DIPLOMARBEIT

GPSEQ MAPS THE RADIAL GENOME ORGANIZATION DURING NEURAL DIFFERENTIATION

ausgeführt am Institut für Quantitative Genombiologie
des Karolinska Institut, Stockholm

unter der Anleitung von
Magda Bienko & Nicola Crosetto
(Karolinska Institute Stockholm)
&
Robert Mach
(TU Wien)

durch

Merula Johanna Stout

Mostgasse 8a/5,
A 1040 Wien

May 26, 2020

Unterschrift BetreuerIn

Unterschrift StudentIn

Contents

1 Project Aim **1**

2 Abstract **1**

3 Kurzfassung **2**

4 Introduction **3**

5 Background **3**

5.1 Packaging of DNA inside the nucleus & folding of the chromatin fiber 3
5.2 Hetero- vs. Euchromatin 5
5.3 Compartmentalization of the genome 6
5.4 TADs & sub-TADs 7
5.5 Chromosome territories 8
5.6 Dynamics of genome organization & chromosome repositioning throughout the cell cycle 9
5.7 Long-range contacts 11
5.8 Changes in the architecture of the nucleus during neural differentiation 12
5.9 iPS technology 14
5.10 Microscopy: FISH 15
5.11 Next Generation Sequencing 16
5.12 Methodological considerations 17
    5.12.1 Detailed explanation of the theory behind GPSeq 18
    5.12.2 Choosing a restriction enzyme 20
    5.12.3 Choosing the number of restriction times 21
    5.12.4 Amplification steps during library preparation 21

6 Materials and Methods **23**

6.1 Differentiation of NES to Neurons 23
6.2 Evaluation of the variation in Hoechst-signal 25
6.3 Genomic loci Positioning by Sequencing 27
    6.3.1 YFISH 27
    6.3.2 Preparation of YFISH and GPSeq adapters 28
    6.3.3 Wide-field epifluorescence microscopy 28
    6.3.4 Image processing 29
    6.3.5 Sample preparation for Sequencing 29
    6.3.6 Variation of experimental parameters 30
    6.3.7 Processing of GPSeq sequencing data GPSeq score calculation 32

6.3.8 Summary of experimental data 33

7 Results and Discussion 37

7.1 Evaluation of variation in Hoechst-Signal 37

7.2 GPSeq 39

7.2.1 YFISH reveals gradual digestion with HindIII and DpnII 40

7.2.2 Sequencing showed satisfying numbers of bins but issues with linkers were observed 46

7.2.3 GPSeq score correlation shows reproducibility of results 51

7.2.4 Principal Component Analysis shows clustering of samples from same cell type 53

7.2.5 GC-content and chromosome size 55

7.2.6 Finding regions of interest 57

8 Conclusions and Future Recommendations 61

9 Acknowledgements 63

References 65

Appendices 78

A Full GPSeq Protocol 78

B Differentiation Protocol 90

C List of Fluorescent Oligonucleotides 92

D List of Primer Sequences 94

List of Abbreviations

3C	chromosome conformation capture
4C	circular chromosome conformation capture
5C	chromosome conformation capture carbon copy
bFGF	fibroblast growth factor-basic
BSA	Bovine Serum Albumin
CBS5	CCTF binding site 5
chr	chromosome
CRT	cyclic reversible termination
CT	chromosomal territory
CTCF	CCCTC-binding factor
DMEM	Dulbecco's Modified Eagles Medium
DSB	double strand break
ESC	embryonic stem cell
EGF	epidermal growth factor
FA	formamide
FISH	fluorescence *in situ* hybridization
FPS	a buffer: 50 % FA/50 mM phosphate buffer/2X SSC
gDNA	genomic DNA
GPSeq	Genomic loci Positioning by Sequencing
HOX	homeobox
HSA18	human chromosome 18
HSA19	human chromosome 19
H1	Histone 1

iPSCs	induced pluripotent stem cells
IVT	*in vitro* transcription
LAD	lamina-associated domain
NES	neuroepithelial stem cells (NES)
NGS	Next-generation Sequencing
NPC	Neural Progenitor Cell
NOR	nucleolar organising regions
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PFA	para-formaldehyde
PNK	polynuclotide kinase
PTM	post-translational modification
RA5	RNA 5' adapter
RS	recognition site
RT	room temperature
SINE	short interspersed nuclear element
SSC	a buffer: saline sodium citrate
TAD	topologically associated domain
TSS	transcription start site
UMI	unique molecular identifier
YFISH	Y- Fluorescence *in situ* Hybridization where the adapter is Y-shaped

List of Tables

1 Experimental set-up to evaluate the dapi channel-signal variation . . . 26
2 Microscopy settings to evaluate dapi channel-signal variation 27
3 Key Specifications Ultra-888 Andor Camera 29
4 Summary of experimental data I 34
5 Summary of experimental data II 35
6 DNA concentrations of neuron samples 47
7 Number of reads from sequencing runs 47
8 GPSeq score correlation with chromosome size and GC-content 55
9 List of oligos used to make YFISH and GPSeq adapters 93

List of Figures

1 Packaging of DNA inside the nucleus 4
2 Representaiton of CTs 8
3 Template Immobilization 17
4 Scheme discribing the YFISH procedure 19
5 Scheme describing GPSeq library preparation 20
6 Illustration of Differentiation 23
7 Brightfield images of NES, NPC and Neurons 24
8 Plots showing signal variation in dapi channel 38
9 Close-up of a digesiton process 40
10 Digestion of NES samples 42
11 Digestion of neurons 43
12 Comparison of neurons and NPC 45
13 Profiles of NPC samples 45
14 BioA profiles 46
15 Number of reads after filtering steps (NES) 48
16 Number of reads after filtering steps (neurons) 49
17 GPSeq score correlation 51
18 Principal component analysis 53
19 Correlation between GPSeq score and Chromosome size 55
20 Case study: chr21 57
21 Identification of "strongly" different regions 58

Note

Due to the outbreak of COVID-19, experimental work had to be discontinued at the beginning of March 2020. The result was that one sample was not sequenced (second replicate of neurons) and thus no detailed analysis on variation of radiality between differentiation stages has been performed yet. The results outlined in this thesis are based on the experiments carried out before shut-down of the laboratory.

1 Project Aim

The aim of this diploma thesis is to assess radial genome organization during neural differentiation. This is done using both a microscopy-based assay (YFISH = Y- Fluorescence in Situ Hybridization), as well as high-throughput sequencing methods (GPSeq) to obtain genome-wide maps of the radial location of DNA loci and to quantify the radial distribution of chromatin.

2 Abstract

As stem cells progress from an undifferentiated to a differentiated state, their purpose and functions change. Processes such as DNA replication and transcription are crucial for cell differentiation and are modulated by the three-dimensional organization of chromatin. Here, the method Genomic loci Positioning by Sequencing (GPSeq) is used to investigate the radial rearrangement of chromatin during neural differentiation as the cells progress from neuroepithelial stem cells (NES) to neuroprogenitor cells (NPC) and finally to neurons. It was shown, that both the enzyme HindIII as well as DpnII can be used for digestion where restriction with DpnII achieves higher resolution. Novel features, such as tubular structures and small dots spread through the nuclei, were observed in the YFISH experiment. Moreover, the generated GPSeq score profiles acquired from the sequencing data, show high correlation at the same differentiation stage but lower correlation at different differentiation stages, indicating a difference in radial arrangement of chromatin between NES, NPC and neurons. Principal Component Analysis could differentiate the cell types and to compare them "strongly different regions" were inferred using a global threshold. The data from the performed experiments suggest a rearrangement of chromatin during differentiation.

3 Kurzfassung

Durch das Fortschreiten der Differenzierung von Stammzellen zu spezialisierten Zellen ändert sich deren Zweck und Funktion. Prozesse wie DNA Replikation und Transkription sind entscheidend für die Differenzierung von Zellen und werden durch die drei-dimensionale Organisation von Chromatin gesteuert. In dieser Arbeit wird die Methode Genomic loci Positioning by Sequencing (GPSeq) verwendet, um die radiale Anordnung von Chromatin während neuronaler Differenzierung - also wenn die Zellen sich von neuroepithelialen Stammzellen bis hin zu Neuronen entwickeln - zu untersuchen. Es wurde gezeigt, dass sowohl das Enzym HindIII als auch DpnII für den enzymatischen Verdau verwendet werden kann, wobei man unter Verwendung von DpnII eine höhere Auflösung erzielt. Neue Entdeckungen, wie zum Beispiel röhrenförmige Strukturen und kleine Punkte, die sich durch den ganzen Nucleus ziehen, wurden durch die YFISH(Y-Fluoreszenz in situ Hybridisierung)-Experimente gemacht. Außerdem zeigen die aus den Sequenzierdaten generierten GPSeq score-Profile hohe Korrelation in der gleichen Differenzierungsphase, aber niedrigere Korrelation in verschiedenen Differenzierungsphasen. Dies weist auf unterschiedliche radiale Anordnungen des Chromatins abhängig vom untersuchten Zelltyp hin. Durch eine Hauptkomponentenanalyse konnten neuroepitheliale Stammzellen, neurale Vorläuferzellen und Neuronen unterschieden werden. Um die unterschiedlichen Zelltypen zu vergleichen, wurden „stark unterschiedliche Regionen“ unter Zuhilfenahme eines globalen Grenzwertes gesucht. Die Daten, die durch die durchgeführten Experimente erhoben wurden, deuten auf eine Reorganisation des Chromatins während der Differenzierung hin.

4 Introduction

The three-dimensional organization of chromatin modulates biological processes such as DNA replication, transcription, cell division and meiosis. These processes are crucial for cell differentiation and animal development. While it is clear that the nuclear periphery and centre represent distinct environments characterised by a different chromosomal content and by different levels of chromatin compaction and function, we do not know how the genome is spatially arranged along individual nucleic radii and how the radial organisation of the nucleus is established and maintained. The newly established method Genomic loci Positioning by Sequencing (GPSeq)[1] allows us to generate reproducible maps of the radial organization of the human genome. This can reveal fundamental aspects of genome architecture, that have important implications on genome function and stability.

Neuroepithelial Stem Cells (NES), Neuroprogenitor Cells (NPC) and Neurons represent distinct stages in neuronal differentiation. How and if the radial arrangement changes between these stages and what the implications of this on processes such as DNA replication and transcription could be, is not well understood yet. Multidisciplinary work combining cell and molecular biology assays, sequencing and microscopy methods and bioinformatic tools is needed to unravel these fundamental questions.

5 Background

5.1 Packaging of DNA inside the nucleus & folding of the chromatin fiber

For the regulation of nuclear processes such as DNA transcription, replication, recombination and repair, the structure of chromatin plays a fundamental role[2]. Gene expression is not only modulated by the DNA sequence, that is the order of nucleotides, but also by the three-dimensional (re-)arrangement of chromatin. To understand the details of how these processes are shaped and what happens during differentiation, it is therefore of importance to elucidate the radial organization of DNA in the nucleus.

Rather than being solely defined by its linear DNA sequence, the genome shows a far more complex 3D architecture. Chromosomes (chr) are formed through hierarchical folding of the DNA helix to compact the DNA for it to fit into the limited space that is the cell nucleus[3]. In eukaryotes, the nucleus is composed of chromosomes and other sub-nuclear structures that constrain how the genome is spatially organised[4].

During the last few decades, technical advances have dramatically improved our understanding on the organization and packaging of DNA. Evidence from for example electron microscopy[5–9], x-ray and neutron scattering[7,9–11], x-ray crystallography[12,13] and nuclease digestion have led to a number of different proposed models[2]. However there is some consensus about the general organization of chromatin. At the first level of organization, nucleosomes are formed by DNA wrapped around the histone octamer containing two histone H2A/H2B dimers flanking a central histone $(H3/H4)_2$ tetramer (figure 1)[2,13–15]. Then the chromatin is further compacted into chromosome domains, compartments and finally the chromosome[14,15]. Epigenetics and post-translational modifications (PTMs) have essential roles in maintaining genome stability[16] and the way the chromosome is packaged can additionally be altered by PTMs (e.g. acetylation or methylation). All histones of the nucleosome core particle are subject to post-translational acetylation[17] and histone H3 was shown to be prone to inducible phophorylation[18]. Changes in the genomic location of specific histone PTMs or alteration in levels of PTM impact genome stability[16].

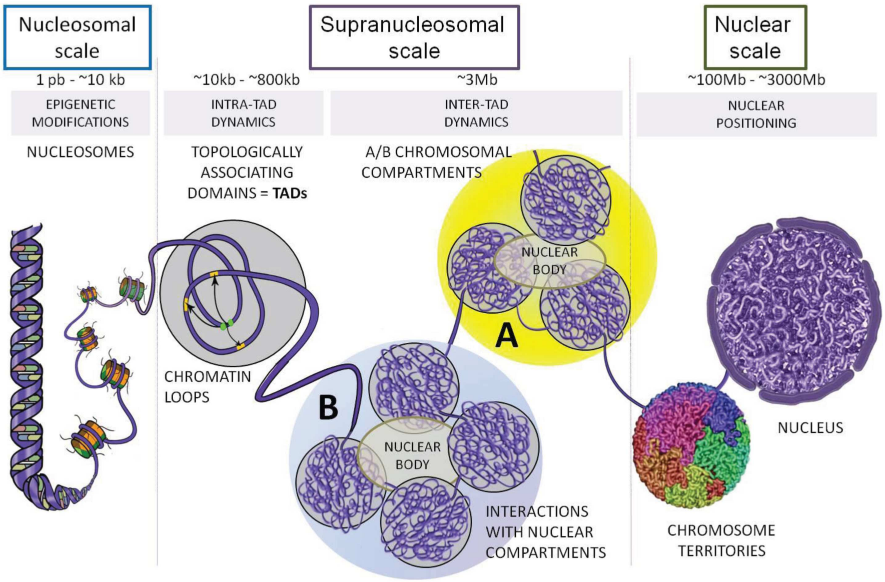

Nucleosomal scale
Supranucleosomal scale
Nuclear scale
1 pb - ~10 kb
~10kb - ~800kb
~3Mb
~100Mb - ~3000Mb
EPIGENETIC MODIFICATIONS
INTRA-TAD DYNAMICS
INTER-TAD DYNAMICS
NUCLEAR POSITIONING
NUCLEOSOMES
TOPOLOGICALLY ASSOCIATING DOMAINS = TADs
A/B CHROMOSOMAL COMPARTMENTS
NUCLEAR BODY
A
CHROMATIN LOOPS
B
NUCLEAR BODY
INTERACTIONS WITH NUCLEAR COMPARTMENTS
NUCLEUS
CHROMOSOME TERRITORIES

Figure 1: DNA is wrapped around the histone octamer to form nucleosomes which are connected by linker DNA. This structure is then packed into a 30 nm fiber. These fibres are further compacted into chromosome domains, compartments and finally the chromosome. Image from Ea et al.[14]

While individual chromosomes appear to be randomly localized with respect to each other[19–23], specific examples of non-random localization of groups of chromosomes[24], selected pairs of chromosomes[25], sub-chromosomal regions[26,27] as well as individual genomic loci[28–30] have been reported.

Clearly, the structure of nucleosome as well as possible modifications of the histones are important to maintaining a functional structure. The level of compaction of chromatin however, is also crucial to the transcriptional activity of a certain region.

5.2 Hetero- vs. Euchromatin

The spatial arrangement of the nucleosome and its modifications has been correlated to either active transcription or gene repression[31]. However, not only the structure of the nucleosome but also level of compaction of chromatin will influence how the DNA sequence is read[32]. Emil Heitz first used the term "heterochromatin" to describe densely stained chromatin with carmin acetic acid throughout the cell cycle in moss[33]. Euchromatin on the other hand, was the name for chromatin that underwent decondensation from metaphase to interphase.

Peripheral, compact heterochormatin is associated with a low density of genes, which are mostly inactive[32]. Frequent epigenetic modifications of heterochromatin are for example tri-methylation of the 9th lysine residue of the histone H3 (H3K9me3) and tri-methylation at the 27th lysine on H3 (H3K27me3), that are associated with gene silencing. These histone modifications give rise to two subtypes of heterochromatin: constitutive (H3K9me3) and facultative (H3K27me3)[34,35]. Facultative heterochromatin may form at various chromosomal regions, which usually contain genes that must be kept silent upon developmental cues[34]. This type of heterochromatin can be either condensed or decondensed depending on the cell cycle phase. It is also compatible with gene expression. Constitutive heterochromatin on the other hand is believed to occur in the same genomic regions in every cell type and these usually do not contain genes[34]. It is condensed throughout the cell cycle and is not transcribed. Most of constitutive chromatin forms at gene poor areas such as at pericentromeric regions and at telomeres. Both heterochromatin types are preferentially found at the nuclear periphery and around nucleoli[34–36].

Euchromatin is thought to predominantly occupy internal nuclear regions[36] and open euchromatin often correlates with high density of transcriptionally active genes[32]. Marks and features of active chromatin, such as DNA accessibility, are for example H3K27ac and H3K4me3[1].

Many transcriptionally inactive genes have however also been found in open chromatin fiber domains. In regions of low gene density, active genes can reside in large domains of compact fibers[32].

Though it has mainly been observed, that euchromatin occupies internal nuclear regions, while heterochromatin is mostly found in peripheral regions, exceptions to this chromatin arrangement were observed in different cell types. In mouse olfactory neurons for example, facultative heterochromatin co-localizes with constitutive heterochromatin in the central part of the nucleus, where it plays an important role in regulating exclusive expression of single odorant receptor genes[37].

Generally, heterochromatin seems to be located at the nuclear periphery and is associated with a low density of mostly inactive genes. Euchromatin on the other hand is thought to occupy mainly internal nuclear regions and frequently correlates with a high density of transcriptionally active genes. Other than a classification in euchromatin and heterochromatin, each chromosome can also be viewed as two sets of compartments.

5.3 Compartmentalization of the genome

Experimental results from Hi-C suggest that each chromosome can be decomposed into two sets of loci, active (A) and inactive (B) compartments[38]. In these compartments, contacts within each set are enriched and contacts between sets are depleted. It has also been shown that open and closed chromatin domains throughout the genome occupy different spatial compartments in the nucleus[38,39].
Compartment A is characterized by presence of genes, higher transcriptional activity and accessible chromatin (measured by DNaseI sensitivity). Moreover, compartment A showed enrichment for activating (H3K36 trimethylation) chromatin marks. Compartment B is more densely packed and genomic regions here are generally gene-poor and contain late-replicated domains that exhibit lower DNaseI accessibility. The mentioned characteristics suggest that the A compartment represents euchromatin and that the B compartment represents heterochromatin[38,39].
Additionally, the nuclear lamina, a thin fibrous layer that coats the inner nuclear membrane, is also involved in spatial organization. Chromatin domains, known as lamina-associated domains (LADs) are characterized by gene repression where the repositioning of the gene locus from the inner nucleus to the nuclear lamina can repress expression of certain genes[40,41]. These LADs have a high probablility to reside in contact with the nuclear lamina and are constrained in a peripheral zone of the nucleus[42–45]. While they contain the majority of inactive chromatin, the nuclear lamina is also thought to be an important reference point for radial genome organization in interphase nuclei[46]. The LADs are of interest as their anchoring to the nuclear lamina helps establish interphase chromosome topology, which in turn establishes overall spatial organization of the genome[46]. It has also been reported, that LAD and non-LAD domains were clearly spatially segregated[47]. Specialised sub-chromosomal regions, such as centromeres, telomeres and nucleolar organising regions (NORs) are also non-randomly positioned in the nucleus[26,27,48–50]. NORs contain ribosomal RNA gene clusters and

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

cluster to form the core of the nucleolus[51].

Thus it has been shown, that open and closed chromatin domains occupy distinct spatial compartments in the nucleus. Compartment A seems to have an abundance of genes, while compartment B is more densely packed and gene-poor. Repositioning of gene loci from the inner nucleus to the nuclear lamina, where chromatin domains known as LADs are located, also seem to play an important role in gene repression.
Other than being separated into A and B compartments, smaller levels of organization, namely the folding into topologically associating domains (TADs) have been observed.

5.4 TADs & sub-TADs

Thanks to the development of the Chromosme Confomation Capture (3C) assay[52] and its combination with the Next-Generation Sequencing (NGS) technology, other methods such as 4C, 5C and Hi-C could be developed[38,53–55]. These provided experimental access to the 3D organization of the genome in nuclear space, which revealed, that in interphase cells, chromatin folds into topologically associating domains (TADs) or contact domains[38,39,54]. TADs range in size from 100 kb to 5 Mb with an average size of 1 Mb and are defined as linear stretches of genome that have increased contact frequency when compared to the contact frequency with regions across their boundaries[39,54,56]. These show high stability between cell types. Single-cell Hi-C revealed, that in different cells contacts within TADs (intra-domain contacts) are more frequent than contacts that cross TAD borders (inter-domain contacts) and those that occur between two different chromosomes (trans-chromosomal contacts)[39]. Moreover, TADs containing active genes seem to be positioned on the surfaces of their chromosomal territories[57]. Although TADs are relatively conserved in a variety of cell types and organisms, the folding that occurs inside them seems to be more dynamic. Multiple factors, including the insulator binding factor CTCF, houskeeping genes and SINE elements have been found to be associated with the boundary regions separating topological domains and transcription start sites (TSS) were enriched there. Factors associated with active promoters and gene bodies are enriched at boundaries, while non-promoter-associated marks were not[39].
TADs are divided into smaller domains, sub-TADs, which are around 200-300 kb in size[58]. Target genes and regulatory elements physically interact with each other through contacts between sub-TADs.[59].

We see, that TADs interact with each other through intra or inter-domain contacts and TADs containing active genes seem to be positioned on the surface of their chromosomal territories (CT). Therefore it is interesting to look at the CTs in detail.

5.5 Chromosome territories

At the largest scale, human DNA is stored in the form of 23 pairs of chromosomes. During interphase, each chromosome occupies a discrete space (1-2 µm), a chromosome territory (CTs) inside the nucleus, where the positioning of genomic regions is non-random in respect to the nuclear periphery and correlated with transcriptional activity (see figure 2)[22,38,60,61]. The chromosomal arrangement into CTs has been shown both by microscopy as well as by genome-wide Hi-C analysis[38,61].

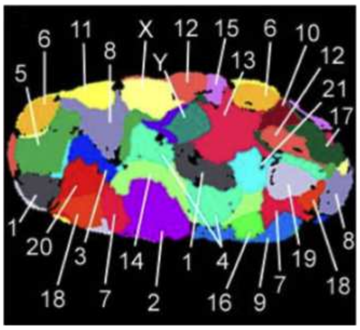


Figure 2: Representation of CTs according to Bolzer et al.[22]

One of the proposed models for radial positioning of chromosomes in interphase cells is order based on size. In this case, small CTs (e.g. chr14, 15, 19-22) are located in the center of the nucleus, while large chromosomes (e.g. chr1,2,3,4,13) are found at the periphery of the nucleus[62–64]. In quiescent human fibroblasts it has been shown that radial chromosome positions correlate with chromosome size. In these cells, territories of small chromosomes, independent of their gene density, were preferentially found close to the center of the nucleus, while the territories of large chromosomes were preferentially located towards the periphery of the nucleus[22]. However, Tanabe et al.[65] also found that in humans, territories of chromosomes 18 and 19 (HSA18 and 19) that are of similar size show a distinctly different position in the cell nuclei of lymphocytes and lymphoblastoid cells. While gene-rich and early replicating HSA19 CTs are typically found close to the nuclear center, gene-poor and later replicating HSA18 CTs are preferentially located at the nuclear periphery. This finding is in line with another proposed model, which states that the location of CTs is dependent on the gene density, where gene-rich chromosomes are located towards the nuclear interior and gene-poor chromosomes occupy peripheral locations[66]. The same has also been observed for embryonic stem cells (ESCs), where a gene density related radial arrangement of chr18 and 19 was found[67]. It has been observed, that in human lymphocytes, phytohemagglutinin-stimulated lymphoblasts, and lymphoblastoid cell lines, gene-rich CTs preferentially position towards the center of the nucleus and gene-poor CTs tend to be located towards the nuclear periphery[66,68–70]. Neighbouring chromosomes can also overlap and chromatin loops can intermingle with neighbouring CTs[61,71,72].

In respect to the radial positioning of chromosomes in interphase cells, both the size as well as the gene density seems to play a role. Smaller chromosomes and gene-rich chromosomes seem to be found at the center of the nucleus, while larger, gene-poor chromosomes seem to be located at the nuclear periphery. However, chromosomes do not always stay in the same spot in the nucleus but rather reposition throughout the cell cycle.

5.6 Dynamics of genome organization & chromosome repositioning throughout the cell cycle

Certain chromosomes have been observed in specific radial locations, such as chr18, which is consistently described as peripheral, and chr19, that seems to be centrally located[66,68,73]. There are various factors thought to contribute to the preferred radial location of chromosomes such as gene density[23,68,74], GC content[75–77], as well as chromosome size[19,22,63,78].

The position of many genes in the nucleus has been correlated with their activity state. This may indicate, that gene expression could be influenced by migration to functional subcompartments. It has been shown, that genes could be recruited to transcription factories and therefore undergo relocation[79], which suggests that genes can move and that those with transcriptional potential often reposition upon activation. Active genes, rather than recruiting and assembling transcription complexes, might migrate to pre-assembled transcription sites.

Mehta et al.[80] showed that in primary fibroblasts genome organization in interphase nuclei is altered when cells leave the proliferative cell cycle and that repositioning on chromosomes is achieved with an active nuclear motor complex that contains nuclear myosin 1β. It was observed, that most chromosomes do change their localization within interphase nuclei following damage. Some repositioned after DNA damage: it was observed that chromosomes 17,19 and 20 relocate from the nuclear interior to the periphery and chromosomes 12 and 15 reposition towards the interior in a majority of cells. Moreover, it was shown, that upon controlled induction of DNA double strand breaks (DSBs), CTs 12, 15, 17 and 19 relocated from their original position to another one. Upon washing off the agent, that induced the DSBs (cisplatin), the original CT arrangement was restored following DNA repair[81].

It has been shown by biochemical analyses and superresolution imaging, that less-compacted chromosomes tend to be more transcriptionally active[82–86]. Wang et al[86] used immunofluorescence analysis together with chromosome FISH for chrs 1-3 and 11 to show that changes in radial positions are accompanied by differential transcriptional activity. Furthermore, they also assessed the intermingling between specific chromosomes. It seems that chromosomes that are in the interior of the nucleus and are less compacted tend to intermingle more.

It was observed, that some chromosomal regions, such as the whole p-arm of chr12, are positioned more towards the center of the nucleus in ESCs but not in differentiated cells[67]. This part of the chromosome contains clustered pluripotency genes such as NANOG and Oct4, which suggests that chromosomal arrangement might be an important factor in controlling the expression of selected genes. Oct4 has been shown to interact with cohesin at enhancer and promotor sites to activate gene expression[87]. It was also reported, that ESCs, when compared to differentiated cells, present less centromeres at the periphery of the nucleus but more around nucleoli. Cell cycle-dependent redistribution of chromatin in the interphase nucleus has also been reported by Vourc'h et al.[88]. Moreover, they report that the profile of the rearrangement is dependent on the domain considered (telomeres, centromeres or chromosome-specific sub-satellite domains). In differentiated cells, centromeres were associated with the nuclear periphery, which has also been reported for most human cells[27,50,89,90]. Weierich et al.[50] report that in human lymphocytes, both centromeres and entire chromosome territories for chr1, 11, 12, 18 and X were found to be located in the periphery of the nucleus, while chr17 and 20 were found to occupy more internal regions. In active X chromosomes and autosomes the centromeres were positioned at the periphery of each CT, while in case of inactive X chromosomes centromeres were situated in the central parts of the CT. The telomeres of most of the mentioned chromosomes were however localized in the nuclear interior. Deposition of histone modifications, which are typically associated with open chromatin, onto subcentromeric regions of the chromosome correlates positively with the frequency of these centromeres to form stable clusters[27].

The cell cycle phase also seems to determine the position of chromosomes in the nucleus. For example, a peripheral localization of chr 18 is established early in the cell cycle and maintained thereafter[66]. In primary dermal fibroblast cells, it was reported that as normal fibroblasts exit the cell cycle, chr 13 and 18 relocate in serum starved medium[80]. It has been shown that CT positions are not inherited through mitosis but established in the G1 phase of the cell cycle[21,91] and are not altered thereafter. Moreover, it has been found, that in HeLa cell nuclei, major changes of CT neighbourhoods were observed during mitosis, while CT arrangements were stably maintained from mid G1 to late G2/early prophase.[21].

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

In summary, genes undergo relocation to facilitate transcription. Chromosomes in the interior of the ucleus are less compacted and tend to intermingle more and rearrangement of the nuclear organization is dependent on the cel cycle stage. So we see, that the chromosomes are not completely static, which is why it is also interesting to consider long-range contacts between genomic loci positioned up to megabases away, which are associated with facilitating transcription.

5.7 Long-range contacts

Proper gene regulation within the constrained nuclear space is important for function of the cell. To ensure this, chromosomes facilitate access to transcribed regions, while compactly packaging other information. Genomic loci positioned tens of kilobases up to megabases away, can undergo long-range interactions. These interactions between specific pairs of loci have in the past been evaluated with 3C, 4C, 5C, Hi-C and ChiP-seq[38,52,53,55,92,93]. These contacts can be found either on the same chromosome (cis) or between two chromosomes (trans). A very well-studied case of long-range chromatin contacts is for example the interaction between enhancers and their target genes in the β-globin locus[94,95].

It was shown that the level of interchromosomal contact differs for different pairs of chromosomes. For example, Lieberman et al. found that loci on chr 1 are most likely to interact with loci on chr 10 and least likely to interact with loci on chr 21[38]. Intrachromosomal interactions are more frequent than interchromosomal contacts. Moreover, small, gene-rich chromosomes tend to interact more with one another, which suggests, that these cluster together in the nucleus[38].

CCCTC-binding factor (CTCF) is a protein involved in the three-dimensional organization of chromatin[96]. CTCF is a transcription factor important for establishing the chromatin architecture, recruiting cell type-specific transcriptional activators, repressors, cohesins, and RNA polymerase II[93]. It is involved in creating boundaries that segregate chromatin and functional domains and in organizing higher-order chromatin structures by establishing chromosomal loops[97]. To form a long-range chromatin loop, two key protein players are needed: CTCF and the ring-shaped cohesin complex[98]. For formation of these loops, these exist together on cohesin/CTCF sites, but not separately. In absence of functional cohesin, global architectural changes associated with loss of cohesin/CTCF contacts and relaxation of topological domains can be observed[98,99]. It was observed, that loop domains disappeared after cohesin degradation, and addition of cohesin recovered their formation. On the other hand, compartment domains and histone marks are not affected by such changes[99]. A model, where CTCF recruits cohesin to form border-to-border loops between TAD boundaries has been proposed. It seems, that not only the presence but also the orientation of the CTCF motifs is of importance[96,100]. It was found, that loop formation preferentially occured between CTCF sites with convergent motifs (i.e. pointing to one another on the linear genome). Moreover, deletion of CTCF binding sites lead to disruption of chromatin loops[96].

All in all, intrachromosomal interactions are more frequent than interchromosomal contacts and small, gene-rich chromosomes tend to cluster together. It seems that the three-dimensional organization of the genome is important for enabling long-range contacts and that the protein CTCF as well as the cohesin are involved in this.
It is evident, that there is no clear formula for the radial position of genes and whole chromosomes. Moreover chromosomes relocate and intermingle but positions also change from a stem cell state to a differentiated state.

5.8 Changes in the architecture of the nucleus during neural differentiation

Nuclear architecture is important for the cell's identity and phenotype as this modulates gene expression. During the process of differentiation the genome is subjected to different kinds of changes. Epigenetic changes such as changes in transcription factors, changes in histone modifications as well as variation in TAD-TAD contacts can be observed.

As described in detail in section 5.2, the nuclear DNA and its associated proteins can be subdivided into heterochromatin and euchromatin. Heterochromatin contains condensed DNA and is found predominantly in the periphery of the nucleus, which makes it inaccessible to the transcriptional machinery and thereby transcriptionally inactive. Euchromatin is loosely packaged and permits active transcription[38]. It has been observed, that ES cells are richer in less compact euchromatin and as differentiation progresses, highly condensed, transcriptionally inactive heterochromatin is more abundant[101].

In terms of changes in transcription factors during neural differentiation, mainly modifications of Oct4 and Ascl1 (Mash1) have been observed.
Oct4 is a protein, that plays a role in induction and maintenance of pluripotency in embryonic stem cells. When investigating embryonic stem cells in both a pluripotent and differentiated neural progenitor state, the loss of Oct4 upon differentiation was observed[102,103]. Additionally Lee et al.[104] showed, that the promoter of Oct4 is also altered during differentiation: They found that the Oct4 promoter is enriched for the

active mark H3K4me3 in undifferentiated but not in differentiating ES cells[104].
Other epigenetic changes during differentiation were observed at the proneural regulator gene Ascl1. Ascl1 (Mash1) is a transcription factor essential for neural differentiation that promotes cell cycle exit and neuronal fate determination[105,106]. It was found that neural induction promotes large-scale chromatin re-organization here[107]. In ES cells, the *Mash1* gene is transcriptionally repressed and is preferentially positioned at the nuclear periphery with other late-replicating genes. At the same time, low levels of histone H3K9 acetylation at the Mash1 promotor and increased H3K27 methylation were observed. Upon neural induction, *Mash1* transcription was upregulated and spatial repositioning to the interior of the nucleus could be observed.

Changes in histone modifications are not only observed locally but also globally throughout the genome. It was shown, that a relatively high proportion of H1 molecules in ES cells are present in an unbound or loosely bound state[101]. Differentiation of mouse ES cells, is accompanied by an increase in H3K9me3[101], a decrease in H3K4me3[104,108,109] and acetylation of histones H3 and H4[101,104]. This means that stem cell neuronal differentiation requires action of chromatin modifying enzymes such as histone acetyl transferases (HATs) and histone deacetylases (HDACs)[110]. ES cell chromatin is globally acetylated and deacetylation is required for differentiation. In NPCs on the other hand, deacetylase activity is required to suppress neuronal genes. This means, that inhibition of HDAC is likely to drive NPCs towards neuronal commitment[110].
During neuronal differentiation, a change in histone variants can also be observed. Histone composition seems to change in core histone variants as well as in subtypes of the linker histone H1. H1a-d decrease during development and are partially or completely replaced by H1e[111]. In mouse cells, it was shown that H3.1 and H3.2, the H3 variants, show an exponential decrease during neuronal differentiation, while the levels of the H3.3, which marks actively transcribed chromatin, increase[112]. In the same study, it was also shown, that the synthesis pattern of immature neurons does not differ from that of mature neurons, indicating that the changes in the synthesis pattern of core histones occur at the arrest of cell proliferation.

TAD-TAD contacts were shown to re-wire during terminal neuronal differentiation in mouse cells[102]. Interestingly, in the same study, they also found that chr4, 6 and 19 have higher similarity in TAD-TAD contacts when comparing NPCs and neurons than when comparing ESCs and NPCs. In another study[113] it was found that long-range contacts between gene bodies of exon-rich, active gene types were formed during mouse neural differentiation. In this process, contacts between active TADs become less pronounced while inactive TADs interact more strongly. Moreover, gene expression is accompanied by the establishment of cell type-specific enhancer-promoter contacts.

In summary, the three-dimensional genome organization changes significantly during neural differentiation. Changes in the level of chromatin compaction, epigenetic changes, histone modifications as well as rewiring of TAD-TAD contacts have been observed. Overall the radial (re-)organization is complex and only little is understood to date. How exactly chromosomes are repositioned and especially how this happens in human neuroepithelial stem cells during differentiation is not known. To investigate this in detail, human stem cells are needed and induced stem cell (iPS) technology has paved the way for this.

5.9 iPS technology

Stem cell biology holds great potential for the study and treatment of neurodegenerative diseases[114]. Due to their potential to propagate indefinitely, as well as giving rise to a multitude of other cell types such as neurons, liver and bone cells, these represent a perfect source for replacing lost or damaged cells. The most well-known type of pluripotent stem cell is the embryonic stem cell (ESC). ESCs however hold the disadvantage of ethical difficulties regarding the use of human embryos or the problem of tissue rejection after transplantation.

Somatic cells can be reprogrammed to embryonic-like stem cells[115]. By expression of transcription factors such as Oct4, Sox2, klf4, c-myc, Nanog and lin28, induced pluripotent stem cells (iPSCs) can be generated from skin fibroblasts, that exhibit a pluripotent phenotype, similar to that of ESC[116]. The resulting cells, exhibit the morphology and growth properties of ESC and express ESC marker genes. Once the cells are pluripotent, the resulting cell colonies can be expanded and differentiated into specific neural populations[117]. Of course, this provides a big opportunity for patient specific ESCs, that can be used in neurological disease modelling, especially since these are patient-matched ESC lines. By deriving neuroepithelial like stem cells (NES) from pluripotent stem cells, disease modelling, drug discovery and the study of the nuclear architecture of differentiating cells is made possible.

The cells used in this thesis are AF-22 cells, a cell line from Falk et al.[116], which was derived from induced pluripotent stem cells (iPSCs). These cells are long-term self-renewing NES. The human iPSC lines were produced by standard retroviral transduction and later these were treated with insulin, transferrin, sodium-selenite and fibronectin to form NES. For the AF22 cell line, the reprogramming factors Oct4, Klf4 and Sox2 were used. To confirm, whether these cells have the potential to differentiate

TU Bibliothek WIEN Your knowledge hub
Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.
The approved original version of this thesis is available in print at TU Wien Bibliothek.

they used growth factor withdrawal and indeed, these NES differentiated into neurons. They also showed, that the cells differentiate asynchronously. Moreover, they observed that the resulting pluripotent stem cell lines exhibit long-term self-renewal capacity, similar growth kinetics, polar organization, expression of neuroepithelial "rosette markers" and a distinct regional specification with a transcription factor expression profile suggestive of a hindbrain identity.

iPS cells provide the opportunity to work with ES-like cells with the advantage of having patient specific cells as well as the elimination of drawbacks from using ESCs. To investigate these cells, microscopy and sequencing techniques can be used.

5.10 Microscopy: FISH

Molecular cytogenetics is often based on fluorescence *in situ* hybridization (FISH), a powerful microscopy technique for visualizing the composition and arrangement of genetic material. Since its development four decades ago, there has been a lot of research and improvement in terms of methods for probe generation and resolution of microscopy techniques[118–120].

FISH exploits the ability of a DNA strand to hybridize specifically to a complementary DNA strand. The basic elements of a FISH setup include a DNA probe and a target sequence. Fluorescently labelled DNA strands called probes that bind to the complementary sequences on the chromosome are used and detected with a fluorescence microscope[121,122]. First step of a FISH-assay is the preparation of the fluorescent probes. In most cases, the next step is the fixation of the cells using e.g. formaldehyde. This causes mainly protein-protein cross links. Next, DNA is denatured and the fluorescent probes hybridize to their complementary sequence. If for example a deletion is present the probe will not hybridize. If a duplication is present more of the probe will be able to hybridize. The last step is detection by fluorescence microscopy[123].

The probe has to be chosen depending on what is investigated as there are many different kinds of probes: dsDNA probes, ssDNA probes, RNA probes and synthetic oligonucleotides probes, which all carry different advantages and disadvantages. The investigated locus also influences the choice of probe and multiple probes are needed if a whole chromosome is explored.

FISH has a variety of applications. Examples include identification of the positions of genes, diagnosis of chromosomal abnormalities and analysis of interphase chromosomes[123].

5.11 Next Generation Sequencing

Next Generation Sequencing (NGS) platforms are on the market since 2004 and allow sequencing of whole genomes in a short time. This however also leads to huge amounts of data that have to be processed, which requires great bioinformatic effort.
The basic principle involves five steps: fragmentation of DNA, adaptor ligation, template immobilization by binding of fragments to a surface, amplification and detection of nucleotides. The templates are usually clonally amplified templates from single DNA molecules. The imaging is done via bioluminescence or one- or four-colour imaging of single molecular events. There are different detection principles such as sequencing by ligation, pyrosequencing, semiconductor sequencing and cyclic reversible dye termination (CRT), which is used by Illumina.

In this thesis the Illumina NextSeq 500 platform was used. With the use of Illumina platforms, the library preparation involves fragmentation, mate-pair and solid-phase amplification. It works via reversible terminator chemistry[124]. The template immobilization is composed of two steps: initial priming and extension of the single-stranded, single-molecule template, and bridge amplification of the immobilized template with immediately adjacent primers to form clusters. The read length in the Illumina NextSeq 500 System is 2 x 150 bp.
The Illumina platform uses the clonally amplified template method (figure 3) coupled with the four-coloured CRT method. Prior to the NGS reaction, single-molecule templates are immobilized on solid supports. This involves the formation of a clonal single-molecule array, where DNA fragments are denatured and single strands are annealed to complementary oligonucleotides on the flow-cell surface[125]. Then, a new strand is copied from the original strand in an extension reaction that is primed from the 3' end of the surface-bound oligonucleotide. After removal of the original strand by denaturation, the adaptor sequence at the 3' end of each copied strand is annealed to a new surface-bound complementary oligonucleotide. This is called bridge amplification, as the annealing process forms a bridge that generates a new site for synthesis of a second strand. This process is repeated multiple times, resulting in the growth of clusters at approx. 1 µm diameter (see figure 3).

Sequencing of DNA templates is done by repeated cycles of polymerase-directed single base extension. Detection is done by cyclic reversible termination. For this, a DNA polymerase bound to the primed template adds one fluorescently modified nucleotide. Unincorporated remaining nucleotides are consequently washed away and imaging is done to determine the identity of the incorporated nucleotide. Afterwards, the termi-

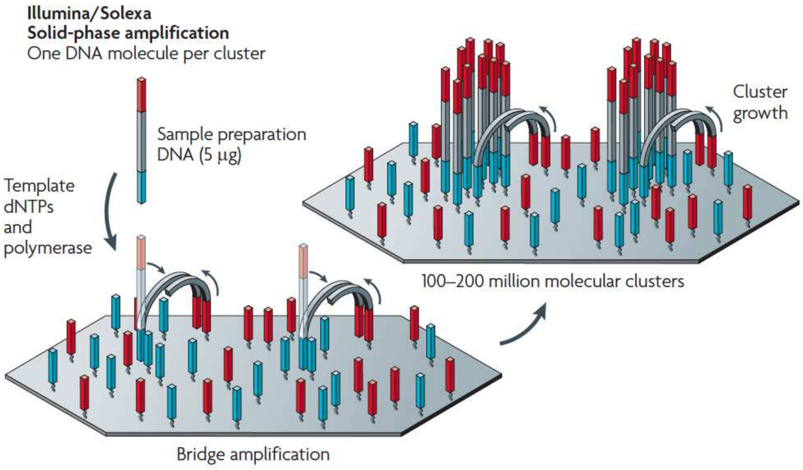

Illumina/Solexa
Solid-phase amplification
One DNA molecule per cluster
Sample preparation
DNA (5 μg)
Template
dNTPs
and
polymerase
Cluster
growth
100–200 million molecular clusters
Bridge amplification

Figure 3: Template immobilization in Illumina platforms where solid-phase amplification is used. The two steps involved in the process are initial priming and extending of the single-stranded, single-molecule template, and bridge amplification of the immobilized template with immediately adjacent primers to form clusters. Image taken from Metzker (2010)[124]

nating group as well as the fluorescent dye is cleaved and after a washing step, the next incorporation step is performed. A set of four reversible terminators, 3'-O-azidomethyl 2'-deoxynucleoside triphosphates (A,C,G and T) are each labelled with a different removable fluorophore. This ensures base-by-base nucleotide incorporation[124,125].
Detection of the four colours is done by total internal reflection fluorescence (TIRF) using two lasers. The most common error type in Illumina platforms are substitutions[124].

NGS therefore provides us with a powerful method to elucidate the nucleic acid sequence and to sequence the whole genome in a relatively short time.

5.12 Methodological considerations

The method, Genomic loci Positioning by Sequencing (GPSeq), which the experiments in this thesis are based on has recently been accepted in the journal Nature Biotechnology[1].
GPSeq, which is a genome-wide method, was developed to investigate the radial genome organization. It allows inferring distances to the nuclear lamina all along the nuclear radius and generates high resolution maps of genome radiality in human cells[1] (details see sections 6.3.1 & 6.3.5).

In this section, first the GPSeq method will be explained in detail and then theory behind the choice of enzymes as well as restriction times and how this affects experimental results will be recapitulated.

5.12.1 Detailed explanation of the theory behind GPSeq

The main idea of GPSeq is to tag genomic DNA (gDNA) in intact nuclei from the nuclear periphery towards the centre. Then the tags are used to assign individual genomic regions to specific radial locations in the nucleus using massively parallel sequencing. One GPSeq experiment consists of a series of conditions (different digestion times) in which the gDNA is restricted. For this a restriction enzyme (in this project HindIII or DpnII) is used to cut, with different probability, peripheral and central loci in fixed nuclei. To allow the restriction enzyme to penetrate the nucleus of cross-linked cells gradually, optimal experimental conditions have to be used[1].

To achieve this, a protocol that contains two main parts was developed. The first part consists of a YFISH experiment, which is an adaptation of a fluorescence *in situ* hybridization (FISH) procedure, and visualizes the digestion. The second part consists of tagging genomic DNA and preparation of samples for sequencing.

For the YFISH assay, the crosslinked DNA is digested with a restriction enzyme. A Y-shaped double-stranded adapter is ligated to each cut site. This is followed by hybridization of fluorescently labeled oligonucleotides to the Y-shaped adapter (see figure 4a). The samples are then imaged using wide-field epifluorescence microscopy. The signal should appear as a fluorescent rim that progressively broadens until finally the whole nucleus shows a uniform signal intensity (see figure 4b). Moreover, signal-intensity scaling should be seen, where the sample digested for the longest time should have the highest signal intensity and the sample digested for the shortest time should have the lowest signal intensity.

In this thesis NES, NPCs and neurons were incubated in the presence of the enzymes HindIII or DpnII for several different durations (e.g. 10 s, 20s, 30s, 45s, 2 min, 5 min, 10 min, 30 min with DpnII; 5 min, 10 min, 15 min, 30 min, 1h, 2h with HindIII).

Rather than just looking at the images from the microscope, a custom-pipeline to process the images is also used to confirm the signal intensity-scaling as well as the rim-progression (details in section 6.3.4). The YFISH method provides us with an effective way to confirm the digestion of the samples and functions as a checkpoint for proceeding with samples for sequencing.

The name YFISH comes from the fact that the adapter that is ligated to the cut sites has a Y-shape. The advantage of using a Y-shaped adapter lies in the fact, that it contains ssDNA regions ready for hybridization without the need to first denature DNA. This provides an easy access to the probes.

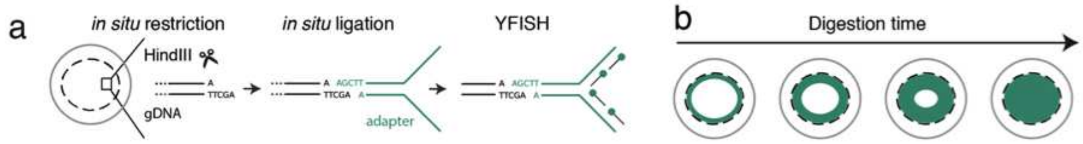


Figure 4: GPSeq procedure: (a) Scheme of YFISH. Cross-linked permeabilized nuclei (dashed circle) are incubated with a restriction enzyme (e.g. HindIII). The cut restriction sites are then ligated to a forked adapter (green) and detected by FISH using fluorescently labeled oligos (green dots) complementary to the fork. (b) Gradual in situ gDNA digestion. Fixed and permeabilized nuclei (dashed circles) are incubated with a restriction enzyme for increasing times. The action of the enzyme is revealed by YFISH and appears as a fluorescent ring (green) that progressively broadens inwards starting at the nuclear periphery. Each schematic cell corresponds to a separate sample.

When the microscopy yields satisfying results (i.e. a progressively broadening rim and signal intensity scaling), the samples can be processed further. Samples for YFISH and sequencing are processed in parallel for digestion as well as adapter ligation. For sequencing, different adapters than the ones used fro YFISH are needed. This "tagging" allows for the acquisition of reads from different radial positions in the nucleus. A scheme illustrating the preparation of the sequencing samples can be seen in figure 5. The protocol for GPSeq includes after ligation a DNA purification and sonication step, an *in vitro* transcription step and finally library preparation and sequencing. The UMIs are the unique molecular identifiers used to add short "tags" to DNA fragments to identify the input DNA molecule i.e. to know how many unique events there are. This is useful due to the fact that there are many amplified reads after PCR amongst which there are also duplicates. The barcodes that are added identify each separate sample and therefore help to distinguish between the different digestion time points. This allows for sequencing several libraries at the same time. RA5 is the Illumina RNA 5' adapter.

Finally the data is pre-processed as described in section 6.3.7. To estimate the radiality of specific loci, the GPSeq score is introduced. The GPSeq score is a probability estimate, and describes how the restriction probability within a given genomic window varies across consecutive digestion times[1] (for details see section 6.3.7).

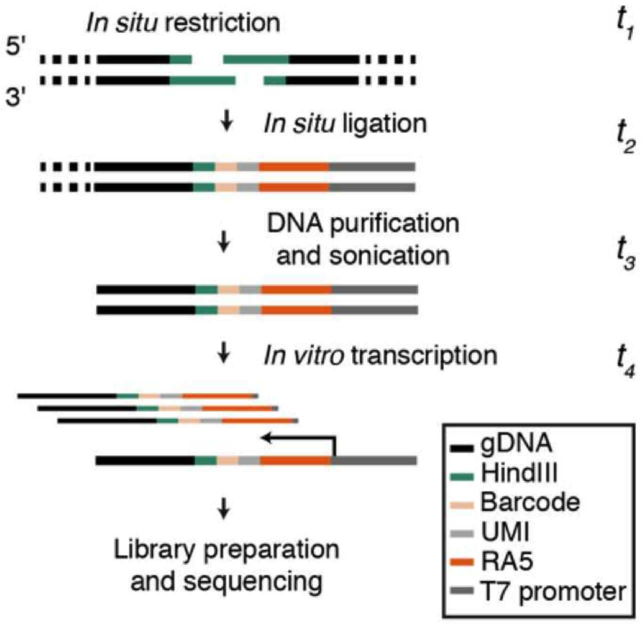

In situ restriction
5'
3'
t_1
In situ ligation
t_2
DNA purification and sonication
t_3
In vitro transcription
t_4
Library preparation and sequencing
gDNA
HindIII
Barcode
UMI
RA5
T7 promoter

Figure 5: Scheme showing the GPSeq library preparation. UMI = unique molecular identifier, RA5 = Illumina RNA 5' adapter

5.12.2 Choosing a restriction enzyme

The first important step for choosing optimal experimental conditions is choosing a restriction enzyme. As already mentioned, the restriction enzymes chosen for the experiments for this thesis were HindIII, a 6-base cutter, and DpnII, a 4-base cutter. These two enzymes have a different frequency and genomic distribution of recognition sites (RS). A 4-base cutter results in a larger number of genomic loci that can potentially yield fragments to be sequenced. A 4-base cutter therefore requires higher sequencing depth compared to using a 6-base cutter. Moreover, this implies that the density of RS will be higher in any genomic window, potentially allowing to achieve a higher resolution in radiality maps[1]. Advantage of using a 6-base cutter is that it allows for more shallow and cost-effective sequencing if a lower resolution is still acceptable. The most important quality of a restriction enzyme is that RSs are as homogeneously as possible along the genome. The number of RSs is accounted for when calculating the GPSeq score. The GPSeq score definition also corrects for linear effects that the number of RSs may have on the number of reads. For HindIII, the RS (AAGCTT) correlates with GC content, which would lead to higher number of reads in high GC-content regions[1]. For DpnII however, anti-correlation is observed, leading to a lower number of reads in high GC-content regions. Nevertheless, HindIII and DpnII experiments were highly correlated at various resolutions.

Both HindIII and DpnII produce 5' overhangs called sticky ends. HindIII has been used for GPSeq before. DpnII is an isoschizomer of MboI, which has also been used for GPSeq before. As it seemed during YFISH experiments, that DpnII produces even sharper rims than MboI, this enzyme was chosen for experiments in this thesis. Both HindIII and DpnII are not impaired by CpG methylation according to the manufacturer[126].

The digestion by HindIII occurs slower than with DpnII. The advantage of faster digestion is the time needed to carry out the experiment. However, since DpnII digests very fast, the timings have to be very precise to get reproducible results. This can especially be a problem if several coverslips are required for the same digestion time (for example when there are not enough cells on one coverslip). The resolution that can be achieved is higher with DpnII as this enzyme cuts more frequently, which is a strong advantage of using this enzyme.

5.12.3 Choosing the number of restriction times

The set of digestion times used in this thesis differs between the two enzymes used. As already mentioned, DpnII digests faster than HindIII and thus shorter digestion times were chosen. The exact times were chosen based on empirical observation of how the restriction enzyme diffuses throughout the nucleus as revealed by YFISH. With DpnII a fluorescent band was already observed after 10 seconds of incubation with the enzyme, while this was only visible after 10 minutes when using HindIII. The longest time point should show a homogeneous signal spread throughout the whole nuclear volume.

5.12.4 Amplification steps during library preparation

There are two amplification steps in the GPSeq protocol: *in vitro* transcription (IVT) and polymerase chain reaction (PCR). While IVT allows for linear amplification of the genomic fragments of interest, PCR is used to add Illumina sequencing adapters and indexes to the IVT-amplified fragments. To minimize the over-representation of certain fragments resulting from PCR, a low number of PCR cycles (typically 9 PCR cycles for 50 ng for DpnII and 300 ng for HindIII) is chosen during the GPSeq library preparation, if enough starting material is available. The complexity of the final sequencing libraries is maximized by splitting the PCR reactions into multiple tubes (e.g. 4 25 µl PCR reactions)[1].

With GPSeq we can infer distances to the nuclear lamina all along the nuclear radius. Studying the role that the spatial and temporal location of genes may elucidate many cell processes that happen in the nucleus. Thus, the question remains how the genome is spatially arranged along individual nucleic radii and how the radial organisation of the nucleus is changing during differentiation from NES to neurons.

6 Materials and Methods

6.1 Differentiation of NES to Neurons

The Neuroepithelial stem cells (NES) were derived from induced pluripotent stem cells (iPSCs) (see section 5.9). The NES were passaged and differentiated into progenitors as well as Neurons (see figure 6), according to an existing protocol (see appendix B). The cells used in this thesis are AF-22 cells, a cell line from Falk et al.[116]. These are cells from a female donor, where the exact patient is confidential according to patient-doctor confidentiality.

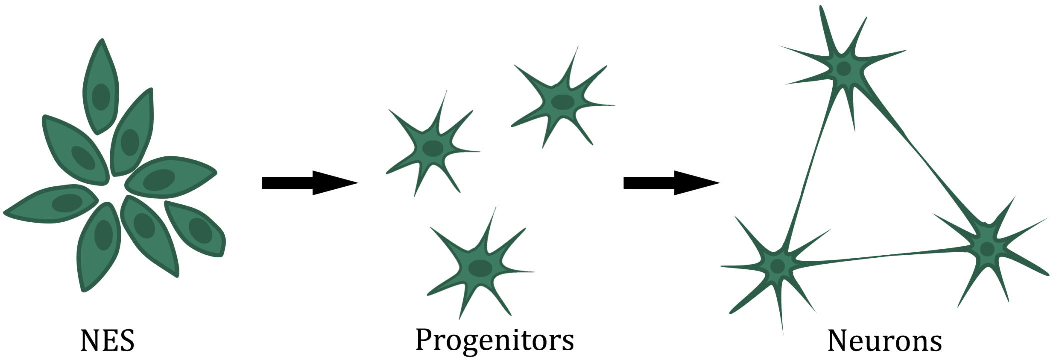


Figure 6: Illustration of the differentiation of Neuroepithelial stem cells (NES) to Neurons with Progenitors as the intermediate.

Experiments were done on 22 x 22 mm glass coverslips. For successful attachment of the cells to the coverslips and culture flasks, a coating with poly-L-ornithine (Sigma, cat no. P4957) for 60 minutes is needed. Subsequently a second coating with either 500x dilution of laminin (Sigma, cat no. L2020) or poly-L-lysine (Merck, cat no. P8920) was performed for 4 h at 37 °C or overnight at 4 °C. For the first experiments, poly-L-ornithine was used as the second coating. Later poly-L-lysine was used for the second coating. The poly-L-lysine step was introduced to make cell-attachment faster and passive, so that the cells could be fixed in PFA shortly after attaching. This ensures that the nuclear structure is preserved as much as possible.

Cells were cultured in NES medium, which contains DMEM/F12+Glutamax (Thermofisher, cat no. 31331-093), Penicillin/Streptamycin (Thermofisher, cat no. 15140-122), N_2 supplement, (Thermofisher, cat no. 17502-001), B27 supplement (Thermofisher, cat no. 17504-044), bFGF (10 ng/ml, Life Technologies, cat no. CTP0261), EGF recombinant human protein (10 ng/ml, Thermo Scientific, cat no. PHG0311). Passaging of cells was done using TrypLE Express Enzyme (Thermofisher, cat no. 12605-036). Washes were performed using DMEM/F12 supplemented with 20 % BSA

(Sigma, cat no. A9418).
For the coating during passaging and differentiation, poly-L-ornithine and laminin2020 (Merck, cat no. L2020) were used. For the differentiation of NES to NPCs and neurons, cells were plated in NES-medium without EGF and bFGF but with ten times higher concentration of B27 supplement (1:100). After 14 days, laminin2020 was added to the medium (1:1000).
Neural progenitor cells were harvested after 6 days of incubation at 37 °C with differentiation medium. Neurons were harvested after a total of 5 weeks of incubation with differentiation medium.

To monitor the progress of differentiation, pictures were taken with a brightfield microscope (see figure 7).

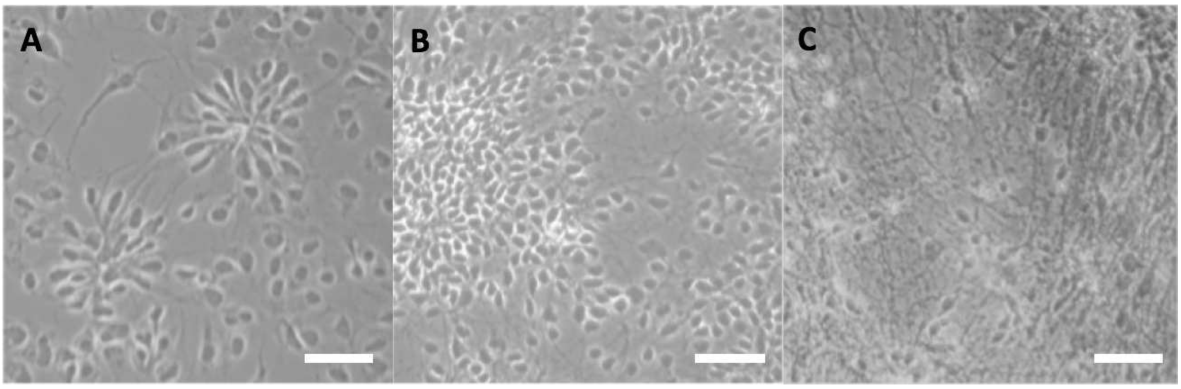


Figure 7: Brightfield microscope images of (A) NES, (B) NPCs and (C) Neurons; scale bar 100 µm

For fixation, cells were washed in 1xPBS containing Ca^{2+} and Mg^{2+} pre-warmed to 37 °C. Then, the coverslips were incubated for 10 min with 4 % paraformaldehyde (EMS, cat. no. 15710)/1X PBS (Thermo Fisher Scientific, cat. no. AM9625), where paraformaldehyde (PFA) acts as a cross-linking agent. This was followed by quenching of unreacted PFA in 1XPBS/125 mM glycine for 5 min at RT. After aspiration of the fixative solution, cells were washed three times, 5 min each with 0.05 % Triton X-100/1xPBS at RT and subsequently permeabilized in 0.5 % Triton X-100/1X PBS for 20 min at RT. Following overnight incubation in 1X PBS/0.5% Triton X-100 at RT, the cells were subjected to four cycles of freeze-and-thaw in liquid nitrogen (30 sec of immersion) for further permeabilization. Then, the cells were washed three times, 1 min each in 1X PBS/0.05 % Triton X-100 at RT. Afterwards, the cells were incubated in 0.1 N HCl for 5 min at RT and quickly rinsed twice in 1X PBS/0.05% Triton X-100 at RT. Finally, cells were rinsed in 2X SSC buffer (Thermo Fisher Scientific, cat. no.

AM9763).
In terms of the time between spotting of the cells and fixation, it was observed that a shorter time (i.e. 10-15 min) is ideal. This gives the cells enough time to attach but not enough time to migrate into rosettes, which is observed in NES cells.

6.2 Evaluation of the variation in Hoechst-signal

As already described in section 5.12, a GPSeq experiment consists of a YFISH experiment followed by preparation of the samples for sequencing. For YFISH, the digestion of the nuclei with the enzyme is monitored using wide-field epifluorescence microscopy. One experiment consists of a series of conditions corresponding to different digestion times. For these YFISH experiments, several coverslips are used which are all treated in exactly the same way, except for the length of digestion. To visualize nuclei in the microscope, all nuclei are stained using 0.1 ng/µl Hoechst. Due to all coverslips being treated in the same way, we would therefore expect the signal of the Hoechst-stain to be of the same median intensity in all imaged coverslips. This however is not the case: When imaging 10-15 fields of view, a variation in the median signal between the different samples in the channel for the Hoechst-stain was observed. This variation seems completely random and not dependent on the digestion time of the specific sample, as no trend was visible. Since images are processed with a custom pipeline after each YFISH experiment, this variation can be problematic. The reason for this is that the script to process the images (as described in section 6.3.4) relies on the signal of the Hoechst-stain for automatic segmentation of the nuclei. Thus, the variation in signal intensity in the channel where the Hoechst-stain is imaged, should be kept to an absolute minimum.

For the above mentioned reasons, an experiment to test at which number of fields of view the variation in signal intensity of the Hoechst-stain would not be observed anymore, was designed. For this, three samples of fixed NES cells were stained with 0.5 ng/ml Hoechst for 30 min. For these samples, the aim was to image 100 fields of view each and then repeat the imaging process to also see if there was variation between the same sample upon re-imaging. The original plan was to use random multipoint scanning to have 100 unbiased fields of view per sample. However, since the cells were not completely evenly spread, this experimental setup did not yield satisfying results (i.e. many empty fields of view). Therefore 100 fields of view per sample (with one repeat each, i.e. in total 600 fields of view) were picked by hand, where the 100 fields of view were split into 5 portions of 20 fields of view each. The set-up of the experiment can be seen in table 1. The settings of the imaging run can be seen in table 2. These settings are camera specific. The camera used was the Andor iXon Ultra 888 EMCCD camera.

Table 1: Set up of Experiment to evaluate variation of signal in dapi channel

Run number	Sample
1	First repeat, Sample 1, Fields of view 1-20
2	First repeat, Sample 1, Fields of view 21-40
3	First repeat, Sample 1, Fields of view 41-60
4	First repeat, Sample 1, Fields of view 61-80
5	First repeat, Sample 1, Fields of view 81-100
6	First repeat, Sample 2, Fields of view 1-20
7	First repeat, Sample 2, Fields of view 21-40
8	First repeat, Sample 2, Fields of view 41-60
9	First repeat, Sample 2, Fields of view 61-80
10	First repeat, Sample 2, Fields of view 81-100
11	First repeat, Sample 3, Fields of view 1-20
12	First repeat, Sample 3, Fields of view 21-40
13	First repeat, Sample 3, Fields of view 41-60
14	First repeat, Sample 3, Fields of view 61-80
15	First repeat, Sample 3, Fields of view 81-100
16	Second repeat, Sample 1, Fields of view 101-120
17	Second repeat, Sample 1, Fields of view 121-140
18	Second repeat, Sample 1, Fields of view 141-160
19	Second repeat, Sample 1, Fields of view 161-180
20	Second repeat, Sample 1, Fields of view 181-200
21	Second repeat, Sample 2, Fields of view 101-120
22	Second repeat, Sample 2, Fields of view 121-140
23	Second repeat, Sample 2, Fields of view 141-160
24	Second repeat, Sample 2, Fields of view 161-180
25	Second repeat, Sample 2, Fields of view 181-200
26	Second repeat, Sample 3, Fields of view 101-120
27	Second repeat, Sample 3, Fields of view 121-140
28	Second repeat, Sample 3, Fields of view 141-160
29	Second repeat, Sample 3, Fields of view 161-180
30	Second repeat, Sample 3, Fields of view 181-200

The images were converted to an uncompressed TIFF format and deconvolved using Huygens Professional v17.04 Software (Scientific Volume Imaging) and afterwards, the images were processed using the gpseq_anim script that is also used for the usual GPSeq-Experiments (see section 6.3.1). A comparison of signal with 20, 40, 60, 80 and 100 fields of view was evaluated (for results see section 7.1).

Table 2: Microscopy settings used to evaluate variation of signal in dapi channel. The camera used was the Andor iXon Ultra 888 EMCCD camera

Power of the 390 nm LED	4 %
Readout mode	EM Gain 10 MHz 16 bit
EM Gain Multiplier	4
Exposure	100 ms
Conversion Gain	Gain 1

6.3 Genomic loci Positioning by Sequencing

Genomic loci Positioning by Sequencing (GPSeq) is a method that allows genome-wide measurements of the distance of genomic loci to the nuclear lamina. It is a technique that uses a method called YFISH to generate growing rings of digestion to identify optimal experimental parameters before sequencing the samples. YFISH and subsequent imaging is used as a validation step to ensure that restriction was successful and that growing rings of digestion were generated. Through sequencing, the restricted region's sequence is obtained. By calculating the so-called GPSeq score, we get a radiality estimate. This way, reproducible maps of radial genome organization in human cells, at various resolutions, can be obtained (detailed description of theory behind method in section 5.12).

6.3.1 YFISH

The experiments were done according to the method by Girelli et al.[1]. A detailed step-by-step YFISH protocol is available in appendix A. Briefly, in situ restriction was performed using either 10 µl of HindIII-HF (NEB, cat. no. R3104S) or 4 µl of DpnII (NEB, cat. no. R0543M) in 400 µl at 37 °C for different durations, ranging from 5 min up to 2 h in the case of HindIII-HF, and 10 s up to 30 min in the case of DpnII. For long time points (i.e. 5 min - 2h) samples were incubated in a large incubator, while short timepoints (i.e. 10 sec - 2 min) were incubated on a thermoblock and the exact times measured with a stop watch. The reaction was stopped by placing the samples in ice-cold 1X PBS/50 mM EDTA/0.01% Triton X-100, and washing them multiple times on ice. Afterwards, the samples were dephosphorylated by incubating them in 400 µl of 1X calf intestinal alkaline phosphatase (CIAP) buffer containing 6 µl of CIAP (Promega, cat. no. M1821) for 2 h at 37 °C. Next, GPSeq adapters were ligated at a final concentration of 0.2 µM in 300 µl of 1X T4 DNA ligase buffer containing 36 µl of T4 DNA ligase (Thermo Fisher Scientific, cat. no. EL0014), by incubating the

TU Bibliothek Your knowledge hub WIEN
Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.
The approved original version of this thesis is available in print at TU Wien Bibliothek.

samples at 16 °C for 18 h. The next day, unligated adapters were washed away by incubating the samples in 10 mM Tris-HCl/1M NaCl /0.5% Triton X-100 pH 8, at 37 °C five times, each for 1 h, while shaking. To prepare the hybridization mix, the labeled oligonucleotides were diluted to 200 nM in a hybridization buffer containing 2X SSC/25% formamide/10% dextran sulfate/1 mg/ml E. coli tRNA/0.02% bovine serum albumin (BSA). Coverslips were placed onto a piece of parafilm, with cells facing a 300µl droplet of hybridization mix, and the samples incubated in a humidity chamber for 18 h at 30 °C. The following day, the samples were washed twice in washing buffer containing 2X SSC/25% formamide for 30 min at 30 °C. Finally, the samples were incubated in 2X SSC/25% formamide/0.1 ng/µl Hoechst 33342 (Thermo Fisher Scientific, cat. no. H3570) for 30 min at 30 °C. All samples were imaged using wide-field epifluorescence microscopy (see section 6.3.3).

The quantification of the YFISH signal was done by examining various features of the distribution of signal intensity calculated along the nuclear radius: i) the peak, which corresponds to the region with the highest intensity in the expanding fluorescent band; ii) the inflection point, which corresponds to the front edge of the fluorescent band; and iii) the ratio (contrast) between the intensity of peak and inflection point.

6.3.2 Preparation of YFISH and GPSeq adapters

The list of oligonucleotides (oligos) used to prepare GPSeq and YFISH adapters is available in appendix C. Individual oligos were purchased from Integrated DNA Technologies as desalted oligos at 100 mM concentration in nuclease-free water. 10 µl of each forward oligo was diluted in 90 µl final volume of a mix containing 10 µl of 10X Polynucleotide Kinase (PNK) buffer and 2 µl of T4 PNK (NEB, cat. no. M0201). Samples were incubated for 1 h at 37 °C, after which 10 µl of the corresponding antisense oligonucleotides was added. This was then incubated for 5 min at 95 °C, followed by gradual cooling down to 25 °C over a period of 45 min (-1.55 °C/min) in a PCR thermocycler.

6.3.3 Wide-field epifluorescence microscopy

The microscopy system used to acquire the images includes the Ultra-888 Andor camera. The key specifications can be found in table 3.

Wide-field microscopy was used to image all the YFISH samples. Briefly, samples were rinsed twice in 2X SSC, equilibrated in RNA equilibration buffer (10 mM TrisHCl ph 7.5/2x SSC/0.4% glucose), and mounted in Anti-bleach buffer (200 mM trolox, 3.7 mg/ml glucose oxidas/catalase/RNA Equilibration buffer) before imaging.

Table 3: Key Specifications Ultra-888 Andor Camera

Active Pixels	1024 x 1024
Pixel size	13 x 13 µm

The samples were imaged using a 100X 1.45 NA objective mounted on a custom-built Eclipse Ti-E inverted microscope system (Nikon) controlled by the NIS Elements software (Nikon) and equipped with an iXON Ultra 888 ECCD camera (Andor Technology). For YFISH, multiple image stacks per sample were aquired, each consisting of 71 focal planes spaced 0.3 µm apart. TetraSpeck Microspheres (0.1 µm, fluorescent blue/green/orange and dark red, Thermo Fisher Scientific, cat. no. T7279) were imaged after each imaging session. These were used to correct for chromatic abberations and shifts between channels using DOTTER (a MATLAB® script).

6.3.4 Image processing

Raw images were converted from ND2 format to an uncompressed TIFF format using the nd2_to_tiff script, from the *pygpseq* Python3 package available on GitHub: https://github.com/ggirelli/pygpseq/. All channels were deconvolved using the Huygens Professional v17.04 Software (Scientific Volume Imaging), with the following parameters: CMLE alogrithm, null background, and signal-to-noise ratio (SNR) of 7, 50 iterations. After this, a report of the experiment was created using another custom-built pipeline available on GitHub: https://github.com/ggirelli/pygpseq with the command *gpseq_anim -d dapi -s cy5 -m ’destination of masks’ –compress -n -r – ’input’ ’output’*. Here the nuclei are automatically segmented in 3D using the signal coming from Hoechst 33342 staining. For each identified cell, the volume, shape, surface, flattened size (in Z-projection), sum of intensity and average intensity are estimated. Then, only G1 nuclei are selected based on the DNA staining. GPSeq profiles are then built by calculating the mean, median and mode of single-voxel intensity values (signal coming from cut-sites) falling in each bin of a binned normalized distance from the nuclear lamina to the center of the nucleus.

6.3.5 Sample preparation for Sequencing

The experiments were done according to the method by Girelli et al.[1]. A detailed step-by-step GPSeq protocol is available in the A. Briefly, DNA was digested and ligated the GPSeq adapters using the same procedure described above for YFISH. After overnight

ligation at 16 °C, unligated adapters were washed away by incubating the samples five times, each for 1 h, in 10 mM Tris-HCl/1 M NaCl/0.5% Triton X-100 pH 8, at 37 °C, while shaking. The cells were then scraped off the coverslips and digested in 110 µl of 10 mM Tris-HCl/100 mM NaCl/50 mM EDTA/1% SDS pH 8, at 25 °C containing 10 µl of Proteinase K (NEB, cat. no. P8107S), for 18 h at 56 °C. The next day, the enzyme was inactivated by increasing the temperature to 96 °C for 10 min. Genomic DNA (gDNA) was purified using phenol-chloroform extraction, and gDNA was precipitated using glycogen (Sigma, cat. no. 10901393001) and sodium acetate, pH 5.5 (Life Technologies, cat. no. AM9740) in ice-cold ethanol (VWR, cat. no. 20816.367) at -80 °C for 18 h. The DNA pellets were resuspended in 100 µl of TE buffer, and sonicated them in a Bioruptor® Plus machine with the following settings: 30 sec ON, 90 sec OFF, high mode, 16 cycles. Afterwards, gDNA was concentrated down to a final volume of 8 µl in nuclease-free water, using AMPure XP (Beckman Coulter, cat. no. A63881). In vitro transcription was performed on each sample separately, with the MEGAscript® T7 Transcription kit (Thermo Fisher Scientific, cat. no. AM1334-5), using the same amount of gDNA (between 50 and 300ng, see Table for each sample in a final volume of 20 µl, and the samples were incubated for 14 h at 37 °C. After IVT, 1 µl of DNAse I (Thermo Fisher Scientific, cat. no. AM2222) was added to the sample, and incubated for 15 min at 37 °C. Then the RNA was purified with Agencourt RNAClean XP beads (Beckman Coulter, cat. no. A63987).
To confirm, whether purification was successful and to measure the concentrations of DNA or RNA in the samples, the samples were analyzed using the Bioanalyzer 2100 (Agilent, cat. no. G2943CA) and measurements of DNA and RNA concentrations were performed using Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, cat. no. Q32866) in combination with the Qubit® dsDNA HS Assay (Thermo Fisher Scientific, cat. no. Q32851).

Lastly, sequencing libraries were prepared using the TruSeq Small RNA Library Preparation kit (Illumina, cat. no. RS-200-0012), following the manufacturer's instructions with some modifications, as described in the step-by-step protocol. All libraries were sequenced on the NextSeq 500 system (Illumina) using the NextSeq 500/550 High Output v2 kit (75 cycles) (Illumina, cat. no. 20024906).

6.3.6 Variation of experimental parameters

As already described in section 6.1, cells were seeded on glass coverslips. The protocol for seeding on glass coverslips had been optimized for 18x18 mm coverslips previously. However due to the amount of cells needed for sequencing, the size of coverslips was changed to 22x22 mm. This meant that more cells were on one coverslip and that the amount of coverslips per condition could be kept to a minimum.

Following this change, the seeding density had to be adjusted. At first, 450 000 cells per coverslip were seeded. As this was a very large amount of cells, and they were also not completely homogeneously spread, this lead to them sometimes overlapping or clumping, which then meant that the enzyme could not properly penetrate all cells equally. A seeding density of 300 000 cells seemed to work well. The application of the cell suspension to the coverslips was also varied between spotting them and covering the whole well of a 6-well-plate with cell suspension. The cells could spread more homogeneously when applying the cell suspension to the whole well, but this meant that very large amounts of cells were needed. For neurons this was not an option, as it was difficult to maintain a neuron culture for a large amount of cells as a lot of flasks were needed for this. Neurons have to differentiate for 5 weeks in large flasks and a part does not survive until the end. Therefore all neurons samples were always spotted. For the other samples (NES and NPC) cells were then also spotted to keep conditions the same and to reduce the amount of cells needed.

As described earlier, two coatings were required for the coverslips. For the first experiments, poly-L-ornithine was used as the second coating. Later the coating was changed to poly-L-lysine, as faster and passive cell-attachment was achieved in this way. This enabled fixation in PFA shortly after attachment and the nuclear structure was preserved as much as possible.

The amount of coverslips used for sequencing for each condition was dependent on the amount of cells present on a coverslip. Usually one coverslip per condition was sufficient. In one experiment, where neurons were used, the amount of cells per coverslip was very low. Therefore two coverslips were used for each condition here, to assure that the amount of DNA was sufficient for further processing.

Preparation of the samples (see section 6.1) before digestion involved incubation in FPS (50% FA/50 mM phosphate buffer/2X SSC) as well as a liquid nitrogen procedure. The time of incubation in FPS was varied between 40 h and 20 h. As this did not seem to yield different results, an incubation time of 20 h was chosen to shorten the time needed to go through the protocol. The liquid nitrogen procedure was also omitted once, however the results achieved with the liquid nitrogen procedure seemed more promising, so this step was kept. This however also resulted in one extra day required for incubation in 20 % glycerol/1xPBS.

During fixation the PBS concentration was varied between 1x and 0.4x. A lower concentration during the fixation step yields rounder cells which is beneficial for the image analysis as the model assumes that nuclei are a perfect sphere. For Neurons however, the fixation had to be done in 1x PBS as these cells are more fragile. Therefore a fixation in 1x PBS was chosen for all experiments.

In the beginning (for the first 11 experiments) HindIII-HF was used for digestion. During one experiment, the concentration of HindIII-HF was changed from 0.5 U/µl to 0.75 Uµl at long time points (i.e. 1h and 2h). This was done to account for the increase in restriction sites when digesting for a longer time. This however did not seem to change the outcome of the experiment, so a concentration of 0.5 U/µl was kept for the remaining experiments.

As discussed in section 5.12, HindIII cuts less frequent than DpnII and therefore generally achieves lower resolution during sequencing. Therefore DpnII was used later as the digestion enzyme of choice as it seemed to produce a sharper rim as visible in the YFISH experiment.

6.3.7 Processing of GPSeq sequencing data GPSeq score calculation

This part was done by Gabriele Girelli, who is working as a Bioinformatician in the group of Magda Bienko and Nicola Crosetto, which is the same group where all experiments mentioned in this thesis were performed. For clarity the data processing and GPSeq score calculation is described here briefly.

Raw sequencing data was demultiplexed based on the RA5 indexes, either using the BaseSpace Sequence Hub cloud service of Illumina, or manually with bcl2fastq (v2.18). Sequencing data was pre-processed using a custom pipeline (gpseq-seq-gg) featuring: quality control, read filtering based on the expected adapter sequence, adapter trimming, mapping, filtering of the mapping output, filtering of reads mapped away from restriction sites, and UMI-based read de-duplication. Restriction sites (AAGCTT with HindIII; GATC with DpnII) associated with an abnormally high number of deduplicated UMIs for a given digestion time (i.e., condition) were discarded, by identifying outliers with a chi-square method and a significance of 0.01 (using the outliers R package).

The first filtering step selects reads containing the full prefix. Then, the reads were trimmed to remove the prefix (including the restriction site), and the remaining part aligned against the human reference genome (Grch37/hg19 GCA_000001405.1) using *bwa-mem*. Primary alignments with a mapping quality equal or higher than 30 were retained, while unmapped reads, chimeric reads and reads mapped to chrY that is not present in the cells used, were discarded. Only the reads whose 5' end is less than 20 bp away from the position of a HindIII or DpnII (depending on which enzyme was used) recognition site (RS) in the reference human genome were retained. The reason why not strictly all the reads are required to align exactly to the position of RS is that the T7 polymerase and reverse transcriptase used during the GPSeq library preparation are prone to occasionally skip some bases, leading to the resulting reads aligining slightly downstream of the RS. Afterwards, the UMI sequence of each aligned read was recovered, and the reads were filtered based on quality of the UMI sequence. Afterwards, UMI sequences mapped to the same recognition site were deduplicated and a BED file was generated containing the genomic coordinate and number of de-duplicated UMIs associated to each restriction site.

The genome was binned using either 1 Mb overlapping windows sliding in steps of 100 kb (1 Mb resolution) or non-overlapping 100 kb windows (100 kb resolution). For each condition, all the restriction sites that had been cut were considered, to calculate a digestion probability, based on which the GPSeq score was calculated. A BED-like file containing the GPSeq score per window was generated and masked based on a manually curated mask of repetitive and low complexity regions. To be able to compare different experiments, the calculated GPSeq score was rescaled. To average the GPSeq score across different experiments, first the score of each window across the experiments was averaged, and then the log2 of these averages was calculated and again rescaled.

6.3.8 Summary of experimental data

Tables 4 and 5 contain a summary of the seven experiments which yielded material that was of a quality suitable for sequencing. Table 4 shows a summary of the experimental data (cell type, enzyme used, condition, IVT input, number of PCR cycles). Table 5 shows the demultiplex index, the Illumina primer, as well as the prefix used in the experiment. The Illumina primer sequences can be found in appendix D. A total of 19 experiments were carried out where experimental conditions varied according to section 6.3.6.

All samples were generated using asynchronized cell populations. The loaded amount was the same for all libraries: 6 nmol. Sequencing was always done on the Illumina NextSeq 500 platform.
In table 5 in the prefix column the following numbers can be seen: [1,0,0]. The first position (here always a 1) represents the mismatches. 1 means that 1 mismatch is allowed when pre-processing the data, as there is a Hamming distance of 2 between two barcodes. The second number (here 0) represents insertions and the last number

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

Table 4: Summary of experimental data part 1 showing: cell type, enzyme used, condition, IVT input and number of PCR cycles

Library ID	Cell type	Enzyme used	Condition	IVT input (ng)	# PCR cycles
MS1	NES	HindIII	5 min	166	10
MS2	NES	HindIII	10 min	166	10
MS3	NES	HindIII	15 min	166	10
MS4	NES	HindIII	30 min	166	10
MS5	NES	HindIII	1 h	166	10
MS6	NES	HindIII	2 h	166	10
MS7	Neurons	DpnII	10 sec	31	10
MS8	Neurons	DpnII	20 sec	31	10
MS9	Neurons	DpnII	30 sec	31	10
MS10	Neurons	DpnII	45 sec	31	10
MS11	Neurons	DpnII	2 min	31	10
MS12	Neurons	DpnII	5 min	31	10
MS13	NES	DpnII	10 sec	253.8	9
MS14	NES	DpnII	30 sec	253.8	9
MS15	NES	DpnII	2 min	253.8	9
MS16	NES	DpnII	5 min	253.8	9
MS17	NES	DpnII	30 min	253.8	9
MS18	NES	DpnII	10 sec	250	9
MS19	NES	DpnII	30 sec	250	9
MS20	NES	DpnII	2 min	250	9
MS21	NES	DpnII	5 min	250	9
MS23	NES	DpnII	20 min	250	9
MS25	NPC	DpnII	10 sec	175	9
MS26	NPC	DpnII	30 sec	175	9
MS27	NPC	DpnII	2 min	175	9
MS28	NPC	DpnII	5 min	175	9
MS29	NPC	DpnII	10 min	175	9
MS30	NPC	DpnII	30 min	175	9
MS32	NPC	DpnII	10 sec	175	9
MS33	NPC	DpnII	30 sec	175	9
MS34	NPC	DpnII	2 min	175	9
MS35	NPC	DpnII	5 min	175	9
MS36	NPC	DpnII	10 min	175	9
MS37	NPC	DpnII	30 min	175	9
MS38	Neurons	DpnII	10 sec	26.8	10
MS39	Neurons	DpnII	20 sec	26.8	10
MS40	Neurons	DpnII	30 sec	26.8	10
MS41	Neurons	DpnII	45 sec	26.8	10
MS42	Neurons	DpnII	2 min	26.8	10
MS43	Neurons	DpnII	5 min	26.8	10

Table 5: Summary of experimental data part 2 showing: demultiplex index, prefix and Illumina primer used

Library ID	Demultiplex Index	Prefix	PCR Primer
MS1	CCGTCC	8...8 CGTGTGAG AAGCTT 1...1000	RPI16
MS2	GTAGAG	8...8 GACACGAG AAGCTT 1...1000	RPI17
MS3	GTCCGC	8...8 AGCCATCA AAGCTT 1...1000	RPI18
MS4	GTGAAA	8...8 TCCGACGA AAGCTT 1...1000	RPI19
MS5	GTGGCC	8...8 ATGTGCGC AAGCTT 1...1000	RPI20
MS6	GTTTCG	8...8 ATGTGGAG AAGCTT 1...1000	RPI21
MS7	CCGTCC	8...8 CATCACGC [1,0,0]] GATC 1...1000	RPI16
MS8	GTAGAG	8...8 GTCGTCGC [1,0,0] GATC 1...1000	RPI17
MS9	GTCCGC	8...8 ACGACCGC [1,0,0] GATC 1...1000	RPI18
MS10	GTGAAA	8...8 TGATGCGC [1,0,0] GATC 1...1000	RPI19
MS11	GTGGCC	8...8 CATCAATC [1,0,0] GATC 1...1000	RPI20
MS12	GTTTCG	8...8 GTCGTATC [1,0,0] GATC 1...1000	RPI21
MS13	GTAGAG	8...8 CATCACGC [1,0,0] GATC 1...1000	RPI17
MS14	GTCCGC	8...8 GTCGTCGC [1,0,0] GATC 1...1000	RPI18
MS15	GTGGCC	8...8 ACGACCGC [1,0,0] GATC 1...1000	RPI20
MS16	GTTTCG	8...8 TGATGCGC [1,0,0] GATC 1...1000	RPI21
MS17	CGTACG	8...8 CATCAATC [1,0,0] GATC 1...1000	RPI22
MS18	CCGTCC	8...8 CATCACGC [1,0,0] GATC 1...1000	RPI16
MS19	GTAGAG	8...8 GTCGTCGC [1,0,0] GATC 1...1000	RPI17
MS20	GTCCGC	8...8 ACGACCGC [1,0,0] GATC 1...1000	RPI18
MS21	GTGGCC	8...8 TGATGCGC [1,0,0] GATC 1...1000	RPI20
MS23	CGTACG	8...8 CATCAATC [1,0,0] GATC 1...1000	RPI22
MS25	CCGTCC	8...8 CATCACGC [1,0,0] GATC [1,0,0] 1...1000	RPI16
MS26	GTAGAG	8...8 GTCGTCGC[1,0,0] GATC[1,0,0] 1...1000	RPI17
MS27	GTCCGC	8...8 ACGACCGC[1,0,0] GATC[1,0,0] 1...1000	RPI18
MS28	GTGGCC	8...8 TGATGCGC[1,0,0] GATC[1,0,0] 1...1000	RPI20
MS29	GTTTCG	8...8 GTCGTATC[1,0,0] GATC[1,0,0] 1...1000	RPI21
MS30	CGTACG	8...8 CATCAATC[1,0,0] GATC[1,0,0] 1...1000	RPI22
MS32	GTAGAG	8...8 CATCACGC [1,0,0] GATC [1,0,0] 1...1000	RPI17
MS33	GTCCGC	8...8 GTCGTCGC[1,0,0] GATC[1,0,0] 1...1000	RPI18
MS34	GTGGCC	8...8 ACGACCGC[1,0,0] GATC[1,0,0] 1...1000	RPI20
MS35	GTTTCG	8...8 TGATGCGC[1,0,0] GATC[1,0,0] 1...1000	RPI21
MS36	CGTACG	8...8 GTCGTATC[1,0,0] GATC[1,0,0] 1...1000	RPI22
MS37	GAGTGG	8...8 CATCAATC[1,0,0] GATC[1,0,0] 1...1000	RPI23
MS38	CCGTCC	not sequenced yet	RPI16
MS39	GTGAAA	not sequenced yet	RPI19
MS40	GTGGCC	not sequenced yet	RPI20
MS41	GTTTCG	not sequenced yet	RPI21
MS42	CGTACG	not sequenced yet	RPI22
MS43	GAGTGG	not sequenced yet	RPI23

(here 0) represents the deletions. This means that no insertions or deletions were observed.

7 Results and Discussion

In this section the results of the successful experiments are discussed. Representative images are shown, as most of the time experimental results were comparable. Also, the images and plots of failed experiments (i.e. the first few experiments that were carried out) are not shown here. Experiments that were carried out during the course of the thesis on other cell types than discussed previously were also omitted in the discussion. The results of the experiment on the variation in the dapi-signal channel will be discussed as well as YFISH and sequencing results from experiments as visible in table 4 and 5. Moreover, first correlation plots as well as results of principal component analysis are shown. Finally, our findings on genomic regions which differ in their radial placement between NES, NPC and neurons are discussed. However, since only one of the two replicates of the neurons has been sequenced to date, the analysis of the differential radial organization of the different cell types is preliminary.

7.1 Evaluation of variation in Hoechst-Signal

As described in section 6.2, 100 fields of view per sample were imaged and then the procedure repeated. One experiment, where random acquisition was used to select fields of view automatically (to avoid bias) did not work at all, as the cells were not completely evenly spread and the experiment yielded many fields of view with a very low number of cells, clumps of cells or no cells at all. Therefore the fields of view were picked by hand. When trying to image more than 20 fields of view (e.g. 40 or more) a drift of the coverslip was observed during the imaging process. This caused cells to be out of focus or not in the imaging plane anymore. Therefore the 100 fields of view had to be imaged in portions of 20 fields of view each (i.e. 5 times 20 fields of view as visible in table 1), which worked very well.

In figure 8 A-E, the comparison between imaging 20, 40, 60, 80 and 100 fields of view can be seen. The figure shows boxplots where the DNA per pixel between different experimental runs is compared. Sample 1.1 and 1.2 correspond to the two repetitions of sample 1, sample 2.1 and 2.2 correspond to the repetitions of sample 2 and samples 3.1 and 3.2 correspond to the repetitions of sample 3. It is visible that there is significant variation in the amount of DNA per pixel, when a low number of fields of view are imaged (i.e. 20 fields of view). At low numbers of nuclei imaged (400-800 in total), the signal between the two repeats of the same sample varies greatly. This is visible in the different value for the median (the horizontal line in each boxplot) of the different conditions. In figure 8B and C, it is visible that the signal intensities start to be more comparable. At 80 fields of view (figure 8D), which corresponds to a mean of 1480 nuclei analyzed, the signal median of each sample are almost equal. Imaging an average of 1700 nuclei as visible in figure 8E, there still is an improvement visible although not such a significant one as when the changes between the previous plots are compared. Interestingly, the difference in median in sample 3 increased from 80 fields of view imaged (figure 8D) to 100 fields of view imaged (figure 8E).

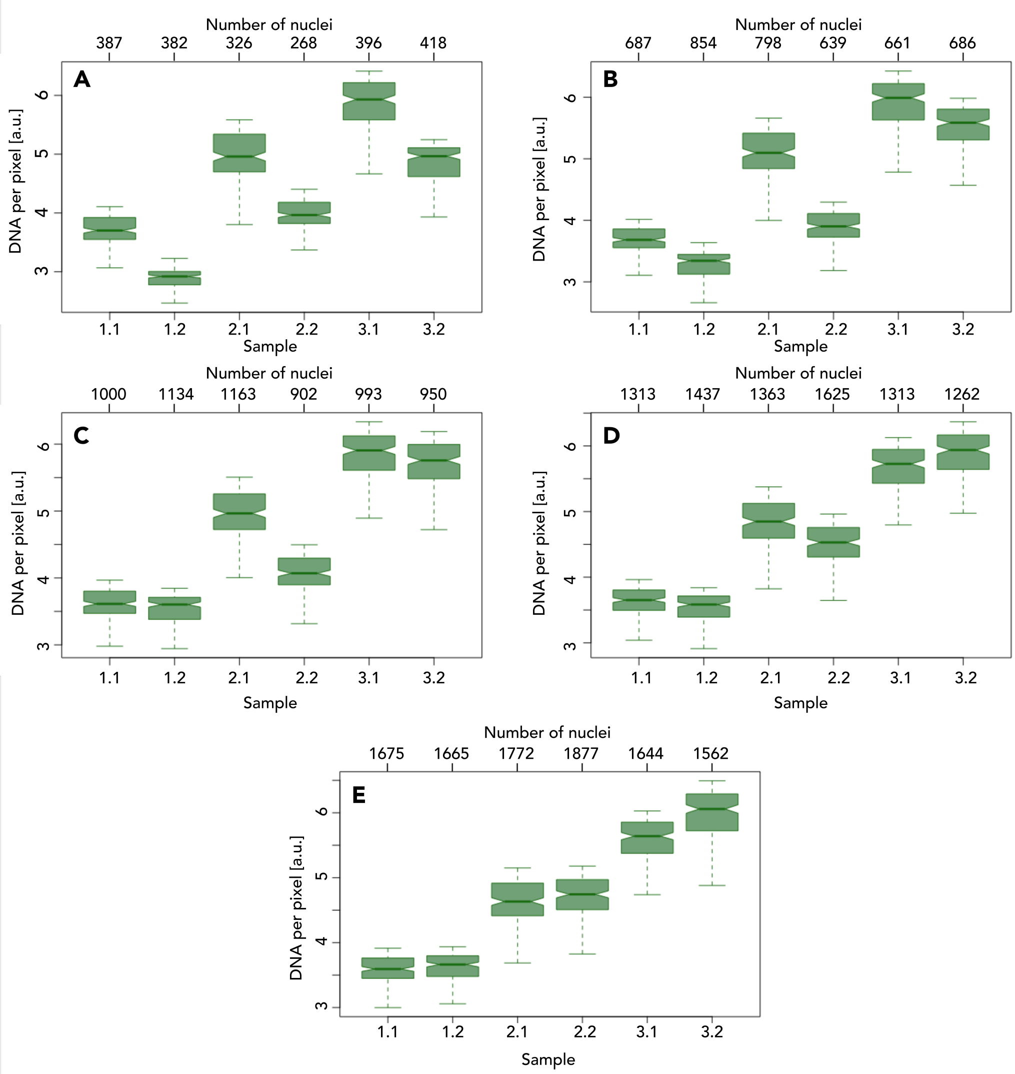


Figure 8: Boxplots showing the variation in Hoechst-Signal as DNA per pixel (a.u. refers to "arbitrary units"). (A) 20 fields of view; (B) 40 fields of view; (C) 60 fields of view; (D) 80 fields of view; (E) 100 fields of view. The number of nuclei imaged can be seen in the number above each boxplot. Sample 1.1 = sample 1, repeat 1; Sample 1.2 = sample 1, repeat 2; Sample 2.1 = sample 2, repeat 1 etc.

What is also interesting is, that all three samples were treated the same but nevertheless show a difference in median DNA per pixel between these samples. Conditions 1/2, 3/4 and 5/6 represent three different coverslips however all have NES from the same passage and day, were treated in the same way at the same time. The cells on the different coverslips were cultivated in the same way and fixed at the same time (in the same 6-well plate). Therefore it is surprising, that, when imaging 100 fields of view, the median of the signal is not the same for all samples, which would be expected from treating all samples the same way. The signal in sample 3 is even almost twice as high as in sample 1.

From this experiment it is clear that a high number of cells (i.e. at least 1500) should be imaged when doing a GPSeq experiment. Since the signal of the Hoechst-stain influences the calculations performed by the script described in section 6.3.4, the variation in the signal coming from this channel should be kept to a minimum.

Having said this, one has to keep in mind that it is sometimes simply not possible to image such a high number of nuclei that is determined here in this experiment. In many experiments that were carried out with neurons, there would not even have been this many nuclei on each coverslip. Moreover, it takes a lot of time to image this many nuclei and even when imaging a lot of cells, it is clear from this experiment, that a variation between samples remains, although the reason for this is not clear.

In summary, one should always image as many fields of view as possible. Ideally at least 1500 cells should be imaged per coverslip as imaging a too low number of cells may lead to errors or misinformation when comparing different samples. The strange issue why the coverslips show different median intensities has not been resolved.

It is important to keep these results in mind when performing GPSeq experiments and when evaluating the data resulting from these.

7.2 GPSeq

This section discusses the outcome of the GPSeq experiments. The use of HindIII vs. DpnII is discussed as well as the outcome from the YFISH experiment as well as from sequencing. First results related to correlation between experiments and between cell types are reported and the correlation between GPSeq score and size as well as GC-content are also shown. Results from principal component analysis as well as diagrams evaluating the difference between the examined cell types are shown.

7.2.1 YFISH reveals gradual digestion with HindIII and DpnII

Generally, the first checkpoint to see if an experiment worked, would be a thickening of the rim as visible in the fluorescence microscope after going through the YFISH procedure. The rim is sharper when a sample was digested with DpnII than with HindIII, but should be visible when any of these two enzymes was used. The DNA was stained with Hoechst, to show the outline of the nucleus. In figure 9 an example of how the digestion looks with a background of the Hoechst signal (blue) is shown. Here, the first three timepoints (10 sec, 20 sec, 30 sec) look almost the same, but then a clear thickening of the rim is visible. The last timepoint (5 min) shows filled in cells. These timepoints are only examples of digestion times used. As already mentioned earlier, the digestion time varies between cell types and even more when using different enzymes. This was a neuron sample, digested with DpnII so restriction happened very fast and the whole nucleus was digested already at 5 min.

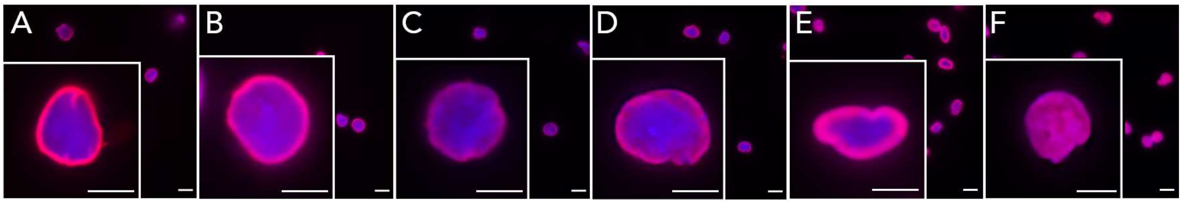


Figure 9: Close-up of a digestion process. In this case it was a neuron sample digested with DpnII. (A)10 sec; (B) 20 sec; (C) 30 sec; (D) 45 sec; (E) 2 min; (F) 5 min. Scalebars: 10 µm. Scalebars in insets: 5 µm.

As already described in section 6.3.6, in the beginning experiments were carried out using HindIII-HF. This ensured better control of digestion as it is a much slower process with this enzyme (rim appears at 10 minutes). In figure 10A, the epifluorescence microscopy images from the samples MS1-6 digested with HindIII-HF (see tables 4 and 5 for experimental conditions) can be seen. The time points used in this experiment were from top left to bottom right: 5 min, 10 min, 15 min, 30 min, 1 h, 2 h. It is clearly visible, that the rim thickens as the digestion time increases. In the 5 minute sample (top left), the signal intensity is very low. No rim is visible yet, however interestingly, small dots can be seen. These dots are not visible in the dapi channel which means that these are regions that are specifically digested first. It could be nucleoli or other regions in the nucleus, that have an exceptionally high or low chromatin density. What exactly these regions are, is not known, however this could be explored in the future using for example immunofluorescence staining. In the 10 minute sample (top center), the rim starts to become visible. Here another interesting observation (also still visible in other time points) is a tube-like structure going through the center of the nucleus.

It seems like there is some inversion or channel present, however again, we do not know what this could be. Like the dots, this is also not visible in the dapi channel. In the other time points (15 minutes to 2 h) a gradual filling in of the nuclei can be seen. In the 30 minute cell only the very center shows a lower intensity. In the 1 h and 2 h samples, the cells are completely filled in. These images show a satisfying result of the experiment and would encourage a continuation of the experiment (i.e. proceeding to sequencing).

In figure 10B, the profiles generated by the pipeline described in section 6.3.1 can be seen. On the left, the signal intensity across the relative distance from the nuclear lamina is plotted. It is clearly visible that the signal intensity scales with higher digestion times (i.e. the shortest digestion time, 5 minutes, shows the lowest intensity and the longest digestion time, 2 hours, shows the highest signal intensity. This is expected, as samples digested for a longer time would have more fluorescent oligonucleotides hybridized to the adapter and therefore yield a stronger signal.

The right plot of figure 10B shows the differentiation of the signal across the relative distance from the nuclear lamina. The curve of the 2 h sample, which is filled in, intercepts the x-axis at 0.4. The exact center of the nucleus would be at a relative distance from the nuclear lamina of 0.5. The reason why the intercept is around 0.4 as opposed to 0.5 can be explained by the fact that maybe not all cells are digested 100 % equally. Moreover the script assumes perfectly spherical nuclei, which is of course not the case and is also clearly visible in figure 10A. This explains the observed difference here. In figure 10B we also observe that the intercept with the x-axis moves to the left with shorter digestion times. The 5 min sample unfortunately differs from this pattern. This however is probably not due to the fact that the rim is thicker here. The problem here is explained by the fact that a clear rim is not visible yet and that low intensities cannot be registered properly by the pipeline. Moreover, in the 10 min sample, the intercept is also not that clear, where again the intensity is very low and cannot be properly processed by the pipeline.

This figure of HindIII was chosen as a representative image of experiments carried out with HindIII-HF. A lot of experiments showed similar results, however these images and profiles correspond to the sequenced samples (as shown in table 4 and 5) and therefore only these images are shown here.

In figure 11, the gradual digestion of neurons with DpnII is shown. The images as well as the profiles look similar for the other cell types (i.e. NES and NPC) and experiments performed, therefore only this figure is shown as a representative image. Figure 11A shows the epifluorescence microscopy images of different digestion time points (top left

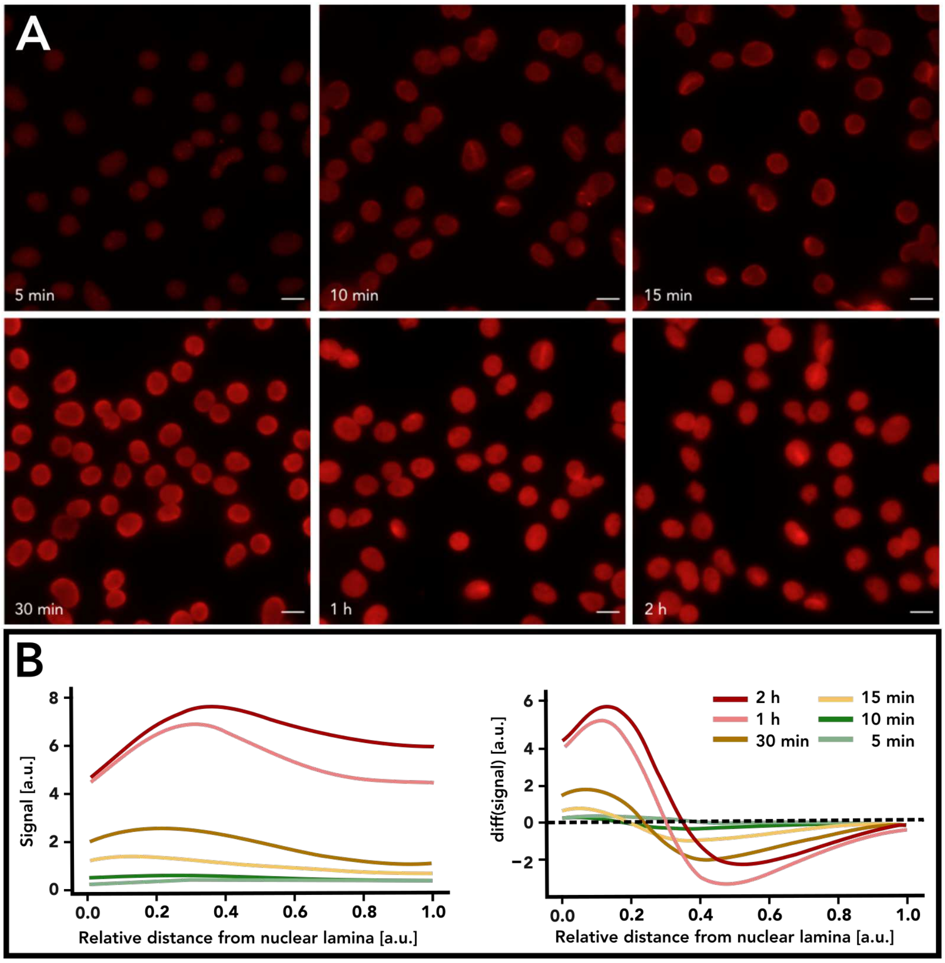


Figure 10: Figure showing the gradual digestion of NES cells (samples MS1-6) with HindIII. Scalebars: 10 µm. (A) Epifluorescence microscopy images showing different digestion time points (top left to bottom right: 5 min, 10 min, 15 min, 30 min, 1 h, 2 h). It is visible that the cells are more filled in the longer the digestion time. In the 1 h and 2 h sample, the cells are completely filled in. (B) Profiles generated using the custom pipeline (see section 6.3.1). It can be seen that the intensity increases as digestion time increases (left plot). The intercept with the x-axis moves to the right as the digestion time increases indicating a filling in of the cell and movement of the peak of the intensity (i.e. the position of the rim). [a. u.] = arbitrary units.

to bottom right: 10 sec, 20 sec, 30 sec, 45 sec, 2 min, 5 min). It is visible that the rim thickens as digestion time increases. The 10 sec sample (top left) for example shows a very thin rim. This rim gradually thickens in the 20 sec, 30 sec and 45 sec sample. In the 2 min sample, the cells are almost filled in where some show lower intensity only in the very center of the nuclei. The 5 min sample shows completely filled in cells. Figure 11B shows the profiles associated with the images in figure 11A. Again, like in the samples treated with HindIII-HF, here the signal intensity increases with increasing digestion time (figure 11B left). The 5 min sample shows highest signal intensity, while the 10 seconds sample shows lowest signal intensity and the samples in between scale accordingly. When looking at the plot of the differentiated signal (figure 11B right), the movement of the intercept with the x-axis can again be observed. As expected, the longest time point (5 min) intercepts at the highest relative distance from the nuclear lamina. The intercept with the x-axis moves to lower relative distance from the nuclear lamina as the digestion time decreases. The curve of the shortest time point, 10 seconds, does not intercept, however this is due to the detail available from segmentation with the pipeline. If signal intensities are very low, the signal cannot be accurately translated.

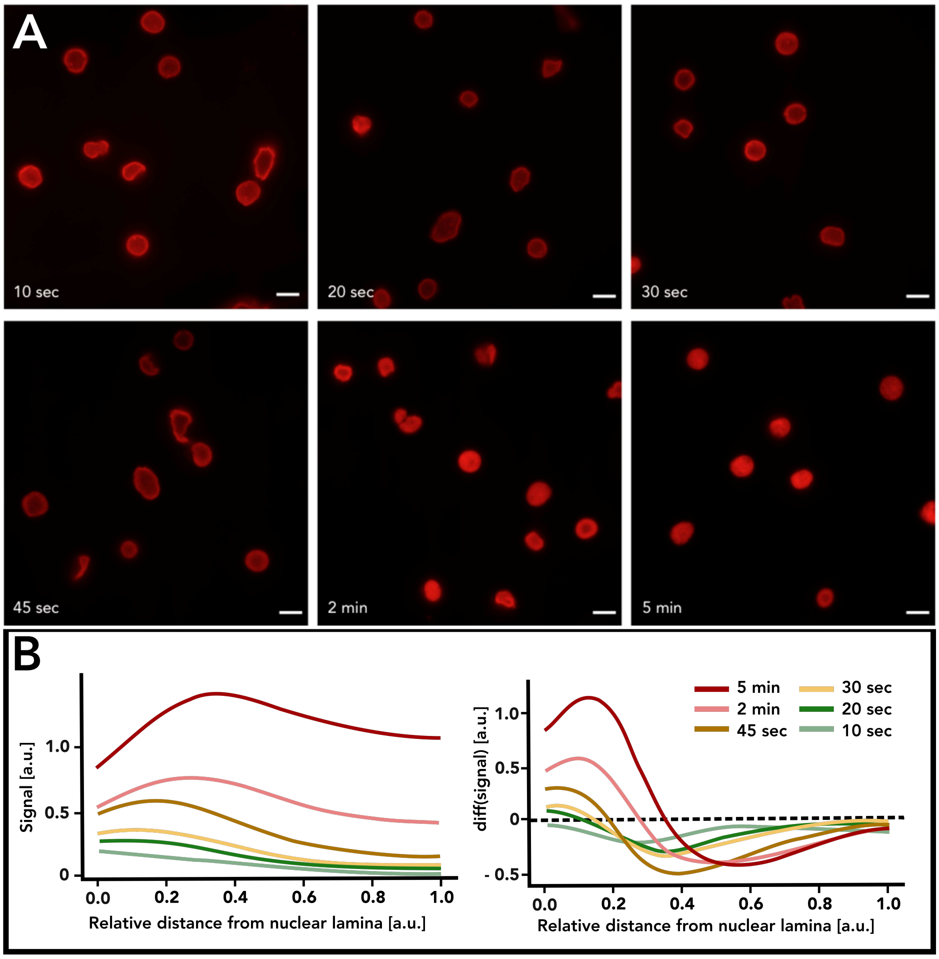

A
10 sec
20 sec
30 sec
45 sec
2 min
5 min
B
Signal [a.u.]
1.0
0.5
0
0.0
0.2
0.4
0.6
0.8
1.0
Relative distance from nuclear lamina [a.u.]
diff(signal) [a.u.]
1.0
0.5
0
- 0.5
5 min
2 min
45 sec
30 sec
20 sec
10 sec

Figure 11: Figure showing the gradual digestion of Neurons with DpnII (MS7-12). Scalebars: 10 µm. (A) Epifluorescence microscopy images showing different digestion time points (top left to bottom right: 10 sec, 20 sec, 30 sec, 45 sec, 2 min, 5 min). It is visible that the cells are more filled in the longer the digestion time. In the 2 min sample the rim is very thick and in the 5 min sample the cells are completely filled in. (B) Profiles generated using the custom pipeline (see section 6.3.1). It can be seen that the intensity increases as digestion time increases (left plot). The intercept with the x-axis moves to the right as the digestion time increases indicating a filling in of the cell and movement of the peak of the intensity (i.e. the position of the rim). [a. u.] = arbitrary units.

When comparing figure 10 and figure 11 we can clearly see that digestion with DpnII produces a much sharper rim than digestion with HindIII. It also seems that already at very short digestion times, higher signal intensities can be achieved. The advantage of slower digestion is, that the choice of timepoints is slightly easier. As we saw, the 5 min sample digested with HindIII did not show a rim yet, so we can be confident, that in the 10 min sample, the rim had just appeared and only the outside of the cell is digested. The lowest time point with DpnII, 10 sec, however already shows a clear rim. This means that is not as easy to say, when exactly the rim starts to appear. It is also difficult to do an even shorter digestion, as the time it takes to transfer the sample from the coverslip to the solution that stops the reaction, also takes some time. As already mentioned, the biggest advantage of using DpnII lies in the fact that it produces higher resolution libraries and it was also found to be more reliable in terms of how often the experiment worked when using a specific enzyme.
When looking at the profiles from experiments with different enzymes (figure 10B and 11B), we see that the results are comparable.

The biggest issue that appeared when using any enzyme was the choice of time points. As already discussed in section 6.3.6 a few parameters were varied to find the perfect experimental conditions. The problem however was that even when the experimental conditions already seemed to be at an optimum, the time points had to be varied for different cell types. After having decided on digestion times of NES with DpnII, the experiment was performed with neurons. There, a much faster digestion was observed. While the nuclei of the NES cells were not fully digested after 5 minutes, the nuclei of the neurons were almost all filled in in the 2 minute sample. Generally it was observed that the digestion of neurons occurs much faster than that of NES and NPCs. Whether the digestion time of NPCs is in between NES and neurons could not clearly be determined as they were very similar and to quantify this, more experiments would have to be performed.

A comparison between selected time points in neurons and NPCs can be seen in figure 12. It is visible, that the nuclei in the 2 min sample of the neurons (top center) are almost completely filled in, while the same time point in NPCs (bottom center) still shows a clear rim. When looking at the image of the 5 min time point, a similar observation can be made, where the neuron sample (top right) is completely filled in and the NPC sample (bottom right) still shows a lower intensity in the center of the nucleus, suggesting it is not completely digested yet.
Due to this issue, samples were usually processed for YFISH first so that not too many

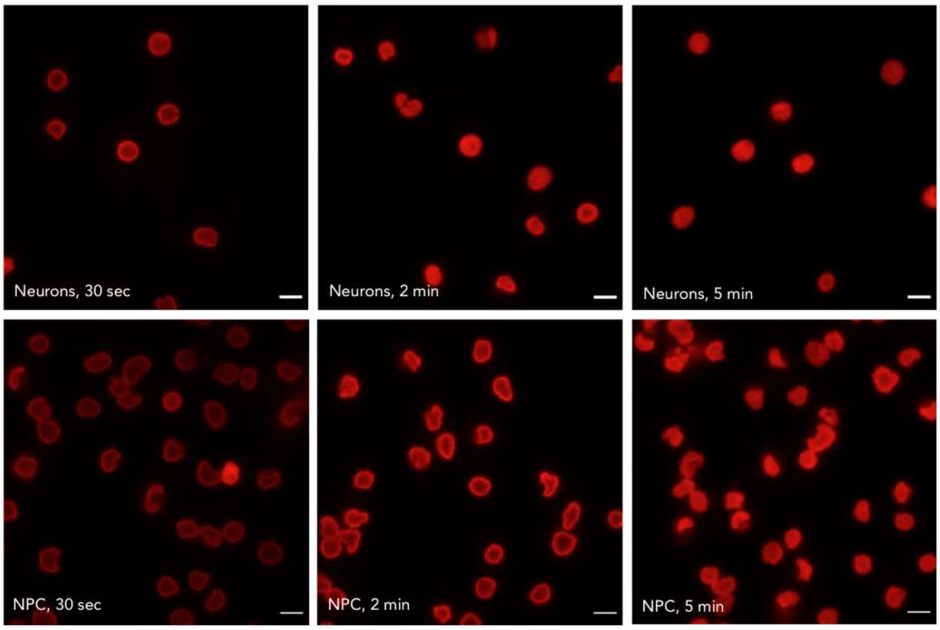


Figure 12: Comparison of neurons and NPCs at selected digestion times (left to right: 30 sec, 2 min, 5 min). The top row shows images of neurons (MS9,11,13), while the bottom row shows images of NPCs (MS26-28). Scalebars: 10 µm.

coverslips were wasted in the search of correct time points.

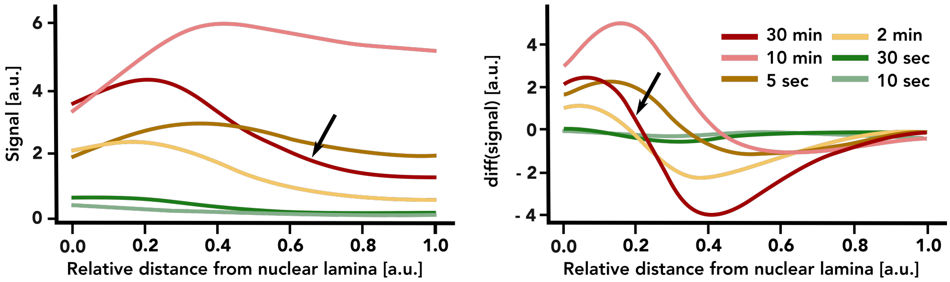


Figure 13: Profiles for one an experiment done with NPC (MS25-30). It can be seen that the 30 min sample is not filled in and shows an intensity drop (arrow). On the left the signal intensity is shown; on the right the differentiated signal is plotted against the distance from the nuclear lamina. [a. u.] = arbitrary units.

Other issues that were encountered during the course of this thesis, were with nuclei not filling in. In some microscopy experiments, nuclei with homogeneous signal were observed in the 30 minute time point, but then not in the 1 hour time point. This problem was observed in two NES samples but also in one NPC sample when using DpnII. A speculation is that after long digestion using a fast-digesting enzyme such as DpnII, there are so many cutsites at the periphery of the nucleus, that the signal from these "overpower" the signal coming from the center. In figure 13 the profiles from an experiment done on NPC (MS25-MS30) are shown. Here the 30 min sample was not filled in while the 10 min sample was. This was also confirmed by the profiles generated. On the left in figure 13 the signal intensity is plotted. Here the intensity of the 30 min sample drops below the intensity of the 10 min sample, while it should actually have a higher signal throughout. On the right, the differentiated signal is plotted. Here, it can be seen that the peak of the 30 min sample moves further to the left, indicating the rim being thinner. The other time points however scale as would be expected (i.e. the shortest time point has the thinnest rim while the longest time point is filled in).

7.2.2 Sequencing showed satisfying numbers of bins but issues with linkers were observed

When the images and profiles obtained from the YFISH procedure showed satisfying results and a thickening rim was observed as well as when the profiles showed the desired trend (see figures 10 and 11) the other half of the samples (i.e. the coverslips with the sequencing linkers) were processed further according to the method described in section 6.3.5.

During the processing according to protocol, the samples were analyzed using the Bioanalyzer 2100 and measurements of DNA and RNA concentrations were performed using Qbit. This was to confirm, whether purification was successful and to measure the concentrations of DNA or RNA in the samples. These checks were done before IVT, after RNA cleanup and after the cleanup of the final libraries. Fragments should be around 300 - 500 bp as this is the optimal length for the Illumina NextSeq 500 platform.

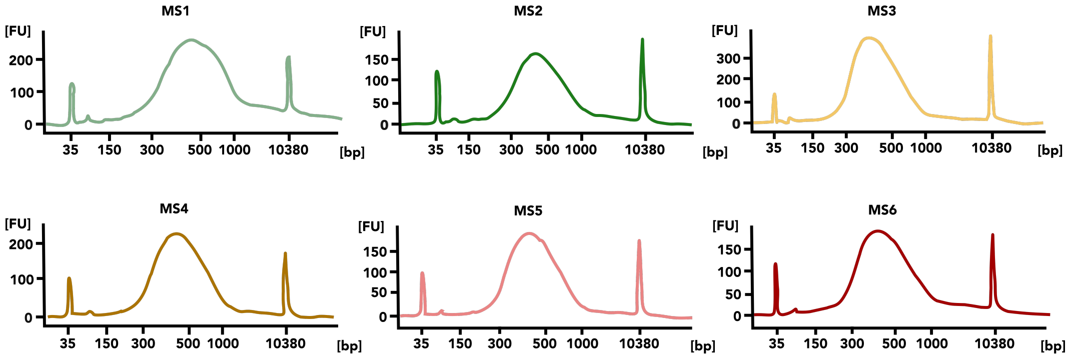


Figure 14: BioA profiles of the NES samples with HindIII (MS1-6)

The RNA concentration as well as the DNA concentration after the cleanup of the final libraries should scale according to digestion time. This means that the sample which was digested longest, should have the highest amount of RNA/DNA as there should be more fragments from more restriction sites. An example of this can be seen in table 6 where the DNA concentrations of NES samples (MS1-MS6) are listed. We see, that

TU Bibliothek WIEN Your knowledge hub
Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.
The approved original version of this thesis is available in print at TU Wien Bibliothek.

the shortest timepoint (5 min) has the lowest DNA concentration with 6.42 ng/µl. The longest digestion time (2 h) does not have the highest DNA concentration, but the 1 h sample does. The reason why the concentrations of the 1 h sample and 2 h sample might not scale, is because the cells are already filled in at 1 h so not much difference is to be expected between these samples. As this serves as a checkpoint, seeing a scaling like this can show that the sequencing will probably show satisfying results. In figure 14 an example of the output of the BioAnalyzer 2100 can be seen. It shows that the fragments have the desired length of about 300 - 500 bp. The peaks at 35 bp as well as at 10380 bp are the markers. BioAnalzer profiles of the other samples (MS7-MS43) had similar profiles.

Table 6: DNA concentration of neuron samples measured with Qbit after final library cleanup

Sample (library names & timepoints)	Concentration (ng/µl)
MS1 (5 min)	6.42
MS2 (10 min)	7.58
MS3 (15 min)	12.0
MS4 (30 min)	39.0
MS5 (1 h)	67.0
MS6 (2 h)	62.4

Across all experiments, where samples were sequenced, the yield in terms of sequencing output, and therefore library sizes, was satisfying. The total number of reads obtained from the sequencing runs is summarized in table 7. The samples MS38-43 are not sequenced yet.

Table 7: Number of reads from sequencing runs

Sequencing run (library names)	Number of reads (in million)
1(MS1-6)	377
2 (MS7-12)	485
3 (MS13-17)	443
4 (MS18-23)	433
5 (MS25-30)	419
6 (MS32-37)	420

In figure 15 the number of reads after each filtering step of the experiments performed on

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

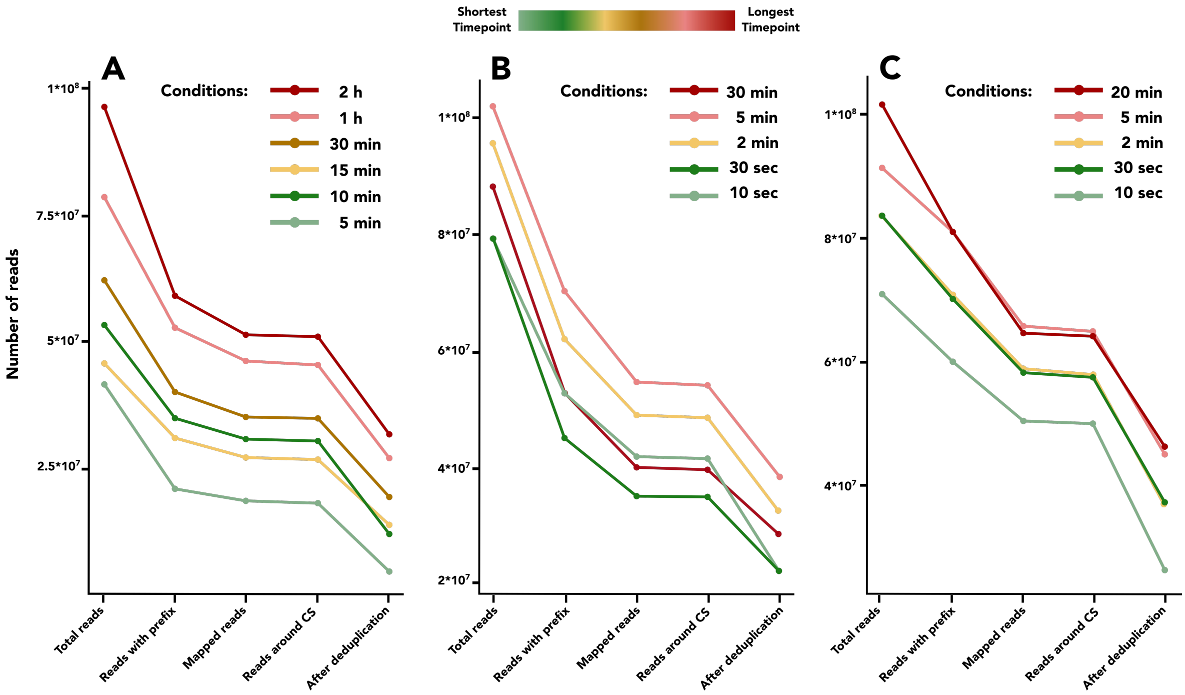


Figure 15: Profiles showing the number of reads after each filtering step (reads with prefix, mapped reads, reads around cut site (cs), reads after de-duplication) of the experiments performed on NES cells. (A) shows the experiment done with HindIII; (B) shows the first experiment done with DpnII; (C) shows the second experiment done with DpnII. Dark red always shows the longest digestion time then shorter digestion times are shown with light red, brown, yellow, dark green and light green (in this order).

NES cells is visualized. Dark red always shows the longest digestion time then shorter digestion times are shown with light red, brown, yellow, dark green and light green (in this order). Note that the specific digestion times vary between the experiments. As already discussed previously, digestion with HindIII is slower and therefore requires longer digestion times (figure 15A). In figure 15B and C experiments with NES and DpnII are shown where the only difference is the longest digestion time point (in 15B it is 30 minutes while it is 20 minutes in figure 15C). The reads are filtered as described in section 6.3.7. It is shown, that the number of reads is reduced after each filtering step. First the total number of reads are shown. Then the next step visible in the plots is the reads that contain the full prefix. Next, the sequences mapped to the human reference genome are shown. Then, the reads are filtered for reads around the cut site. These are reads whose 5’ end is less than 20 bp away from the position of the recognition site. Finally the number of reads after de-duplication, so the reads whose UMI sequences were mapped to the same recognition site, are shown.

The most drastic drop obviously is the filtering for reads with prefixes. The filtering for duplicate reads, which occur due to the PCR step in the protocol, also show quite a

big reduction in number of reads. These resulting numbers, after de-duplication are the ones that should be observed in close detail. They should scale according to digestion time, which means that the longest digestion time should yield the highest number of reads. This is due to the fact that samples that were digested longer have more cut sites and thus result in more reads being generated. In figure 15A, the samples digested with HindIII, this trend is clearly visible. The samples treated for 5 min with HindIII show the lowest number of reads and then a scaling is visible up to the longest digestion time of 2 h, corresponding to the largest number of reads. In figure 15B, which shows the first experiment with DpnII, this scaling is not as clear. The two shortest time points, 10 sec and 30 sec (light green and dark green), have approximately the same number of reads, where the number of reads of the 10 sec sample is slightly lower. The rest of the samples scale nicely, however the longest time point, 30 min (dark red), shows a deviation from this trend. Similar results can be observed in the second replicate with DpnII in figure 15, where the numbers scale according to what would be expected. Only the 30 sec (dark green) and 2 min sample (yellow) are almost the same).
Similar findings were observed in the samples of the NPCs and Neurons.

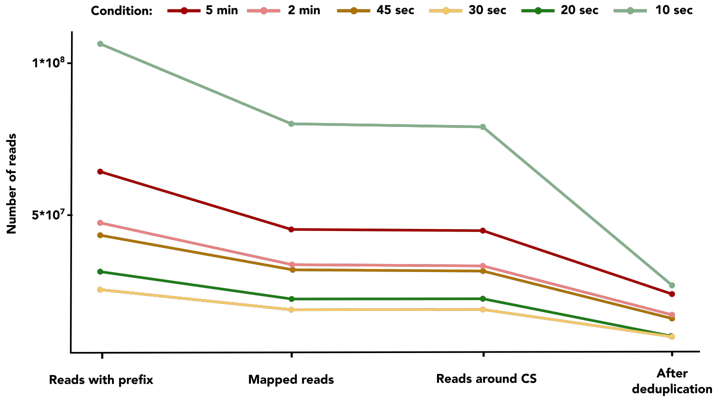


Figure 16: Profiles showing the number of reads after each filtering step (reads with prefix, mapped reads, reads around cut site (cs), reads after de-duplication) of the first experiment performed on Neurons cells. Dark red shows the longest digestion time then shorter digestion times are shown with light red, brown, yellow, dark green and light green (in this order).

In figure 16 the number of reads after the different filtering steps for the first experiment with neurons and DpnII is shown. Here, the reads also scale according to expectations, except for the shortest time point, 10 sec (light green), which does not fit into this pattern. The reason for this however could also simply be, that there were more cells on the coverslip used for this time point or that the DNA from this coverslip was transferred more quantitatively between different steps.

A reoccurring issue was observed in experiments with DpnII when pre-processing the

sequencing data. In all experiments with DpnII (corresponding to samples MS7-37) a strong loss of read quality at the 3' end of the barcode and cut site was observed. This resulted in a mismatch at the last (3' end) position of the expected barcode sequence, which could be observed in a majority of reads. The observed mismatch was a C to A conversion in most cases. The correct barcode sequence was not found at any other position in these fragments. Moreover, approximately 15 % of the library fragments had up to 1 mismatch at the expected cutsite sequence. This problem was resolved by selecting for sequences with up to one mismatch to recover most of the reads. This was feasible, as the barcodes have a minimum Hamming distance of 2 between any pair.
It remains to be investigated whether this is a sequencer-related issue, a linker-related issue or an enzyme-related issue. For checking if it is a linker-related issue, libraries digested with for example MboI, could be sequenced, as these have the same cut site and therefore the same linkers are used. If all future runs show this issue, the issue probably arises from the sequencing machine. If the issue is only observed with DpnII, it would appear to be a problem related to this specific enzyme. With HindIII this issue was not observed, so it is not sequencer related. As mentioned in section 5.11 substitutions are a the most common error type in Illumina Sequencing platforms.

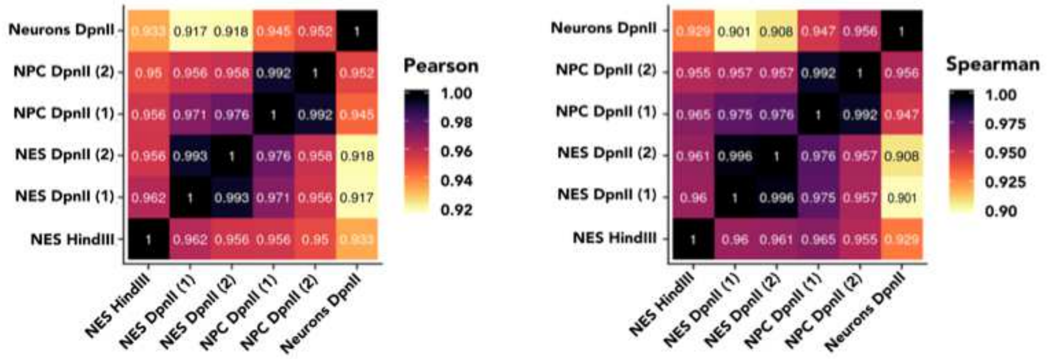
Overall, the yield in terms of sequencing output (i.e. library size) and fraction of duplicated fragments (i.e. low) is very good. A very mild drop of mapped fragments in HindIII of NES and neuron runs was observed. Moreover, there was no change in the fraction of reads mapped far away from known enzyme recognition sites.

In summary, 7 libraries were prepared after showing the desired result in the YFISH experiment. One experiment with NES and HindIII was processed for sequencing as well as 2 experiments of each cell type (NES, NPC and neurons). Digestion with HindIII occurred much slower than with DpnII and digestion times were different for different cell types. While the YFISH experiment worked most of the time, sometimes issues were observed with the rim reappearing in long digested samples.
Sequencing output was good and the fraction of duplicated fragments was low for all sequenced libraries. Issues were observed with a loss of read quality at the 3' end of the barcode but this was fixed during pre-processing.

Following sequencing, first analysis was performed. GPSeq score correlation between experiments was calculated to see if reproducible libraries were generated. As described in section 5.5, different models for radial positioning of chromosomes have been proposed. One is the organization by size, where small CTs are located in the center of the nucleus and large chromosomes are found at the periphery of the nucleus. The other model is an organization based on gene-density, where gene-rich chromosomes are preferentially located towards the nuclear interior, while gene-poor chromosomes occupy peripheral locations. Therefore we also evaluated the correlation between GPSeq score and size as well as GC-content.

7.2.3 GPSeq score correlation shows reproducibility of results

(A) 1 Mb GPSeq score correlation



(B) 100 kb GPSeq score correlation

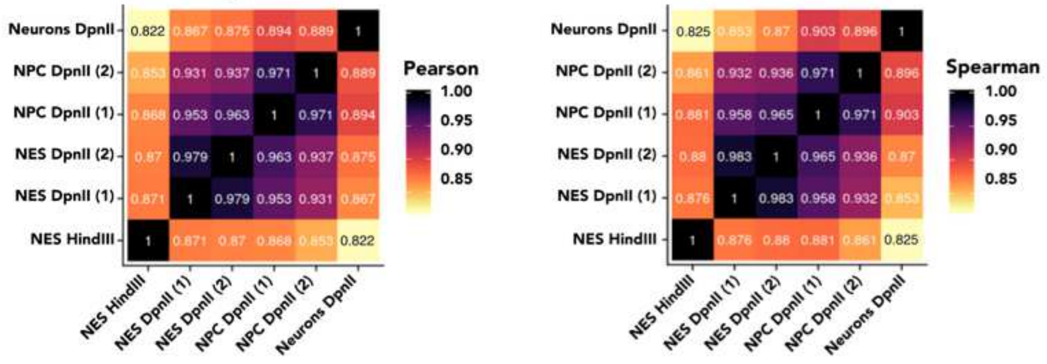


Figure 17: GPSeq score correlation at (A) 1 Mb and (B) 100 kb resolution of all sequenced experiments

In figure 17 the correlations of the GPSeq score at 1 Mb (figure 17A) and at 100 kb (figure 17B) between all performed experiments (see table 4) is visualized. The GPSeq score was calulcated according to the method described in section 6.3.7.

The GPSeq score used here was calculated using all available conditions for the DpnII datasets and all the conditions for the HindIII dataset. Correlation between all datasets is quite high. Overall, as one would expect, the correlations are higher when looking at a lower resolution (1 Mb vs. 100 kb).

In figure 17A it can be seen that the correlation between that samples at the same differentiation stage show higher correlation than between samples at different stages. For example, the correlation between the two replicates of NES with DpnII is 0.993 using Pearson correlation and 0.996 using Spearman correlation. The correlation between the two replicates of NPCs with DpnII is 0.992 and 0.996 using Pearson and Spearman correlation respectively. The correlation between NES and neurons using DpnII is 0.917 (Pearson) and as low as 0.901 (Spearman) which is significantly lower than the correlation between two NES or two NPC samples but also lower than between NES and NPC. This suggests, that in terms of differentiation stage, NPC are more similar to NES than to neurons. The largest difference is observed between NES and neurons. Thus, it appears that the correlation plots reflect the differentiation from NES to NPC and finally to neurons.

When just looking at the colours of the plot, this can easily be seen. The dark squares, indicating high correlation, are clearly the replicates of NPCs and NES. This also shows that these replicates can be compared to each other and that the results are reproducible.

The HindIII experiment shows a distinct drop in correlation, but is still higher with other NES samples than with NPC and neuron samples.

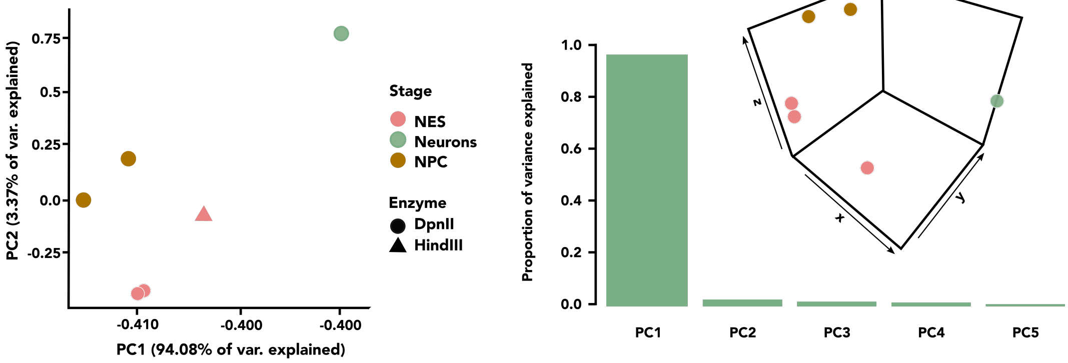
When looking at figure 17B similar results can be observed. The numbers here are slightly lower, which was to be expected at higher resolution. However, still darker squares indicating higher correlation within the same cell type as compared to between cell types can be observed which speaks for a good quality of replicates. The correlation between NES and NPCs is again higher than with neurons and NPCs or NES. Again, we see that the correlation is also higher between neurons and NPCs than between neurons and NES. This indicates higher similarity between NES and NPCs than between neurons and NES or NPCs. The neurons are more similar to NPCs than to NES again demonstrating the differentiation starting from NES and further to NPC and finally to neurons.

Comparing the NES sample digested by HindIII is not really relevant here, as restriction with HindIII achieves much lower resolution than when using DpnII (see section 5.12). This is also visible in the numbers. The correlation between the HindIII sample of NES and the DpnII sample of NES is still slightly higher than with other cell types, but it is not as clear as in the plot where the correlation is calculated at 1 Mb in figure 17A.

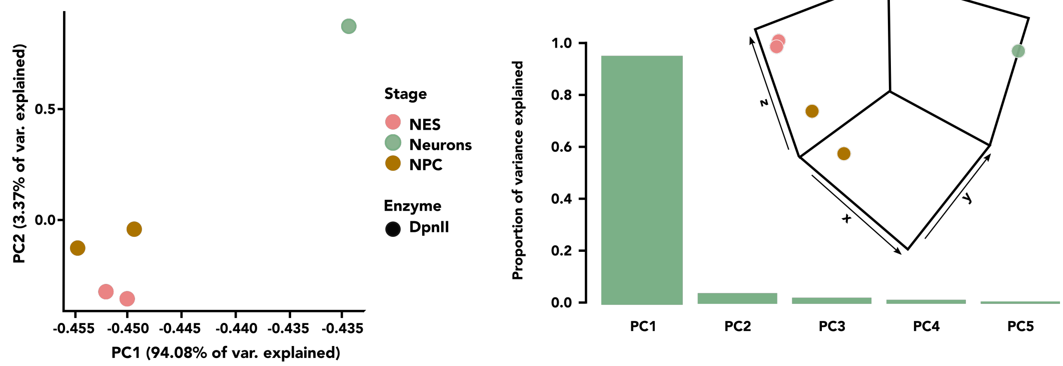
The reason why both Pearson and Spearman correlations are calculated here, is because we do not know yet what exactly the relationship between the two variables is. Generally, the correlation coefficient describes both the strength (value) and direction (+/-) of the change. Pearson correlation describes the linear relationship between two continuous variables. Spearman correlation describes the monotonic relationship between two continuous or ordinal variables, which means that variables tend to change together but not necessarily at a constant rate. Since we do not know the exact relationship and which of these correlations should be used, both are listed here.

To recapitulate, the correlation between samples of the same cell type is higher than between samples of different cell types. This shows that the experiments are reproducible and comparable. The stages of differentiation seem to be clear from these plots, where the NES are more similar to NPCs than to neurons. The neurons are more similar to NPCs than to NES, which shows the differentiation from NES and further to NPCs and finally to neurons.

7.2.4 Principal Component Analysis shows clustering of samples from same cell type



(A) 1 Mb PCA



(B) 100 kb PCA

Figure 18: Principal component analysis (PCA) at (A) 1 Mb resolution and (B) 100 kb resolution. At 100 kb the sample with HindIII is omitted.

In figure 18 the outcome of a principal component analysis (PCA) is shown. In figure

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

18A the analysis was done with a resolution of 1 Mb and in figure 18B at 100 kb. PCA is frequently used for exploratory data analysis such as support for cluster analysis and assignment of cluster. It is also used for the search for influential factors as well as reduction of the dimensionality of data. Therefore it is a method to project a p-dimensional space to 2 or 3 dimensions via orthogonal transformation combined with visual inspection and interpretation.

As can be seen in the PCA performed on these samples, principal component 1 (PC1) explains, regardless of the resolution investigated, more than 94 % of the variance. Therefore there is one main factor determining the radial location of a chromosome. We can also see, that samples at the same differentiation stage seem to cluster together. At 1 Mb (figure 18A, the two NES samples (pink) almost overlap. The two NPC samples (brown) are also quite close together, while the neuron sample (green) is very far away from all others. The NES sample digested with HindIII (pink triangle) is closer to the NES samples than the average of the NPC samples. When looking at the graphic showing the use of three dimensions (on the right) the clustering is even more visible. The NES samples are in close proximity to each other, as are the NPC samples, while the neuron sample is far away from all others.

Similar results are obtained when looking at figure 18B at 100 kb resolution. Here, again samples from different stages seem to cluster together. The neuron sample is again very far away from the other samples. Especially in the plot showing three dimensions, the clustering of the two NES samples can be observed. At this resolution the NES sample digested with HindIII was omitted, as the resolution achieved here was too low to use it as a comparison.

Due to past explorations of radiality of chromatin arrangement[19,22,23,63,68,74–78] (see section 5.6 for details) it is probable that PC1 is GC-content or gene density. Chromosome size might also be one of the determining principle components.

It remains to be explored, if after sequencing the second neuron sample, these observations still hold true and whether the clustering improves or deteriorates upon addition of this sample.

Overall, samples of the same cell type cluster together which was expected. Moreover, NES and NPC seem to be more similar to each other than to neurons. Following this, it is interesting to investigate the relationship between radiality and chromosome size as well as GC-content.

Table 8: Correlation of GPSeq score with chromosome size and GC-content calculated with data from the experiment with NES and HindIII

GPSeq score & chromosome size	
Pearson	Spearman
-0.714	-0.705
GPSeq score & GC-content	
Pearson	Spearman
0.897	0.909

7.2.5 GC-content and chromosome size

In table 8 the outcome of the calculations for the correlation coefficients of the GPSeq score with chromosome size and GC-content are shown. The data is from the experiment performed on NES with HindIII as the restriction enzyme. It is visible, that the GPSeq score and the chromosomes size are anti-correlated, meaning that the larger chromosomes are preferentially located in the periphery of the nucleus, while smaller chromosomes tend to be located at the center of the nucleus. The strength of the anti-correlation is however lower than that of the correlation of GPSeq score and GC-content. This has also been observed previously in other cell types using GPSeq[1].

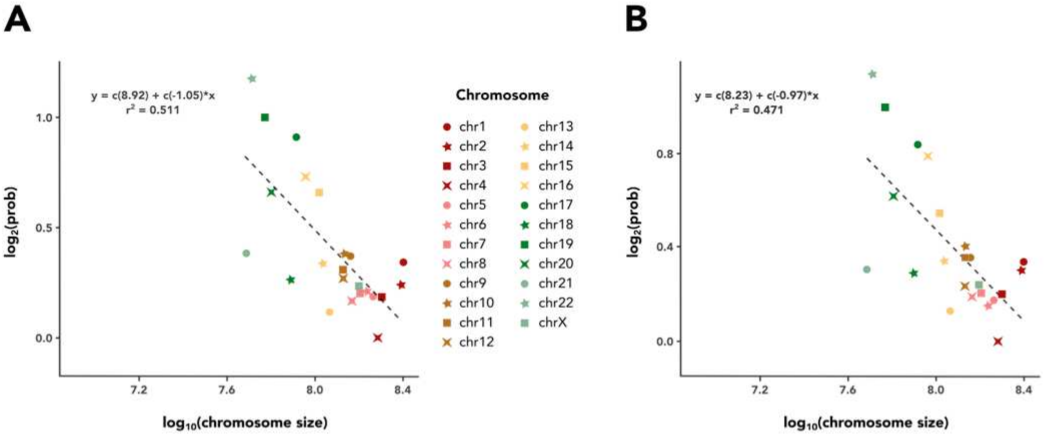


Figure 19: Correlation between GPSeq Score and Chromosomes size per chromosome (A) Experiment with NES and HindIII; (B) Experiment with Neurons and DpnII

In figure 19 the correlation between the GPSeq Score and chromosome size per chromosome is visualized. In figure 19A the plot from the experiment with NES and HindIII is visible and in figure 19B the plot for the experiment with neurons and DpnII is visualized. The x-axis shows the $\log_{10}$ of the chromosome size and the y-axis shows the

GPseq score based on the probability estimate. We can see, that the two plots look really similar. As already mentioned, and also visible here, larger chromosomes tend to occupy the periphery of the nucleus, while small chromosomes tend to be located more centrally. A high GPSeq score corresponds to a central location, while a low GPSeq score corresponds to a peripheral location. For example, chromosome 1 (dark red dot), which is known to be the largest chromosome, seems to be located in the periphery of the nucleus. Chromosome 2, 3 and 4, which are also quite large, are also located in the periphery. This is true both in the NES and the neuron sample. However, as already mentioned, the chromosome size is not always an accurate predictor for the radial location. When looking at chr19 for example, which is a relatively small chromosome, we can see that it has a very high GPSeq score corresponding to a central location in the nucleus. Chr18 on the other hand, which is also a small chromosome, is located in the periphery of the nucleus. These observations have also previously been made with other cell types (HAPI) using GPSeq[1]. The reason for this is probably that chr19 is gene-rich, while chr18 is gene-poor. This would be in line with the assumption, that the radial location is both influenced by the size and by the gene-density as discussed in section 5.5.
Another example is chr21 and 22. Both have approximately the same size (both small), but chr 22 is located in the center of the nucleus, while chr 21 is located towards in the mid region.
This shows, that the chromosome size alone is not an accurate predictor of the radial location of the different chromosomes.

When we go back to table 8 we see that GPSeq score and GC-content show a positive correlation. This suggests, that chromosomes with a high GC-content can be found at more central locations, while chromosomes with a low GC-content are preferentially positioned in the periphery. The strong correlation (0.897 in Pearson correlation and 0.909 in Spearman correlation) suggests, that the GC-content represents an accurate predictor of radiality in these cell types. Probably, the best predictor would be to create a model where a combination of both the size and the GC-content are used.
These results have also previously been observed in other cell types using GPSeq[1].

So we see that here the GPSeq score shows a positive correlation with GC-content, indicating that gene-rich chromosomes tend to be located in the center of the nucleus while gene-poor chromosomes tend to be located at the periphery. GPSeq score and chromosome size is however anti-correlated indicating that larger chromosomes tend to be located in the periphery while smaller chromosomes are usually found in the center

TU Bibliothek WIEN Your knowledge hub
Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.
The approved original version of this thesis is available in print at TU Wien Bibliothek.

of the nucleus. This however is not a strict rule, as exceptions from this are also observed.

Now that we saw that the radial arrangement in the investigated cell types seems to change upon differentiation, the goal is to find regions of interested that can be explored in more detail.

7.2.6 Finding regions of interest

Now we have seen, that the correlation between samples at the same differentiation stage is higher than between different cell types (see section 7.2.3). We also observed, when performing a principal component analysis, that the samples at the same differentiation stage seem to cluster together and there is one PC that explains most of the variance. Therefore it would be interesting to find the regions in the genome, that change or relocate during differentiation in respect to their radial location.
For this, the idea is to automatically identify regions that show "different" GPSeq scores between cells at diverse differentiation stages. The reasoning is that regions with large centrality variation during differentiation should show differences also in terms of for example Hi-C structures, gene expression and epigenetic markers.

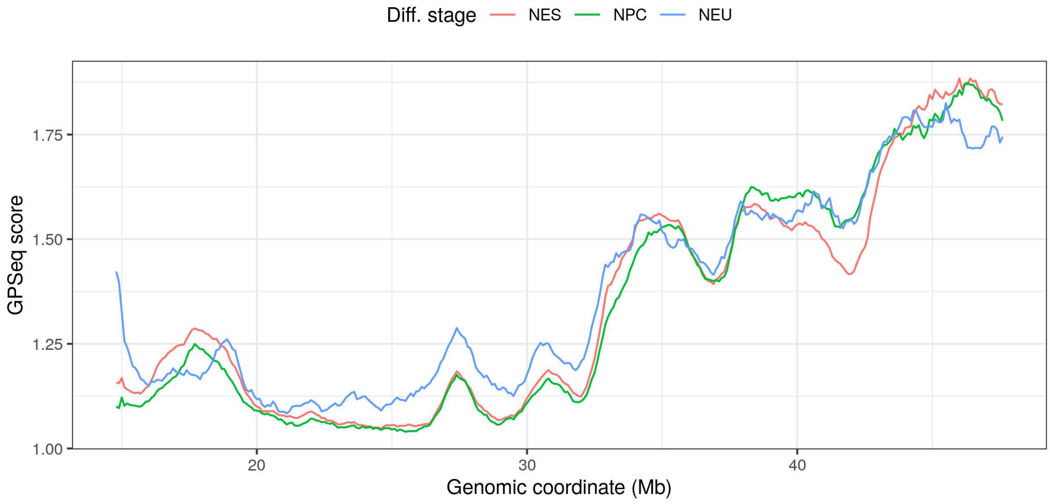


Figure 20: Case study: chr21 at 1 Mb: Average profiles of the GPSeq scores of chr21 at different differentiation stages (NES = red, NPC = blue, Neurons = green).

In figure 20 the average profiles of all the GPSeq scores from the same differentiation stage of chr21 are plotted. This chromosome was chosen as it is a small chromosome and therefore its size makes it easy to visualize. We can see that the profiles often overlap, however there are some regions that seem to differ between NES, NPC and neurons. For example between 25 and 30 Mb, the neuron profile (green) shows on average a higher GPSeq score than the NES and NPC profiles which almost perfectly overlap here. This indicates a more interior radial location of this region of chr21 in neurons. At approximately 42 Mb the NPC and neuron plot overlap, while the NES shows a deviation towards a lower GPSeq score. This would indicate that this region has a more peripheral location as compared to the other differentiation stages, which means that this region relocates to the nuclear interior upon differentiation.

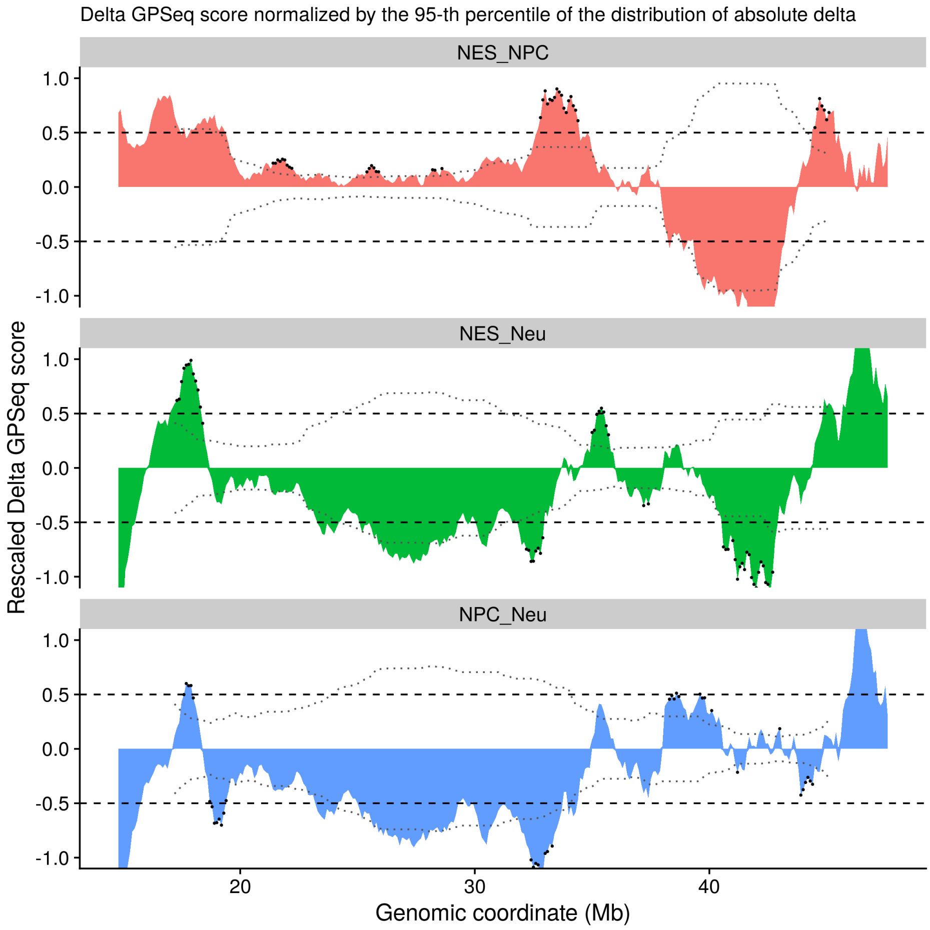


Figure 21: Identificaiton of "strongly" different regions in chromosome 21 using local and global thresholds.

In figure 21 an approach to identifying different regions is visualized. Again, chr21 is analyzed here. For this, the difference between GPSeq profiles (delta) was measured. Then this was devided by the 95th percentile of the absolute data distribution, which allows to visualize the delta profiles at comparable scales and is not as affected by outlier values as normalization by the maximum values. Two different approaches for identifying differential regions can be used: using *local* or *global* differences. The local approach uses the local threshold (e.g. 1.5 times the local median). The global approach utilizes a global threshold (e.g. 0.5, which is half the normalization range). The global threshold is represented by the horizontal dashed line and the local threshold is shown by the dotted grey curve. Black dots represent bins that pass the local threshold. On the top (in red) the comparison between NES and NPC is shown, in the middle, the comparison between NES and neurons is shown (green) and on the bottom, the difference between NPC and neurons (blue) is visualized. The delta GPSeq score is calculated by substracting the GPSeq score of the second cell type from the first (i.e. in the NES_NPC diagram this would be GPSeq Score(NES) - GPSeq Score(NPC) = delta GPSeq score (NES_NPC))

It is visible, that the biggest change from the NES to NPC stage (top plot, red colour) seems to be at approximately 35 Mb. Here the strongly different regions are both above the local and the global threshold. The positive delta GPSeq score here indicates, that these chromosomal regions relocated from a rather peripheral to a more central location. Between 20 and 30 Mb there are some other indicated regions, which are above the local threshold but not above the global one. When differentiating further, we see that even bigger changes happen.

The region between 40 and 45 Mb seems to change a lot as visible in the middle plot (green) where NES and neurons are compared. The changes happening in these regions, probably already happen in the NPC stage, as a strongly negative delta GPSeq score is visible in the same region in the plot at the top (red). These show a negative delta GPSeq score indicating a relocation from peripheral to more central positions.

Changes from the NPC to the neuron stage (blue) mainly happen at around 18 Mb, 33 Mb and 44 Mb. At 18 Mb the delta GPSeq score is positive, thus reporting a move of chromosomal region from periphery to center. At 33 Mb as well as at 44 Mb, the value is negative, which implies a move from peripheral to central regions.

Overall, it seems that the biggest change happens when differentiating from NPC to neurons. This is no surprise, as the differentiation from NES to NPC is only a 6 day process, while the differentiation to neurons takes 5 weeks. Generally it seems, that between each differentiation stage there are different regions that are going through relocation processes.

These findings are interesting, as the position of many genes in the nucleus has been correlated with their activity state and thus, gene expression could be influenced by migration to functional subcompartments. Euchromatin, which is loosely packed and associated with transcriptionally active regions, is usually found in the interior of the nucleus[36], while heterochromatin is more compacted, peripheral and associated with a low gene density which are mostly transcriptionally inactive[32]. In figure 21 it seems that a big part of chr21 is recruited to central regions and thus to a transcriptionally active one. This analysis however provides not a detailed insight, but rather a starting point for more detailed analysis of specific genes in this chromosome.

8 Conclusions and Future Recommendations

In conclusion, we can see that the GPSeq method works both with HindIII and DpnII. However, digesting with DpnII achieves a higher sequencing resolution, as this enzyme cuts more frequently. While DpnII digests very fast and more often, restriction with HindIII occurs much slower and less frequent. A disadvantage when using DpnII that was observed, was a loss of read quality at the barcode-cutsite junction, with a C to A mismatch in most barcodes. The read with the mismatch is the most represented in most libraries. This issue is however easily fixed computationally during pre-processing due to the fact that the minimum Hamming distance between each pair of barcodes is 2.

Moreover, interesting novel observations have been made from YFISH, such as tubular structures or inversions, that were seen in all investigated cell types. Small dots spread through the nucleus were also observed. What the inversions and the dots exactly are, remains to be investigated.

In the course of this thesis, a total of seven libraries were produced. Two replicates of each cell type were produced with DpnII as well as an additional library of NES with HindIII. One neuron library still remains to be sequenced. The libraries at the same differentiation stage showed very high correlations (see section 7.2.3). GPseq score profiles at different differentiation stages however are less correlated, indicating a difference in radial arrangement between NES, NPC and neurons, and therefore a rearrangement of the chromatin during differentiation.

PCA on GPseq score profiles seems to be able to distinguish NES, NPC and neurons in 3 different clusters upon visual inspection. The PCA should however be repeated, once the last neuron sample is sequenced.

We also started to search for regions of interest termed "strongly different regions". These were selected using both a local and a global threshold, where the global threshold was viewed as more reliable. Global selection identified regions with large delta GPSeq score and showed regions in the genome that changed between differentiation stages. In the future, regions of interest, that represent the difference between the distinct differentiation stages (NES, NPC and neuron) should be further analysed. From this, it should be possible to see how the radial locations of chromosomes change during differentiation. This should then also be validated using DNA FISH. It would be interesting for example, to analyse subcompartments, to see if transcriptionally active parts of chromatin show preferential locations and how or if this changes between differentiation stages. Moreover, it might be interesting to analyse genes that are known to reposition upon differentiation, such as for example *Mash1* that has been reported to reposition in neural differentiation[107] (see section 5.8), in more detail. It could also be intriguing to investigate changes in histone modifications during neural differentiation.

Other experiments, that would be interesting to carry out would be to use NES derived from iPSCs from a diseased donor instead of the AF22 cell line, which comes from a healthy donor. Here it could be investigated, whether the radial organization is different in healthy and diseased patients and which regions of the genomes are affected by this.

A lot of data was collected during this thesis. The GPSeq Method was optimized for the used cell types and a new enzyme was introduced (DpnII), that had not been used before. Many new things can be discovered and further detailed analysis remains to be carried out.

9 Acknowledgements

I wish to express my deepest gratitude to Magda Bienko and Nicola Crosetto for giving me the amazing opportunity to join their group at Karolinska Institute. I am thankful for all the meetings we had, especially to Magda, who always had time to answer any spontaneous questions I had and even when in a rush, always found a minute to check images with me in the microscopy room. I really appreciate that you asked me for my opinion on how to proceed and often let me decide what to do, while at the same time giving me valuable advice and sharing her expertise with me. Thank you also for reading my thesis and giving me feedback on it.

Secondly, I would like to thank Robert Mach. Thank you for agreeing to supervise me and being the head of my diploma committee at the University of Technology in Vienna. I am grateful that you are taking the time for this and are always completing things promptly.

I would like to thank Tomasz Kallas. Tomek, thank you for always being there for me, for teaching me how to flip coverslips with grace and for going out of your way to help me. Thanks to you and your beautiful gel mixes I always got magnificent Bioanalyzer profiles, although it sometimes seemed like this instrument didn't want to make me happy. Thank you for helping me with sequencing all my samples, scraping coverslips and coming up with modern interpretations of old Christmas songs. When experiments didn't work, you never failed to make me smile and I am grateful I got to spend disappointment o'clock with you. It was the greatest joy to work with you and I hope in the future you will keep on singing an mermaiding.

I also wish to thank Sólrún Kolbeinsdóttir. Sólrún, when I arrived you were extremely helpful. You taught me all your tricks on how to do GPSeq and how to work in cell culture. You were always there for me, were super patient and I never felt like I was annoying you (although I probably sometimes was). Even after you did not have to help me anymore, you always checked in on me and helped me if I had made too ambitious plans that I couldn't do alone. I won't forget that you spent a whole afternoon imaging my samples and were there timing me when we came up with doing ridiculously short digestion times.

Moreover, I want to thank Gabrielle Girelli. Gabriele, thank you for analyzing my data, changing diagrams when I didn't like font sizes or colours and patiently explaining to

me how to calculate the GPSeq score for the tenth time. Without your analysis my experiments would be nothing. Thank you also for teaching me how to make pasta and for having great dinner parties.

I also want to thank Roberto Ballarino. Thank you, Roberto, for providing me with new material whenever I needed it. Thank you also for the discussions on how to best approach new situations and your willingness to adapt if things did not work out immediately.

Furthermore, I wish to thank Maud and Quim. Maud, I couldn't have wished for a more beautiful face to look at while pipetting. You brightened my incubation times and I wish you all the best being the CEO of Maud Fisher. Remember, no toxic people, only toxic by Britney Spears. Quim, thank you for not letting me die when climbing. You better come and visit my papaya farm in the future.

Thank you Patrick for being such a nice person to sit next to. I hope your new desk neighbour will understand Spongebob jokes.

I would like to thank all the other members of BiCro Lab who made my stay so enjoyable. I really had fun during my time in Stockholm and I appreciated that everyone welcomed me with open arms into the group.

Finally, I want to thank my family, especially my parents who have always supported me both mentally and financially. Thank you for listening to years of complaining and worrying and for proof-reading a lot of applications. Not everyone is so lucky to have parents who would support a decision to study for so long and not everyone has the same opportunities as me. I am a fortunate and privileged student.
Thank you Manuel, for always supporting everything I do. Thank you for always being so understanding, caring and loving. It sure was a challenge sometimes, when we were both writing our thesis at the same time and both were stressed. I am so glad we went on an adventure to Stockholm for half a year and I am the luckiest person to have you by my side. Now I am very much looking forward to starting a new chapter of life together.
Thank you to my best friends Anna, Clara and Anna Clara (I promise, I am not making this up) for distracting me from studying when I needed it and being by my side for over 14 years now (Clara 18 years!). You are the best friends anyone could ever wish for.

References

[1] G. Girelli, J. Custodio, T. Kallas, F. Agostini, E. Wernersson, S. Spanjaard, A. Mota, S. Kolbeinsdottir, E. Gelali, N. Crosetto, and M. Bienko, “Gpseq reveals the radial landscape of chromatin and genomic alterations in the human cell nucleus,” *Nature Biotechnology*, accepted 2020.

[2] P. J. Robinson, L. Fairall, V. A. Huynh, and D. Rhodes, “Em measurements define the dimensions of the 30-nm chromatin fiber: evidence for a compact, interdigitated structure,” *Proceedings of the National Academy of Sciences*, vol. 103, no. 17, pp. 6506–6511, 2006.

[3] C. L. Woodcock, “Chromatin architecture,” *Current opinion in structural biology*, vol. 16, no. 2, pp. 213–220, 2006.

[4] J. E. Sleeman and L. Trinkle-Mulcahy, “Nuclear bodies: new insights into assembly/dynamics and disease relevance,” *Current opinion in cell biology*, vol. 28, pp. 76–83, 2014.

[5] M. F. Smith, B. D. Athey, S. P. Williams, and J. P. Langmore, “Radial density distribution of chromatin: evidence that chromatin fibers have solid centers.,” *The Journal of cell biology*, vol. 110, no. 2, pp. 245–254, 1990.

[6] B. D. Athey, M. F. Smith, D. A. Rankert, S. P. Williams, and J. P. Langmore, “The diameters of frozen-hydrated chromatin fibers increase with dna linker length: evidence in support of variable diameter models for chromatin.,” *The Journal of cell biology*, vol. 111, no. 3, pp. 795–806, 1990.

[7] S. Williams, B. Athey, L. Muglia, R. Schappe, A. Gough, and J. Langmore, “Chromatin fibers are left-handed double helices with diameter and mass per unit length that depend on linker length,” *Biophysical journal*, vol. 49, no. 1, pp. 233–248, 1986.

[8] F. Thoma, T. Koller, and A. Klug, “Involvement of histone h1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin.,” *The Journal of cell biology*, vol. 83, no. 2, pp. 403–427, 1979.

[9] S. Gerchman and V. Ramakrishnan, “Chromatin higher-order structure studied by neutron scattering and scanning transmission electron microscopy,” *Proceedings of the National Academy of Sciences*, vol. 84, no. 22, pp. 7802–7806, 1987.

[10] V. Graziano, S. Gerchman, D. Schneider, and V. Ramakrishnan, “Histone h1 is located in the interior of the chromatin 30-nm filament,” *Nature*, vol. 368, no. 6469, pp. 351–354, 1994.

[11] S. P. Williams and J. P. Langmore, “Small angle x-ray scattering of chromatin. radius and mass per unit length depend on linker length.,” *Biophysical journal*, vol. 59, no. 3, p. 606, 1991.

[12] T. Richmond, J. Finch, B. Rushton, D. Rhodes, and A. Klug, “Structure of the nucleosome core particle at 7 å resolution,” *Nature*, vol. 311, no. 5986, pp. 532–537, 1984.

[13] K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond, “Crystal structure of the nucleosome core particle at 2.8 å resolution,” *Nature*, vol. 389, no. 6648, pp. 251–260, 1997.

[14] V. Ea, M.-O. Baudement, A. Lesne, and T. Forné, “Contribution of topological domains and loop formation to 3d chromatin organization,” *Genes*, vol. 6, no. 3, pp. 734–750, 2015.

[15] T. Igo-Kemenes, W. Hörz, and H. G. Zachau, “Chromatin,” *Annual review of biochemistry*, vol. 51, no. 1, pp. 89–121, 1982.

[16] L. L. Thompson, B. J. Guppy, L. Sawchuk, J. R. Davie, and K. J. McManus, “Regulation of chromatin structure via histone post-translational modification and the link to carcinogenesis,” *Cancer and Metastasis Reviews*, vol. 32, no. 3-4, pp. 363–376, 2013.

[17] B. Turner, L. Franchi, and H. Wallace, “Islands of acetylated histone h4 in polytene chromosomes and their relationship to chromatin packaging and transcriptional activity,” *Journal of cell science*, vol. 96, no. 2, pp. 335–346, 1990.

[18] A. M. Bode and Z. Dong, “Inducible covalent posttranslational modification of histone h3,” *Science Signaling*, vol. 2005, no. 281, pp. re4–re4, 2005.

[19] T. Cremer and C. Cremer, “Chromosome territories, nuclear architecture and gene regulation in mammalian cells,” *Nature reviews genetics*, vol. 2, no. 4, p. 292, 2001.

[20] M. N. Cornforth, K. M. Greulich-Bode, B. D. Loucas, J. Arsuaga, M. Vázquez, R. K. Sachs, M. Brückner, M. Molls, P. Hahnfeldt, L. Hlatky, *et al.*, “Chromosomes are predominantly located randomly with respect to each other in interphase human cells,” *The Journal of cell biology*, vol. 159, no. 2, pp. 237–244, 2002.

[21] J. Walter, L. Schermelleh, M. Cremer, S. Tashiro, and T. Cremer, “Chromosome order in hela cells changes during mitosis and early g1, but is stably maintained during subsequent interphase stages,” *The Journal of cell biology*, vol. 160, no. 5, pp. 685–697, 2003.

[22] A. Bolzer, G. Kreth, I. Solovei, D. Koehler, K. Saracoglu, C. Fauth, S. Müller, R. Eils, C. Cremer, M. R. Speicher, *et al.*, “Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes,” *PLoS biology*, vol. 3, no. 5, p. e157, 2005.

[23] R. Mayer, A. Brero, J. Von Hase, T. Schroeder, T. Cremer, and S. Dietzel, “Common themes and cell type specific variations of higher order chromatin arrangements in the mouse,” *BMC cell biology*, vol. 6, no. 1, p. 44, 2005.

[24] A. R. Barutcu, B. R. Lajoie, R. P. McCord, C. E. Tye, D. Hong, T. L. Messier, G. Browne, A. J. van Wijnen, J. B. Lian, J. L. Stein, *et al.*, “Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells,” *Genome biology*, vol. 16, no. 1, p. 214, 2015.

[25] L. A. Parada, P. G. McQueen, and T. Misteli, “Tissue-specific spatial organization of genomes,” *Genome biology*, vol. 5, no. 7, p. R44, 2004.

[26] I. Alcobia, A. S. Quina, H. Neves, N. Clode, and L. Parreira, “The spatial organization of centromeric heterochromatin during normal human lymphopoiesis: evidence for ontogenically determined spatial patterns,” *Experimental cell research*, vol. 290, no. 2, pp. 358–369, 2003.

[27] H. Tjong, W. Li, R. Kalhor, C. Dai, S. Hao, K. Gong, Y. Zhou, H. Li, X. J. Zhou, M. A. Le Gros, *et al.*, “Population-based 3d genome structure analysis reveals driving forces in spatial genome organization,” *Proceedings of the National Academy of Sciences*, vol. 113, no. 12, pp. E1663–E1672, 2016.

[28] H. Neves, C. Ramos, M. G. da Silva, A. Parreira, and L. Parreira, “The nuclear topography of abl, bcr, pml, and rar? genes: Evidence for gene proximity in specific phases of the cell cycle and stages of hematopoietic differentiation,” *Blood*, vol. 93, no. 4, pp. 1197–1207, 1999.

[29] M. N. Nikiforova, J. R. Stringer, R. Blough, M. Medvedovic, J. A. Fagin, and Y. E. Nikiforov, “Proximity of chromosomal loci that participate in radiation-induced rearrangements in human cells,” *Science*, vol. 290, no. 5489, pp. 138–141, 2000.

[30] J. J. Roix, P. G. McQueen, P. J. Munson, L. A. Parada, and T. Misteli, "Spatial proximity of translocation-prone gene loci in human lymphomas," *Nature genetics*, vol. 34, no. 3, p. 287, 2003.

[31] W. Fischle, Y. Wang, and C. D. Allis, "Histone and chromatin cross-talk," *Current opinion in cell biology*, vol. 15, no. 2, pp. 172–183, 2003.

[32] N. Gilbert, S. Boyle, H. Fiegler, K. Woodfine, N. P. Carter, and W. A. Bickmore, "Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers," *Cell*, vol. 118, no. 5, pp. 555–566, 2004.

[33] E. Heitz, *Das heterochromatin der moose*. Bornträger, 1928.

[34] N. Saksouk, E. Simboeck, and J. Déjardin, "Constitutive heterochromatin formation and transcription in mammals," *Epigenetics & chromatin*, vol. 8, no. 1, p. 3, 2015.

[35] P. Trojer and D. Reinberg, "Facultative heterochromatin: is there a distinctive molecular signature?," *Molecular cell*, vol. 28, no. 1, pp. 1–13, 2007.

[36] I. Solovei, A. S. Wang, K. Thanisch, C. S. Schmidt, S. Krebs, M. Zwerger, T. V. Cohen, D. Devys, R. Foisner, L. Peichl, *et al.*, "Lbr and lamin a/c sequentially tether peripheral heterochromatin and inversely regulate differentiation," *Cell*, vol. 152, no. 3, pp. 584–598, 2013.

[37] L. M. Armelin-Correa, L. M. Gutiyama, D. Y. Brandt, and B. Malnic, "Nuclear compartmentalization of odorant receptor genes," *Proceedings of the National Academy of Sciences*, vol. 111, no. 7, pp. 2782–2787, 2014.

[38] E. Lieberman-Aiden, N. L. Van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, *et al.*, "Comprehensive mapping of long-range interactions reveals folding principles of the human genome," *science*, vol. 326, no. 5950, pp. 289–293, 2009.

[39] J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren, "Topological domains in mammalian genomes identified by analysis of chromatin interactions," *Nature*, vol. 485, no. 7398, pp. 376–380, 2012.

[40] L. E. Finlan, D. Sproul, I. Thomson, S. Boyle, E. Kerr, P. Perry, B. Ylstra, J. R. Chubb, and W. A. Bickmore, "Recruitment to the nuclear periphery can alter expression of genes in human cells," *PLoS genetics*, vol. 4, no. 3, p. e1000039, 2008.

[41] K. L. Reddy, J. Zullo, E. Bertolino, and H. Singh, “Transcriptional repression mediated by repositioning of genes to the nuclear lamina,” *Nature*, vol. 452, no. 7184, p. 243, 2008.

[42] L. Guelen, L. Pagie, E. Brasset, W. Meuleman, M. B. Faza, W. Talhout, B. H. Eussen, A. de Klein, L. Wessels, W. de Laat, *et al.*, “Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions,” *Nature*, vol. 453, no. 7197, pp. 948–951, 2008.

[43] D. Peric-Hupkes, W. Meuleman, L. Pagie, S. W. Bruggeman, I. Solovei, W. Brugman, S. Gräf, P. Flicek, R. M. Kerkhoven, M. van Lohuizen, *et al.*, “Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation,” *Molecular cell*, vol. 38, no. 4, pp. 603–613, 2010.

[44] J. Kind, L. Pagie, S. S. de Vries, L. Nahidiazar, S. S. Dey, M. Bienko, Y. Zhan, B. Lajoie, C. A. de Graaf, M. Amendola, *et al.*, “Genome-wide maps of nuclear lamina interactions in single human cells,” *Cell*, vol. 163, no. 1, pp. 134–147, 2015.

[45] S. A. Quinodoz, N. Ollikainen, B. Tabak, A. Palla, J. M. Schmidt, E. Detmar, M. M. Lai, A. A. Shishkin, P. Bhat, Y. Takei, *et al.*, “Higher-order interchromosomal hubs shape 3d genome organization in the nucleus,” *Cell*, vol. 174, no. 3, pp. 744–757, 2018.

[46] B. Van Steensel and A. S. Belmont, “Lamina-associated domains: links with chromosome architecture, heterochromatin, and gene repression,” *Cell*, vol. 169, no. 5, pp. 780–791, 2017.

[47] T. R. Luperchio, M. E. Sauria, V. E. Hoskins, W. Xianrong, E. DeBoy, M.-C. Gaillard, P. Tsang, K. Pekrun, R. A. Ach, A. Yamada, *et al.*, “The repressive genome compartment is established early in the cell cycle before forming the lamina associated domains,” *bioRxiv*, p. 481598, 2018.

[48] I. Alcobia, R. Dilao, and L. Parreira, “Spatial associations of centromeres in the nuclei of hematopoietic cells: evidence for cell-type-specific organizational patterns,” *Blood, The Journal of the American Society of Hematology*, vol. 95, no. 5, pp. 1608–1615, 2000.

[49] C. Molenaar, K. Wiesmeijer, N. P. Verwoerd, S. Khazen, R. Eils, H. J. Tanke, and R. W. Dirks, “Visualizing telomere dynamics in living mammalian cells using pna probes,” *The EMBO journal*, vol. 22, no. 24, pp. 6631–6641, 2003.

TU Bibliothek WIEN Your knowledge hub
Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.
The approved original version of this thesis is available in print at TU Wien Bibliothek.

[50] C. Weierich, A. Brero, S. Stein, J. von Hase, C. Cremer, T. Cremer, and I. Solovei, "Three-dimensional arrangements of centromeres and telomeres in nuclei of human and murine lymphocytes," *Chromosome research*, vol. 11, no. 5, pp. 485–502, 2003.

[51] A. Németh and G. Längst, "Genome organization in and around the nucleolus," *Trends in genetics*, vol. 27, no. 4, pp. 149–156, 2011.

[52] J. Dekker, K. Rippe, M. Dekker, and N. Kleckner, "Capturing chromosome conformation," *science*, vol. 295, no. 5558, pp. 1306–1311, 2002.

[53] M. Simonis, P. Klous, E. Splinter, Y. Moshkin, R. Willemsen, E. De Wit, B. Van Steensel, and W. De Laat, "Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4c)," *Nature genetics*, vol. 38, no. 11, pp. 1348–1354, 2006.

[54] S. S. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, *et al.*, "A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping," *Cell*, vol. 159, no. 7, pp. 1665–1680, 2014.

[55] J. Dostie, T. A. Richmond, R. A. Arnaout, R. R. Selzer, W. L. Lee, T. A. Honan, E. D. Rubio, A. Krumm, J. Lamb, C. Nusbaum, *et al.*, "Chromosome conformation capture carbon copy (5c): a massively parallel solution for mapping interactions between genomic elements," *Genome research*, vol. 16, no. 10, pp. 1299–1309, 2006.

[56] A. Losada, "Cohesin in cancer: chromosome segregation and beyond," *Nature Reviews Cancer*, vol. 14, no. 6, pp. 389–393, 2014.

[57] T. Nagano, Y. Lubling, T. J. Stevens, S. Schoenfelder, E. Yaffe, W. Dean, E. D. Laue, A. Tanay, and P. Fraser, "Single-cell hi-c reveals cell-to-cell variability in chromosome structure," *Nature*, vol. 502, no. 7469, pp. 59–64, 2013.

[58] L. L. Hanssen, M. T. Kassouf, A. M. Oudelaar, D. Biggs, C. Preece, D. J. Downes, M. Gosden, J. A. Sharpe, J. A. Sloane-Stanley, J. R. Hughes, *et al.*, "Tissue-specific ctcf–cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo," *Nature cell biology*, vol. 19, no. 8, pp. 952–961, 2017.

[59] S. Berlivet, D. Paquette, A. Dumouchel, D. Langlais, J. Dostie, and M. Kmita, "Clustering of tissue-specific sub-tads accompanies the regulation of hoxa genes in developing limbs," *PLoS genetics*, vol. 9, no. 12, 2013.

[60] T. Misteli, “Beyond the sequence: cellular organization of genome function,” *Cell*, vol. 128, no. 4, pp. 787–800, 2007.

[61] J. H. Gibcus and J. Dekker, “The hierarchy of the 3d genome,” *Molecular Cell*, vol. 49, no. 5, pp. 773–782, 2013.

[62] J. Hoo and H. Cramer, “On the position of chromosomes in prepared mitosis figures of human fibroblasts,” *Humangenetik*, vol. 13, no. 2, pp. 166–170, 1971.

[63] H. B. Sun, J. Shen, and H. Yokota, “Size-dependent positioning of human chromosomes in interphase nuclei,” *Biophysical journal*, vol. 79, no. 1, pp. 184–190, 2000.

[64] O. Miller, B. Mukherjee, W. Breg, and A. V. N. Gamble, “Non-random distribution of chromosomes in metaphase figures from cultured human leucocytes i. the peripheral location of the y chromosome,” *Cytogenetic and Genome Research*, vol. 2, no. 1, p. 1?14, 1963.

[65] H. Tanabe, F. A. Habermann, I. Solovei, M. Cremer, and T. Cremer, “Non-random radial arrangements of interphase chromosome territories: evolutionary considerations and functional implications,” *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, vol. 504, no. 1-2, pp. 37–45, 2002.

[66] J. A. Croft, J. M. Bridger, S. Boyle, P. Perry, P. Teague, and W. A. Bickmore, “Differences in the localization and morphology of chromosomes in the human nucleus,” *The Journal of cell biology*, vol. 145, no. 6, pp. 1119–1131, 1999.

[67] A. E. Wiblin, W. Cui, A. Clark, and B. WA, “Distinctive nuclear organisation of centromeres and regions involved in pluripotency in human embryonic stem cells,” *Journal of Cell Science*, vol. 118, p. 3861?3868, Jan 2005.

[68] S. Boyle, S. Gilchrist, J. M. Bridger, N. L. Mahy, J. A. Ellis, and W. A. Bickmore, “The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells,” *Human molecular genetics*, vol. 10, no. 3, pp. 211–220, 2001.

[69] M. Cremer, J. Von Hase, T. Volm, A. Brero, G. Kreth, J. Walter, C. Fischer, I. Solovei, C. Cremer, and T. Cremer, “Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells,” *Chromosome research*, vol. 9, no. 7, pp. 541–567, 2001.

[70] M. Cremer, K. Küpper, B. Wagler, L. Wizelman, J. v. Hase, Y. Weiland, L. Kreja, J. Diebold, M. R. Speicher, and T. Cremer, “Inheritance of gene density–related

higher order chromatin arrangements in normal and tumor cell nuclei," *The Journal of cell biology*, vol. 162, no. 5, pp. 809–820, 2003.

[71] M. R. Branco and A. Pombo, "Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations," *PLoS biology*, vol. 4, no. 5, p. e138, 2006.

[72] T. Misteli, "Higher-order genome organization in human disease," *Cold Spring Harbor perspectives in biology*, vol. 2, no. 8, p. a000794, 2010.

[73] H. Tanabe, S. Müller, M. Neusser, J. von Hase, E. Calcagno, M. Cremer, I. Solovei, C. Cremer, and T. Cremer, "Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates," *Proceedings of the National Academy of Sciences*, vol. 99, no. 7, pp. 4424–4429, 2002.

[74] J. Bridger, S. Boyle, I. Kill, and W. Bickmore, "Re-modelling of nuclear architecture in quiescent and senescent human fibroblasts," *Current Biology*, vol. 10, no. 3, pp. 149–152, 2000.

[75] C. Federico, C. Scavo, C. D. Cantarella, S. Motta, S. Saccone, and G. Bernardi, "Gene-rich and gene-poor chromosomal regions have different locations in the interphase nuclei of cold-blooded vertebrates," *Chromosoma*, vol. 115, no. 2, pp. 123–128, 2006.

[76] F. Grasser, M. Neusser, H. Fiegler, T. Thormeyer, M. Cremer, N. P. Carter, T. Cremer, and S. Müller, "Replication-timing-correlated spatial chromatin arrangements in cancer and in primate interphase nuclei," *Journal of cell science*, vol. 121, no. 11, pp. 1876–1886, 2008.

[77] C. Hepperger, A. Mannes, J. Merz, J. Peters, and S. Dietzel, "Three-dimensional positioning of genes in mouse cell nuclei," *Chromosoma*, vol. 117, no. 6, p. 535, 2008.

[78] G. Kreth, J. Finsterle, J. Von Hase, M. Cremer, and C. Cremer, "Radial arrangement of chromosome territories in human cell nuclei: a computer model approach based on gene density indicates a probabilistic global positioning code," *Biophysical journal*, vol. 86, no. 5, pp. 2803–2812, 2004.

[79] C. S. Osborne, L. Chakalova, K. E. Brown, D. Carter, A. Horton, E. Debrand, B. Goyenechea, J. A. Mitchell, S. Lopes, W. Reik, *et al.*, "Active genes dynamically colocalize to shared sites of ongoing transcription," *Nature genetics*, vol. 36, no. 10, p. 1065, 2004.

[80] I. S. Mehta, M. Amira, A. J. Harvey, and J. M. Bridger, “Rapid chromosome territory relocation by nuclear motor activity in response to serum removal in primary human fibroblasts,” *Genome biology*, vol. 11, no. 1, p. R5, 2010.

[81] I. S. Mehta, M. Kulashreshtha, S. Chakraborty, U. Kolthur-Seetharam, and B. J. Rao, “Chromosome territories reposition during dna damage-repair response,” *Genome biology*, vol. 14, no. 12, p. R135, 2013.

[82] C. J. Nelson, H. Santos-Rosa, and T. Kouzarides, “Proline isomerization of histone h3 regulates lysine methylation and gene expression,” *Cell*, vol. 126, no. 5, pp. 905–916, 2006.

[83] S. L. Berger, “The complex language of chromatin regulation during transcription,” *Nature*, vol. 447, no. 7143, pp. 407–412, 2007.

[84] M. P. van de Corput, E. de Boer, T. A. Knoch, W. A. van Cappellen, A. Quintanilla, L. Ferrand, and F. G. Grosveld, “Super-resolution imaging reveals three-dimensional folding dynamics of the β-globin locus upon gene activation,” *Journal of cell science*, vol. 125, no. 19, pp. 4630–4639, 2012.

[85] Y. Wang, S. Maharana, M. D. Wang, and G. Shivashankar, “Super-resolution microscopy reveals decondensed chromatin structure at transcription sites,” *Scientific reports*, vol. 4, no. 1, pp. 1–7, 2014.

[86] Y. Wang, M. Nagarajan, C. Uhler, and G. Shivashankar, “Orientation and repositioning of chromosomes correlate with cell geometry–dependent gene expression,” *Molecular biology of the cell*, vol. 28, no. 14, pp. 1997–2009, 2017.

[87] M. H. Kagey, J. J. Newman, S. Bilodeau, Y. Zhan, D. A. Orlando, N. L. van Berkum, C. C. Ebmeier, J. Goossens, P. B. Rahl, S. S. Levine, *et al.*, “Mediator and cohesin connect gene expression and chromatin architecture,” *Nature*, vol. 467, no. 7314, pp. 430–435, 2010.

[88] C. Vourc’h, D. Taruscio, A. L. Boyle, and D. C. Ward, “Cell cycle-dependent distribution of telomeres, centromeres, and chromosome-specific subsatellite domains in the interphase nucleus of mouse lymphocytes,” *Experimental cell research*, vol. 205, no. 1, pp. 142–151, 1993.

[89] S. Gilchrist, N. Gilbert, P. Perry, and W. A. Bickmore, “Nuclear organization of centromeric domains is not perturbed by inhibition of histone deacetylases,” *Chromosome research*, vol. 12, no. 5, pp. 505–516, 2004.

[90] M. Ferguson and D. C. Ward, “Cell cycle dependent chromosomal movement in pre-mitotic human t-lymphocyte nuclei,” *Chromosoma*, vol. 101, no. 9, pp. 557–565, 1992.

[91] I. Thomson, S. Gilchrist, W. A. Bickmore, and J. R. Chubb, “The radial positioning of chromatin is not inherited through mitosis but is established de novo in early g1,” *Current biology*, vol. 14, no. 2, pp. 166–172, 2004.

[92] E. Yaffe and A. Tanay, “Probabilistic modeling of hi-c contact maps eliminates systematic biases to characterize global chromosomal architecture,” *Nature genetics*, vol. 43, no. 11, p. 1059, 2011.

[93] R. Fang, M. Yu, G. Li, S. Chee, T. Liu, A. D. Schmitt, and B. Ren, “Mapping of long-range chromatin interactions by proximity ligation-assisted chip-seq,” *Cell research*, vol. 26, no. 12, pp. 1345–1348, 2016.

[94] R.-J. Palstra, B. Tolhuis, E. Splinter, R. Nijmeijer, F. Grosveld, and W. de Laat, “The β-globin nuclear compartment in development and erythroid differentiation,” *Nature genetics*, vol. 35, no. 2, pp. 190–194, 2003.

[95] R. Drissen, R.-J. Palstra, N. Gillemans, E. Splinter, F. Grosveld, S. Philipsen, and W. de Laat, “The active spatial organization of the β-globin locus requires the transcription factor eklf,” *Genes & development*, vol. 18, no. 20, pp. 2485–2490, 2004.

[96] E. de Wit, E. S. Vos, S. J. Holwerda, C. Valdes-Quezada, M. J. Verstegen, H. Teunissen, E. Splinter, P. J. Wijchers, P. H. Krijger, and W. de Laat, “Ctcf binding polarity determines chromatin looping,” *Molecular cell*, vol. 60, no. 4, pp. 676–684, 2015.

[97] J. E. Phillips and V. G. Corces, “Ctcf: master weaver of the genome,” *Cell*, vol. 137, no. 7, pp. 1194–1211, 2009.

[98] S. Sofueva, E. Yaffe, W.-C. Chan, D. Georgopoulou, M. V. Rudan, H. Mira-Bontenbal, S. M. Pollard, G. P. Schroth, A. Tanay, and S. Hadjur, “Cohesin-mediated interactions organize chromosomal domain architecture,” *The EMBO journal*, vol. 32, no. 24, pp. 3119–3129, 2013.

[99] S. S. Rao, S.-C. Huang, B. G. St Hilaire, J. M. Engreitz, E. M. Perez, K.-R. Kieffer-Kwon, A. L. Sanborn, S. E. Johnstone, G. D. Bascom, I. D. Bochkov, *et al.*, “Cohesin loss eliminates all loop domains,” *Cell*, vol. 171, no. 2, pp. 305–320, 2017.

[100] B. A. Bouwman and W. de Laat, “Getting the genome in shape: the formation of loops, domains and compartments,” *Genome biology*, vol. 16, no. 1, p. 154, 2015.

[101] E. Meshorer and T. Misteli, “Chromatin in pluripotent embryonic stem cells and differentiation,” *Nature reviews Molecular cell biology*, vol. 7, no. 7, pp. 540–546, 2006.

[102] J. Fraser, C. Ferrai, A. M. Chiariello, M. Schueler, T. Rito, G. Laudanno, M. Barbieri, B. L. Moore, D. C. Kraemer, S. Aitken, *et al.*, “Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation,” *Molecular systems biology*, vol. 11, no. 12, 2015.

[103] Y. J. Kim, K. R. Cecchini, and T. H. Kim, “Conserved, developmentally regulated mechanism couples chromosomal looping and heterochromatin barrier activity at the homeobox gene a locus,” *Proceedings of the National Academy of Sciences*, vol. 108, no. 18, pp. 7391–7396, 2011.

[104] J.-H. Lee, S. R. Hart, and D. G. Skalnik, “Histone deacetylase activity is required for embryonic stem cell differentiation,” *genesis*, vol. 38, no. 1, pp. 32–38, 2004.

[105] E. J. Kim, J. L. Ables, L. K. Dickel, A. J. Eisch, and J. E. Johnson, “Ascl1 (mash1) defines cells with long-term neurogenic potential in subgranular and subventricular zones in adult mouse brain,” *PloS one*, vol. 6, no. 3, 2011.

[106] T. Kobayashi and R. Kageyama, “Expression dynamics and functions of hes factors in development and diseases,” in *Current topics in developmental biology*, vol. 110, pp. 263–283, Elsevier, 2014.

[107] R. R. Williams, V. Azuara, P. Perry, S. Sauer, M. Dvorkina, H. Jørgensen, J. Roix, P. McQueen, T. Misteli, M. Merkenschlager, *et al.*, “Neural induction promotes large-scale chromatin reorganisation of the mash1 locus,” *Journal of cell science*, vol. 119, no. 1, pp. 132–140, 2006.

[108] B. E. Bernstein, T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, *et al.*, “A bivalent chromatin structure marks key developmental genes in embryonic stem cells,” *Cell*, vol. 125, no. 2, pp. 315–326, 2006.

[109] V. Azuara, P. Perry, S. Sauer, M. Spivakov, H. F. Jørgensen, R. M. John, M. Gouti, M. Casanova, G. Warnes, M. Merkenschlager, *et al.*, “Chromatin signatures of pluripotent cell lines,” *Nature cell biology*, vol. 8, no. 5, pp. 532–538, 2006.

[110] E. Meshorer, “Chromatin in embryonic stem cell neuronal differentiation,” *Histology and histopathology*, 2007.

[111] V. Domínguez, B. Piña, and P. Suau, “Histone h1 subtype synthesis in neurons and neuroblasts,” *Development*, vol. 115, no. 1, pp. 181–185, 1992.

[112] A. Bosch and P. Suau, “Changes in core histone variant composition in differentiating neurons: the roles of differential turnover and synthesis rates.,” *European journal of cell biology*, vol. 68, no. 3, pp. 220–225, 1995.

[113] B. Bonev, N. M. Cohen, Q. Szabo, L. Fritsch, G. L. Papadopoulos, Y. Lubling, X. Xu, X. Lv, J.-P. Hugnot, A. Tanay, *et al.*, “Multiscale 3d genome rewiring during mouse neural development,” *Cell*, vol. 171, no. 3, pp. 557–572, 2017.

[114] O. Lindvall and Z. Kokaia, “Stem cells in human neurodegenerative disorders?time for clinical translation?,” *The Journal of clinical investigation*, vol. 120, no. 1, pp. 29–40, 2010.

[115] K. Takahashi and S. Yamanaka, “Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors,” *cell*, vol. 126, no. 4, pp. 663–676, 2006.

[116] A. Falk, P. Koch, J. Kesavan, Y. Takashima, J. Ladewig, M. Alexander, O. Wiskow, J. Tailor, M. Trotter, S. Pollard, *et al.*, “Capture of neuroepithelial-like stem cells from pluripotent stem cells provides a versatile system for in vitro production of human neurons,” *PloS one*, vol. 7, no. 1, 2012.

[117] V. B. Mattis and C. N. Svendsen, “Induced pluripotent stem cells: a new revolution for clinical neurology?,” *The Lancet Neurology*, vol. 10, no. 4, pp. 383–394, 2011.

[118] S. Boyle, M. J. Rodesch, H. A. Halvensleben, J. A. Jeddeloh, and W. A. Bickmore, “Fluorescence in situ hybridization with high-complexity repeat-free oligonucleotide probes generated by massively parallel synthesis,” *Chromosome research*, vol. 19, no. 7, pp. 901–909, 2011.

[119] B. J. Beliveau, E. F. Joyce, N. Apostolopoulos, F. Yilmaz, C. Y. Fonseka, R. B. McCole, Y. Chang, J. B. Li, T. N. Senaratne, B. R. Williams, *et al.*, “Versatile design and synthesis platform for visualizing genomes with oligopaint fish probes,” *Proceedings of the National Academy of Sciences*, vol. 109, no. 52, pp. 21301–21306, 2012.

[120] M. Bienko, N. Crosetto, L. Teytelman, S. Klemm, S. Itzkovitz, and A. Van Oudenaarden, "A versatile genome-scale pcr-based pipeline for high-definition dna fish," *Nature Methods*, vol. 10, no. 2, p. 122, 2013.

[121] M. R. Speicher and N. P. Carter, "The new cytogenetics: blurring the boundaries with molecular biology," *Nature reviews genetics*, vol. 6, no. 10, p. 782, 2005.

[122] E. Gelali, J. Custodio, G. Girelli, E. Wernersson, N. Crosetto, and M. Bienko, *An Application-Directed, Versatile DNA FISH Platform for Research and Diagnostics*. New York, NY: Springer New York, 2018.

[123] C. O'Connor, "Fluorescence in situ hybridization (fish)," *Nature Education*, vol. 1, no. 1, p. 171, 2008.

[124] M. L. Metzker, "Sequencing technologies?the next generation," *Nature reviews genetics*, vol. 11, no. 1, pp. 31–46, 2010.

[125] D. R. Bentley, S. Balasubramanian, H. P. Swerdlow, G. P. Smith, J. Milton, C. G. Brown, K. P. Hall, D. J. Evers, C. L. Barnes, H. R. Bignell, *et al.*, "Accurate whole human genome sequencing using reversible terminator chemistry," *nature*, vol. 456, no. 7218, pp. 53–59, 2008.

[126] N. E. Biolabs, "Dam-dcm and cpg methylation."

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

Appendices

A Full GPSeq Protocol

Step-by-step YFISH and GPSeq protocol

Reagents

- Poly-L-Lysine (PLL) solution (Sigma, cat. no. P8920-100 ml)
- Methanol-free paraformaldehyde (PFA) 16% (EMS, cat. no. 15710)
- Glycine (Sigma-Aldrich, cat. no. 50046-250G)
- Nuclease-free Phosphate-Buffered Saline (10X) pH 7.4 (Thermo Fisher Scientific, cat. no. AM9625)
- Nuclease-free water (Thermo Fisher Scientific, cat. no. 4387936) (for *in situ* reactions and library preparation)
- CutSmart® buffer (NEB, cat. no. B7204S)
- HindIII-HF® (NEB, cat. no. R3104)
- DpnII (NEB, cat. no. R0543M)
- Calf intestinal alkaline phosphatase (Promega, cat. no. M1821)
- T4 DNA Ligase (NEB, cat. no. M0202M)
- UltraPure™ BSA (50 mg/ml) (Thermo Fisher Scientific, cat. no. AM2616)
- Triton® X-100 (Promega, cat.no. H5142)
- High-salt wash buffer (HSW): 10 mM Tris-HCl, 1M NaCl, 0.5% Triton X-100, pH 8 at 25 °C
- Dextran sulfate (Sigma, cat. no. D8906)
- Formamide (FA) (Ambion, cat. no. AM9342)
- *E.coli* tRNA (Sigma, cat. no. R1753)
- BSA (Ambion, cat. no. AM2616)
- YFISH wash buffer (YWB): 25% FA/2X SSC
- YFISH hybridization buffer (YHB): 10% Dextran sulfate/25% FA/1 mg/ml E.coli tRNA /0.02 % BSA/2X SSC
- Hoechst 33342 (Thermo Fisher Scientific, cat. no. H3570)
- DNA extraction buffer (DEB): 1% SDS, 100 mM NaCl, 50 mM EDTA, 10 mM Tris-HCl, pH 7.5 at 25 °C
- Proteinase K, Molecular Biology Grade (NEB, cat. no. P8107S)

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

- Phenol:Chloroform:Isoamyl Alcohol 25:24:1 Saturated with 10 mM Tris, pH 8.0, 1 mM (Sigma-Aldrich, cat. no. P2069)
- Chloroform (Sigma-Aldrich, cat. no. C2432)
- EDTA (Sigma-Aldrich, cat. no. P2069)
- Glycogen (Sigma-Aldrich, cat. no. 10901393001)
- Nuclease-free TE buffer (Thermo Fisher Scientific, cat. no. AM9849)
- AMPure XP (Beckman Coulter, cat. no. A63881)
- MEGAscript® T7 Transcription Kit (Thermo Fisher Scientific, cat. no. AM1334)
- RiboSafe RNase Inhibitor (Bioline, cat. no. 65027)
- DNase I, RNase-free (Thermo Fisher Scientific, cat. no. AM2222)
- RA3 adaptor and RTP, RP1 and RPI primers are the same as in the TruSeq Small RNA Library Preparation kit (Illumina, cat. no. RS-200-0012), but are custom-synthesized and purified using standard desalting
- RNAClean XP (Beckman Coulter, cat. no. A63987)
- RNaseOUT^{TM} Recombinant Ribonuclease Inhibitor (Invitrogen, cat. no. 10777-019)
- T4 RNA ligase 2, truncated (NEB, cat. no. M0242L)
- Deoxynucleotide (dNTP) Solution Set (NEB, cat. no. N0446S)
- SuperScript® III Reverse Transcriptase (Thermo Fisher Scientific, cat. no. 18080044)
- NEBNext® High-Fidelity 2X PCR Master Mix (NEB, cat. no. M0541L)

Equipment

- Incubator (e.g., Binder incubator, Model KB 53)
- Tabletop centrifuge (e.g., Eppendorf® Microcentrifuge 5424)
- Thermoshaker (e.g., Eppendorf® Thermomixer Compact)
- PCR cycler (e.g., T3 Thermocycler, Biometra)
- Cell scrapers (Corning, Cell Lifter, cat. no. 3008)
- Eppendorf® DNA LoBind microcentrifuge tubes (cat. no. 0030108035 and 0030108051)
- Sonication device (Bioruptor® Plus, Diagenode, cat. no. B01020001)
- DynaMag^{TM}-2 Magnet (Thermo Fisher Scientific, cat. no. 12321D)
- Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, cat. no. Q32866)
- Qubit^{TM} dsDNA HS Assay (Thermo Fisher Scientific, cat. no. Q32851)
- Qubit^{TM} Assay Tubes (Thermo Fisher Scientific, cat. no. Q32856)

- Bioanalyzer 2100 (Agilent, cat. no. G2943CA)
- Agilent RNA 6000 Pico Kit (Agilent, cat. no. 5067-1513)
- Agilent High Sensitivity DNA Kit (Agilent, cat. no. 5067-4626)

Procedure

Cell preparation

Note: a separate set of samples should be prepared for visualization of gradual genome restriction by YFISH and for sequencing. The two sets of samples should be processed in parallel.

1. Wash the cells with 1X PBS containing Ca^{2+} and Mg^{2+} pre-warmed to 37 °C
2. Incubate for 2 min at room temperature (RT)
3. Aspirate the PBS solution from the dish
4. Add 0.4X PBS with Ca^{2+} and Mg^{2+} at RT
5. Incubate for 1 min at RT
6. Aspirate the PBS solution
7. Add 4 % PFA/0.4X PBS at RT
8. Incubate for 10 min at RT
9. Aspirate the fixative solution from the dish
10. Quench the residual PFA by adding 125 mM glycine/1X PBS
11. Incubate for 5 min at RT
12. Briefly wash the cells in 1X PBS at RT
13. Wash the cells in 0.05 % Triton X-100/1X PBS three times, 5 min each at RT, while shaking
14. Aspirate the PBS solution and then add 0.5 % Triton X-100/1X PBS at RT
15. Incubate for 20 min at RT
16. Briefly wash the cells in 1X PBS at RT
17. Briefly wash the cells in 20 % glycerol/1X PBS at RT
18. Incubate the samples in 20 % glycerol/1X PBS for 18 h at RT
19. Fill a cryo-resistant Dewar flask with liquid N_2
20. Remove the glycerol solution from the dish (leave a thin layer on top of the cells)
21. Place the dish on the surface of liquid N_2 for 20-30 sec
22. Place the dish at RT and let the samples thaw gradually, making sure that they do not dry

23. Once all the samples are thawed, fill the dish with fresh 20 % glycerol/1X PBS at RT
24. Incubate for 3 min at RT
25. Repeat steps 20-23 four times
26. Wash the samples in 0.05 % Triton X-100/1X PBS two times, 5 min each at RT, while shaking
27. Briefly rinse the samples in 0.1 M HCl at RT
28. Incubate the samples in 0.1 M HCl for 5 min at RT Note: the samples should be kept in HCl for a maximum of 5 min including the quick rinsing step below
29. Briefly rinse the samples in 0.05 % Triton X-100/1X PBS at RT
30. Wash the samples in 0.05 % Triton X-100/1X PBS two times, 5 min each at RT, while shaking
31. Wash the samples for 5 min in 2X SSC
32. Replace the solution to fresh 2X SSC and process the samples immediately or alternatively store them in 2X SSC/0.05 % NaN_3 for up to 1 month at 4 °C

In situ digestion

1. Bring the 50 % FA/50 mM phosphate buffer/2X SSC (FPS) solution to RT
2. Rinse the samples with FPS at RT
3. Exchange to fresh FPS and store at RT protected from light for 20 h
4. Wash the samples in 2X SSC for 5 min at RT
5. Wash the samples in 1X CutSmart, two times, 5 min each at RT, while gently rocking
6. Prepare the following digestion mix (volumes for one 22x22 mm coverslip):

HindIII-HF (20,000 U/ml)	10 µl
CutSmart (10X)	40 µl
Nuclease-free water	Up to 400 µl

Alternatively:

DpnII (50,000 U/ml)	4 µl
NEB Buffer (10X)	40 µl
Nuclease-free water	Up to 400 µl

Note: The choice of enzyme depends on the desired resolution and affordable sequencing depth. As a rule of thumb, HindIII can be used to obtain highly reproducible radiality maps at 1 Mb resolution using five different digestion durations and one sequencing run on NextSeq 500. DpnII is preferable when a higher resolution is needed.

7. Pre-warm the digestion mix for at least 10 min at 37 °C

8. Dispense the digestion mix on a piece of Parafilm placed inside a humidity chamber Note: we make the chamber by covering the bottom and the edges of a 15 cm dish with tissue paper pre-soaked in distilled water, and then laying a piece of parafilm on top of the soaked tissue on the dish bottom
9. Gently flip one sample at a time, upside-down onto the droplet of digestion mix
10. Incubate the sample at 37 °C in the sealed humidity chamber, for a pre-set duration Note: each sample is incubated in the presence of the restriction enzyme for a different duration. NES & NPC cells: for HindIII, we have used 5, 10, 15, 30 min, 1 and 2 h, whereas, for DpnII, we have used 10, 20, 30, 45 sec, 2, 5, 10 and 30 min, which were sufficient to produce reproducible radiality maps. Neurons: for DpnII, we have used 10, 20, 30, 45 sec, 2 and 5 min. For different cell types, it might be necessary to optimize the number and series of incubation times, by first monitoring the enzyme diffusion by YFISH (see below)
11. Stop the digestion by washing the samples with ice-cold 50 mM EDTA/0.01 % Triton X-100/1X PBS three times, 5 min each, on ice
12. Rinse the samples with 0.01 % Triton X-100/1X PBS at RT
13. Wash with 0.01 % Triton X-100/1X PBS for 5 min at RT

In situ dephosphorylation

1. Wash the samples in 1X Alkaline phosphatase (AP) buffer for 5 min at RT
2. Prepare the following dephosphorylation mix (volumes for one 22x22 mm coverslip):

AP buffer (10X)	40 µl
AP (1 U/µl)	6 µl
Nuclease-free water	354 µl

3. Dispense the dephosphorylation mix on a piece of Parafilm placed inside a humidity chamber
4. Gently flip one sample at a time, upside-down onto the droplet of dephosphorylation mix
5. Incubate the samples for 2 h at 37 °C in the sealed humidity chamber
6. Transfer the samples back to a 6-well plate
7. Wash the samples with 0.1 % Triton X-100/1X PBS two times, 5 min each at RT

In situ ligation

1. Wash the samples in 1X T4 Ligase buffer two times, 5 min each at RT, while gently rocking
2. Prepare the following ligation mix I (volumes for one 22x22 mm coverslip):

Ligase buffer (10X)	21 µl
ATP (10 mM)	24 µl
BSA (5 mg/ml)	6 µl
HindIII GPSeq adapter (10 µM)	6 µl
Alternatively: DpnII GPSeq adapter (10 µM)	18 µl
Nuclease-free water	Up to 210 µl

Note: thaw all reagents slowly on ice or at 4 °C

3. Dispense the ligation mix I on a piece of Parafilm placed inside a humidity chamber
4. Gently flip one sample at a time, upside-down onto the droplet of ligation mix I
5. Incubate the samples for 30 min at RT
6. Prepare the following ligation mix II (volumes for one 22x22 mm coverslip):

Nuclease-free water	45 µl
Ligase buffer (10X)	9 µl
T4 Ligase enzyme	36 µl

7. Gently pipette the ligase mix II under each coverslip in and out 7-8 times, making sure not to make bubbles
8. Incubate the samples for 18 h at 16 °C in the sealed humidity chamber
9. In a new 6-well plate, wash the samples in 1X HSW five times, 1 hour each at 37 °C, while shaking
10. Wash the samples in 1X PBS three times, 5 min each at RT

Breakpoint: at this point, the samples for sequencing can be stored in 1X PBS at 4 °C for several days until the results of YFISH (next section) are available

YFISH

Note: we recommend to check that the genome has been gradually digested by the enzyme in use, from the nuclear periphery towards the nucleus interior, before proceeding to sequencing

1. Equilibrate the samples in 25 % FA/2X SSC
2. Prepare the following hybridization mix (volumes for one 22x22 mm coverslip):

YHB	297 µl
Fluorescently labelled oligonucleotide (2 µM)	3 µl

3. Dispense the hybridization mix on a piece of Parafilm placed inside a humidity chamber
4. Gently flip one sample at a time, upside-down onto the droplet of hybridization mix
5. Incubate the samples for 18 h at 30 °C inside the sealed humidity chamber Note: the samples are now light sensitive, therefore all the subsequent steps should be performed by protecting them from light
6. Transfer the samples into a 6-well plate pre-filled with 2 ml/well of YWB pre-warmed at RT
7. Exchange the YWB solution and incubate for 1 h at 30 °C, while gently shaking
8. Remove the YWB solution and add 2 ml/well of 0.1 ng/µl Hoechst 33342/YWB
9. Incubate for 30 min at 30 °C, while gently shaking

10. Quickly rinse the samples with 2X SSC
11. Mount the samples and image them using wield-field microscopy or, if available, STED microscopy

Library preparation

Note: for each of the restriction times used, an individual sequencing library is prepared

DNA Extraction

1. Wash the samples with nuclease-free water three times, 2 min each at RT
2. Scrape each coverslip in the following DNA extraction mix:

DEB	100 µl
Proteinase K (0.8 U/µl)	10 µl

3. Transfer each sample into a 1.5 ml Eppendorf® DNA LoBind tube
4. Incubate the samples for 18 h at 56 °C, while shaking at 800 rpm
5. Inactivate the enzyme by incubating for 10 min at 96 °C
6. Place the samples on ice

DNA purification

1. To each sample, add an equal volume of phenol:chloroform:isoamyl alcohol 25:24:1, then shake the samples vigorously
2. Centrifuge the samples at 20,000 x g for 5 min at RT
3. Collect the upper phase and transfer it to a new 1.5 ml Eppendorf® DNA LoBind tube
4. Add an equal volume of chloroform to it, and shake vigorously
5. Centrifuge at 20,000 x g for 5 min at RT
6. Collect the upper phase and transfer it to a new 1.5 ml Eppendorf® DNA LoBind tube, then add sodium acetate to a final concentration of 0.3 M
7. Add 3.7 µl of glycogen per each 100 µl of the solution
8. Add 2.5X volumes of absolute ethanol at -20 °C and incubate the samples for 18 h at -80 °C
9. Centrifuge at 20,000 x g for 90 min at 4 °C
10. Discard the liquid and wash the pellet with 600 µl of ice-cold 70% ethanol
11. Centrifuge at 20,000 x g for 15 min at 4 °C
12. Discard the liquid and wash the pellet with 600 µl of ice-cold 70% ethanol

13. Centrifuge at 20,000 x g for 15 min at 4 °C
14. Discard the supernatant and dry the pellet Note: do not over dry the pellet, as this might result in low DNA recovery
15. Dissolve the pellet in 100 µl of TE buffer

Breakpoint: at this point, the samples can be stored at -20 °C until sonication

DNA sonication

Sonicate the samples aiming to achieve a mean DNA fragment size of 300?500 base pairs (bp). We routinely use Bioruptor® Plus with the following settings: 30 sec ON, 90 sec OFF, high mode, 16 cycles. Alternatively, a Covaris S-series sonicator can be used with the following settings: duty 10 %, intensity 4, time 30 sec, cycle/burst 200, 4 cycles in 50 µl TE/water.

DNA concentration using AMPure beads

1. Pre-warm AMPure XP beads for 30 min at RT
2. To each sample, add 80 µl of AMPure XP bead suspension
3. Mix thoroughly by pipetting up-down 5-6 times
4. Incubate for 5 min at RT
5. Place the samples on a magnetic stand
6. Incubate for 5 min until all the beads have attached to the magnet and the liquid is completely transparent
7. Carefully aspirate and discard the supernatant
8. Add 200 µl per sample of ice-cold 80 % ethanol, while the samples are attached to the magnetic stand
9. Aspirate the supernatant
10. Repeat the wash in ice-cold 80 % ethanol once
11. Air-dry the beads for 5 min at RT

    Note: do not over-dry the beads, as this might result in low DNA recovery
12. Remove the samples from the magnetic stand
13. Resuspend each sample in 10 µl of nuclease-free water
14. Incubate for 5 min at RT
15. Place the samples on the magnetic stand
16. Incubate for 5 min until all the beads have attached to the magnet

17. Transfer 8 µl of the cleared solution into a 0.5 ml Eppendorf® DNA LoBind tube
18. Measure DNA concentration using the Qubit® dsDNA HS Assay Kit

Breakpoint: at this point, the samples can be stored at -20 °C until in vitro transcription

In vitro transcription (IVT)

1. In a 0.5 ml Eppendorf® DNA LoBind tube prepare the following IVT mix:

Sonicated DNA	7.5 µl
rNTPs mix*	8 µl
T7 polymerase buffer 10X*	2 µl
T7 polymerase*	2 µl
RiboSafe RNase Inhibitor	0.5 µl

   Note: For samples digested with HindIII, we typically use 300 ng of sonicated DNA as input to the IVT reaction, whereas for those digested with DpnII we use 50 ng.
   Note: all the reagents marked by * are included in the MEGAscript® kit. The rNTPs mix is prepared by mixing equal volumes of each rNTP solution included in the MEGAscript® kit

2. Incubate the samples for 14 h at 37 °C in a thermocycler with the lid set at 70 °C

RNA cleanup

1. To each sample, add 2U (1 µl) of DNase I
2. Incubate the samples for 15 min at 37 °C. Meanwhile, prewarm the Agencourt RNAClean XP beads for 30 min at RT
3. Transfer each IVT product into a 1.5 ml Eppendorf® DNA LoBind tube
4. Bring the volume of each sample up to 30 µl with nuclease-free water
5. Add a 1.8 vol./vol. ratio of RNAClean XP beads suspension to the IVT product
6. Mix thoroughly by pipetting up-down 5-6 times
7. Incubate for 10 min at RT
8. Place the samples on a magnetic stand
9. Incubate for 5 min until all the beads have attached to the magnet and the liquid is completely transparent
10. Aspirate and discard the supernatant
11. Add 200 µl per sample of ice-cold 80 % ethanol, while the samples are attached to the magnetic stand
12. Incubate for 1-2 min at RT

13. Aspirate the supernatant
14. Repeat the wash with ethanol 80 % two times more
15. Airdry the beads for 10 min at RT

    Note: do not over-dry the beads, as this might result in low DNA recovery
16. Remove the samples from the magnetic stand
17. Resuspend the beads in 6 µl of nuclease-free water
18. Incubate for 5 min at RT
19. Place the sample on the magnetic stand
20. Incubate for 5 min until all the beads have attached to the magnet
21. Transfer 6 µl of each sample into a separate 0.5 ml Eppendorf® DNA LoBind tube

Checkpoint: check the RNA fragment size distribution by loading 1 µl of purified IVT product into an RNA 6000 Pico Kit, and run it on a Bioanalyzer 2100

RA3 adapter ligation

Note: all the subsequent steps are done on ice, unless otherwise specified

1. To each sample, add 1 µl of RA3 adapter
2. In a thermocycler, incubate for 2 min at 70 °C with the lid set at 70 °C
3. Immediately place the samples on ice
4. To each sample, add 4 µl of the following RNA ligase mix:

T4 RNA Ligase buffer 10X	2 µl
T4 RNA Ligase 2, truncated (200 U/µl)	1 µl
RNaseOUT™	1 µl

5. Incubate the samples for 2 h at 25 °C in a thermocycler with the lid set at 30 °C
6. Transfer the samples onto ice

Reverse transcription

Note: all the subsequent steps are done on ice, unless otherwise specified

1. To each sample, add 3.5 µl of RTP (10 µM) primer
2. Incubate the samples for 2 min at 70 °C in a thermocycler with the lid set at 70 °C
3. Immediately place the samples on ice
4. Add 11.5 µl per sample of reverse transcription mix:

1st strand buffer*	5 µl
dNTPs mix	0.5 µl
DTT 100 mM*	2 µl

SuperScript® IIII*	2 µl
RNaseOUT™	2 µl

Note: all the reagents marked by * are included in the SuperScript® IIII Reverse Transcriptase kit

5. Incubate the samples for 1 h at 50 °C in a thermocycler

Breakpoint: at this point, the samples can be stored at -20 °C

Library indexing and amplification

1. Transfer each sample into a 200 µl PCR tube
2. To each sample, add 4 µl of the desired RPI (10 µM) primer
3. To each sample, add 71 µl of PCR mix:

NEBNext® 2X PCR Mix	50 µl
RP1 primer	4 µl
Nuclease-free water	17 µl

4. Perform the following cycles in a thermocycler:
   A. 98 °C, 30 sec
   B. 98 °C, 10 sec
   C. 60 °C, 30 sec
   D. 72 °C, 30 sec
   E. Go to B and repeat for 10-14 cycles
   F. 72 °C, 10 min
   G. 4 °C, hold

Note: the number of PCR cycles should be adjusted based on the amount of genomic DNA used in the IVT reaction and the restriction enzyme used. For example, for a 22x22 mm coverslip fully covered with NES, NPC or neurons digested with HindIII or DpnII, we typically use 9 or 10 PCR cycles.

Library cleanup

1. Prewarm the AMPure XP beads for 30 min at RT
2. Transfer each PCR product into a 1.5 ml Eppendorf® DNA LoBind tube
3. Add a 0.8 vol./vol. ratio of AMPure XP bead suspension to each sample
4. Mix thoroughly by pipetting up-down 5-6 times
5. Incubate for 5 min at RT

6. Place the samples on a magnetic stand
7. Incubate for 5 min until all the beads have attached to the magnet and the liquid is completely transparent
8. Aspirate and discard the supernatant
9. Add 200 µl per sample of ice-cold 80% ethanol, while the samples are attached to the magnetic stand
10. Aspirate the supernatant
11. Repeat once the wash in ice-cold 80% ethanol
12. Airdry the beads for 5 min at RT

    Note: do not over-dry the beads, as this might result in low DNA recovery
13. Remove the samples from the magnetic stand
14. Resuspend each sample in 20 µl of nuclease-free water
15. Incubate for 5 min at RT
16. Place the samples on the magnetic stand
17. Incubate for 5 min until all the beads have attached to the magnet
18. For each sample separately, transfer 17-18 µl of the cleared solution into a 1.5 ml Eppendorf® DNA LoBind tube
19. Measure DNA concentration using the Qubit® dsDNA HS Assay Kit
20. Store the libraries at -20 °C

Checkpoint: 1 µl of library can be checked on a Bioanalyzer 2100 using a High Sensitivity DNA Kit
Note: at this point, the libraries can be stored at -20 °C up to one year

B Differentiation Protocol

Cultivation of Neuroepithelial Stem (NES) cells

Reagents

- DMEM/F12+Glutamax (Thermofisher, cat. no. 31331-028)
- Penicilin/Steptamycin (Thermofisher, cat. no. 15140-122)
- N_2 (Thermofisher, cat. no. 17504-001)
- B27 (Thermofisher, cat. no. 17502-044)
- PolyO, (Sigma, cat. no. P3655-100MG)
- Laminin2020 (For NES only!) (Sigma, cat. no. L2020)
- bFGF (Life Technologies, cat. no. CTP0261)
- EGF (10 ng/ml Thermo Scientific, cat. no. PHG0311)
- Trypsin-EDTA (LifeTechnologies, cat. no. R-001-100)
- Trypsin-inhibitor (LifeTechnologies, cat. no. 17075-029)
- BSA (Sigma, cat no. A9418)
- TrypLE Express Enzyme (Thermofisher, cat no. 12605-036)

Chemicals

EGF

Dissolve 1 mg of EGF in 10 ml of dPBS. Aliquot and store at -80 °C. Dilute 10 000 times to final concentration of 10 ng/ml.

bFGF

Dissolve 1 mg in 50 ml dPBS + 0.1 % BSA and aliquot. Store at -80 °C. Dilute 2000 times to a final concentration of 10 ng/ml.

polyOrnithine

Dissolve 100 mg in 10 ml sterile dH_2O and aliquot. Store at -80 °C

Medium

NES medium (for 50 ml)

DMEM/F12+Glutamax	48.5 ml
N_2	500 µl
PE/ST	500 µl

B27	50 µl
bFGF (10 ng/ml)	25 µl
EGF (10 ng/ml)	5 µl

Complete NES medium can be kept in the fridge for one week.

Wash medium (for 50 ml)

DMEM/F12+Glutamax	49.5 ml
BSA (20 % stock)	5 ml

- Prepare a 20 % BSA stock in DMEM/F12+Glutamax, sterile filter and dispense in 5 ml aliquots, store at -20 °C.
- Ready to use wash medium: add 5 ml 20% BSA Stock to 500 ml medium. Keep in the fridge for up to 2 weeks or aliquot in 50 ml tubes and freeze.
- BSA can be replaced with 5% Fetal Bovine Serum

Trypsin-inhibitor

- Dissolve 250 mg of trypsin inhibitor powder into 30 ml of PBS.
- Incubate in the water bath for 1-2 hours.
- Sterile filter into a 500 ml bottle of PBS.
- Aliquot into 50 ml tubes and store at -20 °C.

Methods

Coating

1. Dilute the PolyOrnithin stock x500 in dPBS.
2. Add PolyOrnithin to the wells (0.5 ml/12-well, 1 ml/6-well, 3 ml/T25 and 9 ml/T75)
3. Incubate for 30 minutes at +37 °C
4. Wash twice with PBS
5. Dilute Laminin 2020 x500 in PBS and coat for 4 hours at +37 °C or over night at +4 °C
6. The coated flasks can be stored at +4 °C for at least one week.

Culturing

1. The reagents, Trypsin , Trypsin-Inhibitor, wash medium and medium should be pre-warmed to +37 °C.
2. Aspire the medium
3. Add Trypsin: 500 µl for a T25, 1 ml for a T75 flask.

4. Incubate at +37 °C for 1-2 minutes and check under the microscope if cells have de-attached
5. Add the same amount of Trypsin-Inhibitor and centrifuge the cells, 1300 rpm for 3 minutes.
6. Re-suspend in wash medium and count the cells.
7. Centrifuge the cells, 1300 rpm for 3 minutes.

Alternative method: Instead of using Trypsin, Trypsin Inhibitor and Wash medium:

1. Add TrypLE: 1 ml for a T25, 2 ml for a T75 flask.
2. Wash the cells once in DMEM/F12 supplemented with 5 % Fetal Calf Serum (FCS) .
3. Re-suspend in NES medium and plate the cells, 40 000 cells/cm^2. (Usually 1:3).

Freezing

1. Dissociate the cells as usual.
2. Re-suspend the cells in NES medium and count the cells.
3. Centrifuge again and re-suspend the cells in NES medium supplemented with 10 % DMSO to final concentration of 4 x10^6 cells/ml.
4. Add 500 µl (2 x 10^6 cells) to a vial and freeze in -80 °C over night before moving to liquid N_2.

Thawing

1. Thaw the vial at +37 °C in a water bath until it is almost thawed
2. Add 0.5 ml pre-warmed NES-medium to the vial
3. Transfer the volume to a 15 ml tube with 2-3 ml NES medium
4. Centrifuge the cells at 1300 rpm for 3 minutes
5. Re-suspend the cells in 1 ml of NES medium and add to a pre-coated T12.5 flask
6. Add 1-2 ml NES medium

Differentiation

1. Plate the cells in NES-medium without EGF and FGF2, but with N_2 and ten times higher concentration of B27 (1:100). Plate at a density of 40 000 cells/cm^2
2. Feed every second day Optional: At day 4-5, passage the cells to plates coated with PolyO and Laminin 1:500. Passage 1:2.
3. At days past 14, add Laminin2020 to the medium (1:1000)

C List of Fluorescent Oligonucleotides

TU Bibliothek Your knowledge hub WIEN
Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Table 9: List of oligos used to make YFISH and GPSeq adapters

Oligo name	Sequence (5' → 3')	Purification
	Oligos for YFISH	
YFISH_HindIII_FOR	AATACGACTCACTATAGGGATTAATACGACTCACTATAGGGACCCGAACTCCGTAACGTAGCTGATGAACTAGCCTAGATAGAC	HPLC
YFISH_HindIII_REV	AGCTGTCTATCTAGGCTAGTTCATCAGCTACGTTACGGAGTTCGGGAATACGACTCACTATAGGGATTAATACGACTCACTATAGGGA	HPLC
YFISH_DpnII_FOR	CGGAGACACAGCATACTATGTTCGGAGACACAGCATACTATGCCCGAACTCCGTAACGTAGCTGATGAACTAGCCTAGATAGAC	HPLC
YFISH_DpnII_REV	GATCGTCTATCTAGGCTAGTTCATCAGCTACGTTACGGAGTTCGGGCGGAGACACAGCATACTATGTTCGGAGACACAGCATACTATG	HPLC
	Oligos for GPSeq	
GPSeq_HindIII_FOR_1	AGCTTCGACGAATNNNNNNNNGATCGTCGGACTGTAGAACTCTGAACCCCTATAGTGAGTCGTATTACCGGCCTCAATCGAA	Desalting
GPSeq_HindIII_REV_1	CGATTGAGGCCGGTAATACGACTCACTATAGGGGTTCAGAGTTCTACAGTCCGACGATCNNNNNNNNATTCGTCGA	Desalting
GPSeq_HindIII_FOR_2	AGCTTTGACAAGCNNNNNNNNGATCGTCGGACTGTAGAACTCTGAACCCCTATAGTGAGTCGTATTACCGGCCTCAATCGAA	Desalting
GPSeq_HindIII_REV_2	CGATTGAGGCCGGTAATACGACTCACTATAGGGGTTCAGAGTTCTACAGTCCGACGATCNNNNNNNNGCTTGTCAA	Desalting
GPSeq_HindIII_FOR_3	AGCTTTGATGGCTNNNNNNNNGATCGTCGGACTGTAGAACTCTGAACCCCTATAGTGAGTCGTATTACCGGCCTCAATCGAA	Desalting
GPSeq_HindIII_REV_3	CGATTGAGGCCGGTAATACGACTCACTATAGGGGTTCAGAGTTCTACAGTCCGACGATCNNNNNNNNAGCCATCAA	Desalting
GPSeq_HindIII_FOR_4	AGCTTTGAGTTAGNNNNNNNNGATCGTCGGACTGTAGAACTCTGAACCCCTATAGTGAGTCGTATTACCGGCCTCAATCGAA	Desalting
GPSeq_HindIII_REV_4	CGATTGAGGCCGGTAATACGACTCACTATAGGGGTTCAGAGTTCTACAGTCCGACGATCNNNNNNNNCTAACTCAA	Desalting
GPSeq_HindIII_FOR_5	AGCTTTCGCGAATNNNNNNNNGATCGTCGGACTGTAGAACTCTGAACCCCTATAGTGAGTCGTATTACCGGCCTCAATCGAA	Desalting
GPSeq_HindIII_REV_5	CGATTGAGGCCGGTAATACGACTCACTATAGGGGTTCAGAGTTCTACAGTCCGACGATCNNNNNNNNATTCGCGAA	Desalting
GPSeq_HindIII_FOR_6	AGCTTTGAACCTANNNNNNNNGATCGTCGGACTGTAGAACTCTGAACCCCTATAGTGAGTCGTATTACCGGCCTCAATCGAA	Desalting
GPSeq_HindIII_REV_6	CGATTGAGGCCGGTAATACGACTCACTATAGGGGTTCAGAGTTCTACAGTCCGACGATCNNNNNNNNTAGGTTCAA	Desalting
GPSeq_HindIII_FOR_7	AGCTTCTCGTTAGNNNNNNNNGATCGTCGGACTGTAGAACTCTGAACCCCTATAGTGAGTCGTATTACCGGCCTCAATCGAA	Desalting
GPSeq_HindIII_REV_7	CGATTGAGGCCGGTAATACGACTCACTATAGGGGTTCAGAGTTCTACAGTCCGACGATCNNNNNNNNCTAACGAGA	Desalting
GPSeq_DpnII_FOR_1	GATCGCGTGATGNNNNNNNNGATCGTCGGACTGTAGAACTCTGAACCCCTATAGTGAGTCGTATTACCGGCCTCAATCGAA	HPLC
GPSeq_DpnII_REV_1	CGATTGAGGCCGGTAATACGACTCACTATAGGGGTTCAGAGTTCTACAGTCCGACGATCNNNNNNNNCATCACGC	HPLC
GPSeq_DpnII_FOR_2	GATCGCGACGACNNNNNNNNGATCGTCGGACTGTAGAACTCTGAACCCCTATAGTGAGTCGTATTACCGGCCTCAATCGAA	HPLC
GPSeq_DpnII_REV_2	CGATTGAGGCCGGTAATACGACTCACTATAGGGGTTCAGAGTTCTACAGTCCGACGATCNNNNNNNNGTCGTCGC	HPLC
GPSeq_DpnII_FOR_3	GATCGCGGTCGTNNNNNNNNGATCGTCGGACTGTAGAACTCTGAACCCCTATAGTGAGTCGTATTACCGGCCTCAATCGAA	HPLC
GPSeq_DpnII_REV_3	CGATTGAGGCCGGTAATACGACTCACTATAGGGGTTCAGAGTTCTACAGTCCGACGATCNNNNNNNNACGACCGC	HPLC
GPSeq_DpnII_FOR_4	GATCGCGCATCANNNNNNNNGATCGTCGGACTGTAGAACTCTGAACCCCTATAGTGAGTCGTATTACCGGCCTCAATCGAA	HPLC
GPSeq_DpnII_REV_4	CGATTGAGGCCGGTAATACGACTCACTATAGGGGTTCAGAGTTCTACAGTCCGACGATCNNNNNNNNTGATGCGC	HPLC
GPSeq_DpnII_FOR_5	GATCGATTGATGNNNNNNNNGATCGTCGGACTGTAGAACTCTGAACCCCTATAGTGAGTCGTATTACCGGCCTCAATCGAA	HPLC
GPSeq_DpnII_REV_5	CGATTGAGGCCGGTAATACGACTCACTATAGGGGTTCAGAGTTCTACAGTCCGACGATCNNNNNNNNCATCAATC	HPLC

D List of Primer Sequences

RNA PCR Index Primers (RPI16-23)

as mentioned in table 5
Index sequence is 6 bases as underlined

RNA PCR Primer, Index 16 (RPI16)

5’ CAAGCAGAAGACGGCATACGAGATGGACGGGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA

RNA PCR Primer, Index 17 (RPI17)

5’ CAAGCAGAAGACGGCATACGAGATCTCTACGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA

RNA PCR Primer, Index 18 (RPI18)

5’ CAAGCAGAAGACGGCATACGAGATGCGGACGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA

RNA PCR Primer, Index 19 (RPI19)

5’ CAAGCAGAAGACGGCATACGAGATTTTCACGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA

RNA PCR Primer, Index 20 (RPI20)

5’ CAAGCAGAAGACGGCATACGAGATGGCCACGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA

RNA PCR Primer, Index 21 (RPI21)

5’ CAAGCAGAAGACGGCATACGAGATCGAAACGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA

RNA PCR Primer, Index 22 (RPI22)

5’ CAAGCAGAAGACGGCATACGAGATCGTACGGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA

RNA PCR Primer, Index 23 (RPI23)

5’ CAAGCAGAAGACGGCATACGAGATCCACTCGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA