The role of single-cell genomics in human genetics

Varun K A Sreenivasan ,[1] Saranya Balachandran,[1] Malte Spielmann [1,2]

[1]Institute of Human Genetics, University Hospital Schleswig-Holstein, University of Lübeck and Kiel University, Lübeck and Kiel, Germany
[2]Human Molecular Genetics Group, Max Planck Institute for Molecular Genetics, Berlin, Germany

Correspondence to
Professor Malte Spielmann, University of Lübeck, Lubeck 23562, Germany; Malte.Spielmann@uksh.de

VKAS and SB contributed equally.

ABSTRACT

Single-cell sequencing is a powerful approach that can detect genetic alterations and their phenotypic consequences in the context of human development, with cellular resolution. Humans start out as single-cell zygotes and undergo fission and differentiation to develop into multicellular organisms. Before fertilisation and during development, the cellular genome acquires hundreds of mutations that propagate down the cell lineage. Whether germline or somatic in nature, some of these mutations may have significant genotypic impact and lead to diseased cellular phenotypes, either systemically or confined to a tissue. Single-cell sequencing enables the detection and monitoring of the genotype and the consequent molecular phenotypes at a cellular resolution. It offers powerful tools to compare the cellular lineage between 'normal' and 'diseased' conditions and to establish genotype-phenotype relationships. By preserving cellular heterogeneity, single-cell sequencing, unlike bulk-sequencing, allows the detection of even small, diseased subpopulations of cells within an otherwise normal tissue. Indeed, the characterisation of biopsies with cellular resolution can provide a mechanistic view of the disease. While single-cell approaches are currently used mainly in basic research, it can be expected that applications of these technologies in the clinic may aid the detection, diagnosis and eventually the treatment of rare genetic diseases as well as cancer. This review article provides an overview of the single-cell sequencing technologies in the context of human genetics, with an aim to empower clinicians to understand and interpret the single-cell sequencing data and analyses. We discuss the state-of-the-art experimental and analytical workflows and highlight current challenges/limitations. Notably, we focus on two prospective applications of the technology in human genetics, namely the annotation of the non-coding genome using single-cell functional genomics and the use of single-cell sequencing data for in silico variant prioritisation.

INTRODUCTION

Cells are unique—differing from one another in terms of their copy numbers, transcriptional landscape, physiological functions and spatiotemporal locations.[1] The recent advances of single-cell sequencing (sc-seq) technologies have revolutionised how we profile genetic, epigenetic, transcriptional or proteomic variations in individual cells. Named 'Method of Year' by Nature Methods in 2013 and 'Breakthrough of the Year' by Science in 2018,[2] sc-seq approaches have so far enabled the identification of new cell types, building of cellular atlases of human, mouse, zebrafish and frog embryos (https://www.humancellatlas.org/),[3–11] identification of disease mechanisms,[12] detection and tracing of cancer progression and therapy[13] as well as the screening of genomic regions at scale.[14 15] Organismal-level atlases provide insights into the development, the cell type constitution of mature organs and the spatial context of every cell. They provide a framework to readily investigate developmental disorders as well as pleiotropism arising from ubiquitous cell types present across multiple organs. Merged atlases combining mouse and human data also provide a bridge to link human diseases to mouse models.[3] By mapping the entire human body in adults and in embryonic stages as part of the Human Cell Atlas initiative, sc-seq technologies can also help decipher the cellular-level changes that occur during a specific disease and monitor subsequent anomalies in cellular differentiation.[3]

Of primary importance to human geneticists is the identification of genes, pathways or mechanisms that give rise to disease phenotypes, as well as the detection, monitoring and prediction of disease, its progression or the response to an intervention, which may be confined to a subpopulation of cells within a tissue or an organ. This necessitates characterisation of biopsies and understanding of mechanisms at a cellular resolution. Unlike bulk-sequencing (bulk-seq), which only informs on the tissue averages, sc-seq enables identification and quantification of cellular heterogeneities, such as in the case of neurological diseases[16] or cardiac failure.[17] Moreover, the introduction of high-throughput sequencing approaches in the clinic has led to a massive growth in the number of variants of uncertain significance, new candidate genes and new phenotypes.[18] Screening of unannotated genomic regions and resulting phenotypes using traditional low throughput methods is unfeasible. Sc-seq approaches, in combination with gene-editing technologies, make it now possible to screen thousands of genomic regions in a single experiment [19] and to identify phenotypes, previously overlooked by other technologies.[20] For these reasons, sc-seq is finding increasing interest in human genetics research.

In this review article, we provide an overview of the current sc-seq technologies with an emphasis on the methods relevant to human genetics. To define the scope, we consider sc-seq as a congregation of methods that enables the profiling of individual cells by applying sequencing technologies. In this overarching definition, we include the profiling of the cellular genome, epigenome, transcriptome and proteome. But we exclude technologies that use immunofluorescence, in situ hybridisation, mass-spectroscopy or other readout schemes, such as MERFISH (multiplexed error-robust fluorescence *in situ* hybridization), or CyTOF (cytometry by time of flight), which are reviewed elsewhere.[21 22] We mainly focus our discussions on well-established

methods, but have strived to point out promising recent developments in the field. At first, we discuss the progress made in the experimental workflow and then describe how diverse computational methods help to extract relevant information and construct models. In the latter half, we focus on two prospective applications, chosen for their immediate relevance to human genetics. These include the experimental screening of genomic regions and the prioritisation of variants using computational approaches.

The aim of this review article is (i) to provide a comprehensive overview to a human geneticist entering the field, (ii) to generalise and simplify concepts, where possible, so as to help understand and interpret the literature, (iii) to provide an overview of the current use case scenarios of sc-seq in human genetics basic research and (iv) to forecast on prospective applications of the technology in the human genetics clinic. For focused information beyond the scope of this comprehensive review, we recommend several review articles that discuss the current trends on various topics in detail, including the different sequencing modalities, such as genome,[23–26] epigenome,[27–29] transcriptome,[30–34] multiome[35–37] as well as their applications in cancer,[38] cardiovascular diseases[39] and neurological diseases.[40]

HOW DOES SINGLE-CELL SEQUENCING WORK?
The workflow of single-cell sequencing
Contemporary sc-seq, be it genome, transcriptome or another modality, entails dissociating tissue into cells, capturing and tagging the molecules within each cell with a unique nucleotide sequence (cellular barcode), creating a sequencing library and sequencing. Ground-breaking innovations have been made in all these aspects of the workflow over the last decade (figure 1). As a result, improvements have occurred in the cellular throughput, the diversity of sequence-able molecular species (modalities) as well as many other characteristics (eg, sensitivity, quality, depth and cost) (online supplemental figure 1).[41] In this section, we provide a brief overview of the workflow and summarise the current capabilities and any associated limitations. An exhaustive list of sc-seq techniques and analysis workflow is accessible in several recent reviews.[26 30–32 35 42]

From complex tissues to single cells
The first step in all sc-seq approaches is the dissociation of the tissue into single cells (figure 1A). It is normally achieved by mechanical (eg, cutting, crushing, mincing) and/or enzymatic (eg, trypsin) means, unless a more straightforward cellular separation is possible, like in the case of blood.[16 43–45] This key step influences the outcome of all downstream results, because tissues respond differently to the processing, due to the differences in the cellular composition and the extracellular matrix. This is especially the case in complex tissues (eg, brain), where such biases may give rise to skewed representation of cell types in the final data. For example, cryopreservation has been shown to result in the selective loss of epithelial cells.[46] Enzymatic methods can also affect the omics-profile of the cell by inducing stress responses.[46] FACS sorting may often be required to enrich specific cell types as well as to reduce debris.[16] Due to the difficulty of obtaining intact cells from certain tissues (eg, mature brain, frozen tissue), single-nuclei sequencing (for simplicity collectively denoted hereafter as sc-seq) has been gaining popularity, because intact nuclei are more easily suspended and barcoded than cells.[46–49] For this reason, many large-scale consortia, such as the Human Cell Atlas,[3 50] mainly use single-nuclei sequencing for unbiased sampling.
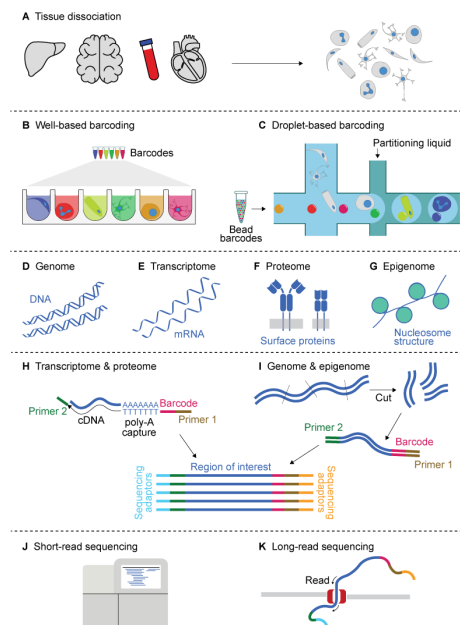

Figure 1 Five stages of single-cell sequencing (sc-seq) experimental workflow. (A) Dissociation of the desired tissue into a cellular suspension is the first step in the sc-seq experimental workflow. (B–C) Cellular barcoding of all required molecules within a cell is realised by segregating individual cells in (B) wells or (C) droplets. Barcodes (demarcated by various colours) are introduced in a solution form or as beads, depending on the technology used. (D–G) The sc-seq allows interrogation of the cellular genome, transcriptome, proteome as well as the epigenome—referred to as sequencing modalities. (H, I) Construction of a sequencing library from the barcoded molecules. (H) The transcriptome and proteome are generally captured by taking advantage of poly-A sequences. This is either the 3′ poly-A tail in the case of mRNA or is part of oligonucleotide sequences conjugated to the antibodies used to recognise the proteins of interest. (I) Genome and epigenome sequencing requires DNA fragmentation, amplification and additional steps depending on the feature of interest. (J, K) Short-read sequencing offers high-accuracy and speed, but is limited to <200 bp read lengths, whereas long-read sequencing generally offers sequence lengths reaching tens of kilo bases.

Labelling the content of each cell by cellular barcoding
Barcoding the molecular contents of each cell allows tracing every sequencing-read back to its cell-of-origin. This is achieved by segregating the dissociated cells or nuclei into small 'reaction microchambers' represented either by microwells (figure 1B) or microdroplets within a microfluidic system (figure 1C).[51] Each reaction chamber contains a reagent cocktail, including enzymes, antibodies and/or primers, in order to permeabilise the cell/nuclear membrane, capture the molecules of interest and to impart a cellular barcode to every captured molecule within the reaction microchamber. The cellular barcodes are 6–20 base nucleotide sequences in length and unique to each microchamber. In many of the high-throughput well or droplet technologies, the reagents for molecular capture and barcoding are either anchored to the bottom of the wells, added to the individual wells or encapsulated in microbeads. Cells are placed into such microchambers at random by means of microfluidics or pipetting. By minimising cases in which more than one cell ends up in a single microchamber (usually below 10%), all of the molecules within each cell, in most cases, receive a unique barcode. This allows in silico allocation of every sequencing-read back to its cell-of-origin.

Among the numerous advances over the years, the cellular throughput—the number of cells that can be barcoded at once—has seen a remarkable increase from a few cells since the initial demonstration[52] to many thousands of cells using well-based and droplet-based techniques.[53–57] Cellular throughput in the order of millions of cells, for example, to sequence organs or even entire embryos, is possible with 'combinatorial indexing' or 'split-pool' barcoding.[7 58–60] The cellular throughput is only one of the many differences between these cellular barcoding approaches. Other differences include the levels of automation and convenience, cost, cellular capture efficiency, molecular capture efficiency, doublet-rate, remote-processing (eg, compatibility with formaldehyde-fixed tissue) and customisation options.[33 61] While generally low in throughput, well-based or plate-based methods allow cellular visualisation by microscopy and are more amenable to customisation.[56] In contrast, droplet-based methods offer higher cellular throughput with automation, where the cell/nuclear lysis, molecular capture and barcoding occur 'on-the-flow'.

Modalities of sc-seq

While the vast majority of sc-seq experiments assay the transcriptome of each cell (figure 1E),[52 62 63] the technology has now been expanded to other omics-modalities, including the genome, epigenome and proteome (figure 1D, F and G). These modalities provide information at the different levels of the *central dogma*: sc-genome-seq helps identify somatic mutations and mosaicism; sc-epigenome-seq measures features such as chromatin accessibility, histone modifications, DNA methylation profile, three-dimensional (3D) genome architecture, etc, which regulate transcription, cellular differentiation and disease progression; sc-transcriptome-sq quantifies the gene-expression, which often is used as a proxy for cell state or type and the sc-proteome-seq provides information about translated products, usually cell-surface proteins.

The innovation in single-cell proteome and epigenome sequencing lies in library preparation, where information of interest (eg, the presence of a protein or the chromatin architecture) is converted to a sequence-able DNA library with cellular barcoding. Stoeckius *et al*,[64] Shahi *et al*[65] and Peterson *et al*[66] demonstrated sc-proteome-seq by using oligonucleotide-conjugated antibodies targeted against various proteins of interest. These oligonucleotides contained adaptor sequences required for library preparation as well as molecular barcodes unique to the antibodies to enable in silico proteome re-construction. sc-proteome-seq offers unmatched scalability (at least theoretically) when compared with other methods with inherent limitations on the bandwidth (eg, microscopy, flow cytometry or mass cytometry): a 10-base oligonucleotide generates sufficient unique molecular barcodes that exceed the number of known human-proteins.[64] In practice, however, the number of detectable proteins is limited by antibodies available. For epigenome sequencing, Cusanovich *et al*[58] and Buenrostro *et al*[67] developed sc-ATAC-seq (assay for transposase-accessible chromatin) to elucidate the mechanisms that give rise to intercellular heterogeneity, despite harbouring the same genetic code. Other epigenomic modalities, such as DNA methylation, 3D genome architecture and histone marks, have also been demonstrated,[68–73] which are discussed in recent reviews.[27] Recent advances have been focused on the combined sequencing of multiple modalities,[36 74] such as co-assays to simultaneously profile chromatin accessibility and gene expression.[60 75 76] Such methods allow linking regulatory elements to gene-expression

patterns within the same cell and help delineate gene regulatory networks, and are becoming increasingly popular techniques for basic research in genetics.

Library preparation and sequencing technologies

The conversion of the molecules of interest within each cell into a barcoded, sequence-able DNA library is at the crux of sc-seq (figure 1H,I). Transcriptome and proteome sequencing usually take advantage of poly-A terminations of transcripts or oligonucleotide-tags on the antibodies to capture the fragments, after which RT and PCR amplifications are performed to generate a sequencing library (figure 1H). To account for bias induced by exponential PCR amplification, certain protocols tag every captured molecule with a unique molecular identifier (UMI—a short random nucleotide sequence) prior to PCR. All sequencing reads from a cell (ie, cellular barcode) associated with a single UMI are pooled.

In general, for genomic and epigenomic sc-seq modalities, including whole genome sequencing, chromatin accessibility, methylation and other epigenetic information, the DNA is segmented, barcoded and amplified (figure 1I). Except for the additional cellular barcoding steps, these procedures are similar and often adapted from developments in bulk-seq. Specifically, several methods have been developed for single-cell whole genome amplification, such as random/degenerate oligonucleotide primed PCR, multiple displacement amplification, multiple annealing and looping-based amplification cycles, linear amplification via transposon insertion, which all result in sequencing libraries with lengths in the range of several hundred bases.[26] A recent development has also demonstrated sc-genome-seq with high-fidelity reads reaching kilobase lengths with the use of a transposase to fragment the DNA.[77] The library preparation varies across the different sc-epigenome-seq techniques,[27 28] depending on the feature of interest. In the case of chromatin accessibility, enzymes such as DNase,[78] Tn5 transposase[58 67] are used to fragment the genome at open chromatin regions, followed by barcoding and adaptor ligation. In the case of profiling histone-marks, earlier methods used free enzymes, such as MNase,[73] to non-specifically fragment the DNA, followed by barcoding and pooled immunoprecipitation. More recent approaches to profile histone-marks without the additional step of immunoprecipitation use antibody-directed cleavage/tagmentation, such as in single-cell CUT&TAG[79] or CoBATCH.[71]

The sc-seq libraries thus generated are sequenced by short-read sequencing technologies (figure 1J) or less often by long-read sequencing methods (figure 1K) for read lengths of tens of kilo bases to up to a few million bases, at the expense sequencing accuracy, duration or cost.[77 80 81] One of the primary challenges in sc-seq library preparation is the imperfect molecular capture efficiency, whereby only a subset of molecules in each cell is represented in the sequencing library. Referred to as 'dropout' events, this can result in bias, misclassification of cell types or batch effects. Further advances in library preparation as well as data analysis strategies are to be expected in the near future, resolving most of these limitations.

Recording additional cellular metadata

In addition to the profiling of the four omics modalities, recent advances also allow the recording of other cellular metadata. Technologies, such as sci-space,[6] transcriptome in vivo analysis[82] and in situ sequencing,[83 84] enable recording of spatial location of each cell, which is lost in the cell suspension-based library preparation described above. Preserving the spatial context can help

in understanding intercellular communication and its alteration in disease contexts by simultaneously describing the cellular subtypes and their physical proximities.[6] Despite limitations in single-cell resolution, techniques such as slide-(RNA-)seq,[85] slide-DNA-seq[86] and spatial-ATAC-seq[87] demonstrate the value of capturing spatial information in characterising diseases. For example, slide-DNA-seq of a cancer model revealed that clonal populations are confined to distinct spatial regions—a discovery not possible with mainstream bulk-seq or sc-seq. Indeed, a recent publication reported on the development of Stereo-seq, which uses a DNA nanoball technology to sequence the sc-transcriptome with submicrometre spatial resolution and map the mouse embryo development to generate a Mouse Organogenesis Spatiotemporal Transcriptomic Atlas.[11] Developmental cell lineage information can be computationally extracted from the omics-profiles in developmentally heterogeneous cellular populations[8 88–92] or it may be measured in individual cells by longitudinally 'recording the cell states'.[93] Niche techniques such as Patch-seq enable associating a neuron's electrophysiological profile with its expression profile. Thus, it is currently possible to profile single cells at many omics-modalities, in parallel and in association with cellular meta-features. Commercial offerings of such multi-omics sc-seq kits are also beginning to appear (Mission Bio and 10x Genomics), foreshadowing the possibility of being able to mix-and-match the modalities and cellular meta-features of interest at ease in the near future.

Considerations when performing sc-seq experiments

The goals of a particular project/experiment, the nature of the sample to be studied, budget, available expertise both in the laboratory and in analysis, available infrastructure, samples variability can all dictate the numerous choices that need to be made in the above described steps of the sc-seq workflow. Here, we summarise a few key considerations that can provide preliminary guidance to a reader planning an sc-seq experiment.

The choice of the sequencing modality is dictated by the goal of the study. However, given the abundance of methods (including commercial kits) and analysis pipelines that enable multimodal sc-seq experiment and analysis, studies may often benefit from the combined data (eg, epigenome, transcriptome and proteome), which has facilitated more accurate cell type annotation and establishing regulatory networks.[94] However, the trade off often includes increased costs and reduced data quality compared with single modality sc-seq. The sample, whether fresh, frozen or fixed, imposes further constraints on the choice of methodology. For example, tissue dissociation into single cells (as opposed to nuclei) is often only possible from fresh tissues. Moreover, the use of fixed cells/nuclei for sc-transcriptome-seq is only possible with a subset of technologies and may require further optimisation.[95 96] The choice of sequencing cells or nuclei can in turn impact downstream analyses, such as RNA velocity, which draws inferences from the difference between spliced and unspliced transcriptome.[97] In terms of time, cost and expertise required to establish and perform the assays, comercial end-to-end solutions often perform better than non-commercial solutions. Another crucial trade off, particularly in sc-transcriptome-seq, is between the number of cells sequenced versus the number of genes detected per cell (online supplemental figure 1). Many well-based methods are aimed at deeper sequencing of a smaller number of cells and combinatorial indexing-based methods fall on the other end of the spectrum, capable of sequencing millions of cells, although at a lower number of transcripts detected per cell. In this regard, the droplet based barcoding methods offer

middle ground. Online supplemental figure 1 also provides an overview of the cost of preparing the sequencing libraries per cell, where non-commercial methods often have the edge. However, when the additional sequencing costs are considered, commercial kits may be on par with the non end-to-end solutions.

Analysis and visualisation

The sc-seq produces large volumes of data (20–120 GB per sample),[98] which necessitates sophisticated computational and statistical techniques for analysis. After sample collection, processing and sequencing as described in the previous section (figure 2A), the sequenced reads are first demultiplexed using the cellular barcodes, tracing each read back to the cell-of-origin, thus effectively creating an omics-profile for each cell. The demultiplexed reads are aligned to the reference genome (sc-genome-seq, sc-epigenome-seq, sc-transcriptome-seq) or feature tag (sc-proteome-seq) and quantified. This is organised into a matrix in which each row contains the feature information and each column the cellular identity, forming the feature-barcode matrix. Due to biological variations, diverse microenvironments and the statistical nature of and inherent biases in the sc-seq experimental workflow (eg, PCR), no two cells, even of the same cell type, will have identical omics-profiles. Normalisation of the feature-barcode matrix makes the data comparable between cells by accounting for the variations, such as the sequencing depth, (PCR) amplification bias, the number of molecules identified and dropout events. At this stage, the feature-barcode matrix is ready for biological variance interpretation. Various computational tools are available for the analysis workflow.[99–104]

Clustering

Clustering refers to the grouping of cells with similar features, for example, accessible chromatin and methylation profiles in the case of epigenome sequencing, mutational landscape in the case of genome sequencing or expression profile in transcriptome or proteome sequencing (figure 2B). For clustering, the features (often several thousands) are condensed, using methods like principal component analysis (PCA), which reduce the data-dimensionality.[105] These condensed data are then clustered using methods like k-means, network clustering, which are adopted and implemented in the single-cell analysis workflows. The clustering results are commonly visualised in a further dimensionality reduced fashion, where the principal features (PCA components) are condensed into two or three dimensions, using methods—Uniform Manifold Approximation and Projection (McInnes,[106] Uniform Manifold Approximation and Projection for Dimension Reduction, 2018, http://arxiv.org/abs/1802.03426, accessed 5 July 2021) or t-distributed stochastic neighbour embedding.[107] Both these methods result in scatter plots, where each dot represents a cell and the cells in a cluster are characterised with a similar cellular profile. The visualisation methods are rapidly evolving and there is still ambiguity about how to infer the underlying biology from the high dimensional data.[108] Annotation of the cell type corresponding to each cluster is then carried out by identifying cluster-specific features (or differential features) that distinguish it from the other clusters.

Differential features

Differential feature analysis helps in identifying the source of variability between clusters as well as across samples. Within a sample, the differential feature analysis is typically used to identify cell types, states or clones based on the cluster-specific features. This is either performed manually based on the
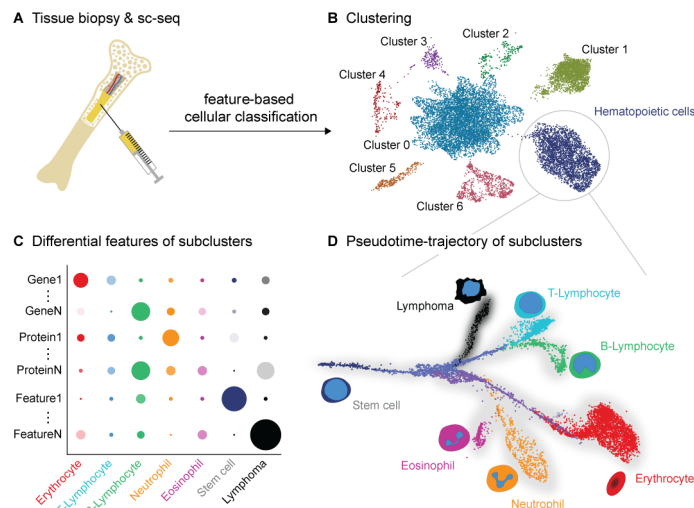
Figure 2 Workflow of single-cell sequencing (sc-seq) data analysis and visualisation. (A) The cell or nuclei suspension from a tissue biopsy is subjected to sc-seq of the desired modality(-ies). After data quality control, (B) cells are clustered and visualised in a map (eg, Uniform Manifold Approximation and Projection), where each cell is represented by a dot and coloured by its cluster identity. The coordinates of the map are arbitrary and represent non-linear reductions of the omics-features. Each cluster may be further subclustered. Cell type or state of each cluster or subcluster is identified by differential features analysis, that is, features which differentiates it from the others. (C) The differential features can be visualised using dot plots, where the size and opacity of the dots scale with the percentage of cells and the quantity at which a feature is present in a cluster. (D) Pseudotime-trajectories enable the visualisation of biological processes, including differentiation or disease development. Early identification of disease-forming, ectopic trajectories (eg, lymphoma, black) is one of the promises of the sc-seq technology. (A–D) Synthesised data, purely for illustrative purposes. (D) Represents a hypothetical scenario adapted from Rajewsky et al.[111]

literature on the identified cluster-specific features (eg, differential features or gene enrichment analysis) or using automated methods based on correlation-based, probabilistic or supervised machine learning models, as reviewed elsewhere.[109] Whereas across samples, it is used to compare the differences between, for example, the control and a diseased tissue, in order to establish genotype-phenotype relationships as discussed below. Thus, sc-seq is able to capture differences stratified into cell types, unlike bulk-seq where the differences are averaged over all cells and often miss smaller changes between cells. There are multiple methods and algorithms available for the differential feature analysis—the choice of the right method will reduce the false discovery rates.[110] Dot plots, heatmap and violin plots are the widely used visualisation methods to portray the differences in features like methylation status, mutational landscape, gene and protein expression between the cell types and also across samples. Figure 2C shows a dot plot, where the hypothetical subclusters of haematopoietic cells are compared to identify features that differentiate lymphoma cells from normal cells.

Trajectory analysis

The omics-profiles of cells even within a cluster (cell type) exhibit diversity, although at a finer level than between clusters. While experimental artefacts may account for some of this, it also contains biologically relevant information, such as their differentiation state, cell cycle state and somatic evolution. Trajectory analysis is an approach to organise cells within a cluster based on their progress along an expected/apparent biological process. The analysis assumes that certain features in these cells vary continuously along a gradient with the progression of biological processes. Therefore, these gradients are represented as branches, showing the minute differences in the state of the cells. Insightful information can often be gained from trajectory analyses and the so-called pseudotime plots, which assign a timestamp to each cell based on its progression along the feature gradient[103]

(figure 2D). In the case of genomic data, such an analysis can help trace the accumulation of somatic mutations within a cell type (eg, cancer) to study clonal evolution. Figure 2D shows the hypothetical organisation of cells from stem cells to mature haematopoietic cell types. Also depicted is a diseased lymphoma trajectory that deviates from normal lymphoid cells. Research initiatives, such as LifeTime, expect that such trajectory analyses can help in early disease detection by monitoring trajectories that deviate from a reference trajectory.[111] Another analysis that extracts temporal information from the static sc-transcriptome-seq data is RNA velocity. It captures the transcriptional dynamics in the cells, for example, a cell differentiating from a progenitor state to a mature state. It infers each cell's future state by comparing the proportion of unspliced versus spliced transcripts of every gene within a cell.[112]

WHAT CAN WE DO WITH SC-SEQ TECHNOLOGIES IN HUMAN GENETICS?

Current implementations of sc-seq in human genetics basic research

Human genetics is in part the identification of the causative reasons behind a diseased state. Before sc-seq, bulk-seq was often used for the molecular phenotyping of genetic disease by comparing the average omics-profiles of the diseased and the control samples. While this approach has been successful in some cases, the resolution of bulk-seq is too low to detect changes in cellular heterogeneity during development. With the advent of sc-seq, such investigations are being performed at a much higher resolution to understand the cellular-level changes pertaining to diseases. sc-genome-seq has been applied in studying genetic mosaicism,[26] aneuploidy[113] and clonal evolution.[114] As an example, sc-genome-seq in patients with acute myeloid leukaemia helped to detect significantly more mutations and also mutations exclusive to clones compared with the
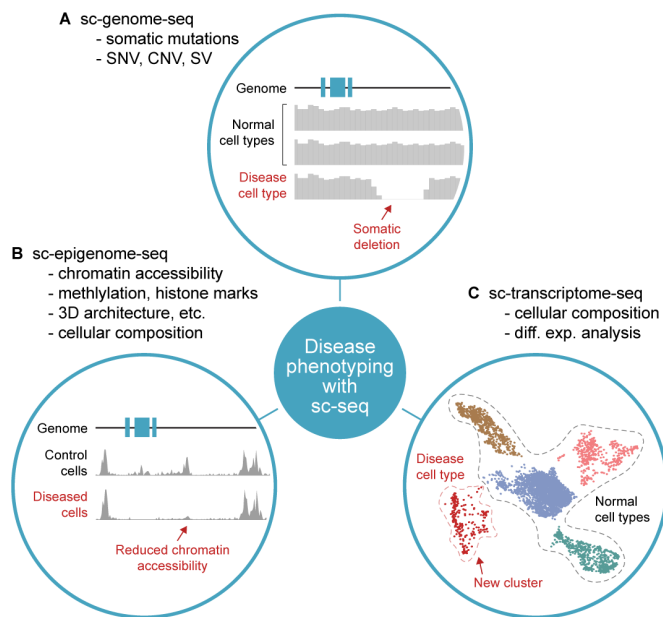
Figure 3 Current use cases of single-cell sequencing (sc-seq) in the field of human genetics. Sc-seq enables unbiased molecular phenotyping of diseases. (A) The mutational landscape leading to clonal evolution can be studied using sc-genome-seq. It provides insight into somatic mutations such as CNVs and single nucleotide variants (SNVs), particularly in cancer. (B) The activities of *cis*-regulatory elements and transcription factors related to disease progression can be explored using sc-epigenome-seq. (C) The cellular composition of a diseased tissue can be uncovered using sc-transcriptome-seq and sc-proteome-seq (not shown). They allow quantification of differential expression at both transcript and protein levels. (A–C) Synthesised data based on Smajić *et al*, Morita *et al* and Corces *et al*.[16 114 115]

bulk-genome-seq, because of its sensitivity to filter at a subpopulation level.[114] Sc-epigenome-seq and sc-transcriptome-seq have been demonstrated to track regulatory and transcriptional cellular heterogeneity, respectively, in patients and control profiles in various diseases,[16 115–121] and therapy strategies,[122 123] which would have been missed with bulk-seq. Sc-ATAC-seq of mitochondrial DNA revealed the kinetics in the levels of heteroplasmy in mitochondrial disorders across cell types, which were difficult to interpret from bulk-seq data.[124] Singh *et al*[125] combined short-read sequencing with targeted long-read sequencing of T cell receptor and B cell receptor at single-cell level. Towards this aim, the cDNA library was split before the usual fragmentation step required for short-read sequencing to allow parallel short-read and long-read sequencing on the same sample. This enabled them to characterise the isoforms due to alternative gene splicing. Various modalities and approaches of sc-seq discussed earlier can help uncover the complexity of diseases as well as monitor therapeutic effects (figure 3). The multiplexing of different sc-seq modalities lends itself for more specific questions to be addressed.[95 126 127] For example, linking gene regulators with the gene regulation is facilitated by combining sc-epigenome-seq with sc-transcriptome-seq.[95 128–131] This can enhance the interpretation of the regulatory pathways involved in development and maturation of cells in a disease perspective. Overall, sc-seq has established itself in human genetics as a tool to characterise diseases, where cellular heterogeneity (at any level of *the central dogma*) plays a role, as well as to disseminate regulatory mechanisms by taking advantage of the rich data. However, the application of the technology remains

confined to exploratory research settings and yet to make its way into routine clinical use.

Prospective applications in human genetics

Here, we describe two prospective applications in human genetics, which take advantage of the latest developments and promises of sc-seq to address current challenges in human genetics. The first relates to our quest to understand the non-coding genome. Specifically, the functional annotation of *cis*-regulatory elements. Here, sc-seq in combination with functional genomics allows thousands of functional experiments to be performed in a single, multiplexed assay (*prospective application 1*). The second addresses the challenges spawned by next-generation sequencing (NGS), namely the thousands of rare variants in each genome, and the vast majority of them being located in the >98% of the human genome that is non-coding.[132] Here, the richness of the sc-seq data combined with machine learning models help prioritise variants based on biological characteristics, such as gene expression and chromatin accessibility (*prospective application 2*).

Prospective application 1: single-cell functional genomics— annotating the non-coding genome

Delineating the function of the 98% of the genome that is non-coding and deciphering the pathogenicity of the variants identified in these non-coding regions is the central challenge of human genetics in the next decade. However, we are far from this goal, because: first, the 'regulatory code' of the non-coding sequence is still unknown; second, an important aspect of physiological gene regulation by *cis*-regulatory elements and enhancers lies in the 3D architecture of the genome and lastly, the number of non-coding variants per generation is so high that traditional functional tests have reached their limit. This underscores the urgent need for high-throughput functional screening technologies to study the endogenous functions of the non-coding genome. Until recently, annotating a non-coding region in a native context (cf. their identification in plasmids, eg, STARR-seq[133]) faced trade-offs between the number of regions simultaneously screened versus the complexity of phenotypes assayed.[134] On the one hand, pooled genetic screens, for example, with CRIPSR/Cas9, have mostly been associated with hypothesis-driven assays of simple phenotypes (eg, proliferation/survival screens, reporter expression).[135–138] This precludes the systematic investigation of the enhancer-driven phenotypes. On the other hand, the unbiased phenotype screening offered by sc-seq has been mostly applied to investigate a handful of mutations or diseases, but often in a non-multiplexed fashion.[139] Here, the effect of one or a few genomic perturbation(s) is assayed separately or at best in an arrayed format, which requires a separate experiment per enhancer and severely restricts scalability.

The integration of sc-seq with CRISPR/Cas9 has resulted in a new set of technologies that eliminate this trade-off, enabling the screening of non-coding (as well as coding) regions and their phenotypic consequences in a multiplexed fashion, at an unprecedented scale. Gasperini *et al* used this approach to functionally characterise ~6000 human candidate enhancers.[19] They used a pool of CRISPR/Cas9-based genetic perturbations to inactivate these candidate enhancers (CRISPRi) in a collection of ~250 000 cells, followed by sc-seq to measure the functional consequence in terms of expression >10 560 genes. In total, they could interrogate ~80 000 potential *cis*-regulatory relationships in a single experiment. The innovation of the technology is in the ability to identify the perturbation(s) present (ie, the
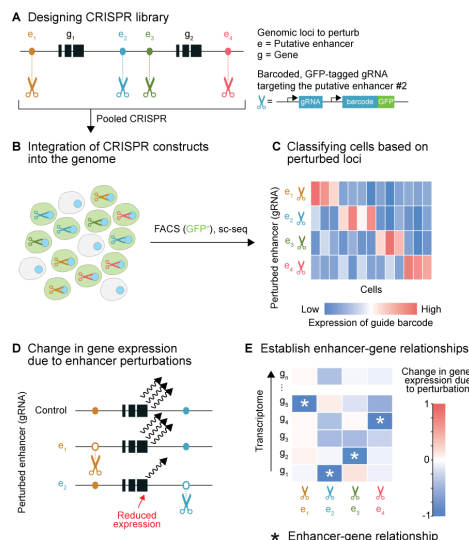
Figure 4 Single-cell sequencing (sc-seq) CRISPR screening of *cis*-regulatory elements. (A) To perturb the various (non-coding) genomic regions of interest, a CRISPR library is created by inserting guide RNAs (gRNAs) targeting these regions, along with barcodes unique to each gRNA, a fluorescence protein (eg, GFP - green fluorescent protein), and necessary promoters into a (lentiviral) vector. Perturbing non-coding regions, such as enhancers, followed by sc-transcriptome-seq can help establish enhancer-gene relationships. The method can also be used to reveal the functions of unannotated regions (not shown), which can help prioritise variants in that region. (B) Transfection/Transduction is followed by the integration of the gRNA library (and the CRISPR/Cas9 machinery; not shown) into the genome of the cells. Depending on the multiplicity of infection (MOI), the number of gRNAs (therefore, perturbations) per cell can be tuned. For simplicity, the figure depicts a maximum of one gRNA per cell. Multiple perturbations within a cell can be used to assess functional cooperativity between regulatory elements (eg, enhancer compensation) or to reduce the number of cells. FACS-sorting or selection for antibiotic resistance enables filtering out cells without any perturbation. (C) The identity and the number of gRNA barcodes per cell detected is identified from the sc-seq data. Cells (columns) can be ordered based on the perturbations they harbour for downstream analysis purposes. Note: each row would have traditionally been a separate experiment and each column would have been a sample or an experimental repeat, which can now be pooled into a single experiment. (D) Enhancer screening—the *cis*-regulatory functions of non-coding loci on the expression profiles of genes of interest can be investigated using sc-transcriptome-seq. Here, perturbation of the enhancer 2, results in the downregulation of gene 1. (E) The entire transcriptome can be assessed for changes in expression on individual (or cooperative) perturbations to establish enhancer-gene relationships. (A–E) Synthesised data based on Gasperini *et al* and Xie *et al*.[19 150]

enhancer targeted) in each cell from the sc-seq data, such that the effect of the perturbations on the omics-profile could be quantified. This is achieved by including a transcribable barcode (guide barcode, unique guide index or the gRNA sequence itself) along with every CRISPR guide RNA (gRNA), which gets integrated into the genome (figure 4A,B). Thus, the read-out of the guide barcode along with the rest of the cell's sequencing library by sc-seq, made it possible to associate the omics-profile of every cell to the perturbation present in that cell (figure 4C). In effect, the technology enabled the transformation of each perturbed cell into a patient harbouring (non-)coding variant(s), from an expression quantitative trait loci (eQTL) perspective.

The initial demonstrations of the technology (called Perturb-seq,[15 140] CRISP-seq[14] and CROP-seq[141]) were focused on elucidating gene functions (as opposed to non-coding regions). They perturbed the genes in a loss-of-function manner by implementing CRISPR/Cas9 in a knockout fashion (CRISPRko) and measured their *trans*-effects on the entire transcriptome. Dixit *et al* knocked out 24 transcription factors in bone marrow-derived dendritic cells and investigated the effect of these perturbations in the single-cell transcriptome. These initial studies screened tens of genes and assessed their trans-effect on the transcriptome of thousands of cells in a single experiment. Since these pioneering demonstrations, the technology has been optimised and adapted, for example, to improve the efficiency,[142] to study the effect of upregulation using CRISPR-activation,[143] to read out the effect of perturbations on the epigenome[144] or proteome[145] to dissect the function of protein domains (sc-Tiling),[146] as well as to elucidate gene functions[147] and gene regulatory networks related to development, diseases[148] or DNA-chromatin structure.[149]

Xie *et al* demonstrated that this approach can be used to measure the function of enhancers (figure 4D,E).[150] They used the CRISPR/dCas9-KRAB system and sgRNAs to epigenetically suppress the activity of 15 super-enhancers (containing a total of 71 constituent enhancers) in topologically associated domains containing highly expressed genes and quantified their function. They dissected the contribution of the individual constituent enhancers within these super-enhancers and concluded that often only a few constituent enhancers contribute to the regulation of target-gene expression. Moreover, by targeting multiple enhancers per cell (3.2 sgRNAs per cell on average), they could evaluate combinatorial enhancer activity and observed evidence of enhancer compensation. However, the design of this pioneering study was limited to quantifying the function of enhancers with known gene associations. Gasperini *et al* extended the approach with a goal to functionally characterise ~5779 candidate enhancers, where each perturbation was found in a median of 900 cells and each cell contained on average 28 gRNAs. They used an eQTL-inspired analysis framework to establish ~600 new enhancer-gene pairs. A similar use of the dCas9-KRAB system was recently demonstrated by Lopes *et al*[151] to evaluate the ~15000 putative regulatory elements engaged by oestrogen receptors in the context of breast cancer. By combining the sc-seq data, HiC maps and functional assays (cell proliferation), they could map the oestrogen receptor-driven oncogenic programme and decipher the role of the respective non-coding regions.

In the above-mentioned studies, the perturbations were introduced in a particular cell type followed by omics-profiling at a particular time-point or, at best, within a time-window. This could, however, lead to missing enhancer functions, because of their time-dependent and cell type-dependent activity. That is, the perturbation of an enhancer active during the development of a specific cell type would go unnoticed if it is perturbed in an unrelated cell type. To overcome this challenge, Jin *et al* applied the Perturb-Seq technology in utero, in the developing brain at E12.5, to study the function of a panel of 35 neurodevelopmental delay-related genes.[12] As a result, a wide variety of cell types, including neurons, microglia and oligodendrocytes, were targeted and the effect of the perturbations on the expression of 14 gene modules could be assessed. Another advantage of this approach was the feasibility to query the effect of the otherwise lethal genetic perturbations, since only 0.1% of the cells were perturbed at a time. Application of

this approach to screen the function of non-coding regions relevant to development or in an in vivo context is, however, yet to be demonstrated.

Taken together, it is currently possible to screen the effects of perturbing multiple coding or non-coding genomic regions on the transcriptome, epigenome or the proteome, with single-cell resolution in a single experiment. While most studies have used this multiplexed screening technology to evaluate the functions of genes, a handful of studies have already applied it to establish regulatory relationships. The technology can be performed in cultured cells or within a developing organism. Commercial kits are readily available from vendors such as 10x Genomics for similar experiments. In human genetics, the functional characterisation of unannotated genomic regions will assist in interpreting variants in these regions. Global efforts to annotate molecular and cellular phenotypes, such as the MorPhiC programme (National Institutes of Health, funding RFA-HG-21-029), would also benefit from the evolution of such technologies.

The next logical advance is the direct experimental screening of the variants, but this is just out of reach of the gene-editing technology, due to limited efficiency and accuracy. However, early demonstrations to profile single nucleotide variants do look promising.[152] Transduction-based overexpression instead of gene-editing has also been used as a method to overcome the limitations for functional annotation of variants.[147] As the gene-editing technologies improve in efficiency, resolution, accuracy and specificity,[153 154] it is increasingly likely that such technologies currently restricted to research will find direct applications in Personal Genomics for high-throughput experimental screening of variants identified in an individual. In the meanwhile, computational approaches combined with sc-seq data can be used to prioritise (if not annotate) variants, as is discussed next.

Prospective application 2: in silico variant prioritisation using sc-seq data

Genome-wide association studies (GWAS) and the widespread introduction of NGS technologies in medical genetics have led to a massive increase in the identification of common and rare variants, respectively.[155] Most of these variants fall into the non-coding genome. By far, not all these variants have an associated disease phenotype yet, and the experimental screening of variants is expensive, laborious and time-consuming. Currently, databases like Clinvar, HPO and OMIM are used to filter for known gene variants (figure 5A). Computation methods play a key role in the interpretation of these unknown variants, but current variant prioritisation methods, like deep-learning methods to prioritise non-coding variants,[156–159] CADD score,[160] SIFT[159] and several other methods[161] use bulk-seq data to rank these candidate variants based on highest disease-association probability. Sc-seq can enhance these methods by providing information at a cell type level rather than at a tissue level. This higher dimensionality information can enhance the interpretation of how subtle changes can lead to diseases. This section focuses on how machine learning models trained on sc-seq data can prioritise a given of variants found either through GWAS or whole genome sequencing.

Corces et al[115] and Trevino et al[162] showed that an sc-ATAC atlas of brain cells can be used to prioritise non-coding variants (figure 5B–D). Corces et al prioritised GWAS variants for Alzheimer's disease and Parkinson's disease by developing a machine learning model (gkm-SVM). They trained their model with sc-ATAC-seq data of adult human brains to predict the importance of each allele for chromatin accessibility. Trevino et al also used the sc-ATAC of developing human cerebral cortex, but to prioritise non-coding de novo mutations from patients with autism spectrum disorder from the Simons Simplex Collection.[163] They used BPNet,[164] which is a deep convolutional neural
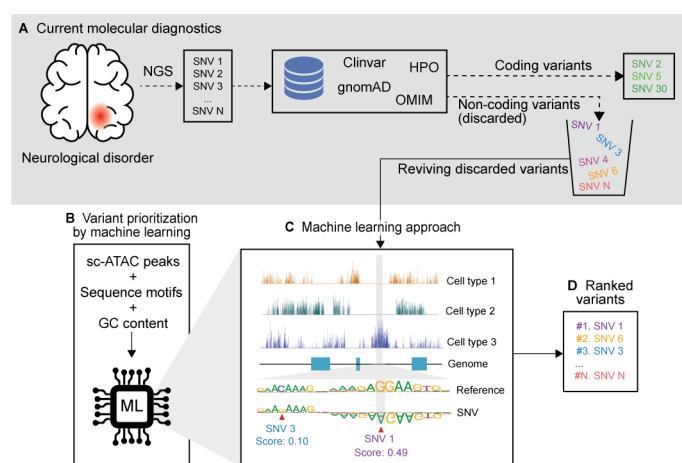

Figure 5 Variant prioritisation workflow with sc-ATAC data. (A) The current molecular diagnostic workflow starts with the next-generation sequencing (NGS) of a patient with a specific disease (eg, autism spectrum disorder), which identifies a large number of variants. Various databases are used to filter and rank the variants in the coding region. Meanwhile, a large number of variants in the non-coding regions are discarded, due to the lack of prioritisation methods. (B–D) Variant prioritisation with sc-seq data helps rank every variant pertinent to a cell type even if it is previously not known. (B) Supervised machine learning approaches are trained on sc-ATAC peaks, transcription factor sequence motifs and matched GC content of the peak and non-peak regions of a control sample. (C) sc-ATAC-seq data provide the machine learning model insights into the chromatin accessibility profile of all the cell types in the tissue and the sequence motifs inform the model on the allelic importance. Pathogenicity of unknown variants are predicted based on the disruption it causes to the accessibility of the loci. For example, the sequence motif shows the significance of the alleles which are accessible only in cell type 3. Single nucleotide variant (SNV)1 (G>A), therefore, causes a large disruption to the epigenome and thus receives a high score. Whereas SNV3 (C>G) is in the same accessible region, but the allele in the motif is not significant and the disruption it causes is low, hence it is scored low. (D) Variants are ranked based on the predicted pathogenicity scores. (A–D) Synthesised data based on Corces et al and Trevino et al.[115 162]

network model that predicts each transcription factor's per base binding signal as counts (ChIP-nexus).[163] Both methods were trained with cluster-specific sc-ATAC-seq peaks, transcription factor sequence motifs in the peak region, along with the regions where GC content of the peak region matches with the non-peak region to reduce the bias due to GC content in the prediction (figure 5B). This enabled them to predict an importance or disruption score based on the changes in the chromatin accessibility with allelic changes (figure 5C and D). This way, variants whether identified de novo or through GWAS approaches could be prioritised. Trevino et al tested their model with sc-ATAC-seq data of cell types from other organs like heart and noticed that there was no change in the case versus control mutations, signifying that the disease state was highly cell type specific. They were also able to predict the most frequently disrupted motifs in autism. Even though the conservation score and the distance to the gene were similar in case and control mutations, sc-ATAC was able to rank the pathogenic mutations, which would have been difficult to identify by other methods. Similar sc-ATAC-seq data-based approaches have been used to prioritise 527 GWAS variants in 48 diseases[5] and in type 2 diabetes.[165] Sc-seq data can also help elucidate the mechanisms in which GWAS variants affect haematopoiesis[166] and type 1 diabetes.[167]

With such applications being validated across various diseases and with the establishment of Human Cell Atlas of disease and healthy individuals, we can be hopeful that the in silico variant prioritisation methods using sc-seq data will evolve further to be able to rank the effect of rare de novo mutations for routine clinical diagnostics. Currently, however, these methods stop short of providing insights into the global mechanistic consequences of the ranked variants, especially when the variants are located in unannotated genomic regions. Methods such as sc-RNA-eQTL[168], sc-ATAC-eQTL[169] and single-cell functional genomic approaches (prospective application 1) can further enable functional annotation of the ranked variants by tracking cellular-level changes, which would be missed by bulk-seq, and help uncover holistic disease biology.

CURRENT LIMITATIONS

While the sc-seq technology and the ways in which it is implemented to address current challenges are advancing at a staggering pace in basic research, several aspects currently hinder their adoption in clinic practice, in particular the cost. Even if the cost was not an issue, several practicalities require that the diagnostics work in collaboration with a research lab specialised in single-cell sequencing, imposing additional barriers. For example, the experimental workflow, especially at the stages of tissue dissociation, requires case-by-case optimisation, depending on the tissue type, whether the tissue is fresh or frozen, as well as the age. Careful consideration of parameters such as sequencing depth and cost are required, especially for multimodal co-assays. As an example, the additional sample processing steps in sc-ATAC-seq can come at the expense of the transcriptome data quality when performing such co-assays. Even with commercial kits, generation of high-quality data may require adaptation or customisation of protocols, which adds to the time and cost. Enormous amounts of data are generated in all the single-cell sequencing approaches, posing data storage issues. The data are also information-rich, making it on the one hand versatile, but on the other hand complex. It necessitates sophisticated analysis to extract biological meaning and clinical significance. Dropout events are an ongoing issue in sc-seq. In sc-transcriptome-seq and sc-proteome-seq, dropouts occur when not all molecules are captured during library preparation.[170] In sc-genome-seq and sc-epigenome-seq, amplification imbalances cause allelic or site dropouts.[101] Dropout events result in the sparsity of the data, making the downstream analysis such as feature selection, dimensionality reduction and differential analysis strenuous. Another pitfall is batch effects between samples processed separately. While batch correction methods can be applied during data analysis, it may lead to the loss of true biological effects. Additionally, many of the analysis methods have not yet been fully standardised and require the judgement and interpretation of an experienced bioinformatician.[171] The field is growing at a rapid pace with many of these practical limitations soon to be addressed.

OUTLOOK

With increasing cellular throughput and the possibility to multiplex samples during the barcoding process, the drop in the cost per sample is expected to continue.[7 172] Computational methods are also evolving daily to overcome the challenges posed by the high data dimensionality, data sparsity due to dropout events, data variability between batches, etc. It is foreseeable that fully integrated and automated instruments will become commercially available over the next decade, streamlining all aspects of the single-cell sequencing resulting in a 'plug-n-play' workflow. As with all high-throughput in vitro technologies for variant testing, functional sc-seq approaches may also generate data of uncertain clinical significance or misclassify individual variants. Therefore, it is important to note that these data cannot stand alone but need to be integrated with all other available clinical and molecular information. The medical genetics community will have to develop frameworks on how to integrate high-throughput variant testing into clinical decision making. The technology is at the doorsteps of human genetics diagnostics. However, similar to whole genome and exome sequencing, it may take time until we see the routine use of these powerful sc-seq technologies in the clinic.

ORCID iDs
Varun K A Sreenivasan http://orcid.org/0000-0002-8667-8394
Malte Spielmann http://orcid.org/0000-0002-0583-4683

REFERENCES

1 Method of the year 2013. *Nat Methods* 2014;11:1.
2 Crespi S. Video: 2018's Breakthrough of the Year and runners-up. *Science* 2018.
3 Cao J, O'Day DR, Pliner HA, Kingsley PD, Deng M, Daza RM, Zager MA, Aldinger KA, Blecher-Gonen R, Zhang F, Spielmann M, Palis J, Doherty D, Steemers FJ, Glass IA, Trapnell C, Shendure J. A human cell atlas of fetal gene expression. *Science* 2020;370. doi:10.1126/science.aba7721. [Epub ahead of print: 13 11 2020].
4 Domcke S, Hill AJ, Daza RM, Cao J, O'Day DR, Pliner HA, Aldinger KA, Pokholok D, Zhang F, Milbank JH, Zager MA, Glass IA, Steemers FJ, Doherty D, Trapnell C, Cusanovich DA, Shendure J. A human cell atlas of fetal chromatin accessibility. *Science* 2020;370. doi:10.1126/science.aba7612
5 Zhang K, Hocker JD, Miller M, Hou X, Chiou J, Poirion OB, Qiu Y, Li YE, Gaulton KJ, Wang A, Preissl S, Ren B. A single-cell atlas of chromatin accessibility in the human genome. *Cell* 2021;184:5985–6001.
6 Srivatsan SR, Regier MC, Barkan E, Franks JM, Packer JS, Grosjean P, Duran M, Saxton S, Ladd JJ, Spielmann M, Lois C, Lampe PD, Shendure J, Stevens KR, Trapnell C. Embryo-scale, single-cell spatial transcriptomics. *Science* 2021;373:111–7.
7 Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, Zhang F, Mundlos S, Christiansen L, Steemers FJ, Trapnell C, Shendure J. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* 2019;566:496–502.
8 Farrell JA, Wang Y, Riesenfeld SJ, Shekhar K, Regev A, Schier AF. Single-Cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* 2018;360. doi:10.1126/science.aar3131. [Epub ahead of print: 01 06 2018].
9 Briggs JA, Weinreb C, Wagner DE, Megason S, Peshkin L, Kirschner MW, Klein AM. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* 2018;360.
10 Qiu C, Cao J, Martin BK, Li T, Welsh IC, Srivatsan S, Huang X, Calderon D, Noble WS, Disteche CM, Murray SA, Spielmann M, Moens CB, Trapnell C, Shendure J. Systematic reconstruction of cellular trajectories across mouse embryogenesis. *Nat Genet* 2022;54:328–41.
11 Chen A, Liao S, Cheng M, Ma K, Wu L, Lai Y, Qiu X, Yang J, Xu J, Hao S, Wang X, Lu H, Chen X, Liu X, Huang X, Li Z, Hong Y, Jiang Y, Peng J, Liu S, Shen M, Liu C, Li Q, Yuan Y, Wei X, Zheng H, Feng W, Wang Z, Liu Y, Wang Z, Yang Y, Xiang H, Han L, Qin B, Guo P, Lai G, Muñoz-Cánoves P, Maxwell PH, Thiery JP, Wu Q-F, Zhao F, Chen B, Li M, Dai X, Wang S, Kuang H, Hui J, Wang L, Fei J-F, Wang O, Wei X, Lu H, Wang B, Liu S, Gu Y, Ni M, Zhang W, Mu F, Yin Y, Yang H, Lisby M, Cornall RJ, Mulder J, Uhlén M, Esteban MA, Li Y, Liu L, Xu X, Wang J. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* 2022;185:1777–92.
12 Jin X, Simmons SK, Guo A, Shetty AS, Ko M, Nguyen L, Jokhi V, Robinson E, Oyler P, Curry N, Deangeli G, Lodato S, Levin JZ, Regev A, Zhang F, Arlotta P. In vivo Perturb-Seq reveals neuronal and glial abnormalities associated with autism risk genes. *Science* 2020;370. doi:10.1126/science.aaz6063. [Epub ahead of print: 27 11 2020].
13 Ren X, Kang B, Zhang Z. Understanding tumor ecosystems by single-cell sequencing: promises and limitations. *Genome Biol* 2018;19:1–14.
14 Jaitin DA, Weiner A, Yofe I, Lara-Astiaso D, Keren-Shaul H, David E, Salame TM, Tanay A, van Oudenaarden A, Amit I. Dissecting immune circuits by linking CRISPR-Pooled screens with single-cell RNA-seq. *Cell* 2016;167:1883–96.
15 Adamson B, Norman TM, Jost M, Cho MY, Nuñez JK, Chen Y, Villalta JE, Gilbert LA, Horlbeck MA, Hein MY, Pak RA, Gray AN, Gross CA, Dixit A, Parnas O, Regev A, Weissman JS. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* 2016;167:1867–82.
16 Smajić S, Prada-Medina CA, Landoulsi Z, Ghelfi J, Delcambre S, Dietrich C, Jarazo J, Henck J, Balachandran S, Pachchek S, Morris CM, Antony P, Timmermann B, Sauer S, Pereira SL, Schwamborn JC, May P, Grünewald A, Spielmann M. Single-Cell sequencing of human midbrain reveals glial activation and a Parkinson-specific neuronal state. 2022;145:964–78.
17 Nomura S, Satoh M, Fujita T, Higo T, Sumida T, Ko T, Yamaguchi T, Tobita T, Naito AT, Ito M, Fujita K, Harada M, Toko H, Kobayashi Y, Ito K, Takimoto E, Akazawa H, Morita H, Aburatani H, Komuro I. Cardiomyocyte gene programs encoding morphological and functional signatures in cardiac hypertrophy and failure. *Nat Commun* 2018;9:1–17.
18 Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of medical genetics and genomics and the association for molecular pathology. *Genet Med* 2015;17:405–24.
19 Gasperini M, Hill AJ, McFaline-Figueroa JL, Martin B, Kim S, Zhang MD, Jackson D, Leith A, Schreiber J, Noble WS, Trapnell C, Ahituv N, Shendure J. A genome-wide framework for mapping gene regulation via cellular genetic screens. *Cell* 2019;176:1516.
20 Kamath T, Abdulraouf A, Burris SJ, Langlieb J, Gazestani V, Nadaf NM, Balderrama K, Vanderburg C, Macosko EZ. Single-Cell genomic profiling of human dopamine neurons identifies a population that selectively degenerates in Parkinson's disease. *Nat Neurosci* 2022;25:588–95.
21 Dries R, Chen J, Del Rossi N, Khan MM, Sistig A, Yuan G-C. Advances in spatial transcriptomic data analysis. *Genome Res* 2021;31:1706–18.
22 Spitzer MH, Nolan GP. Mass cytometry: single cells, many features. *Cell* 2016;165:780–91.
23 Auerbach BJ, Hu J, Reilly MP, Li M. Applications of single-cell genomics and computational strategies to study common disease and population-level variation. *Genome Res* 2021;31:1728–41.
24 Rz R, Ran RZ. Preparation of single cell suspensions from human intestinal biopsies for single cell genomics applications V.1. *Protocols.io* 2020.
25 Tanay A, Regev A. Scaling single-cell genomics from phenomenology to mechanism. *Nature* 2017;541:331–8.
26 Gawad C, Koh W, Quake SR. Single-Cell genome sequencing: current state of the science. *Nat Rev Genet* 2016;17:175–88.
27 Clark SJ, Lee HJ, Smallwood SA, Kelsey G, Reik W. Single-Cell epigenomics: powerful new methods for understanding gene regulation and cell identity. *Genome Biol* 2016;17:72.
28 Shema E, Bernstein BE, Buenrostro JD. Single-Cell and single-molecule epigenomics to uncover genome regulation at unprecedented resolution. *Nat Genet* 2019;51:19–25.
29 Carter B, Zhao K. The epigenetic basis of cellular heterogeneity. *Nat Rev Genet* 2021;22:235–50.
30 Haque A, Engel J, Teichmann SA, Lönnberg T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med* 2017;9:75.
31 Hwang B, Lee JH, Bang D. Single-Cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med* 2018;50:1–14.
32 Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol* 2019;15:e8746.
33 Ziegenhain C, Vieth B, Parekh S, Reinius B, Smets M, Leonhardt H, Hellmann I, Enard W. Comparative analysis of single-cell RNA sequencing methods. *Mol Syst Biol*
34 Zappia L, Theis FJ. Over 1000 tools reveal trends in the single-cell RNA-seq analysis landscape. *Genome Biol* 2021;22:301.
35 Kashima Y, Sakamoto Y, Kaneko K, Seki M, Suzuki Y, Suzuki A. Single-Cell sequencing techniques from individual to multiomics analyses. *Exp Mol Med* 2020;52:1419–27.
36 Lee J, Hyeon DY, Hwang D. Single-Cell multiomics: technologies and data analysis methods. *Exp Mol Med* 2020;52:1428–42.
37 Todorovic V. Single-Cell RNA-seq—now with protein. *Nat Methods* 2017;14:1028–9.
38 Lei Y, Tang R, Xu J, Wang W, Zhang B, Liu J, Yu X, Shi S. Applications of single-cell sequencing in cancer research: progress and perspectives. *J Hematol Oncol* 2021;14.
39 Paik DT, Cho S, Tian L, Chang HY, Wu JC. Single-Cell RNA sequencing in cardiovascular development, disease and medicine. *Nat Rev Cardiol* 2020;17:457–73.
40 Ofengeim D, Giagtzoglou N, Huh D, Zou C, Yuan J. Single-Cell RNA sequencing: unraveling the brain one cell at a time. *Trends Mol Med* 2017;23:563–76.
41 Svensson V, Vento-Tormo R, Teichmann SA. Exponential scaling of single-cell RNA-seq in the past decade. *Nat Protoc* 2018;13:599–604.
42 Stuart T, Satija R. Integrative single-cell analysis. *Nat Rev Genet* 2019;20:257–72.
43 Thomsen ER, Mich JK, Yao Z, Hodge RD, Doyle AM, Jang S, Shehata SI, Nelson AM, Shapovalova NV, Levi BP, Ramanathan S. Fixed single-cell transcriptomic characterization of human radial glia diversity. *Nat Methods* 2016;13:87–93.
44 Fawkner-Corbett D, Antanaviciute A, Parikh K, Jagielowicz M, Gerós AS, Gupta T, Ashley N, Khamis D, Fowler D, Morrissey E, Cunningham C, Johnson PRV, Koohy H, Simmons A. Spatiotemporal analysis of human intestinal development at single-cell resolution. *Cell* 2021;184:810–26.
45 Slyper M, Porter CBM, Ashenberg O, Waldman J, Drokhlyansky E, Wakiro I, Smillie C, Smith-Rosario G, Wu J, Dionne D, Vigneau S, Jané-Valbuena J, Tickle TL, Napolitano S, Su M-J, Patel AG, Karlstrom A, Gritsch S, Nomura M, Waghray A, Gohil SH, Tsankov AM, Jerby-Arnon L, Cohen O, Klughammer J, Rosen Y, Gould J, Nguyen L, Hofree M, Tramontozzi PJ, Li B, Wu CJ, Izar B, Haq R, Hodi FS, Yoon CH, Hata AN, Baker SJ, Suvà ML, Bueno R, Stover EH, Clay MR, Dyer MA, Collins NB, Matulonis UA, Wagle N, Johnson BE, Rotem A, Rozenblatt-Rosen O, Regev A. A single-cell and single-nucleus RNA-seq toolbox for fresh and frozen human tumors. *Nat Med* 2020;26:792–802.
46 Denisenko E, Guo BB, Jones M, Hou R, de Kock L, Lassmann T, Poppe D, Clément O, Simmons RK, Lister R, Forrest ARR. Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genome Biol* 2020;21:130.
47 Grindberg RV, Yee-Greenbaum JL, McConnell MJ, Novotny M, O'Shaughnessy AL, Lambert GM, Araúzo-Bravo MJ, Lee J, Fishman M, Robbins GE, Lin X, Venepally P, Badger JH, Galbraith DW, Gage FH, Lasken RS. RNA-Sequencing from single nuclei. *Proc Natl Acad Sci U S A* 2013;110:19802–7.
48 Habib N, Avraham-Davidi I, Basu A, Burks T, Shekhar K, Hofree M, Choudhury SR, Aguet F, Gelfand E, Ardlie K, Weitz DA, Rozenblatt-Rosen O, Zhang F, Regev

A. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat Methods* 2017;14:955–8.

49 Preissl S, Fang R, Huang H, Zhao Y, Raviram R, Gorkin DU, Zhang Y, Sos BC, Afzal V, Dickel DE, Kuan S, Visel A, Pennacchio LA, Zhang K, Ren B. Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nat Neurosci* 2018;21:432–9.

50 Rozenblatt-Rosen O, Stubbington MJT, Regev A, Teichmann SA. The human cell atlas: from vision to reality. *Nature* 2017;550:451–3.

51 Hashimshony T, Wagner F, Sher N, Yanai I. CEL-Seq: single-cell RNA-seq by multiplexed linear amplification. *Cell Rep* 2012;2:666–73.

52 Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, Lao K, Surani MA, Azim Surani M. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* 2009;6:377–82.

53 Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA, Kirschner MW. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 2015;161:1187–201.

54 Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 2015;161:1202.

55 Hochgerner H, Lönnerberg P, Hodge R, Mikes J, Heskol A, Hubschle H, Lin P, Picelli S, La Manno G, Ratz M, Dunne J, Husain S, Lein E, Srinivasan M, Zeisel A, Linnarsson S. STRT-seq-2i: dual-index 5′ single cell and nucleus RNA-seq on an addressable microwell array. *Sci Rep* 2017;7:1–8.

56 Gierahn TM, Wadsworth MH, Hughes TK, Bryson BD, Butler A, Satija R, Fortune S, Love JC, Shalek AK. Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat Methods* 2017;14:395–8.

57 Yuan J, Sims PA. An automated microwell platform for large-scale single cell RNA-seq. *Sci Rep* 2016;6:1–10.

58 Cusanovich DA, Daza R, Adey A, Pliner HA, Christiansen L, Gunderson KL, Steemers FJ, Trapnell C, Shendure J. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* 2015;348:910–4.

59 Rosenberg AB, Roco CM, Muscat RA, Kuchina A, Sample P, Yao Z, Graybuck LT, Peeler DJ, Mukherjee S, Chen W, Pun SH, Sellers DL, Tasic B, Seelig G. Single-Cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* 2018;360:176–82.

60 Cao J, Cusanovich DA, Ramani V, Aghamirzaie D, Pliner HA, Hill AJ, Daza RM, McFaline-Figueroa JL, Packer JS, Christiansen L, Steemers FJ, Adey AC, Trapnell C, Shendure J. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* 2018;361:1380–5.

61 Zhang X, Li T, Liu F, Chen Y, Yao J, Li Z, Huang Y, Wang J. Comparative analysis of Droplet-Based Ultra-High-Throughput single-cell RNA-seq systems. *Mol Cell* 2019;73:130–42.

62 Islam S, Kjällquist U, Moliner A, Zajac P, Fan J-B, Lönnerberg P, Linnarsson S. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res* 2011;21:1160–7.

63 Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, Cook K, Stepansky A, Levy D, Esposito D, Muthuswamy L, Krasnitz A, McCombie WR, Hicks J, Wigler M. Tumour evolution inferred by single-cell sequencing. *Nature* 2011;472:90–4.

64 Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, Satija R, Smibert P. Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* 2017;14:865–8.

65 Shahi P, Kim SC, Haliburton JR, Gartner ZJ, Abate AR. Abseq: Ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding. *Sci Rep* 2017;7:1–12.

66 Peterson VM, Zhang KX, Kumar N, Wong J, Li L, Wilson DC, Moore R, McClanahan TK, Sadekova S, Klappenbach JA. Multiplexed quantification of proteins and transcripts in single cells. *Nat Biotechnol* 2017;35:936–9.

67 Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, Chang HY, Greenleaf WJ. Single-Cell chromatin accessibility reveals principles of regulatory variation. *Nature* 2015;523:486–90.

68 Smallwood SA, Lee HJ, Angermueller C, Krueger F, Saadeh H, Peat J, Andrews SR, Stegle O, Reik W, Kelsey G. Single-Cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat Methods* 2014;11:817–20.

69 Mulqueen RM, Pokholok D, Norberg SJ, Torkenczy KA, Fields AJ, Sun D, Sinnamon JR, Shendure J, Trapnell C, O'Roak BJ, Xia Z, Steemers FJ, Adey AC. Highly scalable generation of DNA methylation profiles in single cells. *Nat Biotechnol* 2018;36:428–31.

70 Guo H, Zhu P, Wu X, Li X, Wen L, Tang F. Single-Cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Res* 2013;23:2126–35.

71 Wang Q, Xiong H, Ai S, Yu X, Liu Y, Zhang J, He A. CoBATCH for high-throughput single-cell epigenomic profiling. *Mol Cell* 2019;76:206–16.

72 Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, Laue ED, Tanay A, Fraser P. Single-Cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 2013;502:59–64.

73 Rotem A, Ram O, Shoresh N, Sperling RA, Goren A, Weitz DA, Bernstein BE. Single-Cell ChIP-Seq reveals cell subpopulations defined by chromatin state. *Nat Biotechnol* 2015;33:1165–72.

74 Swanson E, Lord C, Reading J, Heubeck AT, Savage AK, Green R, Li X-J TTR, Bumol TF, Graybuck LT, Skene PJ. TEA-seq: a trimodal assay for integrated single cell measurement of transcription, epitopes, and chromatin accessibility. *bioRxiv*.

75 Chen S, Lake BB, Zhang K. High-Throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat Biotechnol* 2019;37:1452–7.

76 Wang Y, Yuan P, Yan Z, Yang M, Huo Y, Nie Y, Zhu X, Qiao J, Yan L. Single-Cell multiomics sequencing reveals the functional regulatory landscape of early embryos. *Nat Commun* 2021;12:1247.

77 Fan X, Yang C, Li W, Bai X, Zhou X, Xie H, Wen L, Tang F. SMOOTH-seq: single-cell genome sequencing of human cells on a third-generation sequencing platform. *Genome Biol* 2021;22:195.

78 Jin W, Tang Q, Wan M, Cui K, Zhang Y, Ren G, Ni B, Sklar J, Przytycka TM, Childs R, Levens D, Zhao K. Genome-Wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. *Nature* 2015;528:142–6.

79 Bartosovic M, Kabbe M, Castelo-Branco G. Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. *Nat Biotechnol* 2021;39:825–35.

80 Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol* 2020;21:30.

81 Payne A, Holmes N, Rakyan V, Loose M. BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics* 2019;35:2193–8.

82 Lovatt D, Ruble BK, Lee J, Dueck H, Kim TK, Fisher S, Francis C, Spaethling JM, Wolf JA, Grady MS, Ulyanova AV, Yeldell SB, Griepenburg JC, Buckley PT, Kim J, Sul J-Y, Dmochowski IJ, Eberwine J. Transcriptome in vivo analysis (TIVA) of spatially defined single cells in live tissue. *Nat Methods* 2014;11:190–6.

83 Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Ferrante TC, Terry R, Turczyk BM, Yang JL, Lee HS, Aach J, Zhang K, Church GM. Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat Protoc* 2015;10:442–58.

84 Wang X, Allen WE, Wright MA, Sylwestrak EL, Samusik N, Vesuna S, Evans K, Liu C, Ramakrishnan C, Liu J, Nolan GP, Bava F-A, Deisseroth K. Three-Dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* 2018;361. doi:10.1126/science.aat5691. [Epub ahead of print: 27 07 2018].

85 Rodriques SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, Welch J, Chen LM, Chen F, Macosko EZ. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* 2019;363:1463–7.

86 Zhao T, Chiang ZD, Morriss JW, LaFave LM, Murray EM, Del Priore I, Meli K, Lareau CA, Nadaf NM, Li J, Earl AS, Macosko EZ, Jacks T, Buenrostro JD, Chen F. Spatial genomics enables multi-modal study of clonal heterogeneity in tissues. *Nature* 2022;601:85–91.

87 Deng Y, Bartosovic M, Ma S, Zhang D, Liu Y, Qin X, Su G, ML X, Halene S, Craft JE, Castelo-Branco G, Fan R. Spatial-ATAC-seq: spatially resolved chromatin accessibility profiling of tissues at genome scale and cellular level. *bioRxiv*.

88 Spanjaard B, Hu B, Mitic N, Olivares-Chauvet P, Janjuha S, Ninov N, Junker JP. Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nat Biotechnol* 2018;36:469–73.

89 Kim IS, Wu J, Rahme GJ, Battaglia S, Dixit A, Gaskell E, Chen H, Pinello L, Bernstein BE. Parallel single-cell RNA-seq and genetic recording reveals lineage decisions in developing embryoid bodies. *Cell Rep* 2020;33:108222.

90 Biddy BA, Kong W, Kamimoto K, Guo C, Waye SE, Sun T, Morris SA. Single-Cell mapping of lineage and identity in direct reprogramming. *Nature* 2018;564:219–24.

91 Raj B, Wagner DE, McKenna A, Pandey S, Klein AM, Shendure J, Gagnon JA, Schier AF. Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat Biotechnol* 2018;36:442–50.

92 Ludwig LS, Lareau CA, Ulirsch JC, Christian E, Muus C, Li LH, Pelka K, Ge W, Oren Y, Brack A, Law T, Rodman C, Chen JH, Boland GM, Hacohen N, Rozenblatt-Rosen O, Aryee MJ, Buenrostro JD, Regev A, Sankaran VG. Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. *Cell* 2019;176:1325–39.

93 Chen W, Guillaume-Gentil O, Dainese R, Rainer PY, Zachara M, Gäbelein CG, Vorholt JA, Deplancke B. Genome-Wide molecular recording using Live-seq. *bioRxiv* 2021.

94 Chen AF, Parks B, Kathiria AS, Ober-Reynolds B, Goronzy JJ, Greenleaf WJ. NEAT-seq: simultaneous profiling of intra-nuclear proteins, chromatin accessibility and gene expression in single cells. *Nat Methods* 2022;19:547–53.

95 Mimitou EP, Lareau CA, Chen KY, Zorzetto-Fernandes AL, Hao Y, Takeshima Y, Luo W, Huang T-S, Yeung BZ, Papalexi E, Thakore PI, Kibayashi T, Wing JB, Hata M, Satija R, Nazor KL, Sakaguchi S, Ludwig LS, Sankaran VG, Regev A, Smibert P. Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. *Nat Biotechnol* 2021;39:1246–58.

96 Gerlach JP, van Buggenum JAG, Tanis SEJ, Hogeweg M, Heuts BMH, Muraro MJ, Elze L, Rivello F, Rakszewska A, van Oudenaarden A, Huck WTS, Stunnenberg HG, Mulder KW. Combined quantification of intracellular (phospho-)proteins and transcriptomics from fixed single cells. *Sci Rep* 2019;9:1469.

97 Bergen V, Soldatov RA, Kharchenko PV, Theis FJ. RNA velocity-current challenges and future perspectives. *Mol Syst Biol* 2021;17:e10282.

98 Angerer P, Simon L, Tritschler S, Wolf FA, Fischer D, Theis FJ. Single cells make big data: new challenges and opportunities in transcriptomics. *Curr Opin Syst Biol* 2017;4:85–91.

99 Andrews TS, Kiselev VY, McCarthy D, Hemberg M. Tutorial: guidelines for the computational analysis of single-cell RNA sequencing data. *Nat Protoc* 2021;16:1–9.

100 Fang R, Preissl S, Li Y, Hou X, Lucero J, Wang X, Motamedi A, Shiau AK, Zhou X, Xie F, Mukamel EA, Zhang K, Zhang Y, Margarita Behrens M, Ecker JR, Ren B. SnapATAC: a comprehensive analysis package for single cell ATAC-seq. *bioRxiv* 2020;615179.

101 Luquette LJ, Bohrson CL, Sherman MA, Park PJ. Identification of somatic mutations in single cell DNA-seq using a spatial model of allelic imbalance. *Nat Commun* 2019;10:1–14.

102 Mimitou EP, Cheng A, Montalbano A, Hao S, Stoeckius M, Legut M, Roush T, Herrera A, Papalexi E, Ouyang Z, Satija R, Sanjana NE, Koralov SB, Smibert P. Expanding the CITE-seq tool-kit: detection of proteins, transcriptomes, clonotypes and CRISPR perturbations with multiplexing, in a single assay. *Nat Methods* 2019;16:409.

103 Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, Trapnell C. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* 2017;14:979–82.

104 Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, Purdom E, Dudoit S. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* 2018;19:477.

105 Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublomme JT, Raychowdhury R, Schwartz S, Yosef N, Malboeuf C, Lu D, Trombetta JJ, Gennert D, Gnirke A, Goren A, Hacohen N, Levin JZ, Park H, Regev A. Single-Cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 2013;498:236–40.

106 McInnes L, Healy J, Melville J. UMAP: uniform manifold approximation and projection for dimension reduction 2018 https://arxiv.org/abs/1802.03426

107 van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res* 2008;9:2579–605.

108 Chari T, Banerjee J, Pachter L. The Specious art of single-cell genomics. *bioRxiv* 2021.

109 Pasquini G, Rojo Arias JE, Schäfer P, Busskamp V. Automated methods for cell type annotation on scRNA-seq data. *Comput Struct Biotechnol J* 2021;19:961–9.

110 Squair JW, Gautier M, Kathe C, Anderson MA, James ND, Hutson TH, Hudelle R, Qaiser T, Matson KJE, Barraud Q, Levine AJ, La Manno G, Skinnider MA, Courtine G. Confronting false discoveries in single-cell differential expression. *Nat Commun* 2021;12:1–15.

111 Rajewsky N, Almouzni G, Gorski SA, Aerts S, Amit I, Bertero MG, Bock C, Bredenoord AL, Cavalli G, Chiocca S, Clevers H, De Strooper B, Eggert A, Ellenberg J, Fernández XM, Figlerowicz M, Gasser SM, Hubner N, Kjems J, Knoblich JA, Krabbe G, Lichter P, Linnarsson S, Marine J-C, Marioni JC, Marti-Renom MA, Netea MG, Nickel D, Nollmann M, Novak HR, Parkinson H, Piccolo S, Pinheiro I, Pombo A, Popp C, Reik W, Roman-Roman S, Rosenstiel P, Schultze JL, Stegle O, Tanay A, Testa G, Thanos D, Theis FJ, Torres-Padilla M-E, Valencia A, Vallot C, van Oudenaarden A, Vidal M, Voet T, LifeTime Community Working Groups. Lifetime and improving European healthcare through cell-based interceptive medicine. *Nature* 2020;587:377–86.

112 La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K, Kastriti ME, Lönnerberg P, Furlan A, Fan J, Borm LE, Liu Z, van Bruggen D, Guo J, He X, Barker R, Sundström E, Castelo-Branco G, Cramer P, Adameyko I, Linnarsson S, Kharchenko PV. RNA velocity of single cells. *Nature* 2018;560:494–8.

113 Cai X, Evrony GD, Lehmann HS, Elhosary PC, Mehta BK, Poduri A, Walsh CA. Single-Cell, genome-wide sequencing identifies clonal somatic copy-number variation in the human brain. *Cell Rep* 2015;10:645.

114 Morita K, Wang F, Jahn K, Hu T, Tanaka T, Sasaki Y, Kuipers J, Loghavi S, Wang SA, Yan Y, Furudate K, Matthews J, Little L, Gumbs C, Zhang J, Song X, Thompson E, Patel KP, Bueso-Ramos CE, DiNardo CD, Ravandi F, Jabbour E, Andreeff M, Cortes J, Bhalla K, Garcia-Manero G, Kantarjian H, Konopleva M, Nakada D, Navin N, Beerenwinkel N, Futreal PA, Takahashi K. Clonal evolution of acute myeloid leukemia revealed by high-throughput single-cell genomics. *Nat Commun* 2020;11:5327.

115 Corces MR, Shcherbina A, Kundu S, Gloudemans MJ, Frésard L, Granja JM, Louie BH, Eulalio T, Shams S, Bagdatli ST, Mumbach MR, Liu B, Montine KS, Greenleaf WJ, Kundaje A, Montgomery SB, Chang HY, Montine TJ. Single-Cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat Genet* 2020;52:1158–68.

116 Rai V, Quang DX, Erdos MR, Cusanovich DA, Daza RM, Narisu N, Zou LS, Didion JP, Guan Y, Shendure J, Parker SCJ, Collins FS. Single-Cell ATAC-Seq in human pancreatic islets and deep learning upscaling of rare cells reveals cell-specific type 2 diabetes regulatory signatures. *Mol Metab* 2020;32:109–21.

117 Ziffra RS, Kim CN, Ross JM, Wilfert A, Turner TN, Haeussler M, Casella AM, Przytycki PF, Keough KC, Shin D, Bogdanoff D, Kreimer A, Pollard KS, Ament SA, Eichler EE, Ahituv N, Nowakowski TJ. Single-Cell epigenomics reveals mechanisms of human cortical development. *Nature* 2021;598:205–13.

118 Elmentaite R, Ross ADB, Roberts K, James KR, Ortmann D, Gomes T, Nayak K, Tuck L, Pritchard S, Bayraktar OA, Heuschkel R, Vallier L, Teichmann SA, Zilbauer M. Single-Cell sequencing of developing human gut reveals transcriptional links to childhood Crohn's disease. *Dev Cell* 2020;55:771–83.

119 Reynolds G, Vegh P, Fletcher J, Poyner EFM, Stephenson E, Goh I, Botting RA, Huang N, Olabi B, Dubois A, Dixon D, Green K, Maunder D, Engelbert J, Efremova M, Polański K, Jardine L, Jones C, Ness T, Horsfall D, McGrath J, Carey C, Popescu D-M, Webb S, Wang X-N, Sayer B, Park J-E, Negri VA, Belokhvostova D, Lynch MD, McDonald D, Filby A, Hagai T, Meyer KB, Husain A, Coxhead J, Vento-Tormo R, Behjati S, Lisgo S, Villani A-C, Bacardit J, Jones PH, O'Toole EA, Ogg GS, Rajan N, Reynolds NJ, Teichmann SA, Watt FM, Haniffa M. Developmental cell programs are co-opted in inflammatory skin disease. *Science* 2021;371. doi:10.1126/science. aba6500. [Epub ahead of print: 22 01 2021].

120 Jardine L, Webb S, Goh I, Quiroga Londoño M, Reynolds G, Mather M, Olabi B, Stephenson E, Botting RA, Horsfall D, Engelbert J, Maunder D, Mende N, Murnane C, Dann E, McGrath J, King H, Kucinski I, Queen R, Carey CD, Shrubsole C, Poyner E, Acres M, Jones C, Ness T, Coulthard R, Elliott N, O'Byrne S, Haltalli MLR, Lawrence JE, Lisgo S, Balogh P, Meyer KB, Prigmore E, Ambridge K, Jain MS, Efremova M, Pickard K, Creasey T, Bacardit J, Henderson D, Coxhead J, Filby A, Hussain K, Dixon D, McDonald D, Popescu D-M, Kowalczyk MS, Li B, Ashenberg O, Tabaka M, Dionne D, Tickle TL, Slyper M, Rozenblatt-Rosen O, Regev A, Behjati S, Laurenti E, Wilson NK, Roy A, Göttgens B, Roberts I, Teichmann SA, Haniffa M. Blood and immune development in human fetal bone marrow and Down syndrome. *Nature* 2021;598:327–31.

121 Mayr CH, Simon LM, Leuschner G, Ansari M, Schniering J, Geyer PE, Angelidis I, Strunz M, Singh P, Kneidinger N, Reichenberger F, Silbernagel E, Böhm S, Adler H, Lindner M, Maurer B, Hilgendorff A, Prasse A, Behr J, Mann M, Eickelberg O, Theis FJ, Schiller HB. Integrative analysis of cell state changes in lung fibrosis with peripheral protein biomarkers. *EMBO Mol Med* 2021;13:e12871.

122 Zhang L, Li Z, Skrzypczynska KM, Fang Q, Zhang W, O'Brien SA, He Y, Wang L, Zhang Q, Kim A, Gao R, Orf J, Wang T, Sawant D, Kang J, Bhatt D, Lu D, Li C-M, Rapaport AS, Perez K, Ye Y, Wang S, Hu X, Ren X, Ouyang W, Shen Z, Egen JG, Zhang Z, Yu X. Single-Cell analyses inform mechanisms of Myeloid-Targeted therapies in colon cancer. *Cell* 2020;181:442–59.

123 Aissa AF, Islam ABMMK, Ariss MM, Go CC, Rader AE, Conrardy RD, Gajda AM, Rubio-Perez C, Valyi-Nagy K, Pasquinelli M, Feldman LE, Green SJ, Lopez-Bigas N, Frolov MV, Benevolenskaya EV. Single-Cell transcriptional changes associated with drug tolerance and response to combination therapies in cancer. *Nat Commun* 2021;12:1628.

124 Walker MA, Lareau CA, Ludwig LS, Karaa A, Sankaran VG, Regev A, Mootha VK. Purifying selection against pathogenic mitochondrial DNA in human T cells. *N Engl J Med* 2020;383:1556–63.

125 Singh M, Al-Eryani G, Carswell S, Ferguson JM, Blackburn J, Barton K, Roden D, Luciani F, Giang Phan T, Junankar S, Jackson K, Goodnow CC, Smith MA, Swarbrick A. High-Throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. *Nat Commun* 2019;10:3120.

126 Vafadarnejad E, Rizzo G, Krampert L, Arampatzi P, Arias-Loza A-P, Nazzal Y, Rizakou A, Knochenhauer T, Bandi SR, Nugroho VA, Schulz DJJ, Roesch M, Alayrac P, Vilar J, Silvestre J-S, Zernecke A, Saliba A-E, Cochain C. Dynamics of cardiac neutrophil diversity in murine myocardial infarction. *Circ Res* 2020;127:e232–49.

127 Golomb SM, Guldner IH, Zhao A, Wang Q, Palakurthi B, Aleksandrovic EA, Lopez JA, Lee SW, Yang K, Zhang S. Multi-Modal single-cell analysis reveals brain immune landscape plasticity during aging and gut microbiota dysbiosis. *Cell Rep* 2020;33:108438.

128 Argelaguet R, Clark SJ, Mohammed H, Stapel LC, Krueger C, Kapourani C-A, Imaz-Rosshandler I, Lohoff T, Xiang Y, Hanna CW, Smallwood S, Ibarra-Soria X, Buettner F, Sanguinetti G, Xie W, Krueger F, Göttgens B, Rugg-Gunn PJ, Kelsey G, Dean W, Nichols J, Stegle O, Marioni JC, Reik W. Multi-Omics profiling of mouse gastrulation at single-cell resolution. *Nature* 2019;576:487–91.

129 Bleckwehl T, Crispatzu G, Schaaf K, Respuela P, Bartusel M, Benson L, Clark SJ, Dorighi KM, Barral A, Laugsch M, van IJcken WFJ, Manzanares M, Wysocka J, Reik W, Rada-Iglesias Álvaro. Enhancer-associated H3K4 methylation safeguards in vitro germline competence. *Nat Commun* 2021;12:1–19. doi:10.1038/s41467-021-26065-6

130 Lake BB, Chen S, Sos BC, Fan J, Kaeser GE, Yung YC, Duong TE, Gao D, Chun J, Kharchenko PV, Zhang K. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat Biotechnol* 2018;36:70-80.

131 Gaiti F, Chaligne R, Gu H, Brand RM, Kothen-Hill S, Schulman RC, Grigorev K, Risso D, Kim K-T, Pastore A, Huang KY, Alonso A, Sheridan C, Omans ND, Biederstedt E, Clement K, Wang L, Felsenfeld JA, Bhavsar EB, Aryee MJ, Allan JN, Furman R, Gnirke A, Wu CJ, Meissner A, Landau DA. Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia. *Nature* 2019;569:576–80.

132 Spielmann M, Lupiáñez DG, Mundlos S. Structural variation in the 3D genome. *Nat Rev Genet* 2018;19:453–67.

133 Arnold CD, Gerlach D, Stelzer C, Boryń Łukasz M, Rath M, Stark A. Genome-Wide quantitative enhancer activity maps identified by STARR-seq. *Science* 2013;339:1074–7.

134 Klein JC, Chen W, Gasperini M, Shendure J. Identifying novel enhancer elements with CRISPR-Based screens. *ACS Chem Biol* 2018;13:326–32.

135 Korkmaz G, Lopes R, Ugalde AP, Nevedomskaya E, Han R, Myacheva K, Zwart W, Elkon R, Agami R. Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat Biotechnol* 2016;34:192–8.

136 Li K, Liu Y, Cao H, Zhang Y, Gu Z, Liu X, Yu A, Kaphle P, Dickerson KE, Ni M, Xu J. Interrogation of enhancer function by enhancer-targeting CRISPR epigenetic editing. *Nat Commun* 2020;11:485.

137 Rajagopal N, Srinivasan S, Kooshesh K, Guo Y, Edwards MD, Banerjee B, Syed T, Emons BJM, Gifford DK, Sherwood RI. High-Throughput mapping of regulatory DNA. *Nat Biotechnol* 2016;34:167–74.

138 Fulco CP, Munschauer M, Anyoha R, Munson G, Grossman SR, Perez EM, Kane M, Cleary B, Lander ES, Engreitz JM. Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* 2016;354:769–73.

139 High-Content CRISPR screening. *Nature Reviews Methods Primers* 2022;2. doi:10.1038/s43586-021-00093-4

140 Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, Marjanovic ND, Dionne D, Burks T, Raychowdhury R, Adamson B, Norman TM, Lander ES, Weissman JS, Friedman N, Regev A. Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* 2016;167:1853–66.

141 Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Traxler P, Klughammer J, Schuster LC, Kuchler A, Alpar D, Bock C. Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods* 2017;14:297–301.

142 Replogle JM, Norman TM, Xu A, Hussmann JA, Chen J, Cogan JZ, Meer EJ, Terry JM, Riordan DP, Srinivas N, Fiddes IT, Arthur JG, Alvarado LJ, Pfeiffer KA, Mikkelsen TS, Weissman JS, Adamson B. Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat Biotechnol* 2020;38:954–61.

143 Alda-Catalinas C, Bredikhin D, Hernando-Herraez I, Santos F, Kubinyecz O, Eckersley-Maslin MA, Stegle O, Reik W. A single-cell transcriptomics CRISPR-Activation screen identifies epigenetic regulators of the zygotic genome activation program. *Cell Syst* 2020;11:25–41.

144 Rubin AJ, Parker KR, Satpathy AT, Qi Y, Wu B, Ong AJ, Mumbach MR, Ji AL, Kim DS, Cho SW, Zarnegar BJ, Greenleaf WJ, Chang HY, Khavari PA. Coupled single-cell CRISPR screening and epigenomic profiling reveals causal gene regulatory networks. *Cell* 2019;176:76.e17.

145 Frangieh CJ, Melms JC, Thakore PI, Geiger-Schuller KR, Ho P, Luoma AM, Cleary B, Jerby-Arnon L, Malu S, Cuoco MS, Zhao M, Ager CR, Rogava M, Hovey L, Rotem A, Bernatchez C, Wucherpfennig KW, Johnson BE, Rozenblatt-Rosen O, Schadendorf D, Regev A, Izar B. Multimodal pooled Perturb-CITE-seq screens in patient models define mechanisms of cancer immune evasion. *Nat Genet* 2021;53:332–41.

146 Yang L, Chan AKN, Miyashita K, Delaney CD, Wang X, Li H, Pokharel SP, Li S, Li M, Xu X, Lu W, Liu Q, Mattson N, Chen KY, Wang J, Yuan Y-C, Horne D, Rosen ST, Soto-Feliciano Y, Feng Z, Hoshii T, Xiao G, Müschen M, Chen J, Armstrong SA, Chen C-W. High-Resolution characterization of gene function using single-cell CRISPR tiling screen. *Nat Commun* 2021;12:4063.

147 Ursu O, Neal JT, Shea E, Thakore PI, Jerby-Arnon L, Nguyen L, Dionne D, Diaz C, Bauman J, Mosaad MM, Fagre C, Lo A, McSharry M, Giacomelli AO, Ly SH, Rozenblatt-Rosen O, Hahn WC, Aguirre AJ, Berger AH, Regev A, Boehm JS. Massively parallel phenotyping of coding variants in cancer with Perturb-seq. *Nat Biotechnol* 2022. doi:10.1038/s41587-021-01160-7. [Epub ahead of print: 20 Jan 2022].

148 Giladi A, Paul F, Herzog Y, Lubling Y, Weiner A, Yofe I, Jaitin D, Cabezas-Wallscheid N, Dress R, Ginhoux F, Trumpp A, Tanay A, Amit I. Single-Cell characterization of haematopoietic progenitors and their trajectories in homeostasis and perturbed haematopoiesis. *Nat Cell Biol* 2018;20:836–46.

149 Liscovitch-Brauer N, Montalbano A, Deng J, Méndez-Mancilla A, Wessels H-H, Moss NG, Kung C-Y, Sookdeo A, Guo X, Geller E, Jaini S, Smibert P, Sanjana NE. Profiling the genetic determinants of chromatin accessibility with scalable single-cell CRISPR screens. *Nat Biotechnol* 2021;39:1270–7.

150 Xie S, Duan J, Li B, Zhou P, Hon GC. Multiplexed engineering and analysis of combinatorial enhancer activity in single cells. *Mol Cell* 2017;66:285–99.

151 Lopes R, Sprouffske K, Sheng C, Uijttewaal ECH, Wesdorp AE, Dahinden J, Wengert S, Diaz-Miyar J, Yildiz U, Bleu M, Apfel V, Mermet-Meillon F, Krese R, Eder M, Olsen AV, Hoppe P, Knehr J, Carbone W, Cuttat R, Waldt A, Altorfer M, Naumann U, Weischenfeldt J, deWeck A, Kauffmann A, Roma G, Schübeler D, Galli GG. Systematic dissection of transcriptional regulatory networks by genome-scale and single-cell CRISPR screens. *Sci Adv* 2021;7. doi:10.1126/sciadv.abf5733. [Epub ahead of print: 02 07 2021].

152 Jun S, Lim H, Chun H, Lee JH, Bang D. Single-Cell analysis of a mutant library generated using CRISPR-guided deaminase in human melanoma cells. *Commun Biol* 2020;3.

153 Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 2016;533:420–4.

154 Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, Chen PJ, Wilson C, Newby GA, Raguram A, Liu DR. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 2019;576:149–57.

155 Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, Martin HC, Lappalainen T, Posthuma D. Genome-Wide association studies. *Nat Rev Methods Primers* 2021;1:1–21. doi:10.1038/s43586-021-00056-9

156 Avsec Žiga, Weilert M, Shrikumar A, Krueger S, Alexandari A, Dalal K, Fropf R, McAnany C, Gagneur J, Kundaje A, Zeitlinger J. Base-resolution models of transcription-factor binding reveal soft motif SYNTAX. *Nat Genet* 2021;53:354–66. doi:10.1038/s41588-021-00782-6

157 Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning-based sequence model. *Nat Methods* 2015;12:931–4.

158 Zhou J, Park CY, Theesfeld CL, Wong AK, Yuan Y, Scheckel C, Fak JJ, Funk J, Yao K, Tajima Y, Packer A, Darnell RB, Troyanskaya OG. Whole-Genome deep-learning analysis identifies contribution of noncoding mutations to autism risk. *Nat Genet* 2019;51:973–80. doi:10.1038/s41588-019-0420-0

159 Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003;31:3812–4. doi:10.1093/nar/gkg509

160 Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;46:310–5. doi:10.1038/ng.2892

161 Eilbeck K, Quinlan A, Yandell M. Settling the score: variant prioritization and Mendelian disease. *Nat Rev Genet* 2017;18:599–612. doi:10.1038/nrg.2017.52

162 Trevino AE, Müller F, Andersen J, Sundaram L, Kathiria A, Shcherbina A, Farh K, Chang HY, Pașca AM, Kundaje A, Pașca SP, Greenleaf WJ. Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution. *Cell* 2021;184:5053–69. doi:10.1016/j.cell.2021.07.039

163 An J-Y, Lin K, Zhu L, Werling DM, Dong S, Brand H, Wang HZ, Zhao X, Schwartz GB, Collins RL, Currall BB, Dastmalchi C, Dea J, Duhn C, Gilson MC, Klei L, Liang L, Markenscoff-Papadimitriou E, Pochareddy S, Ahituv N, Buxbaum JD, Coon H, Daly MJ, Kim YS, Marth GT, Neale BM, Quinlan AR, Rubenstein JL, Sestan N, State MW, Willsey AJ, Talkowski ME, Devlin B, Roeder K, Sanders SJ. Genome-Wide de novo risk score implicates promoter variation in autism spectrum disorder. *Science* 2018;362. doi:10.1126/science.aat6576. [Epub ahead of print: 14 12 2018].

164 Ž A, Weilert M, Shrikumar A, Krueger S, Alexandari A, Dalal K, Fropf R, McAnany C, Gagneur J, Kundaje A, Zeitlinger J. Base-resolution models of transcription-factor binding reveal soft motif SYNTAX. *Nat Genet* 2021;53:354–66.

165 Chiou J, Zeng C, Cheng Z, Han JY, Schlichting M, Miller M, Mendez R, Huang S, Wang J, Sui Y, Deogaygay A, Okino M-L, Qiu Y, Sun Y, Kudtarkar P, Fang R, Preissl S, Sander M, Gorkin DU, Gaulton KJ. Single-Cell chromatin accessibility identifies pancreatic islet cell type- and state-specific regulatory programs of diabetes risk. *Nat Genet* 2021;53:455–66. doi:10.1038/s41588-021-00823-0

166 Ulirsch JC, Lareau CA, Bao EL, Ludwig LS, Guo MH, Benner C, Satpathy AT, Kartha VK, Salem RM, Hirschhorn JN, Finucane HK, Aryee MJ, Buenrostro JD, Sankaran VG. Interrogation of human hematopoiesis at single-cell and single-variant resolution. *Nat Genet* 2019;51:683–93. doi:10.1038/s41588-019-0362-6

167 Chiou J, Geusz RJ, Okino M-L, Han JY, Miller M, Melton R, Beebe E, Benaglio P, Huang S, Korgaonkar K, Heller S, Kleger A, Preissl S, Gorkin DU, Sander M, Gaulton KJ. Interpreting type 1 diabetes risk with genetics and single-cell epigenomics. *Nature* 2021;594:398–402. doi:10.1038/s41586-021-03552-w

168 Yazar S, Alquicira-Hernandez J, Wing K, Senabouth A, Gordon MG, Andersen S, Lu Q, Rowson A, Taylor TRP, Clarke L, Maccora K, Chen C, Cook AL, Ye CJ, Fairfax KA, Hewitt AW, Powell JE. Single-Cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. *Science* 2022;376:eabf3041. doi:10.1126/science.abf3041

169 Örd T, Õunap K, Stolze LK, Aherrahrou R, Nurminen V, Toropainen A, Selvarajan I, Lönnberg T, Aavik E, Ylä-Herttuala S, Civelek M, Romanoski CE, Kaikkonen MU. Single-Cell epigenomics and functional fine-mapping of atherosclerosis GWAS loci. *Circ Res* 2021;129:240–58. doi:10.1161/CIRCRESAHA.121.318971

170 Hicks SC, Townes FW, Teng M, Irizarry RA. Missing data and technical variability in single-cell RNA-sequencing experiments. *Biostatistics* 2018;19:562–78.

171 Lähnemann D, Köster J, Szczurek E, McCarthy DJ, Hicks SC, Robinson MD, Vallejos CA, Campbell KR, Beerenwinkel N, Mahfouz A, Pinello L, Skums P, Stamatakis A, Attolini CS-O, Aparicio S, Baaijens J, Balvert M, Barbanson Bde, Cappuccio A, Corleone G, Dutilh BE, Florescu M, Guryev V, Holmer R, Jahn K, Lobo TJ, Keizer EM, Khatri I, Kielbasa SM, Korbel JO, Kozlov AM, Kuo T-H, Lelieveldt BPF, Mandoiu II, Marioni JC, Marschall T, Mölder F, Niknejad A, Raczkowski L, Reinders M, Ridder Jde, Saliba A-E, Somarakis A, Stegle O, Theis FJ, Yang H, Zelikovsky A, McHardy AC, Raphael BJ, Shah SP, Schönhuth A. Eleven grand challenges in single-cell data science. *Genome Biol* 2020;21:31.

172 Datlinger P, Rendeiro AF, Boenke T, Senekowitsch M, Krausgruber T, Barreca D, Bock C. Ultra-high-throughput single-cell RNA sequencing and perturbation screening with combinatorial fluidic indexing. *Nat Methods* 2021;18:635–42.