Advances and challenges in epigenomic single-cell sequencing applications

Martin Philpott[1], Adam Cribbs[1], Tom Brown jr[2], Tom Brown sr[3], Udo Oppermann[1]

[1] Botnar Research Centre, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, NIHR Oxford BRU, University of Oxford, OX3 7LD, UK

[2] ATDBio, Oxford Science Park, Robert Robinson Ave, Oxford OX4 4GA, UK

[3] Department of Chemistry, University of Oxford, Oxford OX1 3TF, UK

Correspondence to: udo.oppermann@sgc.ox.ac.uk
adam.cribbs@imm.ox.ac.uk
martin.philpott@ndorms.ox.ac.uk

Udo Oppermann
Botnar Research Centre
University of Oxford
Windmill Road, OX3 7LD, UK

Keywords: single-cell epigenomics, single-cell, next generation sequencing, multiplexed single-cell assays

ABSTRACT

Understanding multicellular physiology and pathobiology requires analysis of the relationship between genotype, chromatin organisation and phenotype. In the multiomics era, many methods exist to investigate biological processes across the genome, transcriptome, epigenome, proteome and metabolome. Until recently, this was only possible for populations of cells or complex tissues, creating an averaging effect that may obscure direct correlations between multiple layers of data. Single-cell sequencing methods have removed this averaging effect, but computational integration after profiling distinct modalities separately may still not completely reflect underlying biology. Multiplexed assays resolving multiple modalities in the same cell are required to overcome these shortcomings and have the potential to deliver unprecedented understanding of biology and disease.

The ability of cells to differentiate and their plasticity to adopt new states or identities are central features in the development and homeostasis of multicellular organisms. Cell fate decisions are the results of environmental cues such as cell-cell interactions and binding of soluble factors or ligands with their receptors. These interactions lead to the establishment and maintenance of cell-type specific gene expression programs which are orchestrated by the interplay of genomic and chromatin organisation with cell-type and cell-state specific transcription factor repertoires [1].

Defining cell types and states requires single-cell assays - during the last decade next-generation sequencing, imaging and engineering technologies have provided an unprecedented insight into the biology and heterogeneity on a single-cell level [2] leading to an unprecedented understanding of biology and disease. These technological single-cell advances are important since bulk measurements obliterate crucial information by averaging signals from individual cells. In particular, next-generation sequencing (NGS) has proven to be a remarkably sensitive means of monitoring gene expression, epigenetic modifications, chromatin and nuclear structure, and other aspects of cellular state [1]. This remarkable progress now provides investigators with tools to move further towards mapping and cataloguing critical features such as transcriptomes, epigenomes, metabolomes and proteomes on a single-cell level [3-10]. Whilst single modality interrogation for some of these is possible, the combination of these is at its infancy but would ultimately allow to precisely define cellular states. The construction of comprehensive systems biology models of cellular contexts will eventually provide unparalleled insights into physiology and disease. In this review we will briefly summarize underlying concepts and assays that allow interrogation of cell states and we will indicate the challenges lying ahead.

Current approaches to sequencing-based single-cell technologies - Most single-cell genomics assays have been adapted from similar techniques developed for analysing bulk-cell populations. Nonetheless, most single-cell sequencing based assays require a minimum level of input material that exceeds that of a single cell and accordingly, amplification strategies and development of instruments that physically capture and isolate individual cells provided the first major advancements. Multiplexed single cell sequencing methods can be well-based (where a cell is transferred into an individual well of a multi-well plate, which acts as a discrete reaction vessel for subsequent steps), microfluidics i.e. lab-on-a-chip-based (where

single cells are held at discrete capture sites on a microfluids chip and some steps of library preparation occur in an automated fashion) or droplet-based (where large numbers on cells are individually captured in droplets within an oil emulsion, which then act as enclosed reaction vessels). Well-based and lab-on-chip-based approaches largely remain limited to interrogating hundreds to the low thousands of cells, but may deliver richer information, including coverage of whole transcripts, detection of lower abundance analytes or measurement of analytes not currently amenable to higher throughput approaches. On the other hand, droplet-based multiplexed assays are capable of reporting on many thousands of cells, opening up applications not practical with lower cell numbers. However, the use of barcoded oligo beads in these assays bring their own limitations, such as incomplete analyte capture or restriction to end-sequencing of mRNA transcripts.

Figure 1 illustrates the various sequencing-based assays which we briefly describe here (see figure legend for abbreviations not defined in the text). The first report of combined single cell epigenome and transcriptome profiling was scM&T-seq [11]. scM&T-seq builds upon G&T-seq [12] (which multiplexed single cell genomics and transcriptomics) to provide the first report of combined single cell epigenome and transcriptome profiling. In scM&T-seq, mRNA is captured onto oligo-dT beads, separated and amplified by Switch Mechanism at the 5' End of RNA Templates (SMART-seq), whereas DNA is subjected to bisulphite treatment followed by library preparation using a modified single-cell bisulfite sequencing (scBS-seq) protocol.

Subsequently, scMT-seq [11] and scTrio-seq [13] were reported. These well-based methods involve selective lysis of the cell membrane to release mRNA into solution, followed by physical separation of the nuclei. In both methods, nuclei are subjected to single-cell reduced-representation bisulfite sequencing (scRRBS) to interrogate the DNA methylome, while mRNA libraries are constructed by the SMART-seq2 protocol [14] for scMT-seq or by the method of Tang et al. [15] for scTrio-seq. In addition to DNA methylation, scMT-seq was able to extract (single nucleotide polymorphism, SNP) information from the DNA sequencing, whereas scTrio-seq was able to computationally infer copy number variants (CNV) from the scRRBS.

Similar information is produced by scNMT-seq [16], where single cells are lysed in wells containing GpC methyltransferase, which labels accessible DNA. RNA and DNA libraries are then prepared by the methods of scM&T-seq and scBS-seq [17],

respectively, permitting the measurement of chromatin accessibility, DNA methylation and transcription in single cells.

scCOOL-seq [18] takes this one step further, by combining NOMe-seq [19], which leverages GpC methyltransferase, and post-bisulfite adapter tagging (PBAT), along with lambda DNA spike in, to simultaneously analyse chromatin accessibility/nucleosome positioning, DNA methylation, CNV and ploidy. A subsequent method by the same group, improved scCOOL-seq (iscCOOL-seq) [20] and addresses the low methylome mapping rate observed with previous approaches. Tailing- and ligation-free method for single cells (TAILS) is used to construct methylome libraries and improve mapping efficiencies. However, CNV and ploidy was not demonstrated by iscCOOL-seq and, although scRNA-seq was reported as part of the protocol, this was not performed on the same cells interrogated for epigenomic information.

The ability of Tn5 transposase to cut DNA and append known sequences at the cut sites has made Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) the methods of choice to examine chromatin accessibility [21]. sci-CAR [22] is a well-based method, but makes use of single cell combinatorial indexing (sci), and effectively combines sci-ATAC-seq and sci-RNA-seq into a single protocol. Using this approach, sci-CAR is capable of profiling both chromatin accessibility and the transcriptomes of many thousands of single cells. However, this high throughput profiling, combined with the inherent splitting of mixed RNA and DNA prior to amplification, results in extensive signal loss as well as only reporting on the 3’ ends of RNA transcripts.

scCAT-seq [23] is a well-based method that separates the RNA from the nucleus, before RNA libraries are made by SMART-seq2 and, after Tn5 transposition of the nucleus, scATAC libraries are made using a carrier DNA-mediated protocol. However, while this method overcame the signal loss of sciCAR and reported on full length transcripts, it was only cable of analysing small numbers of cells (74 reported).

ASTAR-seq [24] and T-ATAC-seq [25] are a microfluids lab-on-a-chip assay that use the Fluidigm C1 platform. Both assays return scATAC-seq data, but they differ in that ASTAR-seq also interrogates the transcriptome, whereas T-ATAC-seq enriches small sets of target genes. ASTAR-seq was validated on a number of cell lines, yielding results in the range of hundreds of cells for each cell type, with

improvements in mapping over previous methods, whereas T-ATAC-seq was used to study T-cell receptor-encoding genes in parallel with chromatin accessibility.

SNARE-seq [26] (see Figure 2) describes an innovative droplet-based approach to multiplexing single cell transcriptomics and chromatin accessibility. Pooled extracted nuclei are treated with Tn5 transposase prior to encapsulation on a Drop-seq platform. Standard polyT barcoding beads capture both mRNA directly and transposed DNA via a splint oligo that binds to the polyT oligo at one end and the 5' overhang of transposed DNA at the other end. After droplets are broken, on-bead reverse-transcription with a template switch oligo and covalent ligation of tagmented DNA to the bead are performed in a single step, followed by simultaneous amplification of both cDNA and transposed DNA. Amplified material can then be split without loss of information. Since amplified cDNA and transposed DNA already contain cellular barcodes, library preparation can proceed independently using standard bulk methods. SNARE-seq was used to examine transcriptomes and chromatin accessibility in both cell lines and mouse cortex tissue, reporting on over 10,000 cells for the latter.

Pooled CRISPR-based screens offer tremendous potential to accelerate target discovery in disease and for the dissection of complex biological pathways. However, such screens have largely been restricted to simple readouts such as cell survival or suitable marker proteins. Combining CRISPR screens with single cell transcriptomic and/or epigenomic readouts has the potential to overcome these restrictions.

Perturb-ATAC [27] multiplexes CRISPR screening with chromatin accessibility using pooled lentiviral gRNA libraries and the Fluidigm C1 platform to perform on-chip tagmentation of chromatin and RT of guide barcodes, followed by off-chip library production. Perturb-ATAC was used to dissect gene regulation networks in 2,627 lymphocytes using a library of 40 gRNAs targeting *trans*-factors.

All of these diverse methods demonstrate that there is ample scope for multiplexing at the level of single-cell sequencing. Combinations of some of these existing methods offer pathways for greater levels of multiplexing, such as the inclusion of transcriptomic information in scCOOL-seq using approaches similar to those of scM&T-seq or scMT-seq, or use of the oligo-modifying approach of DART-seq [28] to develop a hybrid assay that eliminates the need for the splint oligo in SNARE-seq, thereby reducing the number of hybridization events needed for target capture.

CITE-seq [29], REAP-seq [30] and ECCITE-seq [31] are novel methods that incorporate DNA-barcoded antibodies into the workflow allowing to perform multimodal single-cell protein, RNA/transcriptome or CRISPR-screen assays. It is conceivable that this approach can be expanded to include chromatin accessibility through e.g. ATAC-seq or possibly in the future with other epigenetic readouts. One area where single cell multiplexing is yet to be demonstrated is with ChIP-seq, which allows the location of specific proteins or protein post-translational modifications to be determined in relation to DNA sequence. However, with the recent description of CUT&Tag [32], which couples Tn5 transposase to ChIP antibodies, a modification of the SNARE-seq protocol to detect chromatin-associated proteins as well as the transcriptome in single cells, multiplexing would seem imminent.

Spatially resolved transcriptomic approaches - in addition to molecular characteristics such as transcriptomes or epigenomes, the spatial organisation of cells is essential to understand their roles. Whilst NGS methods described above provide information on hundreds or thousands of cells, their spatial information is lost during required tissue dissociation to obtain single cell suspensions. In MERFISH [33], an imaging method capable of simultaneously measuring the copy number and spatial distribution of hundreds to thousands of RNA species in single cells, RNA molecules are identified via a combinatorial labelling approach that encodes RNA species with barcodes. This is followed by sequential rounds of single-molecule fluorescence in situ hybridization (smFISH) to read out and map these barcodes onto spatially preserved tissue slices. A conceptionally different approach (spatial transcriptomics) [34] involves spatially arrayed and barcoded capture oligonucleotides, upon which tissue sections are placed followed by cell lysis, leading to conservation and reconstruction of spatially conserved mRNA species upon sequencing. A further development of this concept (high definition spatial transcriptomics, HDST) [35] entails capture of RNAs from tissue sections on a dense, spatially barcoded bead array. These recent developments suggest further imminent advances in multimodal spatial NGS methods.

Computational challenges - As the scale and complexity of new datasets are generated exponentially [36, 37], this presents the computational biology field with the challenge of developing new methodologies. Moreover, new computational approaches for normalisation, data integration and visualisation across often variable datasets will also be required. In the following section we will discuss the

developments, opportunities and challenges that remain in integrating single-cell data, including reference to multi-modal and spatial datasets.

Integration of single-cell datasets across different experiments - Experimental factors, which include both technical elements, as well as biological features make integration of scRNA-seq data challenging. The aim of scRNA-seq data integration is to eliminate the effect of experimental factors driving variation across multiple datasets (Figure 3).

One of the most successful and popular methods for integrating data across different experiments is the Seurat v2 R toolkit [38]. Seurat v2 implements canonical correlation analysis (CCA) to identify sources of variation between different datasets [39], followed by alignment of canonical correlation vectors. Ultimately, the steps in the method project cells into low-dimensional space so that cells are positioned according to their biological state, which is independent of their experimental, donor or species origin. A similar approach is also implemented by mnnCorrect, which accomplishes similar goals to CCA [40]. Because each approach assumes that all datasets share at least one cell type in common or that the gene expression profiles share the same overall population structure across all datasets, these methods are prone to overfitting. This becomes particularly evident upon integrating datasets that have considerable differences in population structure or cellular composition. Overcoming these shortcomings in dataset integration was the motivation for scanorama, a method for integrating multiple scRNA-seq datasets that are composed of highly heterogeneous transcriptional phenotypes [41]. The method is based on computer vision algorithms for panorama stitching and involves identification of nearest neighbours to recognise shared cell types among pairs of datasets. Mutually linked cells from matches are leveraged to correct for batch effects and merge experiments together [41]. This method appears to be a substantial improvement for integrating datasets where there is disparity between the population structure between experiments.

Other specific approaches to integrate different scRNA-seq datasets include methods that utilise factor analysis [42] and cluster based nearest neighbours [43], in addition to normalisation methods such as SCnorm [44] and scran [45] that can also be applied for combining multiple scRNA-seq datasets. Additionally, several groups have demonstrated the utility of neural networks for embedding scRNA-seq datasets in a scalable manner [46-50]. However, the recently published single-cell variational

inference (scVI) framework stands out from other deep learning approaches because of its ability to explicitly model both library size and batch effects. scVI is based on a hierarchical Bayesian model in which the conditional distributions are specified using a deep learning approach to aggregate information across similar cells and genes.

Integration of single-cell datasets across different modalities - Single-cell RNA sequencing (scRNA-seq) is the most common of the single-cell methodologies, with a broad range of technologies that have differing sensitivities, costs and throughput [14, 36, 37, 51]. More recently, other single-cell genomic methods such as chromatin accessibility, chromatin and transcription factor occupancy [32], DNA methylation, proteomic and genomic profiling have complemented the development of scRNA-seq technologies. However, these different types of data present a major computational problem when it comes to attempting to integrate the data across modalities.

There is an extensive number of clustering methods for scRNA-seq or scATAC-seq and most assume that the cells they are sampled from do not represent the same population [52, 53]. However, if cells are sampled from the same population and multiple different single-cell measurements are performed, then it can be assumed that each measurement can inform the analysis of another measurement. Duren *et al* [54] proposed a method based on coupled non-negative matrix factorisation (NMF) to perform coupled clustering of both scRNA-seq and scATAC-seq to infer both the expression profile and the accessibility profile for each subpopulation [55]. Other methods such as self-organising maps [56] and bulk reference guided approaches [4] have also been used to integrate scRNA-seq and scATAC-seq. However, even though the integration of these two profiles reveals a great deal about the active regulatory elements in each subpopulation, a link between active regulatory elements and the active genes cannot be made. This was the motivation for Zeng *et al* [57] to incorporate 3D contact data, such as HiC or HiChIP into their De-Convolution and Coupled Clustering (DC3) NMF model. The authors were able to effectively improve the coupled clustering of the single-cell data and were subsequently able to deconvolve the bulk population profiles of HiChIP data into subpopulation-specific profiles so that they can inform regulatory networks for each subpopulation.

In order to develop a comprehensive framework for integrating different single-cell modalities, defining a shared anchor point between each dataset is required. One approach to define this shared anchor point was proposed for bulk sequencing

integration by Argelaguet *et al* [58]. The Multi-omics Factor Analysis (MOFA) method identifies sets of factors that explain the variance across multiple data modalities [58]. This method was extended to integrate 87 single-cell methylation and transcriptome sequencing profiles performed using scM&T-seq [11]. This revealed organised DNA methylation and transcriptome changes during mouse stem cell embryonic differentiation. This suggests that bulk methods can be repurposed to reveal improved interpretation of single-cell data. Aligned with this joint analysis of multimodal dataset analysis, more recently, two groups have described strategies for accomplishing dedicated multimodal analysis approaches that utilise frameworks that permit identification of shared properties in the gene expression space. Welch *et al* implement linked inference of genomic experimental relationships (LIGER), which leverages an integrative non-negative matrix factorisation (iNMF) strategy to identify reduced dimensionality vectors that describe the major source of variation between two or more datasets [59]. The approach is highly scalable and manually tuneable. Stuart *et al* [39] built upon the CCA alignment methods built into Seurat v2 R toolkit [38]. It implements a method similar to scanorama, in which a set of alignments are generated by finding the mutual nearest neighbours (MNN) across all cells [41]. However, the advantage of applying MNN to Seurat was the ability to perform transfer learning and to project cellular states across different modalities. Both Welch *et al* [60] and Stuart *et al* present a number of extended applications beyond scRNA-seq for their tools. Welch *et al* anti-correlate CpG methylation with scRNA-seq, which allowed further cell type identify refinement and exploration of the DNA methylation-transcriptional relationship. Stuart *et al* apply transfer learning using immune cell data when they used CITE-seq and employed this to impute the protein expression to a larger Human Cell Atlas (HCA) dataset. Furthermore, given the similarity between Seurat v3 and scanorama, it is conceivable that scanorama could be extended to handle multi-modal datasets also.

As datasets expand in their complexity and quantity, powerful approaches for ‘multiview’ machine learning are likely to emerge as single-cell analysis approaches [39]. One ideal approach suggested would be to construct single-cell maps based on a join kernel that incorporates all measured single-cell omics layers (reviewed by [61]). However, it is unclear at present whether this could be performed in a computationally efficient manner in a framework that would be superior to current state of the art methods.

Integration of single-cell and spatial datasets - The spatial organisation of cells in a tissue reflects its function and the cellular localisation can be important for explaining the differences in cellular differentiation and cell state. Spatial localisation of gene expression in single-cells underpins the function of a tissue, as similar gene expression profiles can occupy similar spatial domains *in situ.* A major computational effort is currently focused on integrating single-cell and spatial datasets. Often spatial datasets lack the high resolution achieved by single-cell sequencing and these experimental limitations can be overcome by integrating the two datasets together to reinforce the spatial expression maps. The computational integration of spatial data, gathered using FISH and scRNA-seq, was demonstrated in two seminal publications by Satija *et al* and Achim *et al* [62, 63]. The main idea behind their approach was to use a reference map of informative marker genes as a guide to assign spatial coordinates to single-cell sequenced cells. The methods were then successfully used in a number of tissues including to study stem cell differentiation in *Drosophila* embryos and mammalian liver [64, 65]. Nitzan *et al* 2019 [66] can reconstruct de novo the spatial gene expression profile of tissue, without reliance on any prior information. Despite the advances made so far in single-cell sequencing assays and spatial integration, new methods of integration will provide an unprecedented level of understanding between the spatial and functional organisation of tissue.

Outlook

Whilst significant progress has been achieved over the last 5 years to develop single cell assays and analysis methods with the aim to obtain integrated data across different modalities such as transcriptome, chromatin, epigenome and proteome, several obstacles remain. Sensitivity of single cell sequencing based assays limits the obtainable information from any one cell, hence reliable amplification and detection techniques need further development, especially protein-based NGS single-cell technologies, which are currently restricted to DNA-barcoded antibody detection of relatively few (and not proteome-wide) targets. Furthermore, multiplexing of different NGS assays may hit practical limitations when two or more modalities require the same analyte. For example, it is difficult to envisage assays where scChIP-seq and scATAC-seq or DNA methylation are examined in the same cell, since measuring one may prevent detection of the others. In order to merge different data types from various types of analytes (including in the future metabolomics) computational methods need further development, before robust

deployment. One of the main issues is scalability of computational methods. These demand significant resources e.g. memory availability when computing across millions of cells. This will be more apparent for deep learning approaches, where the running times for model fitting can be significant.

HIGHLIGHTS

- *Progress in microfluidics, imaging and chemistry permit multiple single-cell assays*
- *Multiplexing is possible including epigenetic or chromatin assays*
- *Spatial transcriptomic methods allow identification of cell states in tissues*
- *Computational solutions are developed to achieve integration of different data types*

ACKNOWLEDGMENTS

Work is supported by Cancer Research UK (C41580/A23900), Versus Arthritis (program grant 20522), Leducq Foundation (LEAN program grant), Bone Cancer Research Trust, Chan-Zuckerberg Foundation, EPSRC, Celgene Corporation, Bayer Healthcare and GlaxoSmithKline.

Competing interest statement - TBjr and TBsr are shareholders of ATDBio, a biotech specialising in advanced nucleic acid chemistry. The other authors declare no conflict of interest.

LEGENDS to Figures

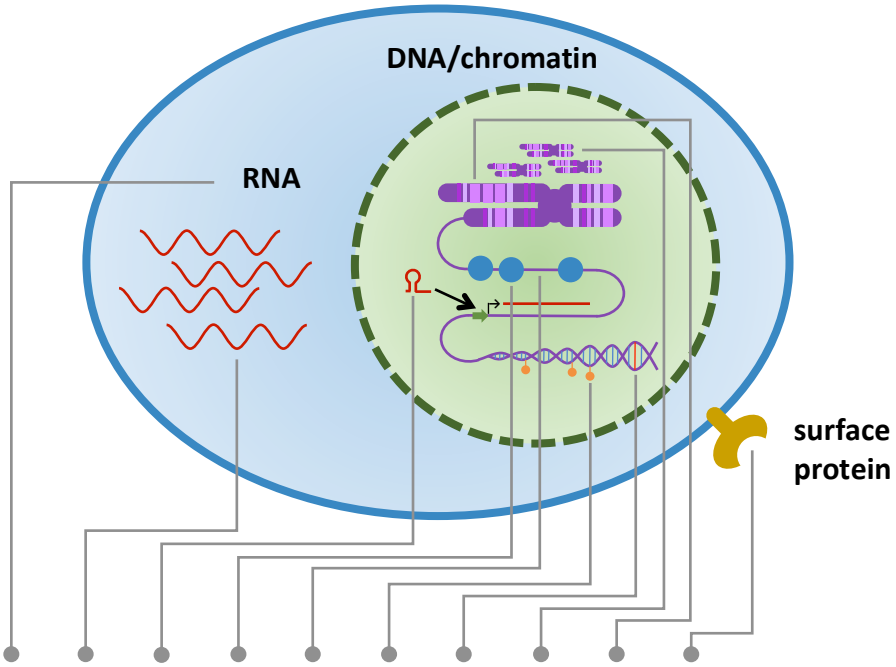
Figure 1: Overview of selected single-cell sequencing-based approaches that interrogate multiple modalities. Abbreviations: G&T-seq, genome and transcriptome sequencing; scM&T-seq, single-cell genome-wide methylome and transcriptome sequencing; scMT-seq, single-cell methylome and transcriptome sequencing; scNMT-seq, single-cell nucleosome, methylation and transcription sequencing; scNOMe-seq, single-cell nucleosome occupancy and methylome-sequencing; scCOOL-seq, single-cell chromatin overall omic-scale landscape sequencing; iscCOOL-seq, improved scCOOL-seq; sci-CAR, single-cell combinatorial indexing - chromatin accessibility and mRNA; scCAT-seq, single-cell chromatin accessibility and transcriptome sequencing; ASTAR-seq, assay for single-cell transcriptome and accessibility regions sequencing; T-ATAC-seq, transcript-indexed ATAC-seq; SNARE-seq, single-nucleus chromatin accessibility and mRNA expression sequencing; CITE-seq, cellular indexing of transcriptomes and epitopes by sequencing; REAP-seq, RNA expression and protein sequencing; ECCITE-seq, expanded CRISPR-compatible cellular indexing of transcriptomes and epitopes by sequencing.

Figure 2: DROP-seq principle, also depicting SNARE-seq modifications (shown in bold). Barcoded beads in a cell-lysis solution, cells and oil are passed through a microfluidics device to create aqueous droplets in an oil emulsion (in SNARE-seq, nuclei are extracted from cells and pre-treated with Tn5, before encapsulation with beads, along with a splint oligo). Within droplets, polyadenylated RNA transcripts are captured onto the beads (in SNARE-seq, the splint oligo also binds to the capture sequence and tagmented DNA binds to the splint oligo). The emulsion is broken and mRNA reverse transcribed (RT) using a template switch oligo (TSO) (in SNARE-seq, the captured DNA is ligated to the barcoding oligo in the same reaction mix as the RT, after which PCR simultaneously amplifies cDNA and transposed DNA before the reaction is split for separate library preparation). cDNA is amplified using the SMART-primer and TSO handles, before fragmentation/adapter addition and indexing using a Nextera XT kit (Illumina) to create scRNA-seq libraries (in SNARE-seq, tagmented DNA is indexed using primers against the SMART primer site and the MEDS sequences added by Tn5 transposase. PAGE-gel purification is used to appropriately size select the final scATAC-seq libraries.

Figure 3: Strategies for the integration of single-cell multimodal datasets above and beyond final result integration, such as those described by Bock *et al* 2016 [67]. **(A)** Multi-view matrix factorisation methods align the datasets into conserved low-dimensional space. Recently, a number of methods have been successfully developed and applied to integrate single-cell multi-omics datasets and include Seurat V3 [39], LIGER [60] and MOFA [58]. **(B)** Similar to multi-omics bulk approaches, dependencies between omics layer can be visualised as interlinking networks, allowing for jointly regulated cores. Thus, biological information can be

inferred through the edges of the network. **(C)** Deep learning approaches have also been suggested as possible multi-view learning approached for single-cell multimodal integration, where they have been successfully applied at the bulk level [68]. **(D)** First proposed by Colome-Tatche *et al* 2018 [61], instead of treating the omics layers as separate, a more suitable approach would be to construct single-cell maps based on a joint kernel that incorporates all measured layers. This approach would join multi-space measurements that deliver a single similarity value between them. The advantage of this method would be that the output data could be analysed directly using standard analysis on the integrated datasets. Figure adapted from [61].

Figure 1



METHOD	V(D)J/targeted sequencing	mRNA	CRISPR	Nucleosome position	Chromatin accessibility	DNA methylation	SNP	Ploidy	CNV	Protein	Format	Reference
G&T-seq		✓					✓		✓		well	[12]
scM&T-seq		✓				✓					well	[11]
scMT-seq		✓				✓	✓				well	[69]
Trio-seq		✓				✓			✓		well	[70]
scNMT-seq		✓			✓	✓					well	[16]
scNOMe-seq				✓		✓					well	[19]
scCOOL-seq				✓	✓	✓		✓	✓		well	[18]
iscCOOL-seq				✓	✓	✓					well	[20]
sci-CAR		✓			✓						well	[22]
scCAT-seq		✓			✓						well	[23]
ASTAR-seq		✓			✓						microfluidics	[24]
T-ATAC-seq	✓				✓						microfluidics	[25]
SNARE-seq		✓			✓						droplet	[26]
Perturb-ATAC			✓		✓						microfluidics	[27]
CITE-seq		✓								✓	droplet	[29]
REAP-seq		✓								✓	droplet	[30]
ECCITE-seq	✓	✓	✓							✓	droplet	[31]
	transcriptomic			epigenomic			genomic					

Figure 2

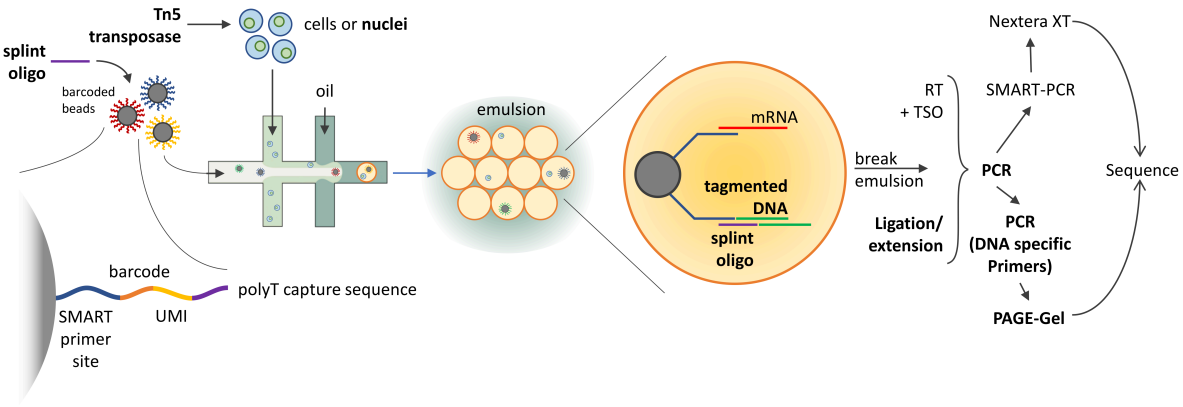
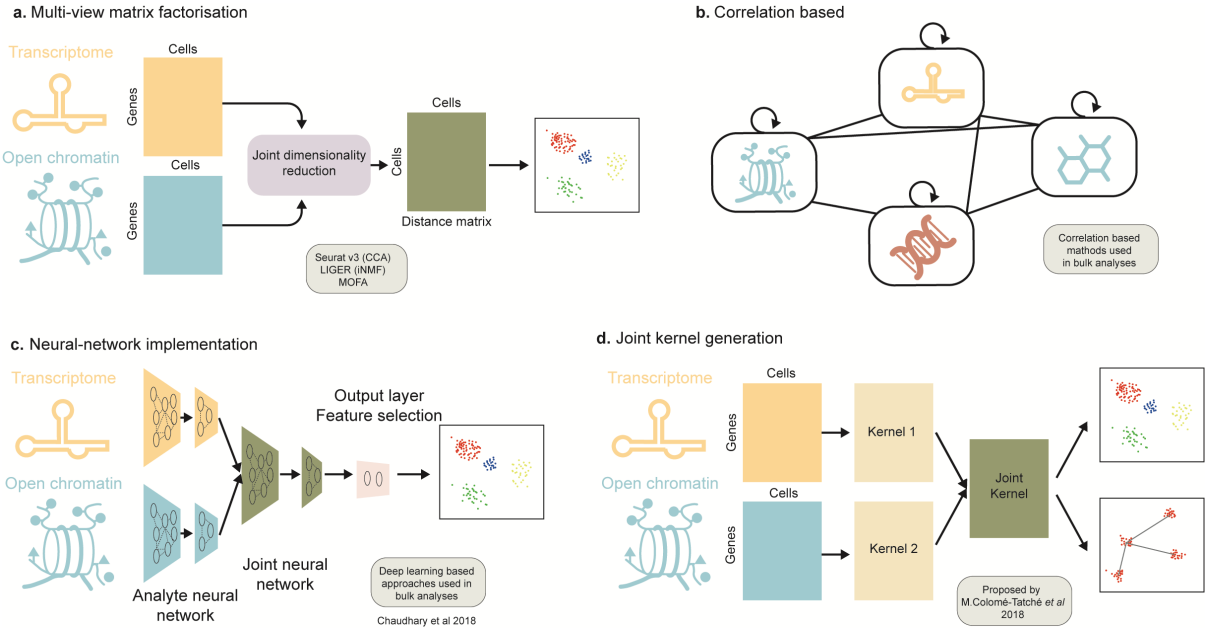

Tn5
transposase
cells or nuclei
splint
oligo
barcoded
beads
oil
emulsion
mRNA
tagmented
DNA
splint
oligo
break
emulsion
RT
+ TSO
Ligation/
extension
Nextera XT
SMART-PCR
PCR
PCR
(DNA specific
Primers)
PAGE-Gel
Sequence
barcode
SMART
primer
site
UMI
polyT capture sequence

Figure 3


a. Multi-view matrix factorisation
Transcriptome
Open chromatin
Cells
Genes
Cells
Genes
Joint dimensionality reduction
Cells
Cells
Distance matrix
Seurat v3 (CCA)
LIGER (iNMF)
MOFA
b. Correlation based
Correlation based mathods used in bulk analyses
c. Neural-network implementation
Transcriptome
Open chromatin
Output layer
Feature selection
Analyte neural network
Joint neural network
Deep learning based approaches used in bulk analyses
Chaudhary et al 2018
d. Joint kernel generation
Transcriptome
Open chromatin
Cells
Genes
Cells
Genes
Kernel 1
Kernel 2
Joint Kernel
Proposed by M.Colomé-Tatché et al 2018

REFERENCES

1. Stadhouders, R., Filion, G.J., and Graf, T. (2019). Transcription factors and 3D genome conformation in cell-fate decisions. Nature *569*, 345-354.
2. Trapnell, C. (2015). Defining cell types and states with single-cell genomics. Genome Res *25*, 1491-1498.
3. Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., et al. (2017). The Human Cell Atlas. Elife *6*.
4. Buenrostro, J.D., Corces, M.R., Lareau, C.A., Wu, B., Schep, A.N., Aryee, M.J., Majeti, R., Chang, H.Y., and Greenleaf, W.J. (2018). Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. Cell *173*, 1535-1548 e1516.
5. Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature *523*, 486-490.

*** This paper describes application of ATAC-seq to single-cell assays**

6. Chihara, N., Madi, A., Kondo, T., Zhang, H., Acharya, N., Singer, M., Nyman, J., Marjanovic, N.D., Kowalczyk, M.S., Wang, C., et al. (2018). Induction and transcriptional regulation of the co-inhibitory gene module in T cells. Nature *558*, 454-459.
7. Duncan, K.D., Fyrestam, J., and Lanekoff, I. (2019). Advances in mass spectrometry based single-cell metabolomics. Analyst *144*, 782-793.
8. Ludwig, L.S., Lareau, C.A., Bao, E.L., Nandakumar, S.K., Muus, C., Ulirsch, J.C., Chowdhary, K., Buenrostro, J.D., Mohandas, N., An, X., et al. (2019). Transcriptional States and Chromatin Accessibility Underlying Human Erythropoiesis. Cell Rep *27*, 3228-3240 e3227.
9. Palii, C.G., Cheng, Q., Gillespie, M.A., Shannon, P., Mazurczyk, M., Napolitani, G., Price, N.D., Ranish, J.A., Morrissey, E., Higgs, D.R., et al. (2019). Single-Cell Proteomics Reveal that Quantitative Changes in Co-expressed Lineage-Specific Transcription Factors Determine Cell Fate. Cell Stem Cell *24*, 812-820 e815.
10. Zenobi, R. (2013). Single-cell metabolomics: analytical and biological perspectives. Science *342*, 1243259.
11. Angermueller, C., Clark, S.J., Lee, H.J., Macaulay, I.C., Teng, M.J., Hu, T.X., Krueger, F., Smallwood, S., Ponting, C.P., Voet, T., et al. (2016). Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. Nat Methods *13*, 229-232.
12. Macaulay, I.C., Haerty, W., Kumar, P., Li, Y.I., Hu, T.X., Teng, M.J., Goolam, M., Saurat, N., Coupland, P., Shirley, L.M., et al. (2015). G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. Nat Methods *12*, 519-522.
13. Goolam, M., Scialdone, A., Graham, S.J.L., Macaulay, I.C., Jedrusik, A., Hupalowska, A., Voet, T., Marioni, J.C., and Zernicka-Goetz, M. (2016). Heterogeneity in Oct4 and Sox2 Targets Biases Cell Fate in 4-Cell Mouse Embryos. Cell *165*, 61-74.
14. Picelli, S., Bjorklund, A.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat Methods *10*, 1096-1098.
15. Tang, F., Barbacioru, C., Bao, S., Lee, C., Nordman, E., Wang, X., Lao, K., and Surani, M.A. (2010). Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. Cell Stem Cell *6*, 468-478.

16. Clark, S.J., Argelaguet, R., Kapourani, C.A., Stubbs, T.M., Lee, H.J., Alda-Catalinas, C., Krueger, F., Sanguinetti, G., Kelsey, G., Marioni, J.C., et al. (2018). scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. Nat Commun *9*, 781.
17. Smallwood, S.A., Lee, H.J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S.R., Stegle, O., Reik, W., and Kelsey, G. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat Methods *11*, 817-820.
18. Guo, F., Li, L., Li, J., Wu, X., Hu, B., Zhu, P., Wen, L., and Tang, F. (2017). Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells. Cell Res *27*, 967-988.

*** Method demonstrating the greatest degree of multiplexing to date, with scope for additional modalities**

19. Kelly, T.K., Liu, Y., Lay, F.D., Liang, G., Berman, B.P., and Jones, P.A. (2012). Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. Genome Res *22*, 2497-2506.
20. Gu, C., Liu, S., Wu, Q., Zhang, L., and Guo, F. (2019). Integrative single-cell analysis of transcriptome, DNA methylome and chromatin accessibility in mouse oocytes. Cell Res *29*, 110-123.
21. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods *10*, 1213-1218.
22. Cao, J., Cusanovich, D.A., Ramani, V., Aghamirzaie, D., Pliner, H.A., Hill, A.J., Daza, R.M., McFaline-Figueroa, J.L., Packer, J.S., Christiansen, L., et al. (2018). Joint profiling of chromatin accessibility and gene expression in thousands of single cells. Science *361*, 1380-1385.

*** This paper builds on previously described strategies for combinatorial indexing to profile chromatin accessibility and transcriptomics, allowing to increase cell numbers for profiling**

23. Liu, L., Liu, C., Quintero, A., Wu, L., Yuan, Y., Wang, M., Cheng, M., Leng, L., Xu, L., Dong, G., et al. (2019). Deconvolution of single-cell multi-omics layers reveals regulatory heterogeneity. Nat Commun *10*, 470.
24. Xing, Q., Farran, C., Yi, Y., Warrier, T., Gautam, P., and Collins, J.e.a. (2019). Parallel Bimodal Single-cell sequencing of transcriptome and chromatin accessibility. BioRxive 2019.
25. Satpathy, A.T., Saligrama, N., Buenrostro, J.D., Wei, Y., Wu, B., Rubin, A.J., Granja, J.M., Lareau, C.A., Li, R., Qi, Y., et al. (2018). Transcript-indexed ATAC-seq for precision immune profiling. Nat Med *24*, 580-590.
26. Chen, S., Lake, B.B., and Zhang, K. (2019). High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. Nat Biotechnol *37*, 1452-1457.

**** The first report to multiplex transcriptome and an epigenomic modality using a high throughput droplet-based approach amenable to greater multiplexing or alternative modalities**

27. Rubin, A.J., Parker, K.R., Satpathy, A.T., Qi, Y., Wu, B., Ong, A.J., Mumbach, M.R., Ji, A.L., Kim, D.S., Cho, S.W., et al. (2019). Coupled Single-Cell CRISPR Screening and

Epigenomic Profiling Reveals Causal Gene Regulatory Networks. Cell *176*, 361-376 e317.

28. Saikia, M., Burnham, P., Keshavjee, S.H., Wang, M.F.Z., Heyang, M., Moral-Lopez, P., Hinchman, M.M., Danko, C.G., Parker, J.S.L., and De Vlaminck, I. (2019). Simultaneous multiplexed amplicon sequencing and transcriptome profiling in single cells. Nat Methods *16*, 59-62.
29. Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P.K., Swerdlow, H., Satija, R., and Smibert, P. (2017). Simultaneous epitope and transcriptome measurement in single cells. Nat Methods *14*, 865-868.
30. Peterson, V.M., Zhang, K.X., Kumar, N., Wong, J., Li, L., Wilson, D.C., Moore, R., McClanahan, T.K., Sadekova, S., and Klappenbach, J.A. (2017). Multiplexed quantification of proteins and transcripts in single cells. Nat Biotechnol *35*, 936-939.
31. Mimitou, E.P., Cheng, A., Montalbano, A., Hao, S., Stoeckius, M., Legut, M., Roush, T., Herrera, A., Papalexi, E., Ouyang, Z., et al. (2019). Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. Nat Methods *16*, 409-412.
32. Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G., Ahmad, K., and Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun *10*, 1930.

**** This paper describes the development and application of transposase mediated tagging for chromatin interrogation beyond ATAC-seq .**

33. Moffitt, J.R., Hao, J., Wang, G., Chen, K.H., Babcock, H.P., and Zhuang, X. (2016). High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. Proc Natl Acad Sci U S A *113*, 11046-11051.

**** This paper, together with [34] and [35], describes methods to resolve transcriptomics in a spatial manner**

34. Stahl, P.L., Salmen, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J.O., Huss, M., et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science *353*, 78-82.

*** see comment for [33]**

35. Vickovic, S., Eraslan, G., Salmen, F., Klughammer, J., Stenbeck, L., Schapiro, D., Aijo, T., Bonneau, R., Bergenstrahle, L., Navarro, J.F., et al. (2019). High-definition spatial transcriptomics for in situ tissue profiling. Nat Methods *16*, 987-990.

**** see comment for [33] and [34].**

36. Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell *161*, 1187-1201.

**** This paper and [37] are landmark investigations that provide the basis for droplet-based and widely used single-cell transcriptomic experiments**

37. Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202-1214.

**** see annotation above for [36]**

38. Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol *36*, 411-420.
39. Stuart, T., and Satija, R. (2019). Integrative single-cell analysis. Nat Rev Genet *20*, 257-272.

*** Excellent review to detail state of the art multimodal single cell experiments**

40. Haghverdi, L., Lun, A.T.L., Morgan, M.D., and Marioni, J.C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat Biotechnol *36*, 421-427.
41. Hie, B., Bryson, B., and Berger, B. (2019). Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. Nat Biotechnol *37*, 685-691.
42. Lin, Y., Ghazanfar, S., Wang, K.Y.X., Gagnon-Bartsch, J.A., Lo, K.K., Su, X., Han, Z.G., Ormerod, J.T., Speed, T.P., Yang, P., et al. (2019). scMerge leverages factor analysis, stable expression, and pseudoreplication to merge multiple single-cell RNA-seq datasets. Proc Natl Acad Sci U S A *116*, 9775-9784.
43. Kiselev, V.Y., Yiu, A., and Hemberg, M. (2018). scmap: projection of single-cell RNA-seq data across data sets. Nat Methods *15*, 359-362.
44. Bacher, R., Chu, L.F., Leng, N., Gasch, A.P., Thomson, J.A., Stewart, R.M., Newton, M., and Kendziorski, C. (2017). SCnorm: robust normalization of single-cell RNA-seq data. Nat Methods *14*, 584-586.
45. Lun, A.T., McCarthy, D.J., and Marioni, J.C. (2016). A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. F1000Res *5*, 2122.
46. Ding, J., Condon, A., and Shah, S.P. (2018). Interpretable dimensionality reduction of single cell transcriptome data with deep generative models. Nat Commun *9*, 2002.
47. Eraslan, G., Simon, L.M., Mircea, M., Mueller, N.S., and Theis, F.J. (2019). Single-cell RNA-seq denoising using a deep count autoencoder. Nat Commun *10*, 390.
48. Gronbach, C.H.e.a. (2019). scVAE: Variational auto-encoders for single-cell gene expression data. BioRxive *10.1101/318295*.
49. Johansen, N., and Quon, G. (2019). scAlign: a tool for alignment, integration, and rare cell identification from scRNA-seq data. Genome Biol *20*, 166.
50. Wang, D., and Gu, J. (2018). VASC: Dimension Reduction and Visualization of Single-cell RNA-seq Data by Deep Variational Autoencoder. Genomics Proteomics Bioinformatics *16*, 320-331.
51. Zheng, G.X., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat Commun *8*, 14049.
52. Kiselev, V.Y., Kirschner, K., Schaub, M.T., Andrews, T., Yiu, A., Chandra, T., Natarajan, K.N., Reik, W., Barahona, M., Green, A.R., et al. (2017). SC3: consensus clustering of single-cell RNA-seq data. Nat Methods *14*, 483-486.
53. Zamanighomi, M., Lin, Z., Daley, T., Chen, X., Duren, Z., Schep, A., Greenleaf, W.J., and Wong, W.H. (2018). Unsupervised clustering and epigenetic classification of single cells. Nat Commun *9*, 2410.
54. Duren, Z., Chen, X., Jiang, R., Wang, Y., and Wong, W.H. (2017). Modeling gene regulation from paired expression and chromatin accessibility data. Proc Natl Acad Sci U S A *114*, E4914-E4923.

55. Duren, Z., Chen, X., Zamanighomi, M., Zeng, W., Satpathy, A.T., Chang, H.Y., Wang, Y., and Wong, W.H. (2018). Integrative analysis of single-cell genomics data by coupled nonnegative matrix factorizations. Proc Natl Acad Sci U S A *115*, 7723-7728.
56. Jansen, C., and al., e. (2019). Building gene regulatory networks from scATAC-seq and scRNA-seq using Linked Self Organizing Maps. PLoS Comput Biol *15*, e1006555.
57. Zeng, W., Chen, X., Duren, Z., Wang, Y., Jiang, R., and Wong, W.H. (2019). DC3 is a method for deconvolution and coupled clustering from bulk and single-cell genomics data. Nat Commun *10*, 4613.
58. Argelaguet, R., Velten, B., Arnol, D., Dietrich, S., Zenz, T., Marioni, J.C., Buettner, F., Huber, W., and Stegle, O. (2018). Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets. Mol Syst Biol *14*, e8124.
59. Adey, A.C. (2019). Integration of Single-Cell Genomics Datasets. Cell *177*, 1677-1679.
60. Welch, J.D., Kozareva, V., Ferreira, A., Vanderburg, C., Martin, C., and Macosko, E.Z. (2019). Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity. Cell *177*, 1873-1887 e1817.
61. Colome-Tatche, M., and Theis, F.J. (2018). Statistical single cell multi-omics integration. Curr Opin Syst Biol *7*, 54-59.

*** Overview on current computational approaches to integrate single cell modalities**

62. Achim, K., Pettit, J.B., Saraiva, L.R., Gavriouchkina, D., Larsson, T., Arendt, D., and Marioni, J.C. (2015). High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. Nat Biotechnol *33*, 503-509.
63. Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat Biotechnol *33*, 495-502.
64. Halpern, K.B., Shenhav, R., Matcovitch-Natan, O., Toth, B., Lemze, D., Golan, M., Massasa, E.E., Baydatch, S., Landen, S., Moor, A.E., et al. (2017). Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. Nature *542*, 352-356.
65. Karaiskos, N., Wahle, P., Alles, J., Boltengagen, A., Ayoub, S., Kipar, C., Kocks, C., Rajewsky, N., and Zinzen, R.P. (2017). The Drosophila embryo at single-cell transcriptome resolution. Science *358*, 194-199.
66. Nitzan, M., Karaiskos, N., Friedman, N., and Rajewsky, N. (2019). Gene expression cartography. Nature *576*, 132-137.
67. Bock, C., Farlik, M., and Sheffield, N.C. (2016). Multi-Omics of Single Cells: Strategies and Applications. Trends Biotechnol *34*, 605-608.
68. Chaudhary, K., Poirion, O.B., Lu, L., and Garmire, L.X. (2018). Deep Learning-Based Multi-Omics Integration Robustly Predicts Survival in Liver Cancer. Clin Cancer Res *24*, 1248-1259.
69. Hu, Y., Huang, K., An, Q., Du, G., Hu, G., Xue, J., Zhu, X., Wang, C.Y., Xue, Z., and Fan, G. (2016). Simultaneous profiling of transcriptome and DNA methylome from a single cell. Genome Biol *17*, 88.
70. Hou, Y., Guo, H., Cao, C., Li, X., Hu, B., Zhu, P., Wu, X., Wen, L., Tang, F., Huang, Y., et al. (2016). Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. Cell Res *26*, 304-319.