Simultaneous multiplexed amplicon sequencing and transcriptome profiling in single cells

Authors: Mridusmita Saikia[1,2,*], Philip Burnham[1,*], Sara H. Keshavjee[1], Michael F. Z. Wang[1], Pablo Moral-Lopez[2], Meleana M. Hinchman[2], Charles G. Danko[2], John S. L. Parker[2], Iwijn De Vlaminck[1]
*These authors contributed equally
To whom correspondence should be addressed: vlaminck@cornell.edu

Affiliations:
[1]Meinig School of Biomedical Engineering, Cornell University, Ithaca, NY 14853, USA
[2]Baker Institute for Animal Health, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA

Abstract: High-throughput single-cell RNA sequencing technology has provided important insights into cellular complexity and transcriptome dynamics. However, current implementations of this technology are limited to capturing information from the ends of A-tailed messenger RNA (mRNA) transcripts. Here, we describe a versatile technology, Droplet Assisted RNA Targeting by single cell sequencing (DART-seq), that surmounts this limitation allowing investigation of all regions of the polyadenylated transcriptome, as well as measurement of other classes of RNA in the cell. We applied DART-seq to simultaneously measure transcripts of the segmented dsRNA genome of a reovirus strain, and the transcriptome of the infected cell. In a second application, we used DART-seq to simultaneously measure natively paired, variable region heavy and light chain (VH:VL) amplicons and the transcriptome of human B lymphocyte cells.

INTRODUCTION

High-throughput single-cell RNA-seq (scRNA-seq) is being widely adopted for phenotyping of cells in heterogeneous populations[1–4]. The most common implementations of this technology utilize droplet microfluidics to co-encapsulate single cells with beads that are modified with barcoded oligos to enable capturing the ends of RNA transcripts[2–4]. Although these approaches provide a means to perform inexpensive single-cell gene expression measurements at scale, they are limited to assaying the ends of mRNA transcripts. Therefore, they are ill-suited for the characterization of non-A-tailed RNA, including the transcripts of many viruses, viral RNA genomes, and non-coding RNAs. They are also uninformative of RNA segments that are located at a distance greater than a few hundred bases from transcript ends that often comprise essential functional information, for example the complementarity determining regions (CDRs) of immunoglobulins (B cell antibody)[5]. Additionally, these techniques are often unable to provide information on low copy number transcripts and splice variants[6].

Here we report DART-seq, a method that combines enriched measurement of targeted RNA sequences, with unbiased profiling of the poly(A)-tailed transcriptome across thousands of single cells in the same biological sample. DART-seq achieves this by implementing a simple and inexpensive alteration of the Drop-seq strategy[2]. Barcoded primer beads that capture the poly(A)-tailed mRNA molecules in Drop-seq are enzymatically modified using a tunable ligation chemistry[7]. The resulting DART-seq primer beads are capable of priming reverse transcription of poly(A)-tailed transcripts as well as any other RNA species of interest.

DART-seq is easy to implement and enables a range of new biological measurements. Here, we explored two applications. We first applied DART-seq to profile viral-host interactions and viral genome dynamics in single cells. We implemented two distinct DART-seq designs to investigate murine L929 cells (L cells) infected by the reovirus strain Type 3 Dearing (T3D). We demonstrate the ability of DART-seq to profile all 10 non-A-tailed viral gene transcripts of T3D reovirus individually, as well as to recover a complete genome segment, while simultaneously providing access to the transcriptome of the infected L cells. In the second application, we applied DART-seq to determine natively paired antibody sequences of human B cells. DART-seq was able to determine B cell clonotypes, as well as variable heavy and light (VH:VL) pairings, even in mixed human peripheral blood mononuclear cells (PBMCs), highlighting the versatility of the approach.

RESULTS

DART-seq primer bead synthesis

Droplet microfluidics based scRNA-seq approaches rely on co-encapsulation of single cells with barcoded primer beads that capture and prime reverse transcription of mRNA molecules expressed by the cell[2–4]. In Drop-seq, the primers on all beads comprise a common sequence used for PCR amplification, a bead-specific cell barcode, a unique molecular identifier (UMI), and a poly-dT sequence for capturing polyadenylated mRNAs and priming reverse transcription. To enable simultaneous measurement of the transcriptome and multiplexed RNA amplicons in DART-seq, we devised a scheme to enzymatically attach custom primers to a subset of poly-dTs on the Drop-seq bead (Fig 1a). This is achieved by annealing a double stranded toehold probe with a 3' ssDNA overhang that is complementary to the poly-dT sequence of the Drop-seq primers. The toehold is then ligated to the bead using T4 DNA ligase. Toeholds with a variety of different sequences can be attached to the same primer beads in a single reaction in this manner. The complementary toehold strand is removed after ligation. We examined the efficiency and tunability of the probe ligation reaction using fluorescence hybridization assays. Here, fluorescently labeled DNA hybridization probes were designed for complementarity to ligated primer sequences. We found that the measured fluorescence signal after ligation is proportional to the number of toeholds included in the ligation reaction (Fig 1b).

After synthesis of DART-seq primer beads, DART-seq follows the Drop-seq workflow without modification (see Methods). Briefly, cells and barcoded primer beads are co-encapsulated in droplets using a microfluidic device. Cellular RNA is captured by the primer beads, and is reverse transcribed after breaking the droplets. The DART-seq beads prime reverse transcription of both A-tailed mRNA transcripts and RNA segments complementary to the custom primers ligated to the beads (Fig 1c). The resulting complementary DNA (cDNA) is PCR-amplified, randomly fragmented via tagmentation, and again PCR amplified to create libraries for sequencing. Sequences of mRNAs and RNA amplicons derived from the same cells are identified by decoding cell-specific barcodes, allowing for gene expression and amplicon measurements across individual cells.

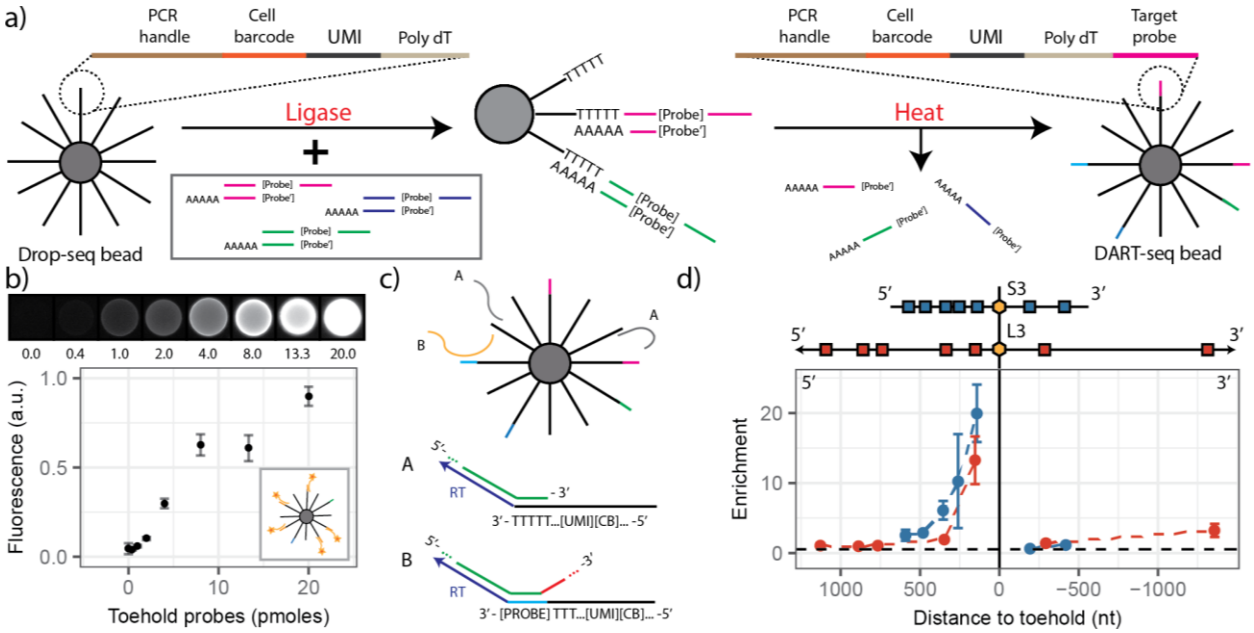

Fig. 1: DART-seq primer bead synthesis and validation of RNA priming. (a) Protocol for converting Drop-seq primer beads (left) to DART-seq primer beads (right). (b) Fluorescence hybridization assay to evaluate probe ligation efficiency. Measured mean fluorescence intensity in the Cy5 channel as function of toehold concentration (> 40 beads per concentration, error bars are 95% CI). (c) Schematic of reverse transcription (RT) priming of poly(A)-tailed mRNA (A) and RNA segments complementary to custom primers (B). (d) The relative enrichment of PCR amplicons as a function of distance to the capture site for two viral genes (S3 in blue and L3 in red, error bars are 95% CI). The black dotted line represents the enrichment of the *Gapdh* host gene.

We measured the abundance of amplicons in the resulting sequencing libraries using quantitative PCR (Fig 1d, qPCR). Here, we compared sequencing libraries of T3D reovirus-infected L cells generated by Drop-seq and libraries for the same cells generated by DART-seq with amplicons targeting all ten genome segments of the virus. Reovirus mRNAs lack poly(A)-tails, and thus should not efficiently captured by Drop-seq. We designed seven PCR assays with 84-120 bp amplicons distributed across the L3 and S3 viral genome segments. We observed significant RNA enrichment upstream (5' end) of the toehold ligation site for both the L3 and S3 segment (Fig 1d). No enrichment was observed in a host transcript (*Gapdh*). As expected, there was no enrichment downstream of the toehold ligation site (3' end). Consistent with sequencing library preparation via tagmentation, we found that the degree of enrichment achieved by DART-seq at a given position decreased exponentially with distance from the target.

DART-seq enables investigation of the heterogeneity of cellular phenotypes and viral genotypes during viral infection.

There is a great need for novel single cell genomics tools that can dissect the heterogeneity in viral genotypes and cellular phenotypes during viral infection[8]. We used DART-seq to examine infection of murine L cells with T3D reovirus. The reovirus polymerase transcribes non-A-tailed mRNAs from each of its 10 dsRNA genome segments[9,10]. We infected L cells at a multiplicity of infection of 10 (MOI 10), and allowed the virus to replicate for 15 hours after inoculation, creating a condition for which nearly all cells are infected (Fig 2a). We performed Drop-seq and DART-seq experiments on infected L cells and non-infected L cells as control. We implemented two distinct DART-seq designs. The first DART-seq design targeted each viral genome segment with a single amplicon. The second DART-seq design was comprised of seven amplicons targeting loci distributed evenly across the S2 genome segment (Fig 2b).

To determine the efficiency by which DART-seq retrieves viral transcripts near the target sequence, we analyzed the per-base coverage of positions upstream of the DART-seq target sites. For DART-seq design-1, we observed a mean enrichment of 34.7x in the gene regions 200 nt upstream of the ten toeholds. In both DART-seq design-1 and 2, all targeted sites were enriched compared to standard Drop-seq beads (Fig 2c,d). Viral transcripts were detected in Drop-seq libraries upstream of A-rich sequences in the viral genome, consistent with spurious priming of reverse transcription by poly-dT sequences on the oligo, as expected for Drop-seq. For example, a 200 nt gene segment upstream of an A_5 sequence on segment M3 (position 1952) was significantly enriched in the Drop-seq dataset (Fig 2c; marked by arrow).

To test the utility of DART-seq to measure the heterogeneity of viral genotypes in single infected cells, we used DART-seq design-2 (Fig 2b), which was tailored to retrieve the complete S2 viral gene segment. The S2 segment encodes inner capsid protein σ2. Across cells with at least 1500 UMIs, DART-seq design-2 increased the mean coverage across the S2 segment 430-fold compared to Drop-seq, thereby enabling the investigation of the rate and pattern of mutations (Fig 2e). 176 single-nucleotide variants (SNVs) were identified across the S2 segment (minor allele frequency greater than 10%, and per-base-coverage greater than 50x). Mutations from guanine-to-adenine (G-to-A) were most common (58%; Fig 2f, top). We did not observe such a mutation pattern in a highly-expressed host transcript (*Actb*). We examined the mutation load of viral transcripts at the single cell level, and observed a wide distribution in mutation load, with a mean G-to-A conversion rate of 13%, and up to 41% (Fig 2f, bottom). The reason for this level of hypermutation is unclear. G-to-A transamidation is an uncommon post-transcriptional modification that is not been previously seen as a host response to viral

infection[11,12]. The high rate of G-to-A transition in the viral transcript could also be secondary to a defect in the fidelity of viral transcription. The T3D strain used in this study has strain-specific allelic variation in the viral polymerase co-factor, μ2, that has been shown to affect the capacity of μ2 to associate with microtubules and the encapsidation of viral mRNAs within capsids[13,14].

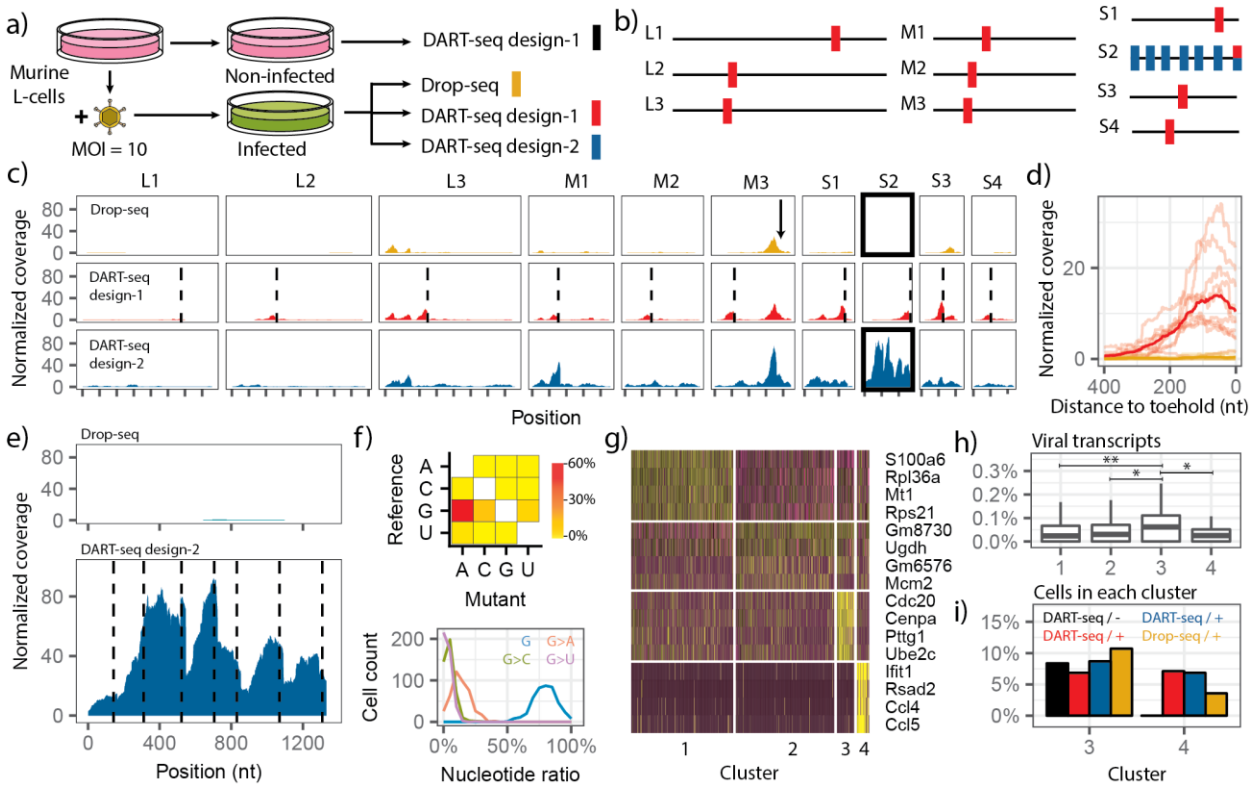

Figure 2 - DART-seq reveals heterogeneity in viral genotypes and host response to infection. (a) Experimental design. Single cell analysis using Drop-seq and two distinct DART-seq designs of murine L cells infected with a reovirus, and a non-infected control. (b) Schematic of two DART-seq designs. Design-1 (red bars) targets all 10 reovirus gene segments (3 x L (Large), 3 x M (Medium), and 4 x S (Small) segments). Design-2 (blue bars) targets seven loci on the S2 gene segment. (c) Comparison of the relative sequence coverage of the 10 reovirus gene segments (columns) for three different library preparations (rows). The arrow indicates an A_5 pentanucleotide sequence part of segment M3. Dotted lines indicate DART-seq target positions. (d) Per-base coverage upstream (5' end) of 10 toeholds of DART-seq design-1 (light red, average shown in dark red), and mean coverage achieved with Drop-seq (yellow). (e) Per-base coverage of the S2 gene segment achieved with DART-seq design-2 (bottom, dashed lines indicate toehold positions) and Drop-seq (top). (f) Frequency and pattern of base mutations. Across all cells, the average nucleotide profile for positions on the S2 segment with SNPs such that the major allele is < 90% are shown (top); the distribution of nucleotide ratios for positions with reference nucleotide G is depicted for single cells (bottom). (g) Clustering analysis of reovirus infected L cells (DART-seq design-1). Hierarchical clustering of clusters displayed as a heatmap (yellow/purple is higher/lower expression). (h) Relative abundance of viral transcripts in L-cell clusters (* and ** indicates significant *p*-value of 10^{-3} and 10^{-4}, respectively). (i) Fraction of cells in meta-clusters for four experiments depicted in panel a with assay type and infection status (+ or -) indicated.

To identify distinct host cell populations based on patterns of gene expression, we performed dimensional reduction and unsupervised clustering using approaches implemented in Seurat[15]. We identified four distinct cell clusters for the monoculture infection model (DART-seq design-1, Fig 2g). Two major clusters comprised of cells with elevated expression of genes related to viral RNA transcription and replication (*Rpl36a*, cluster 1) and metabolic pathways (*Ugdh*, cluster 2). Two additional clusters were defined by the upregulation of genes related to mitotic function (*Cdc20*, *Cenpa*; cluster 3) and innate immunity (*Ifit1*, *Rsad2*; cluster 4), respectively (Fig 2g). The abundance of viral gene transcripts relative to host transcripts was significantly elevated for cells in cluster 3 (n = 69 of 927 total cells) compared to cells in all other clusters (Fig 2h; two-tailed Mann Whitney U test, $p = 1.0 \times 10^{-4}$). We merged datasets for the Drop-seq and three DART-seq assays and quantified the cell type composition for each experiment. We did not observe cells related to cluster 4 (immune response), for the non-infected control, as expected (Fig 2i). These results support the utility of DART-seq to study the single cell heterogeneity in viral genotypes and cellular phenotypes during viral infection.

DART-seq allows high-throughput paired repertoire sequencing of B lymphocytes

As a second application of DART-seq, we explored the biological corollary of viral infection, the cellular immune response. The adaptive immune response is reliant upon the generation of a highly diverse repertoire of B lymphocyte antigen receptors (BCRs) , the membrane-bound form of antibodies expressed on the surface of B cells, as well as antibodies secreted by plasmablasts[16,17]. We applied DART-seq to investigate the B cell antibody repertoire in human PBMCs. We compared the performance of DART-seq and Drop-seq to describe the antibody repertoire (Fig 3a). Antibodies are comprised of heavy (μ, α, γ, δ, ε) and light chains (κ, λ), linked by disulfide bonds (Fig 3b). Each chain contains variable and constant domains. The variable region of the heavy chain is comprised of variable (V), diversity (D) and joining (J) segments, whereas the variable region of the light chain consists of a V and J segment (Fig 3b). We designed DART-seq to target the site where the constant domain is joined to the VDJ gene segment in both heavy and light chain loci[18] (Fig 3b). This design allows us to investigate the complementarity-determining region 3 (CDR3), which plays a key role in antigen binding. This region often goes undetected in regular scRNA-seq methods due to its distance from the 3' end of the transcript (Fig 3b).

To identify the population of B cells within PBMCs, we used dimensional reduction and clustering approaches implemented in Seurat[15]. We identified B cells based on expression of the B cell specific marker *MS4A1*[19]. We mapped transcript sequences obtained from B cells to the immunoglobulin (IG) sequence database, to find matches for the heavy and light chain transcripts in these cells, using MiXCR 2.1.5[20]. We visualized B

cells for which heavy and/or light chain transcripts were detected using t-distributed Stochastic Neighbor Embedding[21] (tSNE, Fig. 3c). The percentage of cells in which we were able to detect the IG transcript sequences was directly correlated to the total number of unique transcripts detected in the cells (Fig 3d). For cells with at least 1200 UMI or more, we identified either a heavy or a light chain transcript in 84% of B cells, and in 42% of B cells both the light and heavy chain transcripts were retrieved (Fig 3d). In contrast, Drop-seq detected either heavy or light chain transcripts in only 50%, and both heavy and light chain transcripts in 6% of B cells (Fig 3d).

B cells derive their repertoire diversity from the variable regions of their heavy (IGHV) and light chains (IGKV, IGLV)[22]. We measured the frequency of variable isoforms captured by DART-seq and compared it to Drop-seq. Our data clearly showed that DART-seq is more efficient than Drop-seq in capturing the diverse variable isoforms population found in B cells, and several isoforms were only detected by DART-seq (Fig 3e). Another significant feature of DART-seq is the capability to sequence the paired variable heavy and light chain transcripts in single cells. We measured clone specific paired usage for the heavy variable regions (IGHV) and light variable regions (IGKV, IGLV) in approximately 200 single B cells (Fig 3f). The observed trend for preferred pairings in single cells was similar to published data[23].

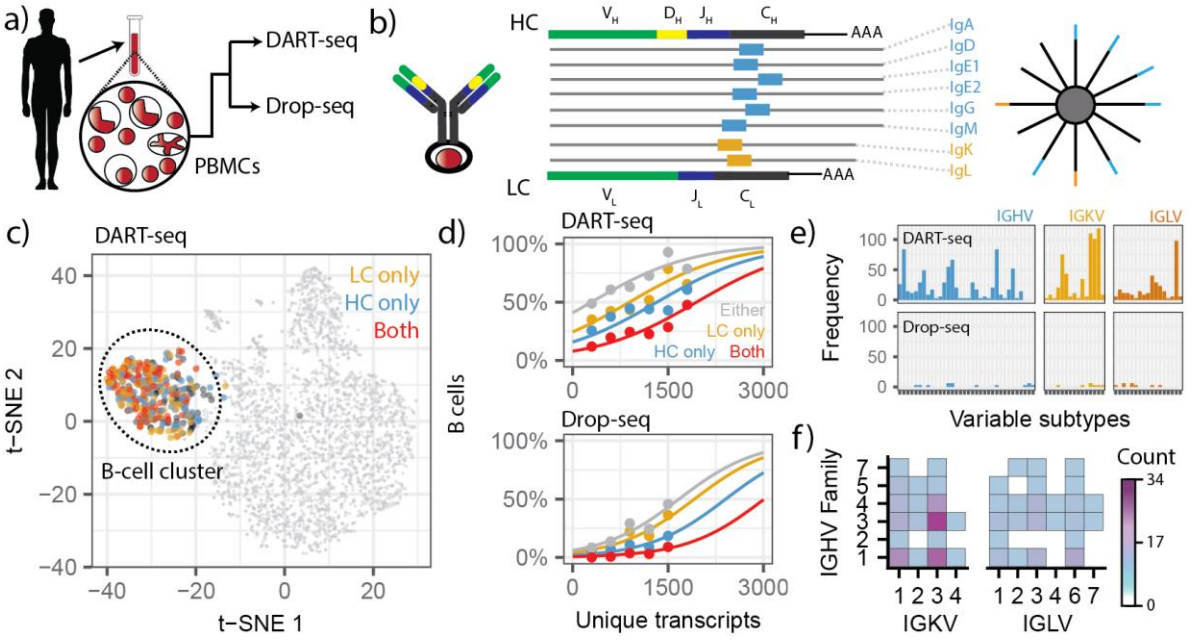

Fig. 3 DART-seq measures paired heavy and light chain B cell transcripts at single cell resolution. (a) PBMCs derived from human blood were subjected to Drop-seq and DART-seq. (b) DART-seq beads comprise probes that target the constant region of all human heavy and light immunoglobulins. (c) Representation of DART-seq data on a tSNE, B cells are highlighted based on heavy and light chain transcript capture. (d) Comparison of fraction of B cells for which heavy and light chain transcripts were detected with DART-seq versus Drop-seq, as a function of the total UMI count measured per cell. Points were fit with sigmoid function. Solid lines are sigmoid fits to the data (see Methods). (e) Comparison of the variable isoforms detected in B cells using DART-seq versus Drop-seq. The various isoforms detected are shown on the x-axis and normalized frequency of reads mapped to isoforms is shown on the y axis (f) Paired heavy (IGHV) and light (IGKV and IGLV) variable chain usage in B cells, pairing data from ~200 single cells was used to generate this collective plot.

DISCUSSION

We have presented an easy-to-implement, high-throughput scRNA-seq technology that overcomes the limitation of 3' end focused transcriptome measurements. DART-seq allows sequencing of all RNA types and all regions of the polyadenylated transcriptome in a single cell while maintaining the ability to perform single-cell transcriptome profiling. A straightforward and inexpensive ligation assay is used to synthesize DART-seq primer beads (Fig 1). The additional experiment time required for DART-seq compared to Drop-seq is minimal (2 hours) as is the cost per experimental design (~ $100 per experiment). DART-seq is compatible with simultaneous querying of many amplicons. Here, we present example designs with 7-10 amplicons. The design and ratio of probes can be tailored to individual applications allowing researchers the flexibility to use their existing scRNA-seq set-up for a wide variety of biological measurements.

We have highlighted two potential applications of DART-seq technology. First, we demonstrated that DART-seq provides a means to study the heterogeneity in viral genotypes and cellular phenotypes during viral infection. We were able to recapitulate a full segment of a dsRNA viral genome, while simultaneously profiling the transcriptome of the infected host cells (Fig 2). DART-seq opens new avenues for studies of host-virus interactions.

We further applied DART-seq to measure endogenously paired, heavy and light chain amplicons within the transcriptome of human B lymphocyte cells in a mixed human PBMC population, while having access to full transcriptome data of all other cell types (Fig 3). Determination of the paired antibody repertoire at depth can provide insights into several medically and immunologically relevant issues, including vaccine design and deployment[24–26].

METHODS

Primer bead synthesis. Single-stranded DNA (ssDNA) probe sequences were designed to complement regions of interest. The probes were annealed to the complementary splint sequences that also carry a 10-12 bp overhang of A-repeats (Supplementary table). All oligos were resuspended in Tris-EDTA (TE) buffer at a concentration of 500 µM. Double-stranded toehold adapters were created by heating equal volumes (20 µL) of the probe and splint oligos in the presence of 50 mM NaCl. The reaction mixture was heated to 95 °C and cooled to 14 °C at a slow rate (-0.1 °C/s). The annealed mixture of toehold probes was diluted with TE buffer to obtain a final concentration of 100 µM. Equal amounts of toehold probes were mixed and the final mixture diluted to obtain the desired probe concentration (2 pmoles for reovirus DART-seq design-1 and B-cell DART-seq, and 10 pmoles for reovirus DART-seq design-2). 16 µL of this pooled probe mixture was combined with 40 µL of PEG-4000 (50% w/v), 40 µL of T4 DNA ligase buffer, 72 µL of water, and 2 µL of T4 DNA Ligase (30 U/µL, Thermo Fisher). Roughly 12,000 beads were combined with the above ligation mix and incubated for 1 hr at 37 °C (15 second alternative mixing at 1800 rpm). After ligation, enzyme activity was inhibited (65 °C for 3 minutes) and beads were quenched in ice water. To obtain the desired quantity of DART-seq primer beads, 6-10 bead ligation reactions were performed in parallel. All reactions were pooled, and beads were washed once with 250 µL Tris-EDTA Sodium dodecyl sulfate (TE-SDS) buffer, and twice with Tris-EDTA-Tween 20 (TE-TW) buffer. DART-seq primer beads were stored in TE-TW at 4 °C.

Cell preparation. Murine L929 cells (L cells) in suspension culture were infected with recombinant Type 3 Dearing reovirus at MOI 10. After 15 hours of infection, the cells were centrifuged at 2300 rpm for 10 minutes and resuspended in PBS containing 0.01% BSA. Two additional washes were followed by centrifugation at 1200 rpm for 8 min, and then resuspended in the same buffer to a final concentration of 300,000 cells/mL. Human PBMCs were obtained from Zen-Bio. Cells were washed three times with PBS containing 0.01% BSA, each wash followed by centrifugation at 1500 rpm for 5 min, and then resuspended in the same buffer. The cell suspension was filtered through a 40 µm filter and resuspended to a final concentration of 120,000 cells/mL.

Single cell library preparation. Single cell library preparation was carried out as described previously[2]. Briefly, single cells were encapsulated with beads in a droplet using a microfluidics device (FlowJEM, Toronto, Ontario). After cell lysis, cDNA synthesis was carried out (Maxima Reverse Transcriptase, Thermo Fisher), followed by PCR (2X Kapa Hotstart Ready mix, VWR, 15 cycles). cDNA libraries were tagmented and PCR amplified (Nextera tagmentation kit, Illumina). Finally, libraries were pooled and sequenced (Illumina Nextseq 500, 20x130 bp).

qPCR measurement of viral gene segments. 0.1 ng DNA from sequencing libraries was used per qPCR reaction. Each reaction was comprised of 1 µL cDNA (0.1 ng/µL), 10

µL of iTaq™ Universal SYBR® Green Supermix (Bio-Rad), 0.5 µL of forward primer (10 µM), 0.5 µL of reverse primer (10 µM) and 13 µL of DNAse, RNAse free water. Reactions were performed in a sealed 96-well plate using the following program in the Bio-Rad C1000 Touch Thermal Cycler: (1) 95 °C for 10 minutes, (2) 95 °C for 30 seconds, (3) 65 °C for 1 minute, (4) plate read in SYBR channel, (5) repeat steps (2)-(4) 49 times, (6) 12 °C infinite hold. The resulting data file was viewed using Bio-Rad CFX manager and the Cq values were exported for further analysis. Each reaction was performed with two technical replicates.

Toehold ligation measurement via fluorescent hybridization. Roughly 6000 DART-seq beads were added to a mixture containing 18 uL of 5M NaCl, 2 µL of 1M Tris HCl pH 8.0, 1 µL of SDS, 78 µl of water, and 1 µL of 100 µM Cy5 fluorescently labeled oligo (see Supplementary Table). The beads were incubated for 45 minutes at 46 ˚C in an Eppendorf ThermoMixer C (15", at 1800 RPM). Following incubation, the beads were pooled and washed with 250 µL TE-SDS, followed by 250 µL TE-TW. The beads were suspended in water and imaged in the Zeiss Axio Observer Z1 in the Cy5 channel and bright field. A custom Python script was used to determine the fluorescence intensity of each bead.

Single cell host transcriptome profiling in viral infected cells. We used previously described bioinformatic tools to process raw sequencing reads[2], and the Seurat package for downstream analysis[15]. Cells with low overall expression or a high proportion of mitochondrial transcripts were removed. For clustering, we used principal component analysis (PCA), followed by k-means clustering to identify distinct cell states. For meta-clustering, host expression matrices from all four experiments were merged using Seurat. Cells with fewer than 2000 host transcripts were excluded. k-means clustering on principal components was used to identify cell clusters.

Viral genotype analysis. Sequencing reads that did not align to the host genome were collected and aligned to the T3D reovirus genome[27] (GenBank Accession EF494435-EF494445). Aligned reads were tagged with their cell barcode and sorted. The per-base coverage across viral gene segments was computed (Samtools[28] depth). Positions where the per-base coverage exceeded 50, and where a minor allele with frequency greater than 10% was observed, were labeled as SNV positions. The frequency of SNVs was calculated across all cells. For the combined host virus analysis, the host expression matrix and virus alignment information were merged. The per-base coverage of the viral genome was normalized by the number of host transcripts. Cells with fewer than 1500 host transcripts were excluded from the analysis.

IG heavy and light chain identification. Sequences derived from B cells (cells that are part of the cluster of B cells identified in Seurat, and that have nonzero expression of the *MS4A1* marker gene) were collected and aligned to a catalog of human germline V, D, J and C gene sequences using MiXCR version 2.1.5[20]. For each cell, the top scoring heavy and light chain variable regions were selected for subtyping and pairing analyses (Fig 3e and Fig 3f).

Sigmoidal fitting heavy/light chain capture. The mapping for the fractions of B cells containing heavy chains or light chains was fit with the following sigmoidal function:

$$y = \frac{1}{1+ e^{-b/(x-c)}} .$$

Where the parameter b was a free parameter for the fit of the light chain or heavy chain data, and then fixed for the light chain only, heavy chain only, and combined light chain and heavy chain data.

Statistical analysis. Statistical tests were performed in R version 3.3.2. Groups were compared using the nonparametric Mann-Whitney U test.

DATA AVAILABILITY

Raw sequencing data and corresponding gene expression matrices have been made available: NCBI Gene Expression Omnibus; Project ID GSE113675.

CODE AVAILABILITY

Custom scripts are available at: https://github.com/pburnham50/DART-seq.

ACKNOWLEDGMENTS

We would like to thank the De Vlaminck and Danko laboratory members for insightful discussions. We thank Peter Schweitzer and his colleagues at the Cornell University Biotechnology Resource Center (BRC) for performing the sequencing. This work was supported by US National Institute of Health (NIH) grant 1DP2AI138242 to IDV and National Science Foundation Graduate Research Fellowship Program (NSF-GRFP) grant DGE-1144153 to PB.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

AUTHOR CONTRIBUTIONS

PB, MS, CGD, JSLP and IDV designed the study. PB, MS, SHK, PML and MMH carried out the experiments. PB, MS, MFZW and IDV analyzed the data. PB, MS and IDV wrote the manuscript. All authors provided comments and edits.

REFERENCES

1. Gawad, C., Koh, W. & Quake, S. R. Single-cell genome sequencing: current state of the science. *Nat. Rev. Genet.* **17,** 175–88 (2016).
2. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161,** 1202–1214 (2015).
3. Gierahn, T. M. *et al.* Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat. Methods* **14,** 395–398 (2017).
4. Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161,** 1187–1201 (2015).
5. Xu, J. L. & Davis, M. M. Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity* **13,** 37–45 (2000).
6. Torre, E. *et al.* Rare Cell Detection by Single-Cell RNA Sequencing as Guided by Single-Molecule RNA FISH. *Cell Syst.* **6,** 171–179.e5 (2018).
7. Gansauge, M.-T. *et al.* Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res.* **45,** e79 (2017).
8. Dolan, P. T., Whitfield, Z. J. & Andino, R. Mapping the Evolutionary Potential of RNA Viruses. *Cell Host Microbe* **23,** 435–446 (2018).
9. Patton, J. T. & Spencer, E. Genome replication and packaging of segmented double-stranded RNA viruses. *Virology* **277,** 217–25 (2000).
10. Joklik, W. K. Structure and function of the reovirus genome. *Microbiol. Rev.* **45,** 483–501 (1981).
11. Niavarani, A. *et al.* APOBEC3A is implicated in a novel class of G-to-A mRNA editing in WT1 transcripts. *PLoS One* **10,** e0120089 (2015).
12. Harris, R. S. & Dudley, J. P. APOBECs and virus restriction. *Virology* **479–480,** 131–145 (2015).
13. Parker, J. S. L., Broering, T. J., Kim, J., Higgins, D. E. & Nibert, M. L. Reovirus core protein mu2 determines the filamentous morphology of viral inclusion bodies by interacting with and stabilizing microtubules. *J. Virol.* **76,** 4483–4496 (2002).
14. Ooms, L. S., Jerome, W. G., Dermody, T. S. & Chappell, J. D. Reovirus replication protein mu2 influences cell tropism by promoting particle assembly within viral inclusions. *J. Virol.* **86,** 10979–10987 (2012).
15. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33,** 495–502 (2015).
16. Georgiou, G. *et al.* The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat. Biotechnol.* **32,** 158–68 (2014).
17. DeKosky, B. J. *et al.* In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. *Nat. Med.* **21,** 86–91 (2015).
18. Vollmers, C., Sit, R. V, Weinstein, J. A., Dekker, C. L. & Quake, S. R. Genetic measurement of memory B-cell recall using antibody repertoire sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 13463–13468 (2013).
19. Tedder, T. F., Streuli, M., Schlossman, S. F. & Saito, H. Isolation and structure of a cDNA encoding the B1 (CD20) cell-surface antigen of human B lymphocytes. *Proc. Natl. Acad. Sci. U. S. A.* **85,** 208–12 (1988).
20. Bolotin, D. A. *et al.* MiXCR: software for comprehensive adaptive immunity

profiling. *Nat. Methods* **12,** 380–1 (2015).
21. van der Maaten, L. & Hinton, G. E. Visualizing data using t-SNE. *J. Mach. Learn.* **9,** 2579–2605 (2008).
22. Mroczek, E. S. *et al.* Differences in the composition of the human antibody repertoire by B cell subsets in the blood. *Front. Immunol.* **5,** 96 (2014).
23. DeKosky, B. J. *et al.* High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nat. Biotechnol.* **31,** 166–9 (2013).
24. Lanzavecchia, A., Frühwirth, A., Perez, L. & Corti, D. Antibody-guided vaccine design: identification of protective epitopes. *Curr. Opin. Immunol.* **41,** 62–67 (2016).
25. Karlsson Hedestam, G. B., Guenaga, J., Corcoran, M. & Wyatt, R. T. Evolution of B cell analysis and Env trimer redesign. *Immunol. Rev.* **275,** 183–202 (2017).
26. Jiang, N. Immune engineering: from systems immunology to engineering immunity. *Curr. Opin. Biomed. Eng.* **1,** 54–62 (2017).
27. Kobayashi, T. *et al.* A plasmid-based reverse genetics system for animal double-stranded RNA viruses. *Cell Host Microbe* **1,** 147–57 (2007).
28. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).