-operativity amongst some pioneer factors. For example, structural studies have indicated that a modified H3K27 can assist the addition of OCT4 and SOX2 to nucleosomal internal sites following initial binding of OCT4[62]. Yet, a similar additive cooperativity is not observed with FOXA2 and GATA4[63].

1.2.4 Chromatin architecture

Although enhancers and promoters are often separated by considerable linear distances, these cis-regulatory elements can still influence gene activation. This interaction is facilitated by long-range communication, which is enabled by a dynamic 3D chromatin structure[64]. Despite the capacity of enhancers to activate various promoters, their specificity is tightly regulated by topological restrictions and TF mediated targeting. Elucidating the mechanisms leading to specificity and interactions between enhancers and their cognate promoters has been advanced through the development of both microscopy[65,66] and chromatin conformation capture techniques[67].

Whilst promoters are often accessible across tissue types, partially due to the abundant nature of CpG islands and the associated demethylated status, their corresponding gene expression is not ubiquitous[68,69]. Whereas, enhancer activation and subsequent gene expression have been linked to increased proximity between the enhancer and the promoter (Fig. 1), as shown by both DNA and RNA FISH techniques such as ORCA[65], and more recently simultaneous detection of 3D genome contacts and gene expression through multiomics assays[70] . This has also led to the development of models such as Activity by Contact (ABC)[71] to understand gene expression.

Contradicting evidence for the necessity of proximity between enhancers and promoters for gene expression have also been proposed. Microscopy studies of particular loci such as the Sox2 region have shown that upon induced expression, the distance between the enhancer and promoter increases[72–74]. This has been attributed to an increase in local protein concentration as a result of an accumulation in transcription machinery and TFs in the region[75]. Other studies have also suggested that proximity is not sufficient for initiation, and additional regulation is required[76].It remains currently unclear if this is exemplary of all transcription initiation events, or limited to the examples studied.

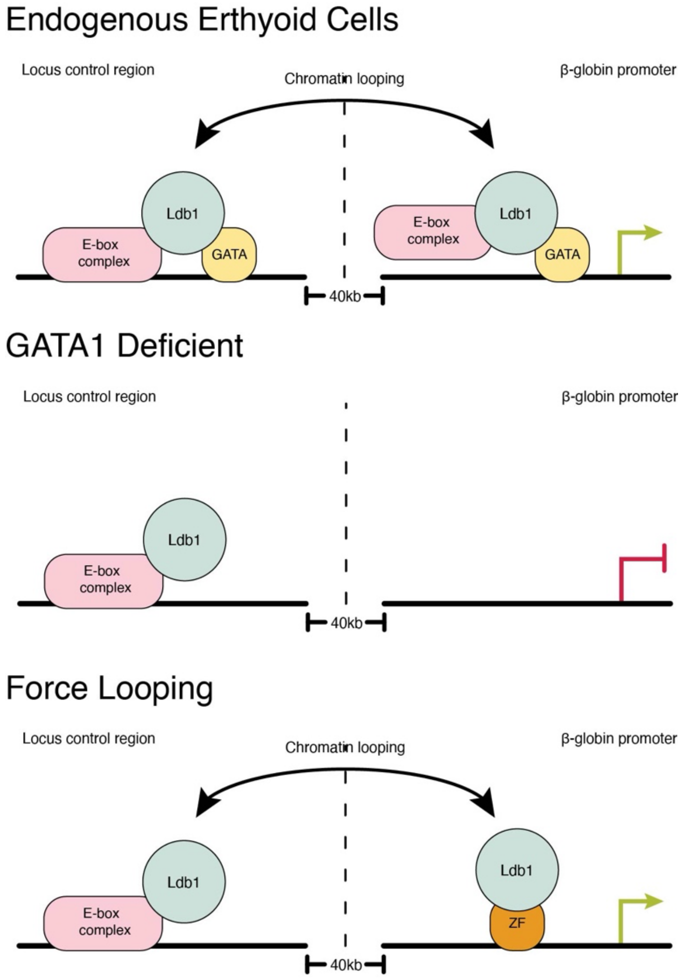

Figure 2 ZF-Ldb1 mediated looping to the B-globin locus.

Top – Wild type scenario where GATA1 and E-box complex recruit Ldb1 to mediate looping. Middle – No GATA1 results in no Ldb1 at the promoter, impaired looping and reduced transcription. Bottom- Ectopic recruitment of Ldb1 by ZF-mediated tethering rescues looping and transcription. Adapted from Deng et al., Cell,2012

Yet, in both models, it is generally agreed that the initiation is facilitated by TFs due to their interaction with architectural proteins like cohesin, the recruitment of co-factors and transcription machinery such as Mediators to facilitate the looping, or the oligomerisation of TFs[77]. The first evidence that TFs/co-factors are capable of mediating looping and henceforth initiating transcription came from early studies in erythroid cells. In this system, tethering the co-factor Ldb1 to the β-globin promoter activates transcription by forming a chromatin loop with the LCR, even without its cognate TF, GATA1(Fig. 2)[78]. Pioneer factors have also shown similar tendencies, such as TCF1, which regulates CTCF binding on enhancers and hence looping[79]. Whereas other factors such as OCT4 and Nanog have been suggested to form protein aggregates to facilitate contact formation[80,81].

1.3 Proneural Transcription Factors

Neurons constitute the most diverse cell population in any organism[82]. The generation of the diverse neural lineages and precise regulation of progenitor proliferation and neuronal differentiation are guided by proneural pioneer TFs belonging to the bHLH family. Proneural proteins primarily function in progenitor cells, but they can also remain transiently expressed in postmitotic neurons, where they play a role in regulating cell migration as well as axonal and dendritic growth[83–85]. Understanding how proneural TFs precisely regulate multiple epigenetic layers to maintain developmental control has recently advanced due to the emergence and prominence of new multiomics tools.

The first studies of proneural factors were performed in *Drosophila* mutant models. Macrochaetes in *Drosophila* are bristle-like structures that function as sensory organs positioned precisely on the thorax and head. In 1916, the first mutant was discovered in a fly lacking some bristles. As more mutations were identified, it became clear that different alleles affected distinct sets of bristles[86]. Subsequent molecular analysis led to the identification of the four genes involved in this process: *achaete (ac), scute (sc), lethal of scute (lsc)* and *asense (ase)*[87]. These proteins were identified to share the structural motif, coined the bHLH, which shows a 70% sequence identity within the *asc* protein family[82]. This discovery prompted further research into the motif's functions, revealing its role in binding to regulatory DNA sequences in enhancers and promoters (referred to as E-box regions) and its capacity for dimerization[88]. Later, a second family of proneural genes were identified consisting of *atonal (ato)*[89], *amos (absent MD neurons and olfactory sensilla)*[90] and *cato (cousin of atonal)*[91], sharing a 45% identity of the bHLH motif.

1.3.1 bHLH motif

The helix-loop-helix (HLH) domain is located at the N-terminus and comprises 40–50 basic amino acid residues. The domain consists of two α-helices joined by a non-conserved loop region that primarily facilitates dimerization, whilst the basic region facilitates DNA binding to the consensus E-box motif CANNTG[92].

Heterodimeric complexes are formed with broadly expressed E proteins, which are encoded by the *Drosophila* gene *da* or one of three mammalian genes: *E2A* (producing E12 and E47), *HEB*, and *E2-2*. Crystal structures of the bHLH domains in proteins like MyoD, Max, and E47 reveal that these dimers form through interactions between two helices in each partner, creating a four-helix bundle[82].

The variability in the E-box sequence, along with the dimerization of various bHLH proteins, plays a key role in regulating a wide range of developmental functions through transcriptional control. According to their expression patterns, bHLH transcription factors are categorized into Class I and Class II groups with Class II bHLH factors characterized by tissue-specific expression[92].

The binding of bHLH proteins to nucleosomal DNA depends on both the positioning of the E-box and the dimerization domain, which together influence histone interactions, binding affinity, and cooperative interactions with other nucleosome-bound factors[93].

Most of these DNA-contacting residues are highly conserved across neural bHLH proteins such as Neurogenin2, enabling binding to the core E-box sequence, although this does not fully explain the DNA-binding specificity and distinct roles seen among different neural bHLH families. Analysis of E-box sequences in various target genes suggests specificity beyond the conserved CA and TG bases, possibly due to interactions between family-specific residues and co-factors[94].

1.3.2 Neurogenin2 – A proneural pioneer factor

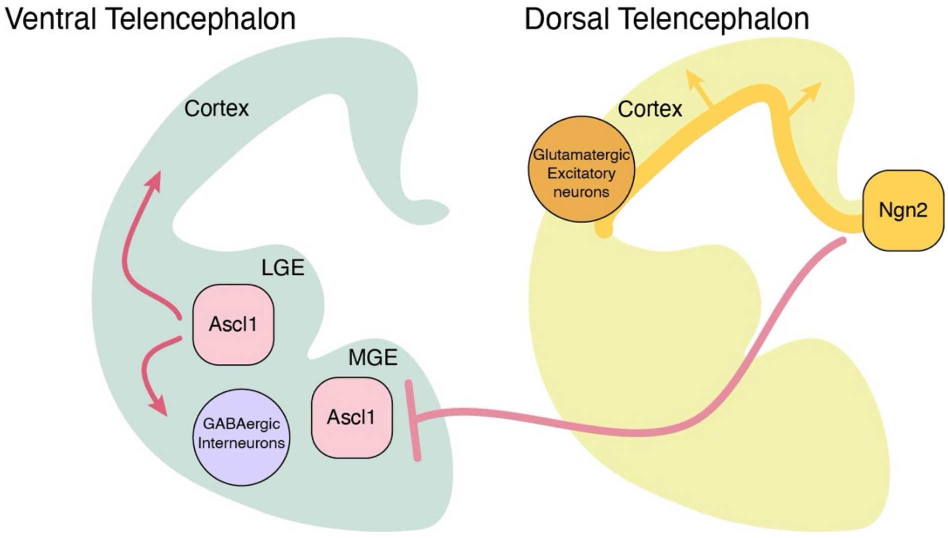

Figure 3 Glutamatergic and GABAergic neuronal origin in the rodent telencephalon.

In the ventral telencephalon, Ascl1 is expressed in the lateral and medial ganglionic eminences (LGE/MGE), specifying the generation of GABAergic interneurons. In the dorsal telencephalon, cortical progenitors induce neocortical development and expression of Ngn2 generating glutamatergic excitatory neurons. A regulatory genetic switch is found between Ngn2 and Ascl1 expression. Adapted from Lee et al. International Journal of Molecular Sciences, 2022

The mammalian neocortex is a six-layered structure in the brain that is responsible for cognitive function, sensory perception and consciousness, and has significantly expanded through evolution. This complex structure consists of hundreds of neuronal and glial subtypes. Cortical neurons are generally classified into two main categories: interneurons and projection neurons. Interneurons predominantly contain GABA (GABA is an inhibitory neurotransmitter) and are generated from progenitors in the ventral telencephalon and function in an inhibitory manner. In contrast, projection neurons are glutamatergic (glutamate is an excitatory neurotransmitter) in nature resulting in an excitatory behaviour and are generated from progenitors in the dorsolateral wall of the telencephalon[95] (Fig. 3).

In neocortical development, the neurons are generated in a specific sequence; multipotent progenitors in the dorsal telencephalon first produce neurons for the cortical preplate, then for the lower layers (V/VI), and finally for the upper layers (II–IV) of the cortical plate[96]. Three main progenitor types contribute to this process: neuroepithelial cells, radial glia and intermediate progenitors. Through a combination of initial symmetric divisions to increase the pool of progenitor cells, followed by asymmetric divisions to generate early neurons, a diverse population of neurons are generated based on temporal and regional identity[97].

In vertebrates, numerous genes related to the *Drosophila* proneural families, *asc* and *ato,* have been identified. In mammals, Ascl1 (previously Mash1) and Neurogenin2 (Ngn2) serve as the primary homologs of the *Drosophila* proneural genes *achaete-scute complex* and *atonal*, respectively[82,98]. In neural progenitor cells, Ascl1 initiates the cell-cycle exit of progenitors, and contributes to their differentiation into mainly GABAergic neurons[99]. Whereas, in the dorsal telencephalon, cortical progenitors transiently express Ngn2 in the ventricular zone producing glutamatergic neurons (Fig. 3)[92].

Proneural bHLH TFs such as Ascl1 and Ngn2 are often only transiently expressed following cortical territory specification by three homeodomain (HD) transcription factors, Lhx2, Emx2 and Pax6, that act either alone or in combination[100,101]. In turn they contribute to inducing other bHLH factors in a “bHLH cascade”[102]. Co-ordinating the positional and timely expression of these TFs is therefore critical. For example, Ngn2 and Ascl1 are thought to form a genetic switch, and when Ngn2 is turned off, Ascl1 is turned on (Fig. 3). As a result, in Ngn2 null cortices, the upregulated expression of Ascl1, results in the misspecification of early-born neurons to an abnormal identity of GABAergic neurons[96,103].

Ngn2 is a pioneer TF[104], considered a master proneural factor, that is essential and sufficient to specify glutamatergic identity of neurons and is expressed in progenitor cells throughout the neurogenic period(E10.5 to E17.5)[105]. However, loss of function experiments have indicated that continued expression of Ngn2 is only needed for specifying deep layer glutamatergic neurons born before E14.5[103], indicating modulation of Ngn2 expression and activity depends on multiple factors.

Three key regulatory influences of Ngn2 include: feedback loops, phosphorylation status and interaction partners. In neural stem cells, Hes1, a transcriptional repressor oscillates in expression autonomously by a negative feedback loop. This oscillation affects expression levels of other factors, including Ngn2, resulting in an inverse oscillation pattern relative to Hes1[106]. These oscillations contribute to the efficient proliferation of the neural stem cells[107]. Upon transition to intermediate progenitor cells, Tbr2 downregulates Hes1, resulting in a stable increase in Ngn2 and a transition to neurogenic gene expression[108,109].

Ngn2 is considered a highly unstable protein that is phosphorylated in a developmental and cell-cycle regulated manner at conserved serine/threonine residues adjacent to prolines (SP/TP sites) at loop-helix 2 region of the bHlH domain[110]. Whilst phosphorylation at these sites does not have a direct effect on the stability of the protein, it contributes to increased neuronal differentiation activity i*n vivo*[111]. During the transition from early to late neocortical development, there is a drastic decrease in Wnt signalling, contributing to increased activity of the proline-directed serine/threonine protein kinase GSK3 by releasing the previously membrane bound endosomes[112]. The increased activity of GSK3 leads to increased phosphorylation of Ngn2 at specific GSK3 phosphoacceptor sites and correlates with a reduction in Ngn2's proneural activity. Through manipulation of GSK3 levels it has been shown that its activation suppresses the proneural activity of exogenous Ngn2 in early cortical progenitors, while inhibiting GSK3 at later stages triggers early neurogenesis. This decrease in proneural activity is attributed to preventing Ngn2 from homodimerizing, instead forming heterodimers with bHLH cofactors E12/E47, highlighting the importance of interaction partners in modulating activity[113].

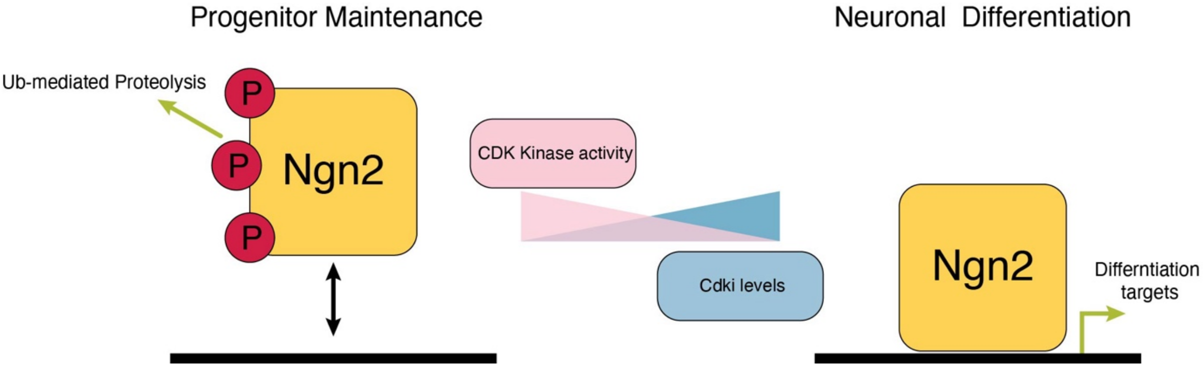

Figure 4 Ngn2 early developmental control by phosphorylation.

Ngn2 phosphorylation (P) occurs during rapid progenitor cell cycles targeting it for ubiquitin (Ub) mediated proteolysis. As the cell cycle lengthens, un(der)phosphorylated Ngn2 accumulates, increasing promoter binding and activating downstream differentiation genes. Adapted from Ali et al., Development, 2011

Conversely, Ngn2 phosphorylation by cell cycle dependent kinases (cdk) in early development during progenitor maintenance prevents Ngn2 from targeting downstream expression of neuronal differentiation genes. Given Ngn2 protein levels are regulated by ubiquitin mediated proteolysis, cdk dependent phosphorylation during rapid progenitor cell cycles targets the protein for degradation. Upon cell-cycle lengthening and accumulation of cdk inhibitors, the reduced phosphorylation of Ngn2 was shown to activate genes involved in differentiation (Fig. 4). A drive towards neuronal differentiation was also shown with functional experiments using a phosphomutant of Ngn2 that cannot be phosphorylated at key cdk target sites[114]. Although the corresponding epigenetic changes were investigated in this study, it was reasoned that by preventing phosphorylation of Ngn2 at key sites, the protein turnover is reduced and the increased stability contributes to longer binding of the TF to E-box sequences, driving genes particular to neuronal differentiation.

Given the wide-ranging effects of Ngn2, its precise regulation is crucial in accurate neocortical development. Understanding how Ngn2 and its variants affect downstream gene regulation will offer valuable insights into its roles in development, disease, neuronal reprogramming, and may also shed light on mechanisms of other proneural TFs. By utilising multiomics tools, we address these questions in the projects tackled in this PhD.

1.4 Proneural TF induced direct neuronal reprogramming

Cellular reprogramming serves as an excellent model system to explore the mechanisms driving gene regulatory network rewiring and their effects on cell identity. Reprogramming involves the forced expression of master transcription factor(s), leading to the conversion of one cell type to another. Ngn2 has been identified as such a master TF that can convert cell types such as astrocytes or fibroblasts towards an induced neuronal identity. However, it remains unresolved how TFs such as Ngn2 can accomplish such widespread remodelling, and how the influence of protein state and genomic/cellular context impact such changes.

In this collaborative project with the lab of Magdalena Götz, we employed single-cell multiomics combined with genome-wide profiling of 3D genome architecture to explore astrocyte-to-neuron reprogramming mediated by Ngn2 and a phosphorylation-resistant variant (PmutNgn2)[115].

1.4.1 Neuronal reprogramming

Brain injuries and neurodegenerative diseases often lead to the loss of neuronal function and eventual cell death. In mammals, the brain lacks regenerative capacity, except in a few specialized niches, preventing the passive recovery or replacement of these neurons[116]. To restore functional neuronal activity in damaged regions, two primary approaches to neuronal replacement therapy have been investigated: utilising exogenous sources or mobilizing endogenous sources of neurons[117].

Replacement therapies that utilise exogenous sources have explored various transplantation options using different neuronal sources of donor cells. For example, as a treatment for Parkinson’s disease, clinical trials have used developing midbrain dopamine cells derived from the human fetal ventral midbrain, with variable results[118], with some cells also being affected by the disease or the patient experiencing disease induced dyskineasias[117]. Alternatively, stem cells can be differentiated into a pure neuronal subtype and subsequently implanted[119]. However, despite significant advancements, transplantation approaches continue to face challenges that include the long-term viability of the grafted cells, tumorigenic risks, immune rejection by the host, and limited to being used in focal loci[120].

As an alternative to overcome the drawbacks of transplantation methods, endogenous sources of cells are of particular interest. Initially, with the identification of adult neurogenesis in the human brain[121], studies investigated recruiting new neurons from neurogenic niches. However, limited success has been achieved with recruiting neuroblasts to the injury sites due to the different spatial specialisation of the neurons, resulting in the differentiation towards unwanted neuronal subtypes and lack of longevity[122]. A compelling variation has been the conversion of local non-neuronal cells, such as astrocytes towards a neuronal fate by overexpressing master neuronal TFs[116].

1.4.2 Proneural TFs in the conversion of glial cells to induced neurons

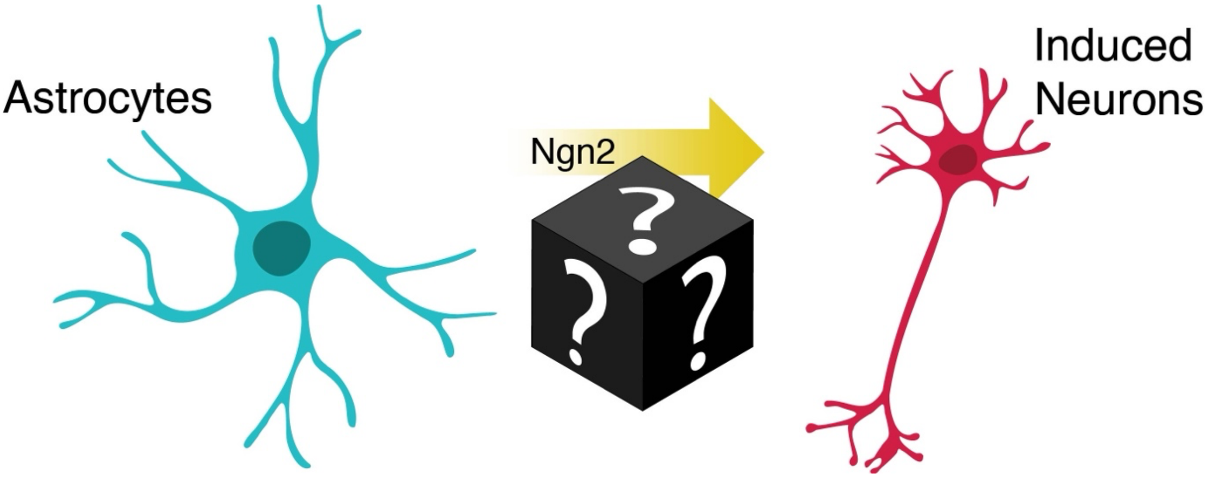

Figure 5 Ngn2 induced neuronal reprogramming.

Ngn2 as a master proneural TF can convert astrocytes into Induced neurons. However, the epigenetic mechanisms underlying this process remains to be explored.

The first demonstration that lineage-specifying TFs can be used to transduce one cell-type to another was the use of MyoD1 in converting fibroblasts to myoblasts[123]. This trans-differentiation has more recently been identified to be accompanied by a broad rewiring of cis-regulatory elements, promoters and insulation preceding transcriptional changes[124]. Adopting a similar strategy, Pax6 was identified as the first TF capable of converting astrocytes into neurons *in vitro*[125], with proof of concept *in vivo* experiments arriving shortly after[126–128]. Astrocytes have been recognised as an ideal target for neuronal reprogramming because they share a common precursor with neurons, exhibit plasticity, and become activated in response to neuronal damage. Their widespread distribution across the CNS also enables neuron generation in diverse regions, making this approach highly versatile for potential therapeutic applications[129].Downstream of Pax6, Ngn2 has been reported as an even more potent reprogramming TF capable of accurately specifying synapse-forming glutamatergic neurons and firing action potentials *in vitro*[130,131] (Fig. 5). To further improve efficiency for *in vivo* purposes, various combinatorial methods have been investigated alongside Ngn2 overexpression.

One major obstacle for successful is premature cell death due to an increase in reactive oxygen species (ROS) in a process shown as ferroptosis. To circumvent this, co-expression of an anti-apoptotic protein, Bcl-2, with Ngn2 greatly improves efficiency and minimises cell death achieving 90% conversion efficiency *in vivo*[132]. Other approaches have used small molecules such as forskolin and dorsomorphin alongside Ngn2 to improve reprogramming efficiency[133], whilst co-expression of Nurr1 has also been explored to generate lamina-specific neuronal subtypes[134].

Despite the additives alongside Ngn2 to improve reprogramming, in this context, Ngn2 remains the lineage-specifying factor that co-ordinates the upregulation of the neuronal genes and downregulation of astrocytic genes. In fibroblast reprogramming studies and *in vitro* differentiation, the pioneering activity Ngn2 has been highlighted by changes in chromatin accessibility[24,135]. By investigating DNA methylation and 3D genome architecture in a separate study, overexpressed *in vivo* Ngn2 also contributed to decreased DNA methylation at binding sites (correlating with increased activity of the region), and with increased chromatin looping amongst Ngn2 bound enhancers and promoters[136].

However, a comprehensive alteration of the epigenomic states during neuronal reprogramming remains unexplored (Fig. 5). This is of particular importance with regard to the changes in 3D genome architecture during cellular reprogramming. Prior to this project, looping changes in reprogramming have only been examined in somatic cells driven towards pluripotency[137,138]. These studies revealed dynamic changes in 3D genome looping, as well as instances where the proper formation of some topologically associating domains (TADs) failed, resulting in the miswiring of regulatory elements with their target genes. Imprecise TADs formation, inappropriate formation of new TADs or retaining characteristics of the original lineage, could all contribute to the establishment of aberrant cell fates[139]. Furthermore, it is understood that a 9S-A phosphomutant form of Ngn2 (PmutNgn2), resistant to phosphorylating by proline-directed serine kinases show stronger neurogenic activity in development[114] and in human IPSC derived glia to neuron reprogramming[140]. The underlying epigenetic changes that drive increased neurogenic activity are not yet fully understood.

Therefore, in this project, we investigated the changes in the epigenome associated with neuronal reprogramming of primary astrocytes mediated by Ngn2 and PmutNgn2. We identified PmutNgn2 as a more potent reprogramming factor, corresponding to more extensive epigenetic rewiring. We further identified the importance of co-factors in facilitating successful reprogramming with Yy1(a co-factor known to play in chromatin looping[141] and neurogenesis[142]), as a direct interactor of Ngn2 important for its activity[143].

1.5 Proneural TF-mediated epigenome remodelling in different cellular contexts

Ngn2 has been shown to direct the generation of glutamatergic cortical neurons when overexpressed in primary astrocytes derived from the same cortical region[143], mirroring the endogenous developmental properties of Ngn2. This neuronal identity is shaped by the shared precursor origin of the astrocytes and neurons in that region[129]. In contrast, Ngn2 in the developing spinal cord functions to generate GABAergic interneurons[144] and motor neurons[145]. Spinal cord astrocytes exhibit a distinct transcriptome compared to astrocytes in the cerebral cortex. Due to the unique developmental role of Ngn2 in the spinal cord and the specific identity of spinal cord astrocytes, reprogramming these cells with Ngn2 results in the generation of V2 interneurons, reflecting their regional identity (Fig. 6)[146]. Ngn2 has also been incorporated into transcription factor cocktails to generate diverse neuronal types including dopaminergic neurons from fibroblasts[147]. The importance of the starting cell-type for neuron generation is clearly reflected by the potential to generate diverse neuronal subtypes. However, the mechanisms involved in driving this specificity remains unexplored.

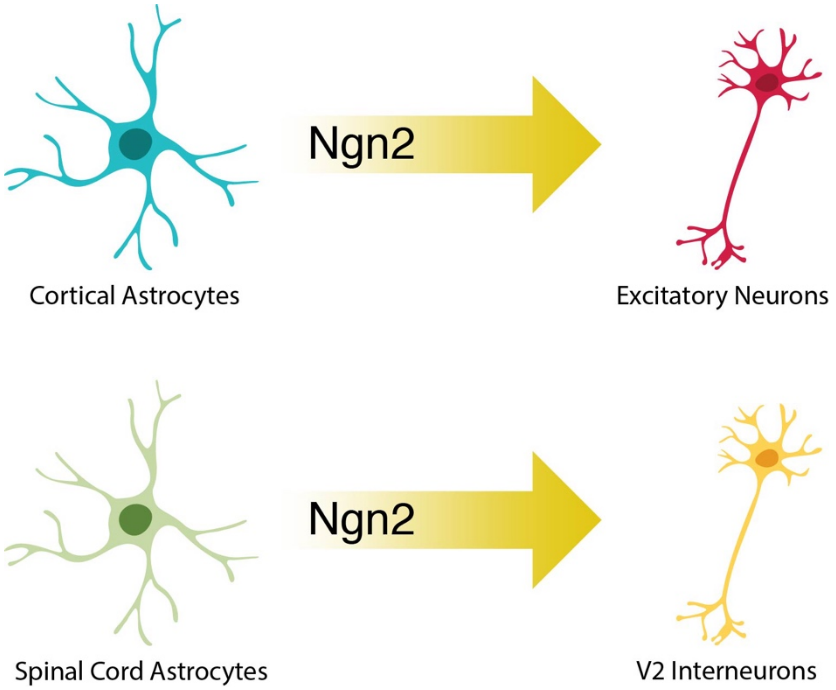

Figure 6 Context dependent neuronal subtype generation

Starting cell type influences the neuronal subtype generated by Ngn2 mediated reprogramming. Ngn2 reprogrammed cortical astrocytes generate excitatory neurons, whereas reprogrammed spinal cord astrocytes generate V2 interneurons.

Ngn2 has also been shown to highly modulated by co-factors and post-translational modifications. In developmental systems, Ngn2–E47 heterodimers exhibit a diminished capacity to activate neuronal differentiation genes compared to Ngn2–Ngn2 homodimers, and in reprogramming we identify that Yy1, a direct interactor of Ngn2, enhances its epigenetic rewiring[113,143]. Thus, we were also intrigued by how the initial cell type impacts the co-factor repertoire of Ngn2 and, consequently, its activity.

To address these questions, in collaboration with Vera Manelli (Bonev Lab, PhD student), we overexpressed Ngn2 in two distinct environments: the more foreign mouse embryonic stem cells (mESCs) and the more native differentiated neural progenitor cell (NPCs). We employed genome-wide multiomics assays to investigate global epigenomic changes and the differential co-factors influencing this process. Consistent with our prior findings, we observed that Ngn2 binding generally promotes increased chromatin accessibility, DNA demethylation, and enhanced chromatin interactions in both cell types. We find that these direct effects are likely mediated by interactions with chromatin modellers and the NuRD complex. Interestingly, we also identify cell type specific indirect effects resulting in a genome-wide increase in DNA methylation, decreased accessibility at sites occupied by REST and influences on global 3D genome organization in mESCS. These effects are likely attributable to ectopic binding and mis-regulation influenced by the cellular and epigenetic context[148].

1.6 Development of single cell multiomics methods

The advent of single-cell techniques has revolutionized our understanding of molecular biology. Previously, our knowledge relied on the average representation of data pooled from cell populations. In contrast, single-cell techniques, instigated with sc-RNA-seq[149], have enabled the discovery of cell types that display intrinsic heterogeneity and dynamic changes within tissues in developmental and disease contexts[150]. The transformative potential of these technologies is exemplified by initiatives such as the Human Cell Atlas (HCA) and subsequent consortium-based efforts. These projects aim to generate comprehensive reference maps of tissues detailing the function and characteristics of every cell type and more recently spatial position[151].

However, to understand the molecular mechanisms that drive the phenotypic changes revealed through gene expression, the development of multiomics techniques to profile the epigenome together with the transcriptome has been necessary. This has led to the emergence of a wide array of single-cell methods designed to profile one or multiple epigenetic layers in combination with the transcriptome[152]. To date, the most prominent single-cell multiomics method is joint measurement of chromatin accessibility alongside transcriptome, due to its relative experimental ease and the commercialisation of the assay. Chromatin accessibility is commonly used to identify active regulatory regions[38] and to examine changes in the regulatory cell state linked to transcription.

As discussed earlier, the genome wide changes in the epigenome mediated by TFs and the relationship across the epigenetic layers are complex, with studies identifying that increase accessibility and gene expression are not simply correlated with a decrease in DNA methylation and increased contact strength, but rather represents a continuum[153]. Furthermore, significant heterogeneity in chromatin accessibility has been observed at the single-cell level. Quantitative analyses have shown that approximately 25% of accessible chromatin regions differ between two individual cells[154]. Such cell-to-cell variations in chromatin accessibility are thought to play a functional role, with corresponding variability observed in the expression of associated genes[155].

Similarly, single-cell methylation assays have shown that variability at regulatory elements contributes to the gene expression heterogeneity[156]. Associated with this, promoter methylation seems to strongly correlate with gene silencing, yet at distal regulatory elements both positive and negative associations have been identified[157]. Furthermore, single-cell HiC[158] and super-resolution microscopy[159] have also reported variance in chromosome, TADs and looping structures from cell to cell[160]. As a result, understanding the relationships across the various epigenetic layers that contribute to this heterogeneity, as well as their functional significance, will be crucial for deeper insights into the molecular mechanisms of cellular behaviour and identity driven by TFs and co-factor dynamics.

However, the integration of chromatin architecture profiling into multiomics approaches that include the transcriptome has remained limited until very recently. Additionally, no high-throughput assays exist that can simultaneously profile 3D genome architecture, DNA methylation, chromatin accessibility, and the transcriptome at single-cell resolution. To address this gap, we are developing a method called sc-3DRAM-seq (single-cell 3D Genome, RNA, Accessibility, and Methylation sequencing) aimed at achieving this comprehensive profiling in a high throughput manner.

1.6.1 Single-cell omics approaches

Different methodologies have been utilised to profile single cells in multiomics assays, each varying in throughput. The three most common methods include: plate based, droplet based and split-pool barcoding based approaches (Fig. 7).

Plate-based methods represent the lowest-throughput variation of single-cell approaches, requiring cells to be sorted into individual wells of a multi-well plate, typically via dilution or Fluorescence-activated Cell Sorting (FACS). While certain

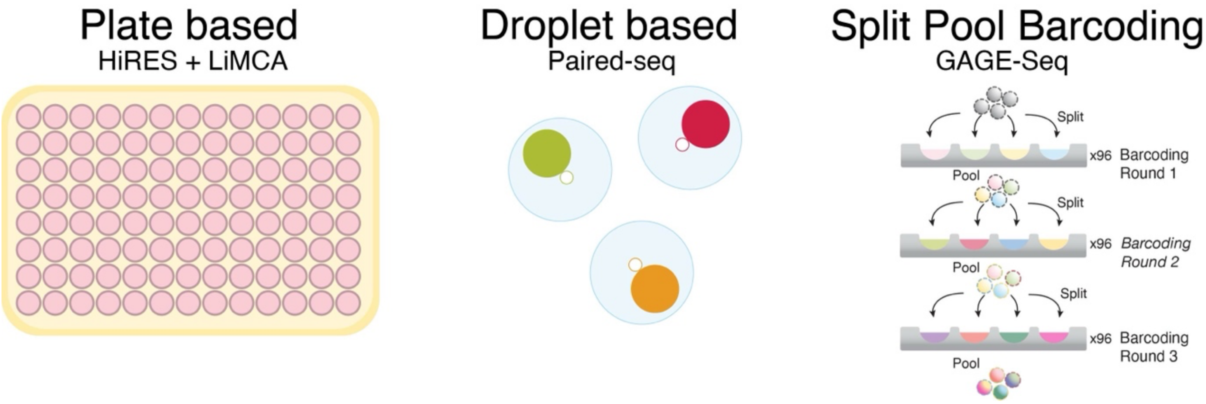

Figure 7 Single-cell omics approaches and corresponding scHi-C + transcriptome techniques

The 3 main categories of methods used in single-cell omics are plate-based, droplet based & split pool barcoding based techniques. Recent scHi-C + transcriptome methods utilise all 3 variations.

reaction steps may be performed in bulk, library preparation is carried out in each well individually, allowing for the incorporation of a unique cell-identifying index. This index, read during Next-Generation Sequencing (NGS), enables the assignment of sequencing reads to specific wells. When assays target both RNA and DNA components, RNA can be distinguished from DNA either through the use of separate RNA/DNA indices or by physically separating the RNA from DNA. The latter is achieved by magnetically pulldown of reverse-transcribed cDNA molecules using streptavidin beads bound to primers containing biotin[151].

Droplet-based approaches are the most common, using microfluidics systems to separate individual cells into nanoliter-sized aqueous droplets, where there are associated with a cell identifying barcode[161]. Capable of profiling thousands of cells, the method has been commercialised by 10x Genomics. Due to the high experimental costs of the instrumentation and kits required, multiplexing of cells with a pre-barcode has also been utilised to increase the number of cells that can be used in each experiment[162,163].

Split-pool barcoding is most recent method providing the highest throughout, allowing >100,000 cells to be assayed in one experiment[164]. This method uses multiple rounds of barcoding, starting with cells distributed into a 96-well plate, where each well contains a unique barcode. The cells are then pooled and redistributed into another plate with a different set of 96 barcodes, with additional rounds performed as needed. Through probabilistic chance, each cell is expected to follow a unique path through the barcoding rounds, resulting in a distinctive barcode combination that serves as its identifier. The modular design of this approach allows for flexibility in the number of barcoding rounds, accommodating the desired scale of cell profiling. For example, three rounds yield 884,736 unique combinations, while four rounds increase this to 84,934,656. Separate RNA and DNA barcodes can be employed to distinguish data from these two nucleic acid types.

Given the high-throughput nature of split-pool barcoding, its cost-effectiveness, and the lack of requirement for specialized equipment and kits, we employ this approach for sc-3DRAM-seq.

1.6.2 The 3D genome and the transcriptome in single cells

The 3D genome architecture of cells has historically been studied with two complementary methods – microscopy and chromosome conformation capture (3C) techniques. The original principles of sub-nuclear organelles and chromosome territory came from microscopy using methods such as Fluorescence RNA and DNA in situ hybridization (FISH), indicating that nuclear positioning can correlate with gene expression levels. Advancement in super-resolution microscopy and live-cell imaging has further allowed direct visualisation of key regulatory loci and how they can change dynamically with transcription[165]. Of particular significance was ORCA, a method that enables simultaneous imaging of DNA loci and RNA expression at high resolution. This approach revealed that the proximity of promoters to known enhancers was predictive of nascent transcription[65]. Although microscopy allows high-resolution study of genomic loci in single cells, it is currently limited by low throughput in cell numbers and loci analysed[166].

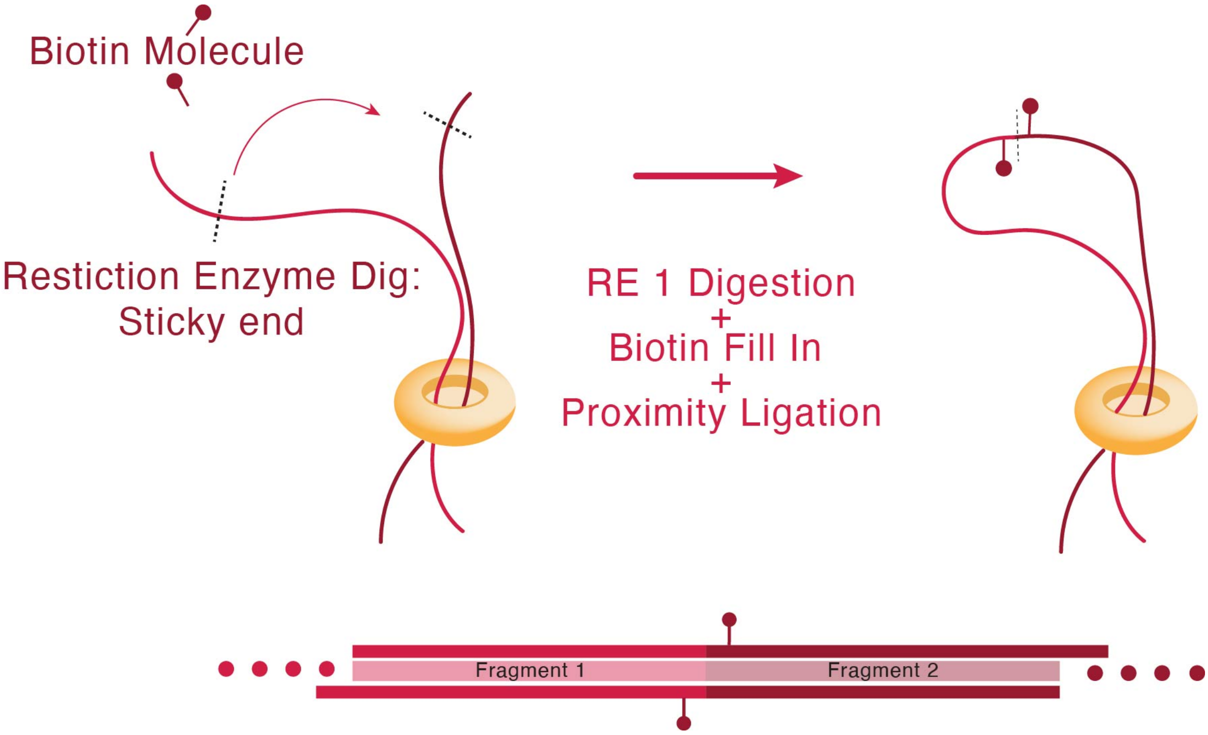

Figure 8 Basic Steps of Hi-C

The main steps of Hi-C following cell fixation are restriction enzyme digestion to create sticky ends, that are end-repaired with the introduction of a biotin molecule for subsequent enrichment, and proximity ligation.

On the other hand, high coverage 3C methods such as Hi-C can be used to map interactions genome wide in fixed cells, using NGS to identify DNA sequences in close spatial proximity[67]. The primary steps of 3C methods are cell fixation, restriction enzyme digestion of the DNA, a biotin nucleotide fill in (to enrich for correctly ligated fragments during library preparation) followed by ligation of DNA sequences in close proximity (Fig. 8). Orthogonal ligation-free methods have also been developed including SPRITE[167] and GAM[168]. Further improvements to resolution have been approached by either using multiple enzymes in the case of Hi-C3.0[169], or Micro-C[170] using MNase instead of restriction enzymes for the digestion – both relying on generating shorter DNA fragments. Furthermore, enrichment of target sites of interest is achievable with techniques such as Promoter Capture Hi-C[171], HiChIP[172], Micro Capture-C[173], and Tiled[174]/Region[175] Micro Capture-C.

Bulk 3C methods provide a snapshot of the population average chromatin structure, driving the advancement of single cell Hi-C. First established as a plate-based method, single-cell Hi-C recapitulated the non-random, but highly variable chromatin conformations observed using FISH[158]. It was subsequently utilised to identify the dynamic changes in structure during the cell-cycle[176]. Advances in the method by omitting the biotin fill in, led to creation of single nucleus Hi-C yielding 1–2 orders of magnitude more contacts[177], whilst other studies imaged the cells prior to single-cell Hi-C allowed the observed structures to be validated[178]. Combinatorial barcoding was employed in sci-HiC to increase the throughput of cells, but due to the design of this particular method utilising bridge adaptors, the number of contacts recovered per cells suffered[179]. Whereas, in Dip-C, the use of multiplexed transposon based library preparation drastically improves detection[180]. SPRITE was also adapted into a single-cell version that allowed multiway contacts to be detected[181], with scNano-HiC achieving similar results with long read nanopore sequencing[182]. Finally and most recently, single-cell HiC was combined with the droplet based 10x system to create Droplet-HiC generating results with increased contact measurement relative to its plate based alternatives[183].

Despite significant advances in these single-cell Hi-C methods, the genomic resolution per cell remains limited, making it challenging to quantitatively assess short-range interactions, such as single enhancer-promoter contacts[165]. Therefore, to improve chromatin contact resolution, cells need to be aggregated based on a common feature. Algorithms such as Higashi[184] and scGHOST[185], have employed imputation techniques to detect multiscale 3D genome features, whilst scHiCluster[186] identifies similarities in interaction domains and clusters in the single cells. However, it remains unclear how accurate such algorithms are due to the inherent heterogeneity in chromatin structure and the lack of a ground truth comparison.

Thus, at the inception of this project, we aimed to develop a method capable of measuring sc-Hi-C and sc-RNA simultaneously. The transcriptome not only enables cell identity classification but also serves as an anchor to pseudo-bulk sc-Hi-C data from similar cells, enhancing chromatin contact resolution. Although we made significant progress using split-pool barcoding to create a high-throughput technique, similar methods were recently published. This included the plate-based HiRES [187]and LiMCA[188], the droplet based Paired-Seq[183] and the split pool barcoding based GAGE-seq[70] (Fig. 7).

GAGE-seq was the closest conceptual counterpart to our method, with a direct comparison revealing it to be the more effective approach. Building on the methodologies developed by GAGE-seq and incorporating insights from our prior work, we are enhancing the technique by integrating NOMe-seq[189], as seen in scNMT-seq[190] and bulk 3DRAM-seq[191]. This had led to the development of sc-3DRAM-seq, a method that not only captures 3D genome and transcriptome information but also includes DNA methylation and chromatin accessibility at the single-cell level.

2. Material and Methods

2.1 Direct neuronal reprogramming materials and methods

Materials and methods associated with the results obtained in the directed neuronal reprogramming project can be found detailed in Pereira et al. Nat Neuro, 2024[115]. Below, I describe the methods I contributed to in this project, specifically Methyl-HiC.

2.1.1 DAPI staining for Methyl-HiC

To conduct the Methyl-HiC, cells were required to be sorted for the G0/G1 cell cycle to prevent cell-cycyle mediated confounding factors of the 3D genome. This was achieved by DAPI staining of the celsl using the previously described staining protocol in Noack et al. 2019[136]. Briefly, cells were fixed with 1% PFA, and permeablished with wash buffer containing 0.1% saponin. This was followed by the cells being stained in wash buffer containing DAPI at a 1 to 1000 dilution. The cells were subsequently washed twice and filtered via a 40-uM cell strainer. The suspension was then sorted, and pelleted down for flash freezing in liquid nitrogen. The cells were stored at -80C and processed for Methyl-HiC at a later time point.

2.1.2 Methyl-HiC

The Methyl-HiC procedure was peformed as described in this detailed protocol - https://www.protocols.io/view/methylhic-bif2kbqe/, established in Noack et al. 2019[136], developed by adapting the original Methyl-HiC protocol from the Ecker Lab[192,193]. In short, the frozen cell pellet previously sorted for the G0/G1cell cycyle was thawed on ice, with nuclear isolation perfomed with lysis buffer containing freshly prepared 0.2% Igepal. This was followed by the standard steps found in chromatin conformation techniques, including SDS permeablisation, restriction enzyme digestion (DpnII), biotin-14-dATP fill in and proximity ligation. The DNA was purified using ethanol preciptation following an overnight reverse crosslinking step and subject to sonicator mediated fragmentation. The DNA was converted the EZ DNA Methylation-Gold Kit (Zymo Research) bisulphite kit, with the an addition of spike in methylation controls to access conversion efficiency downstream. The converted DNA was prepped for sequencing using the using the Accel-NGS Methyl-Seq DNA Library Kit (Swift Biosciences) and EpiMark Hot Start Taq (New England Biolabs) based amplification with specific Methyl-Seq indexing primers (Swift Biosciences).

2.1.3 Hi-C Mapping and QC

Methyl-HiC FASTQ files were mapped to the mouse genome (GRCm38, mm10) using JuiceMe[194]. Only uniquely mapped reads (mapq score >30) were used after eliminating PCR duplicates. MethylDackel assessed CpG methylation, and reads were pooled from replicates. For Hi-C, reads were excluded based on mapped restriction fragments and distance (<1 kb).

Filtered fend-transformed read pairs were imported and normalized using Shaman[153]. Hi-C scores were calculated using kNN.

2.1.4 Analysis of contact probablities, contact probabilities and TAD boundary indetification

Contact probabilities were calculated as a function of genomic distance following established methods described previously[153]. Insulation scores and differential TAD boundaries were determined to quantify local chromatin compaction, using sliding windows to measure interaction frequency drop-offs, as detailed in prior studies[136,153,158].

2.1.5 Compartments Strength

Dominant eigenvectors contact matrices (binned at 250 kb) were computed using Hi-C data to evaluate compartment strength. These matrices were used to assign genomic regions to A or B compartments based on their eigenvector values. Compartment strength was quantified by calculating the log2 ratios of observed-to-expected contact frequencies for intrachromosomal contacts (separated by at least 10 Mb) within the same compartment (A–A or B–B) and between compartments (A–B). The expected contact frequencies were derived from the overall decay of interaction frequencies with genomic distance. All analyses were performed using publicly available scripts from the Dekker lab repository (https://github.com/dekkerlab/cworld-dekker) [195], ensuring consistency with standard methods for compartment analysis.

2.1.6 Average TAD Contact Enrichment

Enrichment within TADs was calculated using insulation and contact enrichment metrics as previously described[136,153].

2.1.7 Assesing Contact Stength at identified genomic features

To determine contact enrichment at pairs of genomic features, such as Ngn2 ChIP-seq peaks or EGPs, two complementary methods were employed. The first method involved generating aggregated Hi-C contact maps and calculating the log2 ratio of observed to expected contacts within a defined window centred on the feature of interest. The contact enrichment was quantified by comparing the contact strength at the feature center to that at its edges. The second method used kNN-based Hi-C scores within a 10-kb window for each feature pair, visualized as scatter plots or box plots to highlight pair-specific contact patterns. Statistical significance was calculated using the Wilcoxon rank-sum test.

2.2 Context dependent epigenome rewiring materials and methods

Materials and methods associated with the results obtained in the context dependent epigenome rewiring project can be found detailed in Manelli et al. bioRxiv, 2024[148]. I describe the methods I contributed to in this project, specifically the cell line generation, preliminary experiments with UMI-4C, Bisulphite Amplicon sequencing and 3DRAM-seq. Data analysis methods are described in this section only if they differ from those previously mentioned.

2.2.1 Cell Culture

MEFs (Gibco, Cat. N: A34181) were cultured on 0.1% gelatin-coated plates (Merck, Cat. N: ES-006-B) following the manufacturer's instructions. The cells were maintained in DMEM (ThermoFisher, Cat. N: #21969-035) supplemented with 10% FBS (ThermoFisher, Cat. N: 16141079), 50 U/mL penicillin-streptomycin (Gibco, Cat. N: 16141079), 0.1 mM non-essential amino acids (Gibco, Cat. N: 11140035), and 0.1 mM 2-mercaptoethanol (Gibco, Cat. N: 31350010), with media changes every other day.

Flag-Neurog2 ES cells were cultured per Bonev et al., 2017[153], with modifications. Cells were grown on MEFs in DMEM (ThermoFisher, Cat. N: 21969-035) supplemented with 15% FBS, 1,000 U/mL LIF, 0.1 mM non-essential amino acids, 50 U/mL penicillin-streptomycin, and 0.1 mM 2-mercaptoethanol, with daily media changes. Cells were passaged every other day at 3×10^5 cells using TryplE (Life Technologies, Cat. N: 12604013) and separated from MEFs using a 30-minute sedimentation step. For Neurog2 overexpression, cells were replated without MEFs, treated with doxycycline the next day, and harvested 24 hours later unless stated otherwise.

2.2.2 Inducible FLAG-Neurog2 cell line generation

A2Lox.cre mES cells[196] were induced with 500 ng/mL doxycycline for 24 hours before electroporation (Amaxa nucleofector, solution VHPH-1001, program 96-CG-104) of 1 µg p2Lox-Flag-Neurog2 plasmid. Electroporated cells were plated on neomycin-resistant MEFs (Gibco, Cat. N: A34963) and selected with 300 µg/mL G418 (Gibco, Cat. N: 11811023) starting 24 hours post-plating. After 10 days, colonies were picked, expanded, and maintained on neomycin-resistant MEFs.

2.2.3 Neural Differentiation

Neural differentiation was adapted from prior protocols[153] with minor changes. Cells were plated at low density (1×10^5 cells/plate) on gelatin-coated 10 cm dishes in ES media without MEFs. After 12 hours, the media was replaced with DDM media (DMEM/F12 + GlutaMAX, supplemented with 1× N2, 1 mM sodium pyruvate, 500 µg/mL BSA, 50 U/mL penicillin-streptomycin, and 0.1 mM 2-mercaptoethanol). Cyclopamine (1 µM) was added every other day from day 2 to day 10, with media changes every two days. On day 11, cells were treated with doxycycline and harvested 24 hours later following Trypsin dissociation unless otherwise specified.

2.2.4 UMI-4C

UMI-4C was performed following the protocol in Schwartzmann et al. Nature Methods, 2016[197]. In brief, 5–10 µg of 3C template DNA was sonicated to 450–550 bp fragments, followed by end-repair, A-tailing, dephosphorylation, and ligation with Illumina-compatible indexed adapters. Libraries were purified using AmpureXP beads, and DNA concentration was measured with a Qubit ssDNA kit. Nested PCR was performed in two steps, with bait and enrichment primers, using 15–20 cycles per reaction, and products were cleaned with AmpureXP beads. The libraries, typically ~500 bp in size, were pooled, diluted to 4–10 nM, and sequenced on an Illumina platform.

2.2.5 Bisulphite amplicon sequencing

BSamp-seq was performed as described in Noack et al. Life Science Alliance, 2019[198]. Briefly, for bisulphite conversion, 0.2–1 µg of DNA was processed using the EpiTect Bisulphite kit (QIAGEN) for BSamp-seq or the TrueMethyl-seq kit (Cambridge Epigenetix) for oxBSamp-seq. Converted DNA served as a template for PCR amplification of target regions with bisulfite-specific primers (Table S1). Amplicons were validated by gel electrophoresis, purified (QIAquick PCR Purification Kit), pooled at 0.2 µM, and prepared for sequencing using the Nextera DNA Library Preparation Kit (Illumina). Libraries were amplified (KAPA High Fidelity Master Mix), purified (Ampure XP-beads), and sequenced on a Miseq-Nanoflowcell, yielding ~100,000 reads per run. Data were analysed with Bismark 0.16.0.

2.2.6 ImmunoFACS

This detailed ImmunoFACS protocol https://www.protocols.io/view/immunofacs-b2a2qage/) was done as previously described with subtle modifications. NPCs were resuspended in PBS at 1 million cells/mL, fixed with 1% PFA for 10 minutes at room temperature, and then quenched with 0.2M glycine for 5 minutes. After centrifugation and washing in PBS with 1% BSA, cells were incubated in wash buffer for 15 minutes at 4°C. Cells were stained with PAX6-Alexa Fluor 488 (1:40; BD Biosciences, Cat.N: 561664). G0-G1 Pax6+ cells were sorted using a FACSAria III (BD Biosciences). Post-sort, cells were snap-frozen and stored at -80°C for subsequent 3DRAM-Seq analysis.

2.2.7 3DRAM-Seq Library generation

3DRAM-Seq libraries were prepared as descirbed in this detailed protocol (https://www.protocols.io/view/3dram-seq-enables-joint-epigenome-profiling-of-spa-brf8m3rw), adapted from the previously published protocol[191]. Briefly, DSG was used as a second fixative and DdeI was jointly used with DpnII in the restriction digestion.

2.2.8 3DRAM-seq Analysis

3DRAM-seq libraries were processed and sequenced using the Illumina NovaSeq 6000 platform[193]. Read mapping was carried out using an adapted version of the TAURUS-MH pipeline197, which allows for splitting of reads at ligation junctions and alignment of bisulfite-converted reads through Bismark[199], as previously described[136].

The TAURUS-MH pipeline mapped the reads and generated fragment-end-transformed read pairs, which were converted into misha tracks and loaded into the mm10 reference genome. The Shaman R Package[200] was employed to shuffle the observed Hi-C contacts and generate expected models that preserved coverage and distance distributions while excluding certain features (e.g., TADs). The k-nearest neighbors (kNN) algorithm was used to calculate Hi-C scores.

2.2.9 CpG and GpC Methylation

CpG and GpC methylation levels were calculated using Bismark's methylation extractor, processing only uniquely mapped reads. To distinguish between CpG and GpC methylation, the coverage2cytosine function in Bismark[199] was applied with the –nome-seq option.

2.2.10 Estimation of Bisulfite Conversion Efficiency

The efficiency of bisulfite conversion was evaluated using Bismark[199] in paired-end mode with the –nome-seq option, focusing on CpG methylation in unmethylated lambda DNA. Detection rates for CpG and GpC methylation were determined using fully methylated pUC19 DNA and in situ GpC-methylated lambda DNA, based on protocols described earlier[136].

2.2.11 Identification of DMRs and DARs

Differentially methylated and accessible regions were identified with gNOMeHMM, as previously described[191]. Accessible peaks from noDox and Dox conditions were merged across cell types to create a unique dataset. MethylKit was then used to identify differentially accessible and methylated regions.

2.2.12 TF Motif Analysis

Motif enrichment analysis was conducted using motifs filtered from the JASPAR2024 database and expressed in ES or NPC cells (FPKM ≥ 1), leveraging the MonaLisa package[201] as previously described[143]. K-mer enrichment analysis was also performed with MonaLisa. Co-occurring TFs in Neurog2 ChIP-seq data were identified using the TF-COMB Python Package[202].

2.3 Development of sc-3DRAM-seq materials and methods

2.3.1 GAGE-seq

GAGE-seq was performed as described in the publication[70], with plate-based steps pooled into one Eppendorf tube for bulk quality control experiments.

Briefly, barcoded adapters for single-cell RNA and Hi-C sequencing were annealed, with barcode designs resembling Split-seq and SHARE-seq. Crosslinked cells from various sources were lysed using high-salt buffers or simplified protocols for certain cell types. Reverse transcription was performed with biotinylated RT primers, followed by thermal cycling. Chromatin was fragmented, proximity ligated and subjected to a second round of fragmentation and barcoding. Cells underwent combinatorial barcoding in 96-well plates using custom ligation mixes, with reverse crosslinking separating scHi-C and scRNA-seq libraries. DNA and RNA were precipitated, purified with MyOne Dynabeads, and prepared for sequencing. Libraries were pooled and sequenced as paired-end reads (PE 150) on Illumina platforms.

2.3.2 sc-3DRAM-seq

Sc-3DRAM-seq was adapted from the GAGE-seq protocol with following key modifications. Barcodes used were devoid of cytosines and the size of the barcode was increased to 12bp. The barcodes were ordered with methyl-cytosines at primer binding sites. 200k nuclei were used for input into the RT reaction. The GpC methyltransferase was performed as previously reported[191], with slight modifications. Briefly, the nuclei were washed once with 1× GpC buffer (New England BioLabs, M0227S) containing 1% BSA (Sigma-Aldrich, B6917), and then incubated for 6 hours at 37°C in a reaction mixture containing 60 U of M.CviPI (New England BioLabs, M0227S) and 0.6 mM SAM (New England BioLabs, B9003). During the incubation, the reaction was supplemented with M.CviPI and SAM every hour. The SDS step was moved from before RT in GAGE-seq to after the GpC methylation here. Bisulphite conversion was performed using an EZ DNA Methylation-Gold kit (Zymo Research, D5005). SPLAT-adapter ligation was performed as previously described[203]. SPLAT-adapter ligation involved a reaction mixture of ss2, T4 DNA ligase buffer, PEG4000, T4 DNA ligase, and nuclease-free water, incubated at 20°C for 1 hour, followed by AMPureXP bead purification. The libraries with amplified with NEBNEXT Q5U Master Mix (New England BioLabs, M0597S). The amplification primers were designed for the library to be ready for sequencing without tagmentation.

2.3.3 sc-3DRAMseq analysis

Raw FASTQ files were processed using *splitcode*[204] to identify and select reads matching the expected barcode 1 and barcode 2 sequences. After adapter sequences were trimmed using *trim_galore*, the reads were aligned to the reference mouse genome mm10 using either *bwa-mem2* or *bwa-meth*. For Hi-C contact identification, *pairtools* was used on reads with unique alignments and a minimum mapping quality of 10, and for deduplication. The resulting unique sorted Hi-C pairs in the .pairs files were used as input for custom Python scripts to visualize the decay of contact frequency as a function of genomic distance (cis-decay plots) or the total number of usable Hi-C pairs (summary grouped barplot). Contacts were binned into short-range (below 1 kb), middle-range (1–20 kb), long-range (above 20 kb) cis contacts, and trans contacts, with normalization by the number of sequenced or barcoded reads.

Methylation data were extracted from BAM files using *allcools*[205] to obtain single-base methylation dataframes. GpC and CpG methylation data were separated based on nucleotide context: only GCAN, GCTN, and GCCN contexts were used for accessibility, while ACGN and TCGN contexts were used for CpG methylation, to avoid biases from nucleotide contexts like GCG, where methylation could derive from either GpC treatment or endogenous CpG methylation. Using a custom Python script, average methylation and accessibility levels were plotted around CTCF sites previously identified by ChIP-Seq.

3. Results

3.1 Multiscale epigenome rewiring in the direct neuronal reprogramming of astrocytes

The results presented in this section are the product of a collaboration with the Götz lab. My major contribution to this study is the Methyl-HiC and support with Yy1 CUT&RUN. Other major experiments were performed by Allwyn Pereira.
This work has been published in Nature Neuroscience[115].

3.1.1 Experimental design

Lentiviral expression constructs were created to code for the green fluorescent protein (GFP), Ngn2-IRES-GFP or PmutNgn2-IRES-GFP. The proteins are downstream of the doxycycline (dox) inducible promoter, facilitating conditional expression. The addition of the IRES-GFP to our Genes of Interest (GOI) allowed imaging of the induced cells (Fig. 9).

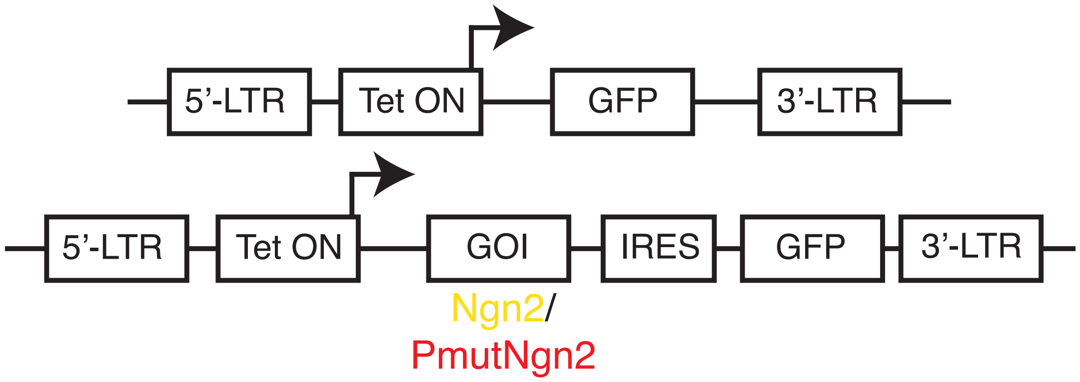

Figure 9 Lentiviral construct design

The generated constructs were transduced in primary mouse cortical astrocytes derived from postnatal day P5-6 following 8 days of *in vitro* culture. Dox was administered the following day to induce express of the control GFP or our GOIs. The induced cells were cultured up to 7 days post induction (dpi), to monitor the astrocyte to neuron reprogramming. Experiments to understand the epigenome changes mediated by Ngn2 or PmutNgn2 activity were carried out at 2dpi to understand the early-stage rewiring associated with their overexpression (Fig. 10).

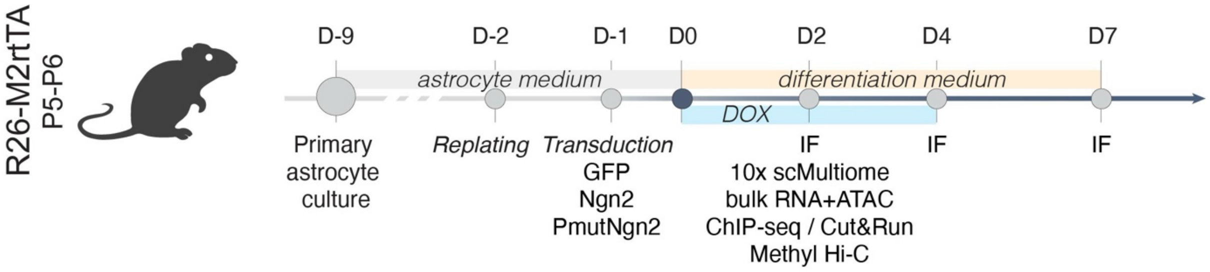

Figure 10 Reprogramming experimental design schematic

3.1.2 PmutNgn2 enhances the efficiency and speed of astrocyte-to-neuron reprogramming

To assess reprogramming efficiency at 2,4 and 7dpi, immunostaining was performed for βIII-tubulin and Gfap to label neurons and astrocytes, respectively. Ngn2 and PmutNgn2 induction demonstrated a progressive increase in the proportion of induced neurons (iNs), with PmutNgn2 achieving a higher iN ratio and more pronounced Gfap reduction compared to Ngn2 by 7 dpi (Fig. 11). The GFP control as expected retained astrocytic morphology and expression profiles and did not gain any neuronal expression, indicating that the transduction, dox treatment or culture mediums do not influence the reprogramming.

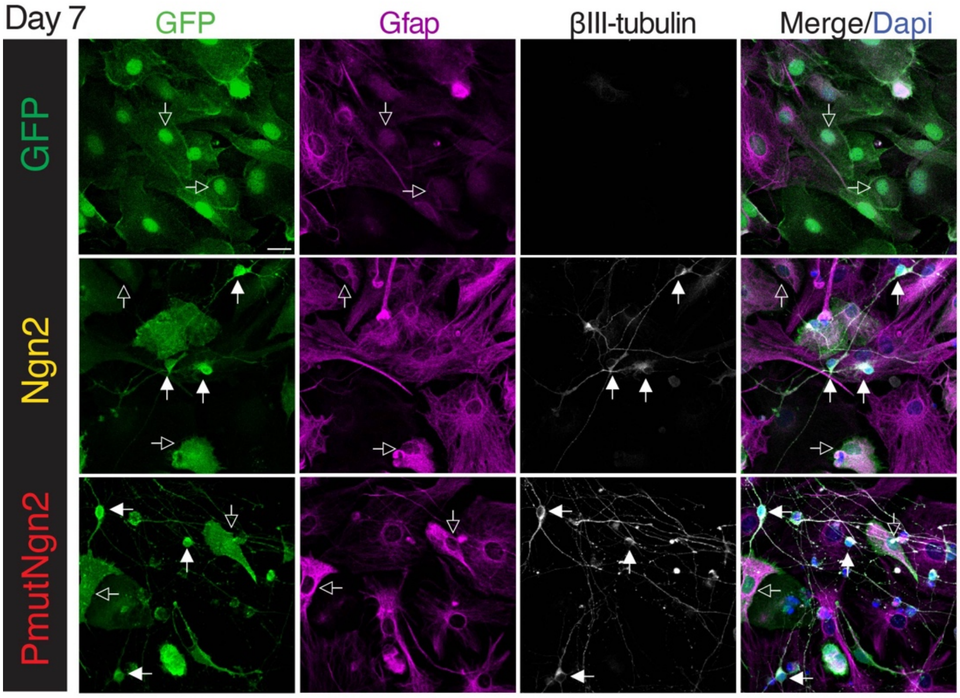

Figure 11 Immunostained astrocytes 7dpi.

Representative micrographs of astrocytes immunostained with βIII-tubulin and Gfap at 7 dpi, aligned with the experimental conditions outlined on the left. Scale bar: 20 µm. Filled arrows highlight induced neurons (iNs), whereas empty arrows identify cells devoid of neuronal markers (n = 3).

Quantitatively measuring the proportion of βIII-tubulin and Gfap cells in the reprogramming conditions relative to the control GFP showed that βIII-tubulin is induced as early as 2dpi, with the PmutNgn2 condition demonstrating significantly higher proportions than the Ngn2 condition. A significant reduction in Gfap-positive cells was also observed starting at 4 dpi, becoming more pronounced by 7 dpi in PmutNgn2 iNs compared to those induced by Ngn2 (Fig. 12).

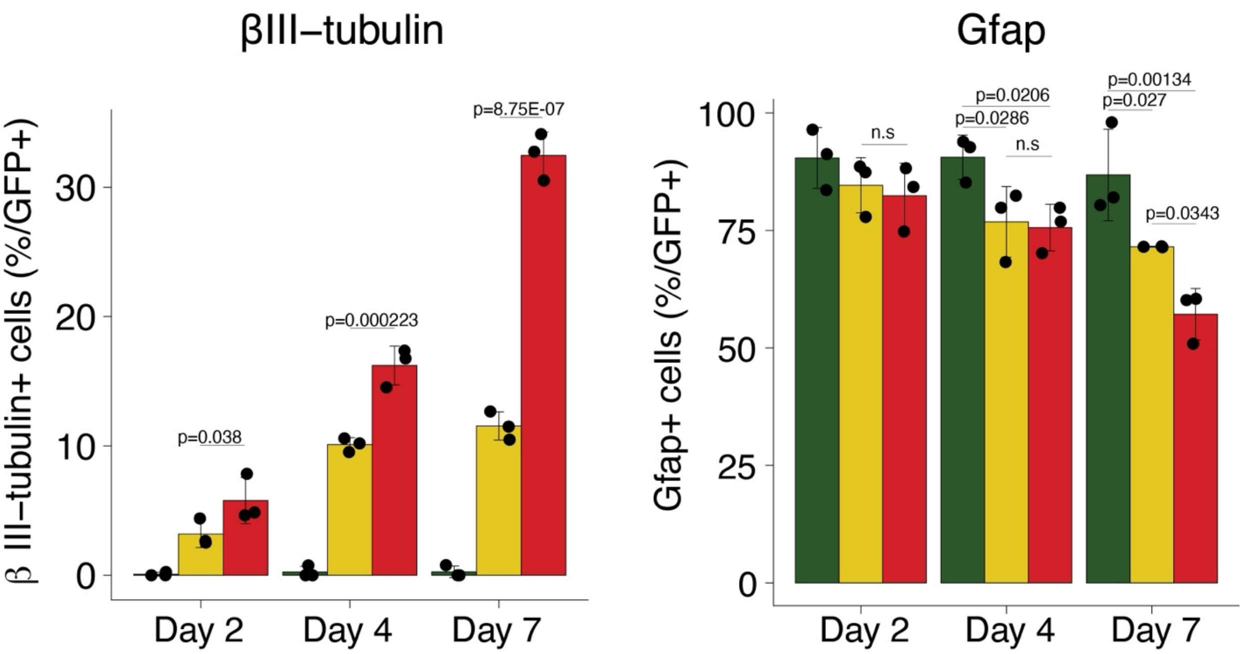

Figure 12 Histogram of βIII-tubulin+ and Gfap+ cell proportion.

Histograms illustrating the percentage of βIII-tubulin+ and Gfap+ cells among transduced populations (y-axis) relative to GFP control, over time (x-axis). Data are presented as mean ± s.e.m., with each dot representing a biological replicate (n = 3).

Furthermore, the speed of reprogramming was quantified using continuous live-cell imaging. Most PmutNgn2 transduced cells acquired neuronal like morphology by 64h, whilst most Ngn2 were only converted by 84h (Fig. 13). These findings indicate that PmutNgn2 drives the generation of a greater number of neuronal cells than Ngn2, attributed in part by its accelerated conversion.

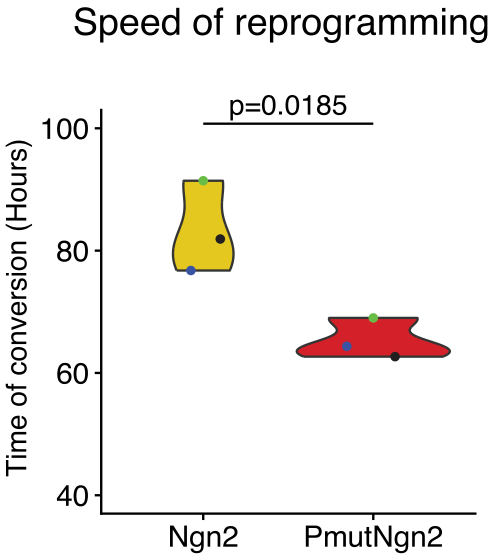

Figure 13 Violin Plot of Speed of Reprogramming.

Violin plot depicting initial timepoint at which each tracked cell displayed neuronal morphology. Each dot is a biological replicate (n=3). Statistical significance determined using a linear regression model.

3.1.3 PmutNgn2 promotes a transcriptional network that drives neuronal maturation

To examine the transcriptional alterations and chromatin accessibility rewiring underlying the enhanced reprogramming efficiency of PmutNgn2 iNs, we utilized the 10x Genomics Multiome platform to simultaneously profile both modalities at single-cell resolution. We performed the experiment on uninduced astrocytes and astrocytes transduced with GFP only, Ngn2 or PmutNgn2 at 2 dpi. Joint sc-RNA and sc-ATAC uniform manifold approximation and projection (UMAP) visualisation[206] labelled by sample identity revealed that the reprogramming Ngn2 and PmutNgn2 transduced cells cluster separately from the untransduced and control astrocytes (Fig. 14).

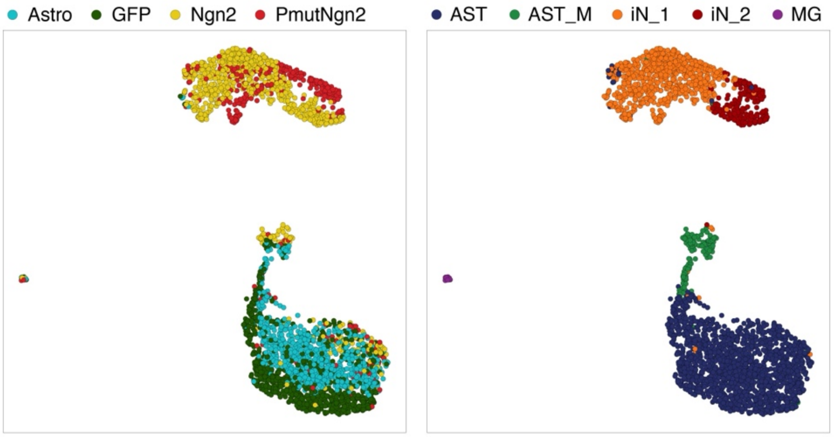

Figure 14 Joint UMAP projection coloured on experimental condition or cluster identity

We subsequently performed Louvain clustering of the joint single-cell RNA and ATAC dataset identified five major clusters: non-dividing astrocytes (AST), dividing (mitotic) astrocytes (AST_M), two neuronal populations (iN_1 and iN_2), and a smaller cluster representing microglial cells (MG) (Fig. 14). Interestingly, more PmutNgn2 cells contribute to iN_2 cluster relative to iN_1 (Fig. 15).

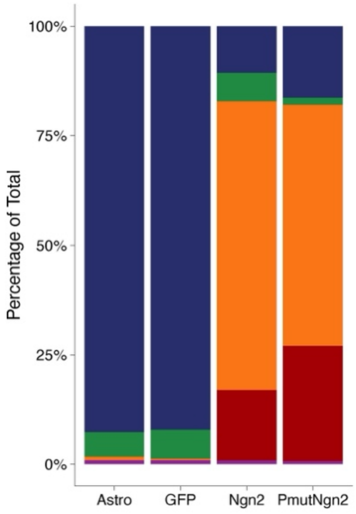

Figure 15 Stacked bar plot of relative proportion of the identified cell types in each experimental condition

Next, we sought to understand the transcriptional differences that delineate iN_1 and iN_2. Pseudotime analysis[207] suggested that the iN_1 is a precursor transitionary state to iN_2 (Fig. 16A). In line with this, we identify that on average the iN_2 cells had a higher expression of neuronal markers such as *Dcx*, *Tubb3* and *Rbfox3* than iN_1 as well as more pronounced downregulation of astrocytic genes (Fig. 16B). Similarly, although pan-neurogenic targets such as *Hes6*, *Prox1* and *Sox11* were expressed in both iN_1 and iN_2, levels were elevated in iN_2 (Fig. 16B). Analysis of the most variable genes across the pseudotime also showed enrichment of neuronal markers in iN_2, corresponding predominantly to the PmutNgn2 condition (Fig. 16C).

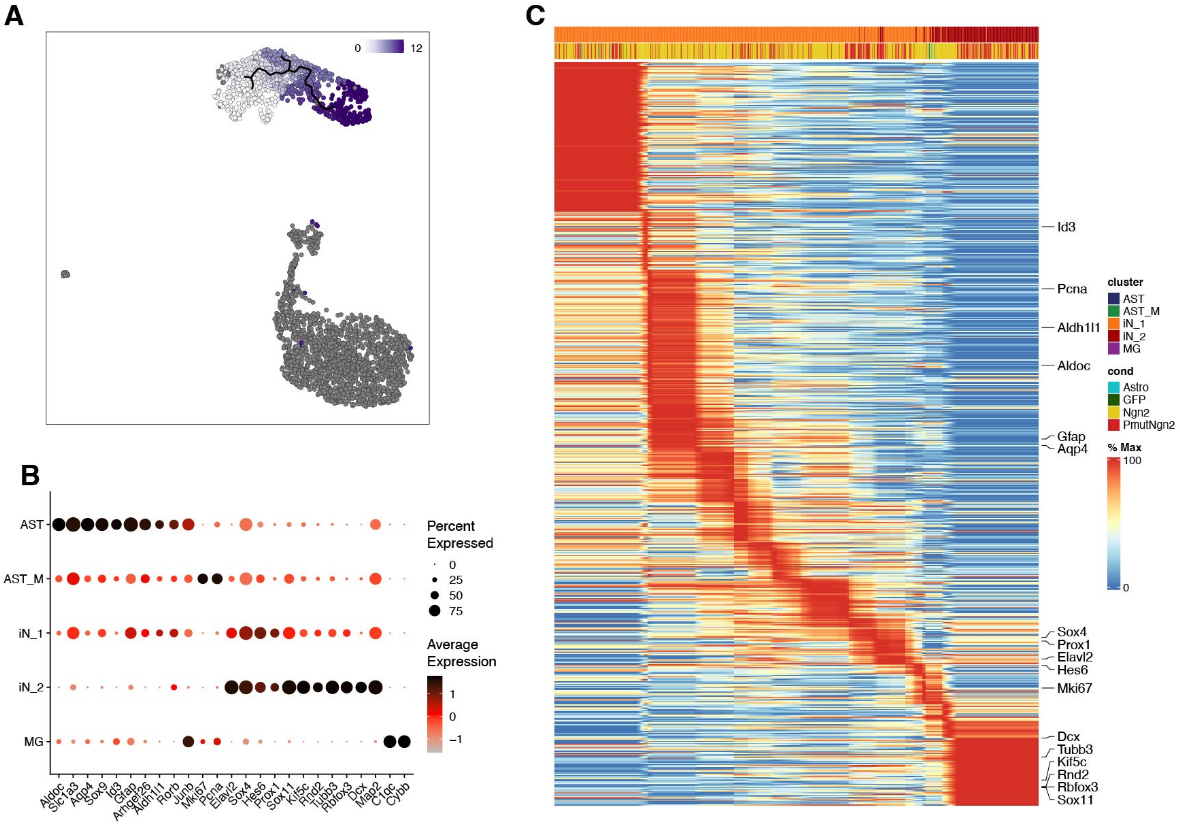

Figure 16 iN_2 cells present more maturation than iN_1

A - UMAP visualization alongside pseudotime progression of neuronal maturation
B - Dot plot illustrating the proportion of cells (represented by dot size) and the expression levels of selected marker genes (indicated by colour) across the respective cell type clusters.
C - Heatmap of the expression levels of the most variable genes across maturation pseudotime.

To further understand the difference between the iNs induced by Ngn2 and PmutNgn2, bulk RNA-seq data was generated at the same time point for pairwise comparison. The Ngn2 iNs retained a more astrocytic transcriptional profile, indicated by the higher expression of astrocytic genes such as *Sox9*, *Gfap* and *Aldoc*. Whereas PmutNgn2 iNs were marked by upregulated neuronal maturation markers (*Reln* and *Brsk2*), neurogenic TFs (*Bhlhe22)* and TF co-factors/chromatin regulators (*Yy1*[208]*)* (Fig. 17). Thus, PmutNgn2 accelerates both the conversion process and the maturation of the iNs.

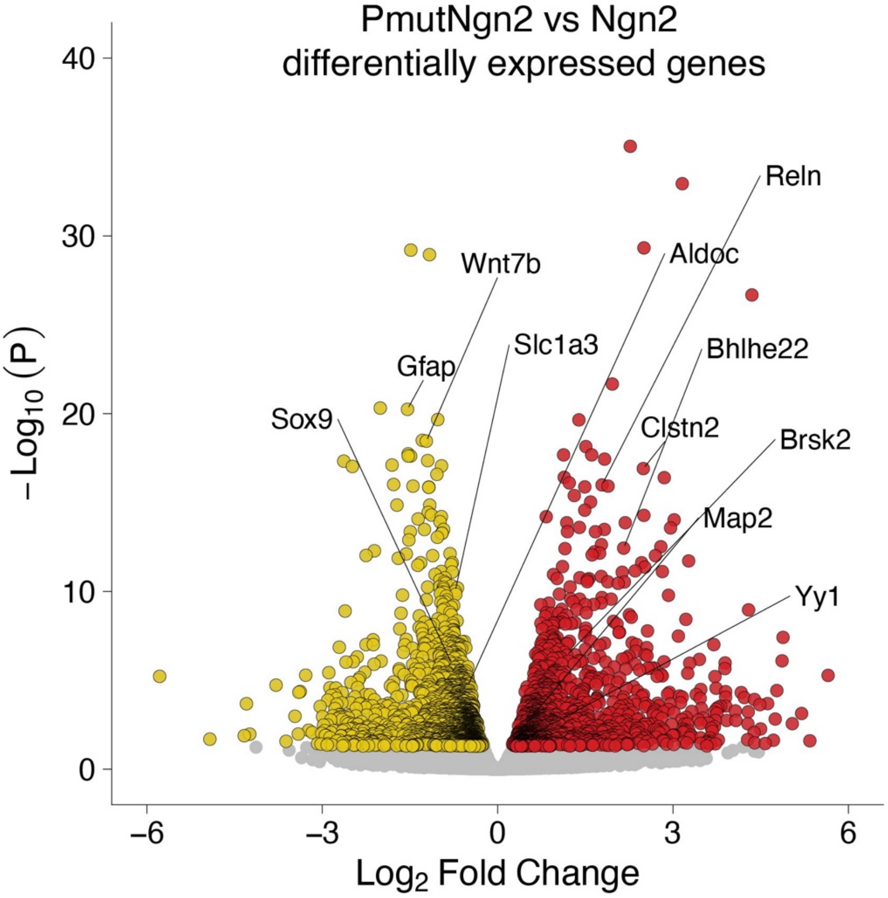

Figure 17 Volcano plot of PmutNgn2 vs Ngn2 differentially expressed genes

Volcano plot showing differentially expressed genes (FDR < 0.05) from a pairwise comparison between PmutNgn2 and Ngn2 using bulk RNA-seq (n = 3).

3.1.4 Ngn2 modulates chromatin accessibility at regulatory sites

Having identified the TF-specific differences in transcriptional profile, we analysed the chromatin accessibility data from the 10x multiome to understand the underlying mechanisms that may contribute to the variation. Using pseudobulk aggregation of the single-cell data, we assessed the differential accessibility at promoter and distal regions for the experimental conditions, clustered into 5 groups (Fig. 18A). Between Ngn2 and PmutNgn2, there are many shared accessible sites corresponding to neuronal sites. However, there are also many sites that are differentially more accessible specifically in the PmutNgn2 cells.

TF motif analysis of the differentially accessible sites annotated by the five clusters revealed that for the control astrocytes there is an enrichment of astrocyte TFs, such as Tead3 and Rfx4 (Fig. 18B). In contrast, clusters 4 and 5, which exhibited greater accessibility in iNs, contained motifs of established neurogenic regulators like NeuroD2, Meis2, and Tcf12, alongside the previously unreported Tgif2 motif (Fig. 21). Uniquely accessible distal sites upon PmutNgn2 induction were enriched for additional E-box protein motifs, such as Tcf3 and Tcf12[209], known to heterodimerize with Ngn2 to bind and transactivate target genes.

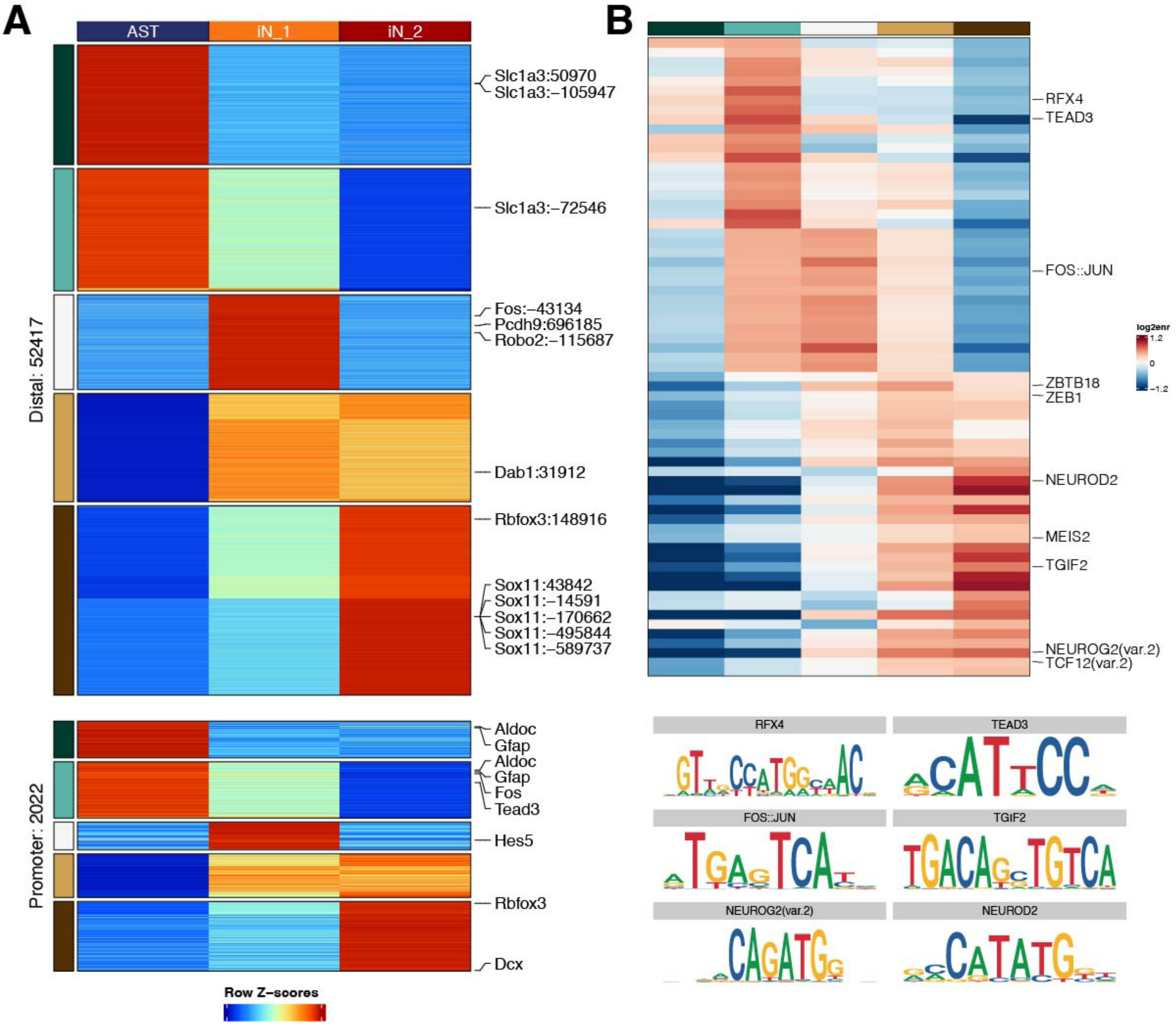

Figure 18 Differential accessibility across experimental conditions and enriched TF motifs

A - Heatmap of pseudobulk differential accessibility (z-score) across experimental conditions at distal and promoter regions

B - Heatmap of log2 fold enrichment of TF motifs in the five clusters identified by k-means clustering between conditions at distal and promoter regions.

To assess the effects of Ngn2 and PmutNgn2 at native sites, we investigated the NERUOG2 motif (Fig. 19). At this motif, we identify greater accessibility in PmutNgn2 induced cells relative to Ngn2 expressing cells. This is more pronounced when comparing the accessibility of the NEUROG2 motif at the more mature iN_2 cluster relative to iN_1, defined by the UMAP cell clustering (Fig. 14). To link the chromatin accessibility dynamics with gene expression changes more broadly in the cell clusters, we identified 7,917 positively correlated enhancer–gene pairs (EGPs)[136,210] (Fig. 20). These EGPs include known direct targets of Ngn2, such as Rbfox3 and Cplx2. Collectively, our findings demonstrate that Ngn2 drives extensive chromatin remodelling, which is further enhanced by PmutNgn2.

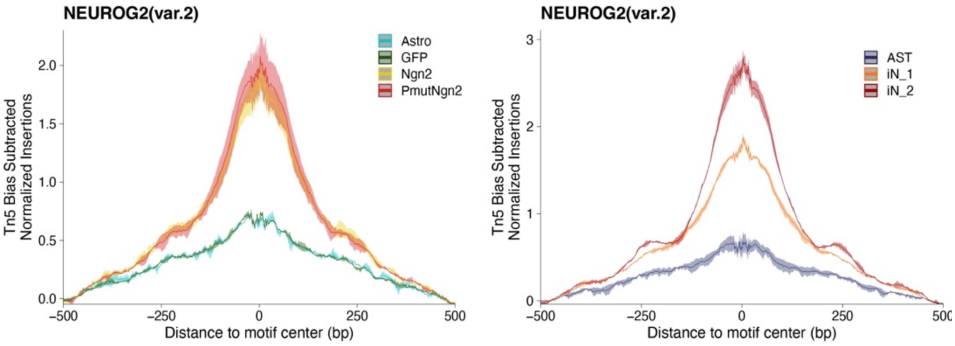

Figure 19 NEUROG2(var.2) motif accessibility.

Motif footprint normalised for Tn5 insertion bias at sc-ATAC peaks for experimental condition or cell type clusters.

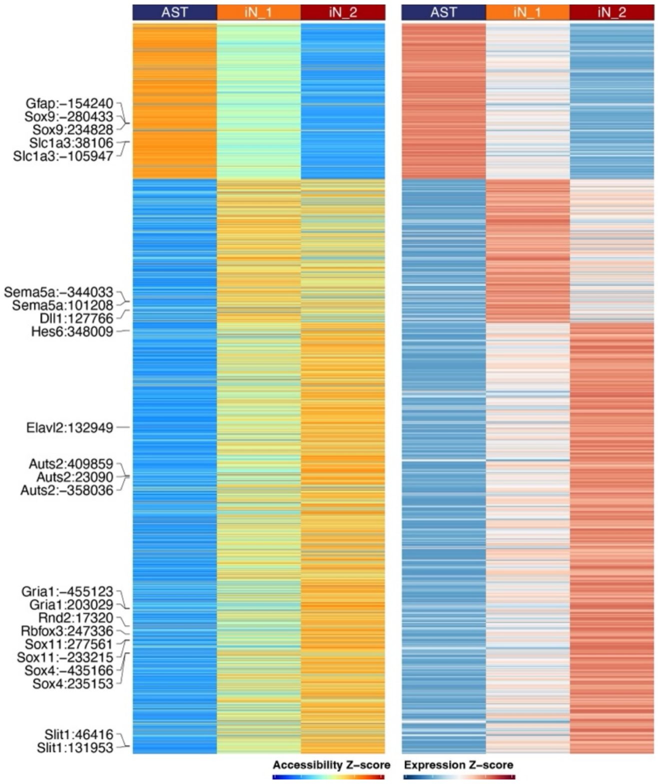

Figure 20 Heatmap of positively correlated EGPs

Heatmaps displaying aggregated accessibility of putative enhancer elements (left) alongside the expression levels of their associated genes (right) for positively correlated EGPs.

3.1.5 Ngn2 and PmutNgn2 exhibit distinct binding patterns to remodel chromatin

To more accurately understand the primary chromatin rewiring mediated by the TFs, we performed chromatin immunoprecipitation sequencing (ChIP-seq) for both the Ngn2 and PmutNgn2. We identified 25,352 shared peaks, but also 5,655 Ngn2-specific and 20,552 PmutNgn2-specific peaks (Fig. 21A). Interestingly, even at shared peaks, the PmutNgn2 showed stronger enrichment.

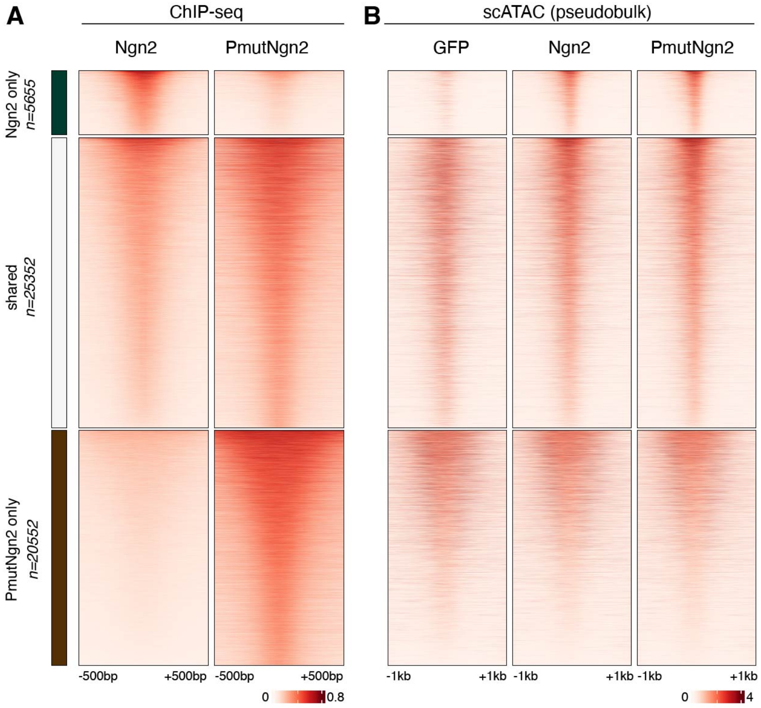

Figure 21 Differential TF binding and associated chromatin accessibility.

Heatmaps illustrating the enrichment of ChIP-seq **(A)** or pseudobulk single-cell ATAC-seq **(B)** signals around differentially bound or shared peaks in murine iNs.

By comparing the changes in accessibility associated in TF binding relative to the control, we find that Ngn2-specific peaks are actively opened upon induction (Fig. 21B). We also find some Ngn2-specific sites being opened in the PmutNgn2 condition which we attribute to the upregulated expression of the endogenous *Ngn2* as seen in our bulk RNA-seq data in this condition (Fig. 22). On the other hand, shared peaks are already lowly accessible in astrocytes and increase in accessibility upon forced TF expression. In contrast, most PmutNgn2-specific peaks were pre-accessible in astrocytes and retained accessibility throughout the reprogramming process.

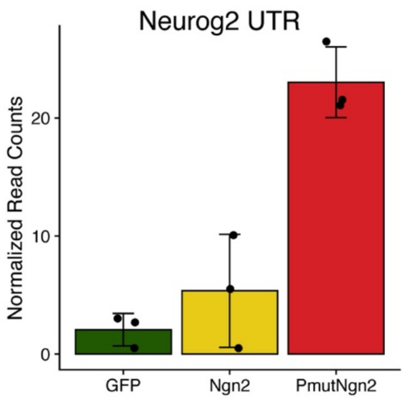

Figure 22 Endogenous Ngn2 expression levels

Barplots with mean ± s.d. showing the normalized read counts mapping to either the 5’ or the 3’UTR of the endogenous Ngn2 locus. Individual biological replicates (n = 3) are represented as dots.

We analysed the TF-binding motifs in each peak group to help understand the differential binding pattern observed between Ngn2 and PmutNgn2 (Fig. 23). Whilst Ngn2-specific and shared peaks were enriched for the Ngn2 motif itself; no specific motif was identified in the PmutNgn2-specific peaks.

We further investigated if the number of Ngn2 motifs present influences binding with respect to the initial accessibility of the site prior to induction as observed in the GFP-transduced astrocytes (Fig. 24A). We find that at both Ngn2 only and Shared peaks, the

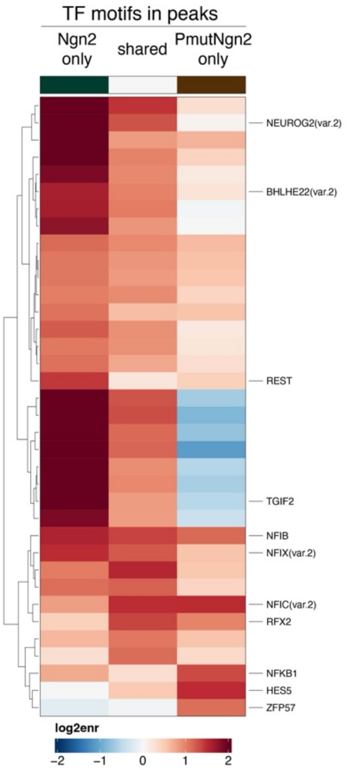

Figure 23 TF motif enrichment in binding peak groups

Heatmaps depicting the motif enrichment in the peak groups identified in Fig. 21.

presence of more Ngn2 motifs in the peak facilitates opening of more closed chromatin, suggesting a synergy amongst the bound proteins. Whereas less motifs are needed for opening if the site is already lowly accessible in astrocytes. PmutNgn2 only sites do not such a relationship between motif number and binding, likely as a result of PmutNgn2-specific binding site preferentially being found in promoter regions, which are largely accessible across cell-types[16,17] (Fig. 24B).

However, binding by Ngn2 or PmutNgn2 both predominantly resulted in gene activation. This is most prominent at Ngn2-specific peaks possibly due to the predominantly distal binding sites bound that could be instructive for the subsequent gene activation (Fig. 24C). Consistent with the observed chromatin accessibility changes, Ngn2-specific genes were also upregulated in the PmutNgn2 condition, likely influenced by endogenous Ngn2 activity.

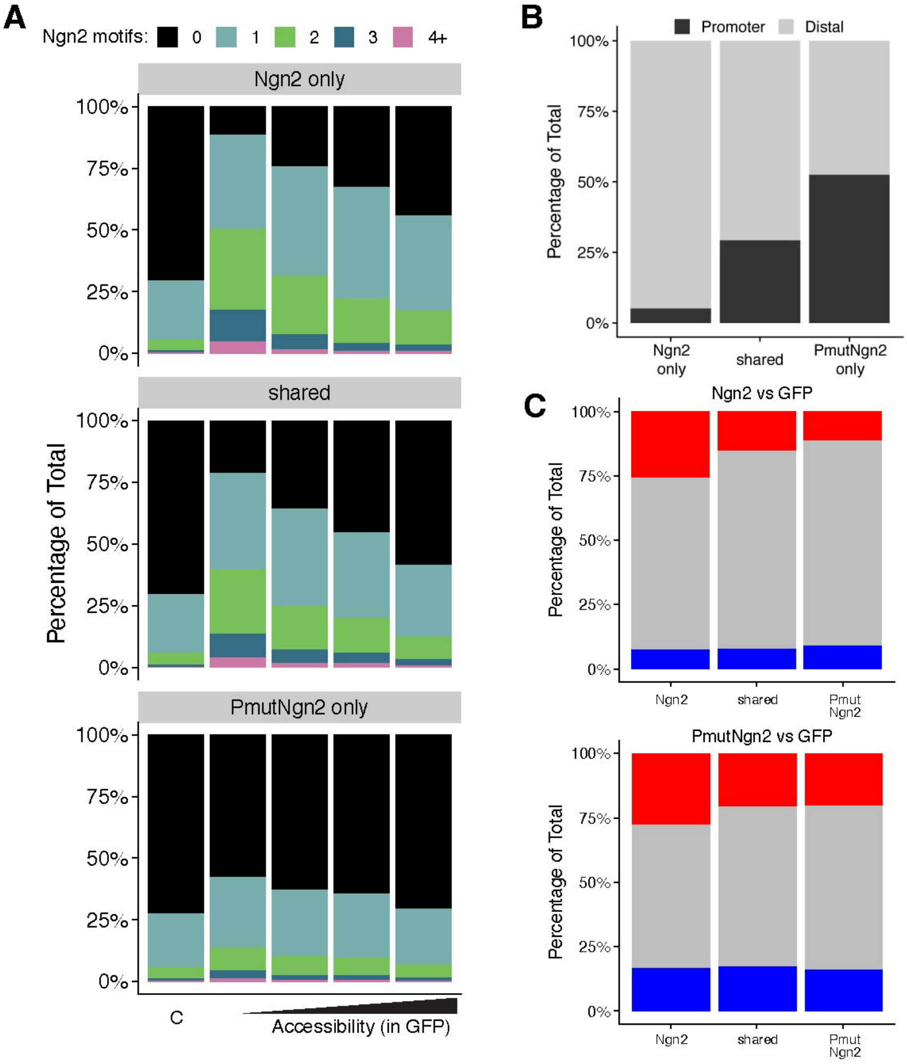

Figure 24 TF binding relationship with number of motifs and gene regulation

A - Bar plot representing the number of Ngn2 motifs across different peak categories, stratified by chromatin accessibility in GFP-transduced astrocytes. The control group ('C') includes randomly sampled accessible regions not bound by Ngn2 or PmutNgn2, serving as a baseline.
B - Percentage overlap between peaks and gene promoters (± 5 kb from TSS).
C - Percentage of differentially regulated genes, identified from bulk RNA-seq, overlapping with various peak categories within a ±100-kb window around the transcription start site (TSS).

3.1.6 PmutNgn2 enhances 3D genome and DNA methylation changes

For examining the 3D genome and DNA methylation remodelling, we utilised a modified version of Methyl-HiC, resulting in good coverage of Hi-C contacts and DNA methylation[115]. The experiment was performed on cell-cycle sorted G0/G1 cells to minimise the effect on cell-cycle on the 3D genome. Upon investigating the contact probability as a function of genomic distance in a cis-decay plot, we identify a global reorganisation of the 3D genome upon reprogramming, showing increased short-range and decreased long-range interactions (Fig. 25). This shift is more pronounced in the PmutNgn2 condition, more closely resembling the shift as seen in *in vivo* mouse cortical neurons[136].

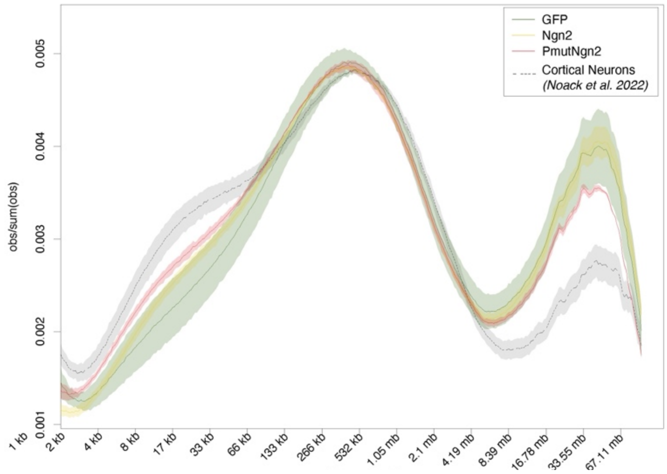

Figure 25 Cis-decay plot of experimental conditions

Contact probability as a function of genomic distance: Lines represent mean values from biological replicates, while semi-transparent ribbons indicate the standard error of the mean (s.e.m.). Observed contact probabilities are normalized by dividing the observed values by the sum of observations (obs/sum).

Reprogramming is also associated with stronger compartmentalisation with more marked interactions within more inactive B compartments, with the effect once again more evident in the PmutNgn2 condition (Fig. 26A). Furthermore, the increased insulation at TAD boundaries associated with the reprogramming is enhanced by the PmutNgn2 condition relative to Ngn2 (Fig. 26B), however, this difference is not associated with a change in the number of TADs. In contrast, the number of loops decrease upon reprogramming, with more loops lost in the PmutNgn2 condition (Fig. 26C). Of the loop anchors identified for each condition, there is an increased ratio of the loops that are bound by Ngn2/PmutNgn2 as reprogramming advances (Fig. 26D).

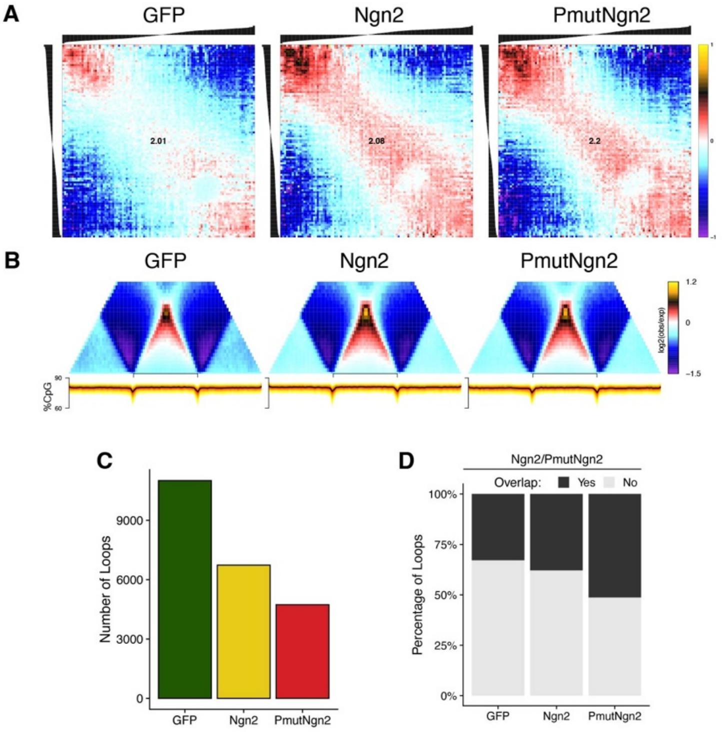

Figure 26 3D Genome changes associated with reprogramming

A - Average contact enrichment between 250-kb loci pairs, ordered by their eigenvalue (shown above). Numbers indicate compartment strength.
B - Average contact enrichment (top) and DNA methylation levels (bottom) across TADs.
C - Barplot showing the number of chromatin loops per condition (FDR < = 0.1)
D - Stacked barplots showing the percentage overlap between Ngn2/PmutNgn2 peaks and loop anchors

Next, we assessed the contact enrichment at intra-TAD of the top 5000 Ngn2-bound sites in each experimental condition (Fig. 27A). The contact frequency between these bound sites is increased upon expression of the reprogramming factor, with the interactions being stronger in the PmutNgn2 condition and quantified to be significant. The same trend was observed by using the top 5000 PmutNgn2-bound sites (Fig. 27B). However, in the control GFP condition, there is a higher starting interaction frequency amongst PmutNgn2 sites likely due to a larger proportion of bound regions being promoters.

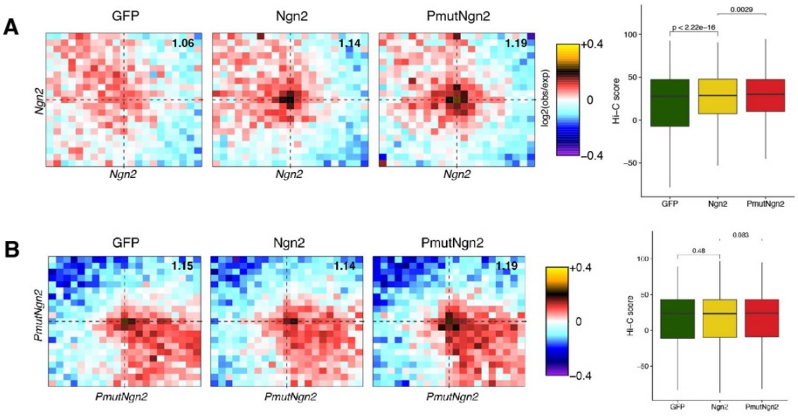

Figure 27 Contact frequency at TF bound sites

A- Aggregated Hi-C plots for intra-TAD pairs of the top 5,000 Ngn2 ChIP-seq peaks and associated quantification
B- Same as A but for PmutNgn2 ChIP-seq peaks

Similarly, using the top 5000 distally bound Ngn2 sites to investigate methylation changes, we identify that DNA methylation is reduced upon binding and the demethylation is stronger in the PmutNgn2 condition, quantified to be significant (Fig. 28A). The same was observed utilising the top 5000 distally bound PmutNgn2 sites (Fig. 28B). Many of the Ngn2-bound sites were already hypomethylated in the GFP condition, but upon Ngn2 or PmutNgn2 binding the sites are further demethylated.

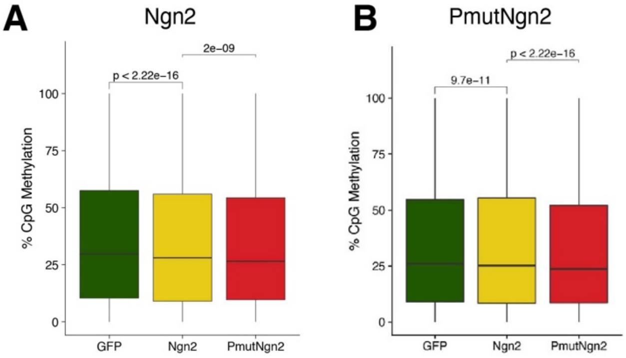

Figure 28 DNA methylation changes at TF bound sites

A - Quantification of the average DNA methylation at top 5,000 Ngn2 ChIP-seq peaks
B - Same as A but for PmutNgn2 ChIP-seq peaks

Beyond Ngn2 or PmutNgn2 bound sites, we wanted to understand if there is a global change at regulatory elements corresponding to gene expression changes. Using the previously identified EGPs within the associated clusters, we examined the enhancer-promoter (E-P) contract strengths (Fig. 29A). The positively correlated E-P pairs were marked by increased contact strength in the AST, iN_1 and iN_2 clusters. The contact strength of the E-P pairs was highest in the condition contributing most of the cell type cluster. However, the E-P contact strength between the Ngn2 and PmutNgn2 conditions were not significantly different. This indicates that the maturation of neurons marked by differential changes in accessibility and subsequently gene expression can be independent of strength of chromatin looping.

Unlike chromatin looping, enhancers in EGPs, exhibited greater demethylation in the PmutNgn2 compared to Ngn2, with the larges decreases at enhances in the iN clusters, in line with the previously observed increased chromatin accessibility at these loci (Fig. 29B).

Despite the complex relationship between the epigenetic modalities, we do identify co-ordinated changes in chromatin interaction and DNA hypomethylation for key neuronal

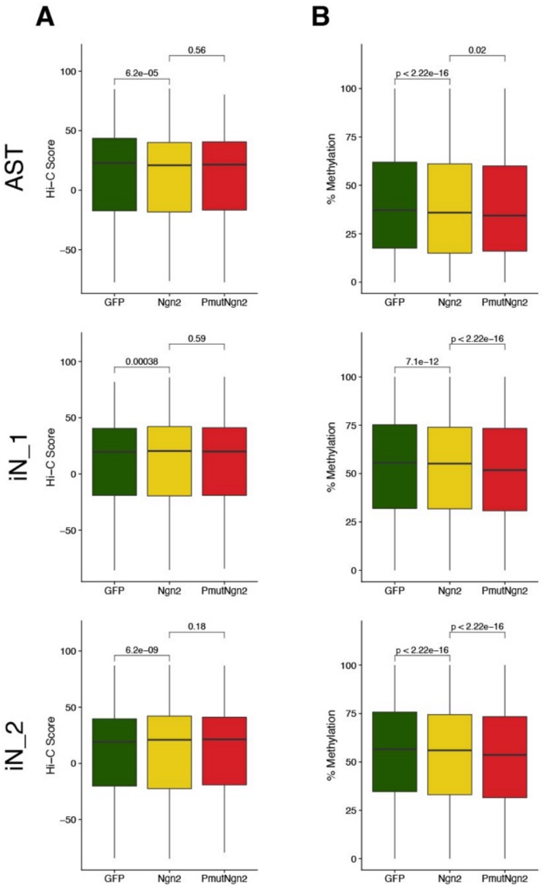

Figure 29 Interaction strength and DNA methylation at positively correlated EGPs

A - Box plots showing the interaction strength of intra-TAD cluster-specific positively correlated EGPs.
B - Quantification of DNA methylation levels at enhancers belonging to cluster-specific EGPs

genes such as *Kirrel3*, *Scna1*, *Plexina2* and *Auts2* as well as known neurogenic TFs such as *Sox11*[115]. Similarly, *Mdga1*, a key player in cell adhesion and synapse formation in the developing brain[211,212], was activated during direct reprogramming and exhibited increased expression throughout neuronal maturation (Fig. 30A). By identifying E-P pairs that are differentially interacting in the iN_2 cluster between PmutNgn2 and Ngn2 condition, as well as differential DNA methylation, we find that *Mdga1's* enhancers strongly loop in with its cognate promoter (Fig. 30B), with some enhancers marked by decreased methylation in the PmutNgn2 condition (Fig. 30C). By looking at the contact map of the locus, we observe new contacts being formed in the PmutNgn2 condition, associated with an increase in accessibility at the marked enhancers (Fig. 30D).

Additionally, in the differential contact and methylation density plots, which reveal increased interactions and decreased methylation specifically in the PmutNgn2 condition, we identify key neuronal maturation factors such as Rbfox3 (which encodes NeuN) and other epigenetic regulators. Notably, Kdm7a, an H3K9 and H3K27 demethylase involved in neural induction, was also identified[213](Fig. 30B/C).

3.1.7 Yy1 contributes to Ngn2-drive epigenetic remodelling during reprogramming

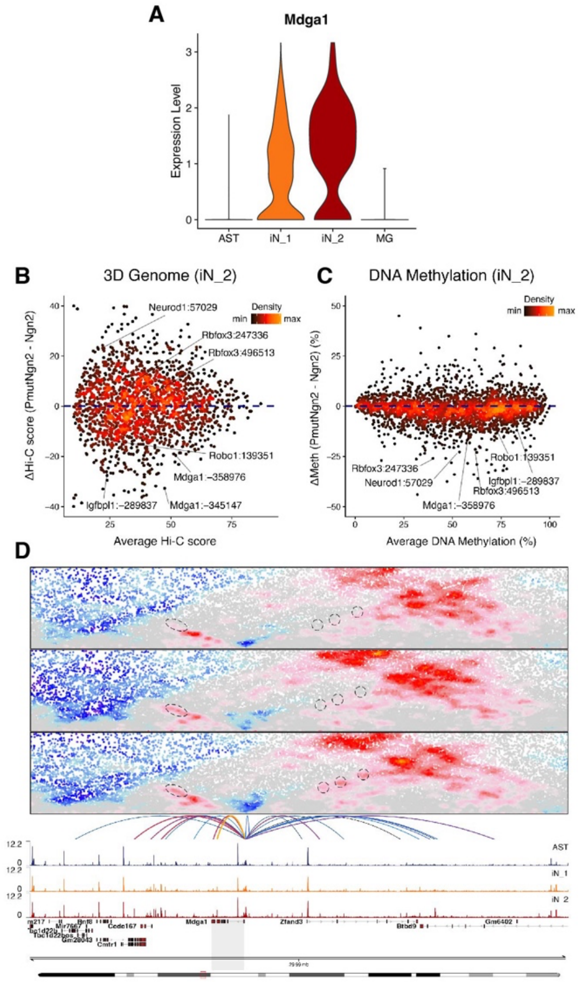

Figure 30 Mdga1 gene and enhancer activation

A - Violin plot showing the expression levels of Mdga1 across the indicated cell-type cluster
B - Density scatter plots showing iN_2 E–P contact strengths
C - Same as C but for DNA methylation
D - Contact map (top) and aggregated accessibility of corresponding single-cell ATAC-seq clusters (bottom) at the Mdga1 locus. Linked enhancers are shown as arcs, coloured by the Pearson correlation between enhancer accessibility and Mdga1 expression. Dashed circles highlight dynamic enhancer-promoter (E–P) interactions, while the shaded region emphasizes the Mdga1 locus along with its associated contact map and aggregated accessibility data.

Although PmutNgn2 exhibits markedly enhanced chromatin remodelling and epigenetic modulation compared to Ngn2, its DNA binding and pioneering activity alone did not fully account for the observed differences. This prompted an investigation into potential co-factors. The previously discussed RNA-seq analysis identified *Yy1* as a differentially upregulated gene in the PmutNgn2 condition (Fig. 19). Yy1 is an epigenetic co-factor known to modulate E-P contact formation to influence gene expression[141,214].

Yy1 CUT&RUN was performed in the control GFP astrocytes and the Ngn2 transduced cells at 2dpi. While the majority of Yy1 peaks were shared across conditions, reprogramming resulted in the loss (GFP-specific) or gain (Ngn2-specific) of a subset of Yy1 sites (Fig. 31A). As Yy1 binding profiles were similar between the PmutNgn2 and Ngn2 conditions, we focused on comparing GFP-specific and Ngn2-specific peaks. Notably, the Yy1 motif was not enriched at Ngn2-specific peaks (Fig. 31B); instead, the Ngn2 motif was enriched at Yy1 peaks uniquely present in the Ngn2 condition (Fig. 31C), implying that Ngn2 actively recruits Yy1 to these loci with most of these Ngn2-specific Yy1 sites overlapped with Ngn2 binding, predominantly at distal regions (Fig. 31D).

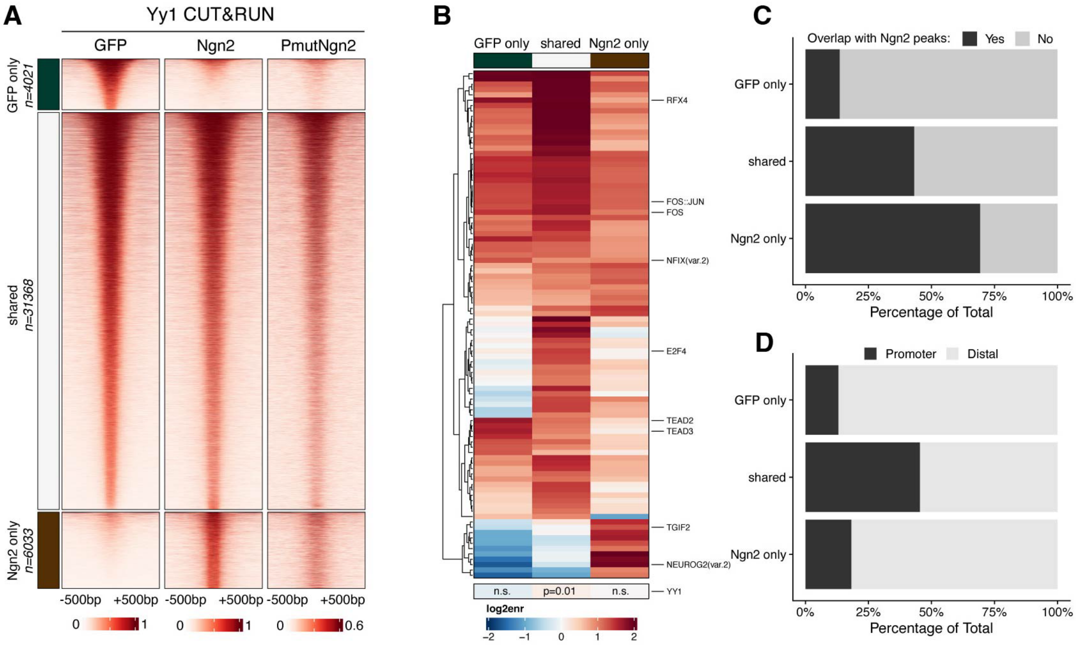

Figure 31 Yy1 and Ngn2 co-binding

A - Heatmaps displaying the enrichment of Yy1 CUT&RUN signal around differentially bound or shared peaks.
B - Heatmap illustrating the TF motif enrichment in the peak groups shown in A.
C - Percentage overlap between Ngn2 peaks and either differentially bound or shared Yy1 peaks.
D - Overlap of differential or shared Yy1 peaks with genomic features.

To address if Ngn2 and Yy1 synergistically function to open chromatin, we studied the top 5000 Ngn2 peaks overlapping differential or shared Yy1 or not bound by Yy1 (Fig. 32A). Relative to the peaks not bound by Yy1, there is increased accessibility at some Ngn2 peaks upon reprogramming. In contrast, co-binding does not seem to influence DNA methylation (Fig 32B). However, the intra-TAD contact frequency analysis of Ngn2 ChIP-seq pairs, categorised based on whether the bound sites overlap with Yy1, reveals a co-operative increase in chromatin looping (Fig. 3C). To determine if this synergy is driven by a direct interaction between Ngn2 and Yy1, we conducted co-immunoprecipitation (IP) in P19 cells (Fig. 33B). Notably, Yy1 was able to pull down both Ngn2 and PmutNgn2.

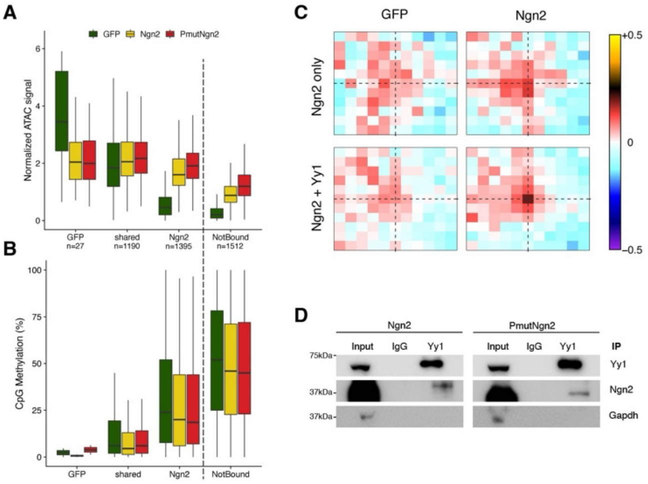

Figure 32 Yy1 and Ngn2 interaction and associated epigenetic changes

A - Box plots showing chromatin accessibility at the top 5,000 Ngn2 peaks, categorized by whether they overlap differential or shared Yy1 peaks, or are not bound by Yy1.
B - As in A, but quantifying DNA methylation at these regions. The box plots display the median (line), 25th and 75th percentiles (box), and 10th and 90th percentiles (whiskers).
C - Aggregated Hi-C plots of intra-TAD pairs from the top 5,000 Ngn2 ChIP-seq peaks, separated based on whether they overlap with Yy1 (Ngn2+Yy1) or not (Ngn2 alone).
D - Co-IP experiments in P19 cells using IgG or Yy1 antibody for pulldown, followed by staining for Yy1, Ngn2, and Gapdh

The functional relevance of the Yy1-mediated enhancement of Ngn2-driven reprogramming remained to be determined. Therefore, a Yy1^{tm2Yshi} homozygous mice mouse was used to conditionally knockout (KO) Yy1 in primary astrocytes in culture by using a Cre-expressing lentivirus, and then subsequently initiating reprogramming by Ngn2 induced expression (Fig. 33A). At 7dpi, Yy1 KO astrocytes displayed a significantly reduced capacity to reprogram, with only 13% of iNs (βIII-tubulin^{+}/GFP^{+}RFP^{+}) generated relative to the 37.1% iNs in the Yy1 WT/Ngn2^{+} condition (Fig. 33B). By performing sc-RNA-seq, it appeared that in the KO condition, that the mature iN_2 and iN_3 clusters identified have strongly reduced proportions (Fig. 33C), associated with an increased stress response. Thus, suggesting that the Yy1 KO affects the transition of early iNs to a mature state. Overall, this suggests that Ngn2 directly recruits Yy1, which plays a crucial role in Ngn2-mediated reprogramming by facilitating changes in chromatin accessibility and looping, but not in DNA methylation.

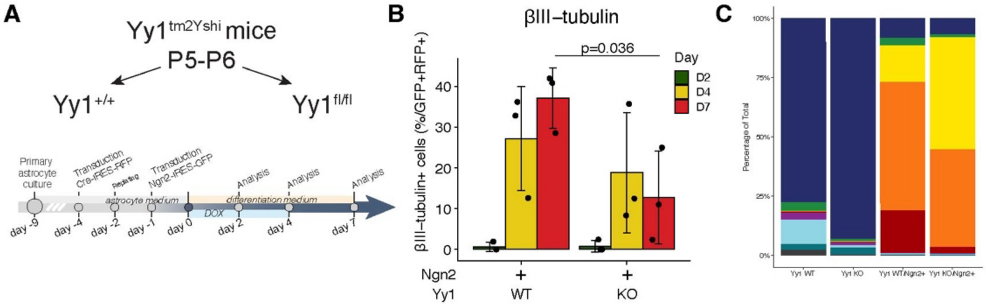

Figure 33 Yy1 KO effect on reprogramming

A – Experimental design
B - Quantification of the proportion of GFP+RFP+ cells positive for βIII-tubulin. Mean ± s.d., with each dot an individual biological replicate (n = 3). Statistical significance determined using linear regression.
C - A stacked bar plot illustrating the proportion of cell types in each experimental condition.

3.2 Epigenome rewiring in a context-dependent manner during neuronal differentiation

The results presented in this section are a product of collaborative work with Vera Manelli (Bonev Lab, PhD student). My major contribution was with the cell-line generation and 3DRAM-seq experiments.
This work is currently a preprint on bioRxiv[148].

3.2.1 Preliminary experimental results

The observed Ngn2(Neurog2)-mediated chromatin remodelling during neuronal reprogramming raised questions about the dynamic interplay among the various epigenetic layers and gene regulation. To investigate this, we generated a stable cell line in A2Lox mESCs by integrating Neurog2 under the control of a doxycycline-inducible Tet Responsive Element (TRE) at a safe harbour locus using an inducible cassette exchange system[196] (Fig. 34A). Utilising such a system provided tight control of Neurog2 expression for subsequent studies. We validated multiple clones of the Neurog2-A2Lox cell line, following neomycin selection using dox treatment and identified a strong upregulation of Neurog2 by qPCR and IF (Fig. 34B).

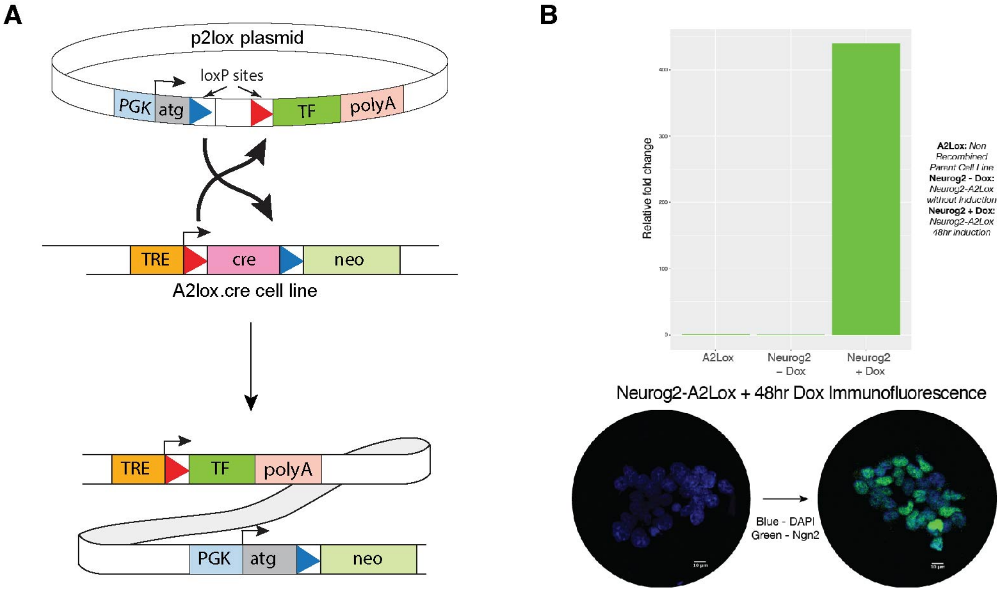

Figure 34 A2Lox cell line generation

A – Inducible cassette exchange experimental schematic. Adapted from Iacovino et al.,Stem Cells, 2011[210].
B - Preliminary classification of Neurog2 upregulation with RT-qPCR and IF.

Embryoid body (EB)-based neuronal differentiation driven by Neurog2 overexpression has been previously utilised to examine TF binding and subsequent chromatin opening[135]. However, these studies were limited to two time points and did not explore alterations in the 3D genome or DNA methylation. Adopting a similar framework, we generated embryoid bodies from the Neurog2-A2Lox cell line. The EBs were treated with dox to induce Neurog2 expression in a time course spanning 0h, 3h, 6h, 9h, 12h, 24h, and 48h. Additionally, a wash condition was included, where dox treatment was halted after 12h, followed by a 36h withdrawal period (Fig. 35). Each time point was paired with an uninduced control. The harvested cells underwent qPCR analysis to examine expression levels of known Neurog2 target genes, UMI-4C[197] to investigate 3D genome changes using target gene promoters as viewpoints, and Bisulphite amplicon sequencing[198] to assess DNA methylation at target gene enhancers.

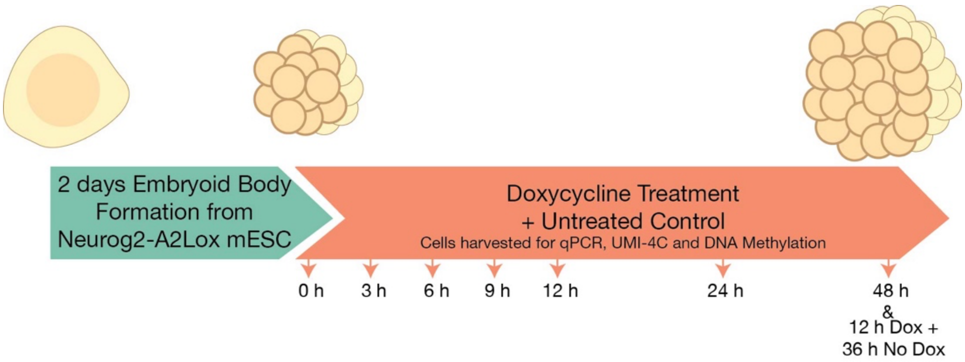

Figure 35 Experimental Schematic for EB differentiation

qPCR showed a strong upregulation of Neurog2 as early 3 hours post induction, increasing in expression up to the 12h time point (Fig. 36). Levels of Neurog2 are 24h and 48h are likely lower due to the impermeability of dox into the growing EBs. The 12h wash condition showed complete loss of Neurog2. Neurog2 target genes showed varying dynamics of induction; Dll1 induction closely resembles Neurog2 dynamics,

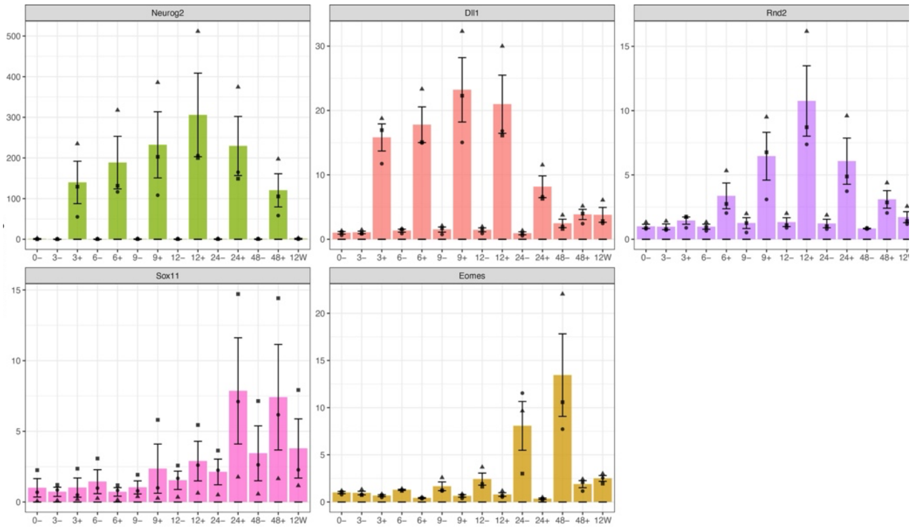

Figure 36 qPCR of Neurog2 target genes in EBs

Barplots showing the relative fold change normalised to expression levels in the no dox condition at 0h. Each shape is a biological replicate (n=3) and the error are stand errors of the mean.

whereas Rnd2 is induced following 6h of treatment and Sox11 at 24h, with the wash condition for all conditions returning the gene expression back close to baseline levels. Interestingly, Eomes was not induced by Neurog2, but seemingly upregulated in the minus dox condition. Neurog2 has been suggested to have a dosage dependent effect, with high expression potentially leading to skipping of certain transitionary states[215], which may contribute to the lack of Eomes upregulation. However, it remains unclear how Eomes is induced without Neurog2 expression.

Despite the changes in expression of three of four target genes tested, UMI-4C identified no alteration in the chromatin architecture between the No Dox and Plus Dox conditions across all genes (Fig. 37). We attribute this result to a lack of sensitivity of the method, as subsequent high-throughput profiling (explored later) indeed identified Neurog2 mediated changes in 3D genome in the mESCs.

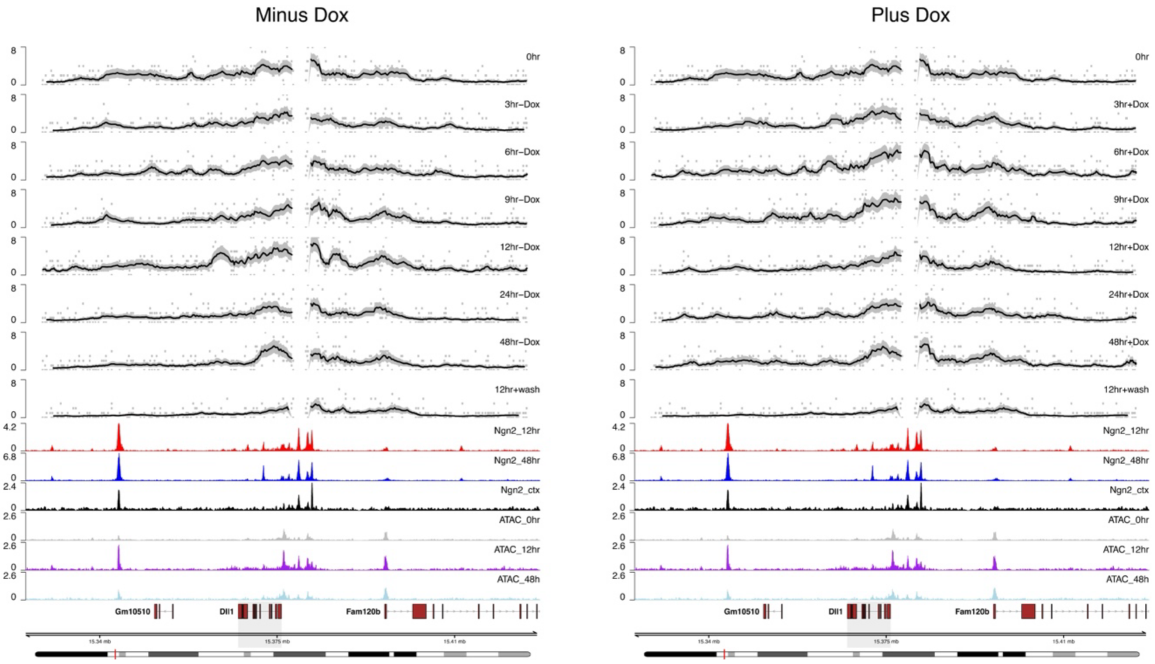

Figure 37 UMI-4C at Dll1

UMI-4C for Dll1 with a promoter viewpoint at no dox (left) and plus dox (right) conditions. Ngn2_12hr and 48hr are ChIP-seq peaks identified in Aydin et al. EB differentiation, and the corresponding ATAC peaks at 0hr, 12hr and 48hr from the same study.

However, Bisulphite-seq amplicon sequencing revealed dynamic methylation changes at two enhancers of Dll1 (Fig. 38). One enhancer, initially hypomethylated, showed further demethylation at 12h, while the other, hypermethylated initially, underwent significant demethylation at 24h. Notably, in the wash condition, methylation levels did not fully revert to minus-dox levels, suggesting active demethylation but passive incomplete remethylation which is still sufficient to downregulate gene expression.

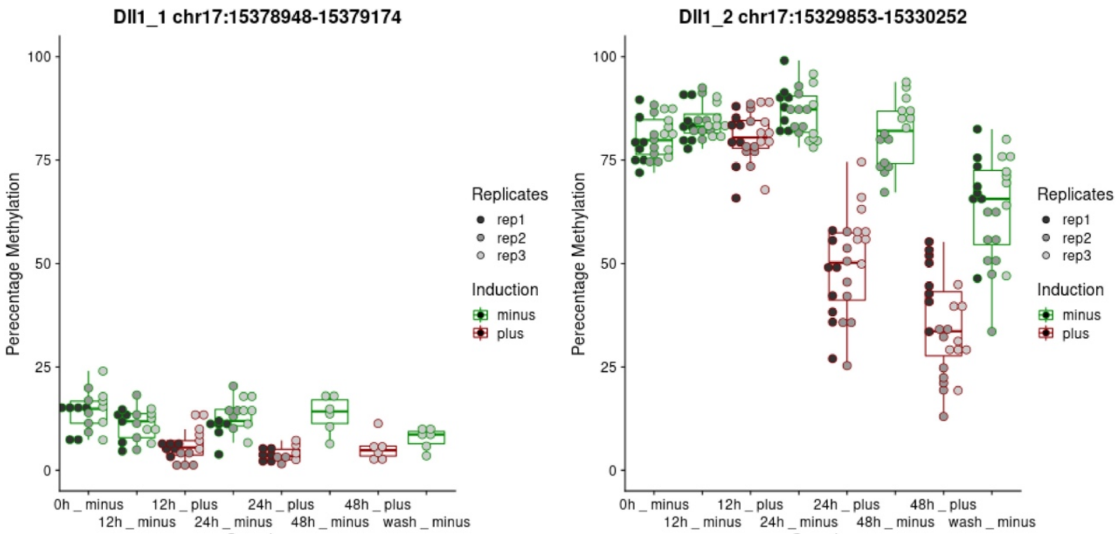

Figure 38 Bisulphite amplicon sequencing at Dll1 enhancers

Whisker box plots showing bisulphite amplicon sequencing at two Dll1 enhancers. Minus dox – green, Plus dox – red. Each dot is a cytosine in the amplicon. n = 3.

Although we find interesting dynamics of gene expression and DNA methylation in this experimental model, some key drawbacks existed. Given the size of EBs at 24h and 48h, the dox is likely impermeable to all cells, resulting in a heterogenous EB population, unsuitable for bulk studies. Furthermore, the lack of sensitivity of UMI-4C resulted in no information obtained regarding the 3D genome. Thus, we adjusted the experimental setup to explore a different question: how cellular context influences the remodelling driven by Neurog2.

3.2.2 Experimental setup for investigating context-dependent activity of Neurog2

To explore the context-dependent epigenetic rewiring initiated by Neurog2, we examined two distinct cell types: neural progenitor cells (NPCs), representing a native environment for Ngn2, and mouse embryonic stem cells (ES), a context more divergent from its natural setting but highly plastic. We generated an A2Lox based ES cell line with a stable integrated Neurog2-FLAG. The addition of the FLAG facilitated subsequent imaging and studies on protein-protein interactions with reliable antibodies. This cell line was differentiated to generate NPCs (Fig. 39). The uninduced ES and NPC provided controls for the 24h dox induced Neurog2 expressing cells in each cell type.

The changes in epigenetic layers upon induction of Neurog2 in these two cell types were primarily investigated using a modified 3DRAM-seq(3D genome, RNA, Accessibility and Methylation sequencing) protocol[191]. To obtain increased resolution on chromatin loops, we incorporated techniques used in Hi-C3.0[169], by the use of two fixatives, PFA and DSG, as well as two restriction enzymes for the digestion, DpnII and DdeI. The double fixation reduces noise due to spurious ligations, with the double digestion generating smaller fragments for increase resolution at shorter ranges[67]. Furthermore, unlike the original 3DRAM-seq, rather than fixing the entire sample prior to RNA isolation, a subset of cells was taken for RNA extraction directly. Additional ChIP-seq and ChIP-Mass Spectrometry (ChIP-MS) experiments were conducted to identify the binding profile of the TF and cell-type specific differences in co-factor repertoire.

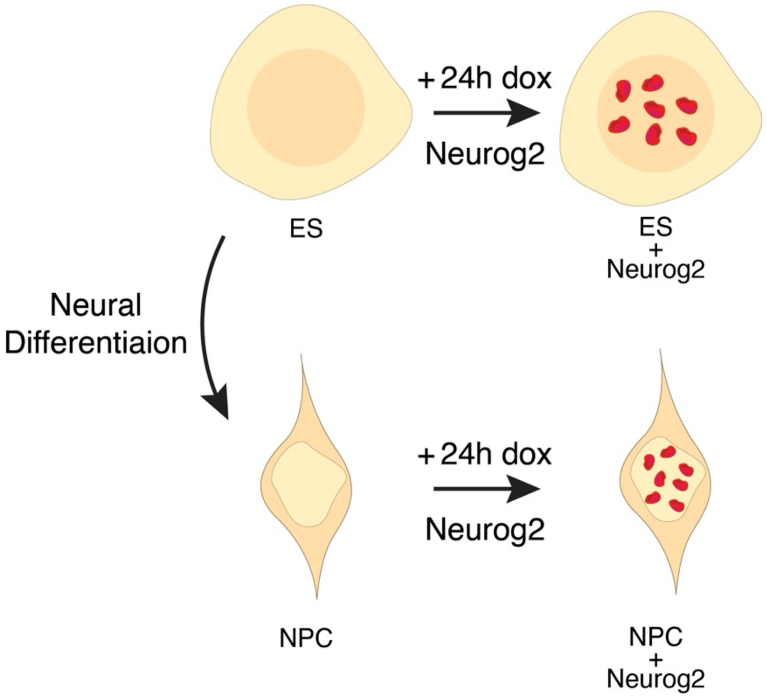

Figure 39 Experimental setup to study context dependent activity of Neurog2

3.2.3 Neurog2 triggers shared and cell-type-specific transcriptional changes

RNA-seq was performed in both ES and NPCs, along with their Neurog2-induced counterparts. We observe that Neurog2 is more strongly upregulated in NPCs upon induction (Fig. 40A), however, Neurog2 protein levels are higher is ES cells (Fig. 40B). Investigating cell type specific markers suggested that upon Neurog2 induction there is a slight loss of pluripotency markers in ES, most prominently seen with Klf4 (Fig. 40C). In contrast, NPC markers remain unaltered but show a gain in neuronal markers such as ßIII-tubulin (Tubb3) and Rbfox3.

A 4-way comparison of the differentially expressed genes (DEGs) revealed that most DEGs are common between induced ES cells and NPCs showing upregulation of Neurog2 and known targets Dll1, Rnd2, Lhx2, Neurod2, Neurod1 and Neurod4[135,136,216–218] (Fig. 40D). In ES specifically we see an upregulation of transcriptional repressors

Id3[219] and Snai2[220], with Id3 known to form heterodimers with bHlH TFs to prevent binding to DNA. Genes upregulated specifically in induced NPCs include Stmn2, which encodes a protein that stabilizes microtubules and plays a crucial role in neurite extension[221], and Neurod6, which is primarily expressed in certain differentiated pyramidal neurons within the deeper cortical layers[222].

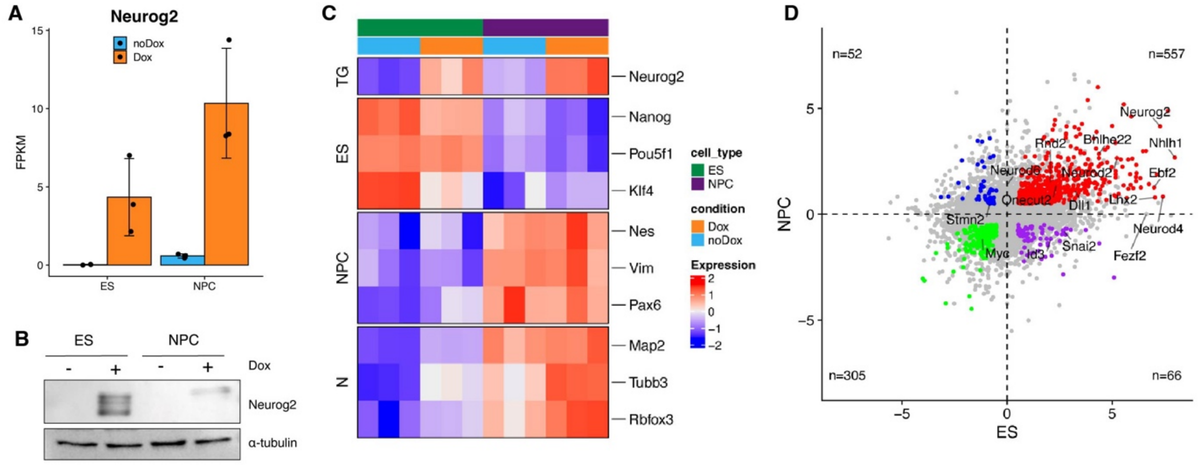

Figure 40 Neurog2 induced transcriptional changes in ES and NPC

A - Expression levels of Neurog2 following Dox induction, represented as dots for FPKM values from individual biological replicates, with error bars indicating standard deviation.
B - Representative western blot of whole-cell lysates showing Neurog2 and α-tubulin expression in ES cells and NPCs.
C - Heatmap showing normalized expression levels of ES, NPC, and neuronal (N) markers during differentiation and/or Neurog2 induction (n = 3).
D - Direct four-way comparison of transcriptional changes across cell types and conditions. Coloured dots indicate significantly upregulated (red), downregulated (green), or anti-correlated genes (blue and purple).

3.2.4 Neurog2 binding sites are mostly shared, with some cell-type specific activity

ChIP-seq for Neurog2 with the attached FLAG for capture in ES and NPCs revealed that most binding sites are shared between the two cell types (Fig. 41A). Stronger enrichment is seen in the ES condition, both for ES-specific and shared peaks likely due to the higher protein levels. TF motif analysis of the peak subsets identified Neurog2 motif enrichment amongst shared peaks (Fig. 41B), with ES-specific peaks enriched in Oct4-Sox2 motifs, indicative of Sox2 known to prime the epigenetic landscape in early neuronal differentiation[223]. Whereas NPC-specific peaks were enriched for motifs belonging to TF family Nf1, known to play a role in cortical development[224].

To link Neurog2 binding to transcription, we analysed the overlap of DEGs with shared or cell-type-specific peaks (Fig. 41C). Neurog2 binding mainly activated genes, particularly at cell-type-specific peaks, with effects correlating to binding strength. Interestingly, we also observe that NPC-specific peaks that are weakly bound by Neurog2 have an increased number of downregulated genes. Together, this suggests that the differential binding on Neurog2 in a cell-type specific manner affects the transcriptional regulation of DEGs.

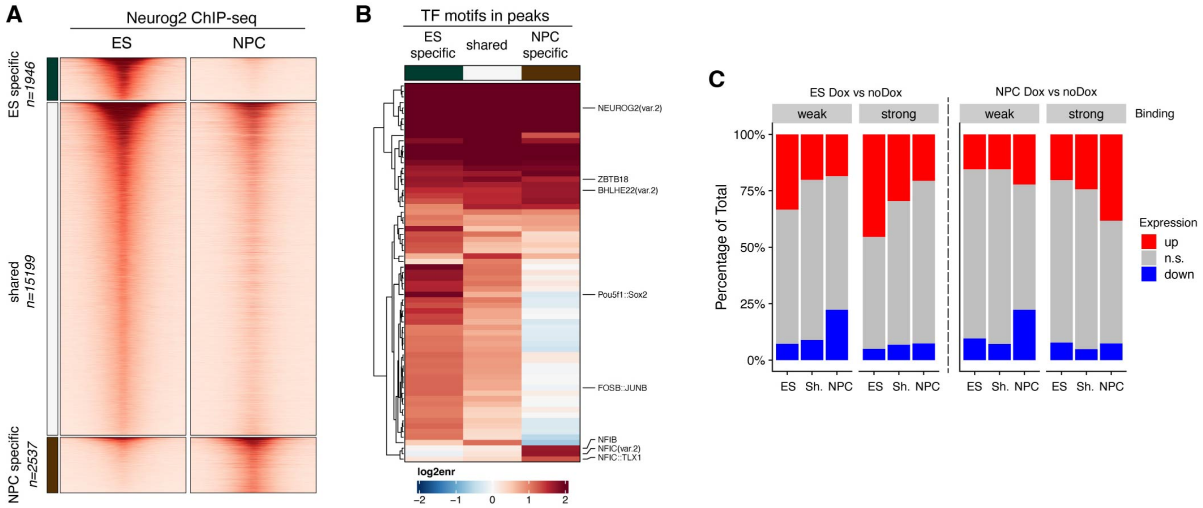

Figure 41 Neurog2 binding in ES and NPC

A - Heatmap showing Neurog2 ChIP-Seq enrichment at shared or differentially bound peaks in ES and NPCs.
B - Heatmaps showing TF motif enrichment analysis for the peak groups presented in A
C – Stacked bar plot showing the percentage of differentially regulated genes with promoters bound by Neurog2 in ES and NPCs, categorized into two quantiles based on Neurog2 binding strength (weak or strong).

3.2.5 Neurog2 mediates direct and indirect effects on chromatin accessibility

3DRAM-seq was subsequently used to obtain the changes in chromatin accessibility, DNA methylation and 3D genome architecture associated with the induction of Neurog2. For both ES and NPCs, cells were FACS sorted for G0/G1 to remove confounding cell cycle effects. In NPC, we additionally selected for a Pax6 positive population to remove inefficiently differentiated cells (Fig. 42).

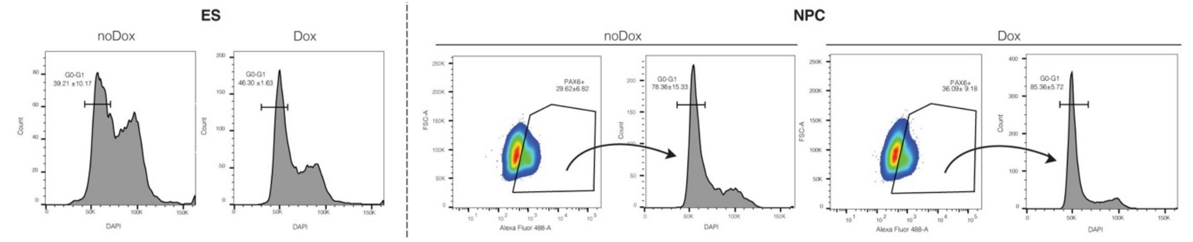

Figure 42 FACS strategy for 3DRAM-seq

Flow cytometry gating strategy for immunoFACS in ES and NPCs. ES cells were gated solely on their cell cycle phase (G0/G1), while NPCs were further selected for neuronal marker, Pax6, expression to minimise differentiation-related heterogeneity. Data represent the mean ± SD from the parent singlet population.

Initially, we looked at the differentially accessibility of Neurog2 peaks at both distal and promoter regions (Fig. 43A). In NPCs, we observed mostly increased accessibility, with the effect strongest at NPC-specific binding sites. Looking specifically at Neurog2 motifs in the top 5000 distal peaks, we also witnessed an increase in accessibility upon induction (Fig. 43B). In contrast, in ES cells alongside many sites that gain accessibility, many regions also lose accessibility at both distal and promoter sites. However, at Neurog2 motifs in distal regions, there is increased accessibility with Neurog2 expression, suggesting that the sites decreasing in accessibility could be a result of indirect effects.

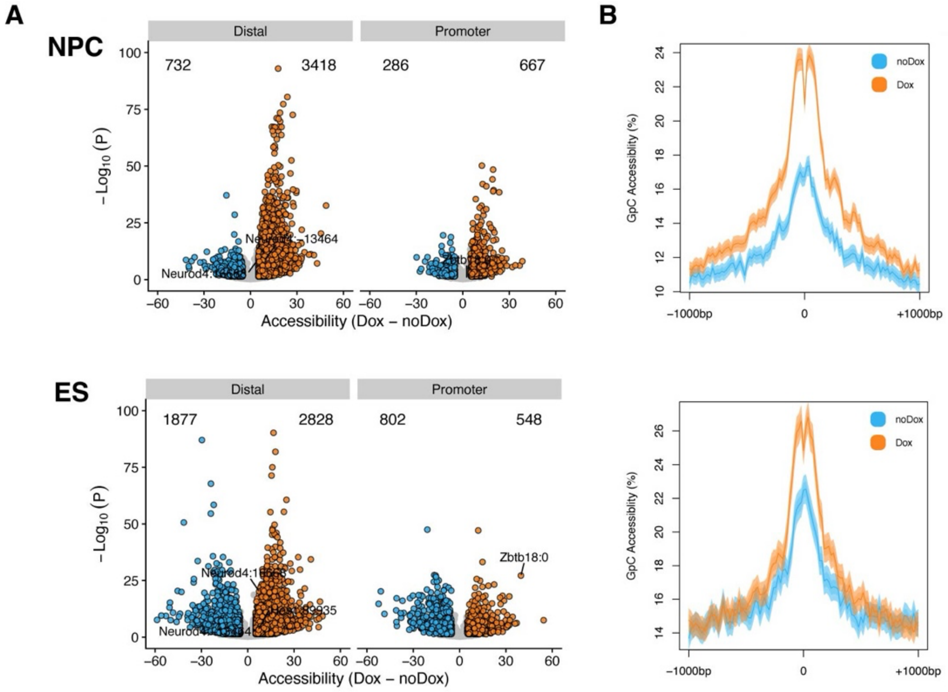

Figure 43 Neurog2 direct effects on chromatin accessibility

A - Volcano plots showing differential accessibility of Neurog2 peaks at distal and promoter regions following Neurog2 overexpression in ES and NPC
B - Neurog2 motif footprinting analysis in the top 5000 distal peaks.

Thus, we next examined global differential accessibility and found that in ES cells, indeed more sites lost accessibility than those that gain accessibility (Fig. 44A/B). To uncover potential drivers of these changes, we analysed TF motifs enriched in peaks categorised by accessibility levels. In differential Neurog2 peaks, ES and NPC sites were predominantly enriched for Neurog2 motifs in sites becoming more accessible (Fig. 44C). In ES cells, we also find that the Snai2 transcriptional repressor motif in enriched in sites that lose accessibility. Similarly, sites that display reduced accessibility globally show an enrichment for the transcriptionally repressive REST complex[216] in both ES and NPCs. While Snai2 expression is increased in ES cells upon Neurog2 induction and could contribute to some of the global decreases in accessibility observed, REST expression decreases upon Neurog2 expression (Fig. 44E). However, at the REST motif, we identified a decreased accessibility at these sites in both ES and NPCs (Fig. 44D). It remains to be seen if reduced REST occupancy leads to compaction of these sites of if de-repressed genes because of REST downregulation contribute to this global compaction.

These findings indicate that Neurog2 influences chromatin accessibility through dual mechanisms. It directly enhances accessibility in both ES and NPC cells. However, Neurog2 also drives notable cell-type-specific effects, particularly in ES cells, including localised compaction via upregulation of repressors like Snai2 and broader compaction linked to reduced REST binding.

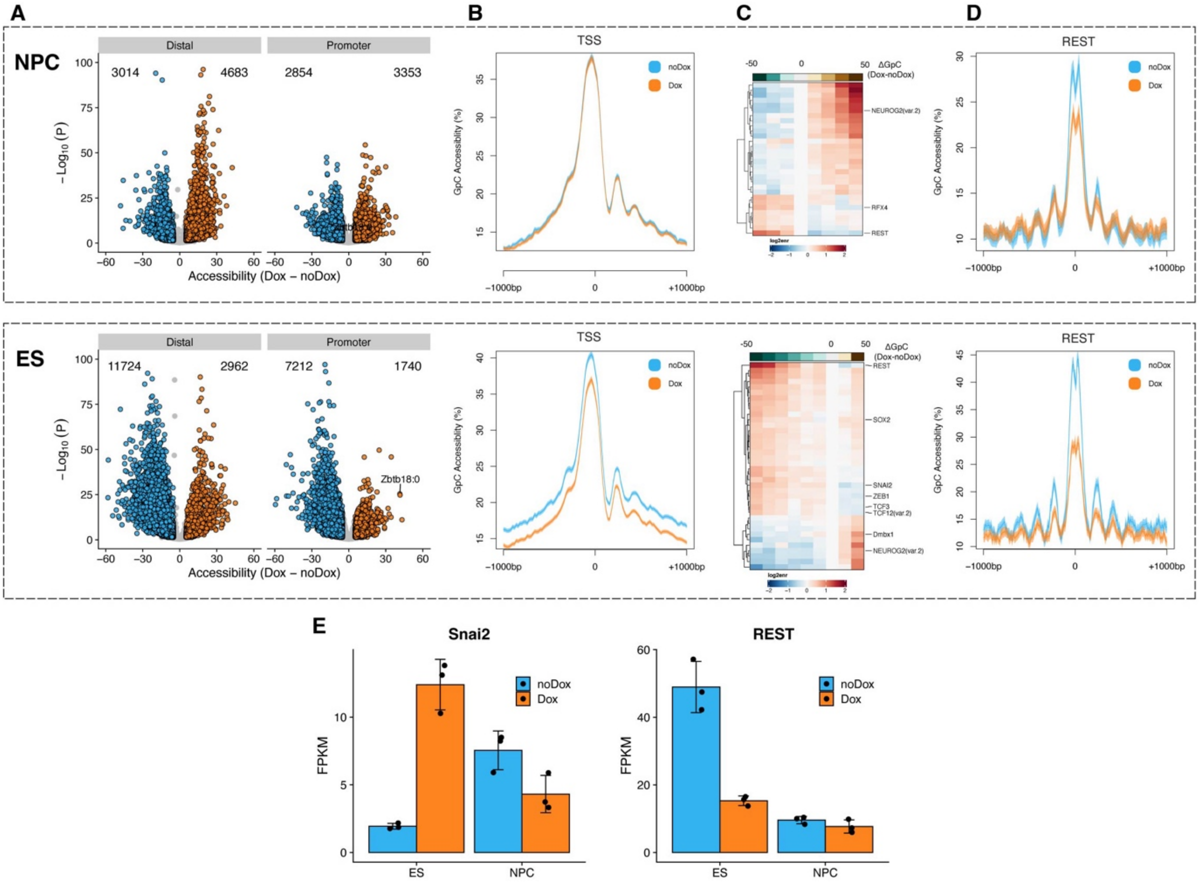

Figure 44 Neurog2 indirect effects on chromatin accessibility

A - Volcano plots showing global differentially accessible regions (DARs) in distal and promoter regions.
B - Chromatin accessibility levels at TSS in ES and NPC.
C - Heatmap illustrating motif enrichment within DARs in ES and NPC, stratified by changes in accessibility upon Neurog2 overexpression.
D - Motif footprinting based on GpC accessibility levels at REST binding sites in ES and NPCs.
E - Expression levels of Snai2, and REST in either ES or NPC, with dots representing individual biological replicates and error bars representing SD.

3.2.6 Neurog2 facilitates direct and global changes on DNA methylation

DNA methylation patterns seem to mirror the direct and global effects identified with chromatin accessibility. At Neurog2 bound peaks, we largely observe DNA demethylation in both ES and NPCs (Fig. 45A/B). However, in ES cells, a proportion of peaks gain methylation, despite methylation specifically at the Neurog2 motif being reduced. Contributing to this ES-specific effect is the global hypermethylation prominent in ES cells upon Neurog2 induction (Fig. 45C/D). This hypermethylation is correlated with the increased expression of the DNA methyltransferase Dnmt3a as well as the downregulated Tet1 enzyme (Fig. 45E). The change is expression patterns of these key methylation modulators is not as prominent in NPCs, reflected by minimal global differences in methylation.

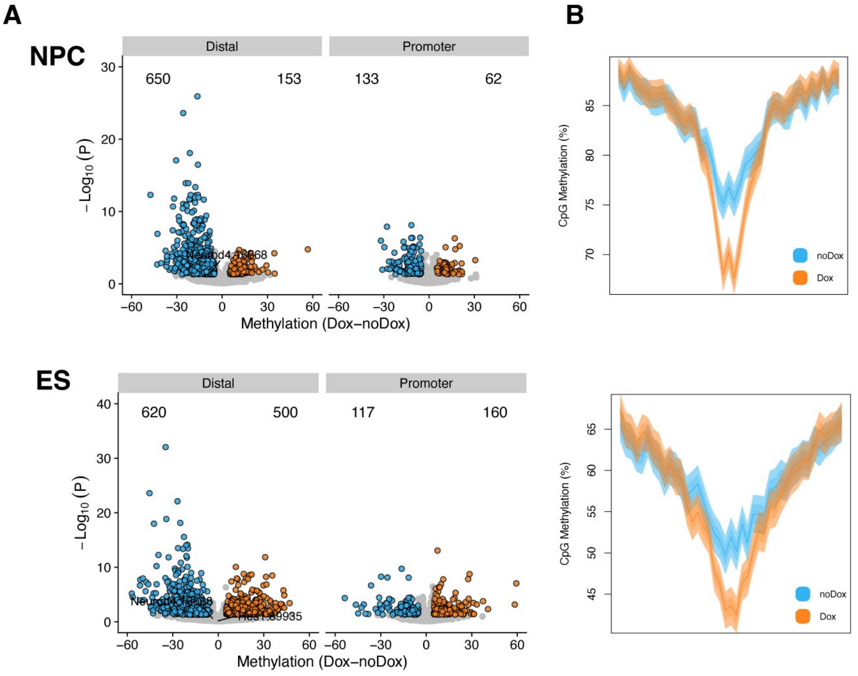

Figure 45 Neurog2 direct effects on DNA methylation

A - Volcano plots showing differentially methylated Neurog2 peaks in distal and promoter regions
B - DNA methylation profile of Neurog2 motifs within the top 5000 distal peaks.

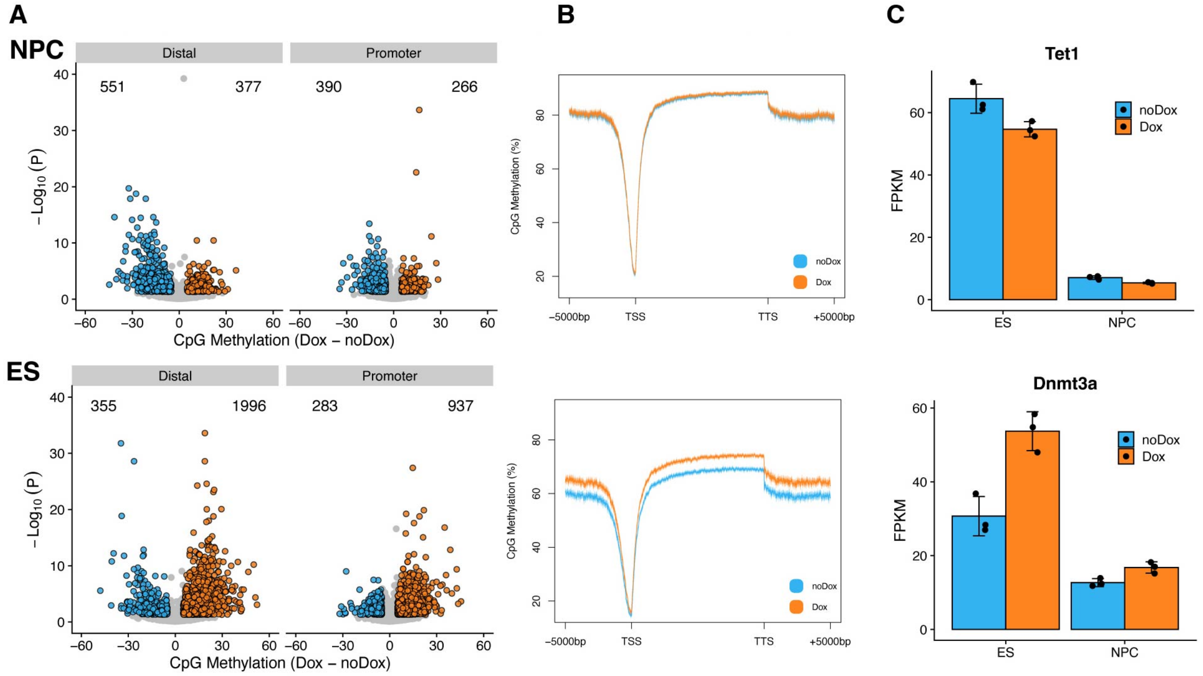

Figure 46 Neurog2 indirect effects on DNA methylation

A - Volcano plots showing global differentially methylated regions (DMRs) at distal and promoter regions.
B - CpG methylation levels at gene bodies in ES and NPC.
C - Expression of Tet1 and Dnmt3a in ES and NPC with dots representing individual biological replicates and error bars representing SD.

3.2.7 Neurog2 leads to direct and indirect rewiring of the 3D genome

Upon investigating the contact probability as a function of genomic distance in a cis-decay plot, we identified an ES-specific global shift towards short-range contacts upon Neurog2 induction, but not in NPCs (Fig. 47A). Compartment strength remained unaltered for both cell types (Fig. 47B), however in induced ES specifically stronger interactions and increased insulation at TAD boundaries were observed (Fig. 47C). Additionally, as previously reported[115,136], chromatin loops at direct Neurog2 binding sites are strengthened upon dox treatment (Fig. 47D).

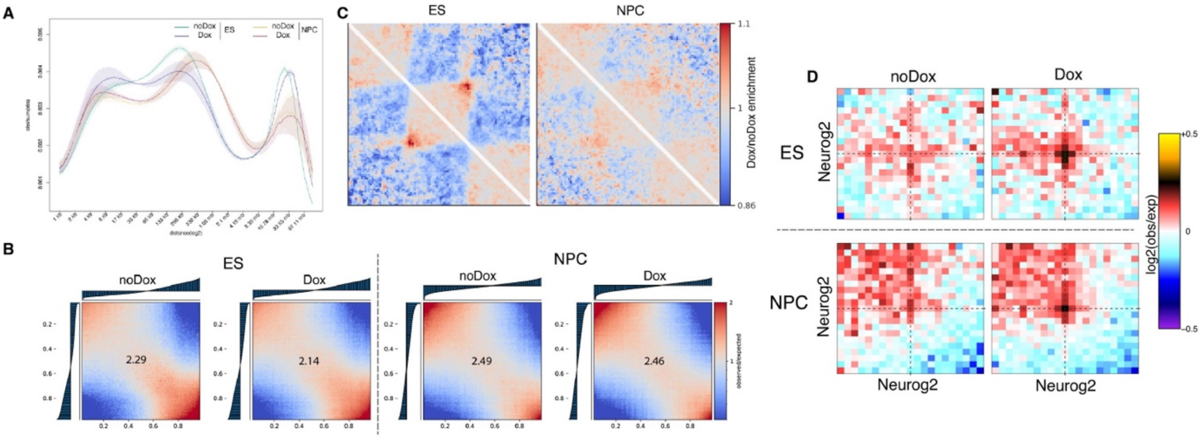

Figure 47 Neurog2 mediated 3D genome rewiring

A - Contact probability as a function of genomic distance: Lines represent mean values from biological replicates, while semi-transparent ribbons indicate the standard error of the mean (s.e.m.). Observed contact probabilities are normalized by dividing the observed values by the sum of observations (obs/sum).
B - Saddle plots illustrating compartment interaction strength, represented as observed/expected contacts in 100 kb bins.
C - Fold change in chromatin contacts across TADs between ES and NPC.
D - Aggregated Hi-C plots between intra-TAD pairs of the top 5000 Ngn2 ChIP-seq peaks

To address if the indirect ES-specific global rewiring is mediated by modulation of architectural proteins, we investigated the role of CTCF and cohesin further. At convergent CTCF sites, upon Neurog2 expression the contacts between pairs are strengthened at both sites (Fig. 48A), likely contributing to the increased insulation at CTCF sites observed. However, as the effects were not ES specific, we performed Rad21 CUT&RUN in the experimental conditions to understand if binding or processivity is altered. Rad21 peaks are enriched at Neurog2 binding sites in the induced conditions, most prominent in ES cells (Fig. 48B). Furthermore, in ES cells, by overlapping Rad21 and CTCF peaks, we find that processivity of Rad21 is reduced, resulting in a sharper binding profile (Fig. 48C). The increased insulation at CTCF peaks and the possible reduced processivity of cohesin, could therefore contribute to the more short-range contacts identified.

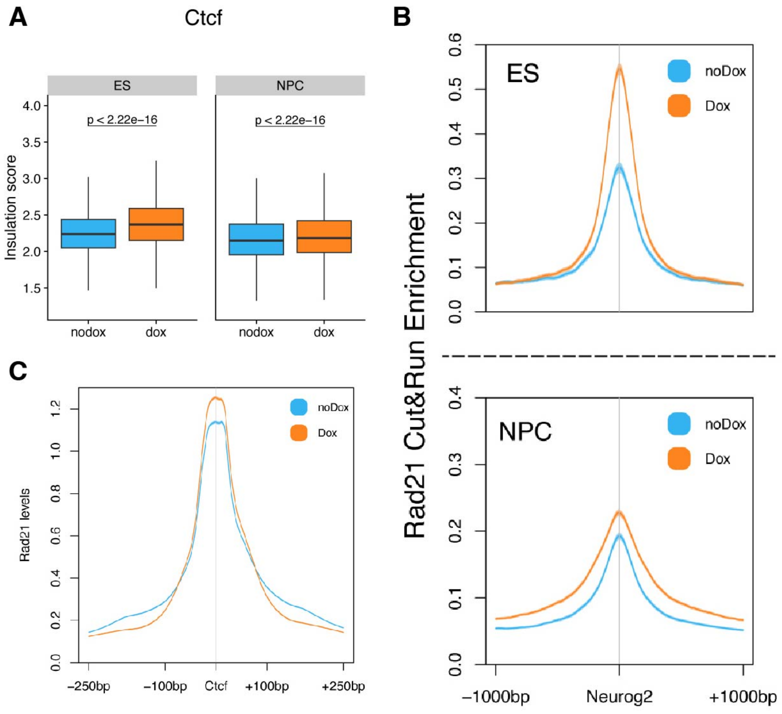

Figure 48 Neurog2 mediated indirect effects on the 3D genome

A - Boxplots displaying insulation scores in ES and NPC at CTCF sites.
B - Rad21 enrichment at Neurog2 ChIP-seq peaks in ES and NPC.
C - Rad21 enrichment at CTCF motifs within CTCF ChIP-seq peaks in ES cells.

3.2.8 Neurog2 interacts with established bHLH transcription factors and chromatin remodellers.

ChIP-MS was performed across experimental conditions to uncover cell-type-specific co-factors potentially driving epigenomic changes[148]. Neurog2 ChIP-MS identified 224 enriched proteins, with 10.3% shared between ES and NPC, including Neurog2 itself and E-box factors like Tcf3/4/12 and Id proteins (Fig 49A/B). Additionally, we found components of the chromatin remodelling NuRD complex enriched in both a shared and cell-type specific manner. Mbd3, Rbbp7 and Rcor2 were shared, whilst Mta2, Rbbp4, Trim28, Gatad2a/b (p66α/β) and Chd4 were ES-specific and Mta2, Rbbp4, Trim28, Gatad2a/b (p66α/β) and Chd4, with Mta1/3 NPC-specific. Other repressive complexes, such as Hdac1/2 and Sall2/4, which are associated with the NuRD complex, were specifically observed in ES cells but not in NPCs. We also detected interactions with Ruvbl1 (Rvb1) – a core subunit of the INO80 complex, functioning as a chromatin remodeler[225]. In NPCs, we find that the SWI/SNF subunit Brg1 (Smarca4) and Crebbp (Cbp) interacts with Neurog2. Interestingly, during neuronal activation, Brg1binding at enhancers is associated with its activation[226], and can influence NuRD interactions.

These findings highlight the cell-type-specific recruitment of chromatin remodelling complexes, supporting the coordinated epigenetic rewiring observed.

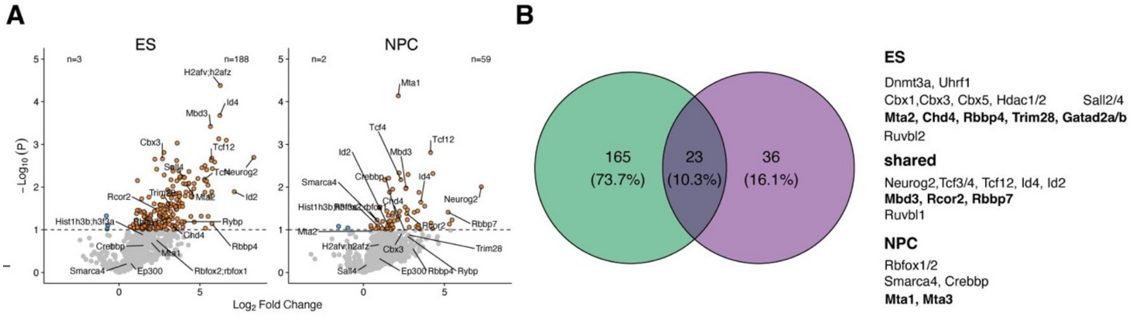

Figure 49 ChiP-MS used to identify Neurog2 interactors

A - Volcano plot showing representative proteins associated with Neurog2 identified via ChIP-MS analysis (n = 3 replicates)

B - Venn diagram illustrating shared and cell type-specific significantly enriched Neurog2-associated proteins.

3.3 Development of single cell multiomics methods to profile multiple epigenetic layers simultaneously

I contributed to all experimental studies carried out in this section. Analysis was supported by Andrea Fratton (Bonev Lab, PhD student).

3.3.1 HiT-seq: single-cell Hi-C + single-cell RNA workflow

At the start of the PhD, in the absence of high-throughput methods to simultaneously profile chromatin conformation and transcription at single-cell resolution, we initiated the development of HiT-seq (Hi-C + Transcription sequencing). The technique combined three main methodologies to achieve the goal of creating a novel method. This included: in-situ Hi-C[227] (for the chromatin conformation capture), SPRITE-seq[167] (for split-pool barcoding) and SHARE-seq[228] (for in-situ reverse transcription).

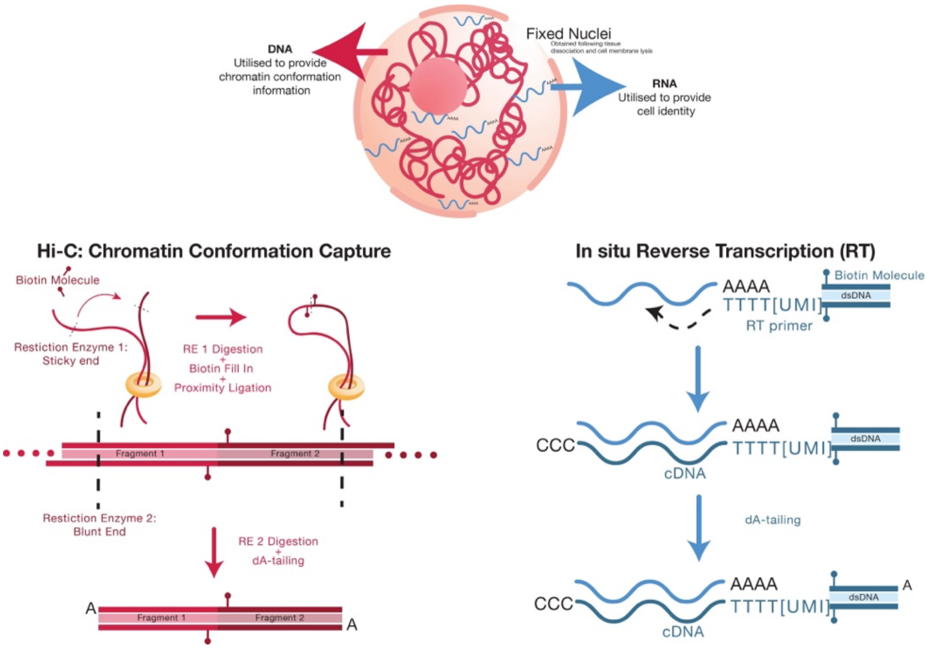

Figure 50 HiT-seq pre-barcoding experimental workflow

3.3.1.1 RNA Workflow

The method is reliant on having fixed and isolated nuclei. To preserve transcriptome quality, the first steps following nuclear isolation is the reverse transcription (RT) of RNA to produce an RNA-cDNA hybrid. As RNA is more prone to degradation in the method, the cDNA molecule generated by RT provides a layer of protection, despite considerable amounts of RNAse inhibitors being used. The RT primer used was a double-stranded poly-dT primer specifically targeting mRNAs. The primer contained a biotin molecule for cDNA pulldown during library preparation. The double-stranded nature of the primer also allowed accurate A-tailing of the primer, providing an overhang for barcoding steps, as well as ligation of the barcodes themselves (Fig. 50).

3.3.1.2 DNA Workflow

Following the RT, the DNA is subject to the conventional in-situ Hi-C process (Fig. 50). The nuclei are treated with 0.1% SDS to eliminate proteins not crosslinked to DNA[229]. The DNA is digested with DpnII producing sticky end fragments. The sticky ends are filled in and made blunt with a fill in step with Klenow Polymerase incorporating biotinylated nucleotides. The biotin would be used to enrich for correctly ligated Hi-C fragments during library preparation. Following the fill-in, proximity ligation is performed to ligate spatially proximal DNA fragments. To facilitate barcoding, new ends must be created within the DNA. Therefore, a second restriction enzyme had to be used to slightly re-digest the DNA. We used HpyCh4V, an enzyme known to function in in-situ assays[167], create blunt ends, and not digest RNA-DNA hybrids[230]. The ability to not cut RNA-DNA hybrids is crucial to prevent fragmentation of the cDNA. Following the generation of blunt ends, the DNA and cDNA fragments are simultaneously A-tailed using a Klenow polymerase without exonuclease activity. Analogous to the cDNA, the A-tail provides an overhang for barcoding.

3.3.1.3 Barcoding

For barcoding, we adopted the strategy used by SPRITE. Relying on split-pool barcoding to achieve a high-throughput (Fig. 51), the barcoding is used to label both sides of the DNA fragment, with only one side being labelled for the RNA. To prevent the unnecessary sequencing of the same barcode sequences on both reads of a DNA fragment, the first barcoding oligo contains 3' spacer in the top strand, preventing it from annealing or ligating to the second barcode. Thus, the free-floating top strand of the first barcode acts as its own primer binding site for amplification during library

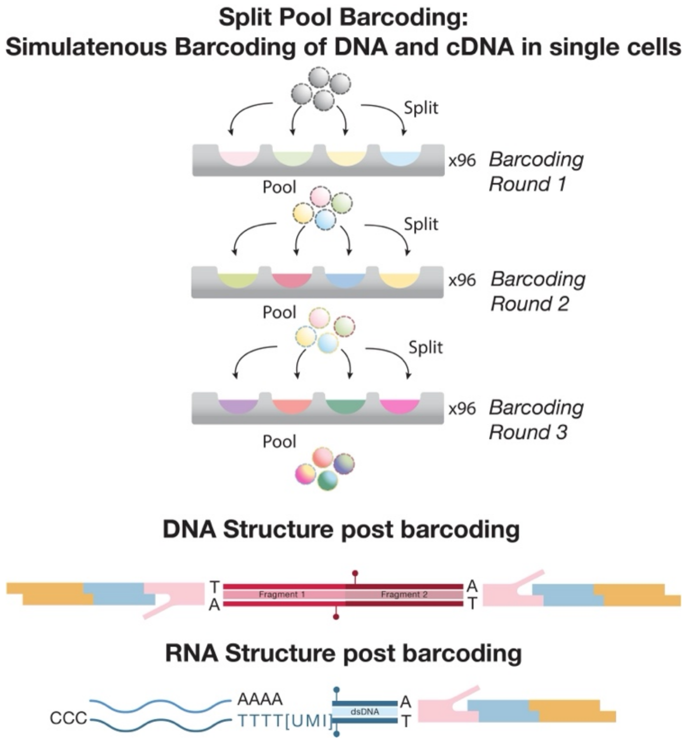

Figure 51 Principles of split pool barcoding

preparation. Consequently, the product created contains all three barcodes on one side, with only barcode on the other.

3.3.1.4 Library preparation

The presence of the biotin molecules on the Hi-C fragments and the cDNA facilitates pulldown with streptavidin beads. The beads were proportionally split 75% for Hi-C library preparation and 25% for cDNA. The DNA already consists of PCR handles on both sides of the fragment, allowing a simple PCR amplification to generate the library. Whereas on the cDNA, template switching is needed to introduce the second PCR handle for pre-amplification. The preamplified cDNA is subsequently tagmented readying the sample for sequencing (Fig. 52).

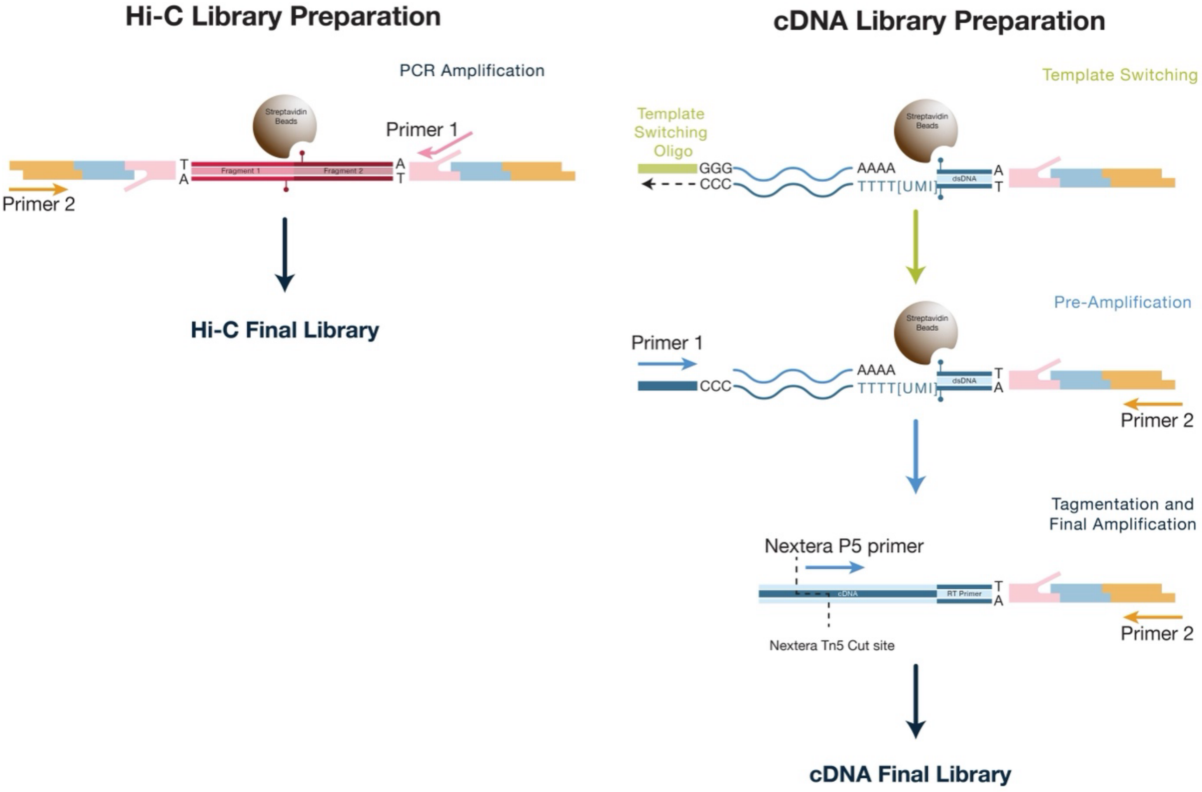

Figure 52 HiT-seq Library Preparation Workflow

3.3.2 HiT-seq: Issues

3.3.2.1 RNA-DNA Hybrid formation

We sequenced bulk tests of HiT-seq to get preliminary data. However, we noticed some peculiarities. We found the presence of the RT primer in a large proportion of the DNA sequencing reads, with the barcode structure not as expected. We deduced that this was a result of ligation of the double stranded RT primer to the DNA fragment (Fig. 53). Following the fill in step of the DNA, if both the cDNA and DNA fragments are not correctly A-tailed, the ends of these fragments remain blunt. Therefore, during barcode ligation, blunt end fragments can ligate to each other.

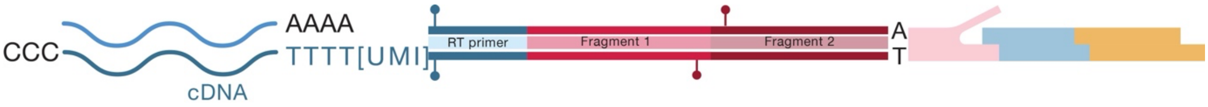

Figure 53 HiT-seq: RNA-DNA hybrid formation

To resolve the issue, we initially designed an alternative RT primer with a pre-existing A-overhang. However, due to the exonuclease activity of the Klenow polymerase the A is removed during the Hi-C fill in, producing the same result. We further tried using an A with a C3 spacer, terminal transferase mediated addition of ddUTP (that can be removed by USER treatment) to the end of the RT primer to prevent ligation, having an inverted T nucleotide that could subsequently be cleaved or using alternative polymerases to Klenow for the fill in. However, none of these approaches yielded promising results.

Therefore, we decided to abandon the use of the biotin fill in to enrich for Hi-C fragments. Single nucleus Hi-C[177], claims that fragment retrieval is restricted by using the biotin enrichment, suggesting that this change could beneficial. However, deeper sequencing would be required as intra-fragment reads would no longer be removed. An additional benefit of removing the fill in step is that during the streptavidin pulldown the beads no longer needed to be split to prepare the libraries, with all the cDNA remaining on the beads and the DNA fragments in the supernatant. Indeed, we found that removing the biotin fill in step, and having a modified RT primer with a pre-existing A overhang resolved the issue of RNA-DNA hybrids.

3.3.2.2 Low proportion of long-range contacts

After incorporating the changes above, we proceeded with the single-cell test of HiT-seq. The experiment was performed on a mixed population of human HEK cells and mouse ESCs to validate doublet ratios and collision rate of the technique. Mapping the cells using the hg38 genome for Human cells and mm10 for Mouse cells, we found that 9% of the cells were of mixed identity. This may point to clumping of cells during split-pool barcoding, requiring more stringent experimental filtration to mitigate.

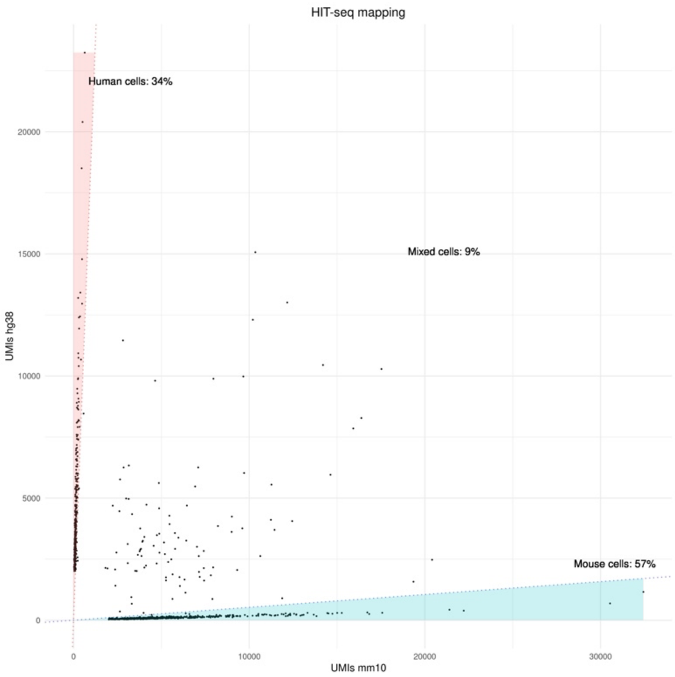

Figure 54 HiT-seq collision rate single-cell test

Unique RNA UMIs aligning to mm10 or hg38 genome was used to identify the collision rate.

Quality control metrics for the cDNA suggested that that large proportion of the reads were correctly barcoded and produced good recovery of genes and UMI's per cell (Fig. 55).

	Proportion of reads
cDNA	
Fully Barcoded	85%
Fully Barcoded with RT primer	73%
Hi-C	
Fully Barcoded	84%
Intra-fragment reads	42%
Hi-C Contacts	39%
Short Range Interactions (<20kb)	37%
Long Range Interactions (>20kb)	2%

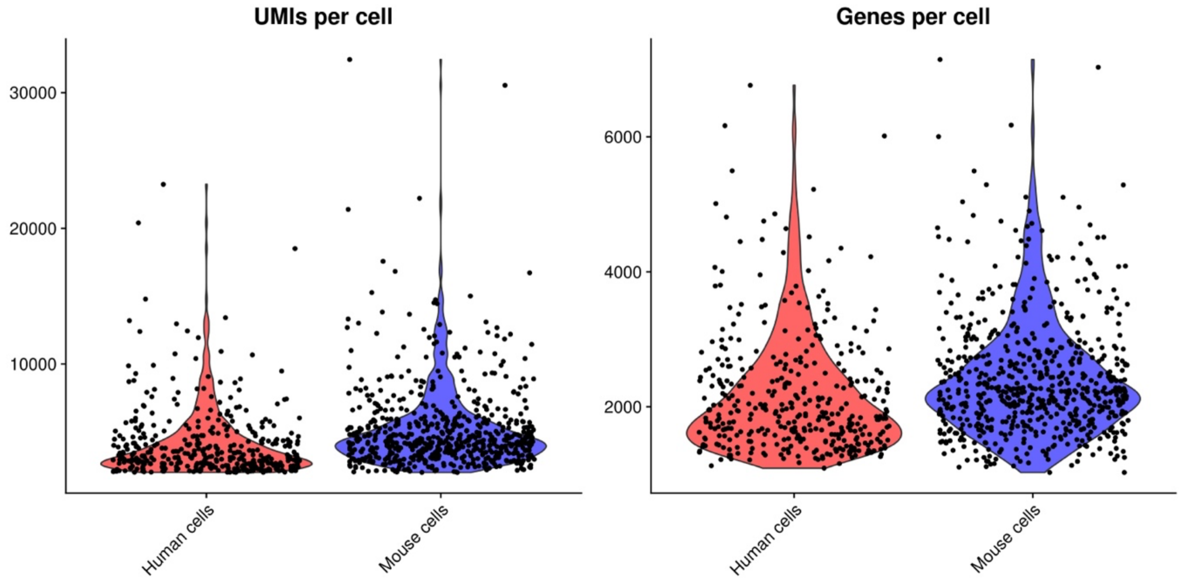

Figure 55 HiT seq - UMIs and genes per cell recovered

Violin plots showing the distribution of UMIs or genes per cells for mouse and human cell. Each black dot is an identified cell.

However, the Hi-C data raised significant concerns. Although the proportion of fully barcoded reads and Hi-C contacts were desirable, the proportion of Long-Range Interactions (>20kb) was very low. This long-range information represents the proportion of the data that is useful for determining regulatory interactions. Whilst HiT-seq only recovers 2% long range interactions, experiments such as Hi-C3.0 recover 25%[67]. Looking at the Hi-C interactions as a function of distance in a histogram showed a dramatic skew for short-range interactions, and zooming in further identified an irregular enrichment of fragments at very specific distances (Fig. 56).

We hypothesised that this irregular pattern and skew towards short range interactions could be result of over digestion with HpyCH4V. However, attempts to shorten the second digestion step proved unsuccessful. In the midst of troubleshooting this defect, other methods such are GAGE-seq[70] were published, prompting a change in strategy.

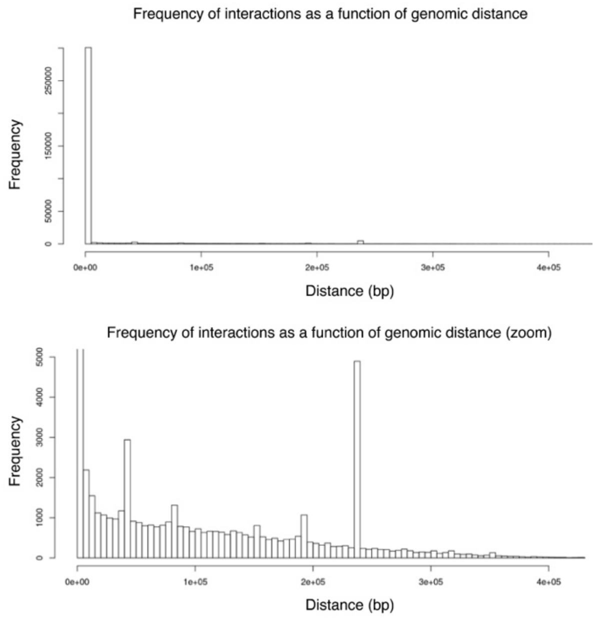

Figure 56 HiT-seq frequency of interactions

Histogram showing the frequency of interactions as a function of genomic distance

3.3.3 GAGE-seq has better recovery of long-range interactions compared to HiT-seq

GAGE-seq[70] conceptually resembled HiT-seq, using a split-pool barcoding based method to capture chromatin conformation and transcriptome simultaneously. However, the methodology contained some key differences which largely appeared to be theoretically beneficial, except for the tagmentation of the Hi-C DNA during library preparation.

HiT-seq	**GAGE-seq**	**Notes**
SDS treatment after RT	SDS treatment before RT	SDS can inhibit RNAses[231]. Early treatment likely beneficial
RT with oligo-dT primer	RT with oligo dT + Random Hexamer primer	Capturing all RNAs, with better recovery of mRNAs
First Digestion – DpnII (GATC) O/N	First Dig – MseI +CviQI (TA) 2hrs	Using two enzymes with different recognition sites but producing same overhangs allows more efficient digestion of the DNA.
Proximity Ligation – 4hrs	Proximity Ligation – O/N	More efficient ligation of Hi-C fragments
Second Digestion – HpyCh4V (Blunt) 4hrs	Second Dig – DdeI (TNA) 1hr	DdeI previously characterised as suitable for Hi-C in Hi-C 3.0[67]

A-tailing to generate overhang for barcode ligation	Barcode ligation to directly to TA / TNA overhang generated previously	One less step. Barcodes also ligated to Hi-C fragments not ligated during proximity ligation
Common barcodes for RNA + DNA (3 rounds)	Different barcodes for RNA + DNA (2 rounds split on multiple plates)	
Reverse Crosslinking after all barcoding	Nuclei lysis before final barcoding round	More efficient second round barcoding due to ligation occurring in solution rather than in-situ
No additional PCR handle ligation	Ligation of additional PCR handles to recover DNA barcoded only on one side of DNA fragment	Allows recovery of DNA fragments even when barcoded only on one side.
No tagmentation of Hi-C DNA during library preparation	Tagments Hi-C DNA with Nextera Tn5	Tagmentation loses 50% of amplifiable reads.
cDNA library preparation reliant of template switching using CCC nucleotides introduced by RT maxima during RT	cDNA library preparation uses terminal transferase in vitro to add poly-G nucleotides for subsequent pre-amplification	In vitro reaction likely more efficient that in situ reaction.

To compare GAGE-seq and HiT-seq, we performed a bulk version of the methods in mESCs. We used a version of HiT-seq with two rounds of barcoding to make it comparable to GAGE-seq. We also investigated a version of GAGE-seq without tagmentation of the DNA to identify if this further improves recovery. Focussing on the Hi-C quality control metrics, both methods have a similar barcoding efficiency (Fig. 57A). However, GAGE-seq captures a significantly higher proportion of long-range interactions, further improved by the non-tagmented version of the protocol (Fig. 57B).

Therefore, we continued with GAGE-seq as the backbone upon which we made further improvements.

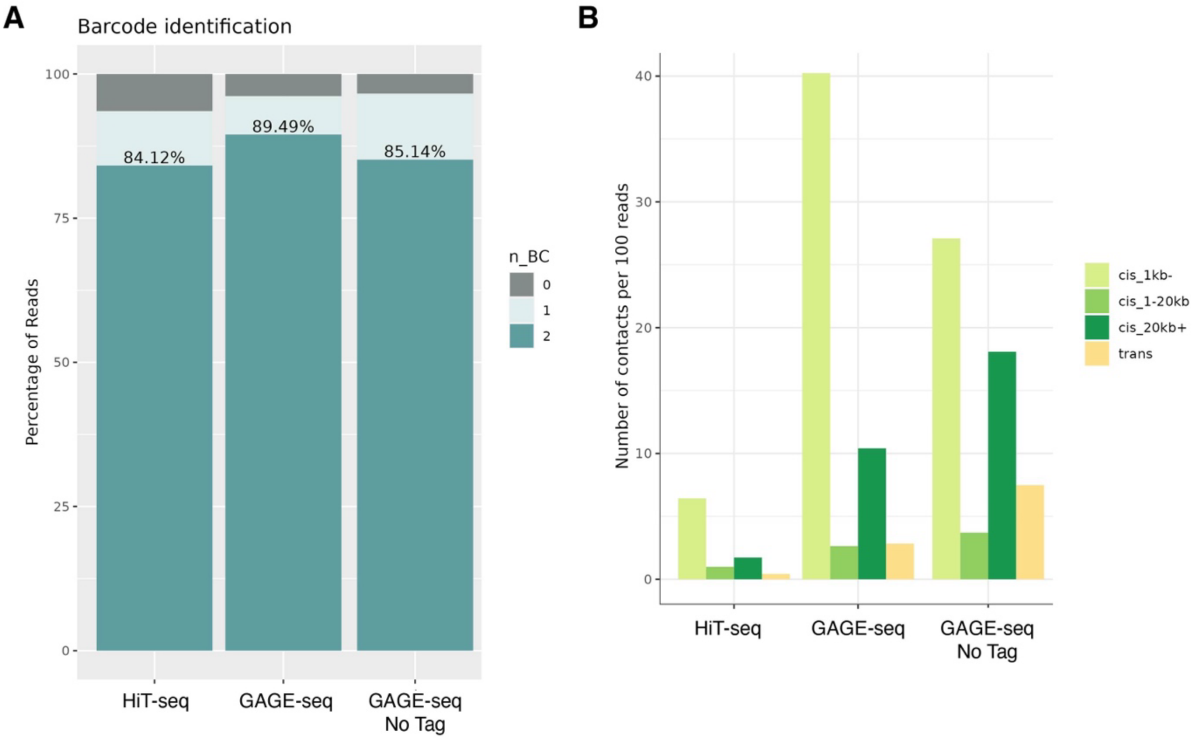

Figure 57 HiT-seq vs GAGE-seq

A – Stacked bar plot showing the barcoding efficiency of the methods.
B – Bar plot indicating the number of contacts stratified by interaction distance across experimental conditions.

3.3.4 sc-3DRAM-seq: Workflow

Building on GAGE-seq we aimed to add DNA methylation and chromatin accessibility as additional modalities to the method. To achieve this, four main modifications were required:

1. Addition of a GpC methyltransferase treatment step to methylate cytosines in the GpC context at accessible sites[189].
2. Bisulphite or enzymatic based conversion of DNA to identify methylated cytosines.
3. Modification of barcode sequences to be compatible with DNA conversion.
4. SPLAT Adapter Ligation[203]: post-conversion PCR handle attachment for damaged DNA fragments

The GpC methyltransferase step was introduced after the RT (Fig. 58). To prevent any changes in accessibility prior to the step, the SDS treatment of the nuclei was rearranged to after the GpC methylation, directly prior to the chromatin conformation steps. To correctly identify endogenous CpG methylation and the exogenously introduced GpC methylation marking accessibility, the DNA needs to be converted. Whilst bisulphite conversion is considered the gold standard for methylation studies[232], more recent methods such as EM-seq[233] allows lower inputs and are considered non-destructive. Unlike bisulphite-seq however, EM-seq relies on the conversion of the methylated cytosine to a carboxymethyl state for protection from subsequent conversion to a thymine. If the conversion to carboxymethyl-cytosine is not accurately achieved, the methyl-cytosine is also converted to thymine. Given these differences, we proceeded by comparing these two conversion techniques for sc-3DRAM-seq.

Following conversion and amplification, non-methylated cytosines are converted to thymines during PCR[232]. This poses an issue for identifying the barcode sequences as complexity would be largely reduced, resulting in possible overlap of sequences and misidentification of cells. Therefore, we modified the 9bp barcodes in each barcoding adapter to be devoid of cytosines and increased the length of the sequence to 12bp to continue to allow accurate discrimination of the sequences despite lower nucleotide complexity. Furthermore, we ordered the final barcode adaptor with methylated cytosines in the PCR primer handle sequence, allowing correct recognition and amplification by the PCR primers during library preparation despite conversion.

However, methylating oligos can be expensive and also it has been suggested that bisulphite chemical treatment can lead to fragmentation of DNA[234], leading to potential loss of the barcoding sequences on one side of the DNA fragment. To recover such fragments, we utilised a SPlinted Ligation Adapter Tag (SPLAT), consiting of a double stranded annealed oligo with a random hexamer overhang that is ligated to the 5' end of converted DNA providing an alternative PCR handle. The SPLAT adapter ligation also helps recover DNA that was only barcoded on one side, analagous to GAGE-seq.

A major advantage of incorporating methylation and accessibility is the ability to repurpose intrafragment reads or short-range contacts, which would otherwise be discarded, to uncover cell-specific DNA methylation and accessibility changes.

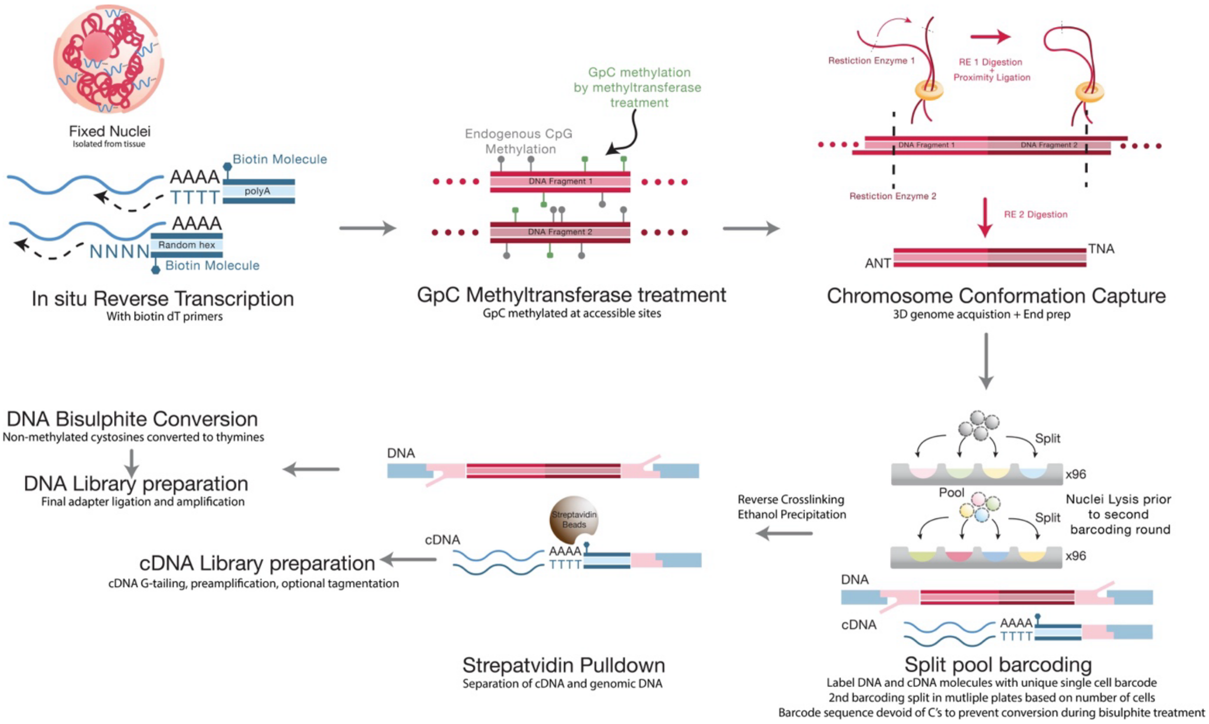

Figure 58 sc-3DRAM-seq experimental workflow

3.3.5 Bisulphite conversion more reliable than EM-seq

On barcoded DNA, we performed Bisulphite conversion or EM-seq conversion with a low input of 15ng of DNA. We find that from the BioA profiles of the libraries (Fig. 59), that the fragment size distribution is largely similar between the two conditions, excluding a strong bias for bisulphite treatment to cause DNA fragmentation.

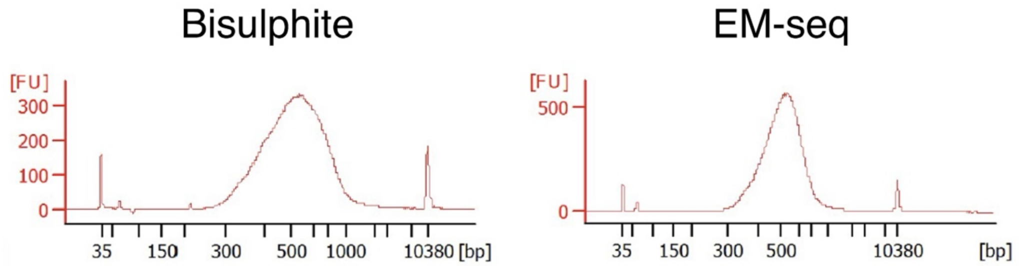

Figure 59 Fragment size comparison of Bisulphite and EM-seq libraries

Sequencing the libraries and examining the spiked in control DNA that was fully methylated on CpG and GpC cytosine revealed that identified that EM-seq had a CpG bias (Fig. 60). This was particularly evident for cytosines in the GpC context that was followed by a thymine, which was often not recognised as fully methylated. This thereby removes a quarter of the usable accessibility information. This CpG bias also resulted in the reduced identification of barcoding sequences (Fig. 61), likely due to inefficient amplification. As a result, we decided to proceed with bisulphite treatment as our conversion method of choice to accurately retain accessibility information and identify barcodes.

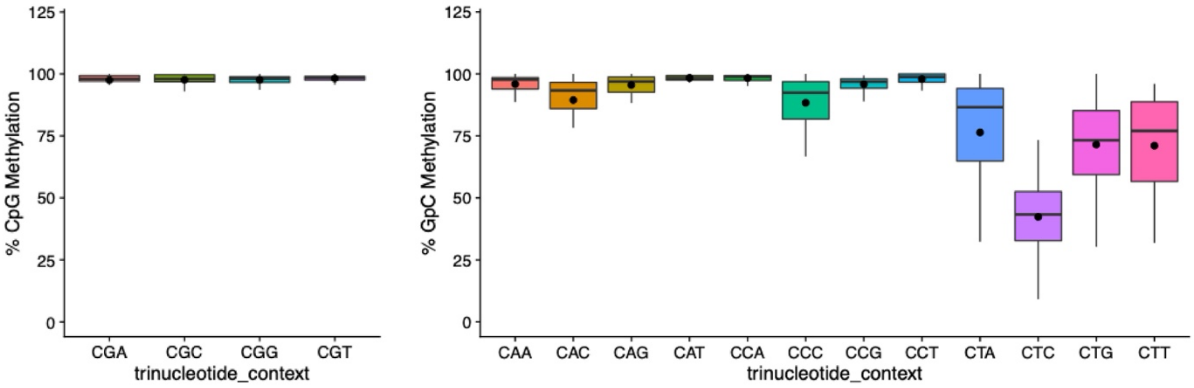

Figure 60 EM-seq CpG bias

Boxplots showing the distribution CpG methylation and GpC methylation across different trinucleotide contexts in fully methylated control DNA. The black dot represents the average methylation.

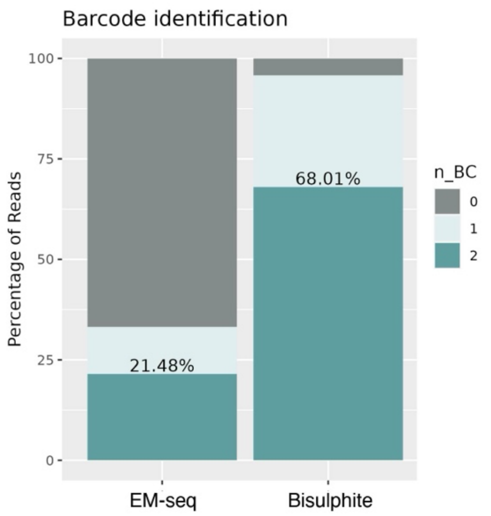

Figure 61 Barcode identification in EM-seq vs Bisulphite-Seq

Stacked bar plot showing the barcoding efficiency of the methods.

3.3.6 sc-3DRAM-seq bulk experiments show promising QC

We performed the bulk adaptation of sc-3DRAM-seq on a variety of cell types including mESCs, E14 Cortex, Human Cortical Organoids and Human Brain Tissue. Due to a sequencing error with cDNA in the most recent experiment, we present below RNA metrics from the mESC experiments.

In mESCs, we find over 87% of the cDNA reads are fully barcoded (Fig.62A), with coverage across the gene body (Fig.62B) because of the use of random hexamers in the RT.

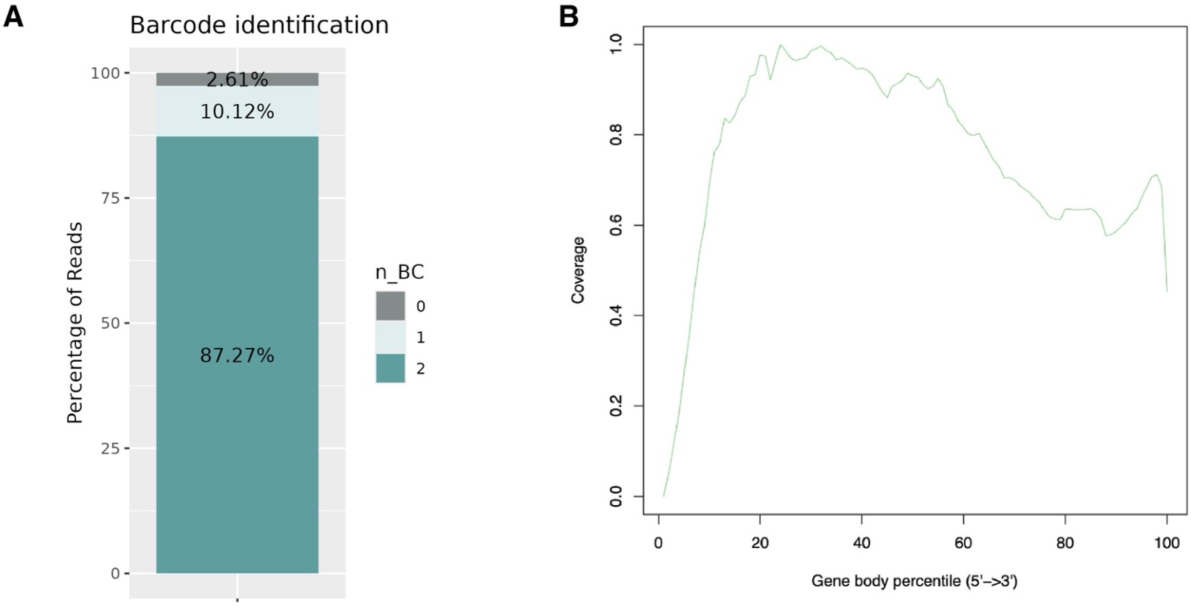

Figure 62 Bulk RNA QC for sc-3DRAM-seq

A - Stacked bar plot showing barcoding efficiency of RNA in sc-3DRAM-seq.
B – RNA read coverage across the gene body.

In human cortical organoids, the DNA also showed high recovery of fully barcoded reads (Fig. 63A). 47% of the reads represented intrafragment reads with 20% of reads retaining Hi-C contact information (Fig.63B). Both the intrafragment and Hi-C reads can be used to study changes in methylation/accessibility.

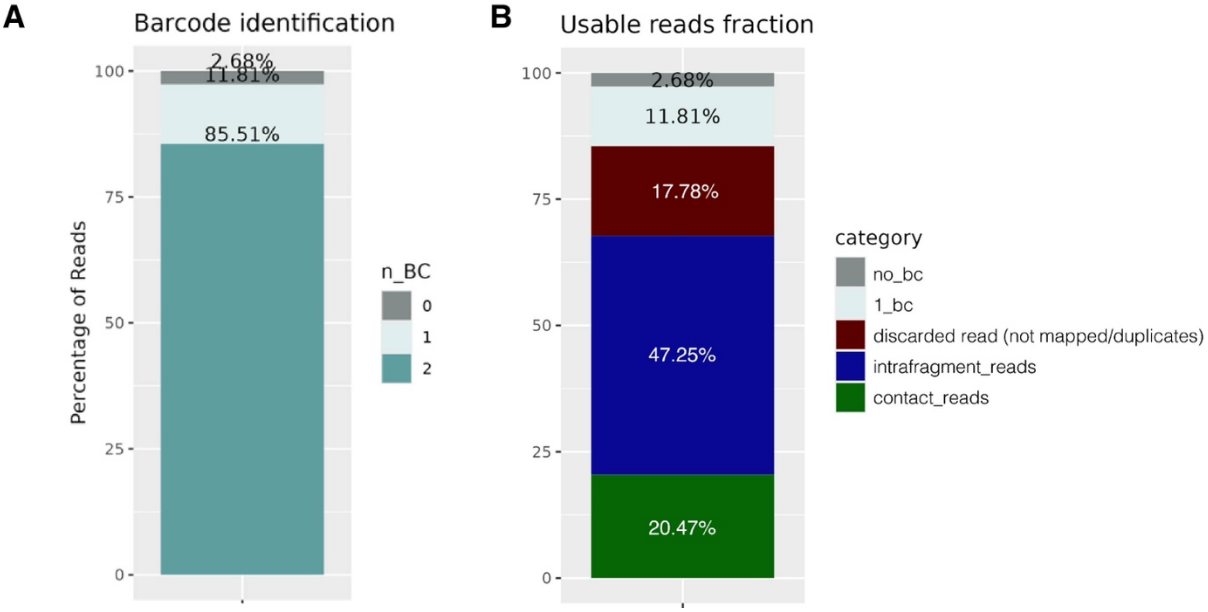

Figure 63 Bulk DNA QC for sc-3DRAM-seq

A - Stacked bar plot showing barcoding efficiency of DNA in sc-3DRAM-seq.
B - Stacked bar plot grouped by read category.

Of the informative Hi-C contacts we capture a good proportion of long-range interactions relative to short range reminiscent of our bulk GAGE-seq tests (Fig. 64A), resulting in a representative cis-decay profile of Hi-C methods (Fig. 64B). Furthermore, examining the methylation and accessibility profiles at CTCF sites revealed a characteristic reduced CpG methylation at the peak (Fig. 65), reciprocal to a

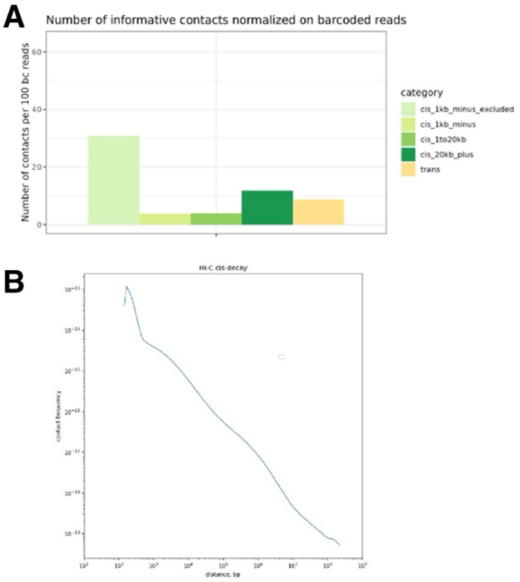

Figure 64 Long-rang interaction in bulk sc-3DRAM-seq

A - Bar plot indicating the number of contacts stratified by interaction distance
B - Contact probability as a function of genomic distance: Lines represent mean values

phased accessibility as represented by GpC methylation. The centre of the peak corresponds to CTCF binding and is associated with a decreased accessibility.

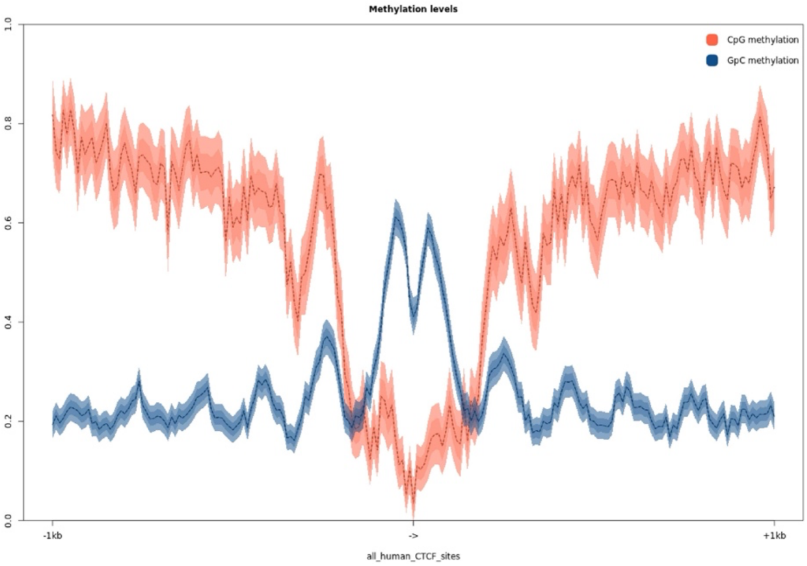

Figure 65 CpG and GpC Methylation centred on CTCF binding sites

Overall, the promising quality control metrics indicate that the method is well-suited for single-cell sequencing, with experiments actively underway.

4. Discussion

4.1 Transcription factor-driven rewiring of epigenetic landscapes

In this PhD, we explored the capacity of proneural pioneer TFs to rewire the epigenome during direct neuronal reprogramming of astrocytes and their cell-type-specific roles in ES and NPCs. In both scenarios, we identified that Ngn2 is capable of opening closed chromatin, demethylating DNA and mediating chromatin loops at bound pairs, in line with previous studies[135,136,235]. Such patterns are characteristic of other lineage-specifying pioneer TFs, like MyoD[236] and Foxa2[237]. However, previous studies on these factors have typically examined only a single epigenetic layer at a time.

We identified that Ngn2 binding at its cognate motif drives many of the direct effects observed. Ngn2's ability to open closed chromatin was influenced by the number of Ngn2 motifs within the region, with closed sites exhibiting increased accessibility when they contain a higher proportion of Ngn2 motifs. It is possible that more TF binding facilitates the recruitment of more chromatin remodellers for chromatin opening. However, it remains unclear whether this relationship is absolute or if closed chromatin with fewer Ngn2 motifs simply takes longer to open, given than Pax7 in pituitary glands has been shown to display delayed opening of chromatin after binding of the TF[235]. Additional time points would be necessary to test this hypothesis.

Whilst bHLH proteins are considered end-binders, preferring DNA sites up to 25–35 bp from the entry/exit site of the nucleosome[29], the relative position of the motifs within a closed region's nucleosome array can potentially affect the dynamics of opening[238]. Consistent with previoulsy identified competition between nucleosomes and TFs[239], we observed preferential binding of Ngn2 to sites with at least low accessibility under control conditions, followed by increased chromatin opening upon Ngn2 engagement. This increased opening appears to be facilitated by the recruitment of chromatin remodellers, having identified that Ngn2 directly interacts with subunits of the INO80, SWI/SWF and NuRD complexes in ES and NPCs. This supports previous studies that have shown *in vivo*, complete opening of the chromatin does still require ATP-dependent remodellers[32,33].

At Ngn2-bound sites, DNA demethylation is observed across different contexts, a phenomenon also recapitulated *in vivo*[136]. This demethylation effect is most prominent at distal sites, as promoters are already largely hypomethylated[45]. Although pioneer transcription factors such as FoxA1 have been shown to directly recruit TET enzymes and reduce 5mC on enhancers[240], we did not detect a direct interaction between Ngn2 and TET enzymes in our ChIP-MS experiments in ES and NPCs. This suggests that Ngn2 likely recruits TET enzymes indirectly to tissue-specific enhancers, similar to other transcription factors[241]. Our preliminary findings from an EB time course indicate that Ngn2 and other recruited factors are required for maintaining the demethylated state. Specifically, under wash conditions, Ngn2 methylation levels increased again but did not return to baseline levels, suggesting passive remethylation is driven by cell cycle dynamics. Ectopic expression of Foxa2 followed by its removal yielded similar results *in vitro*[242]. However, *in vivo*, some epigenetic memory of Foxa2 was observed[240]. If Ngn2 is required for maintenance of demethylated DNA *in vivo* remains to be investigated.

Neuronal differentiation and cell identity changes are closely associated with alterations in the 3D genome's compartmentalization[153,236,243]. Consistent with this, we observed stronger compartmentalisation during neuronal reprogramming, particularly with enhanced interactions within the inactive B compartment. However, in ES and NPC models, compartmentalisation changes were not evident upon Ngn2 expression. This discrepancy may stem from differences in timing, as the 3D genome was analyzed after 2 dpi during reprogramming but only after 1 dpi in the ES and NPC models.

Locally, we observed strengthened TAD boundaries, increased insulation, and a shift toward shorter-range interactions during reprogramming and induction of ES cells. These changes likely result from the enrichment of cohesin peaks at Ngn2 binding sites, facilitating loop extrusion[244]. In contrast, the absence of large-scale changes in 3D genome architecture in NPCs could be due to the similarity of the NPC state to the native endogenous environment of Ngn2, which requires fewer changes to the 3D genome. Alternatively, it may be a reflection of the lower levels of Ngn2 protein upon overexpression in NPCs compared to ES cells.

However, across all contexts, the contact frequency between Ngn2-bound sites is increased, leading to the formation of Ngn2-specific strengthened chromatin loops. Additionally, the positively correlated EGPs were associated with increased E-P contact strength, highlighting a potential link to gene expression regulation. Ngn2 is shown to homodimerize[113], and has previously been shown to form stronger loops when binding is present at both the enhancer and promoter[136], pointing to protein-protein interactions as key mediators of chromatin looping. Further modulation through the recruitment of cohesin as seen with other pioneer TFs[245], co-factors such as Yy1 as identified in the reprogramming study, or known *in vivo* interactors like Lmo4 and Ldb1[246], may also contribute to this process.

4.2 The role of protein stability and post-translational modifications in transcriptional regulation

Ngn2 protein stability is regulated by post-translational modifications, specifically phosphorylation at SP/TP sites, where the underphosphorylated form promotes the expression of neuronal differentiation genes[114]. By utilising a PmutNgn2 variant in our reprogramming study, we uncovered mechanistic insights into its differential activity compared to the wild-type protein.

We found that PmutNgn2 was capable of generating more mature iNs rapidly, with more complex neurite outgrowth, resembling the behaviour observed with the phosphomutant variant of another proneural TF, Ascl1[247]. The transcriptome associated with the PmutNgn2 is distinguished by the upregulation of a broader set of neuronal maturation genes, coupled with a more pronounced downregulation of astrocyte identity genes, thus highlighting the importance of erasure of the starting cell identity for successful maturation[248].

The PmutNgn2 exhibited differential opening of chromatin, more efficient DNA demethylation and increased chromatin looping at bound sites which could be a result of stronger binding to the DNA as previously suggested[114]. Yet, the PmutNgn2-specific binding sites are predominantly promoters that are already largely accessible, and bound genes only mildly increased in expression. This suggests that PmutNgn2 does not have stronger pioneering activity than Ngn2. Furthermore, at these specific binding sites, there is no enrichment of any TF motifs. While the phosphorylation status could potentially lead to higher protein levels, it does not account for the PmutNgn2-specific binding patterns, particularly under conditions of forced overexpression where TF levels are exceedingly high. However, is it a possibility that the scan and search of PmutNgn2 is varied compared to Ngn2, with differential co-factor recruitment stabilising interactions on partially recognised motifs[249].

While PmutNgn2 generates more mature iNs with a significantly rewired epigenome, it remains unclear whether this landscape results from a different regulatory trajectory or simply from the accelerated pace of changes it induces. Therefore, it would be intriguing to explore the epigenetic layers in Ngn2-induced astrocytes at a later time point to determine if their landscape more closely aligns with the one generated by PmutNgn2.

Post-translational modifications may also play a role in ES and NPCs. The increased protein levels in ES cells could indicate a differential phosphorylation status, leading to reduced ubiquitin-mediated protein degradation[114]. This modulation in protein stability could potentially contribute to some of the cell-type specific chromatin remodeling we observe. Further experiments determining the phosphorylation status of Ngn2 in ES and NPCs will be beneficial to address this hypothesis.

4.3 Cell-type context: balancing direct and indirect effects in reprogramming

Ngn2 has previously been shown to generate iNs with different neuronal identity based on the spatial origin of astrocytes used for reprogramming[146]. We therefore sought to explore how the starting cell type influences the activity of this pioneer TF, using ES and NPCs as models. This approach allowed us to investigate whether the inherent chromatin landscape and transcriptional state of the starting cells shape the ability of Ngn2 to drive rewiring.

We observed that in the two cell types the direct binding sites of Ngn2 were largely shared. In both contexts, Ngn2 activated chromatin through similar mechanisms, resulting in a significant overlap of upregulated genes. This suggests that the pioneer activity of Ngn2 and its downstream transcriptional effects are broadly conserved across these distinct starting cell types. Comparing the binding sites to another differentiation study with Ngn2 performed in EBs[135], we find that the ES binding sites are comparable to binding seen after 12h of induction in EBs, with NPCs more closely resembling binding at 48h[148]. This could suggest that the trajectory set by Ngn2 after 24 hours is similar in ES and NPCs, as Ngn2 expression in both cell types have previously been shown to produce differentiated neurons [250]. Continuing Ngn2 expression to the stage of terminally differentiated neurons and assessing differences would allow this to be further explored.

However, in ES cells, we specifically observe upregulation of the transcriptional repressors Id3 and Snai2, potentially reflecting the pluripotent state of Ngn2 and the priming of genes capable of inducing rapid transcriptional repression during differentiation[251]. Furthermore, it is possible that the highly plastic state of ES cells contributes to the global indirect changes[252], including the loss of accessibility and hypermethylation, that we observe. Nevertheless, it was surprising to detect such large-scale indirect changes after only 24h of induction. Interestingly, the sites that globally decreased in accessibility in ES cells were associated with the transcriptionally repressive REST complex[216] motif, despite the downregulation of its expression. It remains to be determined whether reduced REST occupancy leads to compaction of these sites, or if de-repressed genes resulting from REST downregulation contribute to this global compaction. Whether similar indirect changes are observed when overexpressing other TFs in ES cells remains to be carefully studied.

In the reprogramming paradigm, the generation of iNs is associated with a strong downregulation of astrocytic genes. This silencing may be a direct effect, as we found Ngn2 binding at enhancers that regulate gliogenic function. Similar to what was previously described for Ngn1, Ngn2 may sequester transcriptional co-activators, such as CBP and Smad1, away from the astrocytic gene promoters[253]. Alternatively, we find the Id3 motif enriched in early reprogramming iNs, which may indirectly support the repression of the original cell identity.

These widespread changes emphasise the need to thoroughly examine both the direct and indirect effects linked to cell identity shifts, even at early time points of TF induction.

4.4 Co-factors and chromatin remodellers: key players in fate determination

We identified Yy1 as a co-factor directly recruited by Ngn2 at specific sites, enhancing Ngn2's ability to open chromatin more effectively and promote stronger chromatin loops at co-bound sites. Yy1 is known to homodimerize, interact with chromatin remodellers such as INO80 and the BAF complex, and bind to consensus sequences primarily located in enhancers and promoters[142]. Deletion of Yy1 is further known to specifically disrupt E-P contacts and the associated gene expression[141]. These characteristics of Yy1 align with and support our observed results, demonstrating the synergy between Yy1 and Ngn2.

Conditional deletion of Yy1 in reprogramming resulted in a drastically reduced reprogramming efficiency, which was associated by a blockage in the maturation of the iNs. However, loss of Yy1 did not affect Ngn2 at its binding site, suggesting that Yy1 is not essential for the recruitment or retention of Ngn2 at these regions. Given that Ngn2 and Yy1 binding sites primarily overlap at distal regions of neurogenic genes, these regions were most affected by the loss of Yy1. Hence, while Yy1 KO cells were able to initiate reprogramming, the progression to more mature states associated with neurogenic gene expression was impaired. Notably, Yy1 is known to regulate a broad array of metabolic pathways and protein translation during development[208], and defects in these processes may also contribute to the observed inefficiency in reprogramming.

Although we observed a synergistic relationship between Ngn2 and Yy1 in reprogramming, with Yy1 deletion impacting the expression of key Ngn2-regulated neurogenic genes, Yy1 is a ubiquitously expressed protein with diverse functions[142]. To more directly assess the interplay between Ngn2 and Yy1, it would be interesting to overexpress both factors simultaneously and evaluate whether this enhances reprogramming efficiency, potentially mirroring the effects seen with the PmutNgn2. This could be further supported by identifying other co-factors of proneural TFs during neuronal reprogramming.

In ES and NPCs, we used ChIP-MS and identified significant interactions between Ngn2 and chromatin remodeling complexes, particularly with subunits of the SWI/SNF and NuRD complexes. Interestingly, some different subunits were enriched in the two cell types, potentially leading to differential activity of the complexes. For example, while the NuRD complex has traditionally been associated with transcriptional repression[254] and could contribute to the indirect repression of alternative lineages, recent studies have linked the subunits identified in ES cells to transcriptional activation[255]. These subunits may therefore function synergistically with Ngn2 to rewire chromatin[148]. Additionally, we observed cell-type-specific interactions with the repressive proteins Sall2/4 in ES cells. These proteins are known to repress cyclins[256] and are crucial for the neural differentiation of ES cells[257]. The enrichment of such proteins specifically in ES cells may reflect the pluripotent state of the cells, highlighting the need to direct differentiation toward a specific trajectory, a mechanism also shared by FoxA1[258].

While Ngn2 defines the new cell identity, the successful rewiring of the epigenome relies on the recruitment of co-factors and chromatin remodellers to drive the necessary changes for cell fate conversion. These interactions are shaped by the starting cell type and are likely further modulated by the post-translational modifications of Ngn2.

4.5 Integrative analysis multiple epigenetic layers

Using multiomics techniques, we have uncovered the coordinated changes driven by proneural transcription factors in reshaping cell identity across multiple epigenetic layers. However, the interplay between these layers is not always perfectly aligned. For instance, an analysis of E-P contact strength among positively associated EGPs within their respective cell type clusters showed comparable contact strengths in both Ngn2 and PmutNgn2 conditions. Yet, at PmutNgn2-bound sites, contact strengths were notably higher than those observed at Ngn2-bound sites. This indicated that the maturation of neurons is marked by differential changes in accessibility and subsequent gene expression can be independent of strength of chromatin looping. Furthermore, we found that Yy1 supports Ngn2-mediated reprogramming by enhancing chromatin accessibility and looping at co-bound sites, but it does not influence DNA methylation. In ES and NPCs, at the early time point of 1 dpi, the anticorrelation between chromatin accessibility and DNA methylation is relatively weak, particularly in ES cells. However, this anticorrelation may strengthen over time, as previously demonstrated using sc-NMT[190]. These findings highlight the importance of using methods that integrate multiple modalities to more precisely deduce the mechanisms driving the primary effects.

While analytical approaches have significantly advanced the integration of data[152,259,260], the field is evolving rapidly with the development of new experimental methods that enable simultaneous profiling of multiple epigenetic layers[261]. To advance this area of research, we are currently developing sc-3DRAM-seq that can measure the 3D Genome, RNA, Accessibility and DNA Methylation together at single-cell resolution. Building on GAGE-seq[70], we enhanced the recovery of long-range Hi-C contacts by omitting the tagmentation step for the Hi-C DNA library as originally proposed. Additionally, we incorporated GpC methylation to mark chromatin accessibility and implemented bisulphite conversion[190,191] to simultaneously capture endogenous DNA methylation and chromatin accessibility data. By incorporating these two additional modalities into GAGE-seq, reads that would otherwise be discarded due to representing short-range contacts or intra-fragment interactions can now be repurposed to provide valuable information about DNA methylation and chromatin accessibility within the cell. We have generated promising quality control data for all modalities by performing the bulk variation of the protocol in multiple tissues including mESCs, E14 Cortex and Cortical Organoids. The single-cell data is currently being generated.

Recently, methods such as single-nucleus methyl-3C sequencing (snm3C-seq3) have been used to understand heterogeneity within the human brain with temporal[262] and spatial resolution[263]. Paired-Hi-C[183], which simultaneously profiles chromatin conformation and transcription in single cells, has demonstrated ability in detecting copy number variations, structural variations, and extrachromosomal DNA in human glioblastoma, colorectal, and blood cancer cells. These studies highlight the strength of such methods in accurately profiling heterogeneous tissues and enabling meaningful comparisons across diverse cellular contexts.

Similarly, our initial aim with sc-3DRAM-seq is to compare mouse and human fetal brains to uncover the epigenetic modifications that drive cell fate decisions underlying the evolution and expansion of the human neocortex[264]. Moreover, considering the intricate interplay between epigenetic layers and the pivotal roles of chromatin remodellers, architectural proteins, and co-factors in shaping them, a perturbation screen[265] targeting multiple proteins and profiling their primary effects on these layers would provide invaluable insights. Collectively, we hope that this tool will deepen our understanding of the mechanisms that drive cell type-specific phenotypes.

References

1. Hashimshony, T., Zhang, J., Keshet, I., Bustin, M. & Cedar, H. The role of DNA methylation in setting up chromatin structure during development. *Nat Genet* **34**, 187–192 (2003).
2. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet* **20**, 207–220 (2019).
3. Jordan Rowley, M. & Corces, V. G. Organizational Principles of 3D Genome Architecture. *Nat Rev Genet* **19**, 789–800 (2018).
4. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res* **21**, 381–395 (2011).
5. Lenhard, B., Sandelin, A. & Carninci, P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat Rev Genet* **13**, 233–245 (2012).
6. Slattery, M. *et al.* Absence of a simple code: how transcription factors read the genome. *Trends in Biochemical Sciences* **39**, 381–399 (2014).
7. Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13**, 613–626 (2012).
8. Reiter, F., Wienerroither, S. & Stark, A. Combinatorial function of transcription factors and cofactors. *Current Opinion in Genetics & Development* **43**, 73–81 (2017).
9. Zhu, L. & Huq, E. Mapping Functional Domains of Transcription Factors. in *Plant Transcription Factors: Methods and Protocols* (eds. Yuan, L. & Perry, S. E.) 167–184 (Humana Press, Totowa, NJ, 2011). doi:10.1007/978-1-61779-154-3_9.
10. Damante, G. *et al.* Sequence-specific DNA recognition by the thyroid transcription factor-1 homeodomain. *Nucleic Acids Research* **22**, 3075–3083 (1994).
11. Geertz, M., Shore, D. & Maerkl, S. J. Massively parallel measurements of molecular interaction kinetics on a microfluidic platform. *Proceedings of the National Academy of Sciences* **109**, 16540–16545 (2012).
12. Johnson, P. F. & McKnight, S. L. EUKARYOTIC TRANSCRIPTIONAL REGULATORY PROTEINS. *Annual Review of Biochemistry* **58**, 799–839 (1989).
13. Mayran, A. & Drouin, J. Pioneer transcription factors shape the epigenetic landscape. *Journal of Biological Chemistry* **293**, 13795–13804 (2018).
14. Cirillo, L. A. *et al.* Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell* **9**, 279–289 (2002).
15. Yu, X. & Buck, M. J. Pioneer factors and their in vitro identification methods. *Mol Genet Genomics* **295**, 825–835 (2020).
16. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**, 311–318 (2007).

17. Calo, E. & Wysocka, J. Modification of Enhancer Chromatin: What, How, and Why? *Molecular Cell* **49**, 825–837 (2013).
18. Gualdi, R. *et al.* Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. *Genes Dev* **10**, 1670–1682 (1996).
19. Continued activity of the pioneer factor Zelda is required to drive zygotic genome activation - PMC. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6544384/.
20. Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation - PMC. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4617966/.
21. Horisawa, K. & Suzuki, A. The role of pioneer transcription factors in the induction of direct cellular reprogramming. *Regenerative Therapy* **24**, 112–116 (2023).
22. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors - ScienceDirect. https://www.sciencedirect.com/science/article/pii/S0092867406009767.
23. Păun, O. *et al.* Pioneer factor ASCL1 cooperates with the mSWI/SNF complex at distal regulatory elements to regulate human neural differentiation. *Genes Dev* **37**, 218–242 (2023).
24. Lu, C. *et al.* Essential transcription factors for induced neuron differentiation. *Nat Commun* **14**, 8362 (2023).
25. Hansen, J. L., Loell, K. J. & Cohen, B. A. A test of the pioneer factor hypothesis using ectopic liver gene activation. *eLife* **11**, e73358 (2022).
26. Donovan, B. T. *et al.* Basic helix-loop-helix pioneer factors interact with the histone octamer to invade nucleosomes and generate nucleosome-depleted regions. *Molecular Cell* **83**, 1251-1263.e6 (2023).
27. Chen, J. *et al.* Single-Molecule Dynamics of Enhanceosome Assembly in Embryonic Stem Cells. *Cell* **156**, 1274–1285 (2014).
28. Mir, M. *et al.* Dynamic multifactor hubs interact transiently with sites of active transcription in Drosophila embryos. *eLife* **7**, e40497 (2018).
29. Zhu, F. *et al.* The interaction landscape between transcription factors and the nucleosome. *Nature* **562**, 76–81 (2018).
30. Iwafuchi-Doi, M. *et al.* The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Mol Cell* **62**, 79–91 (2016).
31. Tsukiyama, T., Becker, P. B. & Wu, C. AlP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. **367**, (1994).
32. Dodonova, S. O., Zhu, F., Dienemann, C., Taipale, J. & Cramer, P. Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function. *Nature* **580**, 669–672 (2020).
33. Echigoya, K. *et al.* Nucleosome binding by the pioneer transcription factor OCT4. *Sci Rep* **10**, 11832 (2020).

34. King, H. W. & Klose, R. J. The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. *eLife* **6**, e22631 (2017).
35. Lettice, L. A. *et al.* A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* **12**, 1725–1735 (2003).
36. Chatterjee, S. & Ahituv, N. Gene Regulatory Elements, Major Drivers of Human Disease. *Annu Rev Genomics Hum Genet* **18**, 45–63 (2017).
37. Song, L. *et al.* Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res* **21**, 1757–1767 (2011).
38. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213–1218 (2013).
39. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
40. Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766–770 (2008).
41. Ng, H.-H. *et al.* MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat Genet* **23**, 58–61 (1999).
42. Bird, A. P. & Wolffe, A. P. Methylation-Induced Repression— Belts, Braces, and Chromatin. *Cell* **99**, 451–454 (1999).
43. Moore, L. D., Le, T. & Fan, G. DNA Methylation and Its Basic Function. *Neuropsychopharmacol* **38**, 23–38 (2013).
44. Kiselev, I. S., Kulakova, O. G., Boyko, A. N. & Favorova, O. O. DNA Methylation As an Epigenetic Mechanism in the Development of Multiple Sclerosis. *Acta Naturae* **13**, 45–57 (2021).
45. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev* **25**, 1010–1022 (2011).
46. Bell, J. S. K. & Vertino, P. M. Orphan CpG islands define a novel class of highly active enhancers. *Epigenetics* **12**, 449–464 (2017).
47. Suzuki, T. *et al.* RUNX1 regulates site specificity of DNA demethylation by recruitment of DNA demethylation machineries in hematopoietic cells. *Blood Advances* **1**, 1699–1711 (2017).
48. Yang, Y. A. *et al.* FOXA1 potentiates lineage-specific enhancer activation through modulating TET1 expression and function. *Nucleic Acids Res* **44**, 8153–8164 (2016).
49. Kreibich, E. & Krebs, A. R. Relevance of DNA methylation at enhancers for the acquisition of cell identities. *FEBS Letters* **597**, 1805–1817 (2023).
50. Maurano, M. T. *et al.* Role of DNA Methylation in Modulating Transcription Factor Occupancy. *Cell Reports* **12**, 1184–1195 (2015).

51. Kreibich, E., Kleinendorst, R., Barzaghi, G., Kaspar, S. & Krebs, A. R. Single-molecule footprinting identifies context-dependent regulation of enhancers by DNA methylation. *Molecular Cell* **83**, 787-802.e9 (2023).
52. Izzo, F. *et al.* DNA methylation disruption reshapes the hematopoietic differentiation landscape. *Nat Genet* **52**, 378–387 (2020).
53. Rahnamoun, H. *et al.* Mutant p53 regulates enhancer-associated H3K4 monomethylation through interactions with the methyltransferase MLL4. *J Biol Chem* **293**, 13234–13246 (2018).
54. Park, Y.-K. *et al.* Interplay of BAF and MLL4 promotes cell type-specific enhancer activation. *Nat Commun* **12**, 1630 (2021).
55. Raisner, R. *et al.* Enhancer Activity Requires CBP/P300 Bromodomain-Dependent Histone H3K27 Acetylation. *Cell Rep* **24**, 1722–1729 (2018).
56. Heintzman, N. D. *et al.* Histone Modifications at Human Enhancers Reflect Global Cell Type-Specific Gene Expression. *Nature* **459**, 108–112 (2009).
57. Cui, K. *et al.* Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. *Cell Stem Cell* **4**, 80–93 (2009).
58. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931–21936 (2010).
59. Zhang, T., Zhang, Z., Dong, Q., Xiong, J. & Zhu, B. Histone H3K27 acetylation is dispensable for enhancer activity in mouse embryonic stem cells. *Genome Biology* **21**, 45 (2020).
60. Raisner, R. *et al.* Enhancer Activity Requires CBP/P300 Bromodomain-Dependent Histone H3K27 Acetylation. *Cell Reports* **24**, 1722–1729 (2018).
61. Narita, T. *et al.* Acetylation of histone H2B marks active enhancers and predicts CBP/p300 target genes. *Nat Genet* **55**, 679–692 (2023).
62. Sinha, K. K., Bilokapic, S., Du, Y., Malik, D. & Halic, M. Histone modifications regulate pioneer transcription factor cooperativity. *Nature* **619**, 378–384 (2023).
63. Donaghey, J. *et al.* Genetic determinants and epigenetic effects of pioneer-factor occupancy. *Nat Genet* **50**, 250–258 (2018).
64. Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nat Rev Genet* **17**, 661–678 (2016).
65. Mateo, L. J. *et al.* Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* **568**, 49–54 (2019).
66. Park, D. S. *et al.* High-throughput Oligopaint screen identifies druggable 3D genome regulators. *Nature* **620**, 209–217 (2023).
67. Akgol Oksuz, B. *et al.* Systematic evaluation of chromosome conformation capture assays. *Nat Methods* **18**, 1046–1055 (2021).
68. Reddington, J. P. *et al.* Lineage-Resolved Enhancer and Promoter Usage during a Time Course of Embryogenesis. *Developmental Cell* **55**, 648-664.e9 (2020).

69. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75 (2012).
70. Zhou, T. *et al.* GAGE-seq concurrently profiles multiscale 3D genome organization and gene expression in single cells. *Nat Genet* **56**, 1701–1711 (2024).
71. Fulco, C. P. *et al.* Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nat Genet* **51**, 1664–1669 (2019).
72. Benabdallah, N. S. *et al.* Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. *Molecular Cell* **76**, 473 (2019).
73. Alexander, J. M. *et al.* Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *eLife* **8**, e41769 (2019).
74. Acuña, L. I. G., Flyamer, I., Boyle, S., Friman, E. T. & Bickmore, W. A. Transcription decouples estrogen-dependent changes in enhancer-promoter contact frequencies and spatial proximity. *PLOS Genetics* **20**, e1011277 (2024).
75. Banigan, E. J. *et al.* Transcription shapes 3D chromatin organization by interacting with loop-extruding cohesin complexes. *bioRxiv* 2022.01.07.475367 (2022) doi:10.1101/2022.01.07.475367.
76. Pollex, T. *et al.* Enhancer–promoter interactions become more instructive in the transition from cell-fate specification to tissue differentiation. *Nat Genet* **56**, 686–696 (2024).
77. Uyehara, C. M. & Apostolou, E. 3D Enhancer-promoter interactions and multi-connected hubs: Organizational principles and functional roles. *Cell reports* **42**, 112068 (2023).
78. Deng, W. *et al.* Controlling Long-Range Genomic Interactions at a Native Locus by Targeted Tethering of a Looping Factor. *Cell* **149**, 1233–1244 (2012).
79. Wang, W. *et al.* TCF-1 promotes chromatin interactions across topologically associating domains in T cell progenitors. *Nat Immunol* **23**, 1052–1062 (2022).
80. Kuznetsova, K. *et al.* Nanog organizes transcription bodies. *Current Biology* **33**, 164-173.e5 (2023).
81. Boija, A. *et al.* Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* **175**, 1842-1855.e16 (2018).
82. Bertrand, N., Castro, D. S. & Guillemot, F. Proneural genes and the specification of neural cell types. *Nat Rev Neurosci* **3**, 517–530 (2002).
83. Hassan, B. A. *et al.* *atonal* Regulates Neurite Arborization but Does Not Act as a Proneural Gene in the *Drosophila* Brain. *Neuron* **25**, 549–561 (2000).
84. Hand, R. *et al.* Phosphorylation of Neurogenin2 Specifies the Migration Properties and the Dendritic Morphology of Pyramidal Neurons in the Neocortex. *Neuron* **48**, 45–62 (2005).
85. Pacary, E. *et al.* Proneural Transcription Factors Regulate Different Steps of Cortical Neuron Migration through Rnd-Mediated Inhibition of RhoA Signaling. *Neuron* **69**, 1069–1084 (2011).

86. Ghysen, A. & Dambly-Chaudière, C. From DNA to form: the achaete-scute complex. *Genes Dev.* **2**, 495–501 (1988).

87. Villares, R. & Cabrera, C. V. The *achaete-scute* gene complex of D. melanogaster: Conserved Domains in a subset of genes required for neurogenesis and their homology to *myc*. *Cell* **50**, 415–424 (1987).

88. Murre, C., McCaw, P. S. & Baltimore, D. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, *daughterless, MyoD*, and *myc* proteins. *Cell* **56**, 777–783 (1989).

89. Jarman, A. P., Grau, Y., Jan, L. Y. & Jan, Y. N. atonal is a proneural gene that directs chordotonal organ formation in the Drosophila peripheral nervous system. *Cell* **73**, 1307–1321 (1993).

90. Goulding, S. E., zur Lage, P. & Jarman, A. P. amos, a proneural gene for Drosophila olfactory sense organs that is regulated by lozenge. *Neuron* **25**, 69–78 (2000).

91. Goulding, S. E., White, N. M. & Jarman, A. P. cato encodes a basic helix-loop-helix transcription factor implicated in the correct differentiation of Drosophila sense organs. *Dev Biol* **221**, 120–131 (2000).

92. Lee, D. G., Kim, Y.-K. & Baek, K.-H. The bHLH Transcription Factors in Neural Development and Therapeutic Applications for Neurodegenerative Diseases. *International Journal of Molecular Sciences* **23**, 13936 (2022).

93. Michael, A. K. *et al.* Cooperation between bHLH transcription factors and histones for DNA access. *Nature* **619**, 385–393 (2023).

94. Chien, C. T., Hsiao, C. D., Jan, L. Y. & Jan, Y. N. Neuronal type information encoded in the basic-helix-loop-helix domain of proneural genes. *Proc Natl Acad Sci U S A* **93**, 13239–13244 (1996).

95. Molyneaux, B. J., Arlotta, P., Menezes, J. R. L. & Macklis, J. D. Neuronal subtype specification in the cerebral cortex. *Nat Rev Neurosci* **8**, 427–437 (2007).

96. Schuurmans, C. *et al.* Sequential phases of cortical specification involve Neurogenin-dependent and -independent pathways. *EMBO J* **23**, 2892–2902 (2004).

97. Götz, M. & Huttner, W. B. The cell biology of neurogenesis. *Nat Rev Mol Cell Biol* **6**, 777–788 (2005).

98. Guillemot, F. & Hassan, B. A. Beyond proneural: emerging functions and regulations of proneural proteins. *Curr Opin Neurobiol* **42**, 93–101 (2017).

99. Altbürger, C., Rath, M., Wehrle, J. & Driever, W. The proneural factors Ascl1a and Ascl1b contribute to the terminal differentiation of dopaminergic GABAergic dual transmitter neurons in zebrafish. *Dev Biol* **505**, 58–74 (2024).

100. Bulchand, S., Grove, E. A., Porter, F. D. & Tole, S. LIM-homeodomain gene *Lhx2* regulates the formation of the cortical hem. *Mechanisms of Development* **100**, 165–175 (2001).

101. Muzio, L. *et al.* Conversion of cerebral cortex into basal ganglia in Emx2−/− Pax6Sey/Sey double-mutant mice. *Nat Neurosci* **5**, 737–745 (2002).

102. Mattar, P. *et al.* A screen for downstream effectors of *Neurogenin2* in the embryonic neocortex. *Developmental Biology* **273**, 373–389 (2004).
103. Fode, C. *et al.* A role for neural determination genes in specifying the dorsoventral identity of telencephalic neurons. *Genes Dev.* **14**, 67–80 (2000).
104. Wapinski, O. L. *et al.* Hierarchical mechanisms for transcription factor-mediated reprogramming of fibroblasts to neurons. *Cell* **155**, 10.1016/j.cell.2013.09.028 (2013).
105. Wilkinson, G., Dennis, D. & Schuurmans, C. Proneural genes in neocortical development. *Neuroscience* **253**, 256–273 (2013).
106. Hirata, H. *et al.* Oscillatory Expression of the bHLH Factor Hes1 Regulated by a Negative Feedback Loop. *Science* **298**, 840–843 (2002).
107. Imayoshi, I. *et al.* Oscillatory Control of Factors Determining Multipotency and Fate in Mouse Neural Progenitors. *Science* **342**, 1203–1208 (2013).
108. Kowalczyk, T. *et al.* Intermediate Neuronal Progenitors (Basal Progenitors) Produce Pyramidal–Projection Neurons for All Layers of Cerebral Cortex. *Cerebral Cortex* **19**, 2439–2450 (2009).
109. Shimojo, H., Masaki, T. & Kageyama, R. The Neurog2-Tbr2 axis forms a continuous transition to the neurogenic gene expression state in neural stem cells. *Developmental Cell* **59**, 1913-1923.e6 (2024).
110. Quan, X.-J. *et al.* Post-translational Control of the Temporal Dynamics of Transcription Factor Activity Regulates Neurogenesis. *Cell* **164**, 460–475 (2016).
111. McDowell, G. S., Hindley, C. J., Lippens, G., Landrieu, I. & Philpott, A. Phosphorylation in intrinsically disordered regions regulates the activity of Neurogenin2. *BMC Biochemistry* **15**, 24 (2014).
112. Hur, E.-M. & Zhou, F.-Q. GSK3 signalling in neural development. *Nat Rev Neurosci* **11**, 539–551 (2010).
113. Li, S. *et al.* GSK3 Temporally Regulates Neurogenin 2 Proneural Activity in the Neocortex. *J. Neurosci.* **32**, 7791–7805 (2012).
114. Ali, F. *et al.* Cell cycle-regulated multi-site phosphorylation of Neurogenin 2 coordinates cell cycling with differentiation during neurogenesis. *Development* **138**, 4267–4277 (2011).
115. Pereira, A. *et al.* Direct neuronal reprogramming of mouse astrocytes is associated with multiscale epigenome remodeling and requires Yy1. *Nat Neurosci* **27**, 1260–1273 (2024).
116. Bocchi, R. & Götz, M. Neuronal Reprogramming for Brain Repair: Challenges and Perspectives. *Trends in Molecular Medicine* **26**, 890–892 (2020).
117. Barker, R. A., Götz, M. & Parmar, M. New approaches for brain repair—from rescue to reprogramming. *Nature* **557**, 329–334 (2018).
118. Barker, R. A., Drouin-Ouellet, J. & Parmar, M. Cell-based therapies for Parkinson disease—past insights and future potential. *Nat Rev Neurol* **11**, 492–503 (2015).

119. Steinbeck, J. A. & Studer, L. Moving stem cells to the clinic: potential and limitations for brain repair. *Neuron* **86**, 187–206 (2015).
120. Qian, H. & Fu, X.-D. Brain Repair by Cell Replacement via In Situ Neuronal Reprogramming. *Annual Review of Genetics* **55**, 45–69 (2021).
121. Frisén, J. Neurogenesis and Gliogenesis in Nervous System Plasticity and Repair. *Annu Rev Cell Dev Biol* **32**, 127–141 (2016).
122. Arvidsson, A., Collin, T., Kirik, D., Kokaia, Z. & Lindvall, O. Neuronal replacement from endogenous precursors in the adult brain after stroke. *Nat Med* **8**, 963–970 (2002).
123. Davis, R. L., Weintraub, H. & Lassar, A. B. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**, 987–1000 (1987).
124. Dall'Agnese, A. *et al.* Transcription Factor-Directed Re-wiring of Chromatin Architecture for Somatic Cell Nuclear Reprogramming toward trans-Differentiation. *Molecular Cell* **76**, 453-472.e8 (2019).
125. Heins, N. *et al.* Glial cells generate neurons: the role of the transcription factor Pax6. *Nat Neurosci* **5**, 308–315 (2002).
126. Ohori, Y. *et al.* Growth factor treatment and genetic manipulation stimulate neurogenesis and oligodendrogenesis by endogenous neural progenitors in the injured adult spinal cord. *J Neurosci* **26**, 11948–11960 (2006).
127. Buffo, A. *et al.* Expression pattern of the transcription factor Olig2 in response to brain injuries: implications for neuronal repair. *Proc Natl Acad Sci U S A* **102**, 18183–18188 (2005).
128. Kronenberg, G. *et al.* Modulation of fate determinants Olig2 and Pax6 in resident glia evokes spiking neuroblasts in a model of mild brain ischemia. *Stroke* **41**, 2944–2949 (2010).
129. Lei, W., Li, W., Ge, L. & Chen, G. Non-engineered and Engineered Adult Neurogenesis in Mammalian Brains. *Front. Neurosci.* **13**, (2019).
130. Heinrich, C. *et al.* Directing astroglia from the cerebral cortex into subtype specific functional neurons. *PLoS Biol* **8**, e1000373 (2010).
131. Berninger, B. *et al.* Functional properties of neurons derived from in vitro reprogrammed postnatal astroglia. *J Neurosci* **27**, 8654–8664 (2007).
132. Gascón, S. *et al.* Identification and Successful Negotiation of a Metabolic Checkpoint in Direct Neuronal Reprogramming. *Cell Stem Cell* **18**, 396–409 (2016).
133. Smith, D. K., Yang, J., Liu, M.-L. & Zhang, C.-L. Small Molecules Modulate Chromatin Accessibility to Promote NEUROG2-Mediated Fibroblast-to-Neuron Reprogramming. *Stem Cell Reports* **7**, 955–969 (2016).
134. Mattugini, N. *et al.* Inducing Different Neuronal Subtypes from Astrocytes in the Injured Mouse Cerebral Cortex. *Neuron* **103**, 1086-1095.e5 (2019).
135. Aydin, B. *et al.* Proneural factors Ascl1 and Neurog2 contribute to neuronal subtype identities by establishing distinct chromatin landscapes. *Nat Neurosci* **22**, 897–908 (2019).

136. Noack, F. *et al.* Multimodal profiling of the transcriptional regulatory landscape of the developing mouse cortex identifies Neurog2 as a key epigenome remodeler. *Nat Neurosci* **25**, 154–167 (2022).
137. Krijger, P. H. L. *et al.* Cell-of-Origin-Specific 3D Genome Structure Acquired during Somatic Cell Reprogramming. *Cell Stem Cell* **18**, 597–610 (2016).
138. Beagan, J. A. *et al.* Local Genome Topology Can Exhibit an Incompletely Rewired 3D-Folding State during Somatic Cell Reprogramming. *Cell Stem Cell* **18**, 611–624 (2016).
139. Ninkovic, J. & Götz, M. Understanding direct neuronal reprogramming — from pioneer factors to 3D chromatin. *Current Opinion in Genetics & Development* **52**, 65–69 (2018).
140. Sonsalla, G. *et al.* Direct neuronal reprogramming of NDUFS4 patient cells identifies the unfolded protein response as a novel general reprogramming hurdle. *Neuron* **112**, 1117-1132.e9 (2024).
141. Weintraub, A. S. *et al.* YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* **171**, 1573-1588.e28 (2017).
142. Verheul, T. C. J., van Hijfte, L., Perenthaler, E. & Barakat, T. S. The Why of YY1: Mechanisms of Transcriptional Regulation by Yin Yang 1. *Front Cell Dev Biol* **8**, 592164 (2020).
143. Pereira, A. *et al.* Direct neuronal reprogramming of mouse astrocytes is associated with multiscale epigenome remodeling and requires Yy1. *Nat Neurosci* **27**, 1260–1273 (2024).
144. Henke, R. M. *et al.* Neurog2 is a direct downstream target of the Ptf1a-Rbpj transcription complex in dorsal spinal cord. *Development* **136**, 2945–2954 (2009).
145. Ma, Y.-C. *et al.* Regulation of motor neuron specification by phosphorylation of neurogenin 2. *Neuron* **58**, 65–77 (2008).
146. Kempf, J. *et al.* Heterogeneity of neurons reprogrammed from spinal cord astrocytes by the proneural factors Ascl1 and Neurogenin2. *Cell Reports* **36**, (2021).
147. Liu, X. *et al.* Direct reprogramming of human fibroblasts into dopaminergic neuron-like cells. *Cell Res* **22**, 321–332 (2012).
148. Manelli, V. *et al.* Context-dependent epigenome rewiring during neuronal differentiation. 2024.10.18.618996 Preprint at https://doi.org/10.1101/2024.10.18.618996 (2024).
149. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* **6**, 377–382 (2009).
150. Baysoy, A., Bai, Z., Satija, R. & Fan, R. The technological landscape and applications of single-cell multi-omics. *Nat Rev Mol Cell Biol* **24**, 695–713 (2023).
151. Vandereyken, K., Sifrim, A., Thienpont, B. & Voet, T. Methods and applications for single-cell and spatial multi-omics. *Nat Rev Genet* **24**, 494–515 (2023).

152. Lee, J., Hyeon, D. Y. & Hwang, D. Single-cell multiomics: technologies and data analysis methods. *Exp Mol Med* **52**, 1428–1442 (2020).
153. Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**, 557-572.e24 (2017).
154. Jin, W. *et al.* Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. *Nature* **528**, 142–146 (2015).
155. Carter, B. & Zhao, K. The epigenetic basis of cellular heterogeneity. *Nat Rev Genet* **22**, 235–250 (2021).
156. Song, Y. *et al.* Dynamic Enhancer DNA Methylation as Basis for Transcriptional and Cellular Heterogeneity of ESCs. *Mol Cell* **75**, 905-920.e6 (2019).
157. Angermueller, C. *et al.* Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat Methods* **13**, 229–232 (2016).
158. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).
159. Bintu, B. *et al.* Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**, eaau1783 (2018).
160. Stevens, T. J. *et al.* 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59–64 (2017).
161. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).
162. Stoeckius, M. *et al.* Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biology* **19**, 224 (2018).
163. Lobato-Moreno, S. *et al.* Scalable ultra-high-throughput single-cell chromatin and RNA sequencing reveals gene regulatory dynamics linking macrophage polarization to autoimmune disease. 2023.12.26.573253 Preprint at https://doi.org/10.1101/2023.12.26.573253 (2024).
164. Rosenberg, A. B. *et al.* Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* (2018) doi:10.1126/science.aam8999.
165. McCord, R. P., Kaplan, N. & Giorgetti, L. Chromosome Conformation Capture and Beyond: Toward an Integrative View of Chromosome Structure and Function. *Molecular Cell* **77**, 688–708 (2020).
166. Lando, D. *et al.* Combining fluorescence imaging with Hi-C to study 3D genome architecture of the same single cell. *Nat Protoc* **13**, 1034–1061 (2018).
167. Quinodoz, S. A. *et al.* SPRITE: a genome-wide method for mapping higher-order 3D interactions in the nucleus using combinatorial split-and-pool barcoding. *Nat Protoc* **17**, 36–75 (2022).
168. Beagrie, R. A. *et al.* Complex multi-enhancer contacts captured by Genome Architecture Mapping (GAM). *Nature* **543**, 519–524 (2017).
169. Lafontaine, D. L., Yang, L., Dekker, J. & Gibcus, J. H. Hi-C 3.0: Improved Protocol for Genome-Wide Chromosome Conformation Capture. *Curr Protoc* **1**, e198 (2021).

170. Krietenstein, N. *et al.* Ultrastructural Details of Mammalian Chromosome Architecture. *Mol Cell* **78**, 554-565.e7 (2020).
171. Schoenfelder, S., Javierre, B.-M., Furlan-Magaril, M., Wingett, S. W. & Fraser, P. Promoter Capture Hi-C: High-resolution, Genome-wide Profiling of Promoter Interactions. *J Vis Exp* 57320 (2018) doi:10.3791/57320.
172. Mumbach, M. R. *et al.* HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods* **13**, 919–922 (2016).
173. Hua, P. *et al.* Defining genome architecture at base-pair resolution. *Nature* **595**, 125–129 (2021).
174. Aljahani, A. *et al.* Analysis of sub-kilobase chromatin topology reveals nano-scale regulatory interactions with variable dependence on cohesin and CTCF. *Nat Commun* **13**, 2139 (2022).
175. Goel, V. Y., Huseyin, M. K. & Hansen, A. S. Region Capture Micro-C reveals coalescence of enhancers and promoters into nested microcompartments. *Nat Genet* **55**, 1048–1056 (2023).
176. Nagano, T. *et al.* Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **547**, 61–67 (2017).
177. Flyamer, I. M. *et al.* Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* **544**, 110–114 (2017).
178. Stevens, T. J. *et al.* 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59–64 (2017).
179. Ramani, V. *et al.* Sci-Hi-C: A single-cell Hi-C method for mapping 3D genome organization in large number of single cells. *Methods* **170**, 61–68 (2020).
180. Tan, L., Xing, D., Chang, C.-H., Li, H. & Xie, X. S. Three-dimensional genome structures of single diploid human cells. *Science* **361**, 924–928 (2018).
181. Arrastia, M. V. *et al.* Single-cell measurement of higher-order 3D genome organization with scSPRITE. *Nat Biotechnol* **40**, 64–73 (2022).
182. Li, W. *et al.* scNanoHi-C: a single-cell long-read concatemer sequencing method to reveal high-order chromatin structures within individual cells. *Nat Methods* **20**, 1493–1505 (2023).
183. Chang, L. *et al.* Droplet Hi-C enables scalable, single-cell profiling of chromatin architecture in heterogeneous tissues. *Nat Biotechnol* 1–14 (2024) doi:10.1038/s41587-024-02447-1.
184. Zhang, R., Zhou, T. & Ma, J. Multiscale and integrative single-cell Hi-C analysis with Higashi. *Nat Biotechnol* **40**, 254–261 (2022).
185. Xiong, K., Zhang, R. & Ma, J. scGHOST: Identifying single-cell 3D genome subcompartments. *Nat Methods* **21**, 814–822 (2024).
186. Zhou, J. *et al.* Robust single-cell Hi-C clustering by convolution- and random-walk–based imputation. *Proceedings of the National Academy of Sciences* **116**, 14011–14018 (2019).

187. Liu, Z. *et al.* Linking genome structures to functions by simultaneous single-cell Hi-C and RNA-seq. *Science* **380**, 1070–1076 (2023).
188. Wu, H. *et al.* Simultaneous single-cell three-dimensional genome and gene expression profiling uncovers dynamic enhancer connectivity underlying olfactory receptor choice. *Nat Methods* **21**, 974–982 (2024).
189. Lay, F. D., Kelly, T. K. & Jones, P. A. Nucleosome Occupancy and Methylome Sequencing (NOMe-seq). *Methods Mol Biol* **1708**, 267–284 (2018).
190. Clark, S. J. *et al.* scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat Commun* **9**, 781 (2018).
191. Noack, F. *et al.* Joint epigenome profiling reveals cell-type-specific gene regulatory programmes in human cortical organoids. *Nat Cell Biol* **25**, 1873–1883 (2023).
192. Li, G. *et al.* Joint profiling of DNA methylation and chromatin architecture in single cells. *Nat Methods* **16**, 991–993 (2019).
193. Lee, D.-S. *et al.* Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. *Nat Methods* **16**, 999–1006 (2019).
194. Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95–98 (2016).
195. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
196. Iacovino, M. *et al.* Inducible cassette exchange: a rapid and efficient system enabling conditional gene expression in embryonic stem and primary cells. *Stem Cells* **29**, 1580–1588 (2011).
197. Schwartzman, O. *et al.* UMI-4C for quantitative and targeted chromosomal contact profiling. *Nat Methods* **13**, 685–691 (2016).
198. Noack, F. *et al.* Assessment and site-specific manipulation of DNA (hydroxy-)methylation during mouse corticogenesis. *Life Science Alliance* **2**, (2019).
199. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
200. Cohen, N. M. *et al.* SHAMAN: bin-free randomization, normalization and screening of Hi-C matrices. 187203 Preprint at https://doi.org/10.1101/187203 (2017).
201. Machlab, D. *et al.* monaLisa: an R/Bioconductor package for identifying regulatory motifs. *Bioinformatics* **38**, 2624–2625 (2022).
202. Bentsen, M., Heger, V., Schultheis, H., Kuenne, C. & Looso, M. TF-COMB - Discovering grammar of transcription factor binding sites. *Comput Struct Biotechnol J* **20**, 4040–4051 (2022).
203. Raine, A., Manlig, E., Wahlberg, P., Syvänen, A.-C. & Nordlund, J. SPlinted Ligation Adapter Tagging (SPLAT), a novel library preparation method for whole genome bisulphite sequencing. *Nucleic Acids Res* **45**, e36 (2017).
204. Sullivan, D. K. & Pachter, L. Flexible parsing, interpretation, and editing of technical sequences with splitcode. *bioRxiv* 2023.03.20.533521 (2023) doi:10.1101/2023.03.20.533521.

205. Liu, H. *et al.* DNA methylation atlas of the mouse brain at single-cell resolution. *Nature* **598**, 120–128 (2021).
206. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).
207. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
208. Zurkirchen, L. *et al.* Yin Yang 1 sustains biosynthetic demands during brain development in a stage-specific manner. *Nat Commun* **10**, 2192 (2019).
209. Oproescu, A.-M., Han, S. & Schuurmans, C. New Insights Into the Intricacies of Proneural Gene Regulation in the Embryonic and Adult Cerebral Cortex. *Front Mol Neurosci* **14**, 642016 (2021).
210. Granja, J. M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet* **53**, 403–411 (2021).
211. Connor, S. A. *et al.* Loss of Synapse Repressor MDGA1 Enhances Perisomatic Inhibition, Confers Resistance to Network Excitation, and Impairs Cognitive Function. *Cell Rep* **21**, 3637–3645 (2017).
212. Lee, K. *et al.* MDGAs interact selectively with neuroligin-2 but not other neuroligins to regulate inhibitory synapse development. *Proc Natl Acad Sci U S A* **110**, 336–341 (2013).
213. Wang, Y., Hong, Q., Xia, Y., Zhang, Z. & Wen, B. The Lysine Demethylase KDM7A Regulates Immediate Early Genes in Neurons. *Adv Sci (Weinh)* **10**, 2301367 (2023).
214. Beagan, J. A. *et al.* YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome Res* **27**, 1139–1152 (2017).
215. Lin, H.-C. *et al.* NGN2 induces diverse neuron types from human pluripotency. *Stem Cell Reports* **16**, 2118–2127 (2021).
216. Masserdotti, G. *et al.* Transcriptional Mechanisms of Proneural Factors and REST in Regulating Neuronal Reprogramming of Astrocytes. *Cell Stem Cell* **17**, 74–88 (2015).
217. Heng, J. I.-T. *et al.* Neurogenin 2 controls cortical neuron migration through regulation of Rnd2. *Nature* **455**, 114–118 (2008).
218. Kovach, C. *et al.* Neurog2 simultaneously activates and represses alternative gene expression programs in the developing neocortex. *Cereb Cortex* **23**, 1884–1900 (2013).
219. Bohrer, C. *et al.* The balance of Id3 and E47 determines neural stem/precursor cell differentiation into astrocytes. *The EMBO Journal* **34**, 2804–2819 (2015).
220. Vrenken, K. S. *et al.* The transcriptional repressor SNAI2 impairs neuroblastoma differentiation and inhibits response to retinoic acid therapy. *Biochim Biophys Acta Mol Basis Dis* **1866**, 165644 (2020).
221. Menge, S., Decker, L. & Freischmidt, A. Restoring expression of Stathmin-2: a novel strategy to treat TDP-43 proteinopathies. *Sig Transduct Target Ther* **8**, 1–2 (2023).

222. Bormuth, I. *et al.* Neuronal Basic Helix–Loop–Helix Proteins Neurod2/6 Regulate Cortical Commissure Formation before Midline Interactions. *J. Neurosci.* **33**, 641–651 (2013).
223. Amador-Arjona, A. *et al.* SOX2 primes the epigenetic landscape in neural precursors enabling proper gene activation during hippocampal neurogenesis. *Proc Natl Acad Sci U S A* **112**, E1936-1945 (2015).
224. Bunt, J. *et al.* Combined allelic dosage of Nfia and Nfib regulates cortical development. *Brain and Neuroscience Advances* **1**, 2398212817739433 (2017).
225. Jin, J. *et al.* A Mammalian Chromatin Remodeling Complex with Similarities to the Yeast INO80 Complex *. *Journal of Biological Chemistry* **280**, 41207–41212 (2005).
226. Kim, B. *et al.* Neuronal activity-induced BRG1 phosphorylation regulates enhancer activation. *Cell Reports* **36**, (2021).
227. Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**, 1665–1680 (2014).
228. Ma, S. *et al.* Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. *Cell* **183**, 1103-1116.e20 (2020).
229. Belaghzal, H., Dekker, J. & Gibcus, J. H. HI-C 2.0: AN OPTIMIZED HI-C PROCEDURE FOR HIGH-RESOLUTION GENOME-WIDE MAPPING OF CHROMOSOME CONFORMATION. *Methods* **123**, 56–65 (2017).
230. Molloy, P. L. & Symons, R. H. Cleavage of DNA.RNA hybrids by Type II restriction enzymes. *Nucleic Acids Research* **8**, 2939–2946 (1980).
231. Mendelsohn, S. L. & Young, D. A. Inhibition of ribonuclease. Efficacy of sodium dodecyl sulfate, diethyl pyrocarbonate, protein ase K and heparin using a sensitive ribonuclease assay. *Biochim Biophys Acta* **519**, 461–473 (1978).
232. Li, Y. & Tollefsbol, T. O. DNA methylation detection: Bisulfite genomic sequencing analysis. *Methods Mol Biol* **791**, 11–21 (2011).
233. Vaisvila, R. *et al.* Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. *Genome Res.* **31**, 1280–1289 (2021).
234. Kint, S., De Spiegelaere, W., De Kesel, J., Vandekerckhove, L. & Van Criekinge, W. Evaluation of bisulfite kits for DNA methylation profiling in terms of DNA fragmentation and DNA recovery using digital PCR. *PLoS One* **13**, e0199091 (2018).
235. Mayran, A. *et al.* Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate. *Nat Genet* **50**, 259–269 (2018).
236. Wang, R. *et al.* MyoD is a 3D genome structure organizer for muscle cell identity. *Nat Commun* **13**, 205 (2022).
237. Cernilogar, F. M. *et al.* Pre-marked chromatin and transcription factor co-binding shape the pioneering activity of Foxa2. *Nucleic Acids Research* **47**, 9069–9086 (2019).
238. Carminati, M., Vecchia, L., Stoos, L. & Thomä, N. H. Pioneer factors: Emerging rules of engagement for transcription factors on chromatinized DNA. *Current Opinion in Structural Biology* **88**, 102875 (2024).

239. Nishimura, M., Takizawa, Y., Nozawa, K. & Kurumizaka, H. Structural basis for p53 binding to its nucleosomal target DNA sequence. *PNAS Nexus* **1**, pgac177 (2022).
240. Reizel, Y. *et al.* FoxA-dependent demethylation of DNA initiates epigenetic memory of cellular identity. *Developmental Cell* **56**, 602-612.e4 (2021).
241. Sardina, J. L. *et al.* Transcription Factors Drive Tet2-Mediated Enhancer Demethylation to Reprogram Cell Fate. *Cell Stem Cell* **23**, 727-741.e9 (2018).
242. Donaghey, J. *et al.* Genetic determinants and epigenetic effects of pioneer-factor occupancy. *Nat Genet* **50**, 250–258 (2018).
243. Li, H., Playter, C., Das, P. & McCord, R. P. Chromosome compartmentalization: causes, changes, consequences, and conundrums. *Trends in Cell Biology* **34**, 707–727 (2024).
244. Davidson, I. F. *et al.* DNA loop extrusion by human cohesin. *Science* **366**, 1338–1345 (2019).
245. Fournier, M. *et al.* FOXA and master transcription factors recruit Mediator and Cohesin to the core transcriptional regulatory circuitry of cancer cells. *Sci Rep* **6**, 34962 (2016).
246. Asprer, J. S. T. *et al.* LMO4 functions as a co-activator of neurogenin 2 in the developing cortex. *Development* **138**, 2823–2832 (2011).
247. Ali, F. R. *et al.* The phosphorylation status of Ascl1 is a key determinant of neuronal differentiation and maturation in vivo and in vitro. *Development* **141**, 2216–2224 (2014).
248. Bocchi, R., Masserdotti, G. & Götz, M. Direct neuronal reprogramming: Fast forward from new concepts toward therapeutic approaches. *Neuron* **110**, 366–393 (2022).
249. Barral, A. & Zaret, K. S. Pioneer factors: roles and their regulation in development. *Trends in Genetics* **40**, 134–148 (2024).
250. Hulme, A. J., Maksour, S., St-Clair Glover, M., Miellet, S. & Dottori, M. Making neurons, made easy: The use of Neurogenin-2 in neuronal differentiation. *Stem Cell Reports* **17**, 14–34 (2021).
251. Efroni, S. *et al.* Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell* **2**, 437–447 (2008).
252. Atlasi, Y. & Stunnenberg, H. G. The interplay of epigenetic marks during stem cell differentiation and development. *Nat Rev Genet* **18**, 643–658 (2017).
253. Sun, Y. *et al.* Neurogenin promotes neurogenesis and inhibits glial differentiation by independent mechanisms. *Cell* **104**, 365–376 (2001).
254. Xue, Y. *et al.* NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Mol Cell* **2**, 851–861 (1998).
255. Bornelöv, S. *et al.* The Nucleosome Remodeling and Deacetylation Complex Modulates Chromatin Structure at Sites of Active Transcription to Fine-Tune Gene Expression. *Mol Cell* **71**, 56-72.e4 (2018).

256. E Hermosilla, V. *et al.* SALL2 represses cyclins D1 and E1 expression and restrains G1/S cell cycle transition and cancer-related phenotypes. *Mol Oncol* **12**, 1026–1046 (2018).
257. Xiong, H. *et al.* SALL2 regulates neural differentiation of mouse embryonic stem cells through Tuba1a. *Cell Death Dis* **15**, 1–15 (2024).
258. Matsui, S. *et al.* Pioneer and PRDM transcription factors coordinate bivalent epigenetic states to safeguard cell fate. *Mol Cell* **84**, 476-489.e10 (2024).
259. Argelaguet, R. *et al.* MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biology* **21**, 111 (2020).
260. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat Methods* **18**, 1333–1341 (2021).
261. Baysoy, A., Bai, Z., Satija, R. & Fan, R. The technological landscape and applications of single-cell multi-omics. *Nat Rev Mol Cell Biol* **24**, 695–713 (2023).
262. Heffel, M. G. *et al.* Temporally distinct 3D multi-omic dynamics in the developing human brain. *Nature* **635**, 481–489 (2024).
263. Tian, W. *et al.* Single Cell DNA Methylation and 3D Genome Architecture in the Human Brain. *Science* **382**, eadf5357 (2023).
264. Rakic, P. Evolution of the neocortex: a perspective from developmental biology. *Nat Rev Neurosci* **10**, 724–735 (2009).
265. Dixit, A. *et al.* Perturb-seq: Dissecting molecular circuits with scalable single cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853-1866.e17 (2016).

Acknowledgements

I would like to express my sincere gratitude to all those who supported me throughout this academic journey.

I am deeply thankful to my supervisor, Dr. Boyan Bonev, for his trust in me and my work. His guidance, patience, and willingness to engage in thoughtful discussions were essential to my development as a researcher. During challenging moments, his support and encouragement propelled both me and the project to new levels of achievement.

I am also grateful to the members of the lab, whose collaborative spirit and dedication fostered an environment of mutual respect and intellectual growth. It has been an honour to work alongside such a talented and dynamic group of individuals.

My sincere thanks extend to all my collaborators, experts, and the many individuals I encountered in this scientific journey. Your insights and inspiration have been invaluable in shaping the direction of this work.

I would like to acknowledge my TAC members, Prof. Gunnar Schotta and Prof. Magdalena Götz, for their thoughtful feedback and stimulating discussions, which have been instrumental in refining my research.

Additionally, I am deeply appreciative of my scientific mentors, whose guidance has been pivotal in my academic journey. From Dr. Samantha Hughes, who provided me with my first laboratory internship, to Dr. Allison Wollard, who supported me as a tutor during my undergraduate studies, your mentorship has been foundational. I am also grateful to Prof. Wolf Reik and Dr. Aled Parry, who ignited my passion for epigenetics, and to Prof. Rob Klose and Dr. Neil Blackledge, whose insights have further fuelled my curiosity and academic growth in this field.

Finally, I wish to express my heartfelt appreciation to my parents. Their unwavering support and encouragement have been a constant source of strength, and I could not have completed this journey without them.

May this PhD represent not the conclusion of a chapter, but the beginning of a new and exciting phase. As Steve Jobs once said, “The people who are crazy enough to think they can change the world are the ones who do.”

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

Dean’s Office Medical Faculty
Faculty of Medicine

Affidavit

Diwakar, Jei

Surname, first name

I hereby declare, that the submitted thesis entitled

Development and use of multiomics tools to dissect drivers of neuronal identity

is my own work. I have only used the sources indicated and have not made unauthorised use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the dissertation presented here has not been submitted in the same or similar form to any other institution for the purpose of obtaining an academic degree.

Munich, 08/12/24

Place, Date

Jeisimhan Diwakar Shunmugapriya

Signature doctoral candidate

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

Dean's Office Medical Faculty
Doctoral Office

Confirmation of congruency between printed and electronic version of the doctoral thesis

Diwakar, Jei

Surname, first name

I hereby declare that the electronic version of the submitted thesis, entitled

Development and use of multiomics tools to dissect drivers of neuronal identity

is congruent with the printed version both in content and format.

Munich, 08/12/24

Place, Date

Jeisimhan Diwakar Shunmugapriya

Signature doctoral candidate

List of Publications

Pereira, A., **Diwakar, J.,** Masserdotti, G., Beşkardeş, S., Simon, T., So, Y., Martín-Loarte, L., Bergemann, F., Vasan, L., Schauer, T., Danese, A., Bocchi, R., Colomé-Tatché, M., Schuurmans, C., Philpott, A., Straub, T., Bonev, B., & Götz, M. (2024). Direct neuronal reprogramming of mouse astrocytes is associated with multiscale epigenome remodeling and requires Yy1. Nature Neuroscience, 27(7), 1260–1273. https://doi.org/10.1038/s41593-024-01677-5

Noack, F., Vangelisti, S., Raffl, G., Carido, M., **Diwakar, J.**, Chong, F., & Bonev, B. (2022). Multimodal profiling of the transcriptional regulatory landscape of the developing mouse cortex identifies Neurog2 as a key epigenome remodeler. Nature Neuroscience, 25(2), 154–167. https://doi.org/10.1038/s41593-021-01002-4

Preprint

Manelli, V., **Diwakar, J.**, Beşkardeş, S., Alonso-Gil, D., Forné, I., Chong, F., Imhof, A., & Bonev, B. (2024). Context-dependent epigenome rewiring during neuronal differentiation (p. 2024.10.18.618996). bioRxiv. https://doi.org/10.1101/2024.10.18.618996