ANNUAL REVIEWS

*Annual Review of Genomics and Human Genetics*

# Single-Cell (Multi)omics Technologies

Lia Chappell,[1,*] Andrew J.C. Russell,[1,*] and Thierry Voet[1,2]

[1]Wellcome Sanger Institute, Cambridge CB10 1SA, United Kingdom; email: lc5@sanger.ac.uk, ar19@sanger.ac.uk, tv2@sanger.ac.uk

[2]Department of Human Genetics, KU Leuven, B-3000 Leuven, Belgium; email: thierry.voet@kuleuven.be



Annu. Rev. Genom. Hum. Genet. 2018. 19:15–41

First published as a Review in Advance on May 4, 2018

The *Annual Review of Genomics and Human Genetics* is online at genom.annualreviews.org

https://doi.org/10.1146/annurev-genom-091416-035324



*These authors contributed equally to this article

## Keywords

single cell, omics, multiomics, cellular heterogeneity

## Abstract

Single-cell multiomics technologies typically measure multiple types of molecule from the same individual cell, enabling more profound biological insight than can be inferred by analyzing each molecular layer from separate cells. These single-cell multiomics technologies can reveal cellular heterogeneity at multiple molecular layers within a population of cells and reveal how this variation is coupled or uncoupled between the captured omic layers. The data sets generated by these techniques have the potential to enable a deeper understanding of the key biological processes and mechanisms driving cellular heterogeneity and how they are linked with normal development and aging as well as disease etiology. This review details both established and novel single-cell mono- and multiomics technologies and considers their limitations, applications, and likely future developments.

## INTRODUCTION

A cell's state is defined largely by the interplay between its genome, epigenome, transcriptome, and proteome. Many of these "omic" layers can now be read at single-cell resolution, thanks to efforts that address challenges associated with the isolation of single cells and the analysis of minute quantities of molecules present in each cell. These methods enable the exploration of heterogeneity between cells and within populations previously assumed to be homogeneous: Even two daughter cells that have just been produced from the same precursor cell during mitosis can exhibit differences in their genomes (45), transcriptomes (149), and proteomes (101, 107).

For each class of molecule, there are a range of single-cell methods that can profile a particular omic layer, though these methods differ in their sensitivity and specificity as well as their throughput and ease of use. Researchers should be aware of the limitations of the methods used in their studies and their potential effects on the interpretation of the results. Single-cell technology is developing rapidly, with many of the studies described in this review representing the edge of what is currently possible. Some protocols are well established and widely used [such as several single-cell RNA sequencing (scRNA-seq) protocols], while others are newer and have been used primarily by the original authors. As more researchers develop these tools, it is likely that most existing omic protocols will be adapted into robust single-cell methods.

Building on the development of these omics technologies, multiomics technologies for single cells have emerged that are able to capture multiple omic layers from the same cell. These include technologies for genomics plus transcriptomics, epigenomics plus transcriptomics, and transcriptomics in combination with targeted proteomics. Profiling multiple omic layers makes it possible to capture a more complete set of information about each cell than is possible from any single omic layer alone, which better reflects the complex networks of interactions that are responsible for cellular functions. Single-cell multiomics can therefore be more powerful than a stand-alone omics (or mono-omics) technique performed on a set of single cells as well as different mono-omics techniques performed on different sets of similar single cells. Furthermore, analysis of associations between variation in multiple omic layers within the same single cells allows relationships to be identified unambiguously. For example, a comparison of the transcriptome of one cell with the genomic sequence of another cell can be confounded by somatic genetic variation between the cells as well as variation in cellular states and external environment.

A further advantage of collecting multiple molecular layers from the same cell is that it enables one to study the developmental history of tissues and identify key regulatory mechanisms of phenotypic cell states. The DNA of a cell effectively contains a historical record of mutations acquired with each cell cycle, beginning with the fertilized egg and revealing the cell's lineage in relation to a population of cells it shares ancestry with. Layering such cell lineage trees with the transcriptomes of the same cells can reveal the developmental history of specific cell types and states within healthy and diseased organs, including capture of clonal structures within organs and tissues, the number of stem cells contributing to functional units in organs, and the differentiation trajectories available to a given adult stem cell. Similarly, scRNA-seq data can be used to infer cellular differentiation programs (68, 88, 126, 130, 144) and future transcriptomic states (96) if all cellular stages of differentiation are represented in the data set. The ability of single-cell multiomics technologies to link such information with epigenomic measurements of the same single cells will allow the key regulatory drivers and mechanisms of cellular differentiation to be inferred. In conclusion, the power to link molecular layers using single-cell multiomics technologies will enable the characterization of cells and tissues at an unprecedented molecular and architectural resolution.

Single-cell omics are often defined as technologies that capture an entire molecular layer of a cell. However, additional layers of data associated with a single cell that are traditionally not

described as omics will add information that is critical to obtaining a full understanding of the biology of cellular heterogeneity. For instance, because most human organ functions are executed by the concerted action of individual cells in a spatially organized context, it is crucial to study individual cells in their native dimensional context. This is important not only to understand normal organ development and function but also to investigate how (subpopulations of) cells are perturbed in disease conditions, including cancer and neurological disorders. Therefore, we suggest that spatial information may be regarded as an additional omic layer because it contributes to a complete understanding of cells within tissues.

In this review, we provide an overview of single-cell omics and multiomics methods, including recently developed protocols. We focus primarily on the principles of the protocols and their advantages and limitations and consider applications of multiomics methods to biological questions. Finally, we review potential future developments that will enable both capture of more data layers and larger-scale application of multiomics techniques.



## CELL ISOLATION AND CELL BARCODING: KEY CONCEPTS IN SINGLE-CELL OMICS AND MULTIOMICS TECHNIQUES

A crucial choice for any single-cell experiment is the method for isolating single cells into a format compatible with the downstream single-cell omics analysis. Techniques for cell isolation differ in their throughput, ease of use, and phenotypic information captured during isolation (63); the proportion of cells retained also varies. Low-throughput methods, such as manual micromanipulation (63), laser capture microdissection (36), Raman tweezers (132), and patch clamp (16), are used to capture specific cells and can retain the spatial information; however, these approaches are labor intensive and limited to isolating tens or hundreds of cells per study.

Higher throughput can be achieved with fluorescence-activated cell sorting (FACS) of cells from single-cell suspensions followed by plate-based processing, where thousands of single cells can readily be physically isolated into microliter-scale volumes of an appropriate lysis buffer. Although larger numbers of cells can be processed, spatial information is lost when tissues are dissociated into a cell suspension.

Use of microfluidic devices can increase throughput to tens of thousands of single cells. Key technologies are devices containing microfluidic channels and reaction chambers controlled by pressure-controlled valves, droplets generated in microfluidics chips, and devices that contain nanowells (for a recent review, see 117). The cost per cell is greatly reduced by scaling the reaction volume per cell to nanoliter- or picoliter-scale volumes, which can move the main cost barrier for a given experiment to sequencing rather than library preparation. These approaches typically require preparation of cell suspensions, which (as with plate-based methods) loses spatial information. Phenotypic information (such as the presence of cell surface proteins) gathered during FACS is also lost. However, newer methods (discussed in the section titled Single-Cell Transcriptome-Plus-Protein Measurements) allow phenotypic information to be captured using alternative technologies, which could in some cases replace information acquired during FACS.

Cell barcoding is also a crucial step, as it allows libraries from multiple single cells to be sequenced together in the same pool, with downstream bioinformatic steps able to deconvolute the sequence data into files representing each cell. In many plate-based techniques, the cell barcode is added in the final PCR step before sequencing, with the combined pool of sequencing libraries generated in the final steps. Many of the techniques enabled by microfluidic devices allow cell barcodes to be added near the start of the protocol; this enables the pool of libraries to be made earlier, often allowing the whole pool to be processed within a single tube. Combinatorial indexing methods (17, 26, 118) use a different approach: Groups of fixed and permeabilized cells are FACS

sorted per well of a plate, with each well receiving a different short barcode that can be associated with the target molecule. Cells from all wells are then pooled and re-sorted into small groups per well, where additional barcodes are added to the target molecules. This process can be repeated. With sufficient barcodes and steps, thousands of cells can be barcoded uniquely using standard plates, as there is a low probability of two cells receiving the same set of barcodes. However, it should be noted that many cells are lost in combinatorial indexing. For example, in single-cell combinatorial indexing RNA sequencing (17), ~150,000 *Caenorhabditis elegans* larvae were used to generate a total of six 96-well plates, where each well contained approximately 1,000 *C. elegans* cells; these plates then produced 42,035 *C. elegans* single-cell transcriptomes.

## TARGETS OF MULTIOMICS

To understand how single-cell mono-omics methods can be combined into current and future multiomics technologies, it is important to understand the key protocol steps as well as their individual strengths and limitations. We briefly present widely used and state-of-the-art methods for single-cell mono-omics techniques.

### Genomic DNA

In each cell division, there is a small but nonzero probability that each base in the genome can acquire a mutation (106, 137). This means that genetic heterogeneity is certain given a sufficient number of cell divisions, leading to somatic genetic variation within the tissues of multicellular organisms. This genetic heterogeneity can cause a number of diseases, including cancer (138, 154), neurological disorders (114), and developmental disorders (8).

Recent studies have shown that the scale of genetic heterogeneity is higher than expected from textbook figures in both normal and diseased tissues (29, 33, 80–82, 97, 148), highlighting the need to study somatic genetic variation and determine its causality in relation to phenotypes, aging, and disease. However, there are both technical and financial limitations that restrict the study of DNA in single cells. A diploid human cell contains approximately 7 pg of genomic DNA (gDNA), necessitating amplification to yield sufficient DNA for sequencing. Conventional single-cell genome sequencing methods rely on whole-genome amplification (WGA) for each cell prior to library preparation. A range of WGA methods have been developed using the principles of PCR or multiple displacement amplification, or a combination of both approaches; each strategy has its own advantages and limitations. However, all WGA methods are prone to artifacts that complicate the discovery of genetic variants in the resulting sequence data. These artifacts include locus and allelic dropouts, which lead primarily to false-negative variant calls, as well as unevenness in amplification, generation of chimeric DNA molecules, and introduction of base copy errors, which can lead to false identification of DNA copy number variants, structural variants, and single-nucleotide variants, respectively. Every WGA method suffers from these artifacts, but one method may outperform another when considering a specific set of artifacts; there are therefore WGA methods preferred for the discovery of DNA copy number as opposed to single-nucleotide variants and structural variants. A thorough review by Gawad et al. (44) provided an in-depth discussion of the advantages and limitations of these methods.

New methods for single-cell genome sequencing are emerging that address these challenges. Improvements include optimization of WGA conditions, the use of tricks to preserve Watson–Crick molecular strand identity during amplification of the cell's double-stranded gDNA, and direct preparation of sequencing libraries from single-cell gDNA, which avoids previous WGA intermediates. Specifically, Dong et al. (32) developed a wet-plus-dry laboratory pipeline to minimize

false single-nucleotide-variant detection by modifying the multiple-displacement-amplification conditions, reducing cytosine-deamination artifacts, and computationally modeling allelic amplification biases. For the detection of single-nucleotide variants known from bulk DNA sequencing, the methodology enabled a sensitivity of 90.1% at a cost of 0.12 false positives per million base pairs. However, the method is likely outperformed by other techniques for DNA copy number analysis. Linear amplification via transposon insertion (LIANTI), developed by Chen et al. (18), randomly fragments the gDNA from a single cell by using transposases loaded with a T7 promoter-containing adaptor. The DNA fragments attached to T7 promoters are then subjected to in vitro transcription, leading to linear rather than PCR-based exponential amplification. Following reverse transcription, RNase treatment, and second-strand DNA synthesis (51), the double-stranded LIANTI amplicons are ready for DNA library preparation; this reduces amplification bias and errors associated with nonspecific priming and exponential amplification in conventional WGA methods. A genomic coverage of 97% could be obtained and single-nucleotide variants could be detected with a false-positive rate of $5.4 \times 10^{-6}$, while DNA copy number detection improved to ~10 kb. Furthermore, Zhang and colleagues (22) developed single-stranded sequencing using microfluidic reactors (SISSOR). A microfluidic device is used to separate the Watson and Crick strands of chromosomes present in the lysate of a single cell in order to randomly partition megabase-size DNA strands into multiple nanoliter-scale compartments for amplification and barcoded library preparation. Because the resulting sequencing reads can then be traced to the Watson and Crick strands of DNA molecules, this method enables curation of the cell's genomic sequences; real mutations must be reported by sequencing reads from both Watson and Crick strands deriving from the same original DNA molecule. The method enabled sequencing of single-cell genomes with error rates as low as $10^{-8}$ but at reduced genomic coverage ($63.8\% \pm 9.8\%$), most likely owing to the loss of some DNA fragments during strand separation and partitioning within the device, and the performance for DNA copy number detection remains to be established. Bakker et al. (7) and Zahn et al. (156) developed approaches to avoid up-front WGA by using fragmentation, adaptor tagging, and subsequent PCR to make a sequencing library directly from the DNA present in single-cell lysates. Despite potential loss of genetic material, these methods are well suited for gDNA copy number profiling using low-coverage sequencing, as they can provide better uniformity of coverage, but they are not yet suitable for full genome interrogation.

## Methylomes

5-Methylcytosine (5mC) at CpG dinucleotides is an epigenetic mark that has historically been associated with the repression of transcription, but this mark can have varying context-specific effects on transcription (71). Multiple methods are used to capture this mark, and most rely on the use of bisulfite, which converts cytosine to uracil while leaving 5mC residues unchanged, allowing the discrimination of methylated from unmethylated cytosines in the resulting sequence data. However, single-cell methylome sequencing remains technically challenging, as not only is DNA starting material limited, but the bisulfite treatment also causes extensive DNA damage (60, 84). This damage can lead to higher rates of locus and allelic dropouts compared with single-cell genome sequencing approaches. In addition, two types of error can occur during bisulfite treatment: over-conversion (leading to false-negative CpGs) and under-conversion (leading to false-positive CpGs), which can be measured using methylated and unmethylated DNA spike-in controls, respectively. Individual cells that fail to pass over- or under-conversion rate thresholds set by the authors are excluded from further analysis; the choice of exact thresholds varies among studies, but a typical under-conversion threshold is between 1% and 5%.

Single-cell reduced-representation bisulfite sequencing (scRRBS) (53–55) covers an average of ~1 million CpGs for a single diploid mouse cell (out of ~21.9 million CpGs in the mouse genome) and can cover a maximum of ~1.5 million CpGs overall when data for single cells are pooled. Samples with an under-conversion rate of less than 2% were analyzed further in these studies. After single-cell isolation, DNA is digested using the Msp1 restriction enzyme, which recognizes C^CGG but is methylation insensitive. RRBS relies on the fact that CpGs are not evenly distributed throughout the genome and tend to cluster in CpG islands close to gene promoter regions (30). The use of Msp1 digestion enables the recovery of a large number of CpG islands, which make up most of the library input material. After end repair and dA tailing, premethylated indexed sequencing adaptors are ligated to the Msp1-digested DNA fragments prior to bisulfite conversion; PCR is then used to amplify the product before sequencing. The use of biotinylated 5′ and 3′ adaptors and rescue steps with this protocol has recently been proposed to improve sample yield (153).

Bisulfite-induced DNA damage after adaptor ligation means that any damage to the insert DNA will interfere with PCR amplification, leading to sample loss. Single-cell bisulfite sequencing (scBS-seq) (24, 131) limits this loss by using a modification of the postbisulfite adaptor tagging protocol (99). Here, DNA is bisulfite treated after cell lysis, simultaneously fragmenting the DNA and converting unmethylated cytosines. A first set of biotin-tagged random primers containing the first sequencing adaptor is then used to produce complementary strands to the bisulfite-converted DNA. Repeating this step five times increases the sensitivity because it maximizes the number of tagged DNA strands and generates multiple copies of each fragment. Biotin-tagged DNA is then purified, which is followed by another random priming with the reverse primer containing the other sequencing adaptor. The resulting fragments are amplified by PCR prior to sequencing. This technique can provide methylation status information for a maximum of 10.1 million CpG sites in the mouse genome when data from single cells are pooled, profiling an average of 3.7 million CpGs for a standard diploid mouse cell. These studies used an under-conversion threshold of less than 5%, with a typical under-conversion rate of less than 2%.

A postbisulfite adaptor tagging approach was also used in single-cell whole-genome bisulfite sequencing (scWGBS) (39), although multiple rounds of first-strand synthesis complementary to the bisulfite-treated DNA were omitted. This technique allows a mean coverage of 1 million CpGs and a maximum of 2.7 million CpGs (39) or, with higher sequencing depth, 2.2 million CpGs and a maximum of 8.5 million CpGs (50) in diploid mouse cells. An over-conversion rate of less than 2% and an under-conversion rate of less than 1% were reported (39). Here, cells are sorted into 96-well plates for bisulfite conversion, which is followed by random-hexamer-primed DNA synthesis, terminal tagging, and enrichment PCR. This approach has also been extended to single nuclei for systems where single-cell isolation is difficult, such as neurons (91). scWGBS does not have a preamplification step (in contrast to scBS-seq), which reduces amplification biases, allowing more accurate identification of PCR duplicates and reducing reagent costs. In comparison with scWGBS, scBS-seq has a higher library complexity (39) and therefore is better suited to deep sequencing at high coverage; scWGBS is better suited to low-coverage methylation profiling of many cells. Unlike scRRBS and scWGBS, scBS-seq does not give stranded information.

A final method for profiling 5mC in single cells is to use methylation-sensitive restriction enzymes, which avoids bisulfite-mediated DNA degradation. The first use of this approach in single cells applied the *BstU*I restriction enzyme, which recognizes CG^CG sites and does not digest methylated DNA (5mC at position 1 or 3 on one or both strands) (89). Subsequent PCR with three primers produces two products for undigested DNA (i.e., the restriction recognition sequence is methylated on one or more CpGs) and one product for digested DNA (i.e., the restriction recognition sequence is unmethylated), as one of the primer pairs spans the restriction site; this was used to analyze six loci. Another technique, single-cell restriction analysis of methylation (SCRAM)

(21), increased the throughput of this technique to analyze up to 24 loci in 48 single cells; further studies could increase this scale. This method also uses *Hpa*II, which recognizes 5′-CC^GG-3′ and is blocked by 5mC at position 1 or 2 on one or both strands; in mammals, however, the second cytosine is more likely to be methylated. Neither of these approaches is genome-wide, and both have only single-base-pair resolution at the cut site.

DNA methylation is complex and regulated: Tet (ten eleven translocation) proteins actively demethylate 5mC and progressively oxidize 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) (66). Although these modifications may seem to be derivatives of 5mC, they may have functional roles (40) and can be used to track the dynamics of DNA methylation. Additionally, there is evidence that 5fC and 5caC can affect transcription by affecting the rate and specificity of RNA polymerase II (74). The bisulfite- and restriction-enzyme-based methods do not discriminate between 5mC and 5hmC, and 5fC and 5caC are converted to uracils during bisulfite treatment (65, 69, 155); novel approaches are therefore needed to probe the role of these modifications.

Methylase-assisted bisulfite sequencing (MAB-seq) allows detection of 5fC and 5caC indirectly using a treatment with CpG methylase M.SssI, which methylates all unmethylated CpG sites, protecting cytosines but not 5fCs and 5caC during bisulfite treatment (105, 151). This method was adapted for single cells (152), capturing cell-specific 5fC/5caC profiles of mice embryonic stem cells and blastomeres, and provided evidence for 5fC/5caC dilution at DNA replication, producing two daughter cells with complementary 5fC/5caC patterns. Single-cell MAB-seq can also be used to map sister chromatid exchange. Furthermore, 5fC can be detected directly with single-cell chemical-labeling-enabled C-to-T conversion sequencing (scCLEVER-seq) (158). This technique uses malononitrile, which specifically reacts with 5fC and therefore directly chemically labels this modification. When the PCR product is amplified, the adduct prevents normal pairing with guanine, meaning labeled 5fC sites are read as thymine.

Mooijman et al. (100) profiled the remaining modification, 5hmC, through glycosylation of 5hmC positions, which generates recognition sites for the restriction endonuclease AbaS1. This study also used 5hmC to trace the lineages of single cells.

## Histone Modifications

Modification of the N/C-terminal tails of histone proteins can act to modify the probability of gene expression occurring near a DNA locus. Capturing information about histone modifications at the single-cell level would be informative for studying the epigenetic programs of cells (such as cellular differentiation trajectories) and predicting possible transcriptional states.

Modifications to histone proteins can be surveyed at single-cell resolution using droplet-based chromatin immunoprecipitation (Drop-ChIP) (120). Bulk ChIP followed by sequencing (ChIP-seq) uses antibodies specific to a histone modification to pull down chromatin associated with that mark. Low-input ChIP-seq is challenging, as small amounts of on-target epitope binding combined with nonspecific antibody binding results in a low signal-to-noise ratio. Drop-ChIP uses microfluidic devices to coencapsulate cells in a droplet with a lysis detergent and micrococcal nuclease (MNase), generating mono-, di-, or trinucleosomes, and these nucleosome droplets are then merged one by one with a droplet containing a cell-specific barcode and an aliquot of DNA ligase solution. This generates barcoded chromatin fragments with a PCR handle. ChIP-seq can then be performed on these pooled fragments in the presence of carrier chromatin. This method was used to identify histone H3 dimethylation at lysine 4 (H3K4me2) and histone H3 trimethylation at lysine 4 (H3K4me3) marks in mixed populations of mouse cells. These profiles have a sparse amount of signal, with an average of ~1,000 unique reads per cell, leading to ~800

peaks per cell. Although this leads to low de novo peak identification sensitivity, the accuracy of these data is high, with ~50% of the reads aligning to known positive peaks from bulk ChIP-seq, which has proved sufficient to cluster cells into cell type.

## Open Chromatin

Profiling open chromatin or DNA accessibility allows researchers to infer usage of gene regulatory elements such as promoters and enhancers and can be used to infer operative transcription factors via footprinting analysis (145). For bulk samples, several approaches have been developed to identify these accessible DNA sequences (98). Most approaches [MNase sequencing (MNase-seq) (123), formaldehyde-assisted isolation of regulatory elements followed by sequencing (FAIRE-seq) (49), and DNase sequencing (DNase-seq) (11)] fragment the accessible DNA, generating small fragments (<200 base pairs) that can be processed into libraries. Single-cell DNase-seq of DNase I hypersensitive sites has been achieved by including circular carrier DNA along with the digested material from the single cell during purification and library preparation to minimize loss of the small quantity of DNase I hypersensitive site DNA (25, 70).

The assay for transposase-accessible chromatin using sequencing (ATAC-seq) uses transposases that can simultaneously fragment and ligate (or tagment) adaptors to DNA in open chromatin; less accessible chromatin should not be fragmented. Subsequent PCR amplification is performed with primers complementary to these adaptors. The DNA fragments in these libraries are enriched for open chromatin and require a relatively shallow sequencing depth (13). ATAC-seq has been adapted for single cells using a programmable microfluidic platform (14), which can process up to 96 samples. Single-nucleus ATAC-seq has also been achieved using combinatorial indexing (26), but this requires barcode-carrying transposases, which are not yet commercially available. These different barcode-carrying transposases are used to tagment small populations of nuclei in plate format. Following transposition, all nuclei are pooled, redistributed, and then lysed. Subsequently, a second barcode is added with PCR using indexed primers complementary to the transposase-introduced adaptors. Finally, all PCR products are pooled and sequenced. Reads containing the same combination of barcodes are highly likely to be derived from the same cell.

## Nuclear Structure

Transcriptomic output is also regulated by the higher-order chromatin structure of the DNA within a cell (85). Single-cell high-throughput chromosome conformation capture (scHi-C) can be used to determine the chromosomal architecture within single cells, revealing thousands of contacts (102, 104). This technique was also used to associate dynamics in topologically associated domains with cell cycle progression (103). Two studies recently improved the sensitivity of the scHi-C technique (41, 103). A related approach, single-cell combinatorial indexed Hi-C (sciHi-C) (118), uses combinatorial barcoding to increase the throughput of the technique without the need to physically isolate thousands of nuclei.

A key step of scHi-C is cross-linking of a bulk cell sample, which is followed by cell permeabilization and the creation of a nuclear suspension. A restriction digest is used to fragment the gDNA; overhanging ends are filled in, using biotin-labeled dATP in the mix of nucleotides. Nearby fragments are then ligated, capturing spatial nuclear organization. Single nuclei are subsequently isolated, and the cross-linking is reversed. Through the use of streptavidin-coated magnetic beads, the informative biotin-containing fragments are captured and converted into a library for sequencing. Downstream analysis identifies which distant sequences are present in the same fragment, enabling deconvolution of spatial genomic relationships.

DNA adenine methyltransferase identification (DamID) allows the identification of DNA–protein interactions in single cells (78). This technique was used to study lamina-associated domains, i.e., regions of interphase chromosomes that interact with the nuclear lamina. The *Escherichia coli* enzyme Dam is fused to a protein of interest using genetic techniques. The fusion protein is usually inducible, and upon induction, Dam methylates the adenines in GATC motifs in close proximity to it—i.e., DNA that is likely to be interacting with the protein of interest. Cells are FACS sorted into a lysis buffer in plates and digested with the restriction enzyme DpnI, which is specific for Dam-methylated GATC sequences. Adaptor ligation, PCR, and library preparation then follow before sequencing. These steps are all carried out in one tube (as much as possible) to minimize DNA loss.

## RNA

The transcriptome is dynamic and tightly linked to cell identity and function, enabling definition of the cell type and cell state. Examining the transcriptome at the single-cell level allows the interrogation of inherent features of transcription, such as transcriptional noise and regulated heterogeneity. Single-cell resolution is essential in elucidating coexpression and mutually exclusive expression of genes within cells. Transcripts are thought to be expressed over several orders of magnitude, with many transcripts having low-level expression [5–20 transcript copies per cell (139)]. A typical human cell contains less than 1 pg of mRNA (72).

Protocols for measuring eukaryotic polyadenylated mRNA are the most developed of the scRNA-seq techniques, although they vary in their sensitivity and accuracy (140, 160). Read counts from synthetic spike-in RNAs, such as ERCCs (named for the External RNA Controls Consortium, which developed them), can be used to compare these protocols; the performance of these assays is strongly dependent on sequencing depth, with the maximum observed sensitivity saturating at approximately 4.5 million reads per cell (140). Eukaryotic mRNA can be selectively amplified, enabling enrichment compared with the highly abundant and relatively uninformative rRNA. This selective reverse transcription can be achieved by using an oligo(dT) primer, which can also add a PCR handle to the 5′ end of the first-strand complementary DNA (cDNA). Many protocols make use of the template-switching activity (159) of Moloney murine leukemia virus (MMLV) reverse transcriptases, which allows the addition of a second PCR handle at the 3′ end of the first-strand cDNA; this, in turn, enables exponential amplification with PCR. Other protocols make use of in vitro transcription from a T7 promoter incorporated into one end of the cDNA, which enables linear amplification (34). Other parameters that vary between protocols include whether the sequence data generated encode the full-length mRNA or a 5′ or 3′ fragment and whether strand information is retained.

Scale is a key variable in scRNA-seq experiments, as the number of cells processed within a single sample has increased exponentially in recent years (121, 141). Plate-based methods are able to process hundreds of cells per experiment, with the potential to process thousands of cells across multiple runs. Widely used protocols include Smart-seq and Smart-seq2 (112, 113), which use template switching in combination with PCR, while cell expression by linear amplification and sequencing (CEL-seq) (59), CEL-seq2 (58), and massively parallel RNA single-cell sequencing (MARS-seq) (67) use in vitro transcription, although there is an even wider range available (111). Optimization of plate-based techniques has been achieved by multiple groups; a recently optimized protocol is single-cell RNA barcoding and sequencing (SCRB-seq) (133), which was improved in molecular crowding single-cell RNA barcoding and sequencing (mcSCRB-seq) (6) by the addition of polyethylene glycol (a crowding agent) and substitution of the PCR polymerase.

Microfluidics-based approaches have enabled thousands of cells to be processed in parallel. Droplet-based platforms such as Drop-seq (95) and InDrop (79, 161) facilitate large-scale cell numbers at low cost; droplet scRNA-seq has also been commercialized at a similar scale (157). A simple and low-cost nanowell-based approach, Seq-well (47), can process up to 10,000 cells per array. Combinatorial indexing (17) can also process thousands of cells in parallel but requires cells to be fixed.

Other studies focus on different aspects of the transcriptome. Single-cell integrated nuclear RNA and cytoplasmic RNA sequencing (SINC-seq) allows differential capture and quantification of mRNA from the nucleus and cytoplasm from the same single cell by physically separating the nucleus and cytoplasm, allowing insights into posttranscriptional dynamics and regulation (1). Noncoding RNAs are underexplored at the single-cell level, although some long noncoding RNAs can be captured via their poly(A) tail (77). Small RNAs have been captured on a single-cell scale (38), but this approach has not yet been widely adopted.

## Proteins

Proteins directly mediate cellular function, and therefore it is crucial to understand their heterogeneity between cells. Analysis of single-cell proteomes is limited by the small amount of starting material as well as the inability to directly amplify proteins.

A simple single-cell proteomic readout uses antibodies conjugated to fluorophores that target proteins of interest and subsequent flow cytometry to measure the levels of fluorescent signal per cell as a proxy for protein abundance per cell. This enables analysis of thousands to millions of cells in a single run and can be multiplexed to tens of proteins by using different dyes for different antibodies; however, spectral overlap limits multiplexing (109). Replacing fluorophores with different stable metal isotope tags enables more than 40 proteins to be multiplexed when analyzed by mass cytometry but is limited by the number of available stable metal isotopes (9, 134).

An alternative method uses antibodies conjugated with DNA barcode sequences, enabling the conversion of the protein signal into an amplifiable DNA sequence. After washes, the bound antibody–oligonucleotide can be amplified, increasing its signal. Subsequently, protein levels can be determined by quantitative PCR (qPCR) or sequencing of the DNA products. The capacity for multiplexing is much greater—$4^n$, where $n$ represents the length of the antibody-specific DNA sequence. However, covering the full proteome [which is estimated to comprise more than 20,000 proteins and many more proteoforms for humans (115)] is unlikely, mainly because of the lack of high-affinity antibodies.

Ullal et al. (146) first introduced the antibody barcoding with photocleavable DNA (ABCD) technique. They fixed and permeabilized human cancer cells so that both intracellular and extracellular proteins could be profiled using a panel of ~90 barcoded antibodies. Following antibody staining and washing, UV light releases the DNA barcode from the antibody, and barcodes are counted using fluorescent hybridization technology, enabling multiplexed quantitation on the NanoString fluorescent DNA barcoding platform. A limitation of this technique is the lack of possible sample/cell multiplexing, which limits throughput.

Abseq (129) overcomes this multiplexing limit with a microfluidic device. Cells are stained with antibodies containing a DNA barcode comprising (*a*) a sequence that is antibody specific, allowing the target protein to be identified, and (*b*) a unique molecular index (UMI) sequence, which allows PCR duplicates to be identified following sequencing, thus improving the accuracy of quantification. A microfluidic device then isolates antibody-bound single cells into droplets that contain the proteinase K lysing agent, releasing the barcode within the droplet. Following

proteinase K inactivation, these droplets are merged with droplets containing oligonucleotides with cell-specific barcodes. Strand overlap extension PCR (SOE-PCR) is used to associate the cell-specific barcode with the antibody barcode within the same droplet. This technique can profile more than 10,000 cells in parallel in an hour. The authors noted that the encapsulation method is similar to InDrop and Drop-seq and therefore may be compatible with multiomic RNA and DNA capture. However, this technique is currently limited to cell surface proteins.

Single-cell proteomics by mass spectrometry (SCoPE-MS) increases the depth of single-cell proteomics compared with antibody-mediated methods, with 583 proteins quantified at the single-cell level (12). SCoPE utilizes tandem mass tag (TMT) labeling to enable TMT-labeled peptides from single cells to be pooled with 100–200 carrier cells in the mass spectrometry run. TMT-labeled peptides can be quantified while using the total pooled peptides to identify its sequence. Improvements in the sensitivity and speed of mass spectrometry will further enhance the method, which may be extended to include protein modifications.

The proteome extends outside the confines of the cellular membrane: Proteins can be secreted and/or deposited on the surface of a cell. Techniques have emerged to trap cells in nanoliter-scale chambers; the introduction and sealing of a miniaturized antibody microarray onto the top of the well allow secreted proteins to bind to their cognate antibody. Secreted protein levels are quantified by a multicolor immunoassay. Use of spatial and spectral positioning of dots allowed this method to analyze a secreted proteome of 42 proteins (90).

## Spatial Transcriptomics

A cell rarely exists in isolation, and environmental signals will be integrated to some extent by every cell. This environment consists of a complex set of factors, including interactions with other cells and signaling molecules. Identifying a cell's context, especially within a primary tissue or organoid model, is therefore essential to understand how a single cell's phenotype relates to its three-dimensional spatial coordinates.

An indirect approach to identify localization is to analyze cells with scRNA-seq and map these data onto a reference map that was generated using RNA in situ hybridization approaches (2, 108, 122, 124). A direct approach is spatial transcriptomics (135), which uses barcoded oligo(dT) primers immobilized and arrayed onto glass slides. The spotted oligo(dT) primers include a cleavage site, an amplification handle, a spatial barcode (ID), a UMI, and an oligo(dT) sequence. Tissue sections are placed onto the slide, fixed, and then permeabilized, which allows free polyadenylated transcripts to anneal to the barcoded oligo(dT) array. Next, cDNA is generated and can be cleaved from the slide and converted into an RNA-seq library. The ID index in the sequence data allows sequence reads to be mapped to spatial coordinates on the array and therefore the tissue; the inclusion of UMIs reduces noise. A disadvantage of the current system is that the oligo(dT) spots on the array have a diameter of 100 μm and thus often capture transcripts from multiple cells rather than a single cell.

Alternative in situ sequencing approaches are being developed that involve directly preparing massively parallel sequencing libraries in fixed cells or tissue sections (73, 83). Subsequently, these libraries (primarily targeted transcripts) are sequenced in situ, providing sequencing reads from the locations of the nucleic acid molecule within the cell and thus the tissue, which results in spatial maps of the acquired sequences. Issues with overcrowding and low sensitivity need to be overcome to enable transcriptome-wide capture of spatial information.

Alternatively, use of sequential fluorescence in situ hybridization (seqFISH) approaches (19, 127, 128) enables highly accurate quantitative detection of transcripts from hundreds to thousands of genes, which are linked with their cellular locations. seqFISH approaches are typically more

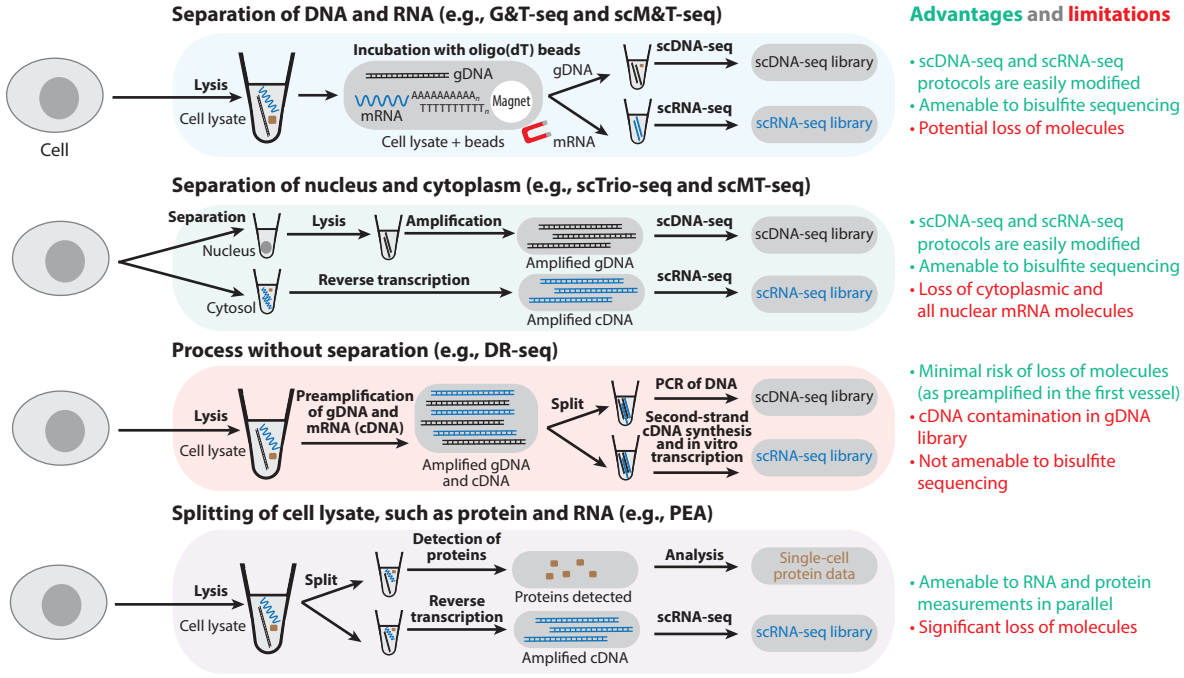


**Figure 1**

Overview of basic single-cell multiomics strategies. Bold text indicates a process; standard text indicates the relevant biological molecules. Abbreviations: cDNA, complementary DNA; DR-seq, DNA–RNA sequencing; G&T-seq, genome and transcriptome sequencing; gDNA, genomic DNA; PEA, proximity extension assay; scDNA-seq, single-cell DNA sequencing; scMT-seq, single-cell methylome and transcriptome sequencing (using a strategy similar to scTrio-seq); scM&T-seq, single-cell methylome and transcriptome sequencing (using a protocol based on G&T-seq); scRNA-seq, single-cell RNA sequencing; scTrio-seq, single-cell triple-omics sequencing.

sensitive than scRNA-seq: scRNA-seq detects 20–40% of mRNA molecules present in the cell, while seqFISH approaches detect 85% of the targeted transcripts in situ.

Spatial information has also been combined with proteomics analysis using mass cytometry. Tissue is stained for mass cytometry, and spatial information is obtained by using a scanning laser that ablates the tissue and sends it in real time through high-temperature plasma and then onto a mass spectrometer. This method was used to identify both proteins (4) and protein modifications (48).

## SINGLE-CELL MULTIOMICS TECHNOLOGIES

A range of strategies following cell lysis enable single-cell multiomics (**Figure 1**, **Table 1**): The cell lysate may be split and subjected to different omics analyses, or the "omes" may be either separated prior to downstream analysis or processed and tagged without a priori separation (10).

### Single-Cell Genome-Plus-Transcriptome Sequencing

Methods that profile DNA and RNA from the same single cell can link acquired genetic variation to transcriptional variation unambiguously. Furthermore, the genomic information may be used to construct cell lineage trees, which is important in understanding normal development (150) as well as disease etiology and progression (28).

**Table 1 Single-cell multiomics technologies**

| Technology | Genomic layer | Chromatin accessibility epigenetic layer | CpG methylation epigenetic layer | Transcriptomic layer | Protein layer | Strategy | Cell isolation[a] | Cell throughput | Automation | Constituent methods |
|---|---|---|---|---|---|---|---|---|---|---|
| DR-seq (31) | Genome | — | — | Transcriptome | — | Preamplification and tagging of DNA and RNA followed by splitting | Mouth pipette | Low | No | Modified CEL-seq (59) and modified MALBAC (161) |
| G&T-seq (92, 94) | Genome | — | — | Transcriptome | — | Separation (DNA and polyadenylated mRNA) | FACS | Medium | Yes | Modified Smart-seq2 (112, 113) and PicoPLEX WGA or MDA WGA (92, 94) |
| scTrio-seq (62) | CNVs (from scRRBS data) | — | Reduced-representation DNA CpG methylation | Transcriptome | — | Separation (nucleus and cytoplasm) followed by bisulfite conversion | Mouth pipette | Low | No | scRNA-seq method of Tang et al. (143) and scRRBS (54) |
| scMT-seq (64) | SNPs (from scRRBS and scRNA-seq data) | — | Reduced-representation DNA CpG methylation | Transcriptome | — | Separation (nucleus and cytoplasm) | Microcapillary pipette | Low | Partial | Modified Smart-seq2 protocol (112, 113) and modified scRRBS (54) |
| scGEM (20) | Targeted genotyping (Sanger and next-generation sequencing) | — | Targeted DNA CpG methylation (qPCR) | Targeted (RT-qPCR) | — | Restriction digestion, preamplification, and splitting | Fluidigm C1 | Medium | Yes | Modified SCRAM (21) |
| scM&T-seq (5) | — | — | DNA CpG methylation | Transcriptome | — | Separation (DNA and polyadenylated mRNA) followed by bisulfite conversion | FACS | Medium | Yes | Modified G&T-seq (92, 94) and modified scBS-seq (24, 131) |

(*Continued*)

**Table 1** (*Continued*)

| Technology | Genomic layer | Chromatin accessibility epigenetic layer | CpG methylation epigenetic layer | Transcriptomic layer | Protein layer | Strategy | Cell isolation[a] | Cell throughput | Automation | Constituent methods |
|---|---|---|---|---|---|---|---|---|---|---|
| scNOMe-seq (116) | — | Chromatin accessibility | DNA CpG methylation | — | — | Bisulfite conversion | FACS | Medium | No | — |
| scCOOL-seq (52) | CNV and ploidy | Chromatin accessibility | DNA CpG methylation | — | — | Bisulfite conversion and prediction of CNVs | Micromanipulator | Low | No | scBS-seq (24, 131) |
| scNMT-seq (23) | — | Chromatin accessibility | DNA CpG methylation | Transcriptome | — | Bisulfite conversion followed by separation | FACS | Medium | Partial | scM&T-seq (5) and scNOMe-seq (116) |
| PEA/STA (46) | — | — | — | Targeted | Tens of proteins | Combination | Fluidigm C1 | Medium | Yes | — |
| PLAYR (42) | — | — | — | Transcriptome | Tens of proteins | Combination | Mass cytometry | High | No | — |
| CITE-seq (136) | — | — | — | Transcriptome | Tens of proteins | Combination followed by separation of libraries | Drop-seq and 10x Genomics Chromium | High | No | Drop-seq (95) and 10x Genomics Chromium (157) |
| REAP-seq (110) | — | — | — | Transcriptome | Tens of proteins | Combination followed by separation of libraries | 10x Genomics Chromium | High | No | 10x Genomics Chromium (157) |

Abbreviations: CEL-seq, cell expression by linear amplification and sequencing; CITE-seq, cellular indexing of transcriptomes and epitopes by sequencing; CNV, copy number variation; DR-seq, DNA–RNA sequencing; Drop-seq, droplet-based sequencing; FACS, fluorescence-activated cell sorting; G&T-seq, genome and transcriptome sequencing; MALBAC, multiple annealing and looping–based amplification cycles; MDA, multiple displacement amplification; PEA, proximity extension assay; PLAYR, proximity ligation assay for RNA; qPCR, quantitative PCR; REAP-seq, RNA expression and protein sequencing assay; RT-qPCR, reverse transcription quantitative PCR; scBS-seq, single-cell bisulfite sequencing; scCOOL-seq, single-cell chromatin overall omic-scale landscape sequencing; scGEM, single-cell analysis of genotype, expression, and methylation; scMT-seq, single-cell methylome and transcriptome sequencing (using a strategy similar to scTrio-seq); scM&T-seq, single-cell methylome and transcriptome sequencing (using a protocol based on G&T-seq); scNMT-seq, single-cell sequencing technique that combines scNOMe-seq and scM&T-seq; scNOMe-seq, single-cell nucleosome occupancy and methylome sequencing; SCRAM, single-cell restriction analysis of methylation; scRNA-seq, single-cell RNA sequencing; scRRBS, single-cell reduced-representation bisulfite sequencing; scTrio-seq, single-cell triple-omics sequencing; STA, shifted termination assay; WGA, whole-genome amplification; —, not applicable.
[a]Isolation technique used in the original publication.

There are three approaches that enable profiling the DNA and RNA from the same single cell. The first approach, DNA–RNA sequencing (DR-seq) (31), preamplifies gDNA and mRNA simultaneously before splitting the reaction in two in order to finalize the gDNA and mRNA sequencing libraries separately. The gDNA is processed using a modified approach for multiple annealing and looping–based amplification cycles (MALBAC)—a WGA method that combines multiple rounds of displacement amplification with PCR (162)—while the mRNA is processed using a modified CEL-seq approach (59). By avoiding up-front separation of DNA and RNA, this method minimizes the risk for loss of nucleic acids, but it is confounded by RNA-derived amplification products contaminating the DNA sequencing (DNA-seq) data.

The second method, genome and transcriptome sequencing (G&T-seq), relies on physical separation of mRNA from gDNA (92, 94). Following cell lysis, the polyadenylated mRNA is separated from the gDNA by binding to oligo(dT)-coated paramagnetic beads. The mRNA is amplified using a modified Smart-seq2 protocol (112, 113), while the DNA in the polyadenylated mRNA–depleted cell lysate is precipitated on-bead and subjected to WGA and conventional library preparation. Physical separation has the advantage of choice of downstream handling methods for both DNA and RNA but could lead to potential loss of nucleic acids, although no major losses have been reported to date. Additionally, G&T-seq has been automated on conventional liquid handling robotics, improving the robustness and throughput of single-cell multiomics. The single-cell transcriptogenomics method published by Li et al. (86) is based on similar principles to G&T-seq.

A third approach separates the nucleus and cytoplasm from the same cell, allowing genomic (57, 147) or epigenomic (62, 64) analysis of the nucleus and transcriptomic analysis of the cytoplasmic RNA. While these approaches also offer flexibility in the protocols applied downstream of the separation of nuclear DNA and cytoplasmic RNA, the separation in this case involves loss of nuclear transcripts, and often the full content of the cytoplasm cannot be separated from the nucleus, resulting in some loss of cytoplasmic RNA for transcriptomics analysis as well. Furthermore, these approaches are currently mostly manually performed and subject to low throughput, but they are amenable to automation and microfluidics (57, 147).

## Single-Cell Epigenome-Plus-Transcriptome Sequencing

The epigenome largely orchestrates how a cell reads its genome and produces its transcriptome. However, the association of epigenomic heterogeneity with transcriptional heterogeneity remains largely to be determined at single-cell resolution. Additionally, the cause and effect of transcriptomic and epigenetic changes during cell state transitions and cellular differentiation may be elucidated by profiling both simultaneously.

Single-cell triple-omics sequencing (scTrio-seq) (62) allows assessment of the genomic copy number variations, DNA methylome, and transcriptome of a single cell. In this approach, an individual cell is lysed gently, leaving the nucleus intact while releasing cytoplasmic mRNAs. Centrifugation is then used to pellet the nucleus, and the cytoplasmic supernatant is transferred to another tube. The mRNA in the supernatant is converted into an scRNA-seq library using the method published by Tang et al. (143). In parallel, the nucleus is processed using the scRRBS method (54), yielding a reduced-representation overview of 5mC in the genome and a low-resolution DNA copy number landscape. Because of systematic coverage bias in scRRBS data, DNA copy number computation from focal read depth analyses in 10-Mb windows required a normalization factor for each genomic bin, which was determined from sequence depth values of bulk RRBS data.

Single-cell methylome and transcriptome sequencing (scMT-seq) applies a similar strategy to scTrio-seq (64) and also lyses the cellular membrane but not the nuclear membrane, after which

the nucleus is isolated by microcapillary picking. The mRNA in the lysate is amplified using a modified Smart-seq2 protocol (112, 113), while the nuclear genome is processed using a modified version of the scRRBS protocol (53) in parallel.

Single-cell methylome and transcriptome sequencing (scM&T-seq) (5) is based on the G&T-seq protocol but applies scBS-seq on the gDNA following physical separation of DNA and RNA. In contrast to scTrio-seq and scMT-seq, scM&T-seq thus provides a broader CpG methylation landscape of the cell and is not confounded by loss of the entire nuclear transcript pool, but may drop some nucleic acids during the physical separation of gDNA and polyadenylated mRNA.

Single-cell analysis of genotype, expression, and methylation (scGEM) (20) first isolates single cells in a Fluidigm C1 Single-Cell Auto Prep system. Cells are then lysed within the nanoliter-scale reaction chambers, and the polyadenylated mRNA is converted to cDNA. A protease treatment removes chromatin-associated proteins before a methylation-sensitive digest is carried out using a modified SCRAM assay. A PCR preamplification is performed on-chip using primers targeting specific cDNAs and gDNA loci for DNA methylation and genotype profiling. The preamplified products are split to generate qPCR-based targeted readouts of gene expression and methylation status as well as sequencing-based genotypes. In contrast to DNA methylation assays that require bisulfite and thus are prone to stochastic dropouts, scGEM enables a more reproducible assessment of methylation status at specific sites across cells.

Apart from methodologies that capture the DNA methylome together with transcriptome and/or genome information, several methods have recently been devised that also allow one to read multiple layers of the epigenome of a single cell. The latest method further combines this with RNA-seq of the same cell.

Single-cell nucleosome occupancy and methylome sequencing (scNOMe-seq) was reported in 2017 (116). Conventional NOMe-seq (75) employs a GpC methyltransferase, M.CviPI, to probe chromatin accessibility (75, 76). M.CviPI methylates cytosines in accessible GpC dinucleotides (i.e., nonnucleosomal DNA) in vitro. In mammalian cells, the cytosine in CpG dinucleotides is methylated in vivo rather than the other cytosine-containing dinucleotides—CpA, CpC, or CpT (87). Bisulfite sequencing then follows, where unmethylated cytosines are converted to uracils and endogenous CpG methylation is encoded simultaneously with GpC accessibility in the DNA-seq data. This has two advantages over count-based methods for chromatin accessibility profiling (e.g., ATAC-seq or DNase-seq): Inaccessible chromatin is better distinguished from missing data, and the resolution is determined by the frequency of GpC sites in the genome (1 in 16 base pairs) rather than library insert size (>100 base pairs).

Single-cell chromatin overall omic-scale landscape sequencing (scCOOL-seq) (52) is an alternative method for scNOMe-seq (116) and allows multiple layers of the epigenome to be accessed, including not only chromatin accessibility and DNA methylation but also DNA copy number variation and ploidy. To do so, it modifies bulk NOMe-seq (56) with postbisulfite adaptor tagging sequencing, which is the basis for scBS-seq (131). DNA copy number information is derived from focal read depth analyses in 1-Mb genomic bins coupled with a hidden Markov model to determine copy number segments. To infer the single-cell ploidy from single-cell COOL-seq data, the same quantity of λDNA is spiked to each single-cell sample to deduce the ploidy of the cell.

A 2018 article describes a novel three-layer technique, scNMT-seq (23), which combines the principle of scNOMe-seq (116) with scM&T-seq (5). Single cells are sorted by FACS into tubes containing a GpC methylase reaction mixture before the DNA and RNA are physically separated and processed as in the scM&T-seq protocol. The chromatin accessibility and DNA methylation can then be measured by using the principles of scNOMe-seq (116), while the transcriptome information is provided by a modified Smart-seq2 protocol (112, 113).

## Single-Cell Transcriptome-Plus-Protein Measurements

Understanding how transcripts are translated into proteins on a genome-wide scale is important in comprehending how transcriptomic cellular states translate into functional phenotypic states. Additionally, this may reveal phenotypic cell states that may not be observable from scRNA-seq alone owing to potential heterogeneity in posttranscriptional and posttranslational processes.

One technology to simultaneously measure protein and transcript levels in a single cell is enabled by combining the proximity extension assay with targeted cDNA amplification (46). This assay uses a pair of antibodies that bind to the same protein at different epitopes. These antibodies are functionalized with single-stranded DNA oligonucleotides that have complementary 3′ ends. The complementary oligonucleotides on antibodies binding nearby on the same protein will anneal in a cell lysate. The dual DNA polymerase–RNA polymerase activity of reverse transcriptases then allows simultaneous generation of a protein-derived DNA barcode and cDNA from RNA. Protein and transcript levels can subsequently be measured by qPCR or sequencing, with protein detection remaining limited mainly by available antibody pairs. This technique enabled parallel measurement of 38 proteins and 96 transcripts.

In an alternative approach, cells are isolated by FACS and lysed (27), after which the lysate is split to measure RNA and protein separately. The proximity extension assay is performed on proteins and reverse transcription is performed on RNA, both of which are then read out by microfluidic qPCR assays. The correlation of transcripts and proteins from 22 genes in single cells was investigated using this method.

In the proximity ligation assay for RNA (PLAYR) approach (42), cells are fixed and permeabilized, and specific transcripts and proteins are targeted for detection by mass cytometry. The transcripts are hybridized with PLAYR probes, which are in turn detected by insert and backbone probes that form a DNA circle following ligation. This circle is amplified by rolling circle amplification, and the resulting amplified insert sequences per transcript are hybridized with mass cytometry–compatible probes. Antibodies conjugated to distinct metal isotopes are used to target specific proteins, enabling concurrent RNA and protein measurement by mass cytometry. While this enables rapid evaluation of RNA and proteins across thousands of single cells, it is intrinsically limited by the number of metal isotope tags available as well as the number of high-affinity antibodies.

Cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq) (136) uses a cytometry-by-sequencing approach. Single cells in suspension are labeled with oligonucleotide-functionalized antibodies that target cell surface proteins. These oligonucleotides comprise a PCR handle, an antibody barcode, and a poly(A) tail and are conjugated via a streptavidin–biotin interaction to the antibodies. Additionally, a disulfide bond at the 5′ end of the oligonucleotide allows the oligonucleotide to be released in reducing conditions. Cells are coencapsulated in droplets with Drop-seq beads or 10x Genomics Chromium gel beads; during lysis, these antibody-derived tags and the mRNAs of the cell anneal to the oligo(dT) on the beads. After reverse transcription and amplification, the cDNA and antibody-derived tags can be separated by size (<300-nucleotide fragments containing antibody-derived tags and >300-nucleotide fragments representing mRNAs). These can be pooled on the same sequencing lane in a manner that allows for adjustments in depth. This study targeted immune cell surface proteins, but in principle the method could access cytosolic and perhaps nuclear proteins if a permeabilization step was performed prior to antibody staining. Notably, this study could detect subtypes of natural killer cells (CD56 bright and dim) that could not be distinguished by scRNA-seq alone.

The RNA expression and protein sequencing assay (REAP-seq) (110) also uses DNA barcode-conjugated antibodies, which are incubated with a cell suspension, followed by droplet

scRNA-seq. REAP-seq uses a Thunder-Link PLUS oligonucleotide conjugation system to link oligonucleotides to barcodes, which is distinct from CITE-seq. After cell lysis, reverse transcriptase produces cDNA from the mRNAs and extends the antibody barcode [which contains a poly(A) region] through its DNA polymerase activity. Similarly to CITE-seq, the two cell-barcoded libraries can be separated by size (155 versus >500 base pairs). The method enabled the detection of 82 proteins in parallel with scRNA-seq.

## APPLICATIONS: CURRENT AND FUTURE

We present two contrasting application areas for multiomics: cancer, in which the technology is emerging to deliver novel insight in the disease, and atlas projects, in which the technology may have great potential.

### Cancer

Tumors comprise different subpopulations of cancer cells, along with stromal cells that provide support. However, the precise combination of driver mutations present in cancer cells, their order of acquisition, and the extent, nature, and functional consequences of the (epi)genomic diversity that arises in the cancer cells remain largely unknown because bulk DNA or RNA analyses are inadequate. For the same reason, our understanding of the potential plurality of cancer cell states within a tumor—of which some can metastasize and/or evade therapy—fueled by (epi)genetic alterations remains limited. Similarly, a better understanding of the heterogeneous stromal cell population, which is a key contributor to the tumor microenvironment, has been hampered by classic bulk molecular profiling methods. Tumor cells coevolve with the surrounding microenvironment and interact to promote tumor growth and development. The stromal cells are recruited from local host stroma and promote processes such as extracellular matrix remodeling, cellular migration, neoangiogenesis, invasion, drug resistance, and evasion of immune surveillance. Questions remain about the precise cellular composition of the microenvironment, the processes of transdifferentiation from stromal cell to stromal cell (e.g., pericytes transdifferentiating into endothelial cells or fibroblasts) or tumor cell to stromal cell [e.g., transdifferentiation of glioblastoma stem-like cells into mural cells driving vasculogenic mimicry (125)], and the interactions between stromal cells and tumor cells that have a potential role in cancer growth and progression (15). Understanding the nature, extent, and biology of cellular heterogeneity within a cancer is paramount for the design of effective therapies (3).

Single-cell multiomics methods can provide novel insights into the development of intratumor cellular heterogeneity. For instance, methods enabling genome-plus-transcriptome sequencing such as DR-seq (31), G&T-seq (92), and microfluidic versions (57, 147) may allow determination of the subclonal genetic architecture and phylogenetic tree, which exposes the order of acquisition of mutations in the cancer cells from single-cell DNA sequencing (scDNA-seq) analysis. scRNA-seq analysis of the same cells will provide insight into the different transcriptomic cancer cell states with potential functional dissimilarities that arise within a tumor, providing unprecedented insight into the extent to which these different phenotypic cancer cell states are driven by genetic alterations, nongenetic alterations, or both. Layering these data with additional protein information from the same cells will provide deeper insight into the functional dissimilarities among the cells; similarly, reading the epigenomes of these same cells can provide further understanding of the molecular processes driving the different cellular cancer states. Applying similar technologies to circulating tumor cells (i.e., cells that leave the solid tumor and become blood-borne) and disseminated tumor cells (i.e., cells that lead to micrometastases that later may fuel overt metastases in distant organs) may provide further understanding of the genetic and nongenetic mechanisms underlying cancer

cell dissemination and the dormancy of cancer cells, respectively. These multiomics methods can also be applied to the stromal cells, which in combination with spatial methodologies may provide unprecedented insight into the subpopulations that exist among the stromal cells, how these subpopulations are molecularly determined, and how they interact with nearby cancer cells to promote tumor growth and progression. In addition to improving the fundamental understanding of tumor development, single-cell multiomics methods have the capacity to provide a deeper understanding of the diversity of drug resistance mechanisms applied by cells, which can be genetic or nongenetic in nature (37, 43, 142).

### Atlas Projects

The ability of scRNA-seq to reveal (sub)types and phenotypic states of cells has led to the initiation of cell atlas projects. The most ambitious is the Human Cell Atlas, which is being created by an international consortium (119, 121). These projects aim to provide data-driven identification and classification of cells (119), which will enable a more profound fundamental understanding of the normal development of organisms and how dysregulation of cellular processes may contribute to aging and disease. Although these atlases will initially comprise reference sets of human and animal cells based on genes detected as expressed in scRNA-seq data and with spatial transcriptomic methods (61), single-cell multiomics will be required to provide a deeper understanding of the key molecular drivers and mechanisms underpinning cellular states. For example, open-chromatin-plus-transcriptome profiling of single cells, which may be enabled by scNMT-seq (23), at scale will provide a more profound understanding of the gene regulatory networks involved in the definition of cell states and how they may become dysregulated in disease. Similarly, layering phenotypic states of cells with cell lineage will provide an understanding of the normal developmental trajectories of organs and their cellular architecture, the noise within such trajectories during the development of different individuals, and how these trajectories may become perturbed, leading to disease.

## FUTURE DEVELOPMENTS

The rate of development of single-cell multiomics protocols is now increasing rapidly, with many new methods published or posted to a preprint server since this field was last reviewed (10, 93). The current trend of generating larger and larger scRNA-seq data sets (121, 141) is likely to be mirrored in single-cell multiomics data sets.

Further progress is likely to be achieved by refining existing mono-omics techniques, including scRNA-seq, scDNA-seq, and single-cell ATAC-seq, by increasing their sensitivity and accuracy; such improvements are likely to be transferable to multiomics techniques. For example, most scRNA-seq protocols often require more than 10 mRNA molecules in order to reliably detect gene expression (140); lowering this threshold would allow better detection of low-copy-number mRNAs that encode, for instance, transcription factors. The breadth of these omic layers may also improve, with the capture of more classes of RNAs, more target proteins, and more DNA modifications. Metabolomics is currently extremely challenging at the single-cell level, but progress is being made (35).

Methods that enable new combinations of omic layers are likely to appear. Long-read technologies may be able to codetect DNA sequence and DNA methylation as well as other DNA marks and adducts on the same molecule (10). Additionally, long-read sequencing may advance the detection of structural variants along with other types of genetic variation in cells. High-resolution capture of spatial information in combination with scRNA-seq may be extended to include capture

of other nucleic acids or antibody tags. Greater linkage of live-cell imaging data with single-cell sequence data may allow the interrogation of more complex phenotypes. Nuclear structure remains unexplored in conjunction with other omic layers; although these studies use single nuclei, no study has combined these techniques with simultaneous RNA or protein profiling from the cytoplasmic fraction of the same cell. We also anticipate the development of technologies that will profile four or five layers of omics data in parallel by further combining existing techniques. Additionally, the use of combinatorial indexing or microfluidics approaches is likely to result in single-cell multiomics methods that enable high-throughput processing of thousands of cells.

Cost is likely to be a barrier to some single-cell multiomics technologies, particularly those that aim to capture the whole genome. Library costs may be reduced by scaling down volumes of reactions, but linearly scaling sequencing costs may prove a greater barrier for most studies. Compromises may be made to sample gDNA by using targeted sequencing of a panel of genes known to be of interest.

Novel computational methods that allow the integrated study of two or more omic layers per cell of large heterogeneous populations will need to be developed. Most current studies analyze each omics data set separately and compare final results. An example of this approach is scTrio-seq (62), where copy number variations are separately analyzed in conjunction with either the methylome or the transcriptome. Specialist multiomics algorithms should be developed to analyze the different layers of data together.

In the next few years, it is likely that greater progress will be made toward "omni-omics"—the capture of all molecules in a cell. Because the cellular phenome is defined as the set of all phenotypes expressed by a cell, multiomics ultimately aims to characterize these omnisciently for different layers. Future multiomics technologies will eventually need to encompass all single-cell omics technologies or find an analytical way to omnisciently characterize a cell's "ome" layers, three-dimensional coordinates, phenotypes, and cell lineage history.

## DISCLOSURE STATEMENT

T.V. is coinventor on patents WO/2011/157846 (methods for haplotyping single cells), WO/2014/053664 (high-throughput genotyping by sequencing low amounts of genetic material), and WO/2015/028576 (haplotyping and copy number typing using polymorphic variant allelic frequencies).

## ACKNOWLEDGMENTS

L.C., A.J.C.R., and T.V. are funded by Wellcome (UK). L.C. is funded by two Wellcome Strategic Awards (105031/E/14/Z and 105045/Z/14/Z). A.J.C.R. is funded by a Wellcome Sanger Institute PhD studentship. T.V. is also funded by the Research Foundation Flanders (FWO, Belgium; G092415N), the Foundation Against Cancer (Belgium; 2014-145 F/2014/297), and the University of Leuven (KU Leuven, Belgium; PFV/10/016).

## LITERATURE CITED

1. Abdelmoez MN, Iida K, Oguchi Y, Nishikii H, Yokokawa R, et al. 2017. Correlation of gene expressions between nucleus and cytoplasm reflects single-cell physiology. bioRxiv 206672. **https://doi.org/10.1101/206672**
2. Achim K, Pettit J-B, Saraiva LR, Gavriouchkina D, Larsson T, et al. 2015. High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.* 33:503–9

3. Alizadeh AA, Aranda V, Bardelli A, Blanpain C, Bock C, et al. 2015. Toward understanding and exploiting tumor heterogeneity. *Nat. Med.* 21:846–53
4. Angelo M, Bendall SC, Finck R, Hale MB, Hitzman C, et al. 2014. Multiplexed ion beam imaging of human breast tumors. *Nat. Med.* 20:436–42
5. Angermueller C, Clark SJ, Lee HJ, Macaulay IC, Teng MJ, et al. 2016. Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat. Methods* 13:229–32
6. Bagnoli JW, Ziegenhain C, Janjic A, Wange LE, Vieth B, et al. 2017. mcSCRB-seq: sensitive and powerful single-cell RNA sequencing. bioRxiv 188367. **https://doi.org/10.1101/188367**
7. Bakker B, Taudt A, Belderbos ME, Porubsky D, Spierings DCJ, et al. 2016. Single-cell sequencing reveals karyotype heterogeneity in murine and human malignancies. *Genome Biol.* 17:115
8. Biesecker LG, Spinner NB. 2013. A genomic view of mosaicism and human disease. *Nat. Rev. Genet.* 14:307–20
9. Bjornson ZB, Nolan GP, Fantl WJ. 2013. Single-cell mass cytometry for analysis of immune system functional states. *Curr. Opin. Immunol.* 25:484–94
10. Bock C, Farlik M, Sheffield NC. 2016. Multi-omics of single cells: strategies and applications. *Trends Biotechnol.* 34:605–8
11. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, et al. 2008. High-resolution mapping and characterization of open chromatin across the genome. *Cell* 132:311–22
12. Budnik B, Levy E, Slavov N. 2017. Mass-spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. bioRxiv 102681. **https://doi.org/10.1101/102681**
13. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* 10:1213–18
14. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, et al. 2015. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 523:486–90
15. Bussard KM, Mutkus L, Stumpf K, Gomez-Manzano C, Marini FC. 2016. Tumor-associated stromal cells as key contributors to the tumor microenvironment. *Breast Cancer Res.* 18:84
16. Cadwell CR, Palasantza A, Jiang X, Berens P, Deng Q, et al. 2016. Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. *Nat. Biotechnol.* 34:199–203
17. Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, et al. 2017. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* 357:661–67
18. Chen C, Xing D, Tan L, Li H, Zhou G, et al. 2017. Single-cell whole-genome analyses by Linear Amplification via Transposon Insertion (LIANTI). *Science* 356:189–94
19. Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. 2015. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 348:aaa6090
20. Cheow LF, Courtois ET, Tan Y, Viswanathan R, Xing Q, et al. 2016. Single-cell multimodal profiling reveals cellular epigenetic heterogeneity. *Nat. Methods* 13:833–36
21. Cheow LF, Quake SR, Burkholder WF, Messerschmidt DM. 2015. Multiplexed locus-specific analysis of DNA methylation in single cells. *Nat. Protoc.* 10:619–31
22. Chu WK, Edge P, Lee HS, Bansal V, Bafna V, et al. 2017. Ultraaccurate genome sequencing and haplotyping of single human cells. *PNAS* 114:12512–17
23. Clark SJ, Argelaguet R, Kapourani C-A, Stubbs TM, Lee HJ, et al. 2018. scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat. Commun.* 9:781
24. Clark SJ, Smallwood SA, Lee HJ, Krueger F, Reik W, Kelsey G. 2017. Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq). *Nat. Protoc.* 12:534–47
25. Cooper J, Ding Y, Song J, Zhao K. 2017. Genome-wide mapping of DNase I hypersensitive sites in rare cell populations using single-cell DNase sequencing. *Nat. Protoc.* 12:2342–54
26. Cusanovich DA, Daza R, Adey A, Pliner HA, Christiansen L, et al. 2015. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* 348:910–14
27. Darmanis S, Gallant CJ, Marinescu VD, Niklasson M, Segerman A, et al. 2016. Simultaneous multiplexed measurement of RNA and proteins in single cells. *Cell Rep.* 14:380–89

28. Davis A, Gao R, Navin N. 2017. Tumor evolution: linear, branching, neutral or punctuated? *Biochim. Biophys. Acta* 1867:151–61
29. De S. 2011. Somatic mosaicism in healthy human tissues. *Trends Genet.* 27:217–23
30. Deaton AM, Bird A. 2011. CpG islands and the regulation of transcription. *Genes Dev.* 25:1010–22
31. Dey SS, Kester L, Spanjaard B, Bienko M, van Oudenaarden A. 2015. Integrated genome and transcriptome sequencing of the same cell. *Nat. Biotechnol.* 33:285–89
32. Dong X, Zhang L, Milholland B, Lee M, Maslov AY, et al. 2017. Accurate identification of single-nucleotide variants in whole-genome-amplified single cells. *Nat. Methods* 14:491–93
33. Dumanski JP, Piotrowski A. 2012. Structural genetic variation in the context of somatic mosaicism. *Methods Mol. Biol.* 838:249–72
34. Eberwine J, Yeh H, Miyashiro K, Cao Y, Nair S, et al. 1992. Analysis of gene expression in single live neurons. *PNAS* 89:3010–14
35. Emara S, Amer S, Ali A, Abouleila Y, Oga A, Masujima T. 2017. Single-cell metabolomics. In *Metabolomics: From Fundamentals to Clinical Applications*, ed. A Sussulini, pp. 323–43. Cham, Switz.: Springer
36. Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, et al. 1996. Laser capture microdissection. *Science* 274:998–1001
37. Fallahi-Sichani M, Becker V, Izar B, Baker GJ, Lin J-R, et al. 2017. Adaptive resistance of melanoma cells to RAF inhibition via reversible induction of a slowly dividing de-differentiated state. *Mol. Syst. Biol.* 13:905
38. Faridani OR, Abdullayev I, Hagemann-Jensen M, Schell JP, Lanner F, Sandberg R. 2016. Single-cell sequencing of the small-RNA transcriptome. *Nat. Biotechnol.* 34:1264–66
39. Farlik M, Sheffield NC, Nuzzo A, Datlinger P, Schönegger A, et al. 2015. Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. *Cell Rep.* 10:1386–97
40. Ficz G. 2015. New insights into mechanisms that regulate DNA methylation patterning. *J. Exp. Biol.* 218:14–20
41. Flyamer IM, Gassler J, Imakaev M, Brandão HB, Ulianov SV, et al. 2017. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* 544:110–14
42. Frei AP, Bava F-A, Zunder ER, Hsieh EWY, Chen S-Y, et al. 2016. Highly multiplexed simultaneous detection of RNAs and proteins in single cells. *Nat. Methods* 13:269–75
43. Garraway LA, Lander ES. 2013. Lessons from the cancer genome. *Cell* 153:17–37
44. Gawad C, Koh W, Quake SR. 2016. Single-cell genome sequencing: current state of the science. *Nat. Rev. Genet.* 17:175–88
45. Gay L, Baker A-M, Graham TA. 2016. Tumour cell heterogeneity. *F1000Research* 5:238
46. Genshaft AS, Li S, Gallant CJ, Darmanis S, Prakadan SM, et al. 2016. Multiplexed, targeted profiling of single-cell proteomes and transcriptomes in a single reaction. *Genome Biol.* 17:188
47. Gierahn TM, Wadsworth MH II, Hughes TK, Bryson BD, Butler A, et al. 2017. Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat. Methods* 14:395–98
48. Giesen C, Wang HAO, Schapiro D, Zivanovic N, Jacobs A, et al. 2014. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat. Methods* 11:417–22
49. Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.* 17:877–85
50. Gravina S, Dong X, Yu B, Vijg J. 2016. Single-cell genome-wide bisulfite sequencing uncovers extensive heterogeneity in the mouse liver methylome. *Genome Biol.* 17:150
51. Guatelli JC, Whitfield KM, Kwoh DY, Barringer KJ, Richman DD, et al. 1990. Isothermal, in vitro amplification of nucleic acids by a multienzyme reaction modeled after retroviral replication. *PNAS* 87:1874–78
52. Guo F, Li L, Li J, Wu X, Hu B, et al. 2017. Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells. *Cell Res.* 27:967–88
53. Guo H, Zhu P, Guo F, Li X, Wu X, et al. 2015. Profiling DNA methylome landscapes of mammalian cells with single-cell reduced-representation bisulfite sequencing. *Nat. Protoc.* 10:645–59

54. Guo H, Zhu P, Wu X, Li X, Wen L, Tang F. 2013. Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Res.* 23:2126–35
55. Guo H, Zhu P, Yan L, Li R, Hu B, et al. 2014. The DNA methylation landscape of human early embryos. *Nature* 511:606–10
56. Han H, Cortez CC, Yang X, Nichols PW, Jones PA, Liang G. 2011. DNA methylation directly silences genes with non-CpG island promoters and establishes a nucleosome occupied promoter. *Hum. Mol. Genet.* 20:4299–310
57. Han L, Zi X, Garmire LX, Wu Y, Weissman SM, et al. 2014. Co-detection and sequencing of genes and transcripts from the same single cells facilitated by a microfluidics platform. *Sci. Rep.* 4:6485
58. Hashimshony T, Senderovich N, Avital G, Klochendler A, de Leeuw Y, et al. 2016. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* 17:77
59. Hashimshony T, Wagner F, Sher N, Yanai I. 2012. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* 2:666–73
60. Hayatsu H. 2008. The bisulfite genomic sequencing used in the analysis of epigenetic states, a technique in the emerging environmental genotoxicology research. *Mutat. Res.* 659:77–82
61. Hon C-C, Shin JW, Carninci P, Stubbington MJT. 2018. The Human Cell Atlas: technical approaches and challenges. *Brief. Funct. Genom.* 17:283–94
62. Hou Y, Guo H, Cao C, Li X, Hu B, et al. 2016. Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res.* 26:304–19
63. Hu P, Zhang W, Xin H, Deng G. 2016. Single cell isolation and analysis. *Front. Cell Dev. Biol.* 4:116
64. Hu Y, Huang K, An Q, Du G, Hu G, et al. 2016. Simultaneous profiling of transcriptome and DNA methylome from a single cell. *Genome Biol.* 17:88
65. Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A. 2010. The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLOS ONE* 5:e8888
66. Ito S, Shen L, Dai Q, Wu SC, Collins LB, et al. 2011. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333:1300–3
67. Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, et al. 2014. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* 343:776–79
68. Ji Z, Ji H. 2016. TSCAN: pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. *Nucleic Acids Res.* 44:e117
69. Jin S-G, Kadam S, Pfeifer GP. 2010. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. *Nucleic Acids Res.* 38:e125
70. Jin W, Tang Q, Wan M, Cui K, Zhang Y, et al. 2015. Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. *Nature* 528:142–46
71. Jones PA. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* 13:484–92
72. Kawasaki ES. 2004. Microarrays and the gene expression profile of a single cell. *Ann. N.Y. Acad. Sci.* 1020:92–100
73. Ke R, Mignardi M, Hauling T, Nilsson M. 2016. Fourth generation of next-generation sequencing technologies: promise and consequences. *Hum. Mutat.* 37:1363–67
74. Kellinger MW, Song C-X, Chong J, Lu X-Y, He C, Wang D. 2012. 5-Formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* 19:831–33
75. Kelly TK, Liu Y, Lay FD, Liang G, Berman BP, Jones PA. 2012. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res.* 22:2497–506
76. Kilgore JA, Hoose SA, Gustafson TL, Porter W, Kladde MP. 2007. Single-molecule and population probing of chromatin structure using DNA methyltransferases. *Methods* 41:320–32
77. Kim DH, Marinov GK, Pepke S, Singer ZS, He P, et al. 2015. Single-cell transcriptome analysis reveals dynamic changes in lncRNA expression during reprogramming. *Cell Stem Cell* 16:88–101
78. Kind J, Pagie L, de Vries SS, Nahidiazar L, Dey SS, et al. 2015. Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* 163:134–47

79. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, et al. 2015. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161:1187–201
80. Kloosterman WP, Cuppen E. 2013. Chromothripsis in congenital disorders and cancer: similarities and differences. *Curr. Opin. Cell Biol.* 25:341–48
81. Kloosterman WP, Guryev V, van Roosmalen M, Duran KJ, de Bruijn E, et al. 2011. Chromothripsis as a mechanism driving complex de novo structural rearrangements in the germline. *Hum. Mol. Genet.* 20:1916–24
82. Korbel JO, Campbell PJ. 2013. Criteria for inference of chromothripsis in cancer genomes. *Cell* 152:1226–36
83. Lee JH. 2017. De novo gene expression reconstruction in space. *Trends Mol. Med.* 23:583–93
84. Leontiou CA, Hadjidaniel MD, Mina P, Antoniou P, Ioannides M, Patsalis PC. 2015. Bisulfite conversion of DNA: performance comparison of different kits and methylation quantitation of epigenetic biomarkers that have the potential to be used in non-invasive prenatal testing. *PLOS ONE* 10:e0135058
85. Li G, Reinberg D. 2011. Chromatin higher-order structures and gene regulation. *Curr. Opin. Genet. Dev.* 21:175–86
86. Li W, Calder RB, Mar JC, Vijg J. 2015. Single-cell transcriptogenomics reveals transcriptional exclusion of ENU-mutated alleles. *Mutat. Res.* 772:55–62
87. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, et al. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462:315–22
88. Liu Z, Lou H, Xie K, Wang H, Chen N, et al. 2017. Reconstructing cell cycle pseudo time-series via single-cell transcriptome data. *Nat. Commun.* 8:22
89. Lorthongpanich C, Cheow LF, Balu S, Quake SR, Knowles BB, Burkholder WF. 2013. Single-cell DNA-methylation analysis reveals epigenetic chimerism in preimplantation embryos. *Science* 341:1110–12
90. Lu Y, Xue Q, Eisele MR, Sulistijo ES, Brower K, et al. 2015. Highly multiplexed profiling of single-cell effector functions reveals deep functional heterogeneity in response to pathogenic ligands. *PNAS* 112:E607–15
91. Luo C, Keown CL, Kurihara L, Zhou J, He Y, et al. 2017. Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* 357:600–4
92. Macaulay IC, Haerty W, Kumar P, Li YI, Hu TX, et al. 2015. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat. Methods* 12:519–22
93. Macaulay IC, Ponting CP, Voet T. 2017. Single-cell multiomics: multiple measurements from single cells. *Trends Genet.* 33:155–68
94. Macaulay IC, Teng MJ, Haerty W, Kumar P, Ponting CP, Voet T. 2016. Separation and parallel sequencing of the genomes and transcriptomes of single cells using G&T-seq. *Nat. Protoc.* 11:2081–103
95. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, et al. 2015. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161:1202–14
96. Manno GL, Soldatov R, Hochgerner H, Zeisel A, Petukhov V, et al. 2017. RNA velocity in single cells. bioRxiv 206052. **https://doi.org/10.1101/206052**
97. McConnell MJ, Lindberg MR, Brennand KJ. 2013. Mosaic copy number variation in human neurons. *Science* 342:632–37
98. Meyer CA, Liu XS. 2014. Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. *Nat. Rev. Genet.* 15:709–21
99. Miura F, Enomoto Y, Dairiki R, Ito T. 2012. Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Res.* 40:e136
100. Mooijman D, Dey SS, Boisset J-C, Crosetto N, van Oudenaarden A. 2016. Single-cell 5hmC sequencing reveals chromosome-wide cell-to-cell variability and enables lineage reconstruction. *Nat. Biotechnol.* 34:852–56
101. Moore DL, Jessberger S. 2017. Creating age asymmetry: consequences of inheriting damaged goods in mammalian cells. *Trends Cell Biol.* 27:82–92
102. Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, et al. 2013. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502:59–64

103. Nagano T, Lubling Y, Várnai C, Dudley C, Leung W, et al. 2017. Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* 547:61–67
104. Nagano T, Lubling Y, Yaffe E, Wingett SW, Dean W, et al. 2015. Single-cell Hi-C for genome-wide detection of chromatin interactions that occur simultaneously in a single cell. *Nat. Protoc.* 10:1986–2003
105. Neri F, Incarnato D, Krepelova A, Parlato C, Oliviero S. 2016. Methylation-assisted bisulfite sequencing to simultaneously map 5fC and 5caC on a genome-wide scale for DNA demethylation analysis. *Nat. Protoc.* 11:1191–205
106. Nussbaum RL, McInnes RR, Willard HF. 2007. *Thompson & Thompson Genetics in Medicine*. Philadelphia: Saunders. 7th ed.
107. Ogrodnik M, Salmonowicz H, Brown R, Turkowska J, Średniawa W, et al. 2014. Dynamic JUNQ inclusion bodies are asymmetrically inherited in mammalian cell lines through the asymmetric partitioning of vimentin. *PNAS* 111:8049–54
108. Peng G, Suo S, Chen J, Chen W, Liu C, et al. 2016. Spatial transcriptome for the molecular annotation of lineage fates and cell identity in mid-gastrula mouse embryo. *Dev. Cell* 36:681–97
109. Perfetto SP, Chattopadhyay PK, Roederer M. 2004. Seventeen-colour flow cytometry: unravelling the immune system. *Nat. Rev. Immunol.* 4:648–55
110. Peterson VM, Zhang KX, Kumar N, Wong J, Li L, et al. 2017. Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.* 35:936–39
111. Picelli S. 2017. Single-cell RNA-sequencing: the future of genome biology is now. *RNA Biol.* 14:637–50
112. Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R. 2013. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* 10:1096–98
113. Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R. 2014. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* 9:171–81
114. Poduri A, Evrony GD, Cai X, Walsh CA. 2013. Somatic mutation, genomic variation, and neurological disease. *Science* 341:1237758
115. Ponomarenko EA, Poverennaya EV, Ilgisonis EV, Pyatnitskiy MA, Kopylov AT, et al. 2016. The size of the human proteome: the width and depth. *Int. J. Anal. Chem.* 2016:7436849
116. Pott S. 2017. Simultaneous measurement of chromatin accessibility, DNA methylation, and nucleosome phasing in single cells. *eLife* 6:e23203
117. Prakadan SM, Shalek AK, Weitz DA. 2017. Scaling by shrinking: empowering single-cell "omics" with microfluidic devices. *Nat. Rev. Genet.* 18:345–61
118. Ramani V, Deng X, Qiu R, Gunderson KL, Steemers FJ, et al. 2017. Massively multiplex single-cell Hi-C. *Nat. Methods* 14:263–66
119. Regev A, Teichmann S, Lander ES, Amit I, Benoist C, et al. 2017. The Human Cell Atlas. *eLife* 6:e27041
120. Rotem A, Ram O, Shoresh N, Sperling RA, Goren A, et al. 2015. Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat. Biotechnol.* 33:1165–72
121. Rozenblatt-Rosen O, Stubbington MJT, Regev A, Teichmann SA. 2017. The Human Cell Atlas: from vision to reality. *Nature* 550:451–53
122. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. 2015. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* 33:495–502
123. Schlesinger F, Smith AD, Gingeras TR, Hannon GJ, Hodges E. 2013. De novo DNA demethylation and noncoding transcription define active intergenic regulatory elements. *Genome Res.* 23:1601–14
124. Scialdone A, Tanaka Y, Jawaid W, Moignard V, Wilson NK, et al. 2016. Resolving early mesoderm diversification through single-cell expression profiling. *Nature* 535:289–93
125. Scully S, Francescone R, Faibish M, Bentley B, Taylor SL, et al. 2012. Transdifferentiation of glioblastoma stem-like cells into mural cells drives vasculogenic mimicry in glioblastomas. *J. Neurosci.* 32:12950–60
126. Setty M, Tadmor MD, Reich-Zeliger S, Angel O, Salame TM, et al. 2016. Wishbone identifies bifurcating developmental trajectories from single-cell data. *Nat. Biotechnol.* 34:637–45
127. Shah S, Lubeck E, Zhou W, Cai L. 2016. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron* 92:342–57
128. Shah S, Lubeck E, Zhou W, Cai L. 2017. seqFISH accurately detects transcripts in single cells and reveals robust spatial organization in the hippocampus. *Neuron* 94:752–58

129. Shahi P, Kim SC, Haliburton JR, Gartner ZJ, Abate AR. 2017. Abseq: ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding. *Sci. Rep.* 7:44447
130. Shin J, Berg DA, Zhu Y, Shin JY, Song J, et al. 2015. Single-cell RNA-seq with waterfall reveals molecular cascades underlying adult neurogenesis. *Cell Stem Cell* 17:360–72
131. Smallwood SA, Lee HJ, Angermueller C, Krueger F, Saadeh H, et al. 2014. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* 11:817–20
132. Snook RD, Harvey TJ, Correia Faria E, Gardner P. 2009. Raman tweezers and their application to the study of singly trapped eukaryotic cells. *Integr. Biol.* 1:43–52
133. Soumillon M, Cacchiarelli D, Semrau S, van Oudenaarden A, Mikkelsen TS. 2014. Characterization of directed differentiation by high-throughput single-cell RNA-Seq. bioRxiv 003236. **https://doi.org/10.1101/003236**
134. Spitzer MH, Nolan GP. 2016. Mass cytometry: single cells, many features. *Cell* 165:780–91
135. Stahl PL, Salmen F, Vickovic S, Lundmark A, Navarro JF, et al. 2016. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 353:78–82
136. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, et al. 2017. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* 14:865–68
137. Strachan T, Read A. 2010. *Human Molecular Genetics*. New York: Garland Sci. 4th ed.
138. Stratton MR, Campbell PJ, Andrew FP. 2009. The cancer genome. *Nature* 458:719–24
139. Subkhankulova T, Gilchrist MJ, Livesey FJ. 2008. Modelling and measuring single cell RNA expression levels find considerable transcriptional differences among phenotypically identical cells. *BMC Genom.* 9:268
140. Svensson V, Natarajan KN, Ly L-H, Miragaia RJ, Labalette C, et al. 2016. Power analysis of single-cell RNA-sequencing experiments. *Nat. Methods* 14:381–87
141. Svensson V, Vento-Tormo R, Teichmann SA. 2017. Exponential scaling of single-cell RNA-seq in the last decade. arXiv:1704.01379 [q-bio.GN]
142. Swanton C. 2012. Intratumor heterogeneity: evolution through space and time. *Cancer Res.* 72:4875–82
143. Tang F, Barbacioru C, Nordman E, Li B, Xu N, et al. 2010. RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nat. Protoc.* 5:516–35
144. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, et al. 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32:381–86
145. Tsompana M, Buck MJ. 2014. Chromatin accessibility: a window into the genome. *Epigenet. Chromatin* 7:33
146. Ullal AV, Peterson V, Agasti SS, Tuang S, Juric D, et al. 2014. Cancer cell profiling by barcoding allows multiplexed protein analysis in fine-needle aspirates. *Sci. Transl. Med.* 6:219ra9
147. van Strijp D, Vulders RCM, Larsen NA, Schira J, Baerlocher L, et al. 2017. Complete sequence-based pathway analysis by differential on-chip DNA and RNA extraction from a single cell. *Sci. Rep.* 7:11030
148. Voet T, Vanneste E, Vermeesch JR. 2011. The human cleavage stage embryo is a cradle of chromosomal rearrangements. *Cytogenet. Genome Res.* 133:160–68
149. Wang R, Jin C, Hu X. 2017. Evidence of drug-response heterogeneity rapidly generated from a single cancer cell. *Oncotarget* 8:41113–24
150. Woodworth MB, Girskis KM, Walsh CA. 2017. Building a lineage from single cells: genetic techniques for cell lineage tracking. *Nat. Rev. Genet.* 18:230–44
151. Wu H, Wu X, Shen L, Zhang Y. 2014. Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing. *Nat. Biotechnol.* 32:1231–40
152. Wu X, Inoue A, Suzuki T, Zhang Y. 2017. Simultaneous mapping of active DNA demethylation and sister chromatid exchange in single cells. *Genes Dev.* 31:511–23
153. Yamane J, Mori T, Taniyama N, Kobayashi K, Fujibuchi W. 2017. Development of enhanced reduced representation bisulfite sequencing method for single-cell methylome analysis. *Genom. Comput. Biol.* 3:49
154. Yates LR, Campbell PJ. 2012. Evolution of the cancer genome. *Nat. Rev. Genet.* 13:795–806
155. Yu M, Hon GC, Szulwach KE, Song C-X, Zhang L, et al. 2012. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* 149:1368–80
156. Zahn H, Steif A, Laks E, Eirew P, VanInsberghe M, et al. 2017. Scalable whole-genome single-cell library preparation without preamplification. *Nat. Methods* 14:167–73

157. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, et al. 2017. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8:14049
158. Zhu C, Gao Y, Guo H, Xia B, Song J, et al. 2017. Single-cell 5-formylcytosine landscapes of mammalian early embryos and ESCs at single-base resolution. *Cell Stem Cell* 20:720–31
159. Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD. 2001. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques* 30:892–97
160. Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, et al. 2017. Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell* 65:631–43
161. Zilionis R, Nainys J, Veres A, Savova V, Zemmour D, et al. 2017. Single-cell barcoding and sequencing using droplet microfluidics. *Nat. Protoc.* 12:44–73
162. Zong C, Lu S, Chapman AR, Xie XS. 2012. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* 338:1622–26

Annual Review of Genomics and Human Genetics

Volume 19, 2018

# Contents

**Errata**

An online log of corrections to *Annual Review of Genomics and Human Genetics* articles may be found at http://www.annualreviews.org/errata/genom