From the Department of Oncology-Pathology
Karolinska Institutet, Stockholm, Sweden

CLONAL STRUCTURES AND CELL INTERACTIONS IN CANCER

Nathanael Andrews

Karolinska Institutet

Stockholm 2023

Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2023

ISBN 978-91-8017-043-7

Clonal Structures and Cell Interactions in Cancer

Thesis for Doctoral Degree (Ph.D.)

By

Nathanael Andrews

The thesis will be defended in public at:
J3:12 Nanna Svartz, Nya Karolinska Sjukhuset, Solnavägen 30
2023-06-20 kl 09:00

Principal Supervisor:
Assistant Professor Martin Enge
Karolinska Institutet
Department of Oncology-Pathology

Co-supervisor(s):
Doctor of Medicine Marco Gerling
Karolinska Institutet
Department of Bioscience and Nutrition

Opponent:
Assistant Professor Andreas Moor
ETH Zurich
Department of Biosystems and Engineering

Examination Board:
Professor Rickard Sandberg
Karolinska Institutet
Department of Cell and Molecular Biology

Assistant Professor Ola Larsson
Karolinska Institutet
Department of Department of Oncology-Pathology

Associate Professor Pelin Sahlén
KTH Royal Institute of Technology
Department of Department of Gene Technology

To my son, my wife, and my parents

Popular science summary of the thesis

Cancer is a disease caused by mutations to our DNA - the genetic blueprint that instructs our cells how to function. Genetic information flows from DNA to RNA to proteins. In other words, our DNA is first transcribed into RNA, which is then translated into proteins, complex molecules that play essential roles in our bodies. When mutations occur, they can affect the function of these proteins, altering cell behavior and potentially leading to the development of cancer.

Cancer is not a single disease but rather a collection of diseases with one common theme: uncontrolled cell growth. Cells within a tumor are not all the same; instead, they display a high degree of diversity. This diversity makes it challenging to target cancer cells effectively with treatments, as different cells may respond differently to the same therapy. This is where single-cell sequencing techniques come into play - enabling the sequencing of DNA or RNA from each cell separately, uncovering unique characteristics and behaviors of individual cells and providing a more detailed and accurate picture of cellular diversity and function.

The development of new single-cell sequencing methods is essential for gaining a deeper understanding of the complex interactions between cancer cells and their environment. In this thesis, we have focused on creating novel single-cell sequencing techniques to study mutations and cell interactions that affect cancer behavior. By improving our ability to analyze the genetic and molecular changes in individual cancer cells, these methods have the potential to unlock new insights into cancer biology, ultimately leading to more effective and targeted treatments for patients.

In **paper I**, we developed a method to enhance the existing single-cell RNA sequencing methods by adding spatial information. This means we can now determine not just what different cells are doing, but also where they are doing it, which helps to identify which cells are likely to interact with each other in a physical space. This method is designed to be unsupervised, high throughput, and compatible with most existing single-cell RNA sequencing methods. In **paper II**, we developed a method that enables us to examine both DNA and RNA from the same cell. This dual profiling is crucial to understanding the relationship between genetic information and its expression, enabling us to study how genetic mutations, changes in the DNA, can influence the behavior of cells.

For **paper III**, we built upon the knowledge gained from the previous studies. We generated high-quality ATAC-seq and RNA-seq data from the same cell. ATAC-seq is a method that tells us which parts of the DNA are active, while RNA-seq reveals which genes are actively being used by the cell. We designed our method to provide improved data quality over previous techniques. Finally, for **paper IV**, we combined the ATAC-seq method with a technology known as CRISPRi screening. CRISPRi is a modification of the well-known CRISPR technology that allows us to 'turn off' specific genes and examine what happens. With this combination, we delved into studying a form of pancreatic cancer called pancreatic ductal adenocarcinoma. We looked at which regions of the DNA were active in the cancer's growth and proliferation, and identified the specific transcription factors, the proteins that turn genes on or off, interacting with these active regions.

The methods provided in this thesis will be significant for enhancing our understanding of cancer biology and help develop effective treatments. By understanding the spatial arrangement of cells, we can explore how cancer cells interact with their neighboring cells, potentially uncovering mechanisms that cancer cells use to spread. Meanwhile, by simultaneously examining mutations or active regions of DNA, and RNA of the same cell we can identify genetic mutations and their direct impact on the cell's behavior. These new methods enable a deeper exploration of the fundamental processes in cancer cells, which can contribute to the discovery of novel strategies for the treatment of cancer.

Abstract

Despite sharing an identical genome, cells of higher order multicellular organisms display a large degree of phenotypic diversity. This diversity is maintained by a sophisticated regulatory machinery that integrates information from both intrinsic and extrinsic factors, ultimately coordinating the appropriate gene expression. Sequencing methods such as RNA and DNA sequencing have become indispensable tools in the pursuit to understand gene regulation. In recent years, the integration of single-cell sequencing techniques and CRISPR-based methods has ushered in a new era of genomic exploration, providing unprecedented opportunities to investigate the intricate interplay between genes, cellular processes, and disease progression. These cutting-edge advances have transformed the research landscape, enabling in-depth studies of gene regulation in single cells, and paving the way for future discoveries in both healthy and malignant tissues.

While cancer has traditionally been studied as a genetic disease, it is now evident that mutations alone do not determine cancer initiation or progression. This notion is supported by two key observations: first, cancer-driving mutations do not always lead to malignancy; and second, identical mutations can yield different outcomes depending on the cell type in which they occur. Consequently, a deeper understanding of gene regulation and the various ways it is modulated is critical for deciphering the complex relationship between genetic changes and cancer initiation.

In this thesis we aimed to develop novel single-cell methodologies applicable to studying biological complex systems. We have developed four techniques: CIM-seq, DNTR-seq, Smart3-ATAC, and ACTI-seq, described in **papers I-IV**, respectively. The methods all capture additional modalities in combination with single-cell RNA-seq data, including spatial information, whole genome sequencing, accessible chromatin, and direct read out of guide RNAs. We applied these methods to investigate biological systems at the single-cell level, offering a more comprehensive understanding of cellular behavior in health and disease. Our approaches have allowed us to characterize stem cell niches and regeneration dynamics in the epithelial layer of the colon, and delve into the effects of gene dosage, quantifying how mutational changes impact transcriptional output. Furthermore, we have explored the complex landscape of gene regulation within pancreatic ductal adenocarcinomas, identifying mechanisms that enable cancer growth and proliferation.

This body of work emphasizes the importance of multimodal and integrative approaches for unraveling the complexities of biological systems at a cellular level. The methods we've developed represent a significant step forward, promising to facilitate the discovery of molecular targets for cancer therapeutics.

List of scientific papers

I. **Nathanael Andrews***, Jason T. Serviss*, Natalie Geyer, Agneta B. Andersson, Ewa Dzwonkowska, Iva Šutevski, Rosan Heijboer, Ninib Baryawno, Marco Gerling, and Martin Enge.
An Unsupervised Method for Physical Cell Interaction Profiling of Complex Tissues.
Nat. Methods 18, 912–920 (2021)
*These authors contributed equally

II. Vasilios Zachariadis, Huaitao Cheng, **Nathanael Andrews,** and Martin Enge.
A Highly Scalable Method for Joint Whole-Genome Sequencing and Gene-Expression Profiling of Single Cells.
Mol. Cell 80: 541-+ (2020)

III. Huaitao Cheng, Han-pin Pui, Antonio Lentini, Solrún Kolbeinsdóttir, **Nathanael Andrews**, Yu Pei, Björn Reinius, Qiaolin Deng, and Martin Enge
Smart3-ATAC: a Highly Sensitive Method for Joint Accessibility and Full-Length Transcriptome Analysis in Single Cells.
Manuscript, bioRxiv (2021) doi: https://doi.org/10.1101/2021.12.02.470912

IV. **Nathanael Andrews**, Huaitao Cheng, and Martin Enge
Simultaneous Sequencing of Full-Length RNA Transcripts, Accessible Chromatin, and Guide RNAs for Isoform-Sensitive CRISPR Perturbation Analysis.
Manuscript

Scientific papers not included in the thesis

Antonio Lentini, Huaitao Cheng, J. C. Noble, Natali Papanicolaou, Christos Coucoravas, **Nathanael Andrews**, Qiaolin Deng, Martin Enge, and Björn Reinius.
Elastic Dosage Compensation by X-chromosome Upregulation.
Nat. Commun. 13, 1854 (2022)

Jason T. Serviss, **Nathanael Andrews**, Jimmy Van den Eynden, Felix Clemens Richter, Miranda Houtman, Mattias Vesterlund, Laura Schwarzmueller, Per Johnsson, Erik Larsson, Dan Grandér, and Katja Pokrovskaja Tamm.
An Antisense RNA Capable of Modulating the Expression of the Tumor Suppressor MicroRNA-34a.
Cell Death Dis. 9, 736 (2018)

Edward C. Harding, Xiao Yu, Andawei Miao, **Nathanael Andrews**, Ying Ma, Zhiwen Ye, Leda Lignos, Giulia Miracca, Wei Ba, Raquel Yustos, Alexei L. Vyssotski, William Wisden, and Nicholas P. Franks.
A Neuronal Hub Binding Sleep Initiation and Body Cooling in Response to a Warm External Stimulus.
Curr. Biol. 28: 2263-+ (2018)

Contents

List of abbreviations

AIDS	Acquired immunodeficiency syndrome
ATAC-seq	Assay for transposase-accessible chromatin using sequencing
BAC	Bacterial artificial chromosomes
Cas9	CRISPR-associated endonuclease 9
CAF	Cancer-associated fibroblast
cDNA	Complementary DNA
CNV	Copy number variation
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
dCas9	Dead Cas9
DLP	Direct library preparation
DNA	Deoxyribonucleic acid
DNA-seq	DNA sequencing
DOP-PCR	Degenerate oligonucleotide primed PCR
ECM	Extracellular matrix
FGF	Fibroblast growth factor
GWAS	Genome wide association study
HGP	Human genome project
LIANTI	Linear amplification via transposon insertion
MALBAC	Multiple annealing and looping-based amplification cycles
MDA	Multiple displacement amplification
MDSC	Myeloid-derived suppressor cell
ncRNA	non-coding RNA
NDR	Nucleosome-deprived region
NGS	Next-generation sequencing
PALS	Periarteriolar lymphoid sheath

PTA	Primary template-directed amplification
RNA	Ribonucleic acid
SBS	Sequence by synthesis
scRNA-seq	Single-cell RNA sequencing
snRNA-seq	Single-nucleus RNA sequencing
SNV	Single nucleotide variation
TAD	Topologically associating domain
TF	Transcription factor
TME	Tumor microenvironment
Treg	Regulatory T cell
TSS	Transcription start site
VEGF	Vascular endothelial growth factor
WGA	Whole-genome amplification
WGS	Whole-genome sequencing

1 Introduction

Cells are the building blocks of multicellular organisms, with multiple cell types constituting distinct tissues capable of carrying out specific functions in a coordinated fashion. These functions vary greatly, and the dramatic differences among tissues are mirrored by the complexity of cell types present in higher order eukaryotes. While cell types describe morphologically or phenotypically distinct cells, cell states describe intermediate stages along trajectories among cell types, but also transient or reversible features such as proliferative, metabolic, or active states, which may differ among cells of the same type. The current generation of single-cell sequencing methods have enabled large-scale studies of such states, revealing nuanced differences between cells of the same type, and set the stage for systematic mapping of cell states and types across organisms in health and disease (Schaum *et al*., 2018; Travaglini *et al*., 2020; Eze *et al*., 2021; Tabula Sapiens Consortium, 2022).

Cancer is a disease of altered cell states. Mutations free the cell from constraints which normally prevent uncontrolled growth, leading to a persistent proliferative cell state and the accumulation of further mutations. This proliferative state is the main basis for chemotherapeutics which target processes such as microtubule assembly and DNA replication, which occur more frequently in proliferative cells. However, targeting cells in this manner can lead to the selection of treatment-resistant cells, as quiescent cancer cells are capable of avoiding apoptosis and will lead to a relapse of disease.

Previous cancer research has largely focused on highly impactful mutations, and usually ones affecting protein coding regions, as they are more easily detected and provide potential drug targets. Unfortunately, most cancers are the result of the accumulation of a large number of lesions, many of which are located in non-coding regions of the genome, and efficient treatments are unlikely to be achieved by targeting singular proteins and pathways. Instead, there is a need for a better understanding of which factors affect cell states and how these are perturbed in cancer progression and during treatment. This will allow the identification of states which confer therapy resistance, as well as the development of therapeutics which modify the cell state into one which is more susceptible to treatment.

In this review I will provide an overview of the technological breakthroughs that have enabled the development of single-cell sequencing techniques. I will describe general mechanisms for transcriptional regulation, ranging from specific interactions between proteins and DNA-motifs to signaling pathways initiated by extracellular cues, and how these are combined to maintain specific cell states. Additionally, I will discuss how these mechanisms fail in cancer cells, leading to states promoting survival and uncontrolled proliferation. Finally, I will discuss how single-cell sequencing techniques have contributed to our current understanding of transcriptional regulation and cell states, future avenues for these techniques, and how they may be implemented to improve cancer therapies.

2 Literature review

2.1 The development of sequencing methods

2.1.1 Sanger sequencing and the dawn of the sequencing era

To gain a complete understanding of the evolution of sequencing methods, we must first explore the roots of DNA sequencing techniques, as modern nucleotide sequencing methods predominantly sequence DNA directly, or rely on the conversion of RNA into complementary DNA (cDNA) before sequencing. Two methods were published in 1977 which allowed for the sequencing of hundreds of bases in a day. Both methods, developed by Sanger and Coulson, and Maxam and Gilbert, were based on generating a radioactively labeled DNA sequence, generating base-specific fragments, and establishing the full sequence through gel-electrophoresis (Maxam and Gilbert, 1977; Sanger *et al., 1977)*. The generation of methods capable of decoding hundreds of bases in an afternoon transformed the field overnight, but despite this leap, DNA sequencing was still a time-consuming process, making the sequencing of bacterial genomes laborious, and the sequencing of larger genomes an impossibility.

Shotgun sequencing significantly improved the rate at which genomes could be sequenced (Staden, 1979). By breaking long DNA molecules into smaller, manageable fragments, and sequencing them simultaneously, a multitude of sequence reads corresponding to different sections of the original DNA molecule are generated. These sequencing reads can then be assembled to reconstruct the original DNA sequence. While initially performed by hand, soon computational algorithms were developed, identifying overlapping regions between the sequence reads and merging them to create contiguous sequences, vastly improving sequencing efficiency and speed.

It is difficult to overstate the impact that Sanger sequencing and shotgun sequencing has had on society. In less than a decade after its inception, Sanger sequencing had already been used for the sequencing of the HIV genome, leading to the development of diagnostic tests and treatment strategies for AIDS (Ratner *et al*., 1985). Through the utilization of Sanger sequencing with shotgun sequencing, the first complete bacterial genome, Haemophilus influenzae, was sequenced in 1995, paving the way for comparative genomics within microbial research (Fleischmann *et al*., 1995). These methods were also pivotal for identifying tumor suppressors and cancer-causing mutations, such as mutations occurring in *BRCA-1* and *BRCA-2*, which were described in the mid-90s (Miki *et al*., 1994; Wooster *et al*., 1995).

2.1.2 Towards sequencing the human genome

Following the successes of these newly developed techniques, scientists began eying a larger target - a comprehensive map of the human genome. The Human Genome Project (HGP) was launched in 1990 to sequence the 3 billion bases of the human genome. The primary method implemented to complete the HGP was the "hierarchical shotgun" approach, involving breaking large portions of the human genome into bacterial artificial chromosomes (BACs). DNA from each BAC was then fragmented, size-selected, and sub-cloned. Selected clones were cultivated, and their DNA was extracted and served as a template for automated Sanger sequencing.

To further improve the scale and efficiency of sequencing methods, a series of additional advancements were made during HGP. These advancements included: (1) the development of dye-labeled terminators and mutant T7 DNA polymerase, which more readily incorporated them, reducing the number of required sequencing reactions from four to one; (2) the introduction of linear amplification reactions, significantly lowering template requirements; (3) methods that facilitated the sequencing of double-stranded DNA; and (4) the automation and standardization of operating procedures to maximize efficiency and minimize errors (Craxton, 1991; Tabor and Richardson, 1987). For a more extensive read on the challenges associated with and overcome during HGP see Hood and Rowen, 2013, and Shendure *et al*. 2017.

The completion of the Human Genome Project led to the comprehensive discovery and cataloging of most human genes, and through inference, of most human proteins, along with other elements such as non-coding RNAs (ncRNA). This resource became invaluable for fields such as evolutionary and comparative genomics, where genes of other organisms could be compared to the human genome, charting evolutionary relationships, and shedding light on conserved and non-coding regions and their functional importance. The HGP also provided a foundation for future sequencing techniques, with the mapped genome obtained from HGP being used as a reference genome for Next Generation Sequencing (NGS) techniques. This strategy was first demonstrated in bacteria before seeing wide use for human genomics, circumventing some of the challenges associated with *de novo* genome assembly and facilitating larger scale sequencing studies (Margulies *et al*., 2005; Shendure *et al*., 2005).

2.1.3 The $1000 genome and beyond

While the HGP was being completed, attempts were made to develop alternatives to Sanger sequencing, to generate larger libraries and reduce costs. The “$1000 genome” was motivated by the promise of personalized medicine, where sequencing an individual's genome to drive medical decision-making would become a possibility. This led to the development of massively parallel NGS methods which have since replaced Sanger sequencing almost entirely. NGS uses a variety of technologies to generate massive amounts of sequencing data in a relatively short time and at a lower cost. Current NGS technologies sequence complex libraries of DNA fragments immobilized onto a two-dimensional surface (i.e., flow cell). Through *in vitro* amplification and repeated cycles of biochemistry and imaging, a process known as "sequencing-by-synthesis" (SBS), the DNA sequence can be determined (Mardis, 2008). NGS platforms can generate millions of sequencing reads in parallel, allowing for the analysis of entire genomes, transcriptomes, and epigenomes. While NGS is faster and more cost-effective than Sanger sequencing, the shorter read lengths and higher error rates require more complex bioinformatics analysis.

While multiple strategies for SBS exist (Ronaghi *et al*., 1996; Brenner *et al*., 2000; Drmanac *et al*., 2010), the most extensively used is polymerase-mediated incorporation of fluorescently labeled deoxynucleotides, which are added stepwise during the sequencing process (Braslavsky *et al*., 2003; McKernan *et al*., 2009). The key to the success of this method was the development of reversibly terminating, reversibly fluorescent dNTPs ensuring the incorporation of only one dNTP per cycle. After each cycle, the DNA templates are imaged to determine the fluorescent color of the incorporated base, and the blocking and fluorescent groups are removed to enable the next extension. While the industry is currently dominated by Illumina, recent challengers have emerged, improving cost efficiency by mainly using unmodified nucleotides, with a low fraction of fluorescently labeled nucleotides in a process termed mostly natural SBS (Simmons *et al*., 2023), or even by abandoning SBS completely, relying instead on novel sequencing chemistries, such as avidity binding of dye labeled polymers to DNA fragments on the flow cell surface (Arslan *et al*., 2022; Li *et al*., 2022).

The development and improvement of NGS methods led to the realization of the $1,000 genome in 2014 (Check Hayden, 2014), and further development has since pushed the field toward even cheaper genomes (See **Fig. 1**). Much like the development of Sanger sequencing, NGS methods led to the birth of new fields, including single-cell sequencing technologies, which rely on the ability to capture many sequences from a multiplexed sample.

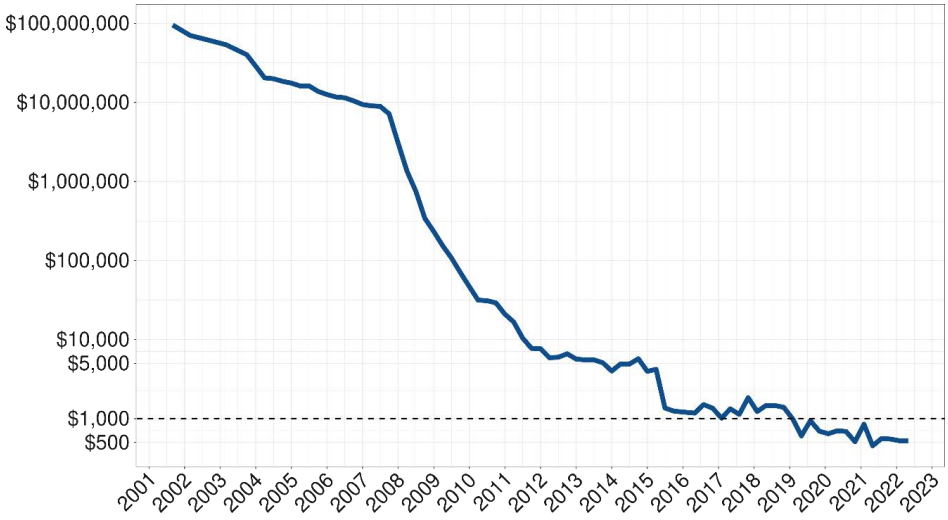


Figure 1: Cost of sequencing a whole genome by year. *Data obtained from the National Human Genome Research Institute.*

2.1.4 Third generation sequencing

The work comprising this thesis makes use of second-generation NGS techniques. However, it is essential to acknowledge the exciting progress occurring in the realm of third-generation sequencing methods, often referred to as long-read sequencing techniques. Third-generation sequencing techniques are capable of sequencing fragments with 20,000 bases or more, far longer than previous technologies, enabling the resolution of repetitive regions, structural variants, and other genomic features that are often difficult to discern with short-read approaches. Currently, two major platforms exist: Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT). PacBio Single-Molecule Real-Time (SMRT) sequencing uses a zero-mode waveguide system to observe the incorporation of fluorescently labeled nucleotides in real-time, generating long reads with high accuracy (Rhoads and Kin Fai, 2015). ONT's nanopore sequencing, on the other hand, measures changes in ionic current as DNA or RNA molecules pass through a protein nanopore, providing direct, real-time sequencing of DNA or RNA without the need for amplification or synthesis (Wang *et al.*, 2021).

As these technologies are in their infancy, current iterations are still trailing second-generation techniques with respect to high error rates, higher costs, and lower throughput. Despite this, their ability to generate long reads facilitate genome assembly, detection of structural variants, and the analysis of full-length transcripts, making third generation techniques attractive for specific research questions. It is likely that further adoption third generation sequencing methods will occur in the coming years, when some of the drawbacks are ironed out.

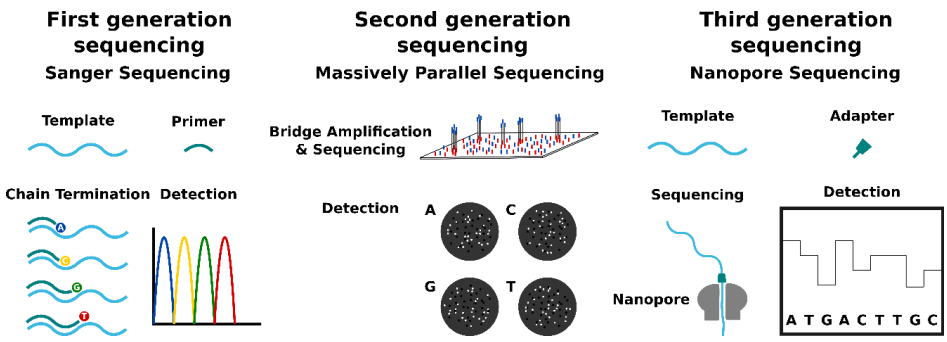


Figure 2: Examples of first, second, and third generation sequencing methods.

2.2 Single-cell sequencing

In the late 2000s, advancements in sequencing technologies were dramatically changing the landscape of biological research. Motivated by overcoming limitations of traditional bulk sequencing methods and encouraged by the technological advancements made in the previous decade, the first single-cell sequencing methods were developed. These methods provided new opportunities to study cell heterogeneity and map the transcriptional profiles of rare cell types, uncovering the roles of individual cells in health and disease.

2.2.1 Single-cell RNA sequencing

The first scRNA-seq method was published in 2009. Driven by the need for a method capable of sequencing the RNA of primordial germ cells from early mouse embryos, where cells were in short supply, a method was developed for performing RNA-seq at a single-cell resolution (Tang *et al*., 2009). The results demonstrated the feasibility of profiling gene expression in individual cells, laying the foundation for future advancements in the field. Despite its relatively low sensitivity, relying on poly-A tailing and adaptor ligation to complete libraries, this work highlighted the potential of single-cell sequencing to reveal cellular heterogeneity and investigate gene expression dynamics in unprecedented detail.

Soon after, a plethora of methods was developed, each with their own approach to produce single-cell libraries. Smart-seq implemented a template switching oligo (TSO), making use of the Moloney murine leukemia virus (MMLV) reverse transcriptase, which adds three cytosines to the 5' end of the transcript (Ramsköld *et al*., 2012). This strategy was subsequently incorporated in STRT-seq (Single-Cell Tagged Reverse Transcription sequencing) which incorporated a barcode via the TSO, allowing for the pooling of samples prior to final library preparation (Islam et al., 2012). In a similar fashion, CEL-seq (Cell Expression by Linear amplification and sequencing), introduced in 2012, implemented an early barcoding approach, introducing a barcode via the dT primer, pooling cDNA prior to fragmentation and ligation with sequencing adapters (Hashimshony *et al*., 2012). These methods improved the sensitivity of gene expression profiling in single cells and significantly improved throughput compared to previous methods.

Further improvements were made through the incorporation of unique molecular identifiers, of which MARS-seq (Massively parallel RNA single-cell sequencing) was an early adopter of (Jaitin *et al*., 2014). MARS-seq combined the use of UMIs with a multi-tiered barcoding strategy, enabling the parallel processing of hundreds of cells in a single run, and reducing amplification noise, making it a popular choice for large-scale single-cell transcriptomics studies (Ziegenhain *et al*., 2017). In 2013, Smart-seq2, an improved version of the proprietary Smart-seq method, became the gold standard for scRNA-seq methods (Picelli *et al*., 2013). By performing Tn5-mediated tagmentation of amplified cDNA, Smart-seq2 enabled the sequencing of full-length transcripts from single cells. Smart-seq2 was further modified to incorporate UMIs in Smart-seq3 and Smart-seq3xpress, resulting in increased sensitivity with a drastically improved throughput and cost-efficiency (Hagemann-Jensen *et al*., 2020; Hagemann-Jensen *et al*., 2022).

In addition to plate-based methods, microfluidics-based RNA-seq methods have gained increased popularity, due to their ease of use, and the fact that they do not require specialized tools such as liquid handling robots for sequencing a large number of cells. By using microfabricated devices and micro-scale fluid manipulation, single cells are processed in micro to nanoliter volumes in an automated fashion, reducing reagent consumption. One of the earliest examples of a microfluidics-based RNA-seq method is the Fluidigm C1 system, introduced in 2012 (Wu *et al*., 2014). This system automated the capture, lysis, and reverse transcription of individual cells in microfluidic chambers, improving the throughput of scRNA-seq experiments. More recently, methods such as Drop-seq and the 10x Genomics Chromium have been developed, encapsulating cells into droplets to considerably improve throughput (Macosko *et al*., 2015; Zheng *et al*., 2017). These methods have made single-cell RNA sequencing widely available, through commercially available platforms such as 10x Chromium. Recently, droplet sequencing has been made even simpler, based on particle-templated emulsification, enabling single-cell encapsulation and barcoding of cDNA in droplet emulsions using only a vortexer, removing all requirements of specialized microfluidic devices, expertise, or hardware (Clark *et al*., 2023).

Despite the numerous advantages of microfluidics and droplet-based RNA-seq methods, some drawbacks persist. Importantly, while offering throughput orders of magnitude larger than what is available for plate-based methods, microfluidics-based RNA-seq methods suffer from lower sensitivity than plate-based methods. Additionally, microfluidic devices often have fixed capture chamber sizes, which may not be suitable for all cell types and sizes; and loading cells into microfluidic chambers can be stressful for cells, potentially affecting their viability. Finally, plate-based methods offer more flexibility in terms of experimental design and protocol customization, whereas microfluidics-based methods often require adherence to predefined protocols and chip formats. Despite these drawbacks, microfluidic-based approaches maintain their relevance due to their user-friendly nature and widespread accessibility, with the major portion of the field adopting these approaches.

2.2.2 Single-cell DNA sequencing

The completion of HGP and the development of NGS methods also led to the development of larger scale DNA sequencing methods, sequencing large amounts of DNA quickly and cost-effectively. While NGS technologies have been successfully applied to various genomic tasks, their application to *de novo* genome assembly has presented certain challenges. First, NGS methods generate short read lengths, typically ranging from 100 to 300 base pairs, which can make the assembly of complex genomes difficult. Additionally, sequencing errors and biases introduced during library preparation and sequencing can affect the accuracy of the final assembly. In contrast, early studies using NGS methods to resequence bacterial genomes demonstrated how useful even very short reads are, given a reference genome to which to map them (Shendure *et al*., 2015). These studies demonstrated how performing DNA-sequencing in NGS platforms could easily be used to identify small mutational changes such as single nucleotide variations (SNV), and suggested that they could be used for identifying larger structural variants as well, such as copy number variations (CNV).

Performing single-cell whole genome sequencing (WGS) is challenging due to the minute amounts of DNA contained in a single cell. In order to obtain sufficient material for sequencing many WGS

methods rely on whole genome amplification (WGA) prior to library preparation. Historically, WGA methods have been associated with challenges such as amplification biases, amplification errors and chimeric sequences. However, through the refinement of existing techniques and the development of new approaches, multiple methods have been designed for amplifying genomic material for sequencing. Each method offers its distinct advantages and drawbacks, allowing the user to select the most appropriate method to fulfill their specific goals.

Degenerate oligonucleotide primed PCR (DOP-PCR) was developed in the early 90s with the aim to amplify small amounts of genomic DNA (Telenius *et al*., 1992). In this method degenerate primers are used, which contain random nucleotides at specific positions, allowing the primers to bind to multiple target sequences in a promiscuous manner. DOP-PCR starts with several pre-amplification cycles at a reduced initial annealing temperature, which promotes random primer binding. The pre-amplified DNA fragments are then subjected to further PCR amplification at higher temperatures. As DOP-PCR relies on PCR amplification of a small portion of the genome its use is mainly limited to determining large copy number variations. Additionally, uneven coverage and a high error rate pose challenges for calling SNVs.

To overcome some of the limitations of DOP-PCR multiple displacement amplification (MDA) was developed. MDA uses isothermal random priming and extension with a Φ29 polymerase, which has high processivity, low error rate, and strand displacement activity (Dean *et al*., 2001). MDA produces a greater genome coverage and a lower error rate, but still suffers from overrepresentation of loci that are amplified first, with greater exponential amplification exacerbating this effect (Bourcy *et al*., 2014). As a result, MDA is more suitable for calling SNVs than DOP-PCR but may still pose challenges for calling CNVs.

More recently, methods building on MDA have been developed to reduce amplification biases while controlling error rates. These methods use quasi-linear amplification and include multiple annealing and looping-based amplification cycles (MALBAC) and primary template-directed amplification (PTA). MALBAC employs a pool of random primers, each with a common sequence and 8 variable nucleotides capable of hybridizing to the template at low temperatures (Zong *et al*., 2012). DNA polymerases with strand-displacement activity are then used to generate semiamplicons of varying lengths. Semiamplicons are subsequently amplified, producing full amplicons with complementary ends, which are looped to prevent further amplification. Finally, exponential PCR is performed on full amplicons, generating large amounts of DNA for sequencing. Meanwhile, PTA makes use of the Φ29 polymerase, but incorporates exonuclease-resistant terminators, creating smaller double-stranded amplification products that subsequently undergo limited amplification (Gonzalez-Pena *et al*., 2021). As this process leads to a bias towards amplification of the primary template, the propagation of errors in daughter amplicons is limited.

The challenge of producing cheap libraries of whole genomes from single cells has also led to the development of methods where amplification is reduced or avoided altogether. Transposase-based methods rely either on the direct library preparation (DLP) of unamplified genomic DNA, or linear amplification via transposon insertion (LIANTI). In DLP, extracted DNA is incubated with a transposase enzyme that carries sequencing adapters. The transposase cleaves the double-stranded DNA and simultaneously integrates the adapters into the cut sites. After tagmentation, a limited number of PCR cycles are performed to amplify the library and incorporate sample-specific barcodes (Zahn *et al*., 2017). Similarly, LIANTI is based on the direct tagmentation of genomic DNA, but uses a transposase loaded with adapters containing a T7 promoter (Chen *et al*., 2017). This enables the linear reverse transcription of fragments into self-priming RNAs, which are subsequently reverse transcribed into DNA, before RNAse digestion, second strand synthesis with a unique molecular barcode, and final library preparation.

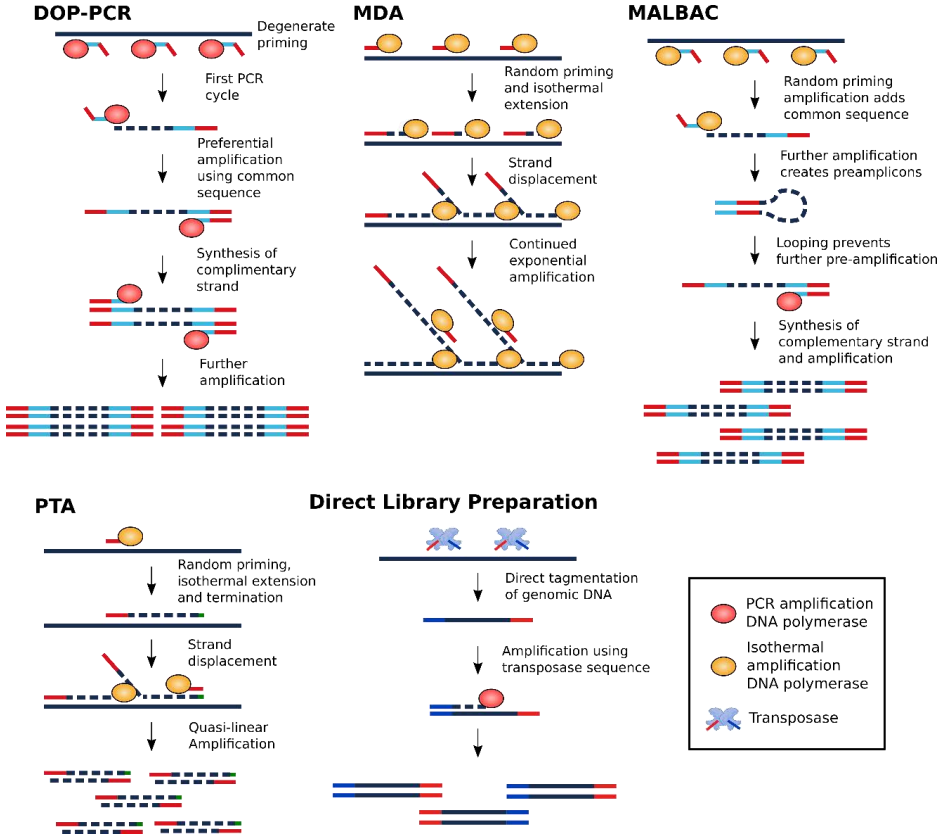


Figure 3 Examples of strategies for whole genome amplification of single cells. *Earlier methods such as DOP-PCR and MDA suffer from lower coverage or higher error rates, compared to newer methods such as MALBAC, PTA, and DLP.*

While most WGA and WGS methods aim to produce libraries with high and even coverage, the choice of method is still dependent upon which specific goals are prioritized, with different methods having their own strengths and weaknesses. As the earliest methods, DOP-PCR and MDA are largely obsolete, suffering from uneven coverage and amplification biases. Methods like MALBAC and PTA combine features of PCR-based and isothermal amplification methods to reduce amplification biases and improve coverage uniformity. Tagmentation-based methods such as LIANTI or DLP offer improved coverage uniformity, reduced amplification biases, and minimized costs (See **Fig. 3** for an overview).

The improvement of WGA methods that have occurred in the last ten years have allowed for the detection of genetic variations and mutations in low-input samples, such as single cells, which was previously challenging or impossible.

2.2.3 Epigenetic profiling using single-cell sequencing

Parallel to the development of scRNA-seq and WGA methods, the development of NGS techniques allowed for the development of novel techniques aiming to perform epigenetic profiling of cells by using sequencing as a final readout, many of which have subsequently been adapted for use in single cells. DNase-seq and ATAC-seq were developed in 1981 and 2013, respectively, and both rely on a similar mechanism (McGhee *et al*., 1981; Buenrostro *et al*., 2013). By treating intact nuclei with DNase or transposase, accessible chromatin is processed, while dense chromatin is inaccessible for the enzymes. The resulting fragments can then be processed for sequencing. While DNase-seq and ATAC-seq produce highly similar results, ATAC-seq confers the advantage that fragments are cut and ligated in one step, simplifying protocols for library preparation (Karabacak Calviello *et al*., 2019). Other methods include bisulfite sequencing, developed in 1992 (Frommer *et al*., 1992), revealing DNA methylation patterns, and Hi-C, developed in 2009 (Lieberman-Aiden *et al*., 2009), elucidating the three-dimensional architecture of the genome.

In later years, epigenetic profiling methods have moved from use on bulk samples to single cells, enabling the study of complex and dynamic regulation of gene expression at the individual cell level. Single-cell ATAC-seq, which captures open chromatin regions, provides insights into active regulatory elements. This high-throughput method requires relatively low input material but provides limited information about specific histone modifications and has lower coverage compared to bulk ATAC-seq. Despite this, single-cell scATAC-seq has seen wide use, being implemented in droplet-based format, and well formats with a high throughput to compensate for data sparsity. This has led to the publication of several studies producing vast atlases of accessible chromatin on a single-cell level (Cusanovich *et al*., 2018), and revealing tissue and cell-type specific regulatory elements (Preissl *et al*., 2018).

While the adaptation of ATAC-seq to a single-cell format has been largely successful, the transition of other epigenome profiling methods to a single-cell scale has proven difficult. Single-cell bisulfite sequencing requires substantial sequencing depth, but harsh reaction conditions result in substantial DNA degradation, compromising the quality of the final libraries. Similarly, the single-cell variant of Hi-C, suffers from significant limitations as it generates sparse data and high levels of technical noise, limiting its broader application in the field.

2.3 Single-cell multiomics

While the majority of single-cell research has relied on techniques measuring either DNA, RNA, or chromatin accessibility (see previous sections), several methods have emerged where two or more modalities are analyzed in the same cell. These multimodal methods enable the analysis of how modalities such as regulatory element accessibility, mutational burden, cell interactions, or tissue architecture act to affect gene transcription and cell states.

2.3.1 Joint DNA and RNA sequencing of single cells

The first method that attempted to analyze both mRNA and genomic DNA (gDNA) from the same cell was gDNA-mRNA sequencing (DR-seq). DR-seq relies on the reverse transcription of mRNA to single stranded DNA and the subsequent amplification of both gDNA and cDNA in the same reaction based on MALBAC (Dey *et al*., 2015). The product is split into two, and gDNA and cDNA are specifically processed in two separate reactions before sequencing. In genome and transcriptome sequencing (G&T seq), cells are lysed and mRNAs are physically separated from gDNA using biotinylated dT-primers, before Smart-seq2 processing of mRNA, and MDA-based WGS of gDNA (Macaulay *et al*., 2015). While these early methods provided a proof of concept, they did not see widespread adoption, due to suffering from technical challenges limiting throughput and increasing costs, and suffering technical biases associated with the WGA methods they applied.

With direct nuclear tagmentation and RNA-sequencing (DNTR-seq), we developed a highly sensitive method, basing the protocol on the physical separation of nuclear and cytosolic compartments into separate multi-well plates directly after cell lysis (Zachariadis *et al*., 2020). This improved the quality of scRNA-seq data compared to bead separation, or co-processing of gDNA and mRNA in the same reaction. Additionally, as the name implies, DNTR-seq applied direct tagmentation of gDNA prior to amplification, leading to reduced amplification bias and lower cost.

2.3.2 Joint epigenetic profiling and RNA sequencing

Several methods have been developed in order to sequence the genome or accessible regulatory regions together with RNA expression in the same cell, and have been applied to multiple tissue types, with multiple approaches achieving this goal. Split and pool approaches such as Paired-seq (parallel analysis of individual cells for RNA expression and DNA accessibility by sequencing) and sci-CAR (single-cell combinatorial indexing chromatin accessibility and mRNA) are based on multiple rounds of combinatorial indexing, tagging both the open chromatin fragments and cDNA molecules (Cao *et al*., 2018; Zhu *et al*., 2019). This enables the sequencing of hundreds of thousands to millions of cells at low cost but produces low sensitivity libraries and a degree of cross contamination among cells. Other methods combine single nucleus RNA sequencing (snRNA-seq) and ATAC sequencing by performing indexing in droplets, limiting cross-contamination, but limit RNA read-out nuclear RNAs (Chen *et al*., 2019). In addition to these methods, we have recently developed Smart3-ATAC, based on the physical separation of compartments from a cell and subsequent scRNA-seq and scATAC-seq (Cheng *et al*., 2021). This approach is more time consuming and lower through-put but generates higher quality data.

2.3.3 Spatial sequencing

In addition to methods attempting to read out multiple modalities from the same cell, spatial sequencing methods have recently gained much attention (Marx *et al*., 2021). These methods involve *in situ* sequencing or imaging-based approaches, keeping the tissue intact during the measurement process. Through these techniques, gene expression data can be mapped back to their original location in the tissue, providing insights into the organization and structure of tissues, the interactions between different cell types, and the local effects of the tissue environment on gene expression. While the work in this thesis does not focus on these types of methods, recent developments within the field demand that some of these methods are at least mentioned.

Spatial transcriptomics (ST) was one of the first spatial transcriptomics methods developed and relies on the *in situ* capturing of mRNA (Ståhl *et al*., 2016). Here, tissue sections are placed on a barcoded array, with each spot on the array capturing RNA from the section of tissue above it. Following sequencing, each read can then be traced back to its original location in the tissue. Multiplexed Error-Robust Fluorescence In Situ Hybridization (MERFISH) is an example of a method based on fluorescent in situ hybridization (Moffitt *et al*., 2016). This method relies on examining the individual RNA molecules with fluorescent probes to enable simultaneous detection of many different RNA species in their native context within cells. Additionally, methods such as Slide-seq exist, based on the *in situ* capturing of mRNA using barcoded micro beads (Rodriques *et al*., 2019). *in situ* capturing methods have recently become widely available with the introduction of 10X Visium, offering high resolution capture of 1-10 cells per barcode, and have recently been expanded to measure the full transcriptome, including non-polyadenylated transcripts (McKellar *et al*., 2023). The methods mentioned here are merely scraping the surface of the field of spatial sequencing technologies, which are not only limited to RNA-seq, but today also cover DNA-seq, ATAC-seq, and even multimodal readout (Zhao *et al*., 2022; Deng *et al*., 2022; Zhang *et al*., 2023). For a more complete review of spatial sequencing methods, their strengths and weaknesses, and applications, see Moses and Pachter, 2022.

In addition to spatial transcriptomics methods, alternatives using conventional single-cell sequencing have been developed to obtain spatial information from single-cell transcriptomics data. These methods rely on the dissociation of tissues into multiplets, clusters of two or more cells, and subsequently deconvolving cell type composition of the multiplet transcriptional profiles. While these methods do not provide direct information on higher-order tissue structures as do conventional spatial sequencing approaches, they do provide the benefit of high quality, full transcriptome data, and are capable of measuring the effects of direct cell interactions. Such approaches have recently been successfully applied in a semi-supervised manner to predetermined pairs of cell types, as well as to hand-picked doublets (Boisset *et al*., 2018; Giladi *et al*., 2020). However, these methods are not without drawbacks, being either limited to interacting cell types with known cell surface markers or requiring laborious microdissection of cells, making high throughput studies impossible. To circumvent these limitations, we recently developed a method where both the number of cells and the constituent cell types for each multiplet is derived directly from the data (Andrews *et al*., 2021). This method confers the advantage that it can be used in an unsupervised manner and is compatible with high throughput droplet-based techniques, making possible the study of cell-interactions in complex tissues.

2.3.4 Combining CRISPR screens with single-cell sequencing

In addition to the development of single-cell multiomics, the last decade has seen the rise of large scale CRISPR-screens. Initial CRISPR methods made use of wild type Cas9 endonucleases, capable of inducing double stranded breaks at target sites and enabling gene editing studies. Since then, the development of the catalytically inactive "dead" Cas9 (dCas9) has further expanded the use of CRISPR studies to the investigation of epigenetic regulatory mechanisms. dCas9 lacks the endonuclease capability of Cas9, but is still guided to target sequences by gRNAs. By linking dCas9 to repressor complexes or transcriptional activators, targeted silencing or activation of regulatory elements such as promoters and enhancers can be achieved.

To maximize the potential of CRISPR screens, multiple methods have been developed capable of reading out gRNAs together with transcriptomic or ATAC-seq read out. This includes PERTURB-seq, CITE/ECCITE-seq, and CROP-seq for RNA-seq, and CRISPR-sciATAC and Spear-ATAC for ATAC-seq (Datlinger *et al*., 2017; Dixit *et al*., 2016; Mimitou *et al*., 2019; Pierce *et al*., 2021; Liscovitch-Brauer *et al*., 2021). Such methods have drastically extended the utility of CRISPR-based screening, allowing the interrogation of all genes in the genome (Replogle *et al*., 2022), and large-scale studies of regulatory elements (Gasperini *et al*., 2019).

2.3.5 Insights into transcriptional regulation gained from single-cell sequencing

Single-cell sequencing technologies have significantly advanced biological and medical research, providing a tool to investigate cellular diversity and functionality in high detail and throughput. These methods have revealed insights into the heterogeneity of tissues in health and disease, highlighting expression patterns across cell subpopulations and revealing rare subpopulations (Haber *et al*., 2017; Ximerakis *et al*., 2019). Additionally, single-cell sequencing technologies have provided insights into the mechanisms underlying transcriptional regulation. In the sections below, I will discuss a variety of such studies and highlight findings that relied on the development of single-cell sequencing techniques.

In a healthy aging cell, there is an accumulation of somatic mutations over time due to intrinsic factors like errors during DNA replication and extrinsic factors such as exposure to environmental mutagens. This mutational burden can contribute to cellular dysfunction and the development of age-related diseases. As somatic mutations occurring in post-mitotic cells are difficult to study, because they cannot be clonally expanded, these processes are best studied using single-cell sequencing. Such approaches have been attempted using both scDNA and scRNA-seq techniques on cells from healthy aging individuals. These approaches have revealed mutational signatures and associated them with cellular phenotypes, quantified the accumulation of cancer-causing mutations in healthy tissues, and compared the rates and patterns of mutations in the stem cells of different tissues (Martincorena *et al*., 2015; Blokzijl *et al*., 2016; Enge *et al*., 2017). By revealing the types and frequencies of mutations that occur in healthy cells as we age, these studies may help us better understand the initial steps in the development of age-related diseases.

Single-cell sequencing has also shed light on the mechanisms controlling transcription, revealing how regulatory elements act to control transcription. Transcriptional bursting refers to the phenomenon where gene transcription occurs in bursts or pulses, rather than at a steady rate. This bursting behavior is thought to be a major source of variability in gene expression across a population of cells. Single-cell sequencing techniques have been instrumental in studying this phenomenon and have enabled the determination of transcriptome-wide burst frequencies and sizes, revealing that enhancers tend to regulate burst frequencies while burst size is defined in core promoter motifs (Larsson *et al*., 2019; Larsson *et al*., 2021). Concordantly, other studies using scRNA-seq have revealed that enhancers predominantly regulate gene expression via transcription initiation, rather than via release of PolII pausing (Larke *et al*., 2021).

Single-cell sequencing techniques have also had massive implications for cancer research, enabling a deeper understanding of the genetic and epigenetic variations within a tumor, as well as allowing remarkable discoveries relating to the TME. This has led to better understanding of tumor evolution, metastasis, and treatment resistance. Much as is has for healthy tissues, single-cell sequencing has been used to elucidate the cellular heterogeneity within tumors and their microenvironment, identifying distinct cellular subpopulations that contribute to tumor progression and therapeutic resistance (Tirosh *et al*., 2016). Additionally, single-cell sequencing technologies have been used to profile the immune cells within tumors, providing insights into anti-tumor immune responses and informing the development of immunotherapies (Legut *et al*., 2022). scRNA-seq has for instance been used to profile immune cells in breast cancer, revealing a diversity of immune cell states and interactions that could be targeted to enhance immunotherapy (Azizi *et al*., 2018). Moreover, scDNA methods have been used to profile cells of the TME, surprisingly revealing CNV:s and clonal expansion of non-malignant stromal cells of the TME (Zhou *et al*., 2020).

2.4 Transcriptional regulation and the cell state

2.4.1 PolII and the core promoter

The most basic form of transcriptional regulation is transcriptional initiation through the binding of RNA polymerase II (PolII) to the transcription start site (TSS) - the first nucleotide of a transcript. The process of TSS identification is a tightly regulated process, preventing promiscuous PolII binding and the transcription of non-functional transcripts. The core promoter, the sequence within 50 base pairs of the TSS, contains binding sites for PolII and general transcription factors (TFs) which together form the pre-initiation complex. Several commonly occurring motifs of the core promoter have been identified, including the TATA box and initiator, as well as other patterns such as the TFIIB recognition element and the downstream core promoter element (Smale and Kadonaga, 2003). Core promoter motifs constitute a mode by which the genetic code directly regulates gene expression, affecting which TFs can promote transcription. As such, it is unsurprising that the perturbation of core promoter motifs can significantly disrupt transcriptional activity (Juven-Gershon, 2008; Parry *et al*., 2010; Kadonaga, 2012).

2.4.2 Regulatory elements

While the core promoter and pre-initiation complex are sufficient to identify the TSS and initiate transcription, PolII activity is further modified by *cis*-regulatory elements, including promoters and enhancers (Shlyueva *et al*., 2014; Spitz and Furlong, 2012; Segert *et al*., 2021). A universal feature of active *cis*-regulatory elements is nucleosome displacement through TF binding, forming nucleosome deprived regions (NDR), where the underlying DNA is accessible. While promoters are easily identifiable due to their proximity to the TSS, NDRs have been important for identifying enhancer regions. In nuclease-based enhancer screens, nucleases such as Dnase I are used to digest accessible regions of the genome, indicating the location of an active regulatory element, which can subsequently be identified through sequencing (Meuleman *et al*., 2020).

The promoter region of a gene is proximal to the TSS, typically extending to 100-1000 base pairs, and contains specific binding sites for TFs. TF binding to these sites activates transcription by recruiting co-activators or through direct interaction with PolII to influence recruitment, initiation, and elongation (Vaquerizas *et al*., 2009). Compared to regions directly adjacent to the TSS, extended promoter regions display a lower degree of sequence similarity, with motifs varying among promoter regions of different genes. This reflects the nature of their function; extended promoter regions enable differential gene expression, as they contain DNA sequences which interact with TFs in a specific manner, thus achieving cell type specific responses to the activation of signaling pathways.

Enhancers are regulatory elements which activate transcription through the recruitment of transcription factors to the TSS. As opposed to promoters, enhancers are located more distally from the TSS and require DNA folding to bring attached TFs to a position where they can affect transcriptional activity. Enhancers are believed to promote gene transcription by acting in concert with other regulatory elements to increase the frequency of gene bursting (Larsson *et al*., 2019). The precise mechanism of how enhancers contribute to transcriptional activation has remained elusive. It has been proposed that enhancers are brought in contact with the TSS in a stable and specific manner. More recently, a phase separation model has been proposed, where regulatory elements bound to TFs fold in a semi-structured manner to form multi-molecular assemblies which compartmentalize and provide a regulatory mechanism (Hnisz *et al*., 2017). Enhancers distinguish themselves further from promoters by their cell type specificity. Large scale studies examining genome accessibility across cell types indicate that enhancers are major regulators of cell states, determining cell type specific gene expression through the integration of external signals (Heinz *et al*., 2010; Meuleman *et al*., 2020).

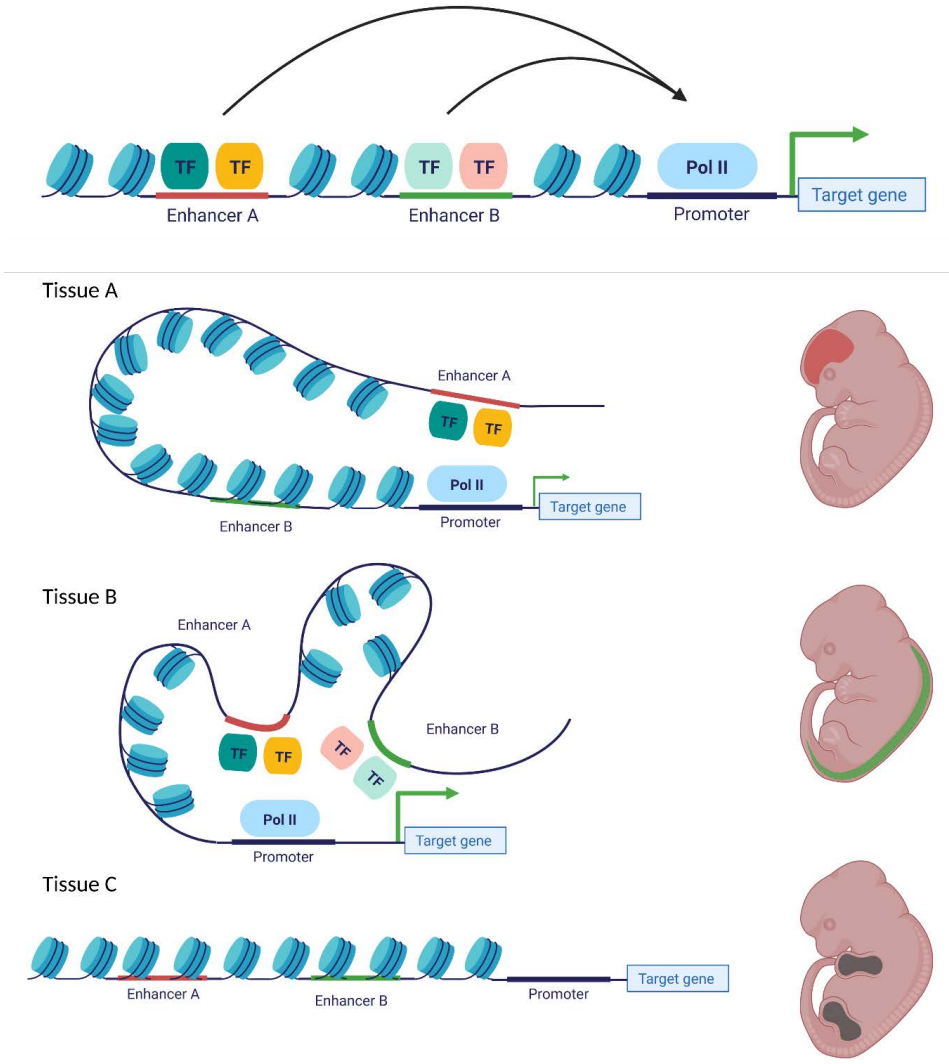


Figure 4: Proposed mechanism by which enhancers confer tissue specific gene expression. *Enhancers located upstream of their target gene are rendered inaccessible by tissue specific chromatin condensation. This proposed mechanism provides an efficient way to control gene expression in different tissues and as a response to different stimuli.*

Despite their notable differences, enhancers and promoters exhibit several shared characteristics. For instance, both types of regulatory elements are located in NDRs that can bind PolII, and frequently share similar sequence structures (Andersson *et al*., 2014a). It is therefore unsurprising that both promoters and enhancers have been observed to perform dual functions, with promoters activating gene expression at distant TSSs and enhancers driving transcription near their own loci (Andersson *et al*., 2014b; Dao *et al*., 2017; Diao *et al*., 2017). The structural and functional similarity between enhancers and promoters has led to attempts at redefining regulatory elements.

In a proposed model, regulatory elements should not be considered a dichotomy of promoters and enhancers, but rather similar structures that perform regulatory functions at varying degrees. In this model a regulatory element should have the following properties: It should be contained within a NDR, which is bound by TFs; it should have some degree of promoter activity, and it may have some degree of enhancer activity (Andersson *et al*., 2020). While such a model provides descriptive characteristics and serves as a reminder that promoter and enhancer-like functions are not mutually exclusive, the commonly used definition where promoters reside at sites where stable transcripts are produced while enhancers regulate transcription from more distal loci, remains a more practical model.

2.4.3 Topologically associating domains

As enhancers relay transcriptional instructions over large distances, there must be a mechanism to adjoin enhancers and promoters appropriately. Chromatin topology, the three-dimensional folding of DNA into topologies, has been suggested as a mode by which enhancer-promoter interactions are regulated. Topologically associating domains (TADs), self-interacting genomic regions, constitute a structural regulatory environment, acting by localizing enhancers to appropriate promoters, while insulating promoters from enhancers located in other TADs (Szabo *et al*., 2019). The importance of TADs has been displayed in studies where inversions at TAD borders lead to aberrant transcriptional activity and disease (Lupiáñez *et al*., 2015). Conversely, TAD disruption in other contexts induce small or partial effects, indicating that the transcriptional regulation of many genes is resistant to TAD disruption (Ghavi-Helm *et al*., 2019; Nora *et al*., 2017), and that additional mechanisms are likely important in enhancer-mediated gene regulation.

2.5 Transcriptional regulation and cell states in cancer

While gene expression is tightly regulated in healthy cells, malignant cells are characterized by increased transcriptional entropy, leading to a larger degree of variation in cell states. This variation is caused by genetic alterations, both in non-coding and protein coding regions of the genome, as well as dysregulation of the epigenetic machinery. Population diversity, in combination with heritable traits and selective pressure, sets the stage for the process of clonal selection, where cells containing beneficial mutations and epigenetic modifications will survive and expand.

2.5.1 Mutations in coding regions

The accumulation of mutations in cancer cells occurs non-randomly, since beneficial mutations affecting cellular programs such as survival, proliferation, and differentiation lead to the preferential expansion of clones harboring them. Historically, a large focus has been placed on mutations affecting the protein coding regions of tumor suppressors and oncogenes. Tumor suppressors tend to be rendered non-functional through insertions, deletions, or nonsense mutations while oncogenes are rendered hyperactive through amplifications, activating point mutations, or translocations leading to hyperactive fusion proteins or increased expression through promoter hijacking. The effects of such mutations can be dramatic, with only a few being sufficient for cancer initiation, and certain cancers are driven by persistently activated signaling pathways as a result of mutated proteins (Look, 1997; Weinstein and Joe, 2006).

The identification of highly impactful mutations has led to the development of treatment strategies targeting specific oncogenes, many of which are used clinically today. Prominent examples include Trastuzumab, targeting HER2 in subsets of breast cancer; Vemurafenib, inhibiting MAPK-signaling in melanoma; and Imatinib, a Tyrosine kinase inhibitor used in the treatment of subsets of leukemias (Hughes *et al*., 2003; Piccart-Gebhart *et al*., 2005; Bollag *et al*., 2010).

Molecular targeting of oncogenic proteins should be considered a success story, leading to drastically improved prognosis for certain cancer types. Despite this, for most cases such approaches have proven unsuccessful. This is due to the fact that most cancers are not the result of a few lesions in highly impactful regions, but involve a large number of perturbations, many of which occur in regulatory portions of the genome. The importance of regulatory elements as drivers of cancer has been highlighted through genome wide association studies (GWAS), showing that the majority of genetic variants associated with cancer risk are located in non-coding regions, and are further enriched in promoters and enhancers (Schaub *et al*., 2012; Maurano *et al*., 2014; He *et al*., 2015; Dunning *et al*., 2016). Additionally, many cancers display elevated mutation rates at TF-binding loci, especially in promoter regions of oncogenes such as *BCL2*, *RBM5* and *WWOX* (Smith *et al*., 2015; Katainen *et al*., 2015).

2.5.2 Non-coding mutations in cancer

Despite the fact that mutations occur in a stochastic manner across cell types, most genes involved are enriched in or specific to distinct cancer-types (Carroll *et al*., 1996; Venkitaraman *et al*., 2002; Bader *et al*., 2006). This suggests that oncogenes require certain conditions to drive cancer. As gene regulatory enhancer activity is highly cell type specific, it has been proposed that cell type specificity provides these conditions with the presence of accessible, poised enhancers near oncogenes and tumor suppressors determining whether or not an activating mutation confers oncogenic activity (Sur and Taipale, 2016; Mertens *et al*., 2015). Additionally, non-coding regions of the genome have been implicated in malignancy through genome-wide association studies (GWAS), revealing that the majority of variants associated with an increased cancer risk are located within non-coding regions, with a significant number situated in candidate cis-regulatory elements (cCREs) (Sud *et al*., 2017), observations which have been corroborated in studies performing WGS of tumors (Wang *et al*., 2014). Taken together, these results emphasize the importance of developing a deeper understanding of the role of non-coding regions in driving the emergence and progression of cancer.

Insights into specific sites have shed light on how such mutations can contribute to cancer progression. For instance, mutations upstream of the TAL1 gene identified in a subset of T-cell acute lymphoblastic leukemias lead to the creation of *de novo* binding sites for the transcription factor MYB, resulting in TAL1 overexpression, a potent driver of tumorigenesis (Mansour *et al*., 2014). Additionally, chromosomal rearrangements have been observed which shift regulatory elements closer to oncogenes, thereby contributing to cancer development. A specific case is found in a subset of leukemias, where a chromosomal inversion moves the enhancer of the GATA2 gene near to the promoter of the adjacent stem-cell regulator EVI1. This leads to EVI1 overexpression and GATA2 haploinsufficiency, exacerbating disease progression (Gröschel *et al*., 2014).

Despite these examples, the study of non-coding mutations and their role in cancer has been historically challenging due to the sheer volume of potential candidates, the absence of suitable tools, and difficulties in examining mutations within the appropriate cellular context. Recent technological advancements, such as the development of massive parallel reporter assays, and large scale CRISPRi-screens, have been used to better address these challenges, signifying a promising shift in the field of cancer genomics (Abell *et al*., 2022; Morris *et al*., 2023).

2.5.3 Acute lymphoblastic leukemias and Pancreatic ductal adenocarcinoma

Pancreatic ductal adenocarcinoma (PDAC) and acute lymphoblastic leukemia (ALL) represent two distinct types of cancer that demonstrate the diversity of cancer biology and the need for individualized approaches. PDAC, a highly aggressive and often lethal solid tumor, arises from the pancreatic ducts and is characterized by rapid progression, late diagnosis, and poor response to treatment. On the other hand, ALL is a blood cancer that originates in the bone marrow, primarily affecting children. Despite its aggressive nature, ALL has seen significant improvements in treatment outcomes in recent years, with survival rates now exceeding 90% for pediatric patients (Hunger and Mullighan, 2015).

The genetic and epigenetic landscapes of PDAC and ALL are quite different, reflecting their distinct tissue origins and pathogenetic mechanisms. PDAC is notorious for its complex genomic alterations, often involving key driver genes. While mutations frequently occur in KRAS, TP53, CDKN2A, and SMAD4, treatment of PDAC presents significant challenges due to the diverse and complex molecular subtypes of this disease. Different subtypes of PDAC are characterized by distinct molecular signatures, often including unique or rare sets of gene mutations (Waddell *et al*., 2015; Bailey *et al*., 2016). These molecular distinctions affect disease progression, responses to treatment, and overall patient prognosis, and often result in significant heterogeneity within the tumor, which might limit the effectiveness of targeted therapies.

On the other hand, ALL is characterized by recurrent chromosomal translocations and focal gene mutations, often involving genes essential for lymphoid development and differentiation. A significant example is the ETV6-RUNX1, caused by a translocation between chromosomes 12 and 21, which leads to the formation of a chimeric transcription factor (Mullighan *et al*., 2007). Such mutations, although destructive, are generally fewer and more predictable than those found in PDAC. This relatively lower level of molecular complexity in ALL, combined with the sensitivity of leukemic cells to chemotherapy, contributes to the generally high cure rates in pediatric ALL. Treatment protocols have been optimized over the years using risk-stratification based on certain molecular and cytogenetic features, enabling high-dose, targeted chemotherapy for high-risk patients, while reducing toxicity for others (Lee *et al*., 2023).

The contrasting features of PDAC and ALL highlight the importance of understanding the specific molecular mechanisms driving each cancer type, as well as the need for single-cell sequencing and improved methods to study the complex interplay between genetic, epigenetic, and transcriptional changes within individual cancer cells. By gaining a deeper understanding of the molecular basis of these cancers, more effective, targeted therapies tailored to the unique characteristics of each disease can be developed.

2.6 Cellular niches and physical interactions in health and disease

2.6.1 Maintenance of tissue homeostasis

The external environment, encompassing the cellular microenvironment and cell-cell interactions within specialized niches, is vital to cellular function and homeostasis. One of the best-studied cellular niches is the epithelial crypts of the small intestine (Allaire *et al*., 2018). The stem cells at the base of these crypts are responsible for the continuous regeneration of the intestinal epithelium, the most rapidly self-renewing tissue in adult mammals. A pivotal actor in this process are Paneth cells, a type of secretory cell intermingled among the stem cells at the crypt base, which has a function in the microbial defense. Paneth cells also provide an important function in supporting the stem cells by providing Wnt, Notch, and other growth factors, creating a niche conducive to stem cell maintenance and function. This process is essential to maintain stem cell identity and to balance self-renewal with differentiation, and the disruption of this niche or its signaling can lead to disorders including cancer.

Another example is the white pulp of the spleen (Mebius *et al*., 2005). The white pulp is structured into zones, with the periarteriolar lymphoid sheath (PALS) primarily housing T cells, and follicles that predominantly contain B cells. The marginal zone, which separates the white pulp from the red pulp, contains antigen-presenting cells, including macrophages and dendritic cells. Dendritic cells, which are adept at capturing, processing, and presenting antigens, play a pivotal role, presenting antigens directly to T cells B cells, initiating adaptive immune responses. The spleen's unique structure and cell composition make it a vital organ for initiating immune responses, particularly against encapsulated bacteria. Removal of the spleen can lead to an increased risk of infections, especially those caused by encapsulated bacteria.

2.6.2 The tumor microenvironment

As these examples illustrate, the effects of cell interactions are critical to tissue function, regulating processes such as regeneration, proliferation and survival. As such, it is unsurprising that perturbations to cellular niches are frequently observed across multiple cancer types, often producing favorable microenvironments for cancer cells. Tumor microenvironments (TME) consist of a variety of cell types, including both cancer cells and non-cancer cells such as immune and stromal cells, as well as other cell types of the primary organ. While clearly implicated in tumor progression and metastasis, cell interactions with their complex natures and heterogeneity between TMEs have made studying cell interactions and their effects in cancer challenging. This challenge is exacerbated by the fact that cells in the TME can either promote or inhibit tumor growth, depending on the specific context in which they act.

Cancer-associated fibroblasts (CAFs) are one of the most abundant cell types in the TME and have been implicated in cancer progression and drug resistance (Kalluri, 2016). CAFs contribute to tumor growth and progression through several mechanisms. CAFs secrete factors such as vascular endothelial growth factor (VEGF) and fibroblast growth factor (FGF), promoting angiogenesis, and supplying nutrients to the growing tumor. The chemokines produced by CAFs also attract a variety of immune cells, including regulatory T cells (Tregs) and myeloid-derived suppressor cells (MDSC), which contribute to immune evasion by suppressing anti-tumor immune responses. CAFs also act by the remodeling of the extracellular matrix (ECM), facilitating tumor growth and metastasis and modulating the TME to promote therapy resistance (Öhlund *et al*., 2014).

While immune cells are pivotal for the early detection and removal of aberrant cells - as illustrated by the increased cancer prevalence in immunocompromised patients (Robison *et al*., 1987; Hayward *et al*., 1997; Shiels *et al*., 2011) - in later stages of cancer, immune cells are often found to mediate effects beneficial to tumor growth. The failure of immune cells to clear cancer cells can lead to a state of chronic inflammation, supporting cancer growth. This is evident in multiple solid tumors, where the degree and cellular composition of immune infiltrate impacts prognosis (Zhang *et al*., 2003; Ribatti *et al*., 2003; Galon *et al*., 2006; Mahmoud *et al*., 2011; Chang *et al*., 2011). Furthermore, interactions between cancer cells and immune cells of the TME can become reciprocal over time, with tumor cells shaping the surrounding immune environment into one promoting growth and tissue remodeling. This is achieved by attracting and expanding supportive immune cells through the secretion of chemokines, suppressing immune mediated cell death, and promoting tissue repair and remodeling (Kohli *et al*., 2021).

2.6.3 Targeting the tumor microenvironment

As the importance of the TME in cancer progression has become apparent, strategies targeting these cells have been devised. Multiple strategies targeting immune cells have been attempted, including the inhibition of immune cell migration and checkpoint inhibition of cytotoxic T-cells, the latter of which has resulted in clinically approved therapies for several cancer types (Qian *et al*., 2011; Rotte 2019). Other therapeutic strategies targeting supportive cell types have shown mixed results. For instance, anti-angiogenic therapies through the antagonism of VEGF signaling are frequently circumvented by compensatory mechanisms, such as an increased capacity of cancer cells to migrate (Bergers and Hanahan, 2008). Similarly, while *in vitro* and *in vivo* models indicate therapeutic potential of targeting stromal cells, clinical trials attempting such methods have had limited success (Sherman *et al*., 2014; Jiang *et al*., 2020). Taken together, these mixed results reflect the complex nature of cell-cell interactions in cancer, and it is clear that a better understanding is required in order to properly utilize therapies targeting supportive cell types.

2.7 Future perspectives

The value of NGS and single-cell sequencing techniques is already clear and has already led to a better understanding of cell states and the gene regulatory mechanisms underlying them, with several studies indicating that these methods may be informative for prognostic purposes (Tirosh *et al*., 2018; Cohen *et al*., 2021). Additionally, while a large number of multimodal methods have been developed, improved study designs will certainly provide additional information, revealing how different cell states determine cell fate in health and disease (Zhou *et al*., 2020; Emert *et al*., 2021).

As recent technical advances have led to an avalanche of new data, a future challenge will be how to best utilize this knowledge to improve treatments and provide better patient outcomes. It is clear that cancer is a diverse class of diseases, differing not only between patients and types, but in many cases showing a great degree of heterogeneity among cancer cells and clones within the same patient. Such a degree of variation provides two challenges: Firstly, as cancer progresses, the dysregulation of programs controlling genome integrity and transcription lead to the accumulation of mutations and epigenetic changes, and discerning drivers of disease from passengers presents a challenge in itself. Secondly, a large degree of diversity provides a base for adaptation and thus therapy evasion, as therapies will affect cells unevenly, allowing cells with a greater capacity for survival to expand. For these reasons it will be important to not only identify targetable traits important for cancer survival and growth, but also to predict and target features capable of rescuing and reconstituting cancer following treatment.

It is becoming increasingly clear that the prevalence of rare clones, nuanced cell states, and TME composition can have a profound impact on patient outcome and risk of relapse. While current diagnostic methods provide limited information on such details, single-cell techniques can discern such features, providing an avenue for improved risk prediction and better decision making regarding therapeutic approaches. While standardization, reduced cost, and increased availability will be required, it seems likely that these methods will see use in the clinic in the near future, bringing us closer to personalized therapeutics.

3 Research aims

The overall aim of this thesis was to develop novel single-cell sequencing methods capable of capturing multiple modalities. While our initial focus was to limit the scope to the study of clonal structures (combined RNA and DNA from a single cell) and cell interactions (spatial inference from incompletely dissociated cells) in cancer, after the initial success of our first two methods we decided to pursue studying regulatory elements and other non-coding regions of the genome. The accomplishment of this is described in the last two, currently unpublished manuscripts, where we develop SMART3-ATAC, for the joint profiling of accessible chromatin and full-length RNA transcripts, and ACTI-seq, where we adapt this for use with CRISPRi screening.

In **Paper I**, we aimed to develop a method capable of extending the utility of single-cell RNA-seq methods to add a layer of spatial information, allowing the user to identify cell types that were more prone to physically interact with each other. Importantly, we wanted this to be a method that could be used in an unsupervised manner, allowed for high throughput, and was compatible with the majority of available single-cell RNA-seq methods.

The aim of **Paper II** was to develop a method that would allow us to jointly profile DNA and RNA from the same cell. Specifically, it was important that we could establish a method where WGS library preparation was uncomplicated in order to reduce cost and allow us to perform sequencing on many cells. We wished to implement this method to study how copy-number variations affect gene expression following copy number gain or loss.

For **Paper III** we aimed to build on our success from paper III, and produce a method capable of producing high-quality ATAC-seq and RNA-seq data from the same cell. Additionally, we chose to elevate the choice of RNA-seq method in order to provide the highest quality of single-cell ATAC-seq and RNA-seq data possible.

Paper IV built further on paper III. Here we aimed to use SMART3-ATAC in combination with CRISPRi screening. We wished to establish this method in order to study pancreatic ductal adenocarcinoma, characterizing regulatory elements involved in maintaining growth and proliferation, and identifying transcription factors binding to these elements in order to regulate transcription.

4 Ethical considerations

In this thesis, we were faced with three primary ethical considerations: the utilization of human pediatric cancer samples for DNA and RNA sequencing, the use of murine animal models, and the employment of patient-derived cell lines. Of these, the use of primary pediatric samples emerges as the most ethically complex.

At the forefront of these complexities is informed consent. This involves comprehensive communication with the parents or guardians, as our samples are typically procured from very young children incapable of fully comprehending the scope of our work. We ensured that all involved parties understand the purpose, process, potential risks, benefits, and outcomes of the study, and are aware of their right to withdraw participation at any time. Equally important is the benefit-risk assessment. By only using surplus samples, we obviate the need for additional invasive procedures, thereby avoiding the infliction of undue distress or harm on our pediatric patients. However, the sensitive nature of these samples, especially considering they are pediatric, warrants extreme caution in privacy and confidentiality measures. Genomic information, while invaluable for research, can reveal sensitive insights into an individual's health, potential diseases, and familial relationships. As such, we ensured stringent anonymization of data and secure storage to prevent the dissemination of private information.

Animal experimentation, our second ethical domain, is fraught with its own challenges. In this study, we used animals specifically bred for research that were scheduled for culling. We adhered strictly to the principle of the 3Rs – Replacement, Reduction, and Refinement – which advocates for the substitution of animal models with alternatives where possible, the minimization of animals used, and the refinement of experimental procedures to lessen discomfort or harm.

Finally, we turned to the use of patient-derived cell lines. Though this poses fewer ethical dilemmas than human samples or animal models, the origins of these cell lines and their informed consent status call for careful attention. Issues like cell line misidentification or contamination also warrant consideration. In this vein, it's important to note that some cell lines, historically obtained under standards that wouldn't meet modern expectations of informed consent, may present ethical quandaries.

In conclusion, while the recent advances in sequencing techniques hold great promise for decoding biological and disease mechanisms, performing large scale sequencing studies on primary tissues from patients comes with its own set of ethical dilemmas. For the research performed in this thesis, these dilemmas have had our ongoing attention and we have taken all the steps possible to ensure the privacy of patients, and that patients have donated samples under informed consent.

5 Results and discussion

5.1 Paper I

An Unsupervised Method for Physical Cell Interaction Profiling of Complex Tissues

To develop a single-cell RNA-seq method capable of adding a level of spatial information, we set out to design a general, high-throughput, and unsupervised method based on multiplet deconvolution to infer cell type interactions in healthy and diseased tissues. scRNA-seq methods generally rely on dissociating cells into single-cell suspensions, with incompletely dissociated cells (multiplets) being produced as a by-product. To design the method called cell interaction by multiplet sequencing (CIM-seq), we decided to make use of this heterogeneous suspension of multiplets to determine cells which are physically interacting with each other in intact tissues.

By preparing RNA-seq libraries from cell multiplets, we obtained transcriptional profiles representing a mix of unknown quantities of cells from the tissue. These quantities could then be approximated *in silico* by combining scRNA-seq data from constituent cell types. In other words, multiplet profiles can be computationally deconvoluted into fractional contributions of single cells, using a set of transcriptional profiles representing all observed cell types in the tissue and estimating the number of cells that constitutes each multiplet. This process is performed in three stages: 1) partial dissociation of the target tissue, cell sorting to separate singlets from multiplets, and scRNA-seq to obtain transcriptional profiles 2) the generation of a blueprint of all cell types and states in the tissue and 3) computational deconvolution, using maximum-likelihood estimation to determine each multiplet's cell-type composition based on the blueprint.

We benchmarked CIM-seq on a series of tissues and artificial multiplets, using both plate-based (Smart-seq2) and droplet based (Chromium Single Cell 3' v.3) methods. As a proof of concept, we first performed CIM-seq on three distinct cell lines (A375, HCT116, HOS) sorted as singlets or multiplets of a known composition. By performing the deconvolution in this controlled setting, we showed that CIM-seq displayed a high level of correspondence between expected and detected connections (<5% average error rate). We subsequently used CIM-seq to identify known structural features of the small intestinal epithelium, spleen, and lung. We used plate-based methods for the small intestinal epithelium and spleen, while 10x was used for the lung, and a repeated run of the small intestinal epithelium was used in order for us to compare results between different methodologies. In all three of these tissues CIM-seq was able to identify cell interactions such as Paneth-stem cell interactions in the base of the intestinal crypt, and interactions between endothelial cells and type II pneumocytes in the pulmonary capillary endothelium. Importantly, when using droplet-based methods with a higher throughput, CIM-seq was also able to detect more subtle enrichments, such as a preference for goblet cells to reside in crypts of the small intestine.

Finally, we used CIM-seq to characterize cell structures in colonic crypts. At the time of the study, it had been established that the colonic epithelium shares a similar crypt structure to that of the small intestine, with Lgr5+ stem cells located at the base of the crypt. However, the equivalent of Paneth cells, responsible for Wnt ligands in the small intestine, had not been identified in the colon. By performing CIM-seq on colonic epithelium we revealed a structure highly similar to that of the small

intestine, but with an increased complexity with regards to goblet cell types and stem cell types. A subset of goblet cells expressed the wound healing marker *Plet1* and interacted strongly with stem cells expressing high levels of *Lgr5*.

To assess whether *Plet1*-expressing goblet cells filled the function of Paneth cells in the colon, we looked at the expression of genes involved in the Wnt signaling pathway. From our scRNA-seq data we could see that Wnt ligands were largely absent from the colonic epithelium with microscopy confirming epithelial Wnt3 expression in small intestinal crypts, while Wnt2b was found in the stroma of both tissues. We could therefore conclude that none of the epithelial cell types in the colon provide canonical Wnt ligands, and that colonic stem cells are supported by a stromal source of Wnt ligands.

In summary, CIM-seq provides an innovative method that leverages multiplet deconvolution to decipher cell-cell interactions in various tissues. By utilizing a previously considered by-product of scRNA-seq, the method provides insights into cellular architecture and interactions in an efficient and cost-effective manner. The proof-of-concept and benchmarking experiments demonstrated the accuracy and versatility of CIM-seq in a variety of tissue types and experimental platforms. Notably, the application of CIM-seq to colonic crypts revealed new insights into the cellular composition and Wnt ligand sources in the colon, highlighting the potential of this approach for advancing our understanding of tissue biology.

5.2 Paper II

A Highly Scalable Method for Joint Whole-Genome Sequencing and Gene-Expression Profiling of Single Cells

To investigate the impact of genetic variation on gene expression in complex cellular mixtures, we developed Direct Nuclear Tagmentation and RNA sequencing (DNTR-seq), a technique that enables the simultaneous sequencing of whole-genome DNA and full-length mRNA in single cells. In order to facilitate widespread adoption of the method, we prioritized four key features: (1) High fidelity, entailing avoiding lossy operations like DNA cleanup and minimizing PCR amplification cycles for genomic sequencing, while striving for high-sensitivity detection of full-length transcripts in mRNA-seq. (2) Individually addressable cells, allowing for sequencing cells at variable depths. (3) Ease of adaptation, precluding the need for non-standard equipment. (4) Minimal positional bias of WGS, facilitating the identification of CNV:s at ultra-low coverage and reducing the sequencing required to achieve desired coverage.

DNTR is a single-well protocol based on direct tagmentation using tn5 transposase. Nuclear proteins are disintegrated by snap freezing and subsequent protease digestion, with the free genomic DNA (gDNA) tagmented prior to PCR amplification. This process is performed without intermediate DNA cleanup, making it highly efficient, and is easily automated. We benchmarked the WGS element of DNTR-seq, by analyzing ultra-low coverage sequencing data from ALL patients and cancer cell lines, including a few cells which were re-sequenced at a higher coverage. At ultra-low coverage duplicate rates were generally low (below 10%), increasing at higher read depths. We also showed that DNTR-seq results in gDNA libraries with much more even coverage and with a lower GC bias than traditional methods such as MALBAC and MDA, and is equivalent to amplification-free bulk sequencing. Additionally, we showed how DNTR-seq can be used to perform SNV calling by performing analysis on groups of clonal cells, leveraging data from multiple cells to identify errors introduced during amplification.

The mRNA sequencing protocol was based on Smart-seq2, providing high-sensitivity and full-length scRNA-seq data. In order to verify that RNA-seq on cytosolic compartment leads to high-quality data, we compared DNTR-seq to conventional Smart-Seq2 in tumor cell lines A375 and HCT116, as well as an ALL-patient sample. We observed a similar number of total read counts and number of detected genes using the two methods, though the number of genes detected was slightly lower in cell lines analyzed using DNTR-seq. We found that Intronic regions, found in unspliced transcripts residing in the nucleus, were less than half as common in DNTR-seq data compared to Smart-Seq2, while exonic reads, were enriched in DNTR-seq data, highlighting the effects of separating the cytosolic fraction from the nucleus.

We subsequently analyzed 607 cells using DNTR-seq, sampled from two pediatric acute lymphoblastic leukemia (ALL) cases, the human colon adenocarcinoma cell line HCT116, and melanoma cell line A375. The results showed that DNTR-seq is capable of producing both high quality gDNA and RNA-seq data from the same cell, with unsupervised classification copy numbers yielding distinct clusters representing variant genomes, and each cell line and leukemic blast clone separating into their own cluster. Additionally, healthy ALL cells from patient samples clustered

according to XX and XY genomes as expected. For the transcriptional data, our two ALL patient samples formed clusters separated by patient, while normal B, T and progenitor cells instead clustered by cell type. Furthermore, the gDNA data produced was of such high quality that we could identify specific subclones within ALL samples, based on CNV at specific sites, and tie these variations to differences in gene expression in genes within the affected regions. These experiments highlighted the complementary features of the two modalities, with joint analysis demonstrating how DNTR-seq can be used to quantify the impact of genomic deletions and amplifications on gene expression.

Finally, we conducted a whole-genome screen to evaluate the impact of CNVs on the expression of affected genes across the genome. To generate a cell population with a highly diverse copy-number profile, we exposed the colon adenocarcinoma cell line HCT116 to two DNA-damaging agents: X-ray ionizing irradiation (X-ray) or etoposide (ETO). We prepared WGS libraries from over 3,000 cells treated with increasing doses of DNA-damaging agents and collected cells at two timepoints (48 or 96 h). Specific transcriptional changes largely depended on DNA damage burden, with upregulated DNA damage pathways and downregulated cell-cycle pathways. Conversely, differential expression analysis between ETO and X-ray conditions with similar DNA damage levels showed only minor differences in a small number of genes.

To assess how genetic dosage effects contribute to transcriptional heterogeneity, we analyzed transcript abundance during copy-number alterations. Most expressed genes were affected by copy-number alterations in an approximately linear fashion, suggesting that random copy-number changes have a specific genetic dosage effect on transcription. However, some genes demonstrated efficient feedback regulation, offsetting the gene dosage effect. Notably, several genes crucial for cancer cell growth, displayed strong dosage compensation and were mostly unaffected by copy-number alterations. Similarly, periodically activated genes, like those involved in mitosis, tended to exhibit robust dosage compensation, consistent with these genes being under strict transcriptional control. Concordantly, genes displaying a strong dosage compensation were generally less subject to purifying selection.

In summary, DNTR-seq enabled us to conduct a large-scale in vitro screen of DNA-damage-induced copy-number alterations. Its flexible study design, stemming from the simplicity and low cost of the WGS analysis, makes DNTR-seq an ideal method for performing extensive screens at low sequencing depth, pinpointing cells of interest for resequencing of gDNA at increased depth and scRNA-seq for in-depth analysis.

5.3 Paper III

Smart3-ATAC: A Highly Sensitive Method for Joint Accessibility and Full-length Transcriptome Analysis in Single Cells

Understanding transcriptional regulation is crucial for deciphering the biology of organisms. To develop a method that can provide high sensitivity measurements of both chromatin accessibility and mRNA expression we designed Smart3-ATAC, a method separating cytosolic and nuclear compartments, before performing scRNA-seq and ATAC-seq on the respective compartments. Smart3-ATAC implements a novel, low-loss, single-cell ATAC (scATAC) protocol to measure chromatin accessibility, while transcriptomic profiling is based on the highly sensitive Smart-seq3 protocol.

To test the performance of Smart3-ATAC on a complex biological process we used a stem-cell based in vitro model for mouse gastrulation. We analyzed over 3000 cells harvested at four different time-points (0hr, 72hr, 96hr and 120hr). Smart3-ATAC generated scATAC libraries had a similar specificity as contemporary scATAC methods, but with a higher yield when compared to commercially available multiomics methods such as 10x genomics multiome. Furthermore, by performing scRNA-seq on cytosolic mRNA rather than nuclear mRNA, we obtained higher quality transcriptional data, as observed when comparing mRNA-seq from Smart3-ATAC with 3' UMI counts from 10x multiomic analysis, where Smart3-ATAC yields 10-fold higher counts.

We displayed the power of leveraging transcriptomic data against chromatin accessibility by calculating the correlation between the accessibility of non-coding regulatory elements and the transcript abundance of nearby genes. We found an enrichment of associations between accessibility near the TSS of genes, with 48% of high confidence cis-peaks occurring within 100bp of their associated TSS, indicating modulation of transcription occurring directly at the promoter. However, many genes completely lacked correlations between promoter accessibility and transcript abundance, suggesting that transcriptional regulation is mainly driven by distal enhancers.

Taken together, our data shows that Smart3-ATAC is a powerful method for the joint analysis of mRNA and chromatin accessibility from single cells, focusing on obtaining the highest-quality data rather than maximizing the number of cells assayed. This approach offers numerous advantages, such as individually addressable cells, enabling multi-tiered experiment setups and enhanced enhancer-gene association analysis. The increased per-cell cost and decreased throughput may limit its use in larger scale projects, but for questions involving limited cell numbers, Smart3-ATAC offers a more cost-effective and insightful solution.

5.4 Paper IV

Simultaneous Sequencing of Full-length RNA Transcripts, Accessible Chromatin, and Guide RNAs for Isoform-sensitive CRISPR Perturbation Analysis

As we had previously established a method for the joint profiling of ATAC-seq and RNA-seq data from a single cell and demonstrated the utility of this method when interrogating regulatory elements, we wished to expand the utility further, combining this method with CRISPRi screening. We developed accessible chromatin and transcriptome sequencing of CRISPR-inhibited cells (ACTI-seq) by modifying Smart3-ATAC, applying a CROP-seq based approach for CRISPRi screening with a modified modified dT primer to include an 8-bp barcode, enabling the pooling and specific PCR of guide-RNAs.

We evaluated ACTI-seq through a series of experiments assessing our ability to generate single-cell RNA and ATAC-seq data and testing the sensitivity of gRNA readouts. To evaluate scRNA-seq results, we compared data obtained using ACTI-seq with that from Smart-seq2 on cytosolic compartments of HPAC cells. We observed a similar number of detected genes per cell for both methods, indicating that our modifications to the dT primer do not significantly affect our ability to generate high quality libraries. Likewise, ACTI-seq yielded high-quality ATAC-seq data, with an 8-fold TSS-enrichment and the fraction of reads in peak (FriP) over 0.6.

To identify potential targets for ACTI-seq, we analyzed three PDAC-derived cell lines in the absence of any perturbations using Smart3-ATAC. From the ATAC-seq data, we called over 40 000 accessible peaks from the three cell lines. By correlating chromatin accessibility with expression levels of nearby genes, we identified more than 3000 candidate CREs. To further assess these candidate CREs, we performed a conventional CRISPRi screen in one of the cell lines (HPAC), targeting all candidate CREs associated with the expression of at least one gene, accessible peaks near PDAC-associated loci or overlapping structural CTCF motifs, and a selection randomly sampled accessible regions. Additionally, we targeted the promoters of all TFs in the genome as a positive control. As expected, the largest effects were observed when targeting promoters of TFs, with few exonic, intronic, or intergenic accessible peaks mediating any significant effects.

Next, we selected 56 targets focusing on TFs with strong negative effects on cell growth in the HPAC cell line, performing a CRISPRi screen and using ACTI-seq as a read out. We selected 56 targets and used 8 non-targeting guides as negative controls. By transducing at a low MOI (~0.4), we saw that the majority of cells expressed one unique gRNA. Excluding targets with very low representation (< 10 cells), we observed an inverse correlation between the number of cells captured targeting a specific TF and the number of differentially expressed genes (DEG) between these cells. By analyzing associations between peak accessibility and gene expression, we identified 545 peaks associated with the expression of 362 genes, with KAT7 and KLF5 having the largest number of peaks significantly associated with differentially expressed genes, based on scRNA-seq data.

To further assess KLF5 binding to associated peaks, we performed a motif enrichment analysis using monaLisa using motifs from the JASPAR2020 database (see methods). A total of 75 motifs were enriched in KLF5-associated peaks with KLF-family motifs being the most highly enriched, indicating binding sites for KLF5. As KAT7 lacks a specific binding motif, we were unable to identify specific

binding sites. Instead, we looked at broad conformational changes to chromatin structures as a result of KAT7 silencing, and examined which TFs were predominantly affected. We identified 18 TF-associated motifs, which could be classified into 12 categories based on motif similarity with top enriched motifs being associated with SNAI1, -2, and -3 and other E-box binding TFs. As these proteins are implicated in epithelial-mesenchymal transition we can speculate that KAT7 is involved in regulating cell motility and polarity in HPAC cells.

In summary, ACTI-seq is a method enabling the simultaneous sequencing of RNA, ATAC-seq data, and gRNAs from single cells. ACTI-seq has broad applications across diverse cell types and can be used as a tool for deepening our understanding of complex regulatory networks governing cellular identity and function, and their roles in malignancy. While our study employed a simple in vitro model, the approach is also adaptable to animal or cancer xenograft models, particularly benefiting systems with limited sample availability.

6 Conclusions and future perspectives

In **paper I** we developed a novel method - cell interaction by multiplet sequencing (CIM-seq) - to obtain spatial information from single-cell RNA-seq data by making use of multiplets produced as a by-product during cell dissociation. CIM-seq enables the identification of physically interacting cells in intact tissues by preparing RNA-seq libraries from cell multiplets and computationally deconvoluting them into fractional contributions of single cells. The process involves partial tissue dissociation, cell sorting, scRNA-seq profiling, generation of a blueprint of all cell types and states in the tissue, and computational deconvolution using maximum-likelihood estimation. The proof-of-concept and benchmarking experiments demonstrated the accuracy and versatility of CIM-seq in a variety of tissue types and experimental platforms, with our application of CIM-seq to colonic crypts revealing insights into their cellular composition.

CIM-seq has the potential to significantly impact the field by providing an innovative approach for deciphering cell-cell interactions in various tissues. Its ability to utilize something previously regarded as a by-product of scRNA-seq allows for the efficient and cost-effective study of cellular architecture and interactions, offering a cheaper alternative to conventional spatial transcriptomics methods. Additionally, CIM-seq offers the full-length whole-transcriptome readout of all cell types in a tissue, eliminating the need for prior knowledge of cell type composition or the pre-selection of genes. This innovative approach will be particularly useful for studying cell type interactions in diseased tissues or other tissues where tissue structure is unknown. By offering a cost-effective and efficient way to explore cellular architecture and interactions, CIM-seq has the potential to significantly advance our understanding of tissue biology and uncover previously unknown cellular interactions and structures.

In **paper II** we developed Direct Nuclear Tagmentation and RNA sequencing (DNTR-seq), a method that allows for simultaneous sequencing of whole-genome DNA and full-length mRNA in single cells. We demonstrated DNTR-seq's ability to produce high-quality genomic DNA and RNA-seq data from the same cell while effectively identifying subclones within ALL samples and linking gene expression differences to genomic variations. DNTR offers a flexible study design due to its simplicity, low cost, and capacity to perform extensive screens at low sequencing depth. This makes DNTR-seq ideal for pinpointing cells of interest for the resequencing of genomic DNA at increased depth and single-cell RNA-seq for in-depth analysis.

As there are currently no commercial methods available for the joint analysis of RNA and genomic DNA on a single cell, DNTR-seq addresses a need within the scientific community, offering a powerful tool for comprehensive analysis of cellular gene expression and genomic variation. Additionally, DNTR-seq outperforms other alternatives by providing higher quality data and a more even coverage. The potential applications for DNTR-seq are vast, with the ability to study not only cancer but also a wide range of other diseases where genomic differences may impact transcription. By enabling the in-depth investigation of genomic and transcriptomic changes within single cells, DNTR-seq can accelerate our understanding of the molecular mechanisms underlying various diseases, paving the way for the development of more effective diagnostic tools and targeted therapies.

For **paper III**, we developed a method for providing high sensitivity measurements of both chromatin accessibility and mRNA expression by separating cytosolic and nuclear compartments, called Smart3-ATAC. Smart3-ATAC utilizes a low-loss, single-cell ATAC (scATAC) protocol to measure chromatin accessibility, while transcriptomic profiling is based on the highly sensitive Smart-seq3 protocol. When tested on a stem-cell-based in vitro model for mouse gastrulation, Smart3-ATAC demonstrated a higher yield compared to commercially available multiomics methods such as 10x genomics' multiome kits and produced higher quality transcriptional data by performing scRNA-seq on cytosolic mRNA rather than nuclear mRNA.

The true power of Smart3-ATAC lies in its ability to leverage transcriptomic data against chromatin accessibility, allowing for the identification of associations between the accessibility of non-coding regulatory elements and transcript abundance of nearby genes. While the method's increased per-cell cost and decreased throughput may limit its use in larger scale projects, it offers a more cost-effective and insightful solution for questions involving limited cell numbers, particularly in the study of transcriptional regulation. The development of Smart3-ATAC has the potential to significantly advance our understanding of the biology of organisms and the intricacies of transcriptional regulation, offering a powerful tool for the joint analysis of mRNA and chromatin accessibility in single cells.

Finally, for **paper IV,** we expanded the utility of Smart3-ATAC by developing accessible chromatin and transcriptome sequencing of CRISPR-inhibited cells (ACTI-seq). ACTI-seq employs a CROP-seq-based approach for CRISPRi screening, using a modified dT primer containing a barcode for pooling and specific amplification of guide-RNAs in a cost-effective manner. Here, we demonstrated that ACTI-seq is capable of generating high-quality data and that even with a limited number of cells, we can use ACTI-seq to identify interacting components in gene regulatory pathways. Additionally, ACTI-seq is the first method to integrate ATAC-seq, RNA-seq, and gRNA readout from the same cell in a single assay, offering an unprecedented opportunity to provide a comprehensive view of complex regulatory networks under the effects of specific gene silencing.

For the projects described in this thesis, our primary focus was to develop and test innovative methods to enhance the current single-cell sequencing toolkit. We believe that the resulting methods will have wide-ranging implications for the scientific community. Together with collaborators we are currently employing CIM-seq to study pancreatic cancer liver metastases, identifying critical cell type interactions that drive tumor spread and growth. DNTR-seq has become a valuable tool in our lab, enabling us to analyze the clonal structures of pediatric leukemias and to catalog CNVs and their transcriptional effects in various cancer types.

While DNTR-seq and CIM-seq were primarily developed for studying tumor and cancer biology, the applications of Smart3-ATAC and ACTI-seq extend far beyond oncology. We are currently using Smart3-ATAC to investigate epigenetic events during early embryology and to identify key factors controlling cell fate in the small intestinal epithelium. As the most recent method developed, ACTI-seq is presently limited to studying regulatory elements and their interacting partners in PDACs. However, much like Smart3-ATAC, ACTI-seq has the potential to be instrumental in elucidating the regulatory machinery underlying a wide array of diseases and biological processes.

In conclusion, this thesis aimed to develop methods for investigating clonal structures and cell interactions in cancer, but the outcome surpassed our initial goals. The techniques presented here hold immense potential for advancing scientific discoveries and making a significant impact on our understanding of various biological processes.

7 Acknowledgements

I would like to express my deepest gratitude to all those who made this thesis possible.

Whenever I'm asked about what's most important when pursuing a PhD, I always highlight three things: a great supervisor, a supportive lab, and an engaging project, in that order. I consider myself incredibly fortunate to have had all three.

I am deeply grateful to my supervisor, **Martin Enge**. Meeting you at a time when I was on the brink of reevaluating my entire career could not have been luckier. Your innovative thinking, guidance, and constant support, especially during unexpected circumstances, have been truly invaluable. Our shared interests and sense of humor have made research not just rewarding, but also incredibly enjoyable.

A big thank you to **Vasilios**. I was always impressed by your willingness to take the time to help others, sharing your knowledge and solving problems. Although I sometimes (always) wish you would stop wasting your time in the clinic, whenever you were around for me to bug you, your help was invaluable. **Huaitao**, for your positive attitude and generosity, it's been a pleasure working alongside someone as hard working, and collaborative as you. **Sólrún**, you've been an amazing colleague, always there to lend a helping hand and share a laugh. Thank you for making our lab environment so enjoyable and enjoy being the senior PhD student when I'm gone. I also want to acknowledge **Jessica** for her unwavering work ethic. What you have achieved in the short time that you've been with us is impressive, and I know that you will achieve great things! **Jason**, to whom I owe a debt of gratitude for being the reason I joined this lab in the first place. We made a decent team during my first years in the lab. Finally, I'd like to acknowledge the students who I've had the great pleasure of supervising, **Caroline**, **Erdem,** and **Ewa**. **Ewa**, you deserve a special mention for putting in countless hours of hard work, meticulously setting up protocols that led to the success of our CIM-seq paper.

I am also grateful to my co-supervisor **Marco Gerling** and the entire team in his lab. **Marco**, your inherent German precision complemented with a generous dose of humor, has always made you a pleasure to work with. Additionally, working alongside **Natalie Geyer**, an outstanding colleague, has been a privilege. I only wish that circumstances had allowed us to work more together.

Additionally, I would be remiss not to acknowledge the long list of individuals whose contributions have been instrumental in making CIM-seq a reality. Notably, **Ninib Baryawno**, **Agneta B. Andersson**, **Iva Šutevski**, and **Rosan Heijboer** deserve special mention. I also want to extend my appreciation to the members of the LCI imaging facility in Flemingsberg, **Gabriela Imreh** and **Sylvie Le Guyader**, for their patient training and support.

I'd also like to extend my profound appreciation to the members of the CRISPR Functional Genomics facility for their essential contributions to the ACTI-seq project. The expertise of **Bernhard**, **Soniya**, and **Miriam** made sure this project was completed in time for this thesis, a feat that would not have been possible without their involvement. I would especially like to thank **Miriam**, who patiently answered my relentless questions and took the time to meet with me to discuss optimal strategies for our screens.

I'd like to express heartfelt gratitude to the members of the Grandér Lab at CCK:

Firstly, I'd like to acknowledge **Danne**, who generously opened his lab to me. An inspiring presence, whose absence is deeply felt to this day. To **Katja**, whose constant helpfulness and kindness created an environment of warmth and collaboration. **Elin Edsbäcker**, **Matheus Dyczynski**, **Linda Viðarsdóttir**, **Yasmin Yu** made the times at CCK some of the most enjoyable I've experienced. I would also like to thank **Sander Busker** for these times, after all, acknowledgements don't cost much.

The importance of my past mentors, who've fueled my passion for science, can't be overstated. Without them, I may have strayed from this field long ago:

Firstly, a big thank you goes out to my old teacher, **Boe Löwbäck**. For sparking a scientific curiosity in me at a young age, which I've carried all the way through to choosing my education. It is not an exaggeration to say that I wouldn't be here if it wasn't for you. I hope you continue to inspire young minds. Also, I must thank **Ed Harding**, for making my first time in a lab an incredible experience. I learned a lot from you, not least of all that research can be fun, and that's something that's stuck with me. Finally, **Helene Rundqvist**. Though we've met infrequently over the past years, I have always appreciated your support and insightful advice.

My gratitude extends to the members of the Department of Oncology-Pathology as well. A huge thank you to **Angelos**, **Mireia**, **Amos**, **Aarren**, and **Stefanie**, who made our lunch breaks a delightful and uplifting part of the day. And of course, **Henri** and **Asimina**, whose positive presences in the lab made every workday more enjoyable.

No journey is complete without the companionship of good friends, and I've been fortunate to have exceptional people by my side. To my old Dungeons and Dragons group - **Kevin**, **Lucas**, and **Anton** - though our adventures in imaginary realms may have ended, our friendship endures. Having you all as friends during both the highs and lows has been invaluable. Venting, as they say, is crucial. A shout out to **Muyi Yang**, the infamous ladies' man. Your reliability and care make you a testament to what true friendship embodies. I'll see you in NYC. **Sara**, my friend from our university days. Our catch-ups, though infrequent, are always a pleasure, filled with chatter about ambitions and future plans. **Daniel** and **Quingyang**, crossing paths with you in Gotland was serendipitous, and I'm truly glad to call you friends. I also want to acknowledge my friends outside the realm of science - **Johan Lindqvist**, **Hjalmar Wide**, **Alexander Andersson**, and **Martin Aspe**. Although I might not always express it, having your perspectives, removed from the world of research, has been a breath of fresh air.

The last acknowledgements I want to make are for the most important people in my life - my family:

First off, my parents, **Ann-Katrin** and **Robert**, deserve a big thanks. You have both been the inspiration that pushed me towards a higher education degree. I've always considered myself exceptionally fortunate to have parents who I can always count on for support, something that has been especially apparent recently, when rushing to complete this thesis. I also want to thank my siblings, **Adam**, **Elias**, and **Mira**. Having you guys around has added immeasurable joy and laughter to my life.

To my son, **Noah**, your existence alone makes me want to do better and work harder.

And last but not least, my wife, **Nicole**. For providing a peaceful and very happy home, and so much more. There are so many things I wouldn't have been able to do without your help. Your support in my work and your faith in me have made all the difference.

8 References

Abell, N. S., *et al.* (2022). Multiple Causal Variants Underlie Genetic Associations in Humans. Science, 375(6586), 1247–1254.

Allaire, J. M., *et al.* (2018). The Intestinal Epithelium: Central Coordinator of Mucosal Immunity. Trends in Immunology, 39(9), 677–696.

Andersson, R. *et al.* (2014a). An Atlas of Active Enhancers across Human Cell Types and Tissues. Nature, 507(7493), 455-461.

Andersson, R. *et al.* (2014b). Nuclear Stability and Transcriptional Directionality Separate Functionally Distinct RNA Species. Nature Communications, 5, 5336.

Andersson, R. and Sandelin, A. (2020). Determinants of Enhancer and Promoter Activities of Regulatory Elements. Nature Reviews Genetics, 21(2), 71-87.

Andrews, N. *et al.* (2021). An Unsupervised Method for Physical Cell Interaction Profiling of Complex Tissues. Nature Methods, 18(8), 912-920.

Arslan, S., *et al.* (2022). Sequencing by Avidity Enables High Accuracy with Low Reagent Consumption. bioRxiv.

Azizi, E., *et al.* (2018). Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. Cell, 174(5), 1293-1308.e36.

Bailey, P., *et al.* (2016). Genomic Analyses Identify Molecular Subtypes of Pancreatic Cancer. Nature, 531(7592), 47–52.

Bader, A. G., *et al.* (2006). Cancer-Specific Mutations in PIK3CA Are Oncogenic in Vivo. Proceedings of the National Academy of Sciences, 103(5), 1475–1479.

Bergers, G. and Hanahan, D. (2008). Modes of Resistance to Anti-Angiogenic Therapy. Nature Reviews Cancer, 8(8), 592-603.

Blokzijl, F. *et al.* (2016). Tissue-Specific Mutation Accumulation in Human Adult Stem Cells during Life. Nature, 538(7624), 260–264.

Boisset, J.-C. *et al.* (2018). Mapping the Physical Network of Cellular Interactions. Nature Methods, 15(7), 547-553.

Bollag, G. *et al.* (2010). Clinical Efficacy of a RAF Inhibitor Needs Broad Target Blockade in BRAF-Mutant Melanoma. Nature, 467(7315), 596-599.

Bourcy, C. F. A. *et al.* (2014). A Quantitative Comparison of Single-Cell Whole Genome Amplification Methods. PLOS ONE, 9(8), e105585.

Braslavsky, I. *et al.* (2003). Sequence Information Can Be Obtained from Single DNA Molecules. Proceedings of the National Academy of Sciences, 100(7), 3960–3964.

Brenner, S. *et al*. (2000). Gene Expression Analysis by Massively Parallel Signature Sequencing (MPSS) on Microbead Arrays. Nature Biotechnology, 18(6), 630–634.

Buenrostro, J. D. *et al*. (2013). Transposition of Native Chromatin for Fast and Sensitive Epigenomic Profiling of Open Chromatin, DNA-Binding Proteins and Nucleosome Position. Nature Methods, 10(12), 1213–1218.

Cao, J. *et al*. (2018). Joint Profiling of Chromatin Accessibility and Gene Expression in Thousands of Single Cells. Science, 361(6409), 1380–1385.

Carroll, M. *et al*. (1996). The TEL/Platelet-Derived Growth Factor β Receptor (PDGFβR) Fusion in Chronic Myelomonocytic Leukemia Is a Transforming Protein That Self-Associates and Activates PDGFβR Kinase-Dependent Signaling Pathways. Proceedings of the National Academy of Sciences, 93(25), 14845–14850.

Chang, D. Z. *et al*. (2011). Mast Cells in Tumor Microenvironment Promotes the in Vivo Growth of Pancreatic Ductal Adenocarcinoma. Clinical Cancer Research: An Official Journal of the American Association for Cancer Research, 17(22), 7015-7023.

Check Hayden, E. (2014). Technology: The $1,000 Genome. Nature, 507(7492), 294–295.

Chen, C. *et al*. (2017). Single-Cell Whole-Genome Analyses by Linear Amplification via Transposon Insertion (LIANTI). Science, 356(6334), 189–194.

Chen, S. *et al*. (2019). High-Throughput Sequencing of the Transcriptome and Chromatin Accessibility in the Same Cell. Nature Biotechnology, 37(12), 1452–1457.

Cheng, H. *et al*. (2021). Smart3-ATAC: A Highly Sensitive Method for Joint Accessibility and Full-Length Transcriptome Analysis in Single Cells. bioRxiv.

Clark, I. C. *et al*. (2023). Microfluidics-Free Single-Cell Genomics with Templated Emulsification. Nature Biotechnology, March, 1–10.

Cohen, Y. C. *et al*. (2021). Identification of Resistance Pathways and Therapeutic Targets in Relapsed Multiple Myeloma Patients through Single-Cell Sequencing. Nature Medicine, 27(3), 491-503.

Craxton, M. (1991). Linear Amplification Sequencing, a Powerful Method for Sequencing DNA. Methods, 3(1), 20–26.

Cusanovich, D. A. *et al*. (2018). A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. Cell, 174(5), 1309-1324.e18.

Dao, L. T. M. *et al*. (2017). Genome-Wide Characterization of Mammalian Promoters with Distal Enhancer Functions. Nature Genetics, 49(7), 1073-1081.

Datlinger, P. *et al*. (2017). Pooled CRISPR Screening with Single-Cell Transcriptome Readout. Nature Methods, 14(3), 297–301.

Dey, S. S. *et al*. (2015). Integrated Genome and Transcriptome Sequencing of the Same Cell. Nature Biotechnology, 33(3), 285-289.

Dean, F. B. *et al*. (2001). Rapid Amplification of Plasmid and Phage DNA Using Phi29 DNA Polymerase and Multiply-Primed Rolling Circle Amplification. Genome Research, 11(6), 1095–1099.

Deng, Y. *et al*. (2022). Spatial Profiling of Chromatin Accessibility in Mouse and Human Tissues. Nature, 609(7926), 375–383.

Diao, Y. *et al*. (2017). A Tiling-Deletion-Based Genetic Screen for Cis-Regulatory Element Identification in Mammalian Cells. Nature Methods, 14(6), 629-635.

Dixit, A. *et al*. (2016). Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell, 167(7), 1853-1866.e17.

Drmanac, R. *et al*. (2010). Human Genome Sequencing Using Unchained Base Reads on Self-Assembling DNA Nanoarrays. Science, 327(5961), 78–81.

Dunning, A. M. *et al*. (2016). Breast Cancer Risk Variants at 6q25 Display Different Phenotype Associations and Regulate ESR1, RMND1 and CCDC170. Nature Genetics, 48(4), 374-386.

Emert, B. L. *et al*. (2021). Variability within Rare Cell States Enables Multiple Paths toward Drug Resistance. Nature Biotechnology.

Enge, M. *et al*. (2017). Single-Cell Analysis of Human Pancreas Reveals Transcriptional Signatures of Aging and Somatic Mutation Patterns. Cell, 171(2), 321-330.e14.

Eze, U. C. *et al*. (2021). Single-Cell Atlas of Early Human Brain Development Highlights Heterogeneity of Human Neuroepithelial Cells and Early Radial Glia. Nature Neuroscience, 24(4), 584-594.

Fleischmann, R. D. *et al*. (1995). Whole-Genome Random Sequencing and Assembly of Haemophilus Influenzae Rd. Science, 269(5223), 496–512.

Frommer, M. *et al*. (1992). A Genomic Sequencing Protocol That Yields a Positive Display of 5-Methylcytosine Residues in Individual DNA Strands. Proceedings of the National Academy of Sciences, 89(5), 1827–1831.

Galon, J. *et al*. (2006). Type, Density, and Location of Immune Cells Within Human Colorectal Tumors Predict Clinical Outcome. Science, 313(5795), 1960–1964.

Gasperini, M. *et al*. (2019). A Genome-Wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. Cell, 176(1), 377-390.e19.

Ghavi-Helm, Y. *et al*. (2019). Highly Rearranged Chromosomes Reveal Uncoupling between Genome Topology and Gene Expression. Nature Genetics, 51(8), 1272-1282.

Giladi, A. *et al*. (2020). Dissecting Cellular Crosstalk by Sequencing Physically Interacting Cells. Nature Biotechnology, 38(5), 629-637.

Gonzalez-Pena, V. *et al*. (2021). Accurate Genomic Variant Detection in Single Cells with Primary Template-Directed Amplification. Proceedings of the National Academy of Sciences, 118(24), e2024176118.

Gröschel, S. *et al*. (2014). A Single Oncogenic Enhancer Rearrangement Causes Concomitant EVI1 and GATA2 Deregulation in Leukemia. Cell, 157(2), 369–381.

Haber, A. L. *et al*. (2017). A Single-Cell Survey of the Small Intestinal Epithelium. Nature, 551(7680), 333–339.

Hagemann-Jensen, M. *et al*. (2020). Single-Cell RNA Counting at Allele and Isoform Resolution Using Smart-Seq3. Nature Biotechnology, 38(6), 708–714.

Hagemann-Jensen, M. *et al*. (2022). Scalable Single-Cell RNA Sequencing from Full Transcripts with Smart-Seq3xpress. Nature Biotechnology, 40(10), 1452–1457.

Hashimshony, T. *et al*. (2012). CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification. Cell Reports, 2(3), 666–673.

Hayward, A. R. *et al*. (1997). Cholangiopathy and Tumors of the Pancreas, Liver, and Biliary Tree in Boys with X-Linked Immunodeficiency with Hyper-IgM. The Journal of Immunology, 158(2), 977-983.

He, H. *et al*. (2015). Multiple Functional Variants in Long-Range Enhancer Elements Contribute to the Risk of SNP Rs965513 in Thyroid Cancer. Proceedings of the National Academy of Sciences of the United States of America, 112(19), 6128-6133.

Heinz, S. *et al*. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime Cis-Regulatory Elements Required for Macrophage and B Cell Identities. Molecular Cell, 38(4), 576-589.

Hnisz, D. *et al*. (2017). A Phase Separation Model for Transcriptional Control. Cell, 169(1), 13-23.

Hood, L., & Rowen, L. (2013). The Human Genome Project: Big Science Transforms Biology and Medicine. Genome Medicine, 5(9), 79.

Hughes, T. P. *et al*. (2003). Frequency of Major Molecular Responses to Imatinib or Interferon Alfa plus Cytarabine in Newly Diagnosed Chronic Myeloid Leukemia. New England Journal of Medicine, 349(15), 1423-1432.

Hunger, S. P., & Mullighan, C. G. (2015). Acute Lymphoblastic Leukemia in Children. New England Journal of Medicine, 373(16), 1541–1552.

Islam, S. *et al*. (2012). Highly Multiplexed and Strand-Specific Single-Cell RNA 5′ End Sequencing. Nature Protocols, 7(5), 813–828.

Jaitin, D. A. *et al*. (2014). Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types. Science, 343(6172), 776–779.

Jiang, B. *et al*. (2020). Stroma-Targeting Therapy in Pancreatic Cancer: One Coin With Two Sides? Frontiers in Oncology, 10.

Juven-Gershon, T., Hsu, J.-Y., and Kadonaga, J. T. (2008). Caudal, a Key Developmental Regulator, Is a DPE-Specific Transcriptional Factor. Genes & Development, 22(20), 2823-2830.

Kadonaga, J. T. (2012). Perspectives on the RNA Polymerase II Core Promoter. Wiley Interdisciplinary Reviews. Developmental Biology, 1(1), 40-51.

Kalluri, R. (2016). The Biology and Function of Fibroblasts in Cancer. Nature Reviews Cancer, 16(9), 582–598.

Karabacak Calviello, A. *et al*. (2019). Reproducible Inference of Transcription Factor Footprints in ATAC-Seq and DNase-Seq Datasets Using Protocol-Specific Bias Modeling. Genome Biology, 20(1), 42.

Katainen, R. *et al*. (2015). CTCF/Cohesin-Binding Sites Are Frequently Mutated in Cancer. Nature Genetics, 47(7), 818-821.

Kivioja, T. *et al*. (2012). Counting Absolute Numbers of Molecules Using Unique Molecular Identifiers. Nature Methods, 9(1), 72–74.

Kohli, K. *et al.* (2022). Key Chemokines Direct Migration of Immune Cells in Solid Tumors. Cancer Gene Therapy, 1-12.

Larke, M. S. C. *et al*. (2021). Enhancers Predominantly Regulate Gene Expression during Differentiation via Transcription Initiation. Molecular Cell, 81(5), 983-997.e7.

Larsson, A. J. M. *et al*. (2019). Genomic Encoding of Transcriptional Burst Kinetics. Nature, 565(7738), 251–254.

Larsson, A. J. M. *et al*. (2021). Transcriptional Bursts Explain Autosomal Random Monoallelic Expression and Affect Allelic Imbalance. PLOS Computational Biology, 17(3), e1008772.

Lee, S. H. R. *et al*. (2023). Pharmacotypes across the Genomic Landscape of Pediatric Acute Lymphoblastic Leukemia and Impact on Treatment Response. Nature Medicine, 29(1), 170–179.

Legut, M. *et al*. (2022). A Genome-Scale Screen for Synthetic Drivers of T Cell Proliferation. Nature, 603(7902), 728–735.

Li, J. H. *et al*. (2022). Low-Pass Sequencing plus Imputation Using Avidity Sequencing Displays Comparable Imputation Accuracy to Sequencing by Synthesis While Reducing Duplicates. bioRxiv.

Lieberman-Aiden, E. *et al*. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. Science, 326(5950), 289–293.

Liscovitch-Brauer, N. *et al*. (2021). Profiling the Genetic Determinants of Chromatin Accessibility with Scalable Single-Cell CRISPR Screens. Nature Biotechnology, 39(10), 1270–1277.

Look, A. T. (1997). Oncogenic Transcription Factors in the Human Acute Leukemias. Science, 278(5340), 1059-1064.

Lupiáñez, D. G. *et al*. (2015). Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. Cell, 161(5), 1012-1025.

Macaulay, I. C. *et al*. (2015). G&T-Seq: Parallel Sequencing of Single-Cell Genomes and Transcriptomes. Nature Methods, 12(6), 519-522

Macosko, E. Z. *et al*. (2015). Highly Parallel Genome-Wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell, 161(5), 1202–1214.

Mahmoud, S. M. A. *et al*. (2011). Tumor-Infiltrating CD8+ Lymphocytes Predict Clinical Outcome in Breast Cancer. Journal of Clinical Oncology, 29(15), 1949–1955.

Mansour, M. R. *et al*. (2014). An Oncogenic Super-Enhancer Formed through Somatic Mutation of a Noncoding Intergenic Element. Science, 346(6215), 1373–1377.

Martincorena, I. *et al*. (2015). High Burden and Pervasive Positive Selection of Somatic Mutations in Normal Human Skin. Science, 348(6237), 880–886.

Mardis, E. R. (2008). Next-Generation DNA Sequencing Methods. Annual Review of Genomics and Human Genetics, 9(1), 387–402.

Margulies, M. *et al*. (2005). Genome Sequencing in Microfabricated High-Density Picolitre Reactors. Nature, 437(7057), 376–380.

Marx, V. (2021). Method of the Year: Spatially Resolved Transcriptomics. Nature Methods, 18(1), 9–14.

Maurano, M. T. *et al*. (2012). Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. Science (New York, N.Y.), 337(6099), 1190-1195.

Maxam, A. M. *et al*. (1977). A New Method for Sequencing DNA. Proceedings of the National Academy of Sciences, 74(2), 560–564.

McGhee, J. D. *et al*. (1981). A 200 Base Pair Region at the 5′ End of the Chicken Adult β-Globin Gene Is Accessible to Nuclease Digestion. Cell, 27(1, Part 2), 45–55.

McKellar, D. W. *et al*. (2023). Spatial Mapping of the Total Transcriptome by in Situ Polyadenylation. Nature Biotechnology, 41(4), 513–520.

McKernan, K. J. *et al*. (2009). Sequence and Structural Variation in a Human Genome Uncovered by Short-Read, Massively Parallel Ligation Sequencing Using Two-Base Encoding. Genome Research, 19(9), 1527–1541.

Mebius, R. E., & Kraal, G. (2005). Structure and Function of the Spleen. Nature Reviews Immunology, 5(8), 606–616.

Mertens, F. *et al*. (2015). The Emerging Complexity of Gene Fusions in Cancer. Nature Reviews. Cancer, 15(6), 371-381.

Meuleman, W. *et al*. (2020). Index and Biological Spectrum of Human DNase I Hypersensitive Sites. Nature, 584(7820), 244-251.

Miki, Y. *et al*. (1994). A Strong Candidate for the Breast and Ovarian Cancer Susceptibility Gene BRCA1. Science, 266(5182), 66–71.

Mimitou, E. P. *et al*. (2019). Multiplexed Detection of Proteins, Transcriptomes, Clonotypes and CRISPR Perturbations in Single Cells. Nature Methods, 16(5), 409–412.

Moffitt, J. R. *et al*. (2016). High-Throughput Single-Cell Gene-Expression Profiling with Multiplexed Error-Robust Fluorescence in Situ Hybridization. Proceedings of the National Academy of Sciences, 113(39), 11046–11051.

Morris, J. A. *et al*. (2023). Discovery of Target Genes and Pathways at GWAS Loci by Pooled Single-Cell CRISPR Screens. Science, 0(0), eadh7699.

Moses, L., & Pachter, L. (2022). Museum of Spatial Transcriptomics. Nature Methods, 19(5), 534–546.

Mullighan, C. G. *et al*. (2007). Genome-Wide Analysis of Genetic Alterations in Acute Lymphoblastic Leukaemia. Nature, 446(7137), 758–764.

Nora, E. P. *et al*. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell, 169(5), 930-944.

Parry, T. J. *et al*. (2010). The TCT Motif, a Key Component of an RNA Polymerase II Transcription System for the Translational Machinery. Genes & Development, 24(18), 2013-2018.

Piccart-Gebhart, M. J. *et al*. (2005). Trastuzumab after Adjuvant Chemotherapy in HER2-Positive Breast Cancer. New England Journal of Medicine, 353(16), 1659-1672.

Picelli, S. *et al*. (2013). Smart-Seq2 for Sensitive Full-Length Transcriptome Profiling in Single Cells. Nature Methods, 10(11), 1096–1098.

Pierce, S. E. *et al*. (2021). High-Throughput Single-Cell Chromatin Accessibility CRISPR Screens Enable Unbiased Identification of Regulatory Networks in Cancer. Nature Communications, 12(1), 2969.

Preissl, S. *et al*. (2018). Single-Nucleus Analysis of Accessible Chromatin in Developing Mouse Forebrain Reveals Cell-Type-Specific Transcriptional Regulation. Nature Neuroscience, 21(3), 432–439.

Qian, B.-Z. *et al*. (2011). CCL2 Recruits Inflammatory Monocytes to Facilitate Breast-Tumour Metastasis. Nature, 475(7355), 222-225.

Ramsköld, D. *et al*. (2012). Full-Length MRNA-Seq from Single-Cell Levels of RNA and Individual Circulating Tumor Cells. Nature Biotechnology 30 (8): 777–82.

Ratner, L. *et al*. (1985). Complete Nucleotide Sequence of the AIDS Virus, HTLV-III. Nature, 313(6000), 277–284.

Replogle, J. M. *et al*. (2022). Mapping Information-Rich Genotype-Phenotype Landscapes with Genome-Scale Perturb-Seq. Cell, 185(14), 2559-2575.e28.

Rhoads, A., & Au, K. F. (2015). PacBio Sequencing and Its Applications. Genomics, Proteomics & Bioinformatics, SI: Metagenomics of Marine Environments, 13(5), 278–289.

Ribatti, D. *et al*. (2003). Tumor Vascularity and Tryptase-Positive Mast Cells Correlate with a Poor Prognosis in Melanoma. European Journal of Clinical Investigation, 33(5), 420-425.

Robison, Ll *et al*. (1987). Hodgkin's Disease in Pediatric Patients with Naturally Occurring Immunodeficiency. The American Journal of Pediatric Hematology/oncology, 9(2), 189-192.

Rodriques, S. G. *et al*. (2019). Slide-Seq: A Scalable Technology for Measuring Genome-Wide Expression at High Spatial Resolution. Science, 363(6434), 1463–1467.

Ronaghi, M. *et al*. (1996). Real-Time DNA Sequencing Using Detection of Pyrophosphate Release. Analytical Biochemistry, 242(1), 84–89.

Rotte, A. (2019). Combination of CTLA-4 and PD-1 Blockers for Treatment of Cancer. Journal of Experimental & Clinical Cancer Research, 38(1), 255.

Sanger, F. *et al*. (1977). DNA Sequencing with Chain-Terminating Inhibitors. Proceedings of the National Academy of Sciences, 74(12), 5463–5467.

Schaub, M. A. *et al*. (2012). Linking Disease Associations with Regulatory Information in the Human Genome. Genome Research, 22(9), 1748-1759.

Schaum, N. *et al*. (2018). Single-Cell Transcriptomics of 20 Mouse Organs Creates a Tabula Muris. Nature, 562(7727), 367-372

Segert, J. A., Gisselbrecht, S. S., and Bulyk, M. L. (2021). Transcriptional Silencers: Driving Gene Expression with the Brakes On. Trends in Genetics, 0(0), 1-13.

Shendure, J. *et al*. (2005). Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome. Science, 309(5741), 1728–1732.

Shendure, J. *et al*. (2017). DNA Sequencing at 40: Past, Present and Future. Nature, 550(7676), 345–353.

Sherman, M. H. *et al*. (2014). Vitamin D Receptor-Mediated Stromal Reprogramming Suppresses Pancreatitis and Enhances Pancreatic Cancer Therapy. Cell, 159(1), 80-93.

Shiels, M. S. *et al*. (2011). Cancer Burden in the HIV-Infected Population in the United States. JNCI: Journal of the National Cancer Institute, 103(9), 753-762.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional Enhancers: From Properties to Genome-Wide Predictions. Nature Reviews Genetics, 15(4), 272-286.

Simmons, S. K. *et al*. (2023). Mostly Natural Sequencing-by-Synthesis for ScRNA-Seq Using Ultima Sequencing. Nature Biotechnology, 41(2), 204–211.

Smale, S. T. and Kadonaga, J. T. (2003). The RNA Polymerase II Core Promoter. Annual Review of Biochemistry, 72(1), 449-479.

Smith, K. S. *et al*. (2015). Signatures of Accelerated Somatic Evolution in Gene Promoters in Multiple Cancer Types. Nucleic Acids Research, 43(11).

Spitz, F. and Furlong, E. E. M. (2012). Transcription Factors: From Enhancer Binding to Developmental Control. Nature Reviews Genetics, 13(9), 613-626.

Staden, R. (1979). A Strategy of DNA Sequencing Employing Computer Programs. Nucleic Acids Research, 6(7), 2601–2610.

Ståhl, P. L. *et al*. (2016). Visualization and Analysis of Gene Expression in Tissue Sections by Spatial Transcriptomics. Science, 353(6294), 78–82.

Sud, A. *et al*. (2017). Genome-Wide Association Studies of Cancer: Current Insights and Future Perspectives. Nature Reviews Cancer, 17(11), 692–704.

Sur, I. and Taipale, J. (2016). The Role of Enhancers in Cancer. Nature Reviews Cancer, 16(8), 483-493.

Szabo, Q. *et al.* (2019). Principles of Genome Folding into Topologically Associating Domains. Science Advances, 5(4), eaaw1668.

Tabor, S., & Richardson, C. C. (1987). DNA Sequence Analysis with a Modified Bacteriophage T7 DNA Polymerase. Proceedings of the National Academy of Sciences, 84(14), 4767–4771.

The Tabula Sapiens Consortium. (2022). The Tabula Sapiens: A Multiple-Organ, Single-Cell Transcriptomic Atlas of Humans. Science 376 (6594), eabl4896.

Tang, F. *et al*. (2009). mRNA-Seq Whole-Transcriptome Analysis of a Single Cell. Nature Methods, 6(5), 377–382.

Telenius, H. *et al*. (1992). Degenerate Oligonucleotide-Primed PCR: General Amplification of Target DNA by a Single Degenerate Primer. Genomics, 13(3), 718–725.

Tirosh, I. *et al*. (2016). Dissecting the Multicellular Ecosystem of Metastatic Melanoma by Single-Cell RNA-Seq. Science, 352(6282), 189–196.

Travaglini, K. J. *et al*. (2020). A Molecular Cell Atlas of the Human Lung from Single-Cell RNA Sequencing. Nature, 587(7835), 619-625.

Vaquerizas, J. M. *et al*. (2009). A Census of Human Transcription Factors: Function, Expression and Evolution. Nature Reviews. Genetics, 10(4), 252-263.

Venkitaraman, A. R. (2002). Cancer Susceptibility and the Functions of BRCA1 and BRCA2. Cell, 108(2), 171–182.

Waddell, N. *et al*. (2015). Whole Genomes Redefine the Mutational Landscape of Pancreatic Cancer. Nature, 518(7540), 495–501.

Wang, K. *et al*. (2014). Whole-Genome Sequencing and Comprehensive Molecular Profiling Identify New Driver Mutations in Gastric Cancer. Nature Genetics, 46(6), 573–582.

Wang, Y. *et al*. (2021). Nanopore Sequencing Technology, Bioinformatics and Applications. Nature Biotechnology, 39(11), 1348–1365.

Weinstein, I. B. and Joe, A. K. (2006). Mechanisms of Disease: Oncogene Addiction—a Rationale for Molecular Targeting in Cancer Therapy. Nature Clinical Practice Oncology, 3(8), 448-457.

Wooster, R. *et al*. (1995). Identification of the Breast Cancer Susceptibility Gene BRCA2. Nature, 378(6559), 789–792.

Wu, A. R. *et al*. (2014). Quantitative Assessment of Single-Cell RNA-Sequencing Methods. Nature Methods, 11(1), 41–46.

Ximerakis, M. *et al*. (2019). Single-Cell Transcriptomic Profiling of the Aging Mouse Brain. Nature Neuroscience, 22(10), 1696–1708.

Zachariadis, V. *et al*. (2020). A Highly Scalable Method for Joint Whole-Genome Sequencing and Gene-Expression Profiling of Single Cells. Molecular Cell, 80(3), 541-553.

Zahn, H. *et al*. (2017). Scalable Whole-Genome Single-Cell Library Preparation without Preamplification. Nature Methods, 14(2), 167–173.

Zhang, D. *et al*. (2023). Spatial Epigenome–Transcriptome Co-Profiling of Mammalian Tissues. Nature, 616(7955), 113–122.

Zhang, L. *et al*. (2003). Intratumoral T Cells, Recurrence, and Survival in Epithelial Ovarian Cancer. New England Journal of Medicine, 348(3), 203–213.

Zhao, T. *et al*. (2022). Spatial Genomics Enables Multi-Modal Study of Clonal Heterogeneity in Tissues. Nature, 601(7891), 85–91.

Zheng, G. X. Y. *et al*. (2017). Massively Parallel Digital Transcriptional Profiling of Single Cells. Nature Communications, 8(1), 14049.

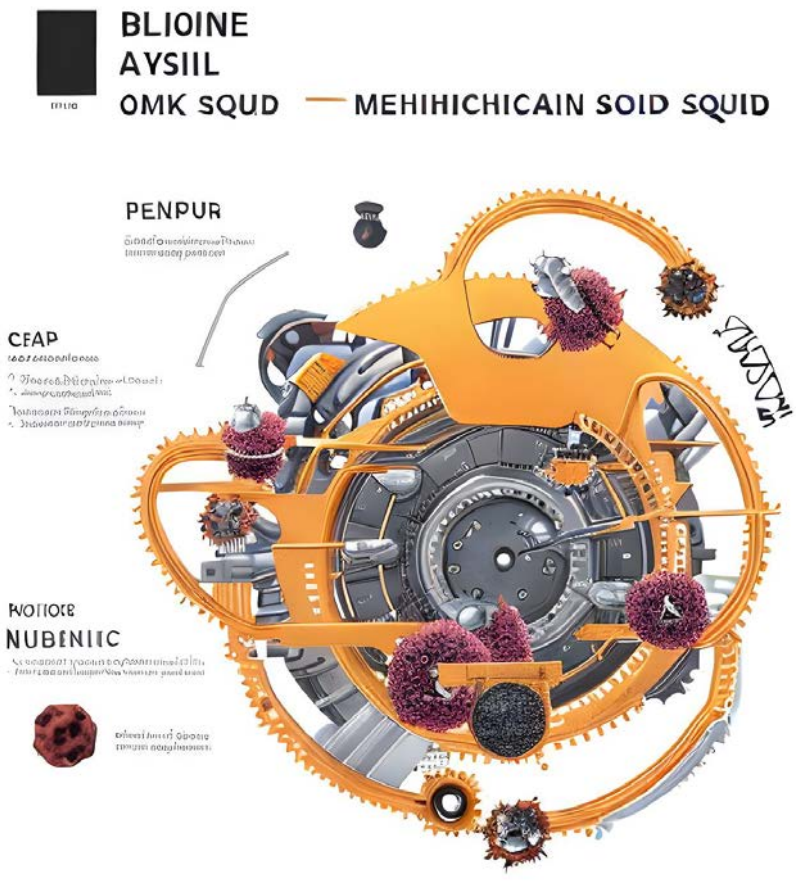
Zhou, Y. *et al*. (2020). Single-Cell Multiomics Sequencing Reveals Prevalent Genomic Alterations in Tumor Stromal Cells of Human Colorectal Cancer. Cancer Cell, 38(6), 818-828.

Zhu, C. *et al*. (2019). An Ultra High-Throughput Method for Single-Cell Joint Analysis of Open Chromatin and Transcriptome. Nature Structural & Molecular Biology, 26(11), 1063–1070.

Ziegenhain, C. *et al*. (2017). Comparative Analysis of Single-Cell RNA Sequencing Methods. Molecular Cell, 65(4), 631-643.e4.

Zong, C. *et al*. (2012). Genome-Wide Detection of Single-Nucleotide and Copy-Number Variations of a Single Human Cell. Science, 338(6114), 1622–1626.

Öhlund, D. *et al*. (2014). Fibroblast Heterogeneity in the Cancer Wound. Journal of Experimental Medicine, 211(8), 1503-23.


BLIOINE
AYSIIL
OMK SQUD
MEHIHICHICAIN SOID SQUID
PENPUR
CFAP
NUBENIIC