ANNUAL REVIEWS

Annual Review of Genomics and Human Genetics

Applications of Single-Cell DNA Sequencing

Gilad D. Evrony,[1,*] Anjali Gupta Hinch,[2,*] and Chongyuan Luo[3,*]

[1]Center for Human Genetics and Genomics, Grossman School of Medicine, New York University, New York, NY 10016, USA; email: gilad.evrony@nyulangone.org

[2]Wellcome Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, United Kingdom; email: anjali@well.ox.ac.uk

[3]Department of Human Genetics, University of California, Los Angeles, California 90095, USA; email: cluo@mednet.ucla.edu

Annu. Rev. Genom. Hum. Genet. 2021. 22:171–97

First published as a Review in Advance on March 15, 2021

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

https://doi.org/10.1146/annurev-genom-111320-090436

*All authors contributed equally to this work

Keywords

single-cell DNA sequencing, single-cell genomics, lineage tracing, germ cells, organismal development, DNA modifications

Abstract

Over the past decade, genomic analyses of single cells—the fundamental units of life—have become possible. Single-cell DNA sequencing has shed light on biological questions that were previously inaccessible across diverse fields of research, including somatic mutagenesis, organismal development, genome function, and microbiology. Single-cell DNA sequencing also promises significant future biomedical and clinical impact, spanning oncology, fertility, and beyond. While single-cell approaches that profile RNA and protein have greatly expanded our understanding of cellular diversity, many fundamental questions in biology and important biomedical applications require analysis of the DNA of single cells. Here, we review the applications and biological questions for which single-cell DNA sequencing is uniquely suited or required. We include a discussion of the fields that will be impacted by single-cell DNA sequencing as the technology continues to advance.

INTRODUCTION

Single-cell DNA sequencing (scDNA-seq) encompasses a suite of technologies and approaches that interrogate DNA at the level of single cells. These technologies contrast with standard DNA sequencing, also known as bulk sequencing, which homogenizes the DNA content of usually thousands to millions of cells. Much like the voice of a single individual or a small number of individuals can be drowned out in a large crowd, genomic signals (variants, DNA modifications, or structural properties of DNA) that are present in only one or a small number of cells in a sample may be undetectable without interrogating single-cell genomes. The development of scDNA-seq methods has at every step been motivated by biological questions that seek to explore the cellular genomic diversity that would otherwise be missed by bulk sequencing.

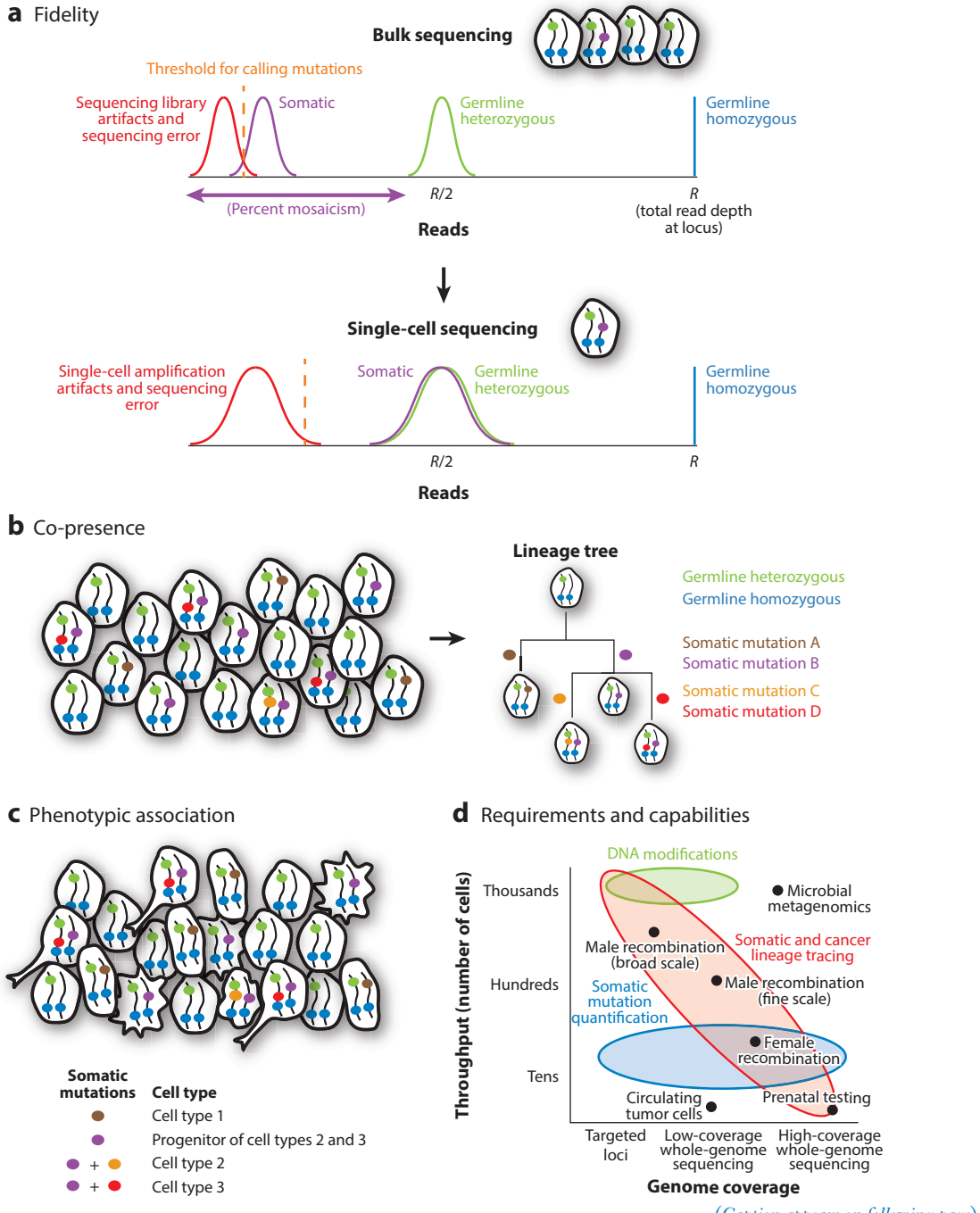
In this review, we focus on the major applications of scDNA-seq rather than its technological aspects. We begin by introducing the basic concepts of scDNA-seq to provide a framework for understanding why certain applications depend on its unique capabilities and why for some important biological and biomedical questions it is the only suitable technology. Following this, we discuss the major biological fields that scDNA-seq has impacted and the discoveries it has enabled. These include a wide array of fields: somatic mutation and mosaicism, organismal development, germ cell mutation and development, fertility, cancer, epigenetic regulation of the genome, genome organization, and microbiology. This review specifically focuses on single-cell genomics methods, i.e., those that profile single-cell DNA at larger genomic scales or genome-wide. Single-cell genomic approaches have only become feasible over the past decade; previously, it was only possible to profile a single locus or small numbers of loci in single cells. In this review, we use the term single-cell genomics interchangeably with scDNA-seq. While single-cell genomics is sometimes used in the literature to refer to single-cell RNA sequencing (scRNA-seq), to avoid terminological confusion, here, single-cell genomics and single-cell transcriptomics refer to scDNA-seq and scRNA-seq, respectively.

A CONCEPTUAL FRAMEWORK FOR SINGLE-CELL DNA SEQUENCING

In this section, we summarize key concepts necessary for understanding the subsequent sections that focus on the applications of scDNA-seq. These concepts not only are important to understand as a terminological and technical reference, but also help explain the unifying characteristics of the applications for which scDNA-seq is suitable. scDNA-seq unlocks a particular set of capabilities, described below, that are advantageous or unique relative to other technologies; therefore, every application of scDNA-seq is distinguished by its reliance on these capabilities. While this section explains this unifying framework, we do not review specific scDNA-seq technologies in detail, as these have been well reviewed previously (53, 68).

scDNA-seq is characterized by three core capabilities, which we term fidelity, co-presence, and phenotypic association, and which together help determine the types of biological questions and applications for which it is best suited (**Figure 1*a–c***). As discussed below, different scDNA-seq technologies may possess different subsets of these three capabilities, and moreover, none of these capabilities are unique to scDNA-seq. However, scDNA-seq is distinguished from other approaches by its potential to achieve all three capabilities on a genome-wide scale.

The fidelity capability of scDNA-seq (**Figure 1*a***) is its ability to overcome the limits of bulk DNA sequencing to detect features of DNA (mutations, modifications, or other properties of DNA) that are at a low level of mosaicism, i.e., features present in only a small subset of cells in a sample. Bulk DNA sequencing is limited by sequencing error that is a constant fraction of the total sequencing coverage. While increasing the depth of bulk sequencing coverage initially



(Caption appears on following page)

Figure 1 (*Figure appears on preceding page*)

The three core capabilities of single-cell DNA sequencing and application requirements of throughput versus genome coverage. (*a–c*) The three core capabilities of single-cell DNA sequencing. (*a*) Fidelity. In bulk sequencing, very-low-level mosaic mutations cannot be distinguished from sequencing library and sequencing error artifacts. By contrast, in single-cell sequencing, the distribution of read depths is the same for somatic mutations and germline (inherited) heterozygous mutations. This enables the detection of somatic mutations regardless of mosaicism, although it requires the same total sequencing costs as a hypothetical bulk sequencing method with perfect fidelity due to the need to sequence many single cells. Additionally, single-cell sequencing generally has an increased level of artifacts due to the single-cell amplification process. (*b*) Co-presence. Single-cell sequencing preserves information regarding which somatic mutations are present together in the same cells. This enables the reconstruction of lineage trees. (*c*) Phenotypic association. Single-cell sequencing, when combined with simultaneous cell phenotyping (either by microscopy or by multiomic single-cell sequencing), preserves information regarding which somatic mutations are present in which cell types. This enables the deconvolution of cell types or species in heterogenous samples and annotation of lineage trees with cell phenotypes. (*d*) Schematic of approximate application-specific requirements and current technological capabilities in terms of throughput (number of single cells per experiment or project) versus genome coverage. Panels *a–c* adapted from Reference 43.

increases sensitivity for lower-level mosaic DNA features by virtue of deeper sampling, the fact that sequencing error is a constant fraction of total coverage means that further increases in coverage eventually reach a hard limit: Mosaic features with less than ~0.5% mosaicism cannot be detected because they cannot be distinguished from sequencing error (18, 150). By contrast, scDNA-seq is not limited by sequencing error, since the sequencing error rate is much lower than the expected 50% signal level of heterozygous DNA features or even 25% single-stranded DNA features. Nevertheless, errors of single-cell DNA amplification can in some cases rival the signal level of true genetic features, motivating the ongoing development of new methods for high-fidelity single-cell genome amplification (53).

The co-presence capability of scDNA-seq (**Figure 1*b***) refers to its ability to ascertain which mosaic DNA features were present together in the same cells. For example, if a tissue sample contains two different low-level mosaic genetic variants, then only a single-cell method can determine whether both variants are present in the same cells or in mutually exclusive subsets of cells. Bulk methods homogenize samples prior to sequencing, so the information about which mosaic DNA features were present in the same cell or subsets of cells is lost. By contrast, scDNA-seq preserves this information. While there exist bulk methods that either directly profile large DNA fragments or preserve long-range phasing information that can technically ascertain co-presence for small numbers of DNA features, these methods have limited phasing distance and cannot ascertain the co-presence of DNA features on different chromosomes or the genome-wide co-presence of a large number of DNA features. Therefore, co-presence is a key capability of scDNA-seq that enables many of its unique applications.

The phenotypic association capability of scDNA-seq (**Figure 1*c***) is its potential to be combined with simultaneous single-cell phenotyping to identify the cell type(s) or cell state(s) in which specific DNA features are present. This capability derives from the fact that scDNA-seq analyzes single cells, thereby preserving the link between a cell's DNA features and its phenotype. However, phenotypic association requires technology that combines scDNA-seq with phenotypic profiling, such as single-cell DNA plus RNA profiling, single-cell DNA plus proteomic profiling, or single-cell DNA plus microscopy/histological profiling. Alternatively, some single-cell DNA features, such as methylation, themselves provide information about cell phenotype. The phenotypic association capability of scDNA-seq is technically challenging since combining genome-wide scDNA-seq with phenotypic profiling can compromise the resolution of the scDNA-seq component (108). However, new multiomic scDNA-seq technologies are emerging (31, 107) and are a focus of current technology development, given their importance for many of the applications we review below.

The co-presence and phenotypic association capabilities are together particularly important for samples composed of genetically distinct heterogeneous cells, a common situation that is often central to the applications of scDNA-seq. Such applications include both profiling cells within an organism, such as distinguishing tumor cells from normal cells in a tumor sample, and profiling mixtures of cells from distinct species, such as soil samples or the gut microbiome.

None of the above three capabilities are unique to scDNA-seq. For example, fluorescence in situ hybridization can ascertain the co-presence of a small number of genetic variants as well as their phenotypic association with specific cell types. Another example is single-cell cloning (180), where a single cell is expanded in vitro to create a bulk sample deriving entirely from one cell, thereby providing both fidelity and co-presence information for mosaic DNA variants that were present in the original single cell. Co-presence can also be assessed using scRNA-seq for the subset of somatic variants that are transcribed (114). The above are only a few examples of existing non-single-cell DNA sequencing methods that can achieve one or more of the three key capabilities of scDNA-seq. However, scDNA-seq is distinguished by the facts that it can potentially achieve all three capabilities simultaneously, it can determine co-presence at a genome-wide scale, and it can be applied directly to primary cells and is not limited to specific cell types. The last feature distinguishes it from single-cell cloning approaches that entail cloning and culture bottlenecks (which may bias results) and are not feasible for some cell types (e.g., mature adult neurons and unculturable bacteria). Every application of scDNA-seq can be traced back to one or more of these three core capabilities, and the most advanced applications leverage all three. It is therefore a useful exercise for any proposed scDNA-seq application or research project to critically evaluate which of these three capabilities—fidelity, co-presence, and phenotypic association—is needed and, consequently, whether scDNA-seq is required as opposed to employing an alternative technology. Conversely, consideration of these three capabilities can help identify new biological questions that can be uniquely addressed by scDNA-seq.

It is also worth noting that the fidelity capability of scDNA-seq is only relative to the fidelity of current DNA sequencers. If the fidelity of DNA sequencers were to advance by approximately four orders of magnitude, then scDNA-seq would be distinguished only by the co-presence and phenotypic association capabilities.

A major limitation of scDNA-seq derives from another limitation of current DNA sequencers: Only a small fraction of the DNA that is input into a sequencer is captured and sequenced. Therefore, scDNA-seq requires amplification of DNA prior to sequencing. And because there is no perfect amplification method, this step can introduce errors that confound analyses (53). Currently, there are several different single-cell DNA amplification methods, each with a different error profile and performance characteristics, which have been reviewed previously (53, 68). The best single-cell DNA amplification method depends on the application, and ideally, the error mode of the amplification method will be orthogonal to the DNA feature(s) of interest.

Finally, throughput—the number of single cells that can be profiled in one experiment—is an important parameter of different scDNA-seq methods. For some applications, tens or hundreds of single cells can suffice, while other applications (and most future applications) will require thousands of cells. New methods utilizing droplet encapsulation (120), combinatorial indexing (177), and nanowell devices (83) are increasing the throughput of scDNA-seq to thousands of cells and expanding the range and depth of questions that can be addressed. However, current scDNA-seq methods and sequencing costs impose a trade-off: Methods with higher throughput have lower genome-wide coverage or profile fewer DNA loci per cell. The applications discussed in this review differ widely in their throughput and coverage requirements (**Figure 1*d***), motivating careful matching between an application's requirements and the capabilities of the chosen scDNA-seq method.

SOMATIC CELLS: MUTATIONAL PROCESSES AND MOSAICISM

Somatic mutations—mutations that occur postzygotically in somatic cells—are ubiquitous in multicellular organisms due to exogenous and endogenous mutagens, DNA replication errors, chromosomal damage, imperfect DNA repair, and other mutagenic processes (159, 186). Indeed, it can be safely assumed that no two cells in an organism have an identical genome sequence, particularly for organisms whose genomes are significantly larger than the inverse of their per-cell division mutation rate. While most somatic mutations are benign, they are the cause of cancer and many noncancer genetic diseases, and they are speculated to contribute to physiological aging (39, 112, 135, 159, 163, 186). Quantifying the rates and spectra of somatic mutations is therefore of fundamental importance. However, within an individual organism, most somatic mutations are either present at very low mosaicism or unique to one cell (163), placing them below the detection limits of current DNA sequencers. The fidelity capability of scDNA-seq allows the detection of somatic mutations regardless of mosaicism level and has provided some of the first direct, genome-wide measurements of somatic mutation rates. Here, we review the application of scDNA-seq in quantifying somatic mutational processes and mosaicism. scDNA-seq of germ cell mutations is reviewed below (see the section titled Germ Cells: Meiotic Recombination and Germline Mutation). For a more in-depth discussion of somatic mosaicism and single-cell sequencing, we also recommend a review by Miller et al. (121) elsewhere in this volume.

Somatic mutation rates vary across tissues, developmental time points, cell types, and mutation types, and scDNA-seq studies have begun to systematically quantify somatic mutations across these dimensions. The first scDNA-seq studies quantifying genome-wide somatic mutation in normal (noncancer) cells profiled LINE-1 (L1) retrotransposon insertions and copy number variants (CNVs) in the human brain (41, 115). Both studies were motivated by long-standing hypotheses that somatic variation may be prevalent and adaptive in the brain (116). These studies successfully identified somatic L1 insertions and CNVs in human cortical neurons, but at lower rates than prior indirect measurements had anticipated: There were less than 0.6 somatic L1 insertions per neuron, less than 2% of neurons exhibited aneuploidy, and 41% of neurons harbored a few (mostly one or two) megabase-scale CNVs (41, 115). Several subsequent scDNA-seq studies were concordant with these findings (21, 40, 43, 79), suggesting that these types of somatic mutations are largely tolerated physiologically while creating a low level of neuronal genomic diversity and in rare cases causing neurological disease (32, 135).

In contrast to these studies of CNVs in mature cells, remarkably high rates of aneuploidy and CNVs have been found in early preimplantation human embryos by single-cell microarray profiling and in macaque embryos by scDNA-seq (29, 161). scDNA-seq of preimplantation macaque embryos found that 74% had at least one blastomere with a whole or partial chromosome copy number abnormality and, remarkably, detected some of the missing chromosomal material in cellular fragments arising from aberrant multipolar divisions at the one- or two-cell stage (29).

The possibility of pathogenic mosaic CNVs in human pluripotent stem cells used for cellular therapies has motivated scDNA-seq studies of these cells. One scDNA-seq study found that 23% of these cells had at least one large megabase-scale CNV, one-third of which had more complex karyotypes. Broader application of scDNA-seq for routine safety assessment and screening of cellular and gene-modifying/editing therapies can be anticipated.

In addition to measuring somatic L1, CNVs, and aneuploidy rates, scDNA-seq studies have quantified single-nucleotide variant (SNV) somatic mutation rates. scDNA-seq SNV studies face a greater analytic challenge separating single-cell amplification artifacts from bona fide somatic variants, and have recently greatly benefited from computational methods utilizing local spatial variation and read-based phasing of candidate variants to germline variants in order to filter artifacts

(16, 106). New amplification methods, such as primary template–directed amplification, that further reduce amplification artifacts will also play an important role in accelerating studies of SNV somatic mutation rates (55). The first genome-wide scDNA-seq SNV study of normal (noncancer) cells examined human cortical neurons and found much higher rates of SNV mutation relative to other mutation types and a mutational strand bias, suggesting that some of these mutations originate from transcription-coupled damage (98). A subsequent study refined these estimates across a larger set of neurons and age groups to quantify 300–900 somatic SNVs per neuron at birth and ~20 new somatic SNVs per year of life, leading to more than triple the number of SNVs (~2,000–3,000) by the time an individual is more than 80 years old (97). Interestingly, these aging mutation rates were higher in hippocampal neurons and in individuals with DNA-repair syndromes (97).

Additional scDNA-seq studies have measured somatic SNV rates of ~900 per fibroblast from a human toddler; ~460 per B lymphocyte in newborns, increasing to ~3,000 in centenarians; and ~1,000 per hepatocyte in newborns, increasing to ~4,000–5,000 in elderly individuals (17, 35, 185). Notably, all of the above studies have found that most SNVs are unique to one cell in the sample and can be computationally assigned to an aging clocklike mutational process (whose mechanism is enigmatic) previously identified in cancer sequencing studies [Catalogue of Somatic Mutations in Cancer (COSMIC) signatures 1 and 5] (3). Altogether, these studies reveal a relatively similar burden of somatic SNVs per cell at birth across cell types and a significant increase with age, albeit at different rates in different cell types. The possible functional impact of accumulating somatic SNVs with age throughout the body is an important open question.

Over the long term, a full view of somatic mutation rates will necessitate scDNA-seq studies spanning numerous developmental time points, cell types, tissues, and mutation types. Most scDNA-seq studies have also focused only on one or a few cell types among hundreds of cell types in the body. Bulk methods for measuring somatic mutation rates have begun to fill this gap, such as clonal expansion of primary single cells followed by bulk sequencing (5), deep sequencing of tissues where clones remain spatially restricted (20), and RNA-seq (178), with findings that are largely consistent with those from scDNA-seq. However, clonal expansion of primary cells is susceptible to culture bottlenecks and cannot be applied to many cell types, deep sequencing of spatial clones is not applicable to many tissues and cannot resolve different cell types, and bulk RNA-seq cannot easily measure cell type–specific mutation rates. Therefore, going forward, the unique advantage of scDNA-seq relative to other methods for somatic mutation quantification will be via multiomic scDNA-seq methods (i.e., utilizing scDNA-seq's phenotypic association capability) to quantify mutation burdens of specific cell types from any tissue.

ORGANISMAL DEVELOPMENT AND LINEAGE TRACING

One of the long-standing goals of biology has been to understand how a single cell (the zygote) transforms into the billions or trillions of cells of a complex multicellular organism (1, 118). The development of an organism can be represented by a lineage tree—a diagram of the cell divisions from the zygote to each of its cells, ideally annotated with spatial data, cell phenotypes, and the molecular determinants of lineage decisions (76). Lineage trees are the blueprints of development, depicting the relationships between progenitor cells and cell types throughout development and providing a framework for understanding stem cell and cell type hierarchies, aberrant lineages in disease, and tissue homeostasis. Lineage tracing—the collection of methods used to determine lineage trees—is therefore among the most fundamental tools of biology (81). In this section, we review the application of scDNA-seq for lineage tracing with a focus on organismal development. scDNA-seq lineage tracing of cancer is reviewed in the section titled Cancer.

Lineage tracing methods are either prospective or retrospective, and scDNA-seq can be used for either approach (7, 173). In prospective lineage tracing, cells are experimentally tagged by a heritable marker (e.g., a genetically encoded barcode), and their progeny are profiled at a later time point. Because prospective lineage tracing requires invasive manipulation of the organism, it is only applicable to animal models. In retrospective lineage tracing, lineage trees are reconstructed from somatic DNA mutations that occurred spontaneously during an organism's lifetime, so it is noninvasive and feasible for studying human development. Regardless of whether DNA mutations are experimentally induced (the prospective approach) or occur spontaneously (the retrospective approach), scDNA-seq can detect these mutations to construct a matrix showing which mutations were present in which single cells (note that this is another way of describing scDNA-seq's co-presence capability). This mutation-by-cell matrix is then transformed into a lineage tree through phylogenetic reconstruction algorithms (118).

Although scDNA-seq can be used for both prospective and retrospective lineage tracing, its greatest long-term potential lies in retrospective lineage tracing. Recently, prospective methods of lineage tracing have been developed in which DNA mutations are created using CRISPR/Cas9 specifically in sites that are transcribed into RNA, such as 3′ untranslated regions (2, 26, 48, 137, 151, 171). This allows DNA mutations to be detected by scRNA-seq, in conjunction with phenotypic profiles, without requiring scDNA-seq. These methods, which have been reviewed elsewhere (7, 76, 118, 164), have enabled high-throughput phenotypic lineage tracing for the first time and have removed the requirement for scDNA-seq in prospective lineage tracing of animal models. By contrast, scDNA-seq is essential for retrospective lineage tracing—the only method that can be applied to humans. This is because most spontaneous somatic mutations, which retrospective lineage tracing relies on, are in the large majority of the genome that is not transcribed and cannot be captured by scRNA-seq. We therefore focus this section on the use of scDNA-seq for retrospective lineage tracing. scRNA-seq-based prospective lineage tracing methods are also currently limited by their inability to induce mutations continuously across broad developmental time points (118). scDNA-seq retrospective lineage tracing, however, leverages somatic mutations that occur in every cell division that can theoretically be used to reconstruct the lineage of any tissue and any developmental time point at the maximum possible resolution (1, 49). Once scDNA-seq retrospective lineage tracing scales to the throughput of the above scRNA-seq prospective lineage tracing methods and is combined with phenotypic (e.g., RNA or protein) profiling—though this may take many years to realize—one can expect that it will become the primary lineage tracing approach for both humans and animal models.

The resolution of retrospective lineage tracing scales with both the number of loci that are genotyped and their somatic mutation rate (14). Microsatellites are abundant genomic elements (>1 million loci in the human genome) consisting of repeated 1–6-base-pair motifs that have the highest mutation rates of any type of genomic element (49). This makes them an attractive target for retrospective lineage tracing (49, 145). In fact, it has been estimated that every cell division may be tagged by at least one microsatellite mutation, theoretically allowing the reconstruction of an organism's complete lineage tree (49). Early pioneering scDNA-seq studies of microsatellite mutations utilized only a small set of less than 100 loci but were able to confirm some basic lineage information, such as clustering liver cells separately from other organs' cell types (144) and closer lineage relationships of muscle satellite cells that are physically proximal (167). Another study of ∼120 microsatellite loci found evidence for periodic replacement of all the cells in a colon crypt with the progeny of a single stem cell, a process termed monoclonal conversion (140).

The number of microsatellite loci profiled per cell has subsequently increased to achieve greater lineage resolution, with ∼1,000–2,000 loci per cell in two different studies (14, 170)

and ~12,000 loci in a third study (157). One of these studies was able to simultaneously profile methylation in the same cells and recovered a known cultured lineage tree with 88% accuracy (170). However, there remain significant challenges associated with artifactual microsatellite mutations that occur during single-cell genome amplification and library preparation (14). These will need to be overcome before the full potential of microsatellites for scDNA-seq lineage tracing can be realized.

Several scDNA-seq studies have used other types of somatic mutations for retrospective lineage tracing, such as SNVs, which are also estimated to occur in every cell division (1), and L1 retrotransposon insertions. Nearly all these studies have so far focused on the human brain. One study traced the spatial distribution of two somatic L1 retrotransposon insertions identified by scDNA-seq in the brain of an individual, finding both a focally restricted lineage in the frontal lobe and a widely distributed lineage consistent with spatial mixing of progenitor cells during early development (42). The L1 insertion of the focal lineage also harbored a poly-A microsatellite that itself mutated somatically at a high rate to mark different subclades in each brain region, consistent with tangential migration and intermingling of radial clones of progenitors in the cortex. A notable study profiled 18 somatic mutations, including 15 SNVs, to reconstruct a lineage tree of 136 single neurons with four clades and additional subclades (98). Remarkably, the four root clades in this individual were dispersed across the cortex at low mosaicism, again indicating significant spatial mixing of early brain progenitors, whereas later sublineages had more focal distributions. The finding of widely dispersed and interspersed early embryonic clades in the human brain was confirmed by another study of somatic SNVs that were initially identified by deep bulk whole-genome sequencing and then profiled in single cells (18). Another recent study performed simultaneous scDNA-seq and RNA profiling of more than 1,000 cortical neurons and found that SNVs of later sublineages were progressively restricted to either excitatory neurons or inhibitory neurons, consistent with these cell types' different developmental origins (67). Upper-layer neurons also made up a greater fraction of later lineage branches, consistent with the inside-out formation of cortical layers. These studies have confirmed processes of brain development previously seen only in animal models using prospective (invasive) approaches. The theme of these studies is that as their throughput has increased, more detailed features of brain development have emerged. Further such studies may eventually reveal human-specific developmental processes.

Because somatic mutations accumulate in every cell division, scDNA-seq retrospective lineage tracing can also be used to estimate for any cell or tissue its approximate number of cell divisions since the zygote, also known as the lineage depth. This can help elucidate organogenesis and tissue turnover. In mismatch-repair-deficient mice with higher microsatellite mutation rates, one scDNA-seq study estimated a depth of 29 for oocytes; a depth of 24–40 for various types of adult stem cells, consistent with their relative quiescence; and a linear increase in depth of ~1 cell division per day for B cells (168). This method could be used to investigate stem cell activity in other organs (such as the brain) and in disease states.

An important recent advance has been scDNA-seq retrospective lineage tracing that uses mitochondrial DNA (mtDNA) instead of nuclear DNA mutations (85, 101, 175). The small size of the mitochondrial genome and the presence of hundreds of copies (or more) per cell enable this approach to scale to thousands of cells via droplet-based methods (85). mtDNA is also readily captured in standard scDNA-seq chromatin accessibility assays, such that mtDNA mutations and cell phenotype information are captured simultaneously from the same cells (85, 101, 175). The throughput of this approach is illustrated by one experiment that traced the lineage relationship of more than 16,000 $CD34^+$ hematopoietic stem cells sampled at one point and peripheral blood mononuclear cells sampled three months later from the same individual (85). The authors

identified more than 250 clonal groups, indicating that a large pool of hematopoietic stem cells is active in healthy blood production, but they also observed variability in the contributions of individual hematopoietic stem cell clones to blood cell production (85). This method has also been used to lineage trace blood and colorectal cancers (85, 101, 175).

Overall, the main advantages of single-cell mtDNA lineage tracing are its significant throughput in terms of cell number and the fact that it does not require a priori knowledge of subclonal somatic mutations. This contrasts with nuclear scDNA-seq lineage tracing, where sequencing the entire genome across thousands of cells is not feasible. However, the advantages of mtDNA lineage tracing are balanced by the limited number of mtDNA mutations that occur in the small mitochondrial genome, even accounting for mtDNA's higher mutation rate (1). In the long run, this will limit the resolution of mtDNA lineage trees relative to nuclear genome lineage trees. Additionally, mtDNA is not a completely faithful lineage marker, because mutations in only a small subset of a cell's mitochondria (heteroplasmy) may not always segregate to both daughter cells. Nevertheless, the biological insight provided by scDNA-seq mtDNA is promising due to its relative ease of integration into established single-cell workflows, and it provides a glimpse into the future of what high-resolution scDNA-seq lineage tracing may achieve. By virtue of its ability to ascertain somatic DNA mutations genome-wide, scDNA-seq—combined with cell phenotyping—has the potential to provide the first systematic and truly general approach for high-resolution lineage tracing.

GERM CELLS: MEIOTIC RECOMBINATION AND GERMLINE MUTATION

Meiosis is a specialized cell division in which haploid gametes are generated from diploid cells through one round of genome replication followed by two rounds of cell division (110). During this process, mutation generates new alleles and recombination generates new combinations of alleles via exchanges of genetic material between chromosomes known as crossovers (110). Together, they create heritable genetic variation and the substrate for evolution via selection. Recombination is mechanistically essential to success in meiosis: It is initiated by programmed DNA double-strand breaks, of which only a fraction are repaired with a crossover (71). Insufficient crossovers or their improper localization leads to aneuploidy, which affects 20–30% of human eggs and 1–8% of human sperm, leading to miscarriage in more than 25% of pregnancies and severe developmental defects in 0.3% of live births (22, 58, 59). Errors in recombination also lead to numerous genetic disorders (23). In this section, we discuss the application of scDNA-seq for understanding meiotic recombination and germ cell mutation.

The most high-resolution approaches to date for studying recombination have been mapping of crossovers in pedigrees (10, 57, 80, 91) or admixed populations (60, 169) and inference of historical recombination via breakdown of linkage disequilibrium (so-called LD-based maps) (38, 47). However, these approaches have significant limitations:

1. Individual- versus population-level maps: Because recombination is a germline process with a strong impact on fitness, it is important to measure it at the level of gametes or in individuals who are unable to reproduce, because pedigree-based and LD-based methods only probe viable offspring. Separately, the fine-scale landscape of recombination in humans and many vertebrates is determined predominantly by the protein PRDM9, whose DNA-binding properties may vary across individuals in the population (124, 130). Interpretation of recombination maps aggregated across individuals is therefore complicated by *PRDM9* allelic diversity in the population.

2. Sample availability: Identification of crossovers at high resolution in pedigrees requires dense genotyping or whole-genome sequencing in large numbers of related individuals (57, 80, 94, 146). Such resources are challenging and expensive to accumulate in humans and model organisms, and this approach is not typically feasible for species in the wild.
3. Gender differences: Recombination-related phenotypes show strong sexual dimorphism in humans and many other species (19, 80, 104). However, understanding of female recombination has lagged behind understanding of male recombination, in part due to the challenges of working with limited amounts of tissue that are difficult to obtain. As discussed below, scDNA-seq of germ cells and gametes has overcome many of these limitations.

scDNA-seq studies of sperm demonstrate that human sperm have high rates of genome instability and aneuploidy (12, 100, 165), in striking contrast to mouse sperm (61). Analysis of tens of thousands of single sperm from multiple human males revealed significant interindividual differences in aneuploidy rates, which ranged from 1% to 5%, with an apparent excess of whole-chromosome losses over gains (12). These data revealed that whole-chromosome gains may result from improper segregation at either of the two meiotic divisions, with a twofold-higher error rate in the second meiotic division for the autosomes. The number and localization of crossovers covary significantly among cells (12), and a larger number of crossovers within a cell appears to be protective against chromosome mis-segregation (12, 100). Further insight into male meiotic segregation was provided by scDNA-seq of spermatocytes (precursors of sperm), which showed an unexpected class of meiotic segregation error in an azoospermic mouse (177). Single-sperm sequencing in mouse has also provided novel insights into the dynamic processes underlying the pairing of homologous chromosomes and the determination of which DNA breaks become crossovers: A high-resolution map of crossovers in mouse sperm showed that DNA breaks that repair more quickly are also more likely to become crossovers (61). scDNA-seq may also enable individualized quantification of sperm mutation rates and the risk of transmitting de novo mutations to offspring (165); however, further technological improvements in throughput, coverage, and accuracy will be required.

Acquisition of oocytes involves a surgical procedure, and only small numbers of oocytes (<10) are typically harvested per individual in humans and other mammals (63, 131). Limited tissue availability, together with the fact that all the products of a single meiosis can be harvested together (oocyte and polar bodies), makes scDNA-seq of oocytes a powerful avenue for studying female meiosis. Single-cell data have confirmed that crossovers exhibit interference (they are farther away from each other than expected by chance) in human oocytes (63, 131), which they also do in sperm (100, 165). Human oocytes have very high aneuploidy rates (18–70%) (63, 131), and in contrast to spermatocytes, they have significantly higher error rates in the first meiotic division. Genome comparisons of the oocyte with the first and second polar bodies led to the identification of a new mode of reverse segregation of chromatids, which was observed to be the dominant signature of aneuploidy (131). This study also showed that crossing over is protective against aneuploidy, with further meiotic drive against nonrecombinant chromatids. Understanding the mechanisms of aneuploidy further will require phasing of missegregated chromosomes (131), which remains a challenge with the relatively low genome coverage obtained per cell. A major focus of recent single-cell research has been to maximize throughput and aggregate small amounts of information per cell across large numbers of cells (12, 177). Generating insights into female recombination, however, will require a parallel emphasis on maximizing the amount of information obtained per cell, e.g., by increasing the genome coverage per cell (61) and performing long-range haplotyping of individual molecules (154).

FERTILITY

A major emerging clinical application of scDNA-seq is in prenatal screening and genetic testing in natural conceptions and in vitro fertilization. During assisted reproduction, multiple embryos are typically created by in vitro fertilization and screened for aneuploidy and highly deleterious mutations, followed by implantation of selected embryos. Single-cell genomic approaches have improved on conventional methods of analyzing embryo biopsies by enabling simultaneous detection of aneuploidy and mutations genome-wide (82, 160, 172). Sequencing of polar bodies can also be used to detect maternally inherited aberrations in oocytes (63). These approaches are adding insight into the nature of aneuploidy and mosaicism in embryos, which may be mitotic or meiotic in origin (117, 149). In natural conceptions, invasive testing carries the risk of procedure-induced miscarriage, and noninvasive sequencing of cell-free fetal DNA in maternal plasma has been rapidly adopted in clinical practice (73, 162). However, the reliability of this method is variable, and recent developments in the extraction of genetic information from fetal cells circulating in maternal blood may lead to future single-cell genomic tools for this application (66, 147). The remaining challenges are reliable isolation of rare fetal cells from maternal blood and concern that the predominant and more readily isolated fetal trophoblast cells are not always representative of the genome of the fetus (25).

CANCER

Cancer is a disease of the genome in which an aberrant lineage of cells accumulates somatic mutations along an evolutionary trajectory of uncontrolled growth. scDNA-seq has revolutionized our ability to study the dynamic evolutionary processes by which tumor lineages evolve and interact with selective pressures during oncogenesis, tumor growth, and treatment (8, 127, 141). This section reviews the most important applications of scDNA-seq in cancer research: intratumoral heterogeneity (ITH), clonal evolution, invasion and metastasis, circulating tumor cells (CTCs), and therapeutic response. We also highlight emerging approaches and clinical applications.

Tumors evolve genetically heterogenous lineages over time, and the resulting ITH plays a significant role in tumor growth, metastasis, and treatment response (28). A tumor's ITH, defined by the subclonal structure of its lineages and driver mutations, is its dramatis personae and of prime importance for understanding its biology. The first study to examine ITH with genome-wide scDNA-seq utilized copy number aberrations (CNAs) to distinguish subclonal lineages within breast tumors (128). Since then, numerous scDNA-seq studies of CNAs and point mutations have cataloged ITH across diverse hematological (6, 52, 65, 69), breast (9, 51, 166), ovarian (119), colorectal (13, 88, 96), renal (176), bladder (92), lung (44), liver (64), and brain (46) cancers. These studies have found that most tumors contain more than one major subclonal lineage, in addition to many tumors that harbor a larger number of lower-frequency sublineages. In some cases, the number of major subclonal lineages (subclonal diversity) identified by scDNA-seq is associated with the tumor subtype. For example, there is greater subclonal diversity in estrogen receptor–negative than in estrogen receptor–positive breast cancers (9) and greater subclonal diversity in acute myeloid leukemias (AML) harboring *FLT3* mutations than in non-*FLT3* AMLs (120). In one study, higher subclonal diversity (four or more subclones) correlated with a worse prognosis (122), suggesting that subclonal diversity may be a useful prognostic biomarker.

While ITH can be inferred from bulk DNA sequencing data (182), scDNA-seq provides a more comprehensive and higher-resolution view of ITH and clonal evolution (8). Subclonal structures derived from bulk DNA and scDNA-seq are largely concordant, but the latter often reveals subclones not found in bulk data (9, 69, 166). scDNA-seq can also profile ITH when sample size is limiting, such as in fine needle aspiration biopsies (83) or when tumor cells are infrequent in the

sample (i.e., low tumor purity). Nevertheless, both bulk and single-cell methods are usually implemented in integrated analyses due to the lower cost and currently higher genome-wide quality of bulk sequencing, especially for broader multiregion and multi-time-point profiling.

Phylogenetic analyses of ITH profiles obtained by scDNA-seq reveal the order in which specific driver mutations and types of mutations occur, which in turn provides important information about the earliest events and mutational processes that initiate cancer. For example, several scDNA-seq studies have shown that in some breast cancers, colorectal cancers, and acute lymphoblastic leukemias, large numbers of structural variants (CNAs and aneuploidies) occur over short time periods early during tumor evolution but infrequently later (termed punctuated clonal evolution), while point mutations accumulate continuously along with a small number of later-occurring focal CNAs involving specific driver genes (51, 52, 96, 128, 141, 166). This suggests that genome-wide, high-impact mutational events underlie the origins of some cases of these cancer types rather than a gradual accumulation of mutations. Indeed, the mechanism of one such mutational process, chromothripsis, has been studied in vitro using scDNA-seq (181). In a large scDNA-seq study of AML, mutations in the epigenetic modifier genes *DNMT3A* and *IDH1/2* were identified as the most common initiating events, while mutations in signaling genes tended to occur later (120). scDNA-seq has also quantified a statistically significant co-occurrence of driver mutations at the single-cell level [e.g., *NPM1* and *FLT3* mutations in AML (122)] as well as marked copy number heterogeneity of focal amplifications within individual tumors: Cells within one brain tumor varied from less than 20 to more than 100 copies of *EGFR* (46).

Interestingly, some of the above studies identified within the tumor sample near-normal cells that were unrelated to the tumor lineage but nevertheless contained a small number of CNAs or only one driver mutation (e.g., *APC*-only mutant cells) (9, 51, 88). This likely reflects baseline CNAs and driver gene abnormalities at low levels in normal cells (9, 51, 88). The relationship of such cells to the origins of cancer is presently unknown and is an important topic for further research. scDNA-seq may eventually allow the identification of the cells of origin of individual tumors and help resolve long-standing debates regarding the existence and role of stem cell hierarchies in tumors (8, 127). Other mechanisms of tumor evolution may be revealed in tumor types that have not yet been studied using scDNA-seq. Systematic application of scDNA-seq to all cancer types—a single-cell DNA Cancer Genome Atlas—in addition to other single-cell omics programs (143) will be needed to achieve a full understanding of cancer evolution.

Much of the mortality of cancer is due to metastatic spread beyond the primary tumor site. scDNA-seq has been used to study the lineage relationships between metastases and the primary tumor in the hopes of finding ways to intercept this lethal process. One scDNA-seq study of ovarian cancer found that most intraperitoneal sites of tumor spread are either monoclonal or derive from a single tumor clade (119). This indicates that once an intraperitoneal metastatic site is established, it is usually not reseeded by new cancer cells. The authors also found one patient with two subclones that were present together in multiple intraperitoneal sites with correlated frequencies, possibly due to co-migration of the two subclones to multiple locations. The mostly monoclonal spread seen in the prior study contrasts with an scDNA-seq study of invasive ductal breast carcinoma that found that most tumor subclones evolve within the breast duct, followed by multiclonal invasion into the surrounding tissue (24). In a study of colorectal cancer, one patient had monoclonal metastatic seeding of the liver, while another patient had polyclonal seeding (88). These divergent findings in different tumor types make clear that the mechanisms and bottlenecks of clonal spread delineated by scDNA-seq vary across tumor types and anatomic environments.

scDNA-seq of CTCs is a promising approach to noninvasively sample tumors (8). CTCs are especially appealing in relation to more widely used cell-free DNA approaches, because they allow examination of complete tumor cell genomes. CTCs have been successfully isolated and

sequenced from diverse cancers, including prostate cancer, where they captured early clonal mutations shared by primary and metastatic sites (99), and lung cancer, where mutations were identified in CTCs that could inform treatment (129). CTCs have also been used to noninvasively monitor treatment response (27). However, CTCs are very rare in blood and require both specific tumor cell surface markers, which are not currently available for most tumor types, and a sensitive and specific isolation procedure. In a study of 51 blood samples from 36 prostate cancer patients, 27% of samples had no detectable CTCs, and the remaining samples had a median of only seven CTCs (99). This poses a challenge for CTC profiling to capture the full clonal heterogeneity of a tumor to reliably inform clinical decision-making. Progress will depend on new methods to purify CTCs from large blood volumes and validation of new markers for CTC isolation. These investments will be worthwhile, as they may enable presymptomatic detection of tumors, noninvasive treatment monitoring, and early detection of recurrence.

scDNA-seq studies have examined how ITH changes during clinical treatment and how tumors evolve treatment resistance. One study showed that treatment-resistant clones in breast cancer were detectable prior to treatment (77). Another used multiomic scDNA-seq plus immune profiling to observe the phenotypic changes that mutant blast cells underwent following treatment with a drug that induces differentiation to a mostly erythroid phenotype (31). This study highlights the potential of scDNA-seq to elucidate how specific lineages with different genotypes can respond differently to the same treatment. If applied broadly, this approach will provide a critical reference for translating subclonal genotypes to predictions of treatment response. However, two limitations of scDNA-seq for treatment monitoring are noteworthy. First, tumor sampling during treatment is usually only possible in discrete, infrequent time points (though CTC profiling may mitigate this). Second, genetics is not the sole driver of ITH or treatment resistance; epigenetic heterogeneity also plays an important role in tumor cell biology (77, 126) (see the section titled DNA Modifications).

While scDNA-seq studies of cancer have already been impactful, new scDNA-seq technologies are emerging with multiomic, spatial, and high-throughput capabilities that will transform cancer research. Multiomic scDNA-seq can build genotype–phenotype correlations to understand how specific subclonal genotypes associate with cellular phenotypes, invasiveness, treatment response, and more. One such method, scTrio-seq, which simultaneously profiles CNAs, methylation, and transcriptomes, found that subclonal lineages defined by CNAs and methylation are concordant, and it further correlated methylation patterns with gene expression (13). Another recent method combined scDNA-seq with oligonucleotide-tagged antibodies to simultaneously profile surface protein expression in order to correlate genotypic and phenotypic ITH in AML (31, 120). Mitochondrial mutations have also been combined with assay for transposase-accessible chromatin using sequencing (ATAC-seq) for phenotypic lineage tracing of leukemia and colorectal cancer (85). Many new multiomic methods continue to be developed (126). scDNA-seq is not only being combined with other molecular modalities but also being applied in ways that preserve spatial information. One study combined laser microdissection with scDNA-seq to show how ITH varies between preinvasive and invasive regions in breast cancer (24). In situ/spatial genotyping or sequencing of large numbers of somatic mutations directly in tissue sections (i.e., without microdissection) would revolutionize cancer research; however, although small numbers of mitochondrial and expressed variants can be readily genotyped because they are present in multiple copies per cell, in situ detection of single-copy variants, especially SNVs, while feasible, is significantly more challenging and cannot currently be scaled to large numbers of variants (183). Finally, the throughput of scDNA-seq is rapidly increasing, an essential step toward achieving high-resolution views of ITH and tumor evolution and for cost-effective profiling of large cohorts. Nanowell- and droplet-based methods have scaled scDNA-seq to thousands of

cells per sample, with a concomitant increase in ITH resolution (83, 120). This scale finally brings scDNA-seq to the throughput regime of scRNA-seq and shifts the bottleneck to sequencing costs, which currently limit these methods to low genome coverage or targeted loci sequencing.

Translation of scDNA-seq to the clinic is a prime goal for delivering on the promise of cancer genomics to improve patient outcomes (8, 127). Three important steps will help make this a reality. First, clinical trials will need to rigorously measure the potential benefits of scDNA-seq profiles of ITH and clonal evolution for assigning patients to personalized or targeted treatment regimens and for assessing treatment response. Second, the current costs and technical challenges associated with scDNA-seq will need to be reduced and simplified to deliver high-throughput and user-friendly workflows that can be implemented reproducibly in a clinical laboratory. Third, while scDNA-seq CNA profiling of formalin-fixed paraffin-embedded (FFPE) samples has been achieved (111), methods for profiling single-cell genomes at high resolution from these samples are needed, as FFPE is the main way that clinical samples are processed. These and additional challenges remain, but the potentially life-saving benefits of scDNA-seq for cancer patients make these worthy goals.

DNA MODIFICATIONS

A sizable fraction of the human genome contains regulatory information that dictates the spatial and temporal expression of genes (37). DNA modifications provide an additional means to precisely control gene expression by modulating the biochemical interaction between genomic DNA and transcriptional machinery (102). The most common mammalian DNA modifications are 5′-methylcytosine (5mC) and its oxidized forms 5′-hydroxymethylcytosine, 5′-formylcytosine, and 5′-carboxylcytosine. Although earlier literature focused on studying 5mC dynamics in CpG dinucleotide islands (gene-regulatory regions with an increased density of 5mC-marked CpG nucleotides), the development of whole-genome bisulfite sequencing and its application to diverse human tissues has revealed pervasive 5mC dynamics in genomic regions outside CpG dinucleotide islands (148, 189). For example, local 5mC depletion is a reliable signature for enhancers and other types of regulatory elements, and whole-genome bisulfite sequencing has uncovered other types of 5mC features, such as partially methylated domains, DNA methylation valleys and canyons, and abundant 5mC in non-CpG contexts (56, 102).

Cell type–specific patterns of 5mC have historically been achieved by isolating bulk populations of specific cell types using fluorescence-activated sorting. However, this approach is limited by the availability of genetically encoded cell type reporters or antibodies used for purification in animal models and human tissues, respectively. These challenges motivated the development of techniques for single-cell profiling of DNA modifications. Currently, there are diverse single-cell cytosine methylome methods that have been well reviewed previously (75). A series of multiomic approaches have also been developed that combine single-cell transcriptome, chromatin accessibility, or chromatin conformation with 5mC profiling (187).

Most single-cell methylome methods produce sparse data that randomly cover 5–40% of the haploid genome. Notably, an important feature of methylome profiling is that coverage dropout can be unequivocally determined and is independent of the methylation state. The analysis of single-cell methylomes typically takes advantage of the local correlation of methylation patterns: The methylation levels of CpG sites within a CpG dinucleotide island region, or in non-CpG contexts in different parts of a gene body, are often correlated. This allows extraction of cell type–specific and dynamic methylation information from sparse genome-wide 5mC measurements and enables unbiased classification of cell types or transient cell populations. Typically, the quantification of 5mC in single cells over discrete features, such as enhancers, requires the reconstruction

of pseudo-bulk profiles derived from pooling the data from a larger number of single cells of each type. Several computational approaches have also been developed to impute missing values in single-cell methylomes and to perform clustering analysis to identify cell populations (30).

Several fields in which single-cell methylome profiling has been applied include embryonic development, brain development, hematopoiesis (70), and cancer—processes involving major 5mC dynamics. A series of studies have investigated the single-cell methylation dynamics of germ cell development (90), pre- and postimplantation embryos of mouse and human (188), and mouse gastrulation (4). The advantage of single-cell methods was demonstrated by the classification of single cells into distinct lineages, enabling the quantification of 5mC and chromatin accessibility at lineage specific enhancers (4).

The human brain contains a large diversity of functional regions and cell types that are established during brain development. Single-cell methylome profiling has provided a means to simultaneously categorize brain cell types and to identify cell type–specific regulatory elements (103). Two methods, single-nucleus methylcytosine sequencing (snmC-seq) and single-cell combinatorial indexing for methylation analysis (sci-MET), have been applied to mammalian brains (103, 105, 123). As part of the Brain Research Through Advancing Innovative Neurotechnologies (BRAIN) Initiative Cell Type Census Network (BICCN), single-cell methylomes are being systematically generated across the entire mouse brain, and the first report analyzed 100,000 cells from 45 regions of the anterior mouse brain (95). These and other studies of the human brain have demonstrated similar resolution of cell type classification using single-cell methylomes as achieved using single-cell transcriptomes (105). Interestingly, a single-cell methylome data set generated from finely dissected mouse brain revealed nuanced 5mC dynamics within cell types that correlates with the spatial organization of the brain (95).

Many cancer types are associated with genomic alterations and aberrant 5mC patterns (72, 126), providing an exciting opportunity to associate epigenomic heterogeneity with tumor lineages. In addition, the possibility of reconstructing a history of epigenetic alterations (or epimutations) has motivated single-cell methylome studies comparing primary and metastatic tumors (13) and CTCs (54). Single-cell methylome profiling of colorectal cancer has found consistent 5mC patterns in both primary and metastatic tumors within each tumor sublineage but different 5mC profiles across lineages (13). Combined single-cell methylation and transcriptome profiling of chronic lymphocytic leukemia cells found a consistently increased rate of 5mC alteration (epimutations) compared with normal B cells, and the tumor cell lineage reconstructed using 5mC changes accurately represented somatic mutation patterns (50). Note that in this section we have focused on single-cell profiling of DNA modifications; applications of single-cell chromatin accessibility have been well reviewed previously (75, 78).

GENOME ORGANIZATION AND CHROMATIN CONFORMATION

The folding of the 2-m-long human diploid genome into a nucleus measuring less than 10 μm in diameter is a complex, hierarchical process. The resulting three-dimensional organization of the genome, also known as chromatin conformation, plays an important role in genome function. In the interphase nuclei of eukaryotes, chromosomes occupy distinct spatial domains called chromosome territories. The development of proximity ligation–based chromatin conformation capture (3C) and the derived Hi-C methods can associate chromatin structures with DNA sequences to obtain a view of genome organization (93). These methods perform restriction digestion of cross-linked nuclei followed by ligation to capture DNA ends that are proximal in three-dimensional space (139). Hi-C methods have revealed a series of hierarchical chromatin domains such as topologically associating domains (TADs), which are hundreds of kilobases to megabases in length

and generally conserved among cell types (34). TADs often contain finer-scale structures such as sub-TADs, domains, or individual loops, which, in contrast to the relative invariance of TADs, can significantly differ among cell types and across development, suggesting that chromatin conformation dynamics is a critical component of gene regulation in normal and diseased tissues (33).

Single-cell profiling of chromatin conformation has generated novel insights into the process of chromatin organization (153, 156). Single-cell studies of chromatin conformation dynamics during the cell cycle and the oocyte-to-zygote transition have revealed that TADs are highly heterogeneous among individual cells and that TADs previously observed in bulk samples arise via chromatin contacts that are constrained but nevertheless stochastic at the single-cell level (45, 125). Another exciting application of single-cell chromatin conformation profiling is to connect distal enhancers to their regulatory target genes in specific cell types. Genome-wide association studies (GWASs) have found a strong enrichment of disease-associated genetic variants in enhancer elements (113). Cell type–specific chromatin loop information therefore provides a valuable resource to mechanistically dissect disease risk loci by connecting noncoding variants to disease risk genes. Single-cell 3C profiling of mouse and human brain cells has demonstrated robust cell type identification using chromatin conformation (87, 155), supporting the feasibility of reconstructing high-resolution chromatin interaction maps for constituent cell types of mammalian tissues that contain heterogeneous cell populations. Single-cell chromatin conformation profiling will become a powerful approach for fine mapping and functional studies of disease-associated variants in the post-GWAS era.

Different strategies for single-cell chromatin conformation profiling continue to be refined and developed. While single-cell Hi-C methods, which use biotin pulldown of contact junctions, enrich for sequencing reads informative of long-range interactions (125, 153), 3C-based methods that perform ligation immediately after restriction digestion (hence avoiding biotin pulldown) provide greater sensitivity and detect substantially more chromatin contacts in individual cells (45, 156). The trade-off is that 3C-based libraries contain a lower fraction of reads containing a ligation junction and thus increase the assay cost. The throughput of single-cell chromatin conformation studies may also be further increased using a combinatorial indexing approach (138). Furthermore, multiomic methods have recently been developed for joint chromatin conformation and methylome profiling of single nuclei (87, 89). The genomic architecture of single cells can also be reconstructed using in situ methods, such as genome architecture mapping (GAM), which slices nuclei into submicrometer-thin sections followed by sequencing of each section (11). Another in situ genome sequencing method has also been recently developed that can spatially localize thousands of genomic regions in individual nuclei (133).

MICROBIOLOGY

The diversity of microorganisms (bacteria, archaea, and unicellular eukaryotes) is vast but remains largely unknown because most (>99%) have not been successfully cultured (152). Bulk DNA metagenomics and ribosomal DNA sequence surveys have advanced taxonomies of microbial diversity, but their reliance on small genomic fragments limits their ability to identify many organisms and to associate genetic pathways with specific microbes (152, 174). scDNA-seq, however, is a culture-independent approach to resolve microbial diversity even in complex microbial communities. The most important applications of scDNA-seq to microbiology have been to expand the census of microorganisms (especially those that are rare and cannot be cultivated), resolve uncertainties in microbial phylogenies, study the physical associations between microorganisms, associate specific genetic pathways to taxa, and discover genetic pathways that may be biomedically useful (15, 86, 152, 174).

scDNA-seq was used initially to sequence small numbers of culturable bacteria as a proof of principle (136, 184), followed by uncultivated microbes from oral and soil samples (109, 134). It has since been used to discover numerous new bacterial clades and even superphyla by profiling diverse environmental samples, such as tropical oceans around the world, hydrothermal sites, freshwater bodies, and underground mines (74, 132, 142). In a large scDNA-seq study, more than 12,000 cells from tropical and subtropical oceans revealed many that could not be assigned to known taxonomies (1.2%, 2.3%, and 11% had no known class, order, and family, respectively) (132). One study identified the first example of lateral gene transfer between eukaryotes and archaea (142), a discovery that could not have been made with bulk metagenomics. scDNA-seq of uncultured protists has even been able to identify genetic sequences of bacteria and viruses that infected the cells (179). The remarkable diversity of microorganisms in almost every environmental sample (174) motivated the creation of a droplet-based method capable of profiling more than 50,000 cells, which was then used to characterize the taxonomic distribution of antibiotic resistance genes within a complex seawater sample (84). While challenges remain, such as varying lysis efficiency and uneven genome amplification of some microbes (15), the continued application and development of high-throughput scDNA-seq methods promises to greatly advance our knowledge of microbial diversity.

CONCLUSIONS AND FUTURE APPLICATIONS

The applications of scDNA-seq reviewed here are only early forays heralding a rapid growth of future applications that will provide deep biological insights and impactful biomedical uses. The continued advancement of single-cell genome coverage, throughput, and multiomic approaches will provide opportunities not only to enhance the resolution of developmental lineage trees, germ cell mutation patterns, cancer evolution, genome function, and microbial communities, but also to explore entirely new questions. While unexpected applications will certainly emerge, here, we briefly speculate regarding future and emerging applications of scDNA-seq that were not addressed in prior sections.

Forensic analysis of low-input DNA samples can be complicated by incomplete genotypes and mixtures of DNA from different individuals. Recent work has shown that scDNA-seq may be useful for separating contributing DNA from mixed semen swab samples (158), and we anticipate increasing use of scDNA-seq in forensics. Outside the setting of forensics, scDNA-seq poses genomic privacy concerns, since near-complete or complete genomes of individuals may feasibly be recovered from even the smallest numbers of cells constantly shed into the environment.

Future environmental uses of scDNA-seq include surveillance of pathogens, ecological changes, and antibiotic resistance of microbial species (e.g., in sewage and diverse ecosystems). In plant biology, scDNA-seq of plant gametes has enabled the generation of crossover maps without an extensive breeding program (36), which may enable more rapid methods of plant breeding. Immunology will benefit from the application of scDNA-seq to characterize immunological lineage expansions and selection in the context of infection and autoimmunity. And further studies of basic genome function, such as the recently developed single-cell genome replication timing (62), are emerging.

The full potential of scDNA-seq for biomedical research is still constrained by the technical limitations of current methods. Many of the technical challenges of scDNA-seq derive from the fact that it is not only a single-cell method but also a single-molecule method: With the exception of mtDNA, each genomic allele is present in one copy per cell. We predict that the field of scDNA-seq will continue to focus heavily on technology development, but that technologies that are robust and scalable will soon be commercialized and adopted by the wider research community,

as has occurred for single-cell transcriptomics. With a larger user base, including clinical adoption, we anticipate that the applications of scDNA-seq will broaden and will continue to expose new dimensions of biology at its most basic unit of life.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

G.D.E is supported by a Sontag Foundation Distinguished Scientist Award, National Institutes of Health grant 1DP5OD028158, a McKnight Foundation Technological Innovations in Neuroscience Award, a Pew Foundation Biomedical Scholars Award, and a Brain & Behavior Research Foundation Young Investigator Grant. A.G.H. is supported by Wellcome Trust grant 095552/Z/11/Z (to Peter Donnelly, Wellcome Centre for Human Genetics, Oxford, United Kingdom). C.L. is supported by the University of California, Los Angeles.

LITERATURE CITED

1. Abyzov A, Vaccarino FM. 2020. Cell lineage tracing and cellular diversity in humans. *Annu. Rev. Genom. Hum. Genet.* 21:101–16
2. Alemany A, Florescu M, Baron CS, Peterson-Maduro J, van Oudenaarden A. 2018. Whole-organism clone tracing using single-cell sequencing. *Nature* 556:108–12
3. Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, et al. 2015. Clock-like mutational processes in human somatic cells. *Nat. Genet.* 47:1402–7
4. Argelaguet R, Clark SJ, Mohammed H, Stapel LC, Krueger C, et al. 2019. Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature* 576:487–91
5. Bae T, Tomasini L, Mariani J, Zhou B, Roychowdhury T, et al. 2018. Different mutational rates and mechanisms in human cells at pregastrulation and neurogenesis. *Science* 359:550–55
6. 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
7. Baron CS, van Oudenaarden A. 2019. Unravelling cellular relationships during development and regeneration using genetic lineage tracing. *Nat. Rev. Mol. Cell Biol.* 20:753–65
8. Baslan T, Hicks J. 2017. Unravelling biology and shifting paradigms in cancer with single-cell sequencing. *Nat. Rev. Cancer* 17:557–69
9. Baslan T, Kendall J, Volyanskyy K, McNamara K, Cox H, et al. 2020. Novel insights into breast cancer copy number genetic heterogeneity revealed by single-cell genome sequencing. *eLife* 9:e51480
10. Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, et al. 2010. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* 327:836–40
11. Beagrie RA, Scialdone A, Schueler M, Kraemer DCA, Chotalia M, et al. 2017. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* 543:519–24
12. Bell AD, Mello CJ, Nemesh J, Brumbaugh SA, Wysoker A, McCarroll SA. 2020. Insights about variation in meiosis from 31,228 human sperm genomes. *Nature* 583:259–64
13. Bian S, Hou Y, Zhou X, Li X, Yong J, et al. 2018. Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science* 362:1060–63
14. Biezuner T, Spiro A, Raz O, Amir S, Milo L, et al. 2016. A generic, cost-effective, and scalable cell lineage analysis platform. *Genome Res.* 26:1588–99
15. Blainey PC. 2013. The future is now: single-cell genomics of bacteria and archaea. *FEMS Microbiol. Rev.* 37:407–27

16. Bohrson CL, Barton AR, Lodato MA, Rodin RE, Luquette LJ, et al. 2019. Linked-read analysis identifies mutations in single-cell DNA-sequencing data. *Nat. Genet.* 51:749–54
17. Brazhnik K, Sun S, Alani O, Kinkhabwala M, Wolkoff AW, et al. 2020. Single-cell analysis reveals different age-related somatic mutation profiles between stem and differentiated cells in human liver. *Sci. Adv.* 6:eaax2659
18. Breuss MW, Yang X, Antaki D, Schlachetzki JCM, Lana AJ, et al. 2020. Somatic mosaicism in the mature brain reveals clonal cellular distributions during cortical development. bioRxiv 2020.08.10.244814. **https://doi.org/10.1101/2020.08.10.244814**
19. Brick K, Thibault-Sennett S, Smagulova F, Lam K-WG, Pu Y, et al. 2018. Extensive sex differences at the initiation of genetic recombination. *Nature* 561:338–42
20. Brunner SF, Roberts ND, Wylie LA, Moore L, Aitken SJ, et al. 2019. Somatic mutations and clonal dynamics in healthy and cirrhotic human liver. *Nature* 574:538–42
21. Cai X, Evrony GD, Lehmann HS, Elhosary PC, Mehta BK, et al. 2014. Single-cell, genome-wide sequencing identifies clonal somatic copy-number variation in the human brain. *Cell Rep.* 8:1280–89
22. Capalbo A, Hoffmann ER, Cimadomo D, Ubaldi FM, Rienzi L. 2017. Human female meiosis revised: new insights into the mechanisms of chromosome segregation and aneuploidies from advanced genomics and time-lapse imaging. *Hum. Reprod. Update* 23:706–22
23. Carvalho CMB, Lupski JR. 2016. Mechanisms underlying structural variant formation in genomic disorders. *Nat. Rev. Genet.* 17:224–38
24. Casasent AK, Schalck A, Gao R, Sei E, Long A, et al. 2018. Multiclonal invasion in breast tumors identified by topographic single cell sequencing. *Cell* 172:205–17.e12
25. Cayrefourcq L, Vincent M-C, Pierredon S, Moutou C, Imbert-Bouteille M, et al. 2020. Single circulating fetal trophoblastic cells eligible for non invasive prenatal diagnosis: the exception rather than the rule. *Sci. Rep.* 10:9861
26. Chan MM, Smith ZD, Grosswendt S, Kretzmer H, Norman TM, et al. 2019. Molecular recording of mammalian embryogenesis. *Nature* 570:77–82
27. Dago AE, Stepansky A, Carlsson A, Luttgen M, Kendall J, et al. 2014. Rapid phenotypic and genomic change in response to therapeutic pressure in prostate cancer inferred by high content analysis of single circulating tumor cells. *PLOS ONE* 9:e101777
28. Dagogo-Jack I, Shaw AT. 2018. Tumour heterogeneity and resistance to cancer therapies. *Nat. Rev. Clin. Oncol.* 15:81–94
29. Daughtry BL, Rosenkrantz JL, Lazar NH, Fei SS, Redmayne N, et al. 2019. Single-cell sequencing of primate preimplantation embryos reveals chromosome elimination via cellular fragmentation and blastomere exclusion. *Genome Res.* 29:367–82
30. de Souza CPE, Andronescu M, Masud T, Kabeer F, Biele J, et al. 2020. Epiclomal: probabilistic clustering of sparse single-cell DNA methylation data. *PLOS Comput. Biol.* 16:e1008270
31. Demaree B, Delley CL, Vasudevan HN, Peretz CAC, Ruff D, et al. 2020. Joint profiling of proteins and DNA in single cells reveals extensive proteogenomic decoupling in leukemia. bioRxiv 2020.02.26.967133. **https://doi.org/10.1101/2020.02.26.967133**
32. D'Gama AM, Walsh CA. 2018. Somatic mosaicism and neurodevelopmental disease. *Nat. Neurosci.* 21:1504–14
33. Dixon JR, Gorkin David U, Ren B. 2016. Chromatin domains: the unit of chromosome organization. *Mol. Cell* 62:668–80
34. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, et al. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485:376–80
35. 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
36. Dreissig S, Fuchs J, Himmelbach A, Mascher M, Houben A. 2017. Sequencing of single pollen nuclei reveals meiotic recombination events at megabase resolution and circumvents segregation distortion caused by postmeiotic processes. *Front. Plant Sci.* 8:1620
37. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, et al. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489:57–74

38. Durbin RM, Altshuler D, Durbin RM, Abecasis GR, Bentley DR, et al. 2010. A map of human genome variation from population-scale sequencing. *Nature* 467:1061–73
39. Erickson RP. 2010. Somatic gene mutation and human disease other than cancer: an update. *Mutat. Res.* 705:96–106
40. Erwin JA, Paquola ACM, Singer T, Gallina I, Novotny M, et al. 2016. L1-associated genomic regions are deleted in somatic cells of the healthy human brain. *Nat. Neurosci.* 19:1583–91
41. Evrony GD, Cai X, Lee E, Hills LB, Elhosary PC, et al. 2012. Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell* 151:483–96
42. Evrony GD, Lee E, Mehta BK, Benjamini Y, Johnson RM, et al. 2015. Cell lineage analysis in human brain using endogenous retroelements. *Neuron* 85:49–59
43. Evrony GD, Lee E, Park PJ, Walsh CA. 2016. Resolving rates of mutation in the brain using single-neuron genomics. *eLife* 5:e12966
44. Ferronika P, van den Bos H, Taudt A, Spierings DCJ, Saber A, et al. 2017. Copy number alterations assessed at the single-cell level revealed mono- and polyclonal seeding patterns of distant metastasis in a small-cell lung cancer patient. *Ann. Oncol.* 28:1668–70
45. 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
46. Francis JM, Zhang C-Z, Maire CL, Jung J, Manzo VE, et al. 2014. *EGFR* variant heterogeneity in glioblastoma resolved through single-nucleus sequencing. *Cancer Discov.* 4:956–71
47. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, et al. 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449:851–61
48. Frieda KL, Linton JM, Hormoz S, Choi J, Chow K-HK, et al. 2017. Synthetic recording and in situ readout of lineage information in single cells. *Nature* 541:107–11
49. Frumkin D, Wasserstrom A, Kaplan S, Feige U, Shapiro E. 2005. Genomic variability within an organism exposes its cell lineage tree. *PLOS Comput. Biol.* 1:e50
50. Gaiti F, Chaligne R, Gu H, Brand RM, Kothen-Hill S, et al. 2019. Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia. *Nature* 569:576–80
51. Gao R, Davis A, McDonald TO, Sei E, Shi X, et al. 2016. Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nat. Genet.* 48:1119–30
52. Gawad C, Koh W, Quake SR. 2014. Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics. *PNAS* 111:17947
53. Gawad C, Koh W, Quake SR. 2016. Single-cell genome sequencing: current state of the science. *Nat. Rev. Genet.* 17:175–88
54. Gkountela S, Castro-Giner F, Szczerba BM, Vetter M, Landin J, et al. 2019. Circulating tumor cell clustering shapes DNA methylation to enable metastasis seeding. *Cell* 176:98–112.e14
55. Gonzalez V, Natarajan S, Xia Y, Klein D, Carter R, et al. 2020. Accurate genomic variant detection in single cells with primary template-directed amplification. bioRxiv 2020.11.20.391961. **https://doi.org/10.1101/2020.11.20.391961**
56. Greenberg MVC, Bourc'his D. 2019. The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* 20:590–607
57. Halldorsson BV, Palsson G, Stefansson OA, Jonsson H, Hardarson MT, et al. 2019. Characterizing mutagenic effects of recombination through a sequence-level genetic map. *Science* 363:eaau1043
58. Hassold T, Abruzzo M, Adkins K, Griffin D, Merrill M, et al. 1996. Human aneuploidy: incidence, origin, and etiology. *Environ. Mol. Mutagen.* 28:167–75
59. Hassold T, Hunt P. 2001. To err (meiotically) is human: the genesis of human aneuploidy. *Nat. Rev. Genet.* 2:280–91
60. Hinch AG, Tandon A, Patterson N, Song Y, Rohland N, et al. 2011. The landscape of recombination in African Americans. *Nature* 476:170–75
61. Hinch AG, Zhang G, Becker PW, Moralli D, Hinch R, et al. 2019. Factors influencing meiotic recombination revealed by whole-genome sequencing of single sperm. *Science* 363:eaau8861
62. Hiratani I, Takahashi S. 2019. DNA replication timing enters the single-cell era. *Genes* 10:221
63. Hou Y, Fan W, Yan L, Li R, Lian Y, et al. 2013. Genome analyses of single human oocytes. *Cell* 155:1492–506

64. 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
65. Hou Y, Song L, Zhu P, Zhang B, Tao Y, et al. 2012. Single-cell exome sequencing and monoclonal evolution of a *JAK2*-negative myeloproliferative neoplasm. *Cell* 148:873–85
66. Hu P, Liang D, Chen Y, Lin Y, Qiao F, et al. 2019. An enrichment method to increase cell-free fetal DNA fraction and significantly reduce false negatives and test failures for non-invasive prenatal screening: a feasibility study. *J. Transl. Med.* 17:124
67. Huang AY, Li P, Rodin RE, Kim SN, Dou Y, et al. 2020. Parallel RNA and DNA analysis after deep sequencing (PRDD-seq) reveals cell type-specific lineage patterns in human brain. *PNAS* 117:13886–95
68. Huang L, Ma F, Chapman A, Lu S, Xie XS. 2015. Single-cell whole-genome amplification and sequencing: methodology and applications. *Annu. Rev. Genom. Hum. Genet.* 16:79–102
69. Hughes AEO, Magrini V, Demeter R, Miller CA, Fulton R, et al. 2014. Clonal architecture of secondary acute myeloid leukemia defined by single-cell sequencing. *PLOS Genet.* 10:e1004462
70. Hui T, Cao Q, Wegrzyn-Woltosz J, O'Neill K, Hammond CA, et al. 2018. High-resolution single-cell DNA methylation measurements reveal epigenetically distinct hematopoietic stem cell subpopulations. *Stem Cell Rep.* 11:578–92
71. Hunter N. 2015. Meiotic recombination: the essence of heredity. *Cold Spring Harb. Perspect. Biol.* 7:a016618
72. Jones PA. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* 13:484–92
73. Kagan KO, Sonek J, Wagner P, Hoopmann M. 2017. Principles of first trimester screening in the age of non-invasive prenatal diagnosis: screening for chromosomal abnormalities. *Arch. Gynecol. Obstet.* 296:645–51
74. Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, et al. 2014. Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science* 344:416–20
75. Kelsey G, Stegle O, Reik W. 2017. Single-cell epigenomics: recording the past and predicting the future. *Science* 358:69–75
76. Kester L, van Oudenaarden A. 2018. Single-cell transcriptomics meets lineage tracing. *Cell Stem Cell* 23:166–79
77. Kim C, Gao R, Sei E, Brandt R, Hartman J, et al. 2018. Chemoresistance evolution in triple-negative breast cancer delineated by single-cell sequencing. *Cell* 173:879–93.e13
78. Klemm SL, Shipony Z, Greenleaf WJ. 2019. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* 20:207–20
79. Knouse KA, Wu J, Whittaker CA, Amon A. 2014. Single cell sequencing reveals low levels of aneuploidy across mammalian tissues. *PNAS* 111:13409
80. Kong A, Thorleifsson G, Gudbjartsson DF, Masson G, Sigurdsson A, et al. 2010. Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* 467:1099–103
81. Kretzschmar K, Watt FM. 2012. Lineage tracing. *Cell* 148:33–45
82. Kumar A, Ryan A, Kitzman JO, Wemmer N, Snyder MW, et al. 2015. Whole genome prediction for preimplantation genetic diagnosis. *Genome Med.* 7:35
83. Laks E, McPherson A, Zahn H, Lai D, Steif A, et al. 2019. Clonal decomposition and DNA replication states defined by scaled single-cell genome sequencing. *Cell* 179:1207–21.e22
84. Lan F, Demaree B, Ahmed N, Abate AR. 2017. Single-cell genome sequencing at ultra-high-throughput with microfluidic droplet barcoding. *Nat. Biotechnol.* 35:640–46
85. Lareau CA, Ludwig LS, Muus C, Gohil SH, Zhao T, et al. 2021. Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. *Nat. Biotechnol.* 39:451–61
86. Lasken RS. 2012. Genomic sequencing of uncultured microorganisms from single cells. *Nat. Rev. Microbiol.* 10:631–40
87. Lee D-S, Luo C, Zhou J, Chandran S, Rivkin A, et al. 2019. Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. *Nat. Methods* 16:999–1006
88. Leung ML, Davis A, Gao R, Casasent A, Wang Y, et al. 2017. Single-cell DNA sequencing reveals a late-dissemination model in metastatic colorectal cancer. *Genome Res.* 27:1287–99

89. Li G, Liu Y, Zhang Y, Kubo N, Yu M, et al. 2019. Joint profiling of DNA methylation and chromatin architecture in single cells. *Nat. Methods* 16:991–93
90. Li L, Li L, Li Q, Liu X, Ma X, et al. 2021. Dissecting the epigenomic dynamics of human fetal germ cell development at single-cell resolution. *Cell Res.* 31:463–77
91. Li R, Bitoun E, Altemose N, Davies RW, Davies B, Myers SR. 2019. A high-resolution map of non-crossover events reveals impacts of genetic diversity on mammalian meiotic recombination. *Nat. Commun.* 10:3900
92. Li Y, Xu X, Song L, Hou Y, Li Z, et al. 2012. Single-cell sequencing analysis characterizes common and cell-lineage-specific mutations in a muscle-invasive bladder cancer. *GigaScience* 1:12
93. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326:289–93
94. Liu EY, Morgan AP, Chesler EJ, Wang W, Churchill GA, Pardo-Manuel de Villena F. 2014. High-resolution sex-specific linkage maps of the mouse reveal polarized distribution of crossovers in male germline. *Genetics* 197:91–106
95. Liu H, Zhou J, Tian W, Luo C, Bartlett A, et al. 2020. DNA methylation atlas of the mouse brain at single-cell resolution. bioRxiv 2020.04.30.069377. **https://doi.org/10.1101/2020.04.30.069377**
96. Liu M, Liu Y, Di J, Su Z, Yang H, et al. 2017. Multi-region and single-cell sequencing reveal variable genomic heterogeneity in rectal cancer. *BMC Cancer* 17:787
97. Lodato MA, Rodin RE, Bohrson CL, Coulter ME, Barton AR, et al. 2018. Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science* 359:555–59
98. Lodato MA, Woodworth MB, Lee S, Evrony GD, Mehta BK, et al. 2015. Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* 350:94–98
99. Lohr JG, Adalsteinsson VA, Cibulskis K, Choudhury AD, Rosenberg M, et al. 2014. Whole-exome sequencing of circulating tumor cells provides a window into metastatic prostate cancer. *Nat. Biotechnol.* 32:479–84
100. Lu S, Zong C, Fan W, Yang M, Li J, et al. 2012. Probing meiotic recombination and aneuploidy of single sperm cells by whole-genome sequencing. *Science* 338:1627–30
101. Ludwig LS, Lareau CA, Ulirsch JC, Christian E, Muus C, et al. 2019. Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. *Cell* 176:1325–39.e22
102. Luo C, Hajkova P, Ecker JR. 2018. Dynamic DNA methylation: in the right place at the right time. *Science* 361:1336–40
103. 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
104. Luo C, Li X, Zhang Q, Yan J. 2019. Single gametophyte sequencing reveals that crossover events differ between sexes in maize. *Nat. Commun.* 10:785
105. Luo C, Liu H, Xie F, Armand EJ, Siletti K, et al. 2019. Single nucleus multi-omics links human cortical cell regulatory genome diversity to disease risk variants. bioRxiv 2019.12.11.873398. **https://doi.org/10.1101/2019.12.11.873398**
106. Luquette LJ, Bohrson CL, Sherman MA, Park PJ. 2019. Identification of somatic mutations in single cell DNA-seq using a spatial model of allelic imbalance. *Nat. Commun.* 10:3908
107. 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
108. Macaulay IC, Ponting CP, Voet T. 2017. Single-cell multiomics: multiple measurements from single cells. *Trends Genet.* 33:155–68
109. Marcy Y, Ouverney C, Bik EM, Lösekann T, Ivanova N, et al. 2007. Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *PNAS* 104:11889–94
110. Marston AL, Amon A. 2004. Meiosis: cell-cycle controls shuffle and deal. *Nat. Rev. Mol. Cell Biol.* 5:983–97
111. Martelotto LG, Baslan T, Kendall J, Geyer FC, Burke KA, et al. 2017. Whole-genome single-cell copy number profiling from formalin-fixed paraffin-embedded samples. *Nat. Med.* 23:376–85

112. Martincorena I, Campbell PJ. 2015. Somatic mutation in cancer and normal cells. *Science* 349:1483–89
113. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, et al. 2012. Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337:1190–95
114. McCarthy DJ, Rostom R, Huang Y, Kunz DJ, Danecek P, et al. 2020. Cardelino: computational integration of somatic clonal substructure and single-cell transcriptomes. *Nat. Methods* 17:414–21
115. McConnell MJ, Lindberg MR, Brennand KJ, Piper JC, Voet T, et al. 2013. Mosaic copy number variation in human neurons. *Science* 342:632–37
116. McConnell MJ, Moran JV, Abyzov A, Akbarian S, Bae T, et al. 2017. Intersection of diverse neuronal genomes and neuropsychiatric disease: the Brain Somatic Mosaicism Network. *Science* 356:eaal1641
117. McCoy RC. 2017. Mosaicism in preimplantation human embryos: when chromosomal abnormalities are the norm. *Trends Genet.* 33:448–63
118. McKenna A, Gagnon JA. 2019. Recording development with single cell dynamic lineage tracing. *Development* 146:dev169730
119. McPherson A, Roth A, Laks E, Masud T, Bashashati A, et al. 2016. Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. *Nat. Genet.* 48:758–67
120. Miles LA, Bowman RL, Merlinsky TR, Csete IS, Ooi A, et al. 2020. Single cell mutational profiling delineates clonal trajectories in myeloid malignancies. bioRxiv 2020.02.07.938860. **https://doi.org/10.1101/2020.02.07.938860**
121. Miller MB, Reed HC, Walsh CA. 2021. Brain somatic mutation in aging and Alzheimer's disease. *Annu. Rev. Genom. Hum. Genet.* 22:239–56
122. Morita K, Wang F, Jahn K, Hu T, Tanaka T, et al. 2020. Clonal evolution of acute myeloid leukemia revealed by high-throughput single-cell genomics. *Nat. Commun.* 11:5327
123. Mulqueen RM, Pokholok D, Norberg SJ, Torkenczy KA, Fields AJ, et al. 2018. Highly scalable generation of DNA methylation profiles in single cells. *Nat. Biotechnol.* 36:428–31
124. Myers S, Bowden R, Tumian A, Bontrop RE, Freeman C, et al. 2010. Drive against hotspot motifs in primates implicates the *PRDM9* gene in meiotic recombination. *Science* 327:876–79
125. 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
126. Nam AS, Chaligne R, Landau DA. 2020. Integrating genetic and non-genetic determinants of cancer evolution by single-cell multi-omics. *Nat. Rev. Genet.* 22:3–18
127. Navin NE. 2015. The first five years of single-cell cancer genomics and beyond. *Genome Res.* 25:1499–507
128. Navin NE, Kendall J, Troge J, Andrews P, Rodgers L, et al. 2011. Tumour evolution inferred by single-cell sequencing. *Nature* 472:90–94
129. Ni X, Zhuo M, Su Z, Duan J, Gao Y, et al. 2013. Reproducible copy number variation patterns among single circulating tumor cells of lung cancer patients. *PNAS* 110:21083–88
130. Oliver PL, Goodstadt L, Bayes JJ, Birtle Z, Roach KC, et al. 2009. Accelerated evolution of the *Prdm9* speciation gene across diverse metazoan taxa. *PLOS Genet.* 5:e1000753
131. Ottolini CS, Newnham LJ, Capalbo A, Natesan SA, Joshi HA, et al. 2015. Genome-wide maps of recombination and chromosome segregation in human oocytes and embryos show selection for maternal recombination rates. *Nat. Genet.* 47:727–35
132. Pachiadaki MG, Brown JM, Brown J, Bezuidt O, Berube PM, et al. 2019. Charting the complexity of the marine microbiome through single-cell genomics. *Cell* 179:1623–35.e11
133. Payne AC, Chiang ZD, Reginato PL, Mangiameli SM, Murray EM, et al. 2021. In situ genome sequencing resolves DNA sequence and structure in intact biological samples. *Science* 371:eaay3446
134. Podar M, Abulencia CB, Walcher M, Hutchison D, Zengler K, et al. 2007. Targeted access to the genomes of low-abundance organisms in complex microbial communities. *Appl. Environ. Microbiol.* 73:3205–14
135. Poduri A, Evrony GD, Cai X, Walsh CA. 2013. Somatic mutation, genomic variation, and neurological disease. *Science* 341:1237758
136. Raghunathan A, Ferguson HR, Bornarth CJ, Song W, Driscoll M, Lasken RS. 2005. Genomic DNA amplification from a single bacterium. *Appl. Environ. Microbiol.* 71:3342–47

137. Raj B, Wagner DE, McKenna A, Pandey S, Klein AM, et al. 2018. Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotechnol.* 36:442–50
138. Ramani V, Deng X, Qiu R, Gunderson KL, Steemers FJ, et al. 2017. Massively multiplex single-cell Hi-C. *Nat. Methods* 14:263–66
139. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, et al. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159:1665–80
140. Reizel Y, Chapal-Ilani N, Adar R, Itzkovitz S, Elbaz J, et al. 2011. Colon stem cell and crypt dynamics exposed by cell lineage reconstruction. *PLOS Genet.* 7:e1002192
141. Ren X, Kang B, Zhang Z. 2018. Understanding tumor ecosystems by single-cell sequencing: promises and limitations. *Genome Biol.* 19:211
142. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, et al. 2013. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499:431–37
143. Rozenblatt-Rosen O, Regev A, Oberdoerffer P, Nawy T, Hupalowska A, et al. 2020. The Human Tumor Atlas Network: charting tumor transitions across space and time at single-cell resolution. *Cell* 181:236–49
144. Salipante SJ, Horwitz MS. 2006. Phylogenetic fate mapping. *PNAS* 103:5448–53
145. Salipante SJ, Horwitz MS. 2007. A phylogenetic approach to mapping cell fate. *Curr. Top. Dev. Biol.* 79:157–84
146. Sandor C, Li W, Coppieters W, Druet T, Charlier C, Georges M. 2012. Genetic variants in *REC8*, *RNF212*, and *PRDM9* influence male recombination in cattle. *PLOS Genet.* 8:e1002854
147. Sato T, Ito Y, Samura O, Aoki H, Uchiyama T, et al. 2020. Direct assessment of single-cell DNA using crudely purified live cells: a proof of concept for noninvasive prenatal definitive diagnosis. *J. Mol. Diagn.* 22:132–40
148. Schultz MD, He Y, Whitaker JW, Hariharan M, Mukamel EA, et al. 2015. Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* 523:212–16
149. Shi Q, Qiu Y, Xu C, Yang H, Li C, et al. 2020. Next-generation sequencing analysis of each blastomere in good-quality embryos: insights into the origins and mechanisms of embryonic aneuploidy in cleavage-stage embryos. *J. Assist. Reprod. Genet.* 37:1711–18
150. Sloan DB, Broz AK, Sharbrough J, Wu Z. 2018. Detecting rare mutations and DNA damage with sequencing-based methods. *Trends Biotechnol.* 36:729–40
151. Spanjaard B, Hu B, Mitic N, Olivares-Chauvet P, Janjuha S, et al. 2018. Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nat. Biotechnol.* 36:469–73
152. Stepanauskas R. 2012. Single cell genomics: an individual look at microbes. *Curr. Opin. Microbiol.* 15:613–20
153. Stevens TJ, Lando D, Basu S, Atkinson LP, Cao Y, et al. 2017. 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* 544:59–64
154. Sun H, Rowan BA, Flood PJ, Brandt R, Fuss J, et al. 2019. Linked-read sequencing of gametes allows efficient genome-wide analysis of meiotic recombination. *Nat. Commun.* 10:4310
155. Tan L, Ma W, Wu H, Zheng Y, Xing D, et al. 2021. Changes in genome architecture and transcriptional dynamics progress independently of sensory experience during post-natal brain development. *Cell* 184:741–58.e17
156. Tan L, Xing D, Chang C-H, Li H, Xie XS. 2018. Three-dimensional genome structures of single diploid human cells. *Science* 361:924–28
157. Tao L, Raz O, Marx Z, Gosh M, Huber S, et al. 2020. Retrospective cell lineage reconstruction in humans using short tandem repeats. bioRxiv 191296. **https://doi.org/10.1101/191296**
158. Theunissen G. 2020. *DNA profiling of single sperm cells and single skin flakes in forensics using micromanipulation and whole genome amplification*. PhD Thesis, Robert Gordon Univ., Aberdeen, Scotl.
159. Thorpe J, Osei-Owusu IA, Avigdor BE, Tupler R, Pevsner J. 2020. Mosaicism in human health and disease. *Annu. Rev. Genet.* 54:487–510
160. Treff NR, Fedick A, Tao X, Devkota B, Taylor D, Scott RT Jr. 2013. Evaluation of targeted next-generation sequencing-based preimplantation genetic diagnosis of monogenic disease. *Fertil. Steril.* 99:1377–84.e6

161. Vanneste E, Voet T, Le Caignec C, Ampe M, Konings P, et al. 2009. Chromosome instability is common in human cleavage-stage embryos. *Nat. Med.* 15:577–83
162. Vermeesch JR, Voet T, Devriendt K. 2016. Prenatal and pre-implantation genetic diagnosis. *Nat. Rev. Genet.* 17:643–56
163. Vijg J, Dong X. 2020. Pathogenic mechanisms of somatic mutation and genome mosaicism in aging. *Cell* 182:12–23
164. Wagner DE, Klein AM. 2020. Lineage tracing meets single-cell omics: opportunities and challenges. *Nat. Rev. Genet.* 21:410–27
165. Wang J, Fan HC, Behr B, Quake SR. 2012. Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. *Cell* 150:402–12
166. Wang Y, Waters J, Leung ML, Unruh A, Roh W, et al. 2014. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* 512:155–60
167. Wasserstrom A, Adar R, Shefer G, Frumkin D, Itzkovitz S, et al. 2008. Reconstruction of cell lineage trees in mice. *PLOS ONE* 3:e1939
168. Wasserstrom A, Frumkin D, Adar R, Itzkovitz S, Stern T, et al. 2008. Estimating cell depth from somatic mutations. *PLOS Comput. Biol.* 4:e1000058
169. Wegmann D, Kessner DE, Veeramah KR, Mathias RA, Nicolae DL, et al. 2011. Recombination rates in admixed individuals identified by ancestry-based inference. *Nat. Genet.* 43:847–53
170. Wei CJ-Y, Zhang K. 2020. RETrace: simultaneous retrospective lineage tracing and methylation profiling of single cells. *Genome Res.* 30:602–10
171. Weinreb C, Rodriguez-Fraticelli A, Camargo FD, Klein AM. 2020. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science* 367:eaaw3381
172. Wells D, Kaur K, Grifo J, Glassner M, Taylor JC, et al. 2014. Clinical utilisation of a rapid low-pass whole genome sequencing technique for the diagnosis of aneuploidy in human embryos prior to implantation. *J. Med. Genet.* 51:553–62
173. 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
174. Woyke T, Doud DFR, Schulz F. 2017. The trajectory of microbial single-cell sequencing. *Nat. Methods* 14:1045–54
175. Xu J, Nuno K, Litzenburger UM, Qi Y, Corces MR, et al. 2019. Single-cell lineage tracing by endogenous mutations enriched in transposase accessible mitochondrial DNA. *eLife* 8:e45105
176. Xu X, Hou Y, Yin X, Bao L, Tang A, et al. 2012. Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell* 148:886–95
177. Yin Y, Jiang Y, Lam K-WG, Berletch JB, Disteche CM, et al. 2019. High-throughput single-cell sequencing with linear amplification. *Mol. Cell* 76:676–90.e10
178. Yizhak K, Aguet F, Kim J, Hess JM, Kübler K, et al. 2019. RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues. *Science* 364:eaaw0726
179. Yoon HS, Price DC, Stepanauskas R, Rajah VD, Sieracki ME, et al. 2011. Single-cell genomics reveals organismal interactions in uncultivated marine protists. *Science* 332:714–717
180. Yoshida K, Gowers KHC, Lee-Six H, Chandrasekharan DP, Coorens T, et al. 2020. Tobacco smoking and somatic mutations in human bronchial epithelium. *Nature* 578:266–72
181. Zhang C-Z, Spektor A, Cornils H, Francis JM, Jackson EK, et al. 2015. Chromothripsis from DNA damage in micronuclei. *Nature* 522:179–84
182. Zhang J, Fujimoto J, Zhang J, Wedge DC, Song X, et al. 2014. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* 346:256–59
183. Zhang K, Deng R, Gao H, Teng X, Li J. 2020. Lighting up single-nucleotide variation in situ in single cells and tissues. *Chem. Soc. Rev.* 49:1932–54
184. Zhang K, Martiny AC, Reppas NB, Barry KW, Malek J, et al. 2006. Sequencing genomes from single cells by polymerase cloning. *Nat. Biotechnol.* 24:680–86
185. Zhang L, Dong X, Lee M, Maslov AY, Wang T, Vijg J. 2019. Single-cell whole-genome sequencing reveals the functional landscape of somatic mutations in B lymphocytes across the human lifespan. *PNAS* 116:9014–19

186. Zhang L, Vijg J. 2018. Somatic mutagenesis in mammals and its implications for human disease and aging. *Annu. Rev. Genet.* 52:397–419
187. Zhu C, Preissl S, Ren B. 2020. Single-cell multimodal omics: the power of many. *Nat. Methods* 17:11–14
188. Zhu P, Guo H, Ren Y, Hou Y, Dong J, et al. 2018. Single-cell DNA methylome sequencing of human preimplantation embryos. *Nat. Genet.* 50:12–19
189. Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LTY, et al. 2013. Charting a dynamic DNA methylation landscape of the human genome. *Nature* 500:477–81