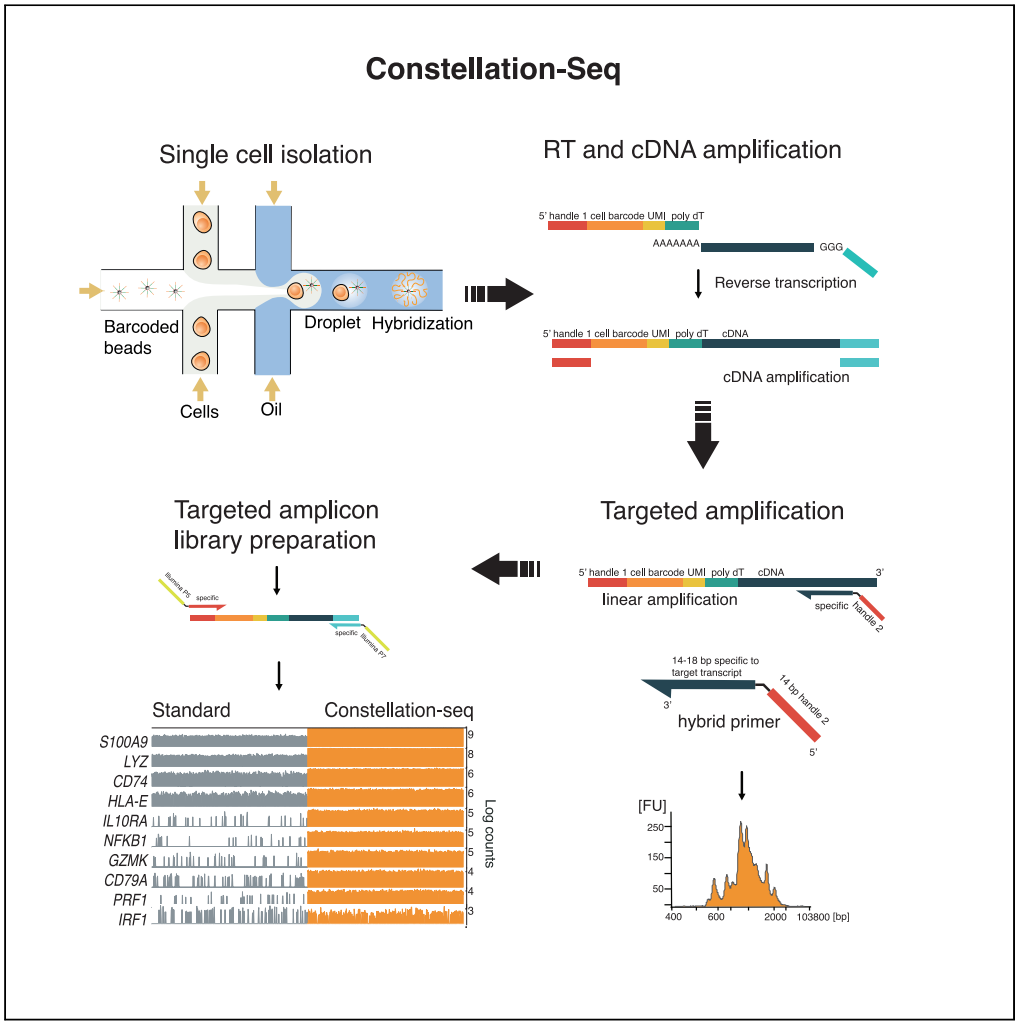
iScience

CellPress
OPEN ACCESS

Article

Resolving cellular systems by ultra-sensitive and economical single-cell transcriptome filtering



Andres F. Vallejo, James Davies, Amit Grover, Ching-Hsuan Tsai, Robert Jepras, Marta E. Polak, Jonathan West

afvp1f17@soton.ac.uk (A.F.V.)
m.e.polak@soton.ac.uk (M.E.P.)
j.j.west@soton.ac.uk (J.W.)

HIGHLIGHTS

Constellation-Seq addresses the issues of sensitivity and sparsity in scRNA-sequencing

Constellation-Seq can be used with DropSeq and chromium 3′ chemistry

Constellation-Seq resolves rare dendritic cell populations from PBMC

Vallejo et al., iScience 24, 102147
March 19, 2021
https://doi.org/10.1016/j.isci.2021.102147

iScience

CellPress
OPEN ACCESS

Article

Resolving cellular systems by ultra-sensitive and economical single-cell transcriptome filtering

Andres F. Vallejo,[1,6,*] James Davies,[1] Amit Grover,[2] Ching-Hsuan Tsai,[3] Robert Jepras,[2] Marta E. Polak,[1,4,6,7,*] and Jonathan West[4,5,6,*]

SUMMARY

Single-cell transcriptomics suffer from sensitivity limits that restrict low abundance transcript identification, affects clustering and can hamper downstream analyses. Here, we describe Constellation sequencing (Constellation-Seq), a molecular transcriptome filter that delivers two orders of magnitude sensitivity gains by maximizing read utility while reducing the data sparsity and sequencing costs. The technique reliably measures changes in gene expression and was demonstrated by resolving rare dendritic cell populations from a peripheral blood mononuclear cell sample sample and exploring their biology with extreme resolution. The simple and powerful method is fully compatible with standard scRNA-Seq library preparation protocols and can be used for hypothesis testing, marker validation or investigating pathways.

INTRODUCTION

The dramatic uptake and expansion of single-cell transcriptome analysis tools has transformed biological research, enabling reconstruction of population architectures and underlying processes to be revealed. The tools rely on compartmentalization of single cells with the introduction of unique genetic barcodes during library preparation (Ziegenhain et al., 2017). Though formidable, not unexpectedly these methods have sensitivity limits, with associated transcript absence events (dropouts) that restrict the faithful delineation of cell subtypes and especially overlook low abundant transcripts such as transcription factors, receptors, and signaling molecules that are often pivotal for accurately describing cell processes and fate (Bacher and Kendziorski, 2016; Vallejos et al., 2017). This is a consequence of high abundance transcripts occupying the available NGS read space and is exacerbated by exponential PCR-directed library preparation routines.

Targeted approaches forgo global transcriptome screens, preferring to select transcripts of known utility and are especially favored for mechanistic studies. Diverse targeted strategies have emerged; physical recovery of transcriptome subsets (Riemondy et al., 2019), coupling custom primers to poly (dT) capture beads (DART-seq) (Saikia et al., 2019) and panel selection by PCR as with the Rhapsody workflow (BD) (Salomon et al., 2019). These methods are technically challenging and introduce substantial costs. In order to overcome these limitations, we developed a fast, easy to use, accurate, and highly flexible method for targeted single cell transcriptomics, while imparting extreme sensitivity to overcome data sparsity problems. We call the method Constellation-Seq and demonstrate its power by application to investigations of a specific, rare population of immune cells: dendritic cells (DCs). DCs play a central role in pathogen sensing, phagocytosis, and antigen presentation (Steinman, 2003). Historically DCs have been defined by a combination of morphology, localization, functions, and expression of a restricted set of surface markers (Fromm et al., 2016; Muzaki et al., 2016; Haniffa et al., 2012; Polak et al., 2008). Single cell RNA sequencing technologies have opened the opportunity for in depth investigation and redefining the classification of these elusive, yet critically important cells. Villani and colleagues redefined the complexity of blood DC populations, describing 6 transcriptomically unique subsets (Villani et al., 2017). However, investigations of their identities and respective roles they play in immune response regulation are limited by their low abundance in tissues and blood. Constellation-Seq enables tracking of the rare DC population without disruptive processing of the PBMCs, expands our knowledge about the

[1]Clinical and Experimental Sciences, Sir Henry Wellcome Laboratories, Faculty of Medicine, University of Southampton, SO16 6YD Southampton, UK

[2]GlaxoSmithKline, Gunnels Wood Road, Stevenage SG1 2NY, UK

[3]GlaxoSmithKline, 200 Cambridge Park Drive, Cambridge, MA 02140, USA

[4]Institute for Life Sciences, University of Southampton, Southampton SO17 1BJ, UK

[5]Cancer Sciences, Faculty of Medicine, University of Southampton, Southampton, UK

[6]These authors contributed equally

[7]Lead contact

*Correspondence: afvp1f17@soton.ac.uk (A.F.V.), m.e.polak@soton.ac.uk (M.E.P.), j.j.west@soton.ac.uk (J.W.)

https://doi.org/10.1016/j.isci.2021.102147

prevalence and activation status of sub-populations of blood DCs in health and disease, and presents an attractive diagnostic means linking to future therapeutic strategies.

The limitation of current scRNA-sequencing techniques relates to the difficulty differentiating biologically inactive genes from technical drop-outs, which impact interpretation of the results, can confound normalization, marker selection and more importantly, cell type labeling and the discovery of new cell types. Here, we describe Constellation-Seq, a remarkably simple, inexpensive and scalable (e.g. >200 targets) approach. The method introduces a linear amplification stage in advance of conventional library preparation. Superior performance is demonstrated with two orders of magnitude sensitivity gains for describing system architectures and processes with unprecedented resolution.

The capture beads each support 10^{10} probes (Saikia et al., 2019) indicating that sensitivity losses arise from the restricted NGS read space (~10^{4-6}/cell) and also from exponential PCR amplification during library preparation, where abundant and more efficiently replicated transcripts dominate the available reads. In contrast, linear (single primer) amplification provides an unbiased route to enrichment across transcripts (Hashimshony et al., 2012; Tang et al., 2011). Therefore, in our approach we have used linear amplification following cDNA synthesis for the targeted enrichment of transcripts of interest. The method involves replacing the template switching oligo (TSO) with hybrid primers containing a transcript-specific region adjacent to a universal handle to select and barcode desired transcripts in a single linear amplification. The method is illustrated in Figure 1A and compared with Drop-Seq and other targeted methods in Figure S1). We introduced this linear targeted amplification step to the scRNA-Seq pipeline to provide a direct comparison that is amenable to cost-effective, large-scale cell screening campaigns albeit with recognised sparsity limitations (Ziegenhain et al., 2017; Lähnemann et al., 2020). The panel of primers can be selected based on previous knowledge of the system, from the literature or hypothesis driven. In addition, an aliquot of the cDNA can be used for standard, bulk sequencing from which a group of target genes can be identified and then used for interrogating the same sample at high resolution.

RESULTS

Constellation-Seq dramatically reduces the sparsity in scRNA-Seq data

Constellation-Seq was first establish for the DropSeq method and further extended to 10X chromium Single Cell 3′ V3. To exclude biological variation we first used the DropSeq protocol for producing standard beads bearing bulk RNA (Macosko et al., 2015; Svensson, 2019). Using a panel of 20 target genes, sensitivity was compared between single primer linear amplification and dual primer exponential amplification (PCR, requiring an SMART-Seq reverse primer) akin to state of the art methods (*e.g.* Rhapsody, BD) (Salomon et al., 2019). The primer panel contained high, medium and low expression level transcripts specific for peripheral blood mononuclear cells (PBMCs), including markers of newly described blood DC populations (Villani et al., 2017) and activation traits (Table S1). Constellation-Seq is amplification cycle and primer concentration dependent (Figure S2), with straightforward optimisation enabling the selective capture of desired transcripts which produce a characteristically spiny tapestation plot (Figure 1A).

Critically, at 12K reads/bead, linear amplification has a low, 7.6 duplication rate, producing 1,818 UMIs per bead to enable the detection of 17/20 transcripts using a 50% dropout cut-off. In contrast, exponential amplification, at matched depth, has a 33.7 duplication rate, reducing the UMI number to 467 and resulting in only 13/20 transcripts attaining the 50% dropout cut-off. In addition, when the captured UMI were compared, 15/17 genes showed increased sensitivity obtained by linear amplification (Figure S3B).

Next, Constellation-Seq was scaled to 52 targets including 3 negative controls and compared with standard DropSeq (Table S2). Using 15k reads/bead, we demonstrated efficient use of the read space (93.5% reads from target genes) while increasing the average counts/cell 2.7-fold (Figure S3). Constellation-Seq dramatically reduced the degree of sparsity in the data which allows expressed transcripts to be accurately ranked (Figures 1B, 1C, and S4). Individual target transcript counts from Constellation-Seq were on average 83-fold higher. In addition, standard sequencing only detected 41 of the targets, while Constellation-Seq detected all 49 targets and none of the control genes (Figure 1D). The 8 transcripts exclusively detected by Constellation-Seq had average expression levels ranging from 0.03 to 2.60

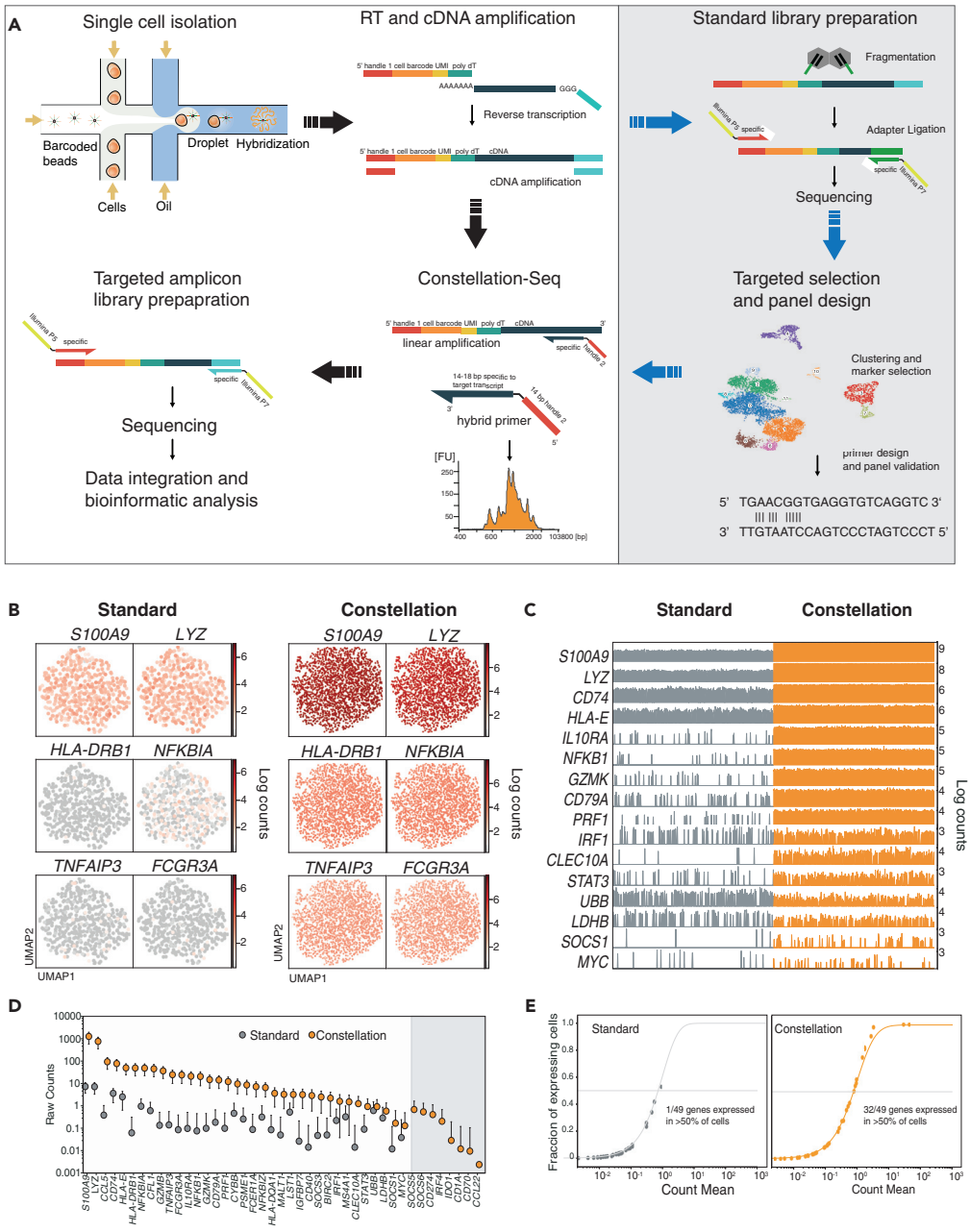


Figure 1. Constellation-Seq methodology and performance

(A) Schematic representation of the method: Constellation-Seq can be applied to any Smart-Seq-like library following the standard cDNA synthesis protocol. With a defined primer panel, Constellation-Seq can be applied directly to the cDNA library (black arrows). Otherwise, an aliquot of the cDNA can be used for bulk sequencing and after data analysis the panel of primers can be selected for hypothesis testing or to reduce the technical zeros (Blue arrows). Constellation-Seq includes a hybrid primer (14–18 bp specific sequence, black, adjacent to a common 14 bp handle 2, red) that binds to a specific target sequence in the cDNA library. Linear amplification of 500–1000 bp stretches of target transcripts allows selective enrichment of targets of interest, and the inclusion of the cell barcode and UMI sequences, leads to generation of the Constellation library, ready to use in next-generation sequencing. (B-E) Constellation-Seq was compared against standard sequencing using a panel of 52 targets on control beads.

(B) UMAP representation of control beads with standard sequencing compared with Constellation-Seq.

(C) A track plot showing the reduction in the data sparsity in a head to head comparison. Each bar represents a gene expression signal from a single cell. A full track plot is included as Figure S4.

(D) Individual target raw counts show ~100-fold sensitivity gains for Constellation-Seq, error bars represent SD.

Figure 1. *Continued*
(E) Dramatic reduction in technical zeros achieved by Constellation-Seq compared with DropSeq. At 2K UMI counts per bead 32/49 genes were detected in half of the beads (3 negative controls were not detected) using Constellation-Seq, whereas only 1 was detected with the same threshold using DropSeq.

counts, without length correlation. In practical terms, when using a 50% dropout cut-off, 32/49 are detected by Constellation-Seq and only 1/49 by standard DropSeq at a sequencing depth of 8k reads/bead (Figure 1E). Of merit, the sensitivity of Constellation-Seq cascades directly into significantly lower read requirements; the 32/49 transcripts above 50% cut-off are detected when reducing the depth to 4k reads/bead, with losses (28/49) only evident at 2k (Figure S5). This striking feature of Constellation-Seq presents the option to reduce the sequencing depth and associated experimental cost or increase the scale of the experiment.

Constellation-Seq reliably measures changes in gene expression

To explore the ability of Constellation-Seq to measure gene expression changes in response to perturbation of a cellular system, we challenged human PBMCs with the super antigen Staphylococcal enterotoxin B (SEB, 100 ng/mL, 16 hr). To compare methods 1,000 cells per treatment were sequenced (200K reads/cell for DropSeq and: 20K reads/cell for Constellation-Seq), Figure 2A). In this context, Constellation-Seq consistently detected low copy transcripts such as *GZMB, IRF4* and *SOCS1* with reduced drop-out and increased UMI counts at 10-fold lower sequencing depth. Differential gene expression was compared between control and stimuli for both standard DropSeq and Constellation-Seq. The fold change measurements correlated well between methods ($r = 0.62$, p value $= 8 \times 10^{-5}$, Figures 2B and 2C). Importantly, Constellation-Seq was 1.6 times more sensitive (assessed by the slope of the correlation between Constellation-Seq and DropSeq) to gene expression changes (Figure 2B), improving the resolution of typical activation features such as *NFKB1/NFKBIA* while maintaining comparable expression levels for stable transcripts unperturbed by stimulation (*e.g. CD74*). In summary, the linear amplification step in Constellation-Seq retains the authentic biological response, while measuring responses with greater sensitivity and resolving greater detail in the underlying process.

Constellation-Seq is compatible with the standard 10x Chromium Single Cell 3′ V3 protocol

Next the Constellation approach was reconfigured for use with the popular Chromium 10x Genomics technology using 6,000 CD14 enriched monocytes and amplified cDNA produced using the standard 10X Chromium protocol as the starting material. Following the linear amplification, the library tapestation plot is spiny, typical of targeted transcriptomics (Figure 3A). The targeted library was processed using the Nextera XT protocol. Constellation-Seq greatly improved the detection of transcripts of interest (Figure 3B). Constellation-Seq applied to the 10X library showed 22-fold greater sensitivity allowing reduction of the sequencing depth from 70k to 1.5K reads/cell, while distinguishing 5 clusters, including an activated monocyte sub-population (*CXCL8*). In comparison standard 10X at 1.5K reads/cell failed to resolve these subpopulations and activation states (Figures 3C and S6). Indeed, standard 10X requires 70k reads/cell to obtain the same results, inflating the experimental costs 46-fold (Figure S6) demonstrating both the sensitivity and financial gains achieved using the Constellation-Seq method.

Constellation- DropSeq can resolve rare DC populations

To demonstrate the applicability of Constellation-Seq for the analysis of specific cell subtypes within complex cellular systems, we designed a primer panel targeting 127 transcripts (Table S3) using a recent molecular classification (Villani et al., 2017) for the identification of DC subpopulations and their activation states. 4,000 human PBMCs were processed following the standard 10X Chromium 3′ protocol. While standard sequencing was able to segregate the blood cell types, including DCs and monocytes (Figure 4A), the technique was not sufficiently sensitive to reliably detect all the markers used for identifying DC sub-populations (Figure 4B), limiting the annotations to DC1 and DC6 subtypes. In contrast, the sensitivity of Constellation-Seq allowed the classification of expression markers for four DC subpopulations (DC1: *IDO1*,CCR7, DC2: PTAFR, DC4: *FCGR3A,AIF1,* and DC6: TCF4, JCHAIN (Villani et al., 2017) Figures 4C and 4D). Furthermore, Constellation-Seq provided greater marker detection sensitivity, increasing the average counts by almost 2 orders of magnitude (i.e *IDO1* detection 10X: 0–4 counts per ten thousand (CPTT), C10X: 01–120 CPTTs). The number of DCs detected using Constellation-Seq was substantially higher, 127 vs 51, due to more cells passing the QC filtering and increased clustering achieved by the increase in sequencing depth.

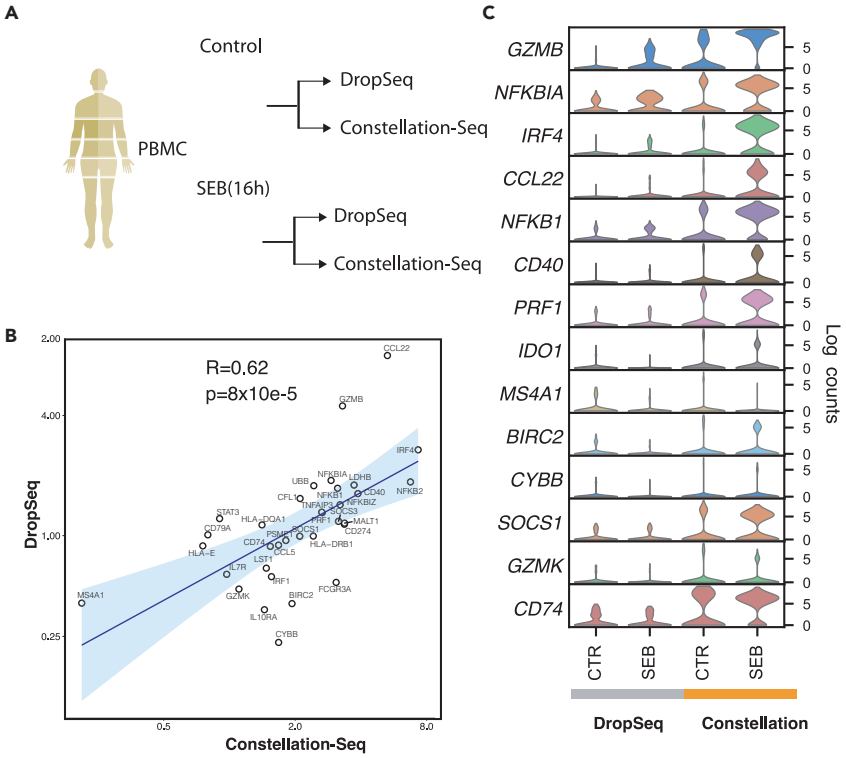


Figure 2. Constellation-Seq reliably measures gene expression changes with higher sensitivity

(A) Experimental design; PBMCs from healthy subjects (n = 3) were stimulated with Staphylococcal enterotoxin B (SEB) or media control for 16hr and analyzed using DropSeq and Constellation-Seq.

(B) Correlation of normalized gene expression fold changes induced by SEB as detected by DropSeq and Constellation-Seq. Pseudo-bulk counts for each gene used for the comparison.

(C) Comparative analysis of selected markers induced by SEB in cultured PBMCs. Violin plots in each row show the distribution and levels of each expressed gene in different culture conditions (CTR – media control, SEB – stimulated cells) and assessed by DropSeq (gray) and Constellation-Seq (orange). y axis represents normalized UMI counts.

DISCUSSION

The current sensitivity limits of single cell sequencing methods restrict the scope of biological investigations and impart substantial costs. The simplicity of Constellation-Seq allows inclusion in almost any single cell transcriptome library preparation pipelines involving SMART-Seq primers (DropSeq, Seq-Well, 10X and potentially InDrop). The multiplex scaling capacity is governed by available volume; a 300-plex assay is feasible for a 50 μL reaction volume (without affecting the normal library preparation pipeline; Figure S7). The highly multiplexed selection of transcripts of interest is at the expense of global transcriptome coverage, yet benefits from maximizing the efficient use of the NGS space to enable ultra-sensitive investigations. In this manner, the architecture of cellular systems can be understood with unprecedented resolution and biological processes can be mapped in exquisite detail. Central to Constellation-Seq is prior knowledge of the cellular system, where specific target selection lends strength to mechanistic studies or allows the prioritization of targets for perturbation studies. Additionally, Constellation-Seq can be implemented in drug discovery, delivering preliminary toxicity and efficacy screens for pharmacological compounds of interest. To gain entry to new biological scenarios and to define the targeted primer library for Constellation-Seq, various standard scRNA-seq approaches or bulk transcriptome analyses can first be applied to provide a global screen of the defining molecules and pathways of interest.

Increasing the sequencing depth on informative genes allows specific clusters in the UMAP space to be clearly resolved. In addition, the reduction in technical dropouts allows better cluster labeling, supported on well established surface markers, even at low mRNA expression levels. In addition, it opens the

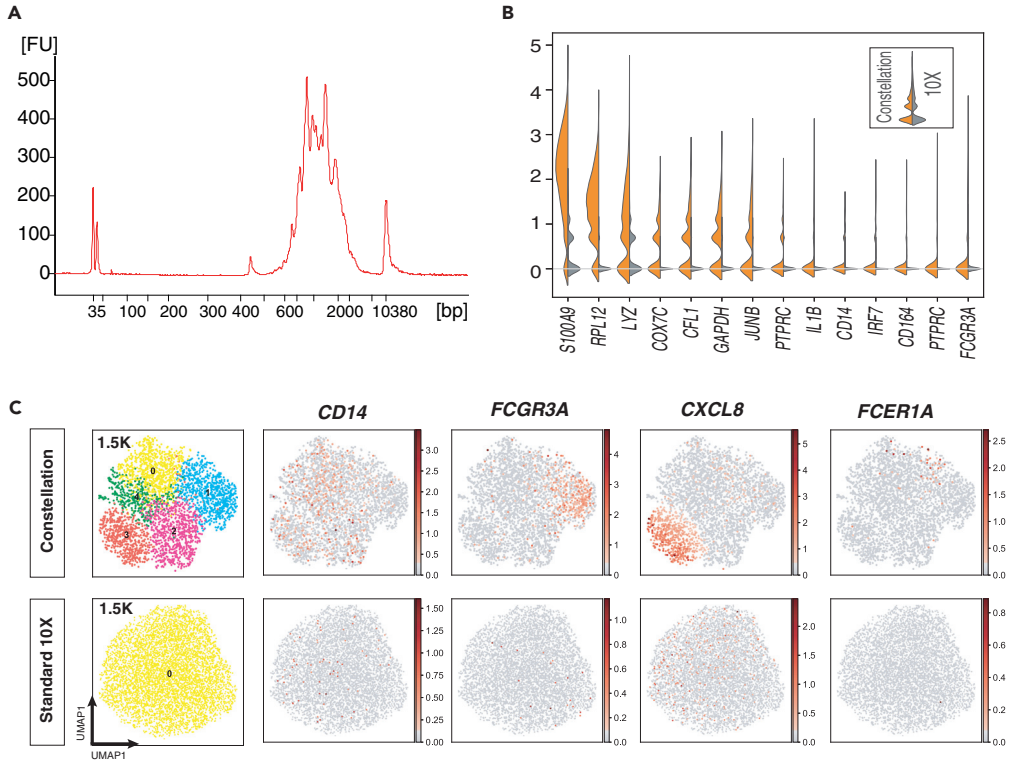


Figure 3. Constellation-Seq is compatible with standard 10-X protocol

(A) DropSeq assay of human monocytes (A) The spikes in the plot are characteristic for the Constellation-Seq method due to the selection of targets with distinct molecular weights.

(B) Direct comparison of gene expression in monocytes using Constellation-Seq vs 10X.

(C) Left: Constellation-Seq (orange) vs 10X (gray), gene expression measured in normalized UMI counts (C). UMAP plot of 6,000 monocyte transcriptomes assessed using 10X and C-10X. At 1.5K UMI counts per cell, C-10X shows more granularity than normal 10X at the same resolution. The enhanced sensitivity of C-10X is demonstrated using monocyte markers.

possibility to group the cells based on transcription factors, which may lead to a functional based cell classification. Our experiments exemplify how Constellation-Seq increases the cluster resolution of a population of interest (DCs) in the context of a mixed cell sample (PBMC), where DCs constitute less that 0.5% of the PBMC population (Ueda et al., 2003). We have not only confirmed the expression of markers proposed by Villani et al., falling into the dropout zone in the standard 10X experiment but were able to identify localization of *IDO1* transcript expression specifically to DC1. Importantly, Constellation-Seq resolved the transcriptomic signal of DCs, identifying 127 cells in contrast with 51 detected by standard 10X. This discrimination power is important for validation assays, where the sparsity of marker expression can be misleading for assigning a cell label, and the signal from more abundant cells dominates the population of interest. Application of Constellation-Seq in laboratory settings and in clinical practice will allow tracking of specific cell populations and their activation in health and disease without incurring significant cost.

Constellation-Seq builds on standard scRNA-Seq pipelines, to provide a cost-effective single cell transcriptomics approach for large-scale experiments, while addressing the issues of sensitivity and sparsity. With Constellation-Seq further savings emerge from shrinking the required sequencing depth to allow substantially larger experiments or simply more experiments. Other methods such as Hybridization of Probes to RNA for sequencing (HyPR-seq) and Seq-FISH can be used for targeted RNA detection method. However HyPR-seq requires multiple rounds of washes for probe hybridization and ligation which reduces cell recovery and may affect other genes (Marshall et al., 2020). Seq-FISH, which requires a spatial targeted method, provides an alternative for laboratories with the required infrastructure (Shah et al., 2016). To inform method selection by end users, the experimental economies, including time-finance trade-offs, of Constellation-Seq are compared with standard DropSeq and 10X approaches

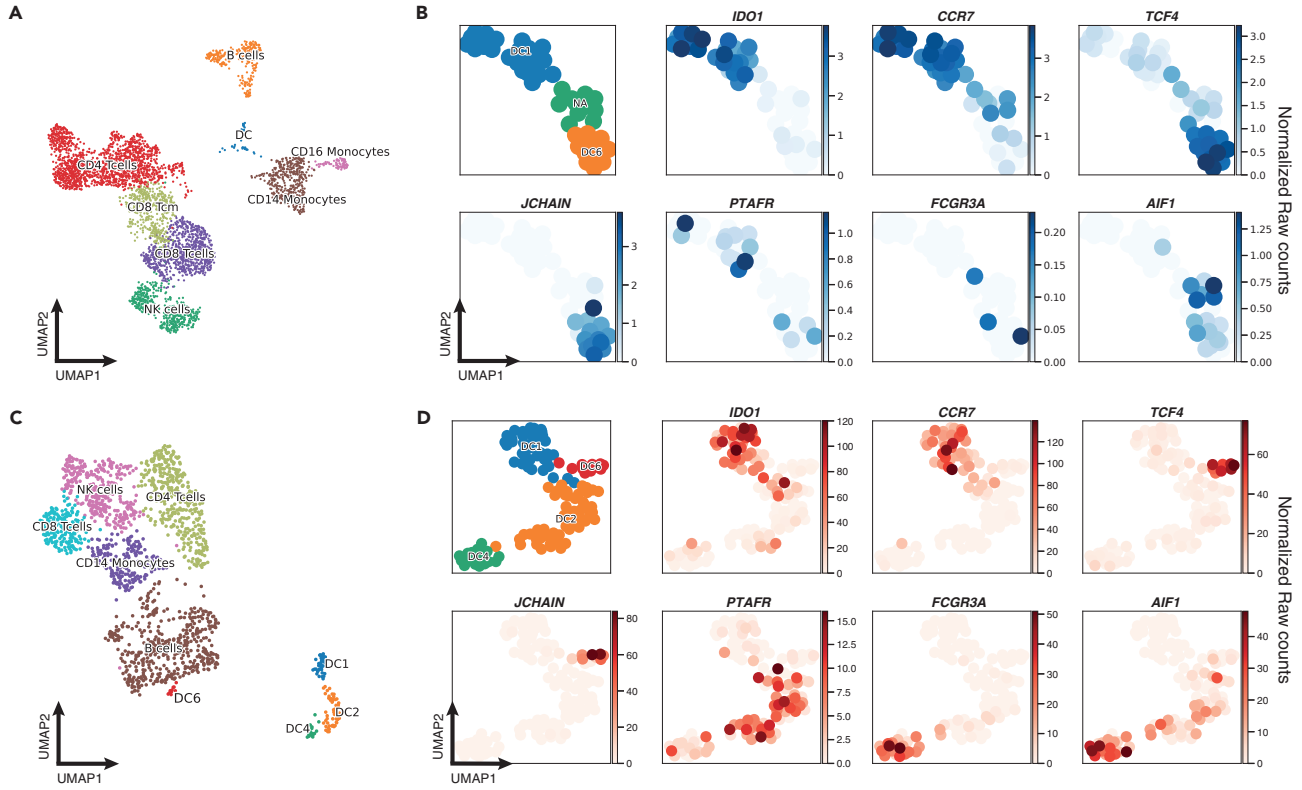


Figure 4. Constellation-Seq can resolve rare cell populations

10X Chromium Single Cell 3′ assay of human PBMC.

(A) UMAP projection of 4,182 PBMCs (Leiden r = 0.5, n_pcs = 20, n_neighbours = 20). Cells were grouped into eight clusters. Classification of PBMCs was inferred from the annotation of cluster-specific genes and based on expression of well-known markers of immune cell types using all detected genes.

(B) Sub-clustering of dendritic cells (n = 51), from the standard 10X protocol showing marker genes for DC cell populations as in the Villani paper (Villani et al., 2017).

(C) Constellation-Seq run of the sample with a panel of 127 genes related to DC biology (n = 1,697 cells).

(D) UMAP plot of the dendritic cell subsets (n = 127).

in Table S4. Beyond this, Constellation-Seq is accessible to resource limited laboratories, overall representing a step toward the democratization of single-cell transcriptomics and the broad-scale expansion of our understanding of biological systems.

Limitations of the study

A potential limitation of Constellation-Seq is that this approach requires previous knowledge for target gene selection. However, because the method can be used in the same cDNA sample used for standard sequencing, primer selection can be done with the standard pipeline and then Constellation-Seq applied to the same sample. Regarding the multiplex capability, up to now, we have multiplexed primers to detect a total of 127 different genes per single cell. Although this will be adequate for exploring a specific pathway, or cell type, complex samples and hypothesis may require a more extensive gene panels. Based on the concentrations, and current primer design capabilities, the method can be straightforwardly expanded to 300 targets in a single reaction. If more targets are needed, it will be possible to set up more reactions in parallel, with cDNA availability being the limiting factor.

Resource availability

Lead contact

Requests for further information and reagents should be directed to and will be fulfilled by the lead contact Marta E. Polak (m.e.polak@soton.ac.uk).

Materials availability

This study did not generate materials than can be shared.

Data and code availability

Raw data is available through ENA ENA:PRJEB41830). All notebooks used for bioinformatic analyses are available through GitHub, https://github.com/afvallejo/Constellation-DropSeq.

METHODS

All methods can be found in the accompanying Transparent methods supplemental file.

ETHICS DECLARATIONS

The PBMCs used in this study were obtained with ethical approval 17/EM/0349.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.102147.

ACKNOWLEDGMENTS

We are grateful to the subjects who participated in this study and Elena Vataga (Computational Modeling Group, University of Southampton) for assistance with the High-Performance Computing | HPC Platform. This study was initially funded by an MRC Discovery grant (MC_PC_15078), M.E.P. is funded by Sir Henry Dale Fellowship, Wellcome Trust. (Grant no 109377/Z/15/Z), and A.F.V. is funded by GSK (project ARCP006668).

AUTHOR CONTRIBUTIONS

A.F.V. designed, performed, and analyzed experiments, performed bioinformatic analyses and wrote the manuscript. M.E.P. and J.W. conceived the idea, planned the experiments, and contributed to writing. J.D. and A.G. performed the experiments. A.G., C.H.T. and R.J. contributed to the experimental plan and reviewed the manuscript.

DECLARATION OF INTERESTS

The authors declare no conflict of interest.

Received: June 16, 2020
Revised: December 9, 2020
Accepted: January 31, 2021
Published: March 19, 2021

REFERENCES

Bacher, R., and Kendziorski, C. (2016). Design and computational analysis of single-cell RNA-sequencing experiments. Genome Biol. *17*, 63.

Fromm, P.D., Kupresanin, F., Brooks, A.E., Dunbar, P.R., Haniffa, M., Hart, D.N., and Clark, G.J. (2016). A multi-laboratory comparison of blood dendritic cell populations. Clin. Transl. Immunol. *5*, e68.

Haniffa, M., Shin, A., Bigley, V., Mcgovern, N., Teo, P., See, P., Wasan, P.S., Wang, X.N., Malinarich, F., Malleret, B., et al. (2012). Human tissues contain CD141hi cross-presenting dendritic cells with functional homology to mouse CD103+ nonlymphoid dendritic cells. Immunity *37*, 60–73.

Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012). CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. Cell Rep. *2*, 666–673.

Lähnemann, D., Köster, J., Szczurek, E., Mccarthy, D.J., Hicks, S.C., Robinson, M.D., Vallejos, C.A., Campbell, K.R., Beerenwinkel, N., and Mahfouz, A. (2020). Eleven grand challenges in single-cell data science. Genome Biol. *21*, 1–35.

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.

Marshall, J.L., Doughty, B.R., Subramanian, V., Wang, Q., Chen, L.M., Rodriques, S.G., Zhang, K., Guckelberger, P., Fulco, C.P., and Nasser, J. (2020). HyPR-seq: single-cell quantification of chosen RNAs via hybridization and sequencing of DNA probes. PNAS *117*, 33404–33413.

Muzaki, A.R., Tetlak, P., Sheng, J., Loh, S.C., Setiagani, Y.A., Poidinger, M., Zolezzi, F., Karjalainen, K., and Ruedl, C. (2016). Intestinal CD103(+)CD11b(-) dendritic cells restrain colitis via IFN-gamma-induced anti-inflammatory response in epithelial cells. Mucosal Immunol. *9*, 336–351.

Polak, M.E., Borthwick, N.J., Gabriel, F.G., Jager, M.J., and Cree, I.A. (2008). Activation of tonsil dendritic cells with immuno-adjuvants. BMC Immunol. *9*, 10.

Riemondy, K.A., Ransom, M., Alderman, C., Gillen, A.E., Fu, R., Finlay-Schultz, J., Kirkpatrick, G.D., Di Paola, J., Kabos, P., Sartorius, C.A., and Hesselberth, J.R. (2019). Recovery and analysis of transcriptome subsets from pooled single-cell RNA-seq libraries. Nucleic Acids Res. *47*, e20.

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.

Salomon, R., Kaczorowski, D., Valdes-Mora, F., Nordon, R.E., Neild, A., Farbehi, N., Bartonicek, N., and Gallego-Ortega, D. (2019). Droplet-based single cell RNAseq tools: a practical guide. Lab Chip *19*, 1706–1727.

Shah, S., Lubeck, E., Zhou, W., and Cai, L. (2016). In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. Neuron *92*, 342–357.

Steinman, R.M. (2003). The control of immunity and tolerance by dendritic cell. Pathol. Biol. (Paris) *51*, 59–60.

Svensson, V. (2019). Droplet scRNA-seq is not zero-inflated. bioRxiv, 582064.

Tang, F., Lao, K., and Surani, M.A. (2011). Development and applications of single-cell transcriptome analysis. Nat. Methods *8*, S6–S11.

Ueda, Y., Hagihara, M., Okamoto, A., Higuchi, A., Tanabe, A., Hirabayashi, K., Izumi, S., Makino, T., Kato, S., and Hotta, T. (2003). Frequencies of dendritic cells (myeloid DC and plasmacytoid DC) and their ratio reduced in pregnant women: comparison with umbilical cord blood and normal healthy adults. Hum. Immunol. *64*, 1144–1151.

Vallejos, C.A., Risso, D., Scialdone, A., Dudoit, S., and Marioni, J.C. (2017). Normalizing single-cell RNA sequencing data: challenges and opportunities. Nat. Methods *14*, 565–571.

Villani, A.-C., Satija, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J., Griesbeck, M., Butler, A., Zheng, S., and Lazo, S. (2017). Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science *356*, eaah4573.

Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., and Enard, W. (2017). Comparative analysis of single-cell RNA sequencing methods. Mol. Cell *65*, 631–643.e4.

iScience, Volume *24*

Supplemental Information

Resolving cellular systems by ultra-sensitive and economical single-cell transcriptome filtering

Andres F. Vallejo, James Davies, Amit Grover, Ching-Hsuan Tsai, Robert Jepras, Marta E. Polak, and Jonathan West

	DropSeq	Targeted PCR	DART-Seq	Constellation-Seq
mRNA Capture	Poly T	Poly T	Specific Probe	Poly T
Library Construction	TSO dependant	TSO independant	TSO dependant	TSO independant
cDNA amplification	Whole	Targets Only	Targets Only	Targets+ Whole
Type of amplification	PCR	PCR	PCR	Linear

Figure S1. Comparison of Constellation-Seq with DropSeq and other targeted methods, Related to figure 1.

The methods use the same poly T capture probes, with the exception of the DART-Seq method that have probes extended with target-specific capture sequences. For targeted PCR and Constellation-Seq, the library construction will not require the TSO, which can improve the library complexity25. Following mRNA capture DropSeq and DART-Seq methods progress directly to PCR library preparation, whereas targeted PCR and Constellation-Seq methods first involve PCR and linear amplification cycles, respectively.

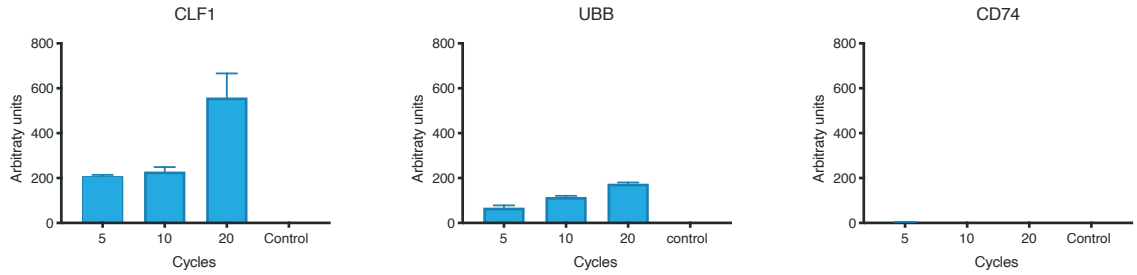


Figure S2 Expression analysis of a Constellation-Seq library containing CLF1 and UBB primers, Related to figure 1.

The library generated from control beads using linear amplification, at a primer concentration of 10 nMol and 65°C annealing temperature was tested with qPCR for expression of CLF1 and UBB as targeted genes and CD74 as a negative control. Data was processed using a semi quantitative approach26. Error bars represent standard deviation (SD).

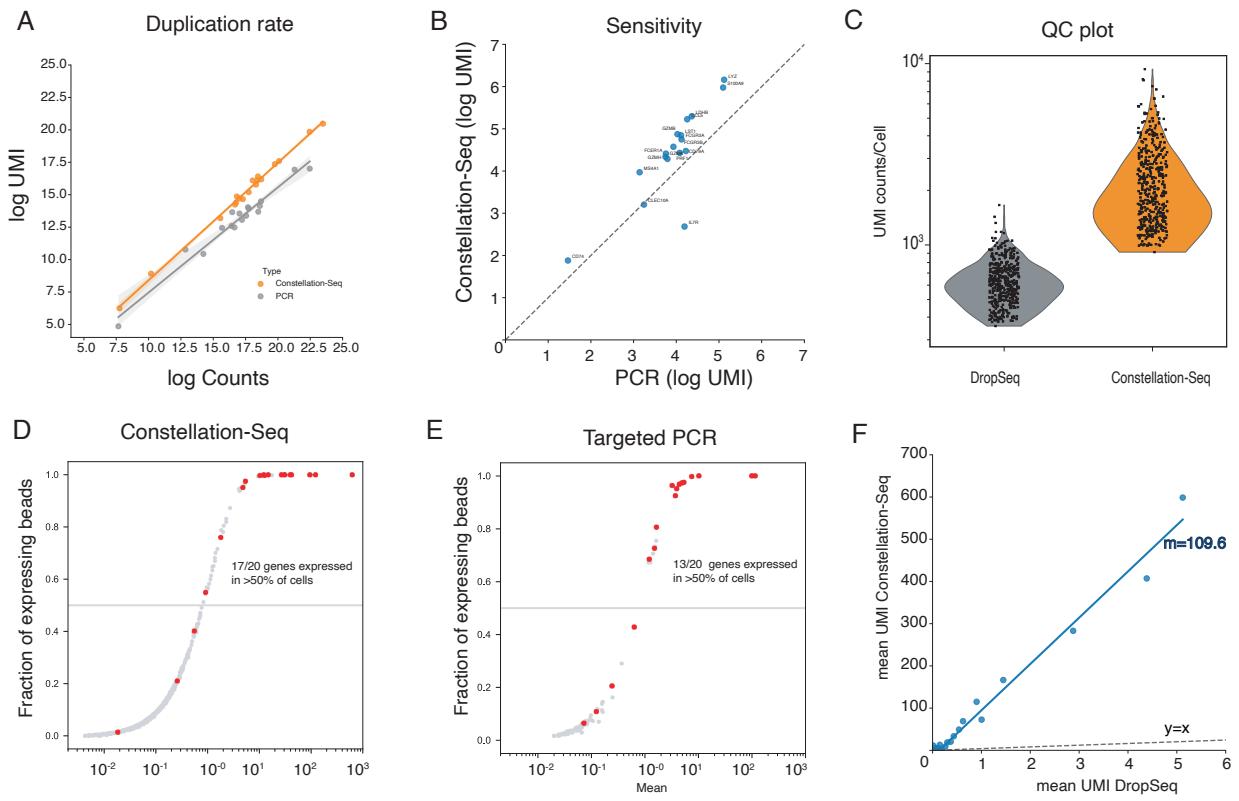


Figure S3 Head to head comparison of detection of 20 targets using linear vs targeted approach, Related to figure 1.

A) Duplication rate analysis at matched sequencing depth. UMI and counts were compared between Constellation-Seq and PCR. The slope was 7.6 and 33.5 for linear amplification and PCR respectively, showing that linear amplification was 4.4 times more sensitive. B) Comparative sensitivity analysis between Constellation-Seq and targeted PCR. Correlation of the UMI captured in both techniques at matched sequencing depth. 15/17 genes above the x=y diagonal demonstrates the increased sensitivity of linear amplification. C) Constellation-Seq counts per bead are 2.7-fold higher than with DropSeq. D-E) Drop-out rate vs mean expression levels in targeted PCR and Constellation-Seq. Red dots represent genes included in the library. F) Comparison of the UMIs captured in DropSeq vs Constellation-Seq.

Figure S4. DropSeq and Constellation-Seq comparison for the detection of a panel of 52 genes, Related to figure 1.

A tracksplot of gene expression for high, medium and low expressed genes detected using Drop-Seq (grey) and Constellation-Seq (orange) with control beads. A total of 41/52 genes were detected in both methods. Each bar shows the UMI counts signal from a single cell.

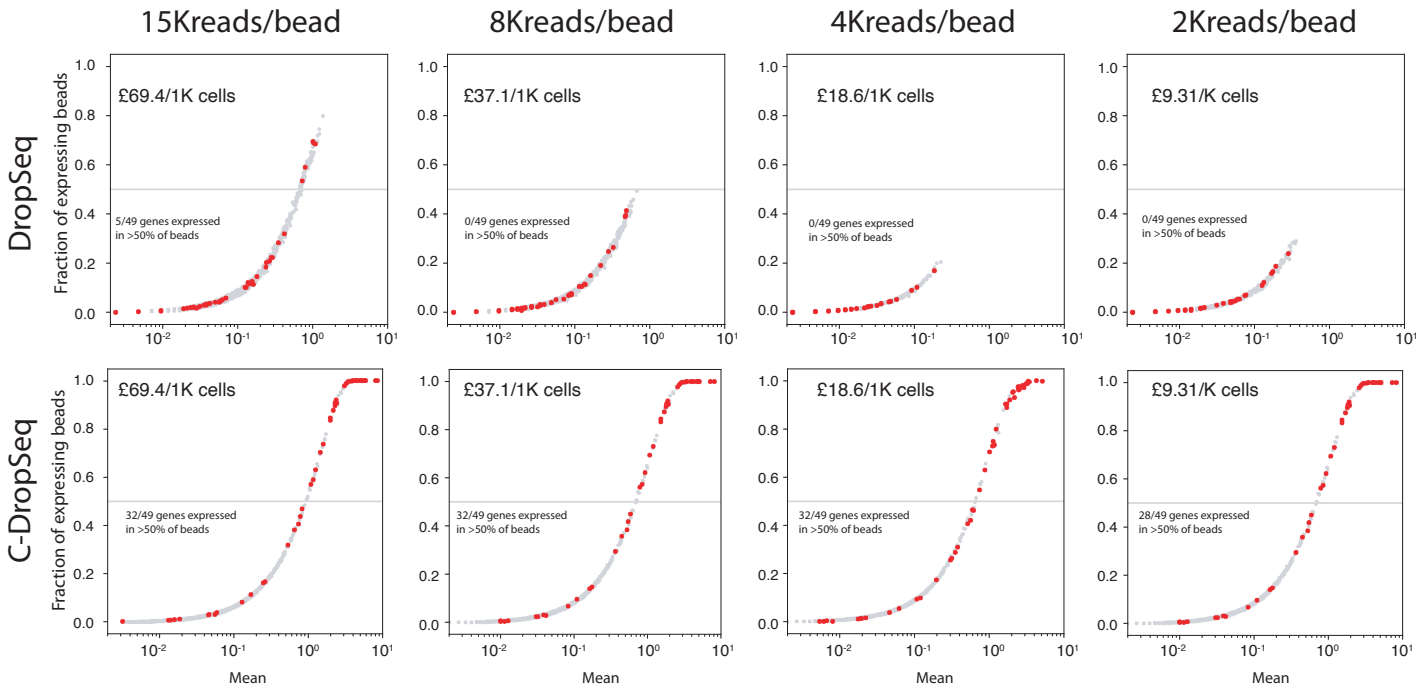


Figure S5 DropSeq and Constellation-Seq sensitivity comparison with varying sequencing depth, Related to figure 1.

The total number of counts for each target was calculated and compared between DropSeq (top) and Constellation-Seq (bottom). The fraction of beads with detected target expression vs mean level of target expression are shown for each gene. The horizontal line indicates the 50% of beads detection threshold. Red: genes from the panel. Grey: genes not included in the panel. Numbers are the predicted effective cost for 1,000 cells.

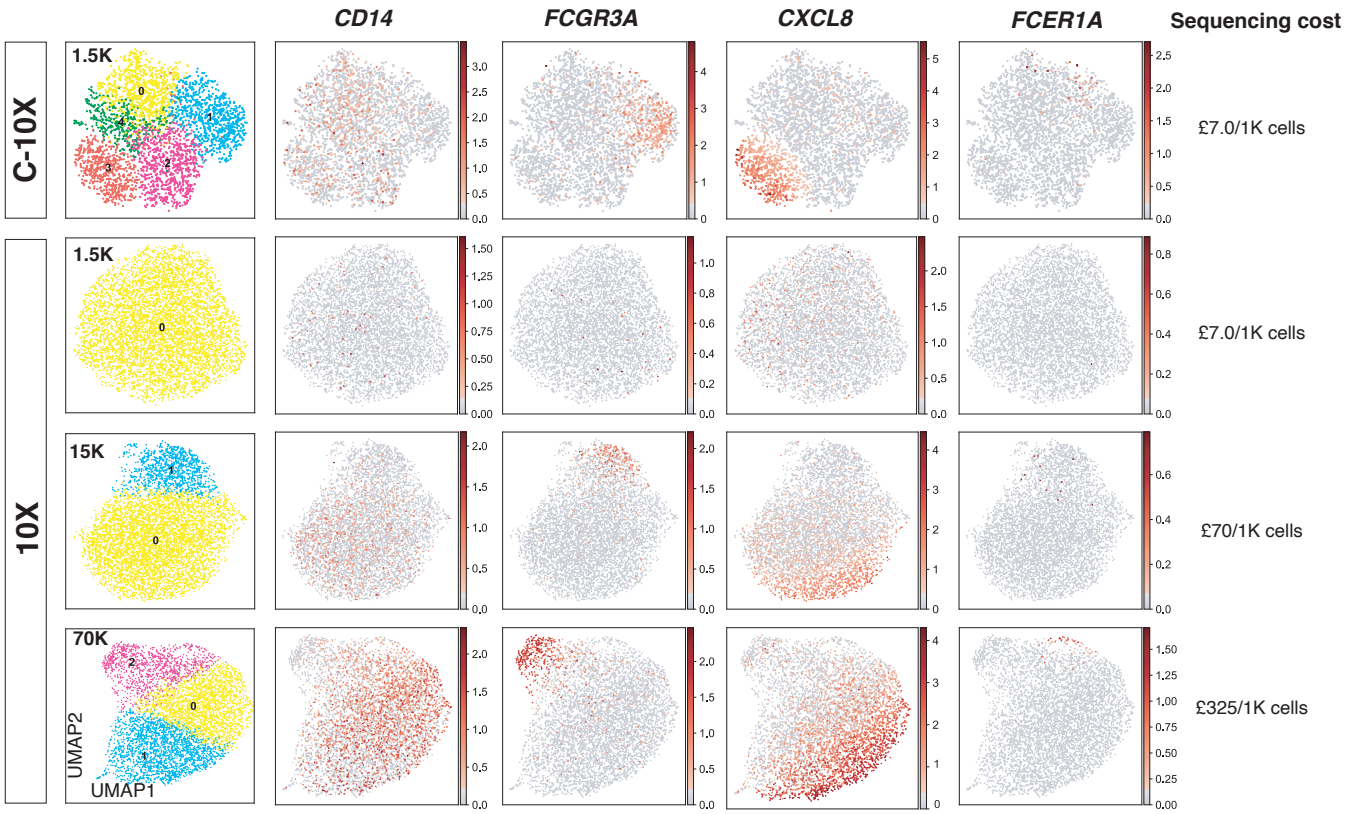


Figure S6 Constellation-Seq can be translated to other single-cell protocols with substantial savings,Related to figure 3.

UMAP plots showing comparison of single cell sequencing of 6,000 monocytes using C-10X at 1,500 reads per cell sequencing depth with standard 10X at varying sequencing depths. Column 1: clustering results, Leiden r=0.5, n_neighbours = 20, columns 2-5: examples of monocyte activation expression markers. Colour denotes gene expression level, as indicated by the legend (normalised UMI counts). Right: the effective cost of sequencing 1,000 cells.

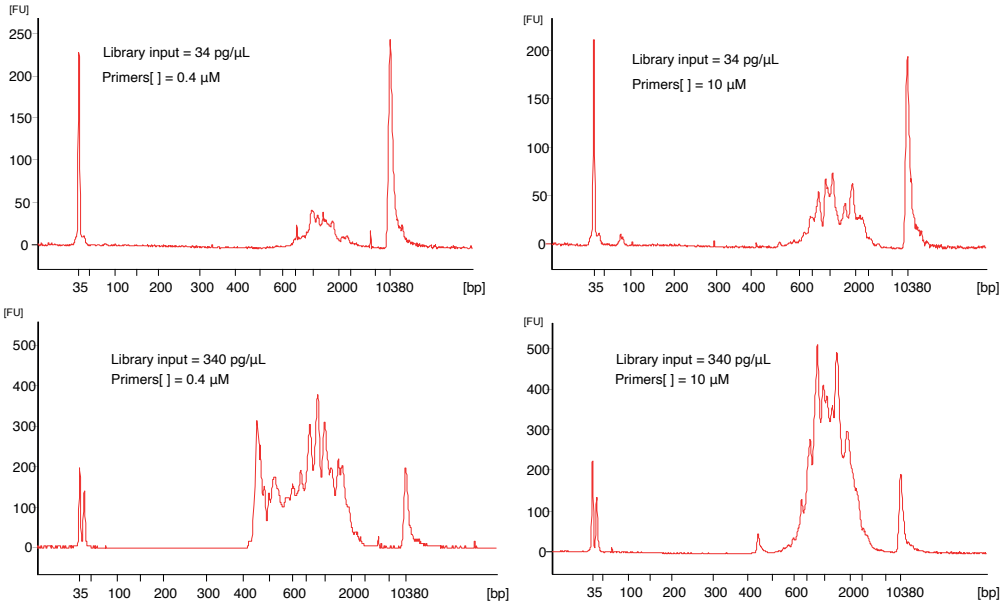


Figure S7 C-10X library optimization,Related to figure 3.

Typical plot from the bioanalyser (Agilent) showing the library input and primer concentration effect on library preparation with C-10X. Top: library input – 34 pg/mL, bottom – library input 340 pg/mL. Left: Primer concentration c= 0.4 µMol, right: Primer concentration c= 10 µMol. Y axis shows fluorescence units (FU) indicating signal intensity and product concentration. The spikes in the plot are characteristic for Constellation-Seq the targeted transcriptomics approach due to the selection of targets with distinct molecular weights.

Transparent Methods

Primer Design

Primers targeting genes of interest were designed using Beacon Designer primer design software (PREMIER Biosoft, California US). The last 14 bases from the SMART primer sequence (TATCAACGCAGAGT) were added to the 5' end of the designed primers. Desired features of primers included: a length between 28-32 base pairs, 40-60% GC content, a primer melting temperature between 52-58°C, and with minimal chance of secondary structures being produced.

Negative control beads

RNA from fresh PBMCs was extracted using RNeasy Plus Mini Kit (Qiagen). Control beads were generated by adding a solution of PBMC RNA at 10 pg/bead, making the RNA content in each droplet equivalent. 200 µL of reverse transcriptase mix (75 µL water, 40 µL Maxima 5x RT buffer, 40 µL 20% Ficoll PM-400, 20 µL 10 mM dNTPs, 5 µL RNase inhibitor and 10 µL Maxima H- RTase) was added to each bead sample. 10 µL of 50 µM TSO was added to the DropSeq controls, whereas for Constellation-Seq no TSO was used. Samples were incubated with rotation at room temperature for 30 minutes followed by 90 minutes at 42°C with continuous rotation. Beads were washed with 1 mL TE-SDS (10 mM Tris, pH 8.0, 1 mM EDTA, 5% SDS) and twice with 1 mL TE-TW (10 mM Tris, pH 8.0, 1 mM EDTA, 0.01% Tween-20). Finally, beads were washed with 1 mL 10 mM Tris pH 8.0, and stored at 4°C.

Cell preparation

Human blood was collected from donors with written consent and ethical approval (study number: 17/EM/0349). PBMCs were extracted immediately using Lymphoprep™ (STEMCELL Technologies) and incubated at 37°C with 5% CO_2. For SEB stimulation experiments cells were cultured in 24 well plates at $2x10^6$ cells/mL for 16h with or without SEB, using a final SEB concentration of 100 ng/mL. For LPS stimulation experiments cells were cultured in 24 well plates at $2x10^6$ cells/mL for 4h with or without

LPS, using a final LPS concentration of 1 μg/mL. Following the incubation period cells were harvested, washed in PBS and counted. 180,000 cells were taken for encapsulation. CD14+ monocytes for the 10X experiment were purchased from Tissue solutions (Glasgow, UK).

DropSeq

DropSeq library preparation and sequencing was performed as described previously(Macosko et al., 2015). Briefly, single cells were co-encapsulated with beads in droplets using the microfluidic design provided by Macosko *et al (Macosko et al., 2015)*. After cell lysis, cDNA synthesis was carried out (Maxima Reverse Transcriptase, Thermo Fisher), followed by PCR (Kapa Hotstart Ready mix, 15 cycles: 4 at 67°C, 11 at 65°C). cDNA libraries were tagmented and PCR-amplified (Nextera tagmentation kit, Illumina). Finally, libraries were pooled and sequenced on an Illumina Nextseq500, (paired end 20x50bp reads).

Constellation-Seq of DropSeq libraries

For Constellation DropSeq, experiments were processed as normal from encapsulation through to extraction and purification of beads from the droplet emulsion. During reverse transcription however, the template switching oligo (TSO) was absent from the reaction*. This resulted in cDNA fragments without SMART primer binding sites at the 3' end of the Macosko bead primers. Hybrid primers were pooled at 10 μM. A 50 μL amplification mix was added (25 μL 2X Kapa HiFi Hotstart Readymix, 10 μM primer pool, 24.6 μL water) to aliquots of 2,000 beads (~100 STAMPs). 20 rounds of linear amplification (at 60°C) were first performed before continuing the standard Drop-Seq protocol for library preparation with PCR amplification and tagmentation. cDNA libraries were purified twice using AMPure XP magnetic beads (Beckman Coulter) (1:0.6) and libraries assessed using the Agilent Bioanalyser (KIT) before tagmentation and Next-seq sequencing.

*Standard reagents including the TSO can be used with the caveat of transcript noise generated by the reverse SMART primer.

10x Chromium Single Cell libraries

Single cell libraries were generated using the Chromium Single Cell 3′ library and gel bead kit v3.1 from 10x Genomics. Briefly, 10,000 cells were loaded onto a channel of the 10x chip to produce Gel Bead-in-Emulsions (GEMs). This underwent reverse transcription to barcode RNA before clean-up and cDNA amplification followed by enzymatic fragmentation and 5′ adaptor and sample index attachment using the Nextera XT Library preparation kit (Illumina). Libraries were sequenced on the MiSeq500 (Illumina) with 28x60 bp paired-end sequencing.

Constellation-Seq of 10X Chromium libraries

For Constellation-Seq of 10X libraries, 395 pg of cDNA were used for linear amplification comprising 20 rounds of linear amplification (60°C) using a pool of primers at 40 nM and 0.4 μM of a P5 3'blocked primer. A 40 μL amplification mix was added (20 μL 2X Kapa HiFi Hotstart Readymix, primer pool and P5 blocked primer) to 10 μL of cDNA library. cDNA libraries were purified twice using AMPure XP (Beckman Coulter) magnetic beads (1:0.6) and libraries assessed using a Bioanalyser before tagmentation and Next-seq sequencing on an Illumina Nextseq500, (paired end 28x60bp reads).

Real Time PCR

Control beads were used to assess the specificity of Constellation-Seq . 400 control beads per well were used as starting material. Constellation-Seq libraries were produced by linear amplification using two control primers (CFL1 and UBB from IDT) for 5, 10 or 20 cycles. Libraries were purified twice using 0.6X AMPure XP magnetic beads (Beckman Coulter) and eluted with 20 μL 1xTE, pH 8.0. Constellation-Seq libraries were tested using specific primers designed within the amplicon region including a negative control, CD74. 2 μL of the Constellation-Seq library was amplified in iTaq™ Universal SYBR (Bio-Rad) containing 200 nM of CFL1, UBB or CD74 primers. Amplification was undertaken in technical triplicates on a HT7900 Fast Real-Time PCR System (Applied Biosystems). Quantification was achieved against a serial dilution calibration curve of the pool of samples in each plate. C_t values were thresholded at 0.1 relative fluorescence units (RFU).

Bioinformatic pipelines

Alignment, read filtering, barcode and UMI counting were performed using kallisto-bustools(Melsted et al., 2019). High quality barcodes were selected based on the overall UMI distribution using emptyDrops(Lun et al., 2019). All further analyses were run using the Python-based Scanpy(Wolf et al., 2018). To remove low quality cells, we filtered cells with a high fraction of counts from mitochondrial genes (20% or more) indicating stressed or dying cells(Macosko et al., 2015). In addition, genes expressed in less than 20 cells were excluded.

Cell by gene count matrices of all samples were concatenated to a single matrix and values log transformed. To account for differences in sequencing depth or cell size UMI counts were normalized using quantile normalization. The top variable genes were selected based on normalized dispersion. This output matrix was input to all further analyses except for differential expression testing where all genes were used.

Visualization and clustering

A single-cell neighbourhood graph was computed on the 50 first principal components that sufficiently explain the variation in the data using 20 nearest neighbours. Uniform Manifold Approximation and Projection (UMAP) was run for visualization. For clustering and cell type identification Leiden-based clustering (Traag et al., 2019) at 0.5 resolution was used. Cell types were annotated based on the expression of known marker genes.

Supplemental References

Lun, A. T. L., Riesenfeld, S., Andrews, T., Dao, T. P., Gomes, T., Participants in the 1st Human Cell Atlas, J. & Marioni, J. C. 2019. EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol,* 20**,** 63.

Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M., Trombetta, J. J., Weitz, D. A., Sanes, J. R., Shalek, A. K., Regev, A. & Mccarroll, S. A. 2015. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell,* 161**,** 1202-1214.

Melsted, P., Booeshaghi, A. S., Gao, F., Da Veiga Beltrame, E., Lu, L., Hjorleifsson, K. E., Gehring, J. & Pachter, L. 2019. Modular and efficient pre-processing of single-cell RNA-seq. *BioRxiv***,** 673285.

Traag, V. A., Waltman, L. & Van Eck, N. J. 2019. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep,* 9**,** 5233.
Wolf, F. A., Angerer, P. & Theis, F. J. 2018. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol,* 19**,** 15.