

Single Cell Technology

An overview

Curated by Aertslab

Latest version on github.com/aertslab/scforest

May 5th 2019



Lab of Computational Biology
KU Leuven/VIB

List of Figures

C.1	PDMS soft lithography	1
C.2	Single-cell atlas of Drosophila brain	2
1.1	Number of publications in single-cell omics	4
1.2	Current single-cell technology landscape	5
1.3	RNA-seq workflow	7
1.4	ATAC-seq concept	9
1.5	Smart-seq library generation	11
1.6	Smart-seq2 library generation	12
1.7	CEL-seq2 library generation	14
1.8	Teichmann scATAC-seq workflow	15
1.9	Fluidigm C1 IFC	18
1.10	Buenrostro scATAC-seq	19
1.11	Seq-Well cell loading and library generation	20
1.12	Drop-seq barcoded bead generation	21
1.13	Drop-seq microfluidic chip	21
1.14	Drop-seq library generation	22
1.15	inDrop barcoded bead generation	24
1.16	inDrop microfluidic chip	24
1.17	inDrop library generation	25
1.18	Single-cell combinatorial indexing	27
1.19	sci-CAR concept	29
1.20	Split-pool ligation-based transcriptome sequencing	30
1.21	Murine cell atlas	32
1.22	Developmental trajectories in Zebrafish embryogenesis	33
1.23	RNA velocity in human neurogenesis	33
1.24	Impact areas of the Human Cell Atlas	34

List of Acronyms

5'UTR	5' untranslated region
5mC	5'-methylcytosine
AA	acrylamide
AA/BIS	acrylamide/bis-acrylamide
AB	acrylamide/bis-acrylamide
APS	ammonium persulfate
ATAC	assay for transposase-accessible chromatin
ATAC-seq	assay for transposase-accessible chromatin using sequencing
BHB	barcoded hydrogel bead
BIS	bis-acrylamide
BLAST	basic local alignment search tool
blat	BLAST-like alignment tool
bp	base-pair(s)
BSA	bovine serum albumin
CCA	canonical correlation analysis
cDNA	complementary DNA
CEL-seq	cell expression by linear amplification and sequencing
CHEQ-seq	captured high-throughput enhancer testing by quantitative sequencing
ChIP-seq	chromatin immunoprecipitation sequencing
CpG	5'-cytosine-phosphate-guanine-3'
CRISPR	clustered regularly interspaced short palindromic repeats

Ct	cycle threshold
DNA	deoxyribonucleic acid
DNase-seq	DNase I hypersensitive sites sequencing
dNTP	deoxynucleosidetriphosphate
Drop-ATAC	droplet assay for transposase accessible chromatin
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
ERCC	external RNA controls consortium
ESC	embryonic stem cell
FACS	fluorescence-activated cell sorting
FAIRE-seq	formaldehyde-assisted isolation of regulatory elements by sequencing
FAM	fluorescein amidite
FISH	fluorescence in situ hybridisation
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
HB	hydrogel bead
HBW	hydrogel bead wash buffer
HCA	human cell atlas
HEK	human embryonic kidney
HFE	hydrofluoroether
IFC	integrated fluidic circuit
IPA	isopropyl alcohol
ISPCR	inverse suppressive PCR
IVT	in-vitro transcription

kb	kilobase(s)
LNA	locked nucleic acid
MARS-seq	massively parallel RNA single-cell sequencing
MMLV	Moloney murine leukemia virus
mRNA	messenger RNA
NGS	next-generation sequencing
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PDMS	polydimethylsiloxane
PEG	polyethylene glycol
PFO	1H, 1H, 2H, 2H-perfluoro-1-octanol
PMMA	poly(methyl methacrylate)
qPCR	quantitative polymerase chain reaction
REML	restricted maximum likelihood
RNA	ribonucleic acid
RNA-seq	RNA sequencing
RT	reverse transcription
SBS	sequencing by synthesis
scATAC-seq	single-cell assay for transposase-accessible chromatin
scBS-seq	single-cell bisulfite sequencing
sci	single-cell combinatorial indexing
sci-ATAC-seq	single-cell combinatorial indexing ATAC-seq
sci-CAR	single-cell combinatorial indexing for chromatin accessibility and RNA

sci-RNA-seq	single-cell combinatorial indexing RNA-seq
scRNA-ATAC-seq	single-cell RNA and ATAC-sequencing
scRNA-seq	single-cell RNA sequencing
SDS	sodium-dodecylsulfate
Smart-seq	switching mechanism at the 5' end of RNA template sequencing
SNP	single-nucleotide polymorphism
SPLiT-seq	split-pool ligation-based transcriptome sequencing
STAMP	single-cell transcriptome attached to microparticle
STAR	spliced transcripts alignment to a reference
	*
TAC	transposase accessible chromatin
TBSET	tris-buffered EDTA Triton
TEMED	tetramethylethylenediamine
TET	tris-EDTA-Tween
TF	transcription factor
tSNE	t-distributed stochastic neighbour embedding
TSO	template-switching oligo
TSS	transcription start site
UMAP	uniform manifold approximation and projection
UMI	unique molecular identifier
UV	ultraviolet

Context

As a technology, microfluidics seems almost too good to be true: it offers so many advantages and so few disadvantages. But it has not yet become widely used. Why not? Why is every biochemistry laboratory not littered with 'labs on chips'? Why does every patient not monitor his or her condition using microfluidic home-test systems? The answers are not yet clear.

— George M. Whitesides, Nature 2006

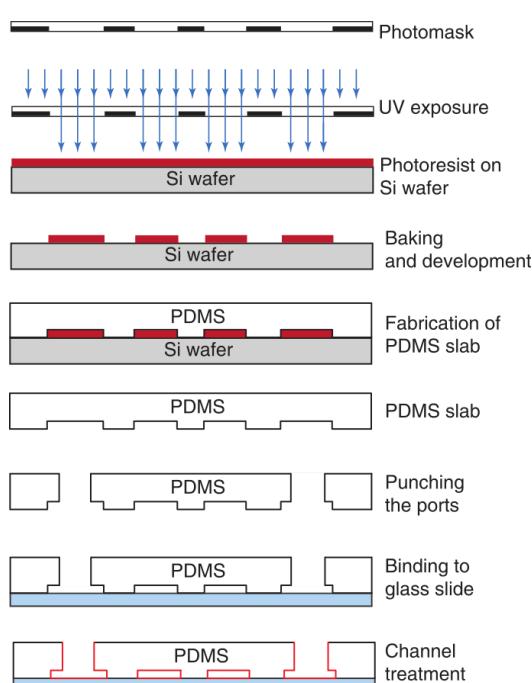


Figure C.1: PDMS soft lithography.

Taken from Mazutis et al. (2013)

produced by focusing two immiscible phases, has been used extensively in bioscience in the past years. The ultra-low volumes associated with droplet microfluidics reduce reagent usage, leading to a reduction in overall cost, and can in some cases increase reaction efficiency (Wu et al., 2014). The technology has been used in a wide array of biological applications: as a platform for high-throughput drug screening assays (Brouzes et al., 2009), in the analysis of protein expression of single cells (Huebner et al., 2007) and even in the engineering of stem cell growth niches (Allazetta and Lutolf, 2015).

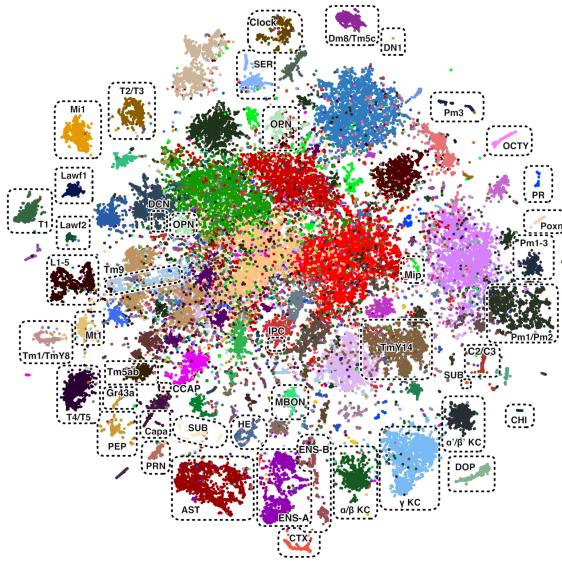


Figure C.2: Single-cell atlas of Drosophila brain. Transcriptomes of 57k single cells are sequenced and computationally separated to discern different cell types. (Davie et al., 2018)

mentation and optimisation of inDrop, a pre-existing droplet-microfluidic single-cell RNA sequencing platform (Klein et al., 2015; Žilionis et al., 2017). In the second line (chapter ??), the droplet microfluidics and molecular engineering expertise acquired in the first part is applied to develop an entirely new protocol. Here, we attempt to prototype a microfluidic approach for ATAC-seq ("Drop-ATAC"), which as of today only exists as a commercial package. As both the optimised inDrop and Drop-ATAC are situated on the far end of the throughput scale, they pose high potential in screening assays where large numbers of cells need to be analysed rapidly.

The end-goal is to explore the new possibilities posed by combining droplet microfluidics and modern molecular cell-analysis. This combination has only recently emerged as a powerful tool in (systems) biology, but is already revolutionising the field. As the droplet-microfluidic single-cell techniques are only in their infancy, there is still a lot of room for improvement and innovation, an opportunity we address in this work.

In the past few years, droplet microfluidics have also been applied in single-cell technology. Single cells are co-encapsulated into nanolitre droplets together with various reagents and barcoded hydrogel beads in order to capture their mRNA content, which is then molecularly barcoded. Due to the barcoding process, mRNA from single cells can be sequenced together and computationally separated. An example of such an application is given in figure C.2. Here, 57 000 fly brain cells were processed using droplet-microfluidic single-cell technology, and computationally analysed to form a comprehensive atlas of the different cell types present in the *Drosophila* brain.

This thesis revolves around optimisation and innovation in the technical/molecular side of the single-cell analysis field. The work can be split in two main lines: the first line (chapter ??) focuses on the imple-

1 | Single-Cell Omics: State of the Art

All life originates from and operates as a massive, interconnected network of molecules. The term 'omics' refers to the systematic characterisation of these elements in order to elucidate their relation to development, function and disease. Since the advent of genomics in the mid 90s shortly after the first genomes of small organisms were sequenced, several new 'omics' fields have emerged. Transcriptomics focuses on the quantification and characterisation of all RNA transcripts. Epigenomics revolves around chromatin changes and their relation to gene expression. Since the focus of omics is a complete overview of an entire class of molecules, the availability of high-throughput analysis methods is critical in these fields (Hood et al., 2004; Patti et al., 2012; Patterson and Aebersold, 2003)

Since the mid 2000s, several bulk omics analysis techniques have allowed researchers to explore the complete transcriptomes and epigenomes of pre-defined cell populations. These techniques often require explicit cell pre-selection, for example using fluorescence-activated cell sorting (FACS) and established marker genes. In the bulk approach, the sampled population is isolated as a group and further treated as a single, homogeneous entity. Any information on heterogeneity within that population is disregarded, masking the presence of previously undiscovered and rare cell types (Bengtsson et al., 2005; Wang and Bodovitz, 2010; Kolodziejczyk et al., 2015; Grün and Van Oudenaarden, 2015).

For decades, however, researchers have acknowledged that morphologically homogeneous cells may exhibit vastly different transcriptional profiles. One of the earliest mentions of this idea occurred in 1992, when Eberwine et al. discovered that in a sample of 15 rat hippocampus pyramidal cells, 2 cells exhibited significantly different messenger RNA (mRNA) profiles than the rest of the population (Eberwine et al., 1992). Realising the power of their discovery, the authors suggested that analysing single cells could help identify new cell types, reveal gene transcript subtypes and even directly determine the influence of external stimuli on a cell's gene expression. However, the single-cell analysis methods employed by Eberwine et al. were expensive and laborious, prohibiting large-scale application of the technique. Indeed, the challenges associated with analysing single cells are numerous and complex:

1. Single cells need to be isolated from complex tissue matrices while preserving their natural state as closely as possible.
2. The minute nucleic acid content of single cells (pg) calls for sensitive library generation protocols before sequencing can occur. This problem is exacerbated by the relative mRNA transcript levels of different genes, which can span orders of

magnitudes between and within single cells from the same tissue (Bengtsson et al., 2005).

3. Due to the large number of cells, the aforementioned challenges need to be tackled in ways that permit high throughput and low reaction volumes to reduce time, cost and labour while maintaining sufficient efficiency and sensitivity.

In the past few years, technological progress in microfluidics and next-generation sequencing have enabled researchers to overcome many of these challenges. Using advanced separation techniques, single cells can now be isolated and analysed individually with increasing accuracy. Ultra-sensitive amplification techniques and transcript barcoding allow sequencing of a single cell's DNA or mRNA content. Since these single-cell sequencing techniques are usually dependent only on basic reagents and machinery, many research institutes have started to adopt them, leading to a rapid increase in the number of publications using them (figure 1.1).

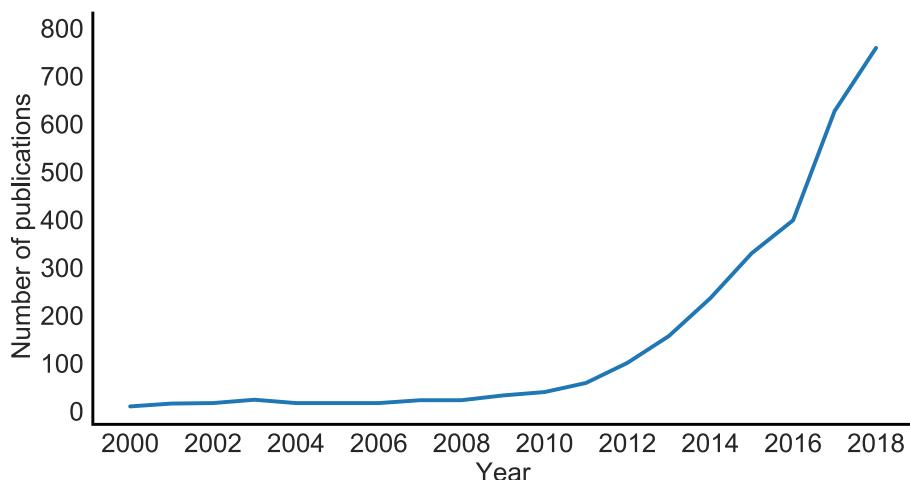


Figure 1.1: Number of abstracts published in scientific journals containing the words "single-cell" and "sequencing". Data gathered from Scopus.

Single-cell analysis techniques have been used to uncover transcriptional heterogeneity in tissues previously deemed homogeneous, to identify new transcripts, map cell state trajectories in (pseudo)time, to study the effect of gene knockdowns and to unravel gene regulatory networks (Tang et al., 2011).

In this literature study, the current state of the art in single-cell research is examined. First, a selection of popular single-cell analysis techniques is explained using their cell separation strategy as a basic classification criterion. Figure 1.2 shows the general subdivision of the reviewed methods. We start with the most straightforward approach: applying established bulk analysis methods on single-cells sorted in microwells. Then, two techniques that automatically load cells into microfluidic arrays in order to decrease labour are

discussed. Third, we look at how droplet microfluidics lead to a 100-fold increase in cell throughput. Fourth, a number of ultra-high throughput well-based protocols is discussed briefly. After the state of the art is established, we go over some of the most important breakthroughs in biology fueled by single-cell research. Finally, a short conclusion is drawn on single-cell research and possible future perspectives of single-cell omics are posed.

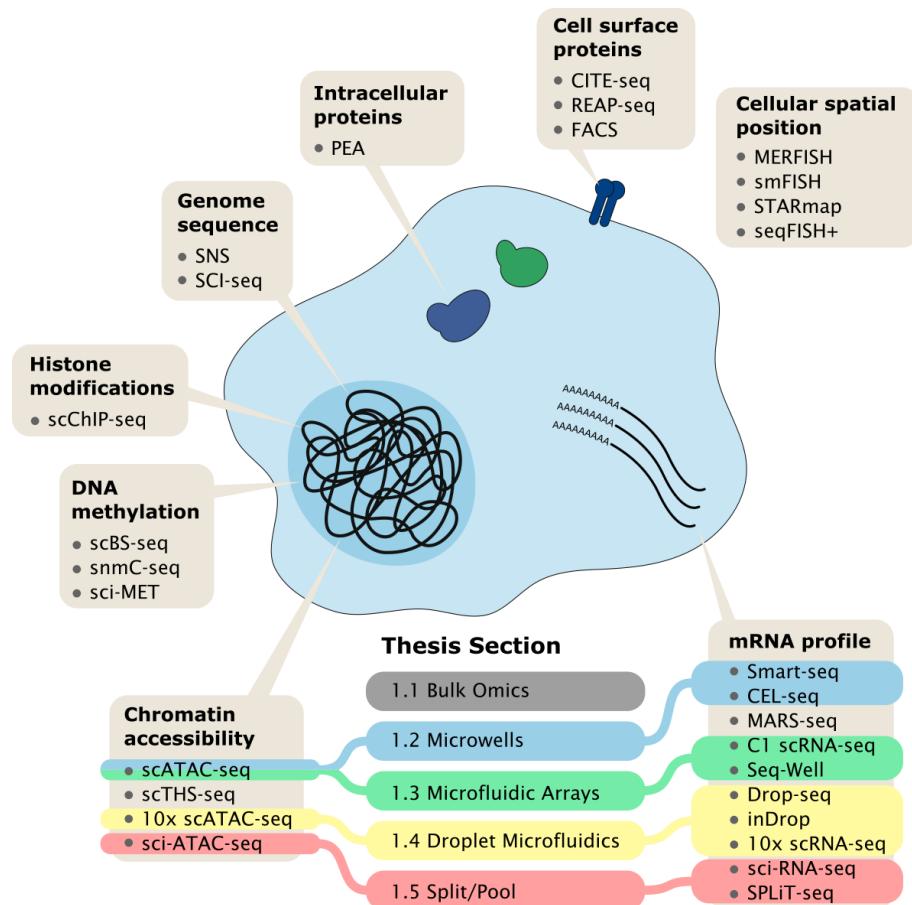


Figure 1.2: Current single-cell technology landscape. An array of analysis techniques can be used to study genomics, transcriptomics, proteomics and epigenomics at single-cell resolution. In this literature study, a number of important single-cell techniques are explained and compared. Modelled after Stuart and Satija (2019).

1.1 Bulk Omics Techniques

Before single-cell analysis techniques can be discussed, the bulk techniques which they are founded on need to be thoroughly understood. The following section will briefly explain RNA-seq and ATAC-seq, which have both become a major point of focus in the single-cell omics landscape and will play a key role in the experimental part of this work.

RNA-seq

The transcriptome comprises all mRNA transcripts, the functional elements of the genome, in a cell population. Transcriptomics focuses on the quantification of these mRNA transcripts and how their levels change during disease and development. Previous generations of transcriptomic techniques, such as expression microarrays, suffer from hybridisation artefacts, cannot detect splice variants or new genes, have a low dynamic range and yield semi-quantitative data due to the limitations of fluorescence (Wang et al., 2009; Tang et al., 2011). RNA-seq, first described in 2008, was the first comprehensive and simple protocol to offer accurate quantification of a cell population's gene expression without requiring cloning of the sample RNA (Mortazavi et al., 2008).

RNA-seq evades the pitfalls of hybridisation and fluorescence-dependent methods by sequencing the cDNA of captured mRNA (fig. 1.3). By using oligo(dT) magnetic beads to capture poly-A tails (repetitions of adenosine nucleotides that is incorporated in all mature mRNA transcripts), transcripts can be detected without explicit previous knowledge of their sequence. In the next step, fragmentation of the captured mRNA destroys secondary structures and mitigates transcript length variation, improving upon random hexamer reverse transcription priming. Optionally, *Arabidopsis* and phage lambda RNA standards can be co-processed with the sample in order to allow absolute transcript quantification. These known RNA sequences are added at set concentrations, yielding a linear standard curve which can be used to relate read count and transcript concentration. In the 140 million reads generated by RNA-seq, Mortazavi et al. detected alternative splice variants for 3 462 genes and identified 596 new candidate transcripts in mouse brain, liver and muscle.

RNA-seq offers uniform transcript coverage, accurate quantification of transcripts up to 1 transcript/cell, a dynamic quantification range of five orders of magnitude, and is able to detect transcripts outside of the prior reference transcriptome. Generated data is highly replicable and correlates well to RNA-microarray data. RNA-seq allows researchers to detect differences in transcriptional profile in response to certain conditions or along development, to catalogue different species of transcripts and to determine the transcriptional structure of genes. Due to its simplicity and low cost, RNA-seq remains one of the main techniques used in transcriptomics to date. However, the method is still sensitive to strongly related sequences in the exome such as gene duplications and paralogs (Wang et al., 2009).

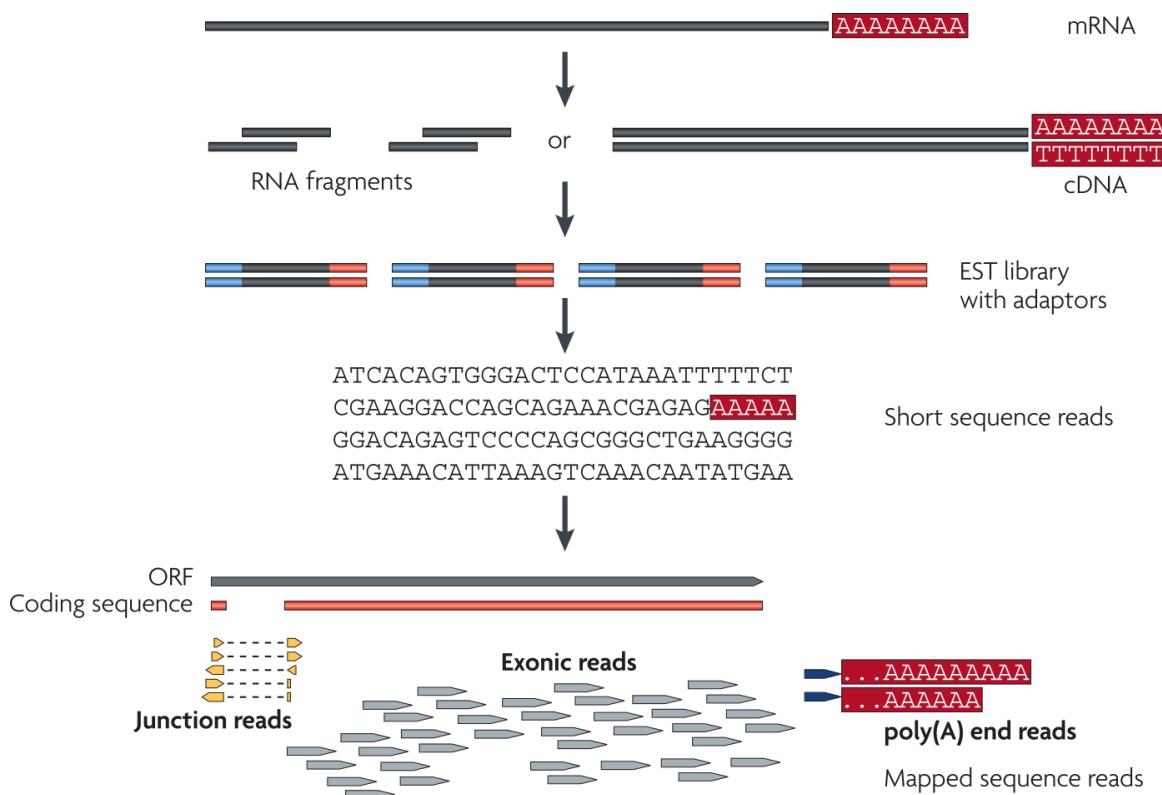


Figure 1.3: RNA-seq workflow. Processed mRNA transcripts are captured using poly-T magnetic beads and hydrolysed to 200-300 bp fragments to remove secondary structures. After reverse transcription using random hexamer primers and second strand synthesis, the fragments are processed into a sequencing library. Sequencing and alignment to a reference transcriptome or exome yields an accurate overview of the sample tissue's expression profile. Additionally, clusters of reads that do not map to previously known exons can be organised into candidate exons in order to identify newly transcribed regions. Adapted from Wang et al. 2009

In 2012, a method for absolute transcript quantification was developed (Kivioja et al., 2012). Here, each reverse transcription primer carries a sequence of 10 additional random nucleotides, called the unique molecular identifier (UMI). This UMI is incorporated into each cDNA transcript. After PCR, each amplification product can now be traced back to its transcript of origin by its UMI, allowing absolute quantification of the original number of transcripts present in the sample.

ATAC-seq

The DNA of eukaryotic cells is organised into chromatin, an organised complex of nucleic acid and histone nucleoproteins. This form of compaction allows the cell's entire genome to fit into the nuclear subspace (Kornberg, 1974). When DNA is packed tightly around these nucleoproteins, the underlying genes sequences are inaccessible to transcription factors, inhibiting transcription of the restricted DNA. However, using an array of mechanisms such as DNA methylation, nucleosome positioning and histone modification, the cell can locally remodel the steric accessibility of chromatin to allow gene transcription. Chromatin accessibility state is therefore a leading indicator of which genes are actively expressed in a cell and, together with DNA methylation and histone modification, forms the physical basis of epigenomics (Jaenisch and Bird, 2003; Kouzarides, 2007; Schones and Zhao, 2008; Bannister and Kouzarides, 2011). Epigenomic variation has been shown to be highly variable in time, between cell populations and across cell generations and thus effectively provides a layer of information "on top of" the cell's genome. Studying epigenomic phenomena can thus help construct models of gene regulatory pathways. A thorough understanding of these pathways may ultimately help answer the question of how different cell phenotypes can arise from genetically identical precursor cells (Johannes et al., 2008).

The chromatin state of cell populations has previously been studied using techniques such as FAIRE-seq and DNase-seq (Giresi et al., 2007; Song and Crawford, 2010; Gaulton et al., 2010; Song et al., 2011). However, these methods involve steps such as chloroform extraction and gel purification which may lead to loss of sensitivity. The assay for transposase-accessible chromatin using sequencing (ATAC-seq), a technique published by Buenrostro et al. in 2013, allows researchers to investigate the chromatin condensation state of a cell population without such potentially loss-prone steps. ATAC-seq relies on the activity of a hyperactive, mutated Tn5 transposase which (near)-randomly fragments only sterically accessible chromatin regions. In contrast to DNase, another enzyme that randomly fragments accessible chromatin, Tn5 ligates oligonucleotide adapters where it cleaves DNA. Strands fragmented by Tn5 are thus flanked by two adapter sequences. These adapters can then be used in PCR as hybridisation sites for specialised sequencing primers, simplifying post-fragmentation processing (fig. 1.4).

Random fragmentation of accessible DNA results in many short fragments of the accessible chromatin and fewer but longer fragments of the inaccessible regions. The fragment length and count can therefore be used to determine the "accessibility" of a given chromatin region. Areas of significant enrichment in the read distribution are detected computationally in a process called peak calling, yielding information on which regions of the genome are accessible in the cell population.

ATAC-seq requires a starting number of cells 2 - 3 orders of magnitude lower than DNase-seq and FAIRE-seq (~50 000 versus ~50 000 000). The protocol takes only ~5 hours from sample collection to sequencing compared to 3.5 days for DNase-seq, but the resulting data is of similar quality (Buenrostro et al., 2013). In 2017, a more widely-

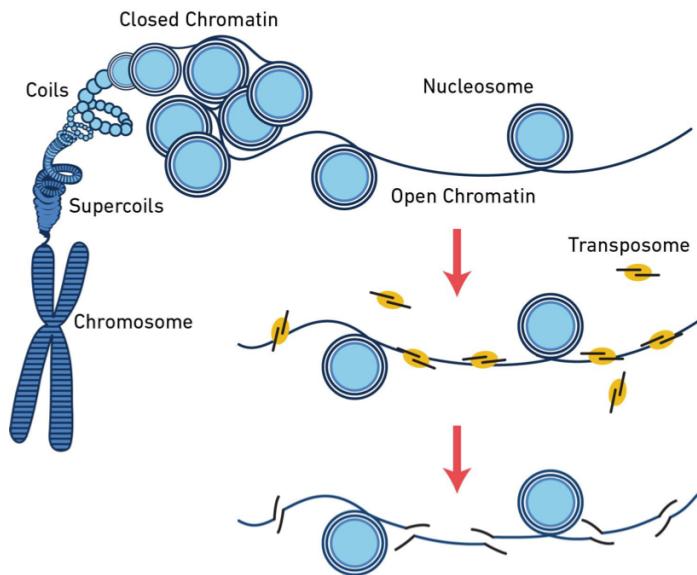


Figure 1.4: ATAC-seq concept. Tn5 transposase fragments DNA only in sterically accessible chromatin and ligates adapters. A subsequent PCR uses these adapters to incorporate sequencing adapters and sample indexes. After sequencing, reads are mapped to a reference. Source: 10x Cell Ranger ATAC Algorithm brochure

applicable bulk ATAC-seq protocol, dubbed "Omni-ATAC-seq", was released. This updated version is applicable to a broader range of cell types and yields a higher fraction of reads in peaks than the standard ATAC-seq. Importantly, the improved protocol also efficiently removes mitochondrial DNA from the transposition reaction, hereby reducing sequencing costs (Corces et al., 2017). As will be shown later, recent efforts have scaled ATAC-seq up to single-cell resolution (Buenrostro et al., 2015; Chen et al., 2018).

1.2 Microwell-based Single-Cell Omics Techniques

Some of the earliest efforts to analyse the nucleic acid content of single cells simply applied bulk analysis techniques to single cell lysate suspended in small compartments. For example, Eberwine et al. carried out reverse transcription of mRNA by injecting a live cell with viral reverse transcriptase and oligo-dT primers and aspirating the cell contents into the glass electrode. Due to the high fixed labour cost and limits of technology at the time, their approach led to extremely low throughput compared to today's standard: 5 days of work for 1 cell then, compared to 5 days of work for 10 000 cells today (Hashimshony et al., 2012). The following section shows recent continuations on the microcompartment approach, which have become widespread due to their ease of use.

Smart-seq

Switching mechanism at the 5' end of RNA template sequencing (Smart-seq), originally formulated by Ransköld et al. in 2012, was one of the first single-cell mRNA-seq protocols and the first to provide full transcript coverage. The technique is modelled after an earlier mRNA-seq protocol by Tang et al. (2009), who sequenced the transcriptome of a single mouse blastomere.

Ransköld et al. first applied Smart-seq to 42 human cells manually picked from dissociated tissues using microscope-assisted micromanipulation. Single cells were further treated separately in microwells. Figure 1.5 shows the steps involved in Smart-seq library generation. An important addition of Smart-seq compared to the Tang et al. protocol is the use of template switching. During reverse transcription, the Moloney murine leukemia virus (MMLV) reverse transcriptase adds extra cytosines to the 3' end of the cDNA strand, allowing a so-called template-switching oligo (TSO) to hybridise. MMLV reverse transcriptase then continues cDNA synthesis using the TSO as a template in a process called template switching. The sequence complementary to the TSO is thus incorporated in the cDNA library and is used as a PCR priming site in further amplification steps. Importantly, the SMART primer can hybridise to both ends of the cDNA transcript. The PCR product of short templates can thus form loops which impede further amplification. This mechanism corrects the natural short-fragment bias associated with PCR amplification. The template-switching reverse transcription approach is more user-friendly and time-efficient than linear amplification by in-vitro transcription. Compared to a regular reverse transcription, template-switching also produces cDNA transcripts of the complete RNA template. This means that any splicing information carried by the very distal ends of the transcript is retained after sequencing.

An improvement on Smart-seq, Smart-seq2, was published by Picelli et al. in 2013. Here, single cell isolation was performed on 262 human and mouse cells. In the improved Smart-seq2 protocol, cDNA-yield was increased twofold by incorporating a locked nucleic

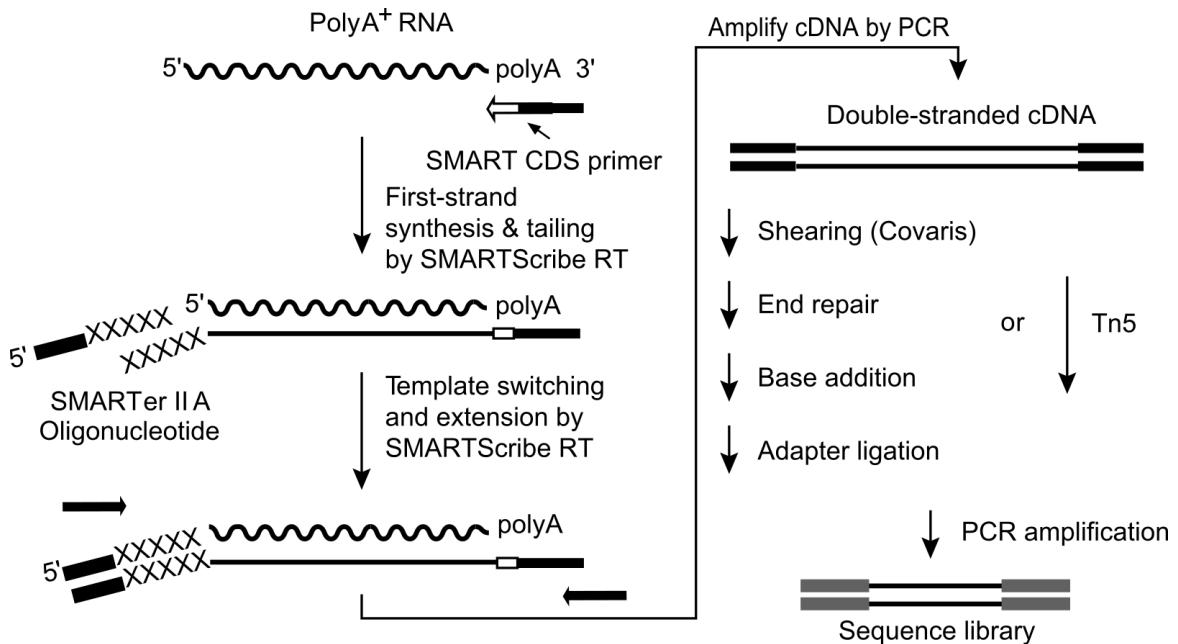


Figure 1.5: Smart-seq library generation. Single cells are manually isolated from dissociated sample tissue and placed in separate microwells. The cells are then lysed and their polyadenylated mRNA content is captured by oligo(dT) reverse transcriptase primers. Full transcript cDNA is synthesised using MMLV reverse transcriptase and pre-amplified by ISPCR. Double-stranded cDNA is then fragmented and tagged with sequencing primers and adapters by Tn5. Alternatively, shearing and adapter ligation can be used. After further PCR amplification, the library is ready for next-generation sequencing. Adapted from Ramsköld et al., 2012.

acid (LNA) in the TSO. This small change led to better thermal stability of the TSO-cDNA duplex. Further reagent concentration optimisations resulted in an overall increase in sensitivity and accuracy relative to Smart-seq. Picelli et al. report that Smart-seq2 detects ~12k genes from HEK cells compared to ~10k genes detected by first generation Smart-seq. Additionally, cells were isolated by distributing μ l volumes of strongly diluted cell suspensions into microwells instead of relying on manual cell picking. This allows for the use of automated liquid handling, but incurs an additional increase in reagent cost due to empty wells. An overview of the Smart-seq2 library generation workflow is given in figure 1.6.

Smart-seq and Smart-seq2 yield full transcript information and thus allow researchers to study both distal and proximal splicing events, novel exon detection, single-cell SNP detection, and allele-specific gene expression (Kolodziejczyk et al., 2015). However, using microwells reduces throughput and increases reaction volumes, leading to an overall increased cost per cell. Cell selection based on dilution also inhibits the detection of rare

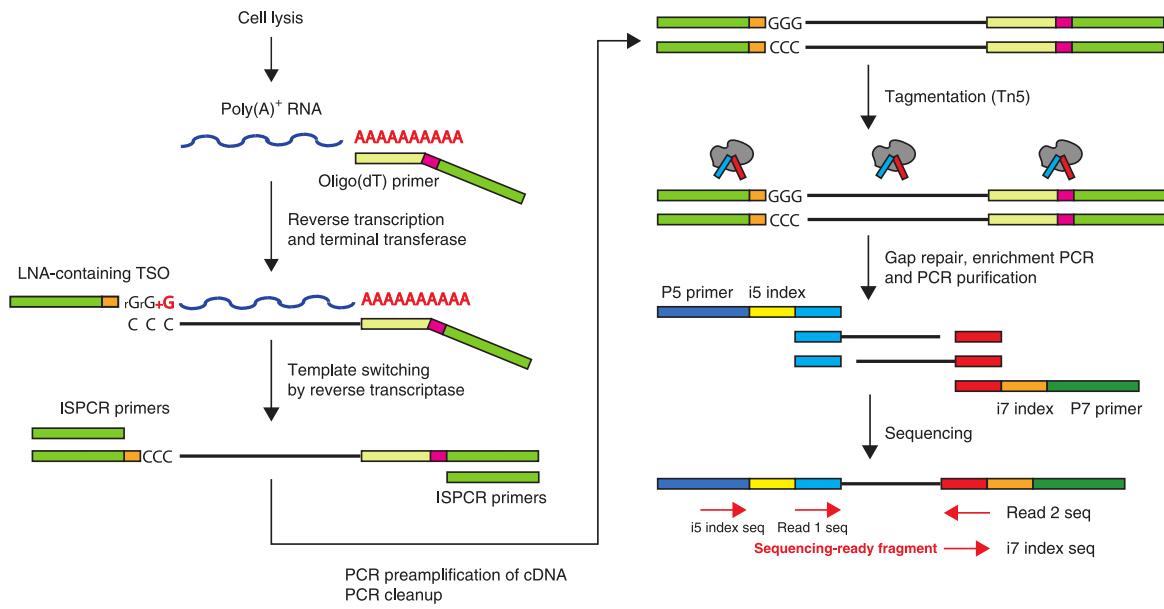


Figure 1.6: Smart-seq2 library generation. Single cells from a dilute cell suspension are distributed into separate microwells. The cells are then lysed and their polyadenylated mRNA content is captured by oligo(dT) reverse transcription primers. After full-transcript reverse transcription by template switching using an LNA-TSO, the resulting cDNA is pre-amplified and subsequently tagmented with sequencing adapters using Tn5 transposase. The library is then further amplified and sequenced. Adapted from Picelli et al., 2014b.

cell types in large samples. The Smart-seq and Smart-seq2 protocols do not incorporate UMIs into the cDNA, making absolute quantification of transcripts impossible. The use of double-stranded cDNA also discards information about transcript strand specificity. Moreover, the random fragmentation nature of the Tn5 tagmentation reaction leads to a reduced sequencing coverage of the very 5' ends of transcripts, which carry the transcription start site (TSS) and 5' untranslated region (5'UTR). Studying these important regions is therefore difficult using Smart-seq. In conclusion, Smart-seq2 is a highly sensitive, but low-throughput method best suited for small cell populations where the complete mRNA transcript is needed to investigate distal splicing or presence of SNPs. Since Smart-seq2 requires only off-the shelf reagents and equipment, it has become a widespread single-cell RNA-seq protocol (Picelli, 2017).

CEL-seq

Cell expression by linear amplification and sequencing (CEL-seq), published in 2012 by Hashimshony et al., is a single-cell transcriptomics technique revolving around mRNA barcoding followed by in-vitro transcription (IVT) for linear amplification. IVT leads to more reproducible and sensitive results compared to exponential PCR amplification, but requires ~400 pg of cDNA compared to the average eukaryotic cell's mRNA content of ~1 pg (Tang et al., 2011; Hashimshony et al., 2012). CEL-seq satisfies this requirement by pooling barcoded cDNA from different cells of origin before applying IVT. Due to the in vitro transcription step which generates only sense RNA from antisense cDNA, CEL-seq generates strand-specific sequencing libraries. For a detailed overview of the CEL-seq protocol, see (Hashimshony et al., 2012). A second generation protocol, CEL-seq2, was published in 2016. Figure 1.7 shows the generalised CEL-seq2 workflow.

CEL-seq2 increases reverse transcription efficiency by shortening the reverse transcription primer. This leads to a higher fraction of mRNA transcripts being detected. Incorporating Illumina sequencing adapters during reverse transcription eliminates a PCR-ligation step, increasing read mappability from 60.9% to 93.8%. CEL-seq2 also incorporates a UMI into every cDNA strand, enabling absolute transcript quantification. These incremental improvements together lead to a 30% increase in number of genes detected compared to the first generation CEL-seq. It is also shown that CEL-seq2 can be readily performed on the Fluidigm C1 microfluidic platform for increased gene detection and reduced labour cost (Hashimshony et al., 2016).

Compared to Smart-seq, CEL-seq2 yields a near 100% increase in genes detected on the same cell sample (Hashimshony et al., 2016). CEL-seq2, however, is strongly 3'-biased and can therefore not provide information on distal splicing as opposed to Smart-seq. Importantly, CEL-seq's in-vitro transcription amplifies at a lower rate and is more contrived than Smart-seq's PCR, taking 13 hours per sample in the original CEL-seq protocol. Regardless, CEL-seq's sensitivity and absolute quantification possibilities often outweigh its disadvantages and make it suitable technique for many transcriptomics applications.

Microwell scATAC-seq

A scATAC-seq approach which does not explicitly require the use of microfluidics was published in 2018 by the Teichmann Group (Chen et al., 2018). Their protocol relies on an interesting peculiarity about Tn5: when Tn5 binds to accessible chromatin, it cleaves chromosomal DNA and inserts its adapters at the 5' end of each opposite strand. However, during this process, Tn5 remains clamped on the DNA strand until it is denatured by either SDS or a heat shock (Picelli et al., 2014a). During this process, the cell is lysed, but its DNA content stays confined to the nucleus. Only upon denaturation of Tn5, the fragmented DNA is released from the enzyme. In the Teichmann protocol, bulk fragmented

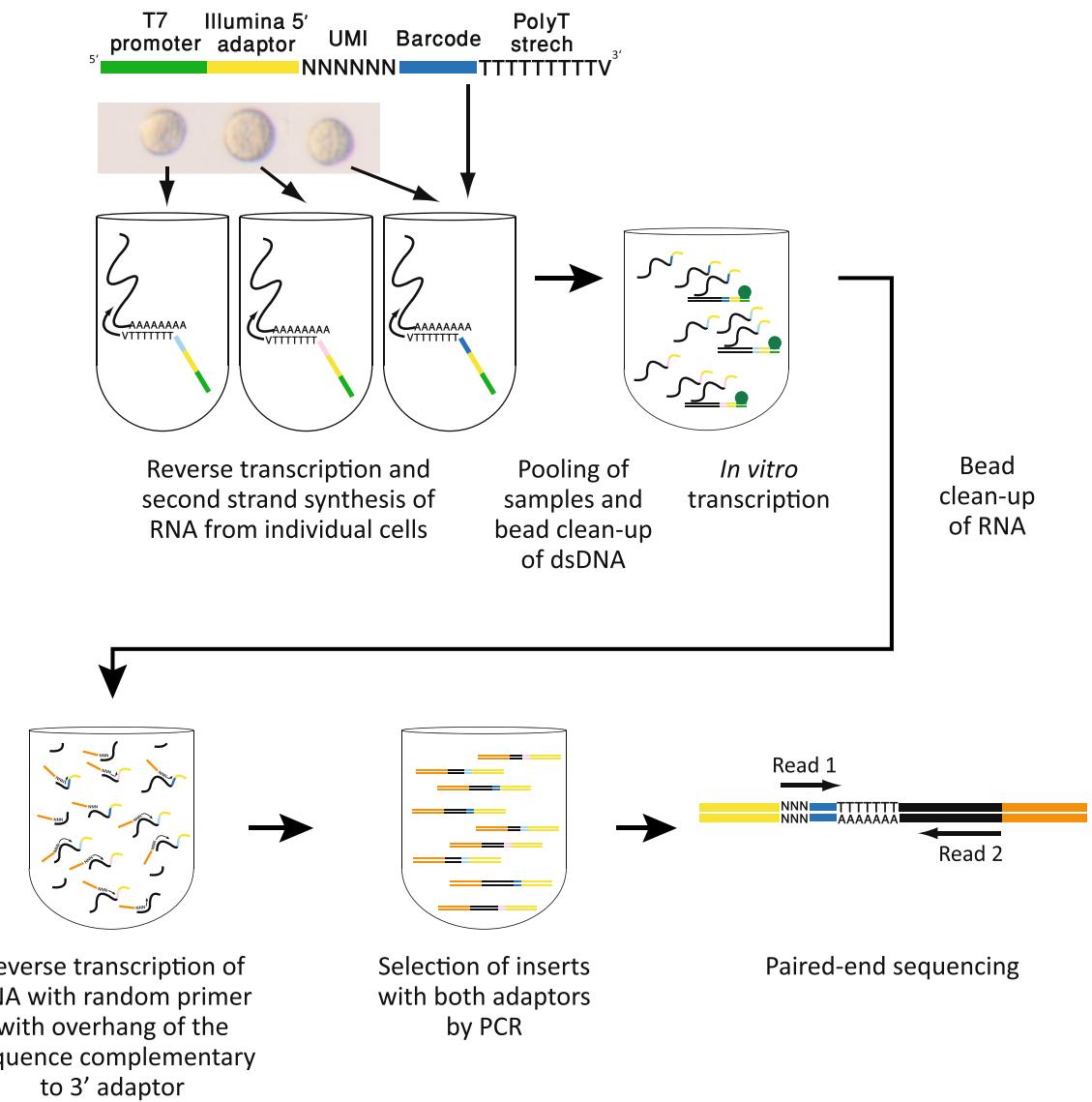


Figure 1.7: CEL-seq2 library generation. Single cells are lysed and poly-A⁺ mRNA is reverse transcribed using a poly-T primer that includes a cell-specific barcode, a transcript-specific UMI, an Illumina 5' adapter and a T7 promoter. Pooled cDNA is transcribed in vitro and the amplified RNA is purified. Then, using a random primer with overhang complementary to the 3'-adapter, inserts flanked with a 3' Illumina sequencing adapter are generated which are then purified and processed for paired-end sequencing. Adapted from Hashimshony et al. (2016)

nuclei are sorted into a microtiter plate containing lysis buffer. The fragmentation reaction is then stopped using SDS and further reactions are performed separately in each microwell. An overview of the technique is shown in figure 1.8.

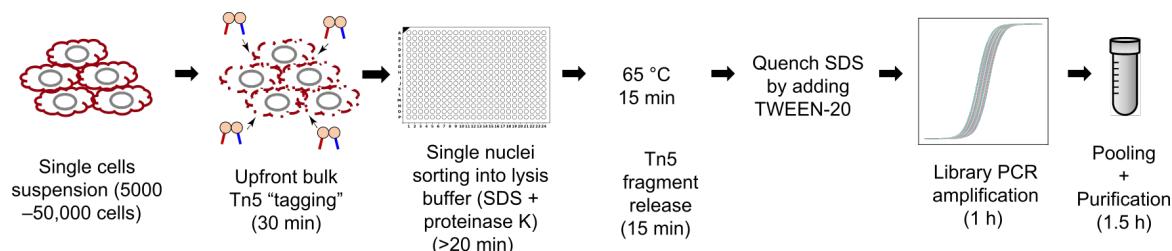


Figure 1.8: Teichmann scATAC-seq workflow. Fragmentation is performed in bulk by incubating 50 000 cells with Tn5 transposase. "Tagmented" nuclei are sorted into 384-well plates containing lysis buffer with SDS and proteinase K, which inactivate Tn5. After Tn5 is inactivated and DNA fragments are released, SDS is quenched by addition of Tween-20, preventing non-specific inactivation of downstream enzymes. Fragments generated by Tn5 are then indexed using indexing PCR, amplified, pooled and sequenced. Taken from Chen et al. (2018).

In contrast to earlier scATAC-seq protocols, the Teichmann protocol does not require intermediate purification of the DNA, simplifying the fragmentation workflow. Chen et al. benchmarked their scATAC-seq method against the Buenrostro et al. C1 scATAC-seq protocol (which will be shown in section 1.3) and found that their own protocol generated higher quality data in a shorter time frame. The whole procedure takes place in the same plate and does not require intermediate purification steps, as opposed to previous scATAC-seq protocols which employ bead purification of DNA after fragmentation. Chen et al. also show that immunostained cells retain their staining after bulk fragmentation, meaning FACS can be used to filter rare cell types from the sample post-fragmentation. Additionally, the bulk fragmentation approach cuts cost of Tn5 per cell considerably but the subsequent microcompartment-based approach still incurs a substantial cost due to the 20 μ l PCR reaction volume.

Microwell Approaches: Key Takeaway

Due to their user-friendliness and modest requirements in terms of equipment, the use of plate-based techniques has become widespread. However, microcompartment-based single-cell analysis techniques suffer from a number of inherent disadvantages:

1. Isolating and/or sorting single cells into microwells and performing reactions on them is tedious, especially when performed manually. FACS and automated liquid handling may be used to partially solve these problems.
2. Performing single-cell reactions in a microwell plate leads to large reagent volumes per cell, resulting in a high cost and, more importantly, decreased reaction efficiency due to dilution effects (Wu et al., 2014).
3. Handling reaction plates can form a throughput bottle neck in parallel processing of a large number of cells, for example during reverse transcription, where one heat block can only process 96 cells.

Despite these drawbacks, microcompartment-based techniques have proven to be viable tools to extract information on single-cell resolution DNA methylation, chromatin condensation, gene expression and genomic profile. Generally speaking, microcompartment techniques such as Smart-seq and CEL-seq yield data of higher quality than the high-throughput methods shown later (Ziegenhain et al., 2017). Together, the protocols explained in this sections have led to a number of groundbreaking results of which a limited selection will be shown in 1.6.

The greatest advantage that microwells offer can be found not in accuracy, cost or throughput, but in flexibility. It is straightforward to manipulate and sample the contents of a microwell plate to optimise the reactions carried out within. Moreover, multi-step methods involving sequential addition and purification can easily be carried out on an accessible well platform. Many high-throughput techniques shown in the following sections were born in a microwell, and most likely some of the more complicated techniques we will see in the future will start in a well as well.

1.3 Microfluidic Arrays

The previous section has shown how single-cell data can be obtained by simply applying bulk analysis methods to a single cell in a microwell. A major bottleneck in the application of these assays is the distribution of single cells into individual microcompartments. Common methods include manual picking, FACS or Poisson-distributed dilution, each with their own set of disadvantages. In the following section, a number of techniques which separate or trap single cells into (sub)nanolitre volume compartments are introduced.

Fluidigm C1

The C1 comprises a benchtop microfluidic controller that contains the pneumatic hardware, and a disposable integrated fluidic circuit (IFC) (figure 1.9), which hosts the microfluidic channels through which the sample is routed. Together, they form a platform that can isolate single cells from dilute cell suspensions into nanolitre capture sites. In section 1.2, we briefly mentioned that the Smart-seq2 protocol can be implemented on the C1. Preceding Smart-seq2, Buenrostro et al. published a C1-based scATAC-seq protocol (Buenrostro et al., 2015). Here, tagmentation and PCR occur on the C1 IFC, after which single-cell libraries are collected and further amplified with cell-identifying barcoded primers. For an overview of the Buenrostro et al. scATAC-seq method, see figure 1.10.

The greatest advantage of the C1 compared to conventional microwells are ease of use and semi-automation at a minimal loss of data quality. During cell trapping, tagmentation and pre-amplification, minimal human interference is necessary. However, barcoding and subsequent amplification of the single-cell libraries still need to happen in a microwell plate. The C1 approach is therefore practically an automated cell-capture and nucleic acid capture method followed by microcompartment-style barcoding. As the IFC can only process 96 cells in parallel, high-throughput application of the C1 is difficult. Moreover, the C1 IFC only accommodates three cell size classes per chip design (5-10, 10-17 and 17-25 μm in diameter), leading to a size-biased selection. The very nature of the C1's cell trapping mechanism also complicates the capture of non-spherical cells. Macosko et al., authors of a competing microfluidic scRNA-seq method covered in section 1.4, report that 30% of C1-generated libraries contain mixed-species contamination, which is unusually high for single-cell methods (Macosko et al., 2015). Shalek et al. report a more forgiving multiplet rate of 11%. Additionally, the 96-cell IFC takes an input sample of at least 1000 cells, imposing an effective capture efficiency of 1-10% (Žilionis et al., 2017). This last drawback prohibits the application of C1-based cell capture on protocols where information on rare cell types is key, such as de novo cell typing. It is due to these limitations and strong competition from droplet microfluidics approaches that the C1 has not been able hold a large market share, appearing in 228 publications over a span of 8 years (Fluidigm Website, 2019), compared to 391 publications for the 10x Chromium, which was released

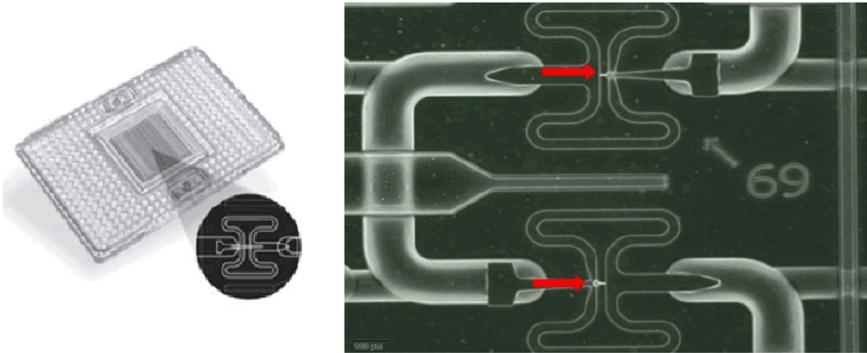


Figure 1.9: Fluidigm C1 IFC. A suspension of single cells is added to the IFC's sample inlet and inserted into the C1 control unit. The control unit pneumatically pumps the suspension through a serpentine channel containing a series of capture sites. Once a cell occupies a capture site, access to the site's collection chamber is blocked, routing subsequent cells to the next capture site. This results in a series of capture sites along the path each containing a single cell. After a run time of 1 hour, cell occupancy is checked using (fluorescence) microscopy. Then, lysis reagents are routed through the serpentine channels, lysing cells and clearing the obstructed path to the collection chambers, resulting in a flow of lysate from each capture site to its respective collection chamber. Next, reverse transcription or amplification reagents can be routed to the chamber. The reaction products from individual cells are then collected into a separate microwell. During each microfluidic unit operation, a set of peristaltic valves seal off every capture site in order to minimise contamination. Adapted from Azizi et al. (2014).

in 2016. Still, the C1 has found niche applications in the automation of specialised and complicated protocols such as sci-CAR (see section 1.5).

Sq-Well

A rather straightforward microfluidic cell loading strategy dubbed Seq-Well was published in 2017 (Gierahn et al., 2017). In this method, cells and beads carrying barcoded poly-dT primers are loaded into an array of picolitre wells. Here, the barcoded poly-dT primers capture cellular mRNA. The beads are then pooled for bulk reverse transcription. An overview of the Seq-Well method is given in figure 1.11.

Seq-Well improves on earlier picowell methods by Fan et al. (2015) and Yuan and Sims (2016) by chemically sealing the array of picolitre wells using semi-permeable membrane which allows for buffer exchange, but prevents cross-contamination between wells. The inner surface of the picowells is also treated to prevent non-specific mRNA binding. Similarly

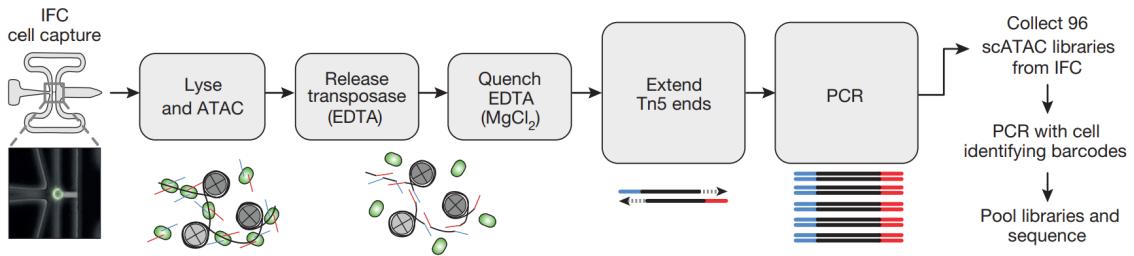


Figure 1.10: Buenrostro scATAC-seq. 5 μ l of a 300 cells/ μ l suspension is placed into the IFC sample inlet. Single cells are trapped into capture sites and lysed. In an individual reaction chamber, each cell is fragmented in 13.5 nl of fragmentation mix. After EDTA-mediated release of Tn5, the tagged fragments are pre-amplified by PCR on-chip. After amplification, individual cell samples are collected and individually barcoded using PCR. Only then are the single-cell libraries collected and pooled for sequencing. Taken from Buenrostro et al. (2015).

to Smart-seq, Seq-Well uses a template-switching library generation strategy to facilitate amplification and to produce full-length transcripts. In short, Seq-Well is a simple and portable protocol for loading a large number of cells into microwells rapidly. However, Žilionis et al. remark that not all mRNA is captured by the barcoded beads, possibly leading to contamination at the pooling step after hybridisation (Žilionis et al., 2017). Gierahn et al. also observed a multiplet rate of 11.4% when 20 000 of the 83 200 available picowells were loaded with cells. 77.5% of the Seq-Well reads could be mapped to a reference exome and ~6 000 human genes were detected. This is higher than droplet-microfluidics techniques such as Drop-seq (~5 000 human genes) and 10x Chromium (~4 600 human genes) but lower than a true microwell protocol such as Smart-seq2 (~12 000 human genes).

Microfluidic Arrays: Key Takeaway

Microfluidic array methods have succeeded in reducing the significant labour costs associated with single-cell experiments. As always, a trade-off is made between data complexity and throughput. Single-cell experiments on the Fluidigm C1 platform offer high quality data consistent with explicit well-based techniques, but at a high cost and low throughput while Seq-Well offers high-throughput, portability and low cost at the expense of sensitivity and specificity. Importantly, both methods still allow customisation and adaptation of the reactions performed on the isolated cells. Microfluidic array techniques therefore form an intermediate between well-based protocols and the droplet-microfluidic techniques shown in section 1.4.

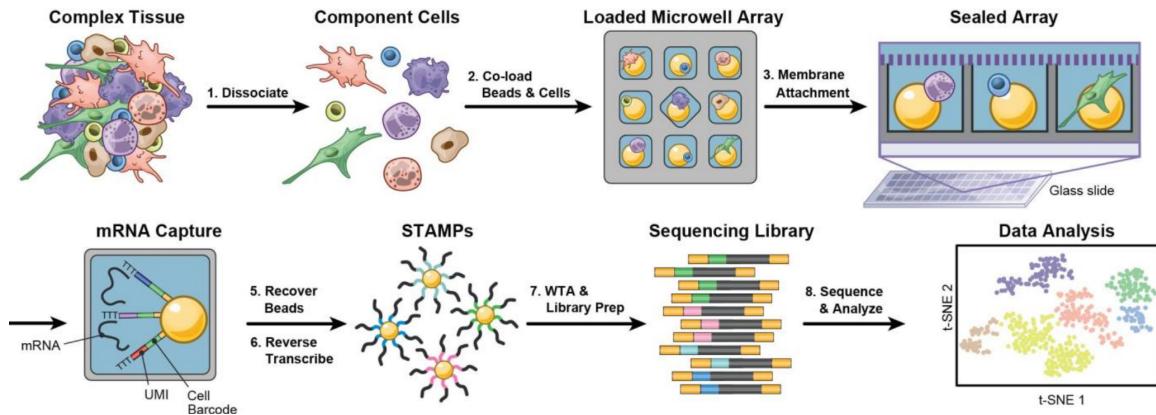


Figure 1.11: Seq-Well cell loading and library generation. Cells are co-loaded together with barcoded microbeads in a PDMS array containing 83 200 picowells. Each bead oligo holds a cellular barcode, a UMI and a poly-dT tail to capture mRNA poly-A tails. The wells are then sealed using a semi-permeable membrane and submerged in lysis buffer. In each well, released cellular mRNA is captured by the poly-dT primers on the barcoded beads. The plates are unsealed, and beads are pooled. A bulk template-switching reverse transcription reaction then generates the single-cell transcriptomes attached to microparticles (STAMPs), which are then amplified and sequenced. Adapted from Gierahn et al. (2017).

1.4 Droplet Microfluidic Single Cell Omics Techniques

In 2015, two remarkable single-cell RNA sequencing methods debuted in the same issue of *Cell* (Klein et al., 2015; Macosko et al., 2015). Two research groups both affiliated with Harvard university had collaboratively come up with the idea of barcoding cells by encapsulating them together with solid primer carriers in microscopic droplets. The methods, aptly named inDrop and Drop-seq, both relied on encapsulating single cells in tiny water-in-oil droplets together with a microbead carrying the barcoded primers used to index every cell's mRNA content. The following section will briefly cover inDrop and Drop-seq and, as the experimental section of this thesis revolves around a hybrid version between the two, their differences will be examined closely.

Drop-seq

Drop-seq, Macosko et al.'s approach to droplet-based single-cell RNA sequencing, is based around co-encapsulating single cells with a barcoded resin bead on a microfluidic chip. Inside every droplet, the cell's mRNA content is captured by the bead's barcoded transcription primers. The ensuing reverse transcription reaction incorporates the primer's barcode and unique molecular identifier (UMI) in the cDNA, which allows researchers to demultiplex the cDNA by cell and mRNA transcript of origin. Drop-seq's bead production process, microfluidic chip and workflow are given in figures 1.12-1.14.

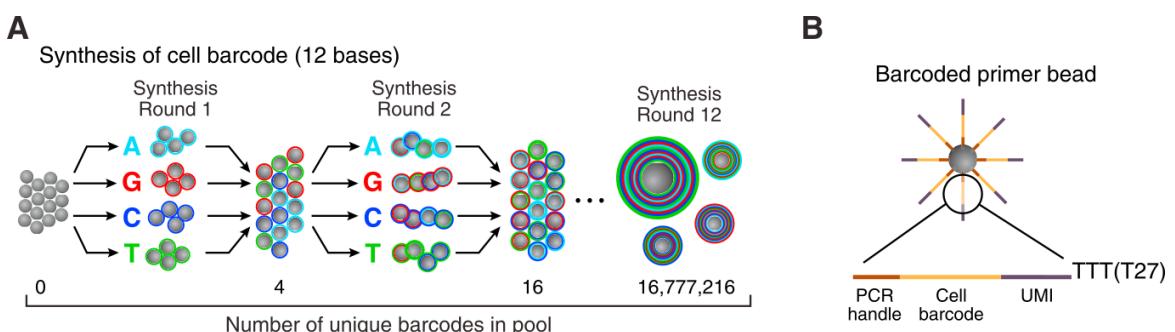


Figure 1.12: Drop-seq barcoded bead generation. (A) Resin beads undergo successive rounds of splitting, oligo synthesis and pooling to generate $2^{12} = 16\ 777\ 216$ possible uniquely barcoded beads. Then, eight steps of degenerate synthesis append an eight nucleotide UMI to the cell barcodes, followed by addition of a poly-T tail which can capture cellular mRNA. (B) Overview of the final bead barcode structure. Taken from Macosko et al. (2015).

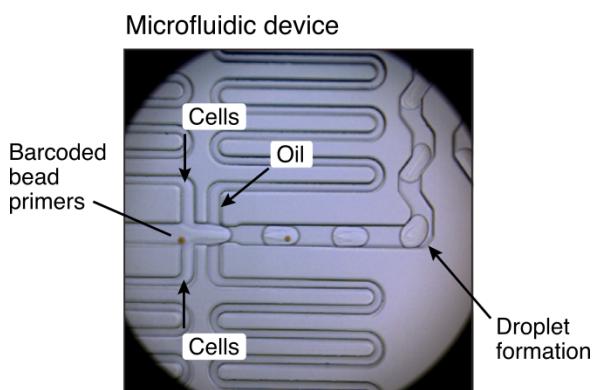


Figure 1.13: Drop-seq microfluidic chip. A flow of barcoded beads suspended in lysis buffer is joined by a single-cell suspension and emulsified into nanolitre droplets by an oil flow. Taken from Macosko et al. (2015).

Using Drop-seq, a single researcher can prepare thousands of single cell transcriptomes for sequencing in a single day, orders of magnitude faster than the 96-well plate assays discussed in section 1.2. In the original Drop-seq paper, Macosko et al. profiled 44808 single-cell mouse retina transcriptomes in just 4 days, at a fraction of the cost of

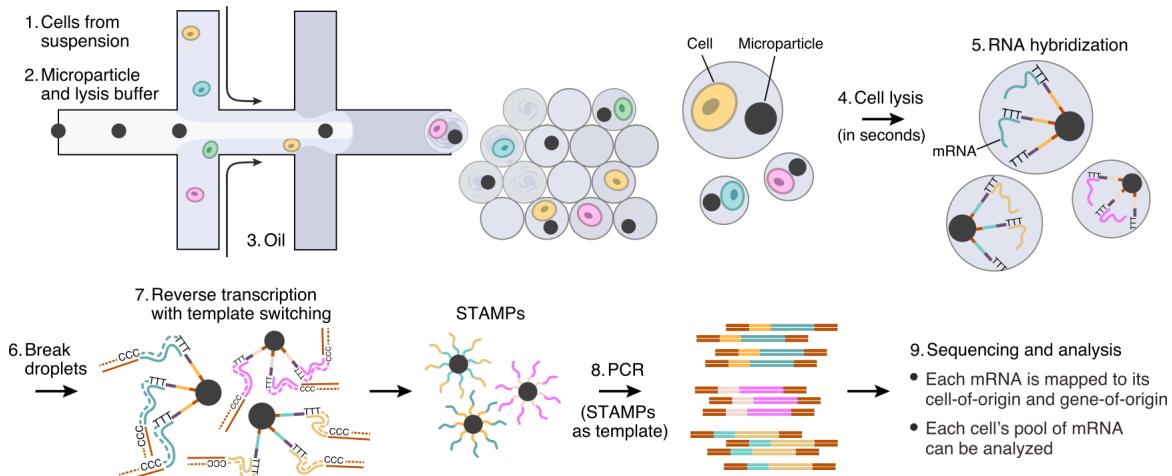


Figure 1.14: Drop-seq library generation. Single cells are encapsulated in a nanolitre volume droplet together with a barcoded resin bead and lysis buffer. The surface of each bead is covered in millions of copies of a bead-specific barcoded poly-dT primer which captures mRNA released by the cells in the droplets. The beads are then pooled in a large volume of water, which prevents further hybridisation outside of the droplet. Immediately after, the hybridised mRNA libraries are copied to the bead primers by template-switching reverse transcription, generating a library of single-cell transcriptomes attached to microparticles (STAMPs). The STAMP libraries are ISPCR-amplified, fragmented, PCR-amplified again and sequenced. The last PCR preferentially amplifies 3' fragments, leading to a 3'-enriched sequencing library. After sequencing, the cellular barcode is used to identify each transcript's cell of origin, and the UMI is used to count individual transcripts. Taken from Macosko et al. (2015).

conventional microwell-based due to the ultra low reaction volumes ($\sim 1 \text{ nl}$). Whereas the price of Smart-seq is estimated at \$3 - \$30 per cell before sequencing, a Drop-seq run indexes single cells at \$0.1 - \$0.65 per cell (Macosko et al., 2015; Ziegenhain et al., 2017). Drop-seq's incredible throughput potential is paired with a number of important drawbacks, a major one being reduced mRNA capture efficiency and sensitivity compared to well-based protocols. Using spiked-in RNA standards, Macosko et al. estimate that Drop-seq captures $\sim 12\%$ of the cellular mRNA and can detect an average of 44 295 mRNA transcripts from 6 722 genes in HEK cells. This is significantly lower than Smart-seq2, which captures $\sim 20\%$ of cellular mRNA and detects 12k genes in the same cell line (Picelli et al., 2013). It is important to note that both Drop-seq and Smart-seq use the same path of template-switching RT followed by ISPCR and subsequent Illumina Nextera XT sequencing library preparation. However, Drop-seq selectively PCR-amplifies 3' fragments since this is where the cellular barcode is located. Other fragments can simply not be assigned to a cell of origin and are therefore suppressed in the final sequencing library.

This means that Drop-seq can absolutely quantify transcripts by mapping the 3' ends to a reference exome, but cannot give information on distal transcript regions.

Another important limitation of Drop-seq is low cell capture efficiency. Since co-encapsulation of two independent particles is a double-Poissonian process, both cell and bead concentrations need to be kept low in order to minimise multiplet encapsulation. At the concentrations used by Macosko et al., the chance for a droplet to receive either a bead or a droplet is ~1-5%. This leads to a dual occupancy rate of only ~0.1% and an effective cell capture efficiency of just ~5% of the input sample. Macosko et al. suggest that Drop-seq's low capture efficiency can be compensated by simply brute-force processing a large number of cells. This is shown in the original publication, where rare cells constituting 0.1%-0.9% of the population are successfully characterised when thousands of cells are analysed.

inDrop

Published concurrently with Drop-seq, inDrop ("indexing droplets") is a different approach to the same concept. Again, single cells are encapsulated in nanolitre droplets together with reverse transcription reagents and beads carrying barcoded poly-dT primers (Klein et al., 2015; Žilionis et al., 2017). The inDrop bead production process, microfluidic chip and workflow are shown in figures 1.15-1.17.

Though both methods share a central concept, several important differences distinguish inDrop from Drop-seq. One of the most important differences can be found in the nature of the primer-loaded beads - whereas Drop-seq uses hard, 30 µm plastic beads, inDrop uses 70 µm soft, deformable hydrogels. These hydrogels are loaded into the microfluidic chip at concentrations ~100 times higher than Drop-seq's beads, allowing them to stack inside the chip's microfluidic channels. Near the end of the funnel, a single, lined-up file of beads is formed and pushed towards the cell flow. As the release of beads can be controlled directly, flows can now be tuned so that nearly 100% of droplets contain at least one bead (Klein et al., 2015; Abate et al., 2009). Due to this "super-Poissonian" stochastic bead loading, inDrop attains a cell-capture rate of near 95%, compared to Drop-seq's 5%. This allows inDrop to be applied to scenarios where the sample does *not* permit brute-force analysis of thousands of cells, for example in a clinical context.

Another important difference between Drop-seq and inDrop is the post-encapsulation library preparation. Whereas Drop-seq is strongly Smart-seq inspired, inDrop essentially follows the CEL-seq/MARS-seq (a high-throughput implementation of CEL-seq) library preparation process (Hashimshony et al., 2012; Jaitin et al., 2014). Like CEL-seq, the inDrop cDNA libraries are pooled, in-vitro transcribed for linear amplification and fragmented/PCR enriched for sequencing. Due to this overnight in-vitro transcription step, inDrop has a higher fixed time cost than Drop-seq. Using ERCC spike-ins, Klein et al. estimate inDrop's mRNA capture efficiency at 7.1% - higher than CEL-seq's 3.4%, but

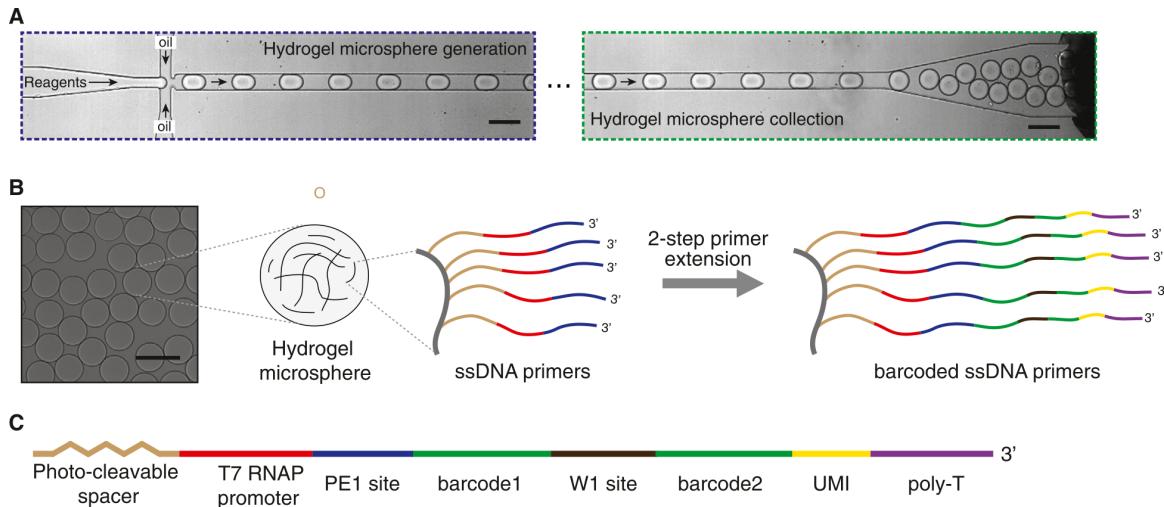


Figure 1.15: inDrop barcoded bead generation. (A) Millions of polyacrylamide microspheres are produced by droplet microfluidics and polymerised to hydrogel beads. (B) Short DNA primers in the acrylamide matrix serve as a PCR priming site and allow a photocleavable spacer, T7 RNA promoter to be appended. After two successive steps of splitting and pooling PCR in 384 wells, each hydrogel carries one of 147 456 (= 384 × 384) possible cellular barcodes. (C) The hydrogel primers also carry a UMI and a poly-T tail. Taken from Klein et al. (2015).

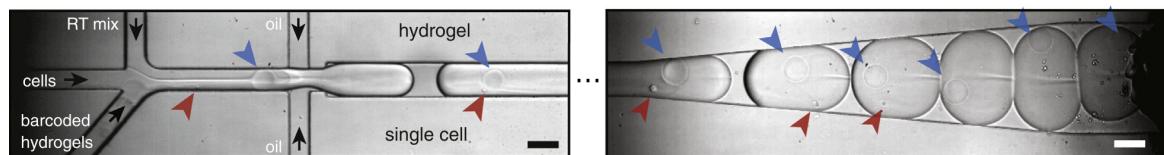


Figure 1.16: inDrop microfluidic chip. A dilute cell suspension is joined by a flow of reverse transcription reagents and stacked hydrogels, and further emulsified by flow-focusing. Taken from Klein et al. (2015).

lower than Smart-seq2 and CEL-seq2's ~20% (Klein et al., 2015; Grün et al., 2014; Picelli et al., 2013; Hashimshony et al., 2016). inDrop is thus plagued by the same low sensitivity issues as Drop-seq.

Droplet Microfluidic Techniques: Key Takeaway

As shown above, droplet-microfluidic single-cell techniques offer several advantages, most notably throughput. A single researcher can now process 10 000 single cells for sequencing in 12 hours, at a fraction of the cost of microwell-based assays. As the actual cell encapsulation step takes only minutes, throughput can be increased further by simply

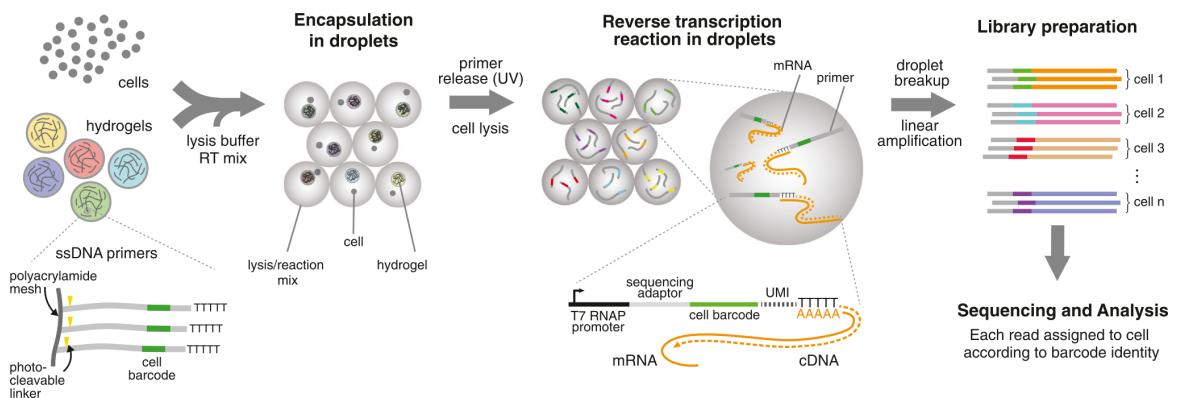


Figure 1.17: inDrop library generation. Similarly to Drop-seq, single cells are encapsulated in a nanolitre volume droplet together with a barcoded hydrogel bead. Each hydrogel bead carries a payload of bead-specific barcoded poly-dT primers covalently incorporated in the gel matrix. These primers are released from the hydrogel using a UV-cleavable chemical linker. The cell is lysed and its mRNA content is reverse transcribed with the released hydrogel primers in the droplet. The resulting barcoded cDNA fragments are pooled and linearly amplified using IVT, similar to CEL-seq. After sequencing, each transcript's cellular barcode and UMI are used to identify cell and transcript of origin. Taken from Klein et al. (2015).

encapsulating more cells and processing their emulsions together. The ultra-low reaction volumes (nl) associated with microfluidics strongly reduce cost, which can be further minimised when cell/bead dual occupancy are optimised. Moreover, these low reaction volumes have been shown to reduce technical noise and increase product yield (Streets et al., 2014). However, despite this supposed increase in reaction efficiency, all droplet microfluidic single-cell techniques exhibit significantly lower sensitivity when compared to their microwell counterparts. Two considerations need to be made regarding this observation. First, sequencing depth, a deciding factor in the data quality of single-cell experiments, is usually much lower in high-throughput methods due to the high number of cells processed per sample. It is often difficult or prohibitively costly to sequence thousands of single-cell transcriptomes from the same tissue sample to saturation. Hopefully, the ongoing decrease of NGS price will mitigate this limitation in the future. Second, Drop-seq and inDrop are often compared with newer generation of microwell methods such as Smart-seq2 and CEL-seq2. Over the years, these methods have been heavily optimised in terms of reagent concentrations, enzymes, incubation times and so on. Several efforts have already been made to improve the sensitivity and capture efficiency of droplet microfluidic methods. We expect that in the coming years, droplet microfluidic single-cell techniques will undergo a similar process of optimisation.

So far, the microfluidics involved in droplet-based single-cell technologies have been very simple, usually consisting of just one round of encapsulation. Žilionis et al. remark

that microfluidic modalities such as droplet splitting, merging and sorting or even multiple encapsulation could lead to droplet versions of more complicated protocols such as ATAC-seq, ChIP-seq, or a combination of pre-existing modalities (Žilionis et al., 2017; Ahn et al., 2006a,b).

1.5 In-Situ Cellular Indexing by Splitting & Pooling

So far, every technique shown has depended on molecular reactions performed on physically separated individual cells. Split-pool methods rely on an entirely different concept - here, groups of cells are in-situ barcoded together in successive rounds of redistributing/splitting and pooling. In these approaches, the cells are never truly individually separated from each other - the nucleic acid content is contained within in the cell itself. Split/pool approaches pose specific advantages and risks, both which will be explained below.

Combinatorial Indexing

In 2015, Shendure Lab produced an interesting cell barcoding strategy which they dubbed single-cell combinatorial indexing (sci) (Cusanovich et al., 2015). In sci-based protocols, a large number of cells is split into microwells where their nucleic acid content is barcoded using in-situ PCR. Then, the cells are pooled and redistributed into new wells where a second barcode is appended to the first. If a sufficiently high number of wells is used in each step, the majority of cells pass through a unique combination of wells, resulting in a unique set of two barcodes nucleic acid content. A visual overview is shown in figure 1.18.

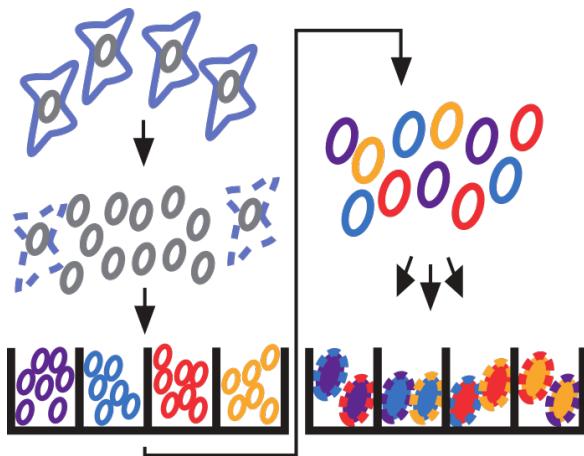


Figure 1.18: Single-cell combinatorial indexing. A suspension of nuclei is equally distributed into wells where they undergo a well-specific barcoding reaction. The nuclei are then pooled and redistributed into a second array of wells, where a second barcode is appended to the first. The majority of nuclei undergo a unique barcoding trajectory, each receiving a unique cellular index that is later used to demultiplex the resulting libraries after sequencing. Adapted from Cusanovich et al. (2015)

The sci approach was first used in Cusanovich et al.'s single-cell combinatorial indexing ATAC-seq (sci-ATAC-seq) method (Cusanovich et al., 2015). In sci-ATAC-seq, a population of ~2000 nuclei is distributed equally into 96 microwells. Here, the nuclei undergo fragmentation with a well-specific custom Tn5 transposase which carries a well-specific barcode. This step results in a chromatin library of which the fragments are tagged with

a first barcode. The fragmented nuclei are then pooled and redistributed into 96 new wells where they undergo PCR amplification with a barcoded primer, resulting in single-cell ATAC-seq libraries. The protocol was later adapted for to sci-RNA-seq (Cao et al., 2017) by performing reverse transcription with well-specific barcodes instead of well-specific Tn5 fragmentation. sci-RNA-seq starts from either single nuclei or fixed and permeabilised cells. The sci approach can thus be used for both scATAC-seq an scRNA-seq.

As with the high-throughput droplet-microfluidic methods described in section 1.4, the sensitivity of sci-based protocols is sub-par. Fiers et al. remark that sci-ATAC-seq retrieves ~2500 chromatin fragment reads per cell compared to ~73 000 reads per cell from the Buenrostro et al. scATAC-seq protocol (see section 1.3) (Fiers et al., 2018). This difference can in part be explained by the difference in per-cell sequencing coverage between the two methods.

A major single-cell multiomics breakthrough was made in 2018 when Cao et al. extracted both RNA-seq and ATAC-seq data from the same nucleus using a sci-based approach (Cao et al., 2018). The method, called single-cell combinatorial indexing for chromatin accessibility and RNA (sci-CAR), simultaneously barcodes the RNA and chromatin content of single nuclei before splitting the sample for both sci-ATAC-seq and sci-RNA-seq processing. Since the cDNA and chromatin reads of the same cell share the same barcode, both libraries can be assigned to their cell of origin. An overview is given in figure 1.19.

Sci-CAR was applied to extract joint transcriptomic and chromatin accessibility data from 4 825 and 11 296 nuclei, a major increase from other multiomics protocols by e.g. Hou et al. (2016) and Clark et al. (2018) respectively, which could only be applied to fewer than 100 cells. To date, sci-CAR remains the only protocol able to simultaneously profile the transcriptome and chromatin accessibility of high numbers of single cells. Despite the intrinsic value of such datasets in the study of for example gene regulatory networks, sci-CAR has not been applied yet by other labs. A hindrance to widespread adoption is the requirement for a custom Tn5 transposase, which is not available commercially. Importantly, sci-CAR also discards half of the available cellular material. Sci-approaches generally require a large number of cells to initiate (10^5 to 10^6 cells), and also retrieve about 7-50% of the sample input, with a high incidence of doublets. A final drawback is that cells/nuclei need to be fixed for sci, prohibiting the use of fresh samples. Ideally, a multiomic method would barcode and sequence the whole transcriptome and chromatin accessibility profile of cells completely separately. Such a protocol has not been published yet.

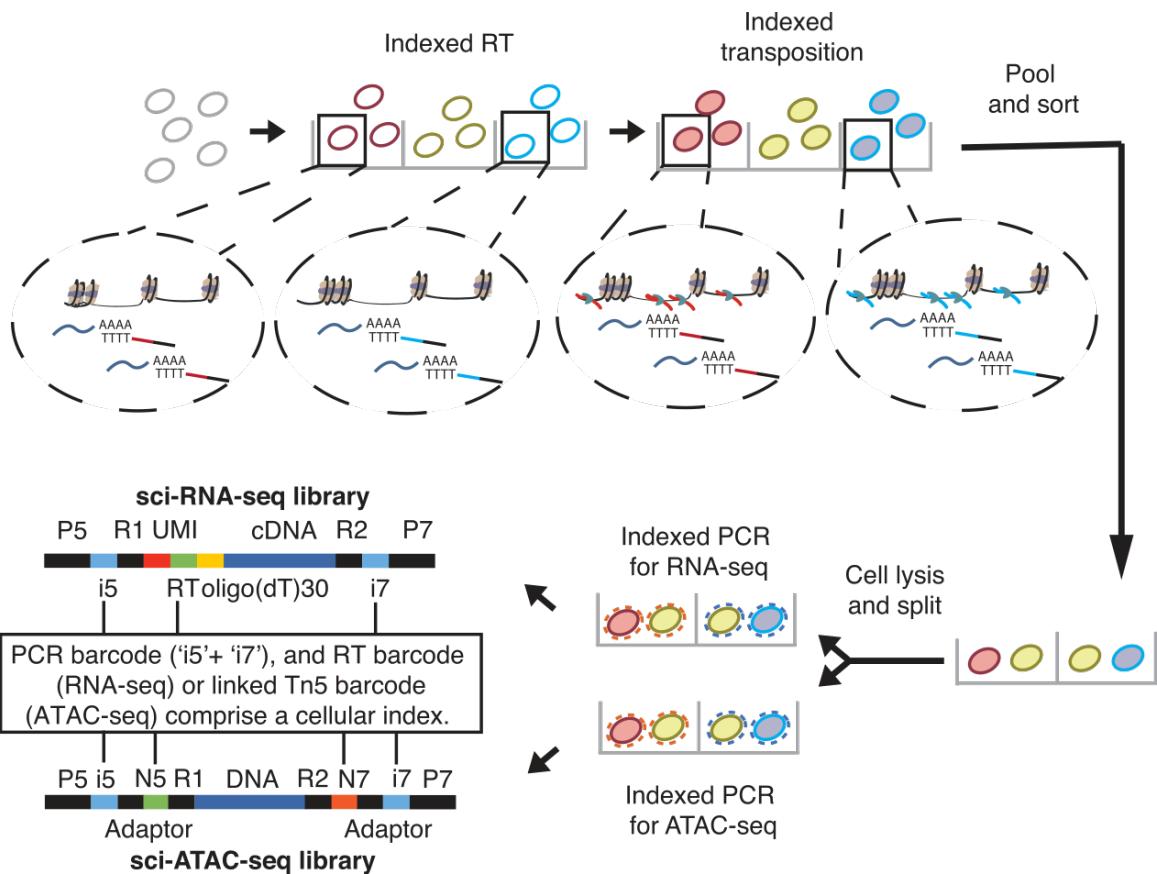


Figure 1.19: Sci-CAR concept. Single nuclei are distributed into a 96 well plate at 5 000 nuclei per well, where their RNA content is barcoded by a reverse transcription reaction with a well-specific primer. Next, the nuclei's chromatin content is fragmented and barcoded in the same well by a custom Tn5 carrying a well-specific adapter. In a crucial step, nuclei are first pooled and then FACS-sorted into 576 new wells where the second cDNA strand synthesis occurs. The nuclei are then lysed and the sample is split equally (without pooling) into dedicated portions for subsequent sci-RNA-seq or sci-ATAC-seq library generation. Both the RNA-seq and ATAC-seq reads can be assigned to the cell of origin based on their barcode combination. Taken from Cao et al. (2018)

SPLiT-seq

In 2018, Rosenberg et al. published split-pool ligation-based transcriptome sequencing (SPLiT-seq), a scRNA-seq technique conceptually identical to the Shendure Lab's single-cell combinatorial indexing scRNA-seq (Cusanovich et al., 2015; Rosenberg et al., 2018). Like sci-RNA-seq, SPLiT-seq indexes fixed cells using successive split/pool in-situ reverse transcription and PCR. However, the addition of two barcode ligation steps leads to four barcoding opportunities per cell, as opposed to sci-RNA-seq's two. This increases overall hands-on time, but allows SPLiT-seq to scale to large numbers of cells more rapidly than sci-RNA-seq. An overview of the SPLiT-seq barcoding process is given in figure 1.20.

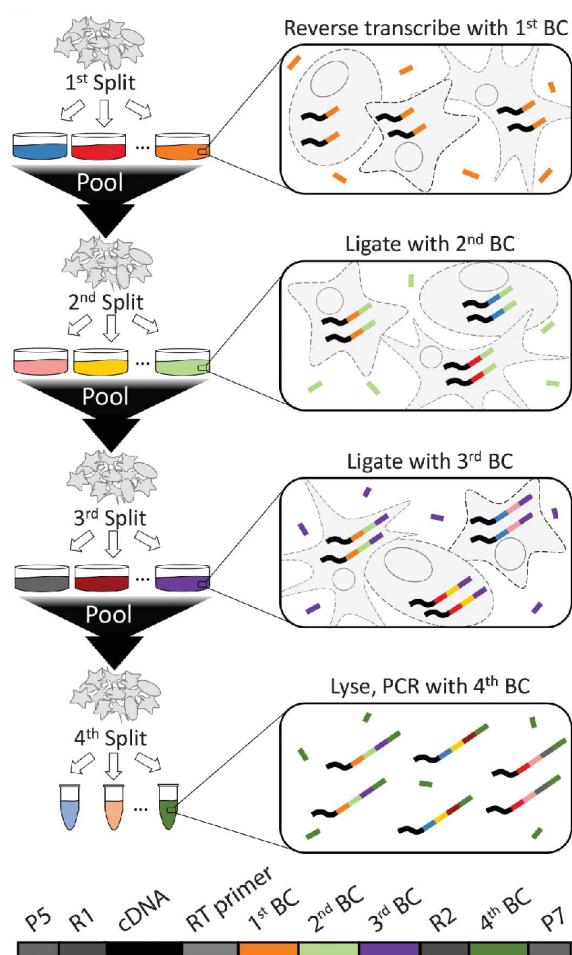


Figure 1.20: Split-pool ligation-based transcriptome sequencing (SPLiT-seq). Methanol-fixed and permeabilised cells are split into 48 wells where they undergo in-situ reverse transcription with a well-specific barcode. Two additional barcodes are in-situ ligated to the cellular cDNA libraries in subsequent splitting and pooling steps. The cells are then lysed and a final barcode is introduced in the sequencing library indexing PCR. Taken from Rosenberg et al. (2018).

SPLiT-seq's main advantage is unprecedented throughput. In theory, SPLiT-seq can barcode more than 6 million cells using just four different microwell plates. It is not unthinkable that future analysis of large cell samples will necessitate experiments of this magnitude. Today, however, the current cost of sequencing and modern computational possibilities prohibit the execution of such ultra-high throughput experiments.

In-Situ Cellular Indexing: Key Takeaway

The ultra-high throughput capabilities of combinatorial indexing approaches hold promises for future experiments where 10^4 - 10^5 cells are analysed. Due to the limits on computation and sequencing technology, such dazzling numbers are unthinkable today. In a sense, the main limiting factor to contemporary single-cell research is by no means the number of cells analysed in a single run. Rather, high-throughput (10^3 - 10^4 cells), high-sensitivity assays seem to be a more important goal. It is for this reason, combined with the low sensitivity associated with combinatorial indexing and the use of proprietary reagents, that sci and SPLiT-seq have not been extensively applied outside of their respective labs yet.

1.6 Applications of Single-Cell Omics

Cell Typing

One of the major applications of single-cell RNA-seq is de-novo cell typing. The *Tabula Muris* is an example of such a large scale effort. The Tabula Muris Consortium sequenced the individual transcriptomes of 55k mouse cells spread over all major organs using 10x Chromium scRNA-seq, and another 45k cells using FACS + Smart-seq2. The resulting cells were computationally separated (figure 1.21) and analysed, identifying several distinct new cell types and transcription factor networks, and revealing previously undiscovered roles of known genes.

Analysing such a large amount of cells spanning over multiple organs allows for direct comparison of the resulting data, but is a major undertaking using the current state of the art technology. The tiered approach employed here, where low sequencing coverage droplet microfluidic methods are used to rapidly form a global picture, and more plate-based methods are used to analyse carefully filtered pre-defined populations at higher sequencing depth, will most likely be the key to undertaking large scale sequencing operations such as these in the future.

Development

During embryogenesis, a single totipotent cell will divide, and its descendants will gradually lose potency and differentiate into the cell types that make up the organism. The acquisition of cell identity, function and morphology is for a large part controlled through differential gene expression (Farrell et al., 2018), making high-throughput scRNA-seq a valuable tool in our effort to understand this complex process.

Farrell et al. analysed the individual transcriptomes of 39k zebrafish cells from 12 different embryonic stages using Drop-seq and developed novel computational methods

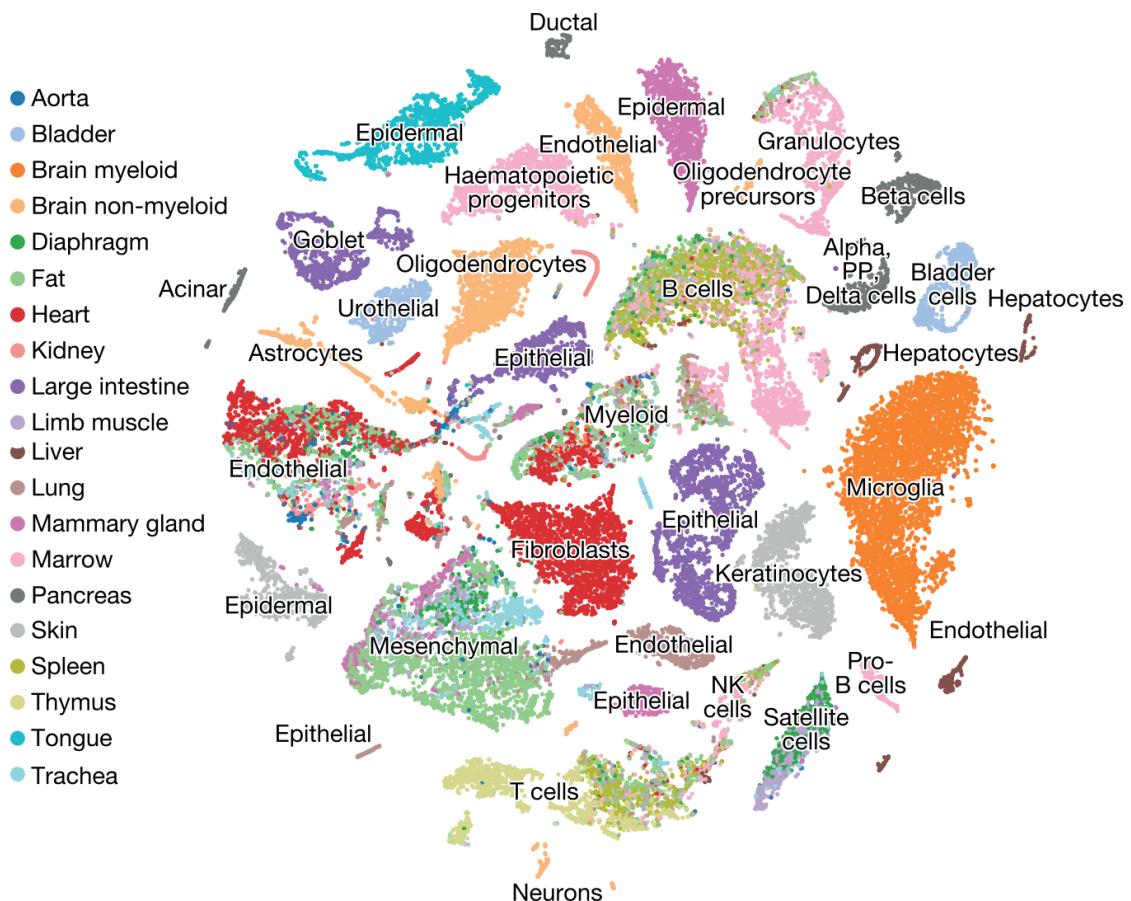


Figure 1.21: Murine cell atlas. tSNE visualisation of 45k single cell transcriptomes from 20 mouse organs. Taken from Tabula Muris Consortium (2018)

(a combination of URD and Seurat) to map the differentiation process of 25 cell lineages during embryonic development (figure 1.22). The tips of the resulting tree corresponded to previously known cell types in terms of marker gene expression, and much of what was already known about embryonic development was reflected in the tree's branching structure.

In addition to agreement with the canonical embryological knowledge, the systems biology approach revealed new candidate regulators of the differentiation process and how the spatial organisation of the developing embryo may be decided earlier in development than previously thought. Importantly, the computational framework developed by Farrell et al. can be used for the reconstruction of the developmental trajectories of other biological model organisms.

Another new computational tool was pioneered by La Manno et al. in 2018, who realised that the time derivative of gene expression profiles across samples taken on multiple time points could be inferred from the identity of unspliced and mature mRNA transcripts.

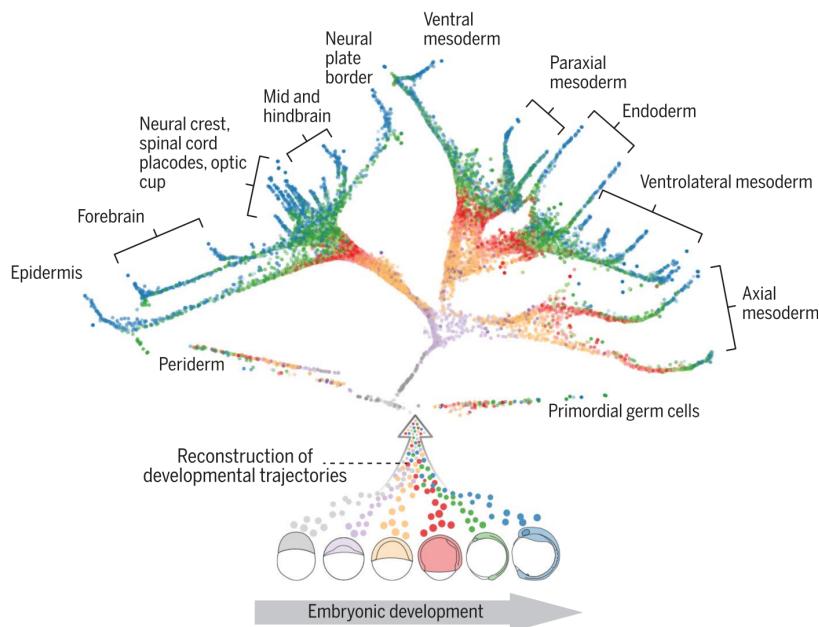


Figure 1.22: Developmental trajectories in Zebrafish.
Taken from Farrell et al. (2018)

They found that in several open-access single-cell RNA-seq datasets from samples taken at different time points, the population of immature mRNA transcripts was present as mature transcripts at the next time point. La Manno et al. then used this information to project the developmental flow on regular tSNE plots (figure 1.23).

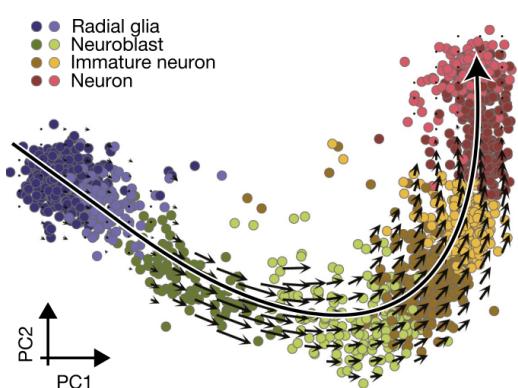


Figure 1.23: RNA velocity in human neurogenesis. Taken from La Manno et al. (2018)

As this method was first performed on previously published data, this publication proves that new information can often be extracted from the datasets generated by single-cell sequencing experiments when approaching them from a different angle. When a sequencing dataset is generated, it exists permanently as a vector space in which future in-silico "experiments" can be performed, as opposed to classical experiments where samples may have a limited lifetime and interrogating the same sample at a later time point may be difficult. In the future, we may thus see more examples where discoveries are made in older datasets thought to be exhausted of information.

Disease

The final application of single cell technology, and its main accelerator today, is human medicine. Several areas of medicine will directly benefit from single-cell resolution data, notably those where cell diversity is most impactful - such as cancer and brain disease. Today's leading effort in pushing single-cell technology to applications in medicine is the human cell atlas (HCA) (Regev et al., 2017). This project, which aims to approach the Human Genome Project in magnitude and scope, aims to comprehensively map all cells present in the human body. Such an atlas could be used as a reference for patient sample comparisons and help understand the cellular mechanisms behind disease. Figure 1.24 shows the key impact aims of the HCA.

A major application of the HCA will be targeted drug discovery. Comparing the genetic profile of healthy cell samples to the reference cell atlas would provide leads for possible new drugs. Single-cell experiments could also be used to compare in-vitro generated cell cultures with the reference profile of the human tissue they attempt to mimic, accelerating the development of engineered tissues for regenerative medicine. The project's first draft, published in 2017, profiles a subset of cells from the tissues and organs that hold the most promise for immediate application. The first draft includes 'only' 30 to 100 million cells, a fraction of the projected 10 billion for the final atlas (The Human Cell Atlas Consortium, 2017).

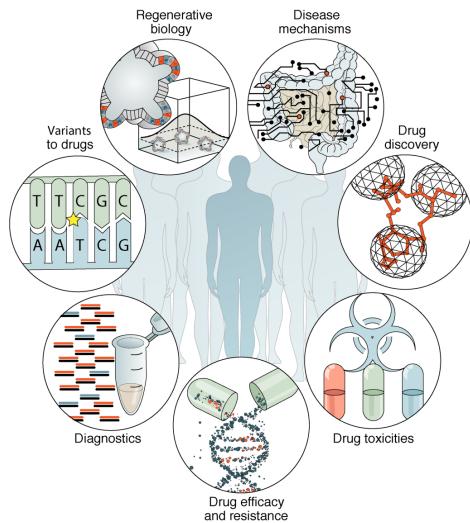


Figure 1.24: Impact areas of the human cell atlas. Taken from The Human Cell Atlas Consortium (2017)

1.7 Single-cell Omics: Current Progress and Future Perspectives

As shown in the previous sections, the past few years have seen an explosion in single-cell research and technology. Single-cell transcriptomics, genomics and epigenomics are already a fact, and proteomics and spatial omics are looming over the horizon. Single cell technology has allowed us to capture snapshots of complex cellular processes such as gene regulation, gene expression, tissue development, and the origins of disease. Even in its infancy, single-cell technology has proven its value in almost every area of biology involving cells. We have uncovered regulatory pathways and mapped the (partial) transcriptome atlases of several model organisms, and are rapidly moving forward to achieving the same in humans. However, much remains to be done. It is painstakingly clear that the vast majority of today's single-cell technology can be further improved upon. Our most cutting-edge high-throughput techniques are able to capture only a small fraction of the information embedded in a single cell, with low reproducibility and high noise levels. Indeed, we are asking much of our bioinformatician colleagues. Equally concerning is how virtually every single-cell technique to date can only interrogate a single omics modality. It is therefore difficult to, for example, systematically relate a cell's transcriptome to its epigenome. The clear-cut next step is therefore to fine-tune existing methods and to move forward to single-cell multi-omics. Extracting information on several omics modalities from the same single cells, at high fidelity *and* close spatial resolution will help us understand the absolute fundamentals of life - and how we may improve it.

Bibliography

- Abate, A. R., C. . Chen, J. J. Agresti, and D. A. Weitz
2009. Beating poisson encapsulation statistics using close-packed ordering. *Lab on a Chip*, 9(18):2628–2631.
- Ahn, K., J. Agresti, H. Chong, M. Marquez, and D. A. Weitz
2006a. Electrocoalescence of drops synchronized by size-dependent flow in microfluidic channels. *Applied Physics Letters*, 88(26).
- Ahn, K., C. Kerbage, T. P. Hunt, R. M. Westervelt, D. R. Link, and D. A. Weitz
2006b. Dielectrophoretic manipulation of drops for high-speed microfluidic sorting devices. *Applied Physics Letters*, 88(2):1–3.
- Allazetta, S. and M. P. Lutolf
2015. Stem cell niche engineering through droplet microfluidics. *Current opinion in biotechnology*, 35:86–93.
- Azizi, F., S. Clouthier, and M. S. Wicha
2014. The promise of single cell omics for onco-therapy. *Journal of Molecular and Genetic Medicine*, 8(3):121–122.
- Bannister, A. J. and T. Kouzarides
2011. Regulation of chromatin by histone modifications. *Cell research*, 21(3):381–395.
- Bengtsson, M., A. Ståhlberg, P. Rorsman, and M. Kubista
2005. Gene expression profiling in single cells from the pancreatic islets of langerhans reveals lognormal distribution of mrna levels. *Genome research*, 15(10):1388–1392.
- Brouzes, E., M. Medkova, N. Savenelli, D. Marran, M. Twardowski, J. B. Hutchison, J. M. Rothberg, D. R. Link, N. Perrimon, and M. L. Samuels
2009. Droplet microfluidic technology for single-cell high-throughput screening. *Proceedings of the National Academy of Sciences of the United States of America*, 106(34):14195–14200.
- Buenrostro, J. D., P. G. Giresi, L. C. Zaba, H. Y. Chang, and W. J. Greenleaf
2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, dna-binding proteins and nucleosome position. *Nature Methods*, 10(12):1213–1218.

- Buenrostro, J. D., B. Wu, U. M. Litzenburger, D. Ruff, M. L. Gonzales, M. P. Snyder, H. Y. Chang, and W. J. Greenleaf
2015. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*, 523(7561):486–490.
- Cao, J., D. A. Cusanovich, V. Ramani, D. Aghamirzaie, H. A. Pliner, A. J. Hill, R. M. Daza, J. L. McFaline-Figueroa, J. S. Packer, L. Christiansen, F. J. Steemers, A. C. Adey, C. Trapnell, and J. Shendure
2018. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science*, 361(6409):1380–1385.
- Cao, J., J. S. Packer, V. Ramani, D. A. Cusanovich, C. Huynh, R. Daza, X. Qiu, C. Lee, S. N. Furlan, F. J. Steemers, A. Adey, R. H. Waterston, C. Trapnell, and J. Shendure
2017. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science*, 357(6352):661–667.
- Chen, X., R. J. Miragaia, K. N. Natarajan, and S. A. Teichmann
2018. A rapid and robust method for single cell chromatin accessibility profiling. *Nature Communications*, 9(1).
- Clark, S. J., R. Argelaguet, C. Kapourani, T. M. Stubbs, H. J. Lee, C. Alda-Catalinas, F. Krueger, G. Sanguinetti, G. Kelsey, J. C. Marioni, O. Stegle, and W. Reik
2018. Scnmt-seq enables joint profiling of chromatin accessibility dna methylation and transcription in single cells e. *Nature Communications*, 9(1).
- Corces, M. R., A. E. Trevino, E. G. Hamilton, P. G. Greenside, N. A. Sinnott-Armstrong, S. Vesuna, A. T. Satpathy, A. J. Rubin, K. S. Montine, B. Wu, A. Kathiria, S. W. Cho, M. R. Mumbach, A. C. Carter, M. Kasowski, L. A. Orloff, V. I. Risca, A. Kundaje, P. A. Khavari, T. J. Montine, W. J. Greenleaf, and H. Y. Chang
2017. An improved atac-seq protocol reduces background and enables interrogation of frozen tissues. *Nature Methods*, 14(10):959–962. Cited By :43.
- Cusanovich, D. A., R. Daza, A. Adey, H. A. Pliner, L. Christiansen, K. L. Gunderson, F. J. Steemers, C. Trapnell, and J. Shendure
2015. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science*, 348(6237):910–914.
- Davie, K., J. Janssens, D. Koldere, M. De Waegeneer, U. Pech, Å. Kreft, S. Aibar, S. Makhzami, V. Christiaens, C. Bravo González-Blas, S. Poovathingal, G. Hulselmans, K. I. Spanier, T. Moerman, B. Vanspauwen, S. Geurs, T. Voet, J. Lammertyn, B. Thienpont, S. Liu, N. Konstantinides, M. Fiers, P. Verstreken, and S. Aerts
2018. A single-cell transcriptome atlas of the aging drosophila brain. *Cell*, 174(4):982–998.e20.

- Eberwine, J., H. Yeh, K. Miyashiro, Y. Cao, S. Nair, R. Finnell, M. Zettel, and P. Coleman
1992. Analysis of gene expression in single live neurons. *Proceedings of the National Academy of Sciences of the United States of America*, 89(7):3010–3014.
- Fan, H. C., G. K. Fu, and S. P. A. Fodor
2015. Combinatorial labeling of single cells for gene expression cytometry. *Science*, 347(6222).
- Farrell, J. A., Y. Wang, S. J. Riesenfeld, K. Shekhar, A. Regev, and A. F. Schier
2018. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science*, 360(6392).
- Fiers, M. W. E. J., L. Minnoye, S. Aibar, C. B. González-Blas, Z. K. Atak, and S. Aerts
2018. Mapping gene regulatory networks from single-cell omics data. *Briefings in Functional Genomics*, 17(4):246–254.
- Fluidigm Website
2019. Fluidigm Website - <https://www.fluidigm.com/publications/c1>.
- Gaulton, K. J., T. Nammo, L. Pasquali, J. M. Simon, P. G. Giresi, M. P. Fogarty, T. M. Panhuis, P. Mieczkowski, A. Secchi, D. Bosco, T. Berney, E. Montanya, K. L. Mohlke, J. D. Lieb, and J. Ferrer
2010. A map of open chromatin in human pancreatic islets. *Nature genetics*, 42(3):255–259.
- Gierahn, T. M., M. H. Wadsworth, T. K. Hughes, B. D. Bryson, A. Butler, R. Satija, S. Fortune, J. Christopher Love, and A. K. Shalek
2017. Seq-well: Portable, low-cost rna sequencing of single cells at high throughput. *Nature Methods*, 14(4):395–398.
- Giresi, P. G., J. Kim, R. M. McDaniell, V. R. Iyer, and J. D. Lieb
2007. Faire (formaldehyde-assisted isolation of regulatory elements) isolates active regulatory elements from human chromatin. *Genome research*, 17(6):877–885.
- Grün, D., L. Kester, and A. Van Oudenaarden
2014. Validation of noise models for single-cell transcriptomics. *Nature Methods*, 11(6):637–640.
- Grün, D. and A. Van Oudenaarden
2015. Design and analysis of single-cell sequencing experiments. *Cell*, 163(4):799–810.
- Hashimshony, T., N. Senderovich, G. Avital, A. Klochendler, Y. de Leeuw, L. Anavy, D. Gennert, S. Li, K. J. Livak, O. Rozenblatt-Rosen, Y. Dor, A. Regev, and I. Yanai
2016. Cel-seq2: Sensitive highly-multiplexed single-cell rna-seq. *Genome biology*, 17(1).

- Hashimshony, T., F. Wagner, N. Sher, and I. Yanai
2012. Cel-seq: Single-cell rna-seq by multiplexed linear amplification. *Cell Reports*, 2(3):666–673.
- Hood, L., J. R. Heath, M. E. Phelps, and B. Lin
2004. Systems biology and new technologies enable predictive and preventative medicine. *Science*, 306(5696):640–643.
- Hou, Y., H. Guo, C. Cao, X. Li, B. Hu, P. Zhu, X. Wu, L. Wen, F. Tang, Y. Huang, and J. Peng
2016. Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell research*, 26(3):304–319.
- Huebner, A., M. Srisa-Art, D. Holt, C. Abell, F. Hollfelder, A. J. DeMello, and J. B. Edel
2007. Quantitative detection of protein expression in single cells using droplet microfluidics. *Chemical Communications*, 12:1218–1220.
- Jaenisch, R. and A. Bird
2003. Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nature genetics*, 33(3S):245–254.
- Jaitin, D. A., E. Kenigsberg, H. Keren-Shaul, N. Elefant, F. Paul, I. Zaretsky, A. Mildner, N. Cohen, S. Jung, A. Tanay, and I. Amit
2014. Massively parallel single-cell rna-seq for marker-free decomposition of tissues into cell types. *Science*, 343(6172):776–779.
- Johannes, F., V. Colot, and R. C. Jansen
2008. Epigenome dynamics: A quantitative genetics perspective. *Nature Reviews Genetics*, 9(11):883–890.
- Kivioja, T., A. Vähärautio, K. Karlsson, M. Bonke, M. Enge, S. Linnarsson, and J. Taipale
2012. Counting absolute numbers of molecules using unique molecular identifiers. *Nature Methods*, 9(1):72–74.
- Klein, A. M., L. Mazutis, I. Akartuna, N. Tallapragada, A. Veres, V. Li, L. Peshkin, D. A. Weitz, and M. W. Kirschner
2015. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*, 161(5):1187–1201.
- Kolodziejczyk, A. A., J. K. Kim, V. Svensson, J. C. Marioni, and S. A. Teichmann
2015. The technology and biology of single-cell rna sequencing. *Molecular cell*, 58(4):610–620.
- Kornberg, R. D.
1974. Chromatin structure: A repeating unit of histones and dna. *Science*, 184.

- Kouzarides, T.
2007. Chromatin modifications and their function. *Cell*, 128(4):693–705.
- La Manno, G., R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, V. Petukhov, K. Lidschreiber, M. E. Kastriti, P. Lönnberg, A. Furlan, J. Fan, L. E. Borm, Z. Liu, D. van Bruggen, J. Guo, X. He, R. Barker, E. Sundström, G. Castelo-Branco, P. Cramer, I. Adameyko, S. Linnarsson, and P. V. Kharchenko
2018. Rna velocity of single cells. *Nature*, 560(7719):494–498.
- Macosko, E. Z., A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev, and S. A. McCarroll
2015. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, 161(5):1202–1214.
- Mazutis, L., J. Gilbert, W. L. Ung, D. A. Weitz, A. D. Griffiths, and J. A. Heyman
2013. Single-cell analysis and sorting using droplet-based microfluidics. *Nature Protocols*, 8(5):870–891.
- Mortazavi, A., B. A. Williams, K. McCue, L. Schaeffer, and B. Wold
2008. Mapping and quantifying mammalian transcriptomes by rna-seq. *Nature Methods*, 5(7):621–628.
- Patterson, S. D. and R. H. Aebersold
2003. Proteomics: The first decade and beyond. *Nature genetics*, 33(3S):311–323.
- Patti, G. J., O. Yanes, and G. Siuzdak
2012. Innovation: Metabolomics: the apogee of the omics trilogy. *Nature Reviews Molecular Cell Biology*, 13(4):263–269.
- Picelli, S.
2017. Single-cell rna-sequencing: The future of genome biology is now. *RNA Biology*, 14(5):637–650.
- Picelli, S., Å. K. Björklund, O. R. Faridani, S. Sagasser, G. Winberg, and R. Sandberg
2013. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nature Methods*, 10(11):1096–1100.
- Picelli, S., Å. K. Björklund, B. Reinius, S. Sagasser, G. Winberg, and R. Sandberg
2014a. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome research*, 24(12):2033–2040.
- Picelli, S., O. R. Faridani, Å. K. Björklund, G. Winberg, S. Sagasser, and R. Sandberg
2014b. Full-length rna-seq from single cells using smart-seq2. *Nature Protocols*, 9(1):171–181.

Ramsköld, D., S. Luo, Y. . Wang, R. Li, Q. Deng, O. R. Faridani, G. A. Daniels, I. Khrebtukova, J. F. Loring, L. C. Laurent, G. P. Schroth, and R. Sandberg

2012. Full-length mrna-seq from single-cell levels of rna and individual circulating tumor cells. *Nature biotechnology*, 30(8):777–782.

Regev, A., S. A. Teichmann, E. S. Lander, I. Amit, C. Benoist, E. Birney, B. Bodenmiller, P. Campbell, P. Carninci, M. Clatworthy, H. Clevers, B. Deplancke, I. Dunham, J. Eberwine, R. Eils, W. Enard, A. Farmer, L. Fugger, B. Göttgens, N. Hacohen, M. Haniffa, M. Hemberg, S. Kim, P. Klennerman, A. Kriegstein, E. Lein, S. Linnarsson, E. Lundberg, J. Lundeberg, P. Majumder, J. C. Marioni, M. Merad, M. Mhlanga, M. Nawijn, M. Netea, G. Nolan, D. Pe'er, A. Phillipakis, C. P. Ponting, S. Quake, W. Reik, O. Rozenblatt-Rosen, J. Sanes, R. Satija, T. N. Schumacher, A. Shalek, E. Shapiro, P. Sharma, J. W. Shin, O. Stegle, M. Stratton, M. J. T. Stubbington, F. J. Theis, M. Uhlen, A. Van Oudenaarden, A. Wagner, F. Watt, J. Weissman, B. Wold, R. Xavier, N. Yosef, and H. Participants

2017. The human cell atlas. *eLife*, 6.

Rosenberg, A. B., C. M. Roco, R. A. Muscat, A. Kuchina, P. Sample, Z. Yao, L. T. Graybuck, D. J. Peeler, S. Mukherjee, W. Chen, S. H. Pun, D. L. Sellers, B. Tasic, and G. Seelig

2018. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science*, 360(6385):176–182.

Schones, D. E. and K. Zhao

2008. Genome-wide approaches to studying chromatin modifications. *Nature Reviews Genetics*, 9(3):179–191.

Shalek, A. K., R. Satija, J. Shuga, J. J. Trombetta, D. Gennert, D. Lu, P. Chen, R. S. Gertner, J. T. Gaublomme, N. Yosef, S. Schwartz, B. Fowler, S. Weaver, J. Wang, X. Wang, R. Ding, R. Raychowdhury, N. Friedman, N. Hacohen, H. Park, A. P. May, and A. Regev

2014. Single-cell rna-seq reveals dynamic paracrine control of cellular variation. *Nature*, 510(7505):363–369.

Song, L. and G. E. Crawford

2010. Dnase-seq: A high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harbor Protocols*, 5(2).

Song, L., Z. Zhang, L. L. Grasfeder, A. P. Boyle, P. G. Giresi, B. . Lee, N. C. Sheffield, S. Gräf, M. Huss, D. Keefe, Z. Liu, D. London, R. M. McDaniell, Y. Shibata, K. A. Showers, J. M. Simon, T. Vales, T. Wang, D. Winter, Z. Zhang, N. D. Clarke, E. Birney, V. R. Iyer, G. E. Crawford, J. D. Lieb, and T. S. Furey

2011. Open chromatin defined by dnasei and faire identifies regulatory elements that shape cell-type identity. *Genome research*, 21(10):1757–1767.

- Streets, A. M., X. Zhang, C. Cao, Y. Pang, X. Wu, L. Xiong, L. Yang, Y. Fu, L. Zhao, F. Tang, and Y. Huang
2014. Microfluidic single-cell whole-transcriptome sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, 111(19):7048–7053.
- Stuart, T. and R. Satija
2019. Integrative single-cell analysis. *Nature Reviews Genetics*.
- Tabula Muris Consortium
2018. Single-cell transcriptomics of 20 mouse organs creates a tabula muris. *Nature*, 562(7727):367–372.
- Tang, F., C. Barbacioru, Y. Wang, E. Nordman, C. Lee, N. Xu, X. Wang, J. Bodeau, B. B. Tuch, A. Siddiqui, K. Lao, and M. A. Surani
2009. mrna-seq whole-transcriptome analysis of a single cell. *Nature Methods*, 6(5):377–382.
- Tang, F., K. Lao, and M. A. Surani
2011. Development and applications of single-cell transcriptome analysis. *Nature Methods*, 8(4 SUPPL.):S6–S11.
- The Human Cell Atlas Consortium
2017. The human cell atlas whitepaper.
- Wang, D. and S. Bodovitz
2010. Single cell analysis: The new frontier in 'omics'. *Trends in biotechnology*, 28(6):281–290.
- Wang, Z., M. Gerstein, and M. Snyder
2009. Rna-seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1):57–63.
- Whitesides, G. M.
2006. The origins and the future of microfluidics. *Nature*, 442(7101):368–373.
- Wu, A. R., N. F. Neff, T. Kalisky, P. Dalerba, B. Treutlein, M. E. Rothenberg, F. M. Mburu, G. L. Mantalas, S. Sim, M. F. Clarke, and S. R. Quake
2014. Quantitative assessment of single-cell rna-sequencing methods. *Nature Methods*, 11(1):41–46.
- Xia, Y. and G. M. Whitesides
1998. Soft lithography. *Angewandte Chemie - International Edition*, 37(5):550–575.
- Yuan, J. and P. A. Sims
2016. An automated microwell platform for large-scale single cell rna-seq. *Scientific Reports*, 6.

Ziegenhain, C., B. Vieth, S. Parekh, B. Reinius, A. Guillaumet-Adkins, M. Smets, H. Leonhardt, H. Heyn, I. Hellmann, and W. Enard

2017. Comparative analysis of single-cell rna sequencing methods. *Molecular cell*, 65(4):631–643.e4.

Žilionis, R., J. Nainys, A. Veres, V. Savova, D. Zemmour, A. M. Klein, and L. Mazutis

2017. Single-cell barcoding and sequencing using droplet microfluidics. *Nature Protocols*, 12(1):44–73.