Handbook of Genetic Diagnostic Technologies in Reproductive Medicine

Improving Patient Success Rates and Infant Health

Second Edition

Edited by

Carlos Simón MD

Professor of Obstetrics and Gynecology at the University of Valencia
Senior Lecturer, Department of Obstetrics and Gynecology
Beth Israel Deaconess Medical Center, Harvard University Medical School
Boston, Massachusetts, USA
Head of the Scientific Advisory Board of Igenomix
Valencia, Spain

Carmen Rubio PhD

Research Senior Director for Embryo Chromosomal Studies, Igenomix
Scientist at Igenomix Foundation, INCLIVA, University of Valencia
Valencia, Spain

First published in 2022

ISBN: 978-0-367-45718-1 (hbk)
ISBN: 978-1-032-24490-7 (pbk)
ISBN: 978-1-003-02494-1 (ebk)

Chapter 2

Genetics at the Cell Level: The Human Cell Atlas

Valentina Lorenzi and Roser Vento-Tormo

DOI: 10.1201/9781003024941-2

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

Genetics at the Cell Level

2

The Human Cell Atlas

Valentina Lorenzi and Roser Vento-Tormo

UNCOVERING CELLULAR HETEROGENEITY IN THE HUMAN BODY

The Human Cell Atlas (HCA) is an international consortium established at the end of 2016 with the mission of mapping and characterizing all cells in the human body in terms of their distinctive patterns of gene expression, physiological states, and location (Rozenblatt-Rosen *et al.*, 2017); (Regev *et al.*, 2017) (www.humancellatlas.org). It is an open and collaborative initiative, bringing together experts across multiple disciplines, and is meant to progress in phases. Recently, the first maps focused on specific organs and tissues (Ordovas-Montanes *et al.*, 2018; Vento-Tormo *et al.*, 2018; Popescu *et al.*, 2019; Ramachandran *et al.*, 2019; Smillie *et al.*, 2019; Stewart *et al.*, 2019; Vieira Braga *et al.*, 2019) have laid the foundations for further work aimed at completing the atlas to include at least ten billion cells that fully represent the world's diversity.

The desire to comprehensively characterize and classify cells into distinct types is not new. Research has long been focused on cataloging cells based on their shape, location, biological function, and molecular components with increasing levels of detail. Only recently, however, have advances in single-cell genomic technologies made it possible to undertake the high-resolution, unbiased, and systematic characterization of cells in the human body which is at the heart of the HCA (Svensson, Vento-Tormo, and Teichmann, 2018). Such revolutionary technologies allow us to profile the genome and the genome products—including chromatin architecture, RNA transcripts, and proteins from single cells (Lander, 1996).

Genomic profiling technologies have been used for many years to describe ensembles of cells in a tissue, called bulk tissue samples, but it was not until the beginning of the last decade that they could be employed in the characterization of individual cells. This represents a dramatic step forward in the study of cellular heterogeneity and has led to the discovery of new cell types, thus resonating with the HCA's main goal of creating a comprehensive, integrated catalog of all the cell types present in the human body.

DOI: 10.1201/9781003024941-2

From Bulk to Single Cells

No two cells in a body are exactly the same. Although the set of genetic instructions is shared across all the cells in an individual, the form and function of each cell, even within the same cell type, are unique. Such uniqueness is accomplished by finely regulating gene expression, meaning the quantity of RNAs and proteins that are produced from each gene in a given cell.

Traditionally, gene expression has been studied by taking thousands of cells in a tissue and treating them as a single, homogeneous entity. This approach measures the *aggregated* expression level for each gene across the population of input cells (Wang, Gerstein, and Snyder, 2009). Although very useful for comparing expression signatures of a tissue across different conditions, such as health and disease, bulk studies inevitably miss out on what makes each cell unique (Raj and van Oudenaarden, 2008).

The last decade has witnessed the rise of new, single-cell sequencing technologies that allow researchers to capture the cell diversity within a tissue, a feature which is lost in bulk populations (Kulkarni *et al.*, 2019). These technologies can measure the *distribution* of expression levels for each gene in each cell across a population, thus providing answers to new biological questions in which cell-specific changes in gene expression play a major role. The ability to discern individual cells within a tissue can, for instance, lead to the identification of new cell types and improve our understanding of organ development and homeostasis (Figure 2.1).

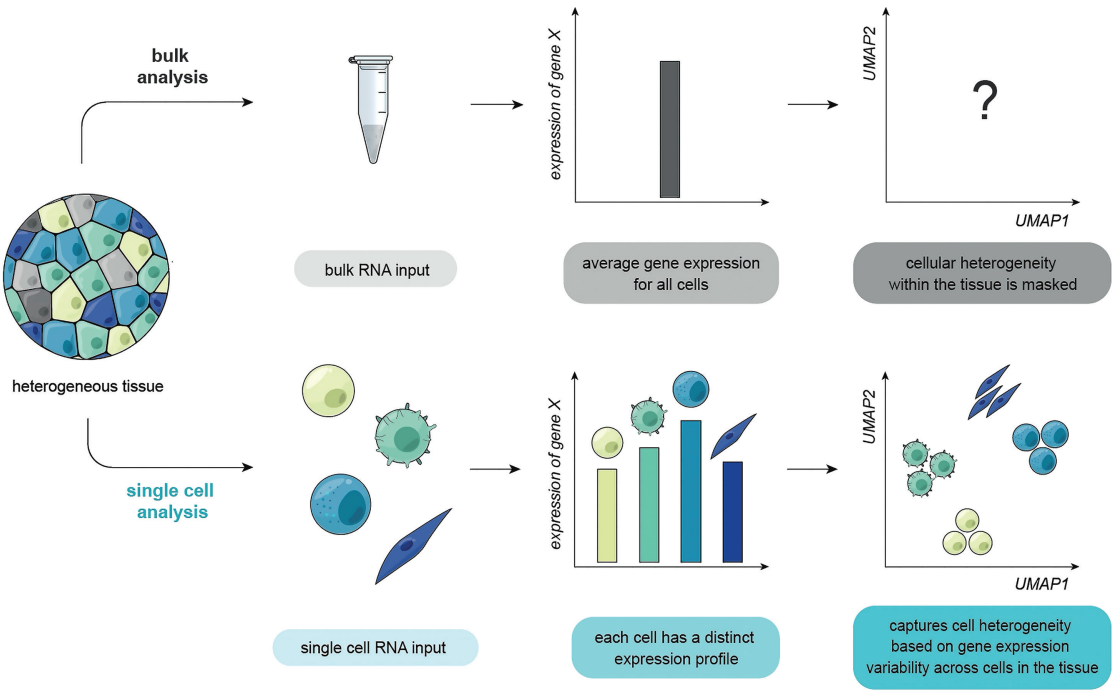

FIGURE 2.1 Schematic representation of the major differences between bulk and single-cell transcriptomic analysis. Bulk transcriptomic analysis measures the average expression for each gene across all cells in a tissue, making it unfit for unraveling cellular heterogeneity. In contrast, single-cell transcriptomic analysis outputs the gene expression profile specific to each cell, thus capturing cellular heterogeneity within a tissue.

Finding New Cell Types

Our knowledge about the types of cells present in the human body is still limited and discovering new, often rare, cell types poses a significant challenge. Indeed, the rarity of these cell types makes them hard to identify *via* bulk transcriptomic technologies, whereas targeted, single-cell molecular techniques that interrogate gene expression of a few selected cells are also ill suited because these new cell types are, by definition, unknown. This is where single-cell sequencing technologies come into play, offering an unbiased and novel understanding of the cellular composition of tissues.

Notably, in 2018, transcriptomic profiling of human bronchial epithelial cells with single-cell sequencing technologies led to the identification of a rare lung cell type that plays a major role in cystic fibrosis (Plasschaert *et al.*, 2018). Named the pulmonary ionocyte, this cell type expresses high levels of the cystic fibrosis transmembrane conductance regulator (CFTR) gene, which codes for a cell membrane protein that allows the outward flow of chloride ions and which is mutated in patients with cystic fibrosis, causing pathogenic accumulation of mucus in the airways. Currently available drugs for treating cystic fibrosis are able to target specific mutated versions of the CFTR protein, and knowing which cells are contributing to the production of the protein could be conducive to more precise delivery of the treatment.

Several projects within the HCA have led to the discovery of new cell subtypes that play an important role during human development (Vento-Tormo *et al.*, 2018; Popescu *et al.*, 2019; Park *et al.*, 2020). A recent study profiling the uterine-placental interface at single-cell resolution identified three distinct subsets of decidual natural killer cells (dNK) and decidual stromal cells with potential roles in early pregnancy in humans (Vento-Tormo *et al.*, 2018). The novel dNK subsets found in the tissue are thought to be key to maintaining the immunomodulatory environment during early pregnancy, helping in the establishment of placentation through invasion of the placental trophoblast. This study has advanced our understanding of the uterine environment during the early stages of pregnancy and will have a direct impact on pregnancy-related pathologies, such as preeclampsia or fetal growth restriction (Figure 2.2).

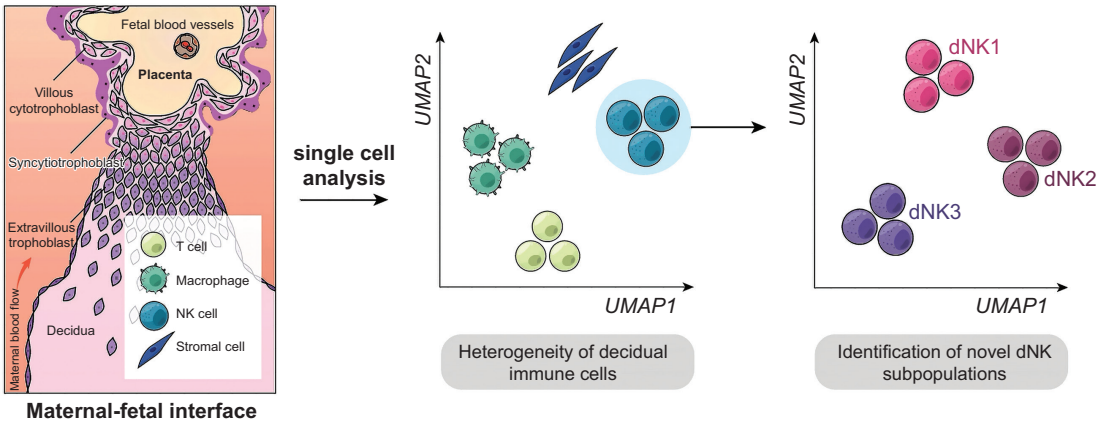

FIGURE 2.2 Single-cell profiling of the maternal–fetal interface during early pregnancy in humans has led to the discovery of three novel subsets of decidual natural killer cells (dNKs). Each subpopulation of dNK was found to have a specific function in the establishment of placentation.

Creation of a Comprehensive Body of Knowledge

Each single-cell study carried out within the framework of the HCA initiative provides invaluable information on how cells are organized into functional organs. However, in order to fully exploit the newly acquired knowledge for further basic research as well as for medical and translational purposes, it is necessary to integrate it into a single, coherent data resource. Moreover, having an atlas means that meta-analyses can be performed across cells, tissues, and organs, thus gaining a comprehensive understanding that goes well beyond what can be gathered from the individual studies.

In more practical terms, the integrated maps of the human body in the HCA allow researchers and practitioners around the world to query organs, tissues, and cells to obtain information about their molecular and organizational features. The analogy used in the White Paper (Regen *et al.*, 2018) to describe the purpose of the HCA is that of Google maps: the atlas allows its users to navigate the human body at various levels of resolution to identify patterns and interactions among its fundamental elements, zooming in and out, depending on the research goals (Figure 2.3).

Building an integrative resource comes with its own challenges. New computational tools are being developed to integrate datasets from distinct batches, including technologies and protocols for processing or sequencing platforms (Lopez *et al.*, 2018; Korsunsky *et al.*, 2019; Polański *et al.*, 2019). In addition, data have to be released on an open-access basis to ensure its accessibility. The Data Coordination Platform (DCP) seeks to coordinate data accessibility through its deposition in a cloud-based platform and integrate it into a user-friendly portal.

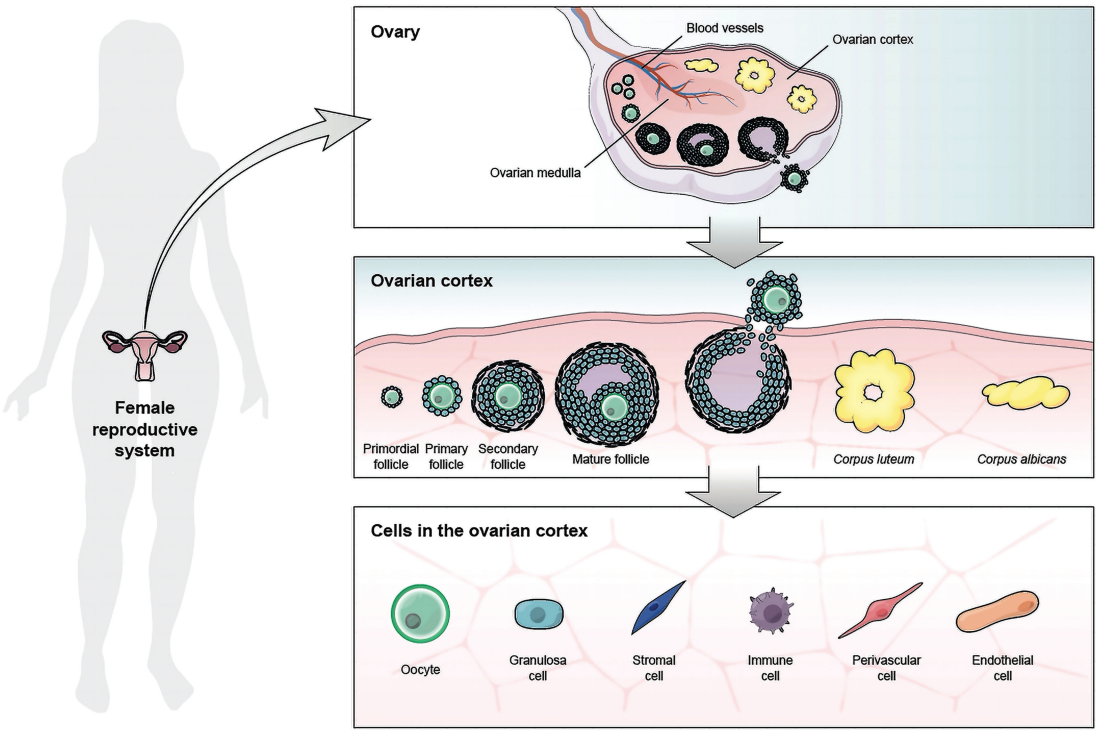

FIGURE 2.3 Example of an integrated map of the human ovary. Such maps can be queried by researchers and practitioners at the organ, tissue, and cell levels to obtain information about the molecular and organizational features of the structure of interest.

SINGLE-CELL RNA SEQUENCING TECHNOLOGIES

Single-cell sequencing is a relatively recent technological development which was named Method of the Year in 2013 by the journal *Nature Methods* (Nawy, 2014). The earliest single-cell sequencing method applicable on a large scale was transcriptome analysis *via* RNA-Seq (single-cell RNA sequencing, scRNA-seq) (Tang *et al.*, 2009). scRNA-seq marked an enormous progression in the definition of cell identities because it offered a comprehensive and unbiased characterization of the cell at the molecular level (Trapnell, 2015).

Currently available scRNA-seq protocols are based on technologies that vary greatly in cost and sensitivity. Individual strengths and weaknesses of scRNA-seq protocols need to be assessed in light of the biological questions of interest (Papalexi and Satija, 2018). Regardless of the underlying technology, they all involve a workflow which can be broken down into three main steps: physical separation of the cells; reverse transcription and PCR amplification of the polyadenylated RNAs; and sequencing of the PCR products. Depending on their approach to cell isolation and transcript amplification, the technologies can be broadly divided into plate-based, droplet-based, and combinatorial indexing protocols. Moreover, scRNA-seq data can now be complemented with spatial data to study the exact location of cells within a tissue.

Plate-Based Protocols

The first protocols for the unbiased quantification of the transcriptome of single cells were developed in a single tube containing lysis buffer (Tang *et al.*, 2009). Shortly after, multiplexing and robotics enabled the processing of hundreds of cells, which meant that cells could be processed in 96-well or 384-well plates (Baugh *et al.*, 2001; Islam *et al.*, 2011; Ramsköld *et al.*, 2012; Jaitin *et al.*, 2014; Zeisel *et al.*, 2015). Whereas the first protocols isolated cells *via* micro-pipetting and placed them into a plate containing lysis buffer, it is now common to isolate them by fluorescent activated cell sorting (FACS). This method allows the user to isolate cells by protein surface markers selected on the basis of prior information about the cell type. Following lysis of the cells, conversion of RNA into cDNA and subsequent cDNA amplification are performed separately on each cell and the PCR products are sequenced *via* next-generation sequencing (NGS) (Figure 2.4, top panel).

One example of plate-based technology that is broadly used is Smart-Seq2 (Picelli *et al.*, 2013, 2014). The advantage of this method is that it enables the quantification of full-length transcriptomes. This information has proven useful for the reconstruction of highly variable T- and B-cell receptors required to control B and T cell clonal expansion in response to antigens (Stubbington *et al.*, 2016; Lindeman *et al.*, 2018). Smart-seq2 has also been used to reconstruct haplotypes of KIR (killer-cell immunoglobulin-like receptor), which are key receptors involved in the activation and inhibitory states of NK cells (Vento-Tormo *et al.*, 2018). Smart-Seq3 is a newly defined protocol that offers even greater sensitivity than its predecessors, enabling isoform identification (Hagemann-Jensen *et al.*, 2020).

Droplet-Based Protocols

Instead of using wells to carry out the reverse transcription and PCR reactions, droplet-based methods, such as Drop-seq (Macosko *et al.*, 2015) and InDrop (Klein *et al.*, 2015), employ microdroplets which encapsulate a single cell and a gel bead covered with barcodes. After reverse transcription and amplification, the mRNAs inside each droplet are pooled together, significantly increasing the multiplexing of the methodology, and sequenced in parallel *via* NGS. This technology has been recently adapted and commercialized by 10x Genomics, increasing the accessibility of the method across the scientific community.

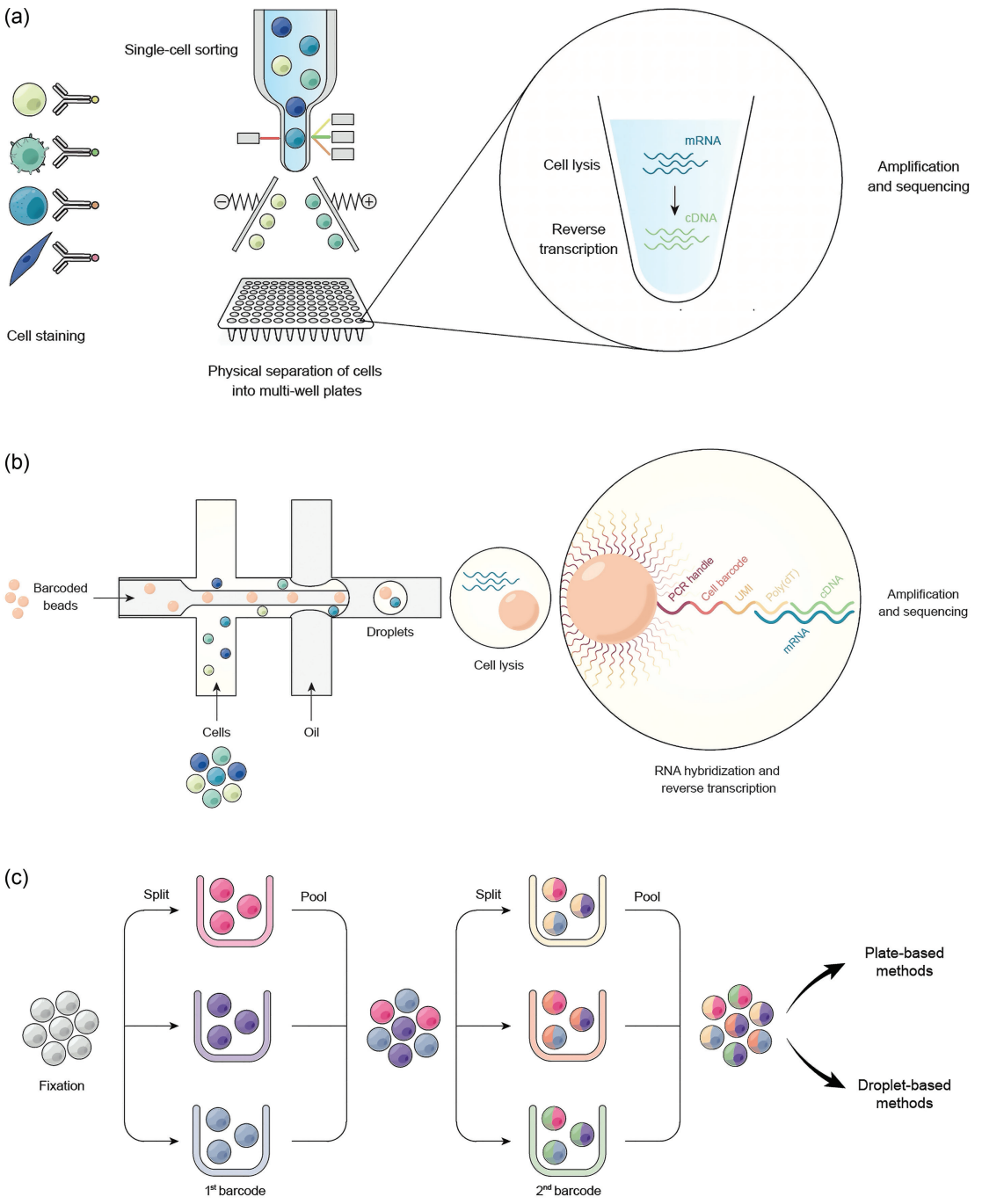

FIGURE 2.4 (a) In plate-based scRNA-seq protocols, cells are physically isolated into a multi-well plate using a cell-sorting technique. Cell lysis, reverse transcription, and amplification are carried out separately within each well. (b) Droplet-based scRNA-seq protocols employ a microfluidics device that enables the formation of microdroplets encapsulating both a single cell and a gel bead covered with barcodes. After reverse transcription and amplification, the mRNAs inside each droplet are pooled together for parallel sequencing. (c) Schematic representation of the principle behind combinatorial indexing protocols. Fixed cells are randomly split into a multi-well plate, where each well contains a unique barcode. After being labeled with the first barcode, the cells are pooled, shuffled, and split again randomly into the same set of wells. This so-called *split-pool* cycle is iterated until the number of combinations of possible barcodes is much higher than the number of cells being profiled.

In these protocols, individual barcoded gel beads and cells are flowed into a microfluidics device (Guo *et al*., 2012) where a mixture of aqueous and oil flows allows for their compartmentalization into a single droplet. Each barcode is composed of three segments: a PCR handle to initiate the PCR reaction; a cell barcode that is unique to each bead; and a unique molecular identifier (UMI) which is different for each cellular mRNA molecule. Cell barcodes and UMIs are generated by randomly assembling nucleotides. UMIs provide a means to remove PCR artifacts during the step of RNA amplification. Once the beads and cells are trapped in the droplets, a lysis buffer breaks the cell membrane and releases the mRNAs, which come into contact with the barcodes and ligate to them. This results in every mRNA molecule present inside the droplet being labeled with the same cell barcode and a different UMI. It is thus possible to subsequently employ massive parallel sequencing without losing the information about the cell of origin of each mRNA (Figure 2.4, central panel).

The main benefit of droplet-based protocols lies in the large number of cells that can be profiled in parallel in a single experiment, the reduction in reagent volumes, and multiplexing, which directly translates into a much lower cost per cell. On the other hand, the sensitivity is often lower than with plate-based methods and only one end of the transcript is sequenced. Recent studies have shown that the definition of cell identity is highly dependent on the number of cells sequenced (Svensson, da Veiga Beltrame, and Pachter, 2019). Therefore, droplet-based methods are the strategies-of-choice for studying heterogeneous cell populations and discovering rare cell types.

Plate-Based Combinatorial Indexing Protocols

Building upon the concept of cellular barcoding, plate-based combinatorial indexing scRNA-seq protocols, such as single-cell combinatorial indexing RNA sequencing (sci-RNA-seq) (Cao *et al*., 2017) and split-pool ligation-based transcriptome sequencing (SPLiT-seq) (Rosenberg *et al*., 2018) allow the profiling of thousands of cells without having to physically isolate each cell. In these methods, the cells are fixed, and their mRNA is manipulated *in situ* by sequential addition of random barcodes in a combinatorial fashion. These strategies overcome the cell isolation inefficiency associated with the droplet-based methods and therefore reduce the pool of barcodes required (Figure 2.4, bottom panel).

In SPLiT-seq, formaldehyde-fixed cells are randomly split into a 96-well plate with unique barcodes for each well. All cells are labeled with the first barcode, and reverse transcription is carried out: at this stage; the chance of two cells having the same barcode is 1/96. The cells are then pooled, shuffled, and split again randomly into the same set of wells. After the second barcode is added, the probability that two cells have the same barcode is 1/9216 (96*96). This procedure can be iterated until the number of combinations of possible barcodes is much higher than the number of cells being profiled, and this is what overcomes the need for physically isolating each cell. The cells are eventually lysed, and their labeled cDNAs are PCR amplified and sequenced.

Droplet-Based Combinatorial Indexing Protocols

The inefficiency of droplet-based methods limits the number of cell barcodes that can be designed, and, consequently, limits their multiplexing capacity. On the other hand, the potential for massive-scale profiling of droplet-based methods still surpasses that of plate-based methods. Newly developed, droplet-based, combinatorial indexing protocols, such as single-cell combinatorial fluidic indexing (scifi-RNA-seq) (Datlinger *et al*., 2019) and droplet single-cell assay for transposase-accessible chromatin using sequencing (dscATAC-seq) (Lareau *et al*., 2019), apply the idea behind plate-based combinatorial indexing protocols to overcome the limitation of the cell isolation inefficiency of droplet-based methods and are thus able to fully exploit their potential for massive-scale profiling.

In scifi-RNA-seq, permeabilized cells are first pre-indexed by reverse transcription in microwell plates, similarly to plate-based combinatorial indexing protocols. Then, the pre-indexed cells are pooled,

randomly mixed, and encapsulated with a microfluidic droplet generator. The encapsulation is carried out such that most droplets are filled with potentially more than one cell, and, inside the droplets, the transcripts are labeled with a second barcode. Because the cells are randomly mixed between the first (plate well) and the second (droplet) round of barcoding, the combination of the two barcodes enables the unique identification of individual cells. Moreover, by allowing each droplet to encapsulate multiple cells, scifi-RNA-seq solves the cell isolation inefficiency inherent to droplet-based scRNA-seq.

The combination of both methods results in an exponential increase in the number of cells that can be sequenced in each reaction and a significant reduction in reagent usage and hence costs. However, the higher number of cells also increases sequencing demands which have not experienced the exponential reduction in costs observed for the reagents. Therefore, it is important to select the method of choice based on the user's budget and the scientific question to address.

Integration of Sequencing and Spatial Data

Because single-cell sequencing methods require cells to be dissociated, information about the spatial location of each cell within the tissue of origin is inevitably lost. Knowledge of how the cells are spatially organized in the tissue, however, is crucial to fully understanding cellular identity. Indeed, both the "absolute" location of a cell within the tissue and its relative position compared to other cells are often linked to the cell's function. Combining such information with single-cell gene expression data, measured by RNA-seq, also allows the inference of complex interactions, such as cell-cell communication. Currently, scRNA-seq data can be integrated with either image-based or sequencing-based spatial data.

A commonly used image-based technique is fluorescence *in situ* hybridization (FISH). Originally designed for detecting chromosomal abnormalities in diseases, it was employed to visualize specific chromosomal locations by using fluorescent oligonucleotide probes complementary to the region of interest (Gall and Pardue, 1969). The application of the original molecular principle to image mRNA in an intact tissue is known as single-molecule FISH (smFISH) and allows for the visualization of individual cells (including their subcellular structure) along with their spatial coordinates within the tissue (Figure 2.5, top panel). Spatial technologies based on smFISH vary greatly in scale, ranging from tens of imaged mRNAs/cell, such as in RNAscope (Wang *et al.*, 2012) and SABER-FISH (Kishi *et al.*, 2019), to hundreds or thousands of imaged mRNAs/cell *via* the use of imageable barcodes, such as in MERFISH (Xia *et al.*, 2019) and seqFISH (Qian *et al.*, 2020). By integrating scRNA-seq and smFISH data, it is possible to transfer cell-type annotations derived from gene expression measurements of the entire transcriptome to spatially resolved cells. Once cell types are mapped onto the tissue, the architecture of the tissue can then be reconstructed in terms of both its overall cellular composition and the spatial relationships among cell types.

More recently, sequencing-based methods, such as the spatial transcriptomic technology (Visium) commercialized by 10x Genomics, allow the whole transcriptome to be mapped directly onto the tissue (Ståhl *et al.*, 2016). In Visium, the tissue slide is first placed onto a chip used to measure gene expression, which is made up of capture areas containing thousands of barcoded spots. Each spot consists of millions of oligonucleotides with unique spatial barcodes. After permeabilization of the tissue, mRNAs can bind to the barcoded oligonucleotides on the underlying spot by diffusion. Reverse transcription is carried out and cDNAs are pooled and sequenced *via* NGS as in massive parallel scRNA-seq protocols (Figure 2.5, bottom panel). Although greatly promising, the Visium technology does not reach single-cell resolution when measuring transcriptome-wide gene expression because the size of each spot (~50 μ) is several times the average size of a cell (approximately 10–50 cells/spot). To overcome this limitation, matched scRNA-seq from the same sample, can be used to deconvolve the Visium data from spatially resolved spots to spatially resolved cells (Kleshchevnikov *et al.*, 2022; Andersson *et al.*, 2019, 2020). More recent, sequencing-based spatial technologies include Slide-seq (Rodriques *et al.*, 2019), which employs barcoded beads spatially indexed by SOLiD sequencing, and high-definition spatial transcriptomics (Vickovic *et al.*, 2019), which uses a high-density bead array and decodes each bead's location *via* rounds of hybridization with complementary, labeled, decoder oligonucleotides.

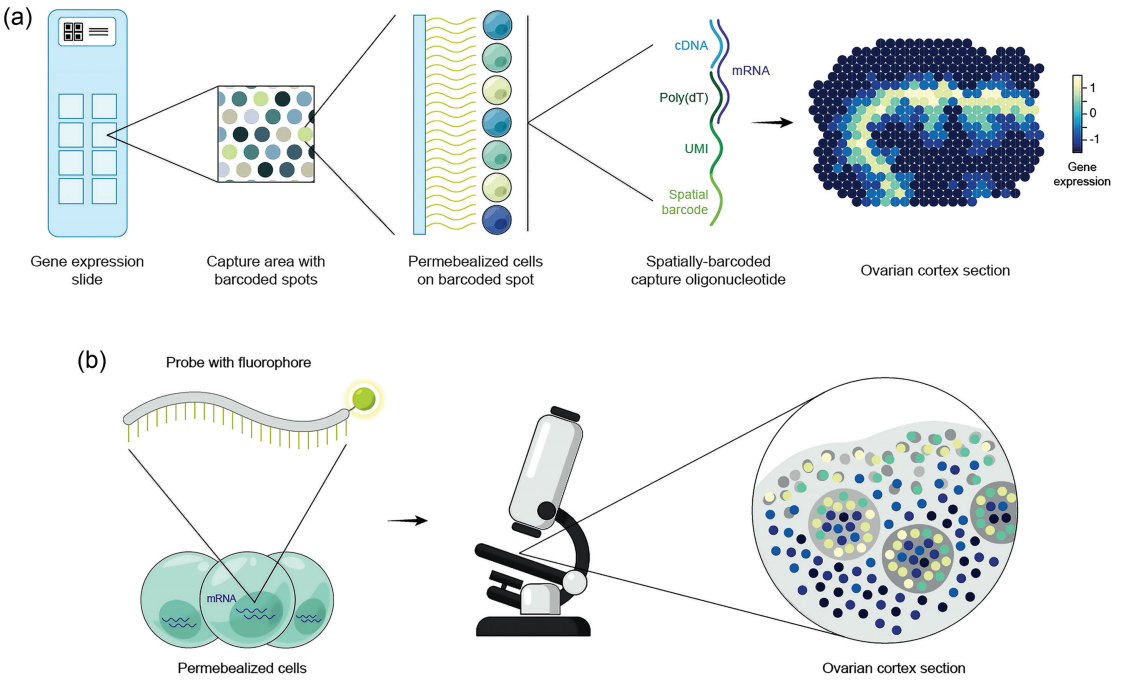

FIGURE 2.5 (a) Schematic representation of the main steps in the workflow for Visium, as an example of sequencing-based spatial technologies. The tissue is first loaded onto the gene expression slide and then permeabilized so that all cellular mRNAs can bind to the barcoded oligonucleotides on the slide. (b) Schematic representation of the main steps in the workflow for imaging-based spatial technologies. Fluorescent oligonucleotides are incubated with the permeabilized tissue, allowing for their hybridization with cellular mRNAs of interest and subsequent imaging under the microscope.

TOWARD SINGLE-CELL REPRODUCTIVE MEDICINE

The potential applications of the HCA to both basic and translational research are limitless. By taking our understanding of human biology to a higher level of resolution, an atlas can help uncover the cellular basis of disease as well as lead to the development of improved *in vitro* models, revolutionizing drug testing. The aforementioned applications are particularly relevant in reproductive medicine, where the physiology of the tissues and many widespread diseases are still poorly understood. In addition, there is a need for better *in vitro* models to accelerate the process of drug development and testing (Figure 2.6).

Single-Cell Atlases of Healthy Tissue

Before we can study the cellular dysregulation brought about by disease, it is imperative to have a comprehensive reference map of the molecular state of the cells in the healthy human tissue. The human reproductive system is responsible for producing gametes and hormones and for accommodating and nurturing the fetus, all of which are highly regulated functions that require timely activation of distinct cellular phenotypes. Single-cell transcriptomics has already helped shed light on the diversity of cells present in the human reproductive system, including primary and secondary reproductive organs across various stages of development, and have hinted at the cellular mechanisms potentially underlying many pathological conditions.

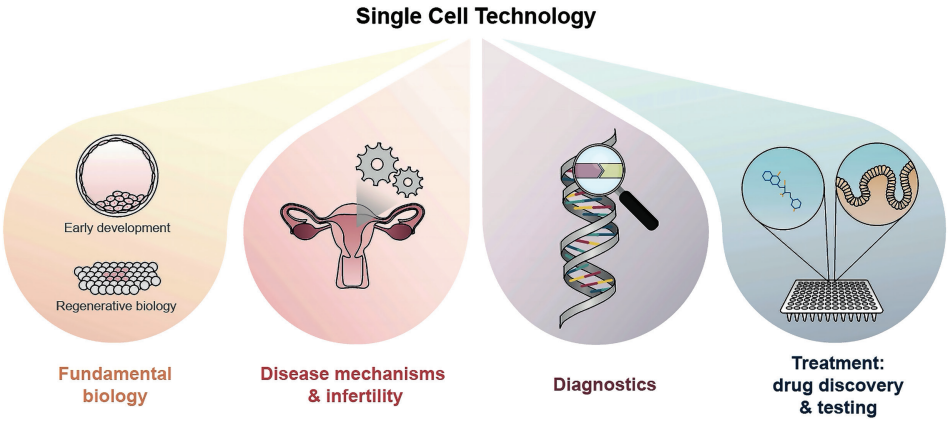

FIGURE 2.6 Areas of impact of single-cell technologies in reproductive medicine. Single-cell technologies have already shed light on the biological mechanisms underlying the functioning of the healthy reproductive system as well as several disease conditions. Moreover, they are instrumental in the development of new tools for testing and treating such conditions.

Single-cell analyses of the neonatal, adolescent, and adult male primary reproductive organs elucidated the transcriptomic and developmental stages of human spermatogenesis and investigated the cellular crosstalk between germline and somatic cells present in the testes (Guo *et al.*, 2018; Sohni *et al.*, 2019; Shami *et al.*, 2020; Xia *et al.*, 2020). Similarly, scRNA-seq profiling of the adult ovaries provided a map of the molecular signatures of the cell types in the inner and outer ovarian cortices at the different stages of follicular development (Fan *et al.*, 2019; Wagner *et al.*, 2020). Altogether, these datasets constitute an invaluable resource with which to dissect the potential mechanisms of male and female infertility as well as other pathologies associated with the reproductive system, such as cancer, and to develop new treatments and assisted reproductive technologies.

Other studies include atlasing of secondary female reproductive organs, such as the endometrium (Suryawanshi *et al.*, 2018; Vento-Tormo *et al.*, 2018; Wang *et al.*, 2018, 2020; Lucas *et al.*, 2020) and the fallopian tube (Dinh *et al.*, 2021; Hu *et al.*, 2020). Single-cell transcriptomic analysis of endometrial biopsies across the menstrual cycle revealed a high degree of heterogeneity in the cellular composition of the tissue, with characteristic signatures for each cell type and phase of endometrial transformation (Suryawanshi *et al.*, 2018; Vento-Tormo *et al.*, 2018; Wang *et al.*, 2018, 2020; Lucas *et al.*, 2020). The fallopian tube also undergoes structural changes in response to the menstrual cycle and is thought to harbor the cell-of-origin for many high-grade serous ovarian cancers (HGSOCs). Droplet-based scRNA-seq of fallopian tubes from healthy individuals revealed the transcriptional programs underlying different epithelial cell populations (Dinh *et al.*, 2020; Hu *et al.*, 2020). Furthermore, computational deconvolution of HGSOCs based on the transcriptional signatures of the epithelial populations present in the healthy tissue revealed that early secretory epithelial cells from the fallopian tubes are likely to be the precursor state for many HGSOCs (Dinh *et al.*, 2020; Hu *et al.*, 2020).

Single-Cell Atlases of Disease

Disease involves the disruption of normal cellular functions and interactions. Initial efforts toward building reference single-cell maps of non-physiological conditions in the field of reproductive medicine, such as ovarian cancer, have already provided invaluable insights. They have helped uncover heterogeneity within these conditions and provided mechanistic insights into their development.

Among the plethora of malignancies affecting the ovary, HGSOC predominates in the clinical setting and is infamous for its high fatality rate and poor prognosis (Lisio *et al.*, 2019). At the time of diagnosis,

one-third of patients with HGSOC present ascites, which act as a reservoir of cell types that provide a tumor-promoting microenvironment for cancer cells (Ahmed and Stenvers, 2013). A recent analysis of ascites samples from HGSOC patients, using droplet-based and plate-based scRNA-seq, resolved the expression profiles of cancer, immune, and stromal cells. It uncovered inter-patient and intra-patient heterogeneity within cancer-associated fibroblasts (CAFs) and macrophages found in ascites and suggested a role for JAK-STAT signaling in both malignant cells and CAFs. It also helped redefine the immunoreactive and mesenchymal subtypes of HGSOCs described by The Cancer Genome Atlas (TCGA), finding that they are derived from macrophages and CAFs, respectively, rather than from malignant cells (Izar *et al*., 2020).

Improving *In Vitro* Models

Mapping the development of malignant as well as non-malignant cells is important to understanding the origin and progression of conditions such as HGSOCs *in vivo* lineage tracing experiments have provided invaluable insights into how cells differentiate in mice although translating these findings to humans is often challenging. A solution to this challenge is offered by *in vitro* models, which allow researchers to study and manipulate developmental processes. Then, scRNA-seq can be used to obtain an unbiased and comprehensive read-out of the changes occurring at the transcriptome level as cells develop or are perturbed in the dish.

Prompted by the *in vivo* findings about the potential role of JAK-STAT signaling in the ascites of HGSOC patients, Izar *et al*. (2020) used primary HGSOC cell lines and patient ascites-derived xenograft models to test the effects of JAK-STAT signaling inhibition. They performed a drug screen with compounds targeting different nodes of the JAK-STAT signaling pathway and identified one compound, JSI-124, as having potent anti-tumor activity. Taken together, the results from the *in vivo* and *in vitro* analyses revealed that inhibition of the JAK-STAT signaling pathway may be a therapeutic option for HGSOC patients (Izar *et al*., 2020).

Other noteworthy studies employed scRNA-seq to disentangle the heterogeneity of endometrial organoids (Fitzgerald *et al*., 2019) and cultures of primary decidualizing endometrial stromal cells (Lucas *et al*., 2020). Future quantitative comparisons of *in vitro* models and their corresponding *in vivo* single-cell references will prove useful in improving the efficiency and accuracy of current *in vitro* models (Boretto *et al*., 2017; Turco *et al*., 2017). In addition, such comparisons will help us develop more accurate *in vitro* models that mirror the fundamental developmental events that give rise to the formation and differentiation of reproductive organs.

CONCLUSION AND FUTURE PERSPECTIVES

While scRNA-seq remains the most widely used single-cell sequencing approach, it is now also possible to achieve single-cell resolution when measuring chromatin accessibility, DNA methylation, cell surface proteins, histone modifications, and chromosomal information (Stuart and Satija, 2019). Moreover, as described in this chapter, image-based and sequencing-based spatial technologies can currently be used to locate cells anatomically within a tissue. Integrating the aforementioned data types or simultaneously measuring multiple modalities is invaluable when defining cell identities. Unsurprisingly, single-cell multimodal omics (i.e., a combination of distinct single-cell genomic methods) was named Method of the Year in 2019 ("Method of the Year 2019: Single-cell multimodal omics," Teichmann, 2020).

Creating single-cell references of the human reproductive system that encompass information from multiple layers of regulation of the genome will prove crucial in further dissecting the cellular mechanisms

underlying both physiological and pathological conditions. Moreover, such multi-omic atlases will serve as a framework for the development of better *in vitro* models. Altogether, single-cell sequencing technologies are taking us towards a comprehensive understanding of the complexity of our body, by—to put it in the words of the HCA—"mapping the human body one cell at a time."

ACKNOWLEDGMENTS

We thank Antonio Garcia, scientific illustrator from Bio-Graphics, for help with the illustrations and Charlotte Wright, Luz Garcia-Alonso, and Regina Hoo for insightful comments. R.V-T is supported by Wellcome Sanger core funding (grant number WT206194), and V.L is supported by the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 874741.

AUTHORS' CONTRIBUTIONS

V.L. and R.V.-T .wrote and edited the manuscript.

BIBLIOGRAPHY

Ahmed, N., *et al.* (2013) *Frontiers in Oncology*, 3, p. 256.
Andersson, A., *et al.* (2019) *bioRxiv*. Available at: https://www.biorxiv.org/content/10.1101/2019.12.13.874495v1.full
Andersson, A. *et al.* (2020) *Communications Biology*, 3(1), p. 565.
Baugh, L. R., *et al.* (2001) *Nucleic Acids Research*, 29(5), p. E29.
Boretto, M., *et al.* (2017) *Development*, 144(10), pp. 1775–1786.
Cao, J., *et al.* (2017) *Science*, 357(6352), pp. 661–667.
Datlinger, P., *et al.* (2019) *bioRxiv*. Available at: https://www.biorxiv.org/content/10.1101/2019.12.17.879304v1.abstract.
Dinh, H. Q., *et al.* (2021) *Cell Reports*, 35(2), p. 108978.
Fan, X., *et al.* (2019) *Nature Communications*, 10(1), p. 3164.
Fitzgerald, H. C., *et al.* (2019) *Proceedings of the National Academy of Sciences of the United States of America*, 116(46), pp. 23132–23142.
Gall, J. G., et al. (1969) *Proceedings of the National Academy of Sciences of the United States of America*, 63(2), pp. 378–383.
Guo, J., *et al.* (2018) *Cell Research*, 28(12), pp. 1141–1157.
Guo, M. T., *et al.* (2012) *Lab on a Chip*, p. 2146.
Hagemann-Jensen, M., *et al.* (2020) *Nature Biotechnology*.
Hu, Z., *et al.* (2020) *Cancer Cell*, 37(2), pp. 226–242.e7.
Islam, S., *et al.* (2011) *Genome Research*, 21(7), pp. 1160–1167.
Izar, B., *et al.* (2020) *Nature Medicine*, 26, pp. 1271–1279.
Jaitin, D. A., *et al.* (2014) *Science*, 343(6172), pp. 776–779.
Kishi, J. Y., *et al.* (2019) *Nature Methods*, 16(6), pp. 533–544.
Klein, A. M., *et al.* (2015) *Cell*, 161(5), pp. 1187–1201.
Kleshchevnikov, V., *et al.* (2022) *Nature Biotechnology*.
Korsunsky, I., *et al.* (2019) *Nature Methods*, 16(12), pp. 1289–1296.
Kulkarni, A., *et al.* (2019) *Current Opinion in Biotechnology*, pp. 129–136.
Lander, E. S. (1996) *Science*, 274(5287), pp. 536–539.

Lareau, C. A., *et al.* (2019) *Nature Biotechnology*, pp. 916–924.
Lindeman, I., *et al.* (2018) *Nature Methods*, 15(8), pp. 563–565.
Lisio, M.-A., *et al.* (2019) *International Journal of Molecular Sciences*, 20(4).
Lopez, R., *et al.* (2018) *Nature Methods*, 15(12), pp. 1053–1058.
Lucas, E. S., *et al.* (2020) *Communications Biology*, 3(1), p. 37.
Macosko, E. Z., *et al.* (2015) *Cell*, 161(5), pp. 1202–1214.
Nawy, T. (2014) *Nature Methods*, 11(1), p. 18.
Ordovas-Montanes, J., *et al.* (2018) *Nature*, 560(7720), pp. 649–654.
Papalexi, E., et al. (2018) *Nature Reviews. Immunology*, 18(1), pp. 35–45.
Park, J.-E., *et al.* (2020) *Science*, 367(6480).
Picelli, S., *et al.* (2013) *Nature Methods*, 10(11), pp. 1096–1098.
Picelli, S., *et al.* (2014) *Nature Protocols*, 9(1), pp. 171–181.
Plasschaert, L. W., *et al.* (2018) *Nature*, 560(7718), pp. 377–381.
Polański, K., *et al.* (2020) *Bioinformatics*, 36(3), pp. 964–965.
Popescu, D.-M., *et al.* (2019) *Nature*, 574(7778), pp. 365–371.
Qian, X., *et al.* (2020) *Nature Methods*, 17(1), pp. 101–106.
Raj, A., et al. (2008) *Cell*, pp. 216–226.
Ramachandran, P., *et al.* (2019) *Nature*, 575(7783), pp. 512–518.
Ramsköld, D., *et al.* (2012) *Nature Biotechnology*, 30(8), pp. 777–782.
Regev, A., *et al.* (2017) *eLife*, 6.
Regev, A., *et al.* (2018) *ArXiv.* Available at https://arxiv.org/abs/1810.05192.
Rodriques, S. G., *et al.* (2019) *Science*, 363(6434), pp. 1463–1467.
Rosenberg, A. B., *et al.* (2018), *Science*, 360(6385), pp. 176–182.
Rozenblatt-Rosen, O., *et al.* (2017) *Nature News*, 550(7677), p. 451.
Shami, A. N., *et al.* (2020) *Developmental Cell*, 54(4), pp. 529–548.e12.
Smillie, C. S., *et al.* (2019) *Cell*, 178(3), pp. 714–730.e22.
Sohni, A., *et al.* (2019) *Cell Reports*, 26(6), pp. 1501–1517.e4.
Ståhl, P. L., *et al.* (2016) *Science*, 353(6294), pp. 78–82.
Stewart, B. J., *et al.* (2019) *Science*, 365(6460), pp. 1461–1466.
Stuart, T., et al. (2019) *Nature Reviews Genetics*, pp. 257–272.
Stubbington, M. J. T., *et al.* (2016) *Nature Methods*, 13(4), pp. 329–332.
Suryawanshi, H., *et al.* (2018) *Science Advances*, 4(10), p. eaau4788.
Svensson, V., *et al.* (2020) *Database: the journal of biological databases and curation*.
Svensson, V., *et al.* (2018) *Nature Protocols*, 13(4), pp. 599–604.
Tang, F. *et al.* (2009) *Nature Methods*, 6(5), pp. 377–382.
Teichmann, S. (2020) *Nature Methods*, 17(1), p. 1.
Trapnell, C. (2015) *Genome Research*, 25(10), pp. 1491–1498.
Turco, M. Y., *et al.* (2017) *Nature Cell Biology*, 19(5), pp. 568–577.
Vento-Tormo, R., *et al.* (2018) *Nature*, 563(7731), pp. 347–353.
Vickovic, S., *et al.* (2019) *Nature Methods*, 16(10), pp. 987–990.
Vieira Braga, F. A., *et al.* (2019) *Nature Medicine*, 25(7), pp. 1153–1163.
Wagner, M., *et al.* (2020) *Nature Communications*, 11(1), p. 1147.
Wang, F., *et al.* (2012), *The Journal of Molecular Diagnostics*, 14(1), pp. 22–29.
Wang, W., *et al.* (2018) *Fertility and Sterility*, p. e2.
Wang, W., *et al.* (2020) *Nature Medicine*, 26(10), pp. 1644–1653.
Wang, Z., *et al.* (2009) *Nature Reviews. Genetics*, 10(1), pp. 57–63.
Xia, B., *et al.* (2020), *Cell*, 180(2), pp. 248–262.e21.
Xia, C., *et al.* (2019) *Proceedings of the National Academy of Sciences of the United States of America*, 116(39), pp. 19490–19499.
Zeisel, A., *et al.* (2015) *Science*, 347(6226), pp. 1138–1142.