

Simultaneous multiplexed amplicon sequencing and transcriptome profiling in single cells

Authors: Mridusmita Saikia^{1,2,*}, Philip Burnham^{1,*}, Sara H. Keshavjee¹, Michael F. Z. Wang¹, Pablo Moral-Lopez², Meleana M. Hinchman², Charles G. Danko², John S. L. Parker², Iwijn De Vlaminck¹

*These authors contributed equally

To whom correspondence should be addressed: vlaminck@cornell.edu

Affiliations:

¹Meinig School of Biomedical Engineering, Cornell University, Ithaca, NY 14853, USA

²Baker 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.

1 INTRODUCTION

2
3 High-throughput single-cell RNA-seq (scRNA-seq) is being widely adopted for
4 phenotyping of cells in heterogeneous populations¹⁻⁴. The most common
5 implementations of this technology utilize droplet microfluidics to co-encapsulate single
6 cells with beads that are modified with barcoded oligos to enable capturing the ends of
7 RNA transcripts²⁻⁴. Although these approaches provide a means to perform inexpensive
8 single-cell gene expression measurements at scale, they are limited to assaying the ends
9 of mRNA transcripts. Therefore, they are ill-suited for the characterization of non-A-tailed
10 RNA, including the transcripts of many viruses, viral RNA genomes, and non-coding
11 RNAs. They are also uninformative of RNA segments that are located at a distance
12 greater than a few hundred bases from transcript ends that often comprise essential
13 functional information, for example the complementarity determining regions (CDRs) of
14 immunoglobulins (B cell antibody)⁵. Additionally, these techniques are often unable to
15 provide information on low copy number transcripts and splice variants⁶.

16
17 Here we report DART-seq, a method that combines enriched measurement of targeted
18 RNA sequences, with unbiased profiling of the poly(A)-tailed transcriptome across
19 thousands of single cells in the same biological sample. DART-seq achieves this by
20 implementing a simple and inexpensive alteration of the Drop-seq strategy². Barcoded
21 primer beads that capture the poly(A)-tailed mRNA molecules in Drop-seq are
22 enzymatically modified using a tunable ligation chemistry⁷. The resulting DART-seq
23 primer beads are capable of priming reverse transcription of poly(A)-tailed transcripts as
24 well as any other RNA species of interest.

25
26 DART-seq is easy to implement and enables a range of new biological measurements.
27 Here, we explored two applications. We first applied DART-seq to profile viral-host
28 interactions and viral genome dynamics in single cells. We implemented two distinct
29 DART-seq designs to investigate murine L929 cells (L cells) infected by the reovirus strain
30 Type 3 Dearing (T3D). We demonstrate the ability of DART-seq to profile all 10 non-A-
31 tailed viral gene transcripts of T3D reovirus individually, as well as to recover a complete
32 genome segment, while simultaneously providing access to the transcriptome of the
33 infected L cells. In the second application, we applied DART-seq to determine natively
34 paired antibody sequences of human B cells. DART-seq was able to determine B cell
35 clonotypes, as well as variable heavy and light (VH:VL) pairings, even in mixed human
36 peripheral blood mononuclear cells (PBMCs), highlighting the versatility of the approach.

37
38
39
40

1 RESULTS

2

3 DART-seq primer bead synthesis

4

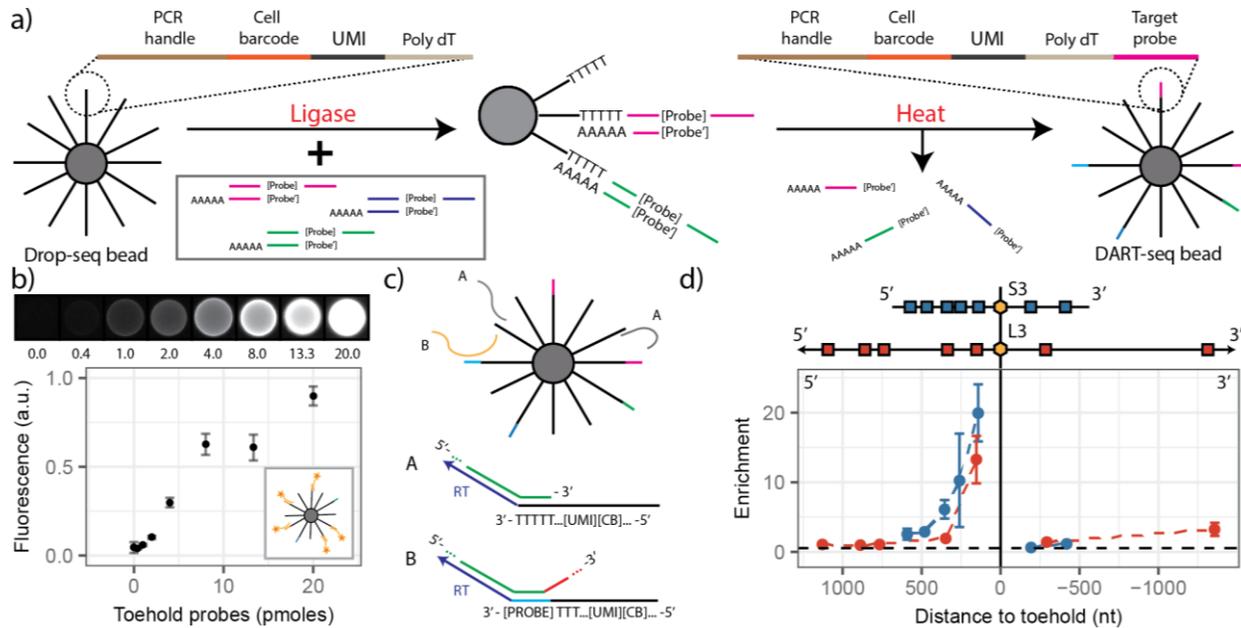
5 Droplet microfluidics based scRNA-seq approaches rely on co-encapsulation of single
6 cells with barcoded primer beads that capture and prime reverse transcription of mRNA
7 molecules expressed by the cell²⁻⁴. In Drop-seq, the primers on all beads comprise a
8 common sequence used for PCR amplification, a bead-specific cell barcode, a unique
9 molecular identifier (UMI), and a poly-dT sequence for capturing polyadenylated mRNAs
10 and priming reverse transcription. To enable simultaneous measurement of the
11 transcriptome and multiplexed RNA amplicons in DART-seq, we devised a scheme to
12 enzymatically attach custom primers to a subset of poly-dTs on the Drop-seq bead (Fig
13 1a). This is achieved by annealing a double stranded toehold probe with a 3' ssDNA
14 overhang that is complementary to the poly-dT sequence of the Drop-seq primers. The
15 toehold is then ligated to the bead using T4 DNA ligase. Toeholds with a variety of
16 different sequences can be attached to the same primer beads in a single reaction in this
17 manner. The complementary toehold strand is removed after ligation. We examined the
18 efficiency and tunability of the probe ligation reaction using fluorescence hybridization
19 assays. Here, fluorescently labeled DNA hybridization probes were designed for
20 complementarity to ligated primer sequences. We found that the measured fluorescence
21 signal after ligation is proportional to the number of toeholds included in the ligation
22 reaction (Fig 1b).

23

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

35

36



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

9
10
11 We measured the abundance of amplicons in the resulting sequencing libraries using
12 quantitative PCR (Fig 1d, qPCR). Here, we compared sequencing libraries of T3D
13 reovirus-infected L cells generated by Drop-seq and libraries for the same cells generated
14 by DART-seq with amplicons targeting all ten genome segments of the virus. Reovirus
15 mRNAs lack poly(A)-tails, and thus should not efficiently captured by Drop-seq. We
16 designed seven PCR assays with 84-120 bp amplicons distributed across the L3 and S3
17 viral genome segments. We observed significant RNA enrichment upstream (5' end) of
18 the toehold ligation site for both the L3 and S3 segment (Fig 1d). No enrichment was
19 observed in a host transcript (*Gapdh*). As expected, there was no enrichment downstream
20 of the toehold ligation site (3' end). Consistent with sequencing library preparation via
21 tagmentation, we found that the degree of enrichment achieved by DART-seq at a given
22 position decreased exponentially with distance from the target.

23
24
25
26

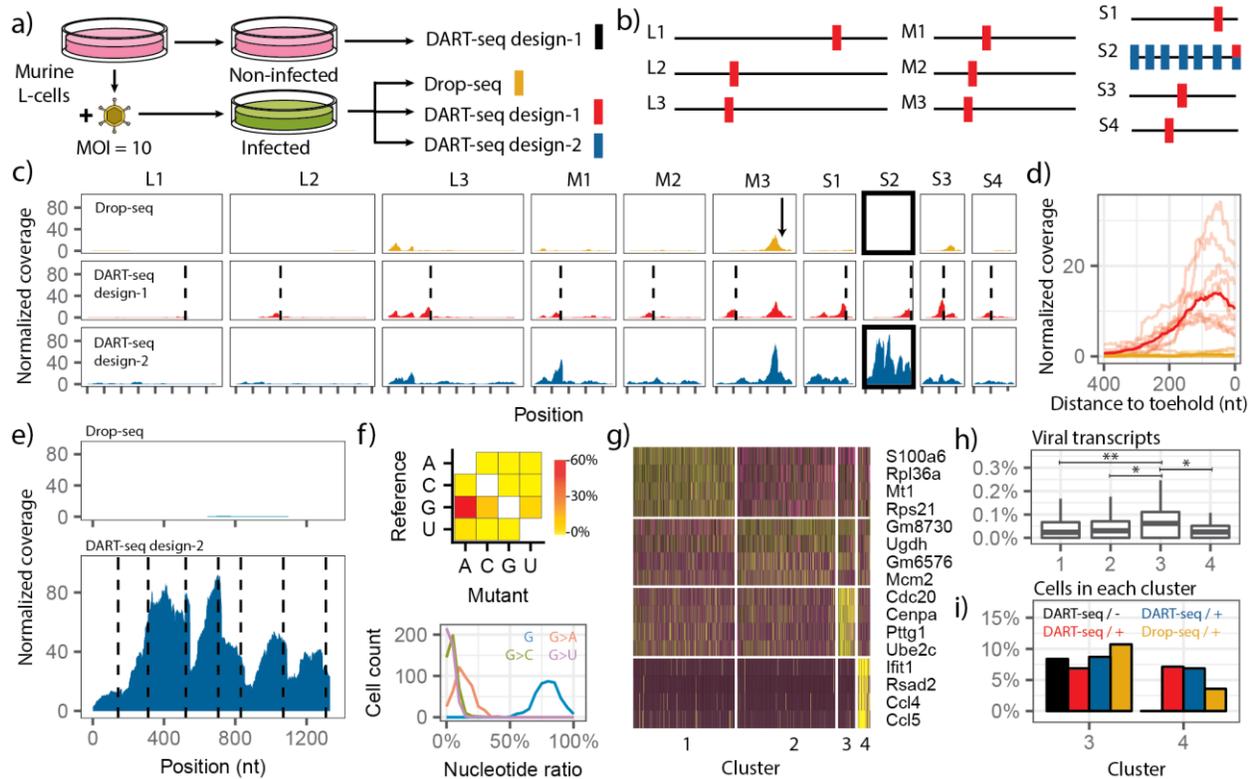
1 **DART-seq enables investigation of the heterogeneity of cellular phenotypes and**
2 **viral genotypes during viral infection.**

3
4 There is a great need for novel single cell genomics tools that can dissect the
5 heterogeneity in viral genotypes and cellular phenotypes during viral infection⁸. We used
6 DART-seq to examine infection of murine L cells with T3D reovirus. The reovirus
7 polymerase transcribes non-A-tailed mRNAs from each of its 10 dsRNA genome
8 segments^{9,10}. We infected L cells at a multiplicity of infection of 10 (MOI 10), and allowed
9 the virus to replicate for 15 hours after inoculation, creating a condition for which nearly
10 all cells are infected (Fig 2a). We performed Drop-seq and DART-seq experiments on
11 infected L cells and non-infected L cells as control. We implemented two distinct DART-
12 seq designs. The first DART-seq design targeted each viral genome segment with a
13 single amplicon. The second DART-seq design was comprised of seven amplicons
14 targeting loci distributed evenly across the S2 genome segment (Fig 2b).

15
16 To determine the efficiency by which DART-seq retrieves viral transcripts near the target
17 sequence, we analyzed the per-base coverage of positions upstream of the DART-seq
18 target sites. For DART-seq design-1, we observed a mean enrichment of 34.7x in the
19 gene regions 200 nt upstream of the ten toeholds. In both DART-seq design-1 and 2, all
20 targeted sites were enriched compared to standard Drop-seq beads (Fig 2c,d). Viral
21 transcripts were detected in Drop-seq libraries upstream of A-rich sequences in the viral
22 genome, consistent with spurious priming of reverse transcription by poly-dT sequences
23 on the oligo, as expected for Drop-seq. For example, a 200 nt gene segment upstream
24 of an A₅ sequence on segment M3 (position 1952) was significantly enriched in the Drop-
25 seq dataset (Fig 2c; marked by arrow).

26
27 To test the utility of DART-seq to measure the heterogeneity of viral genotypes in single
28 infected cells, we used DART-seq design-2 (Fig 2b), which was tailored to retrieve the
29 complete S2 viral gene segment. The S2 segment encodes inner capsid protein σ_2 .
30 Across cells with at least 1500 UMIs, DART-seq design-2 increased the mean coverage
31 across the S2 segment 430-fold compared to Drop-seq, thereby enabling the
32 investigation of the rate and pattern of mutations (Fig 2e). 176 single-nucleotide variants
33 (SNVs) were identified across the S2 segment (minor allele frequency greater than 10%,
34 and per-base-coverage greater than 50x). Mutations from guanine-to-adenine (G-to-A)
35 were most common (58%; Fig 2f, top). We did not observe such a mutation pattern in a
36 highly-expressed host transcript (*Actb*). We examined the mutation load of viral
37 transcripts at the single cell level, and observed a wide distribution in mutation load, with
38 a mean G-to-A conversion rate of 13%, and up to 41% (Fig 2f, bottom). The reason for
39 this level of hypermutation is unclear. G-to-A transamidation is an uncommon post-
40 transcriptional modification that is not been previously seen as a host response to viral

1 infection^{11,12}. The high rate of G-to-A transition in the viral transcript could also be
 2 secondary to a defect in the fidelity of viral transcription. The T3D strain used in this study
 3 has strain-specific allelic variation in the viral polymerase co-factor, $\mu 2$, that has been
 4 shown to affect the capacity of $\mu 2$ to associate with microtubules and the encapsidation
 5 of viral mRNAs within capsids^{13,14}.



6 **Figure 2 - DART-seq reveals heterogeneity in viral genotypes and host response to infection.** (a)
 7 Experimental design. Single cell analysis using Drop-seq and two distinct DART-seq designs of murine L
 8 cells infected with a reovirus, and a non-infected control. (b) Schematic of two DART-seq designs. Design-
 9 1 targets all 10 reovirus gene segments (3 x L (Large), 3 x M (Medium), and 4 x S (Small)
 10 segments). Design-2 (blue bars) targets seven loci on the S2 gene segment. (c) Comparison of the relative
 11 sequence coverage of the 10 reovirus gene segments (columns) for three different library preparations
 12 (rows). The arrow indicates an A₅ pentanucleotide sequence part of segment M3. Dotted lines indicate
 13 DART-seq target positions. (d) Per-base coverage upstream (5' end) of 10 toeholds of DART-seq design-
 14 1 (light red, average shown in dark red), and mean coverage achieved with Drop-seq (yellow). (e) Per-base
 15 coverage of the S2 gene segment achieved with DART-seq design-2 (bottom, dashed lines indicate toehold
 16 positions) and Drop-seq (top). (f) Frequency and pattern of base mutations. Across all cells, the average
 17 nucleotide profile for positions on the S2 segment with SNPs such that the major allele is < 90% are shown
 18 (top); the distribution of nucleotide ratios for positions with reference nucleotide G is depicted for single cells
 19 (bottom). (g) Clustering analysis of reovirus infected L cells (DART-seq design-1). Hierarchical clustering
 20 of clusters displayed as a heatmap (yellow/purple is higher/lower expression). (h) Relative abundance of
 21 viral transcripts in L-cell clusters (* and ** indicates significant *p*-value of 10⁻³ and 10⁻⁴, respectively). (i)
 22 Fraction of cells in meta-clusters for four experiments depicted in panel a with assay type and infection
 23 status (+ or -) indicated.

1 To identify distinct host cell populations based on patterns of gene expression, we
2 performed dimensional reduction and unsupervised clustering using approaches
3 implemented in Seurat¹⁵. We identified four distinct cell clusters for the monoculture
4 infection model (DART-seq design-1, Fig 2g). Two major clusters comprised of cells with
5 elevated expression of genes related to viral RNA transcription and replication (*Rpl36a*,
6 cluster 1) and metabolic pathways (*Ugdh*, cluster 2). Two additional clusters were defined
7 by the upregulation of genes related to mitotic function (*Cdc20*, *Cenpa*; cluster 3) and
8 innate immunity (*Ifit1*, *Rsad2*; cluster 4), respectively (Fig 2g). The abundance of viral
9 gene transcripts relative to host transcripts was significantly elevated for cells in cluster 3
10 ($n = 69$ of 927 total cells) compared to cells in all other clusters (Fig 2h; two-tailed Mann
11 Whitney U test, $p = 1.0 \times 10^{-4}$). We merged datasets for the Drop-seq and three DART-seq
12 assays and quantified the cell type composition for each experiment. We did not observe
13 cells related to cluster 4 (immune response), for the non-infected control, as expected
14 (Fig 2i). These results support the utility of DART-seq to study the single cell
15 heterogeneity in viral genotypes and cellular phenotypes during viral infection.

16

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

18

19 As a second application of DART-seq, we explored the biological corollary of viral
20 infection, the cellular immune response. The adaptive immune response is reliant upon
21 the generation of a highly diverse repertoire of B lymphocyte antigen receptors (BCRs),
22 the membrane-bound form of antibodies expressed on the surface of B cells, as well as
23 antibodies secreted by plasmablasts^{16,17}. We applied DART-seq to investigate the B cell
24 antibody repertoire in human PBMCs. We compared the performance of DART-seq and
25 Drop-seq to describe the antibody repertoire (Fig 3a). Antibodies are comprised of heavy
26 (μ , α , γ , δ , ϵ) and light chains (κ , λ), linked by disulfide bonds (Fig 3b). Each chain contains
27 variable and constant domains. The variable region of the heavy chain is comprised of
28 variable (V), diversity (D) and joining (J) segments, whereas the variable region of the
29 light chain consists of a V and J segment (Fig 3b). We designed DART-seq to target the
30 site where the constant domain is joined to the VDJ gene segment in both heavy and light
31 chain loci¹⁸ (Fig 3b). This design allows us to investigate the complementarity-determining
32 region 3 (CDR3), which plays a key role in antigen binding. This region often goes
33 undetected in regular scRNA-seq methods due to its distance from the 3' end of the
34 transcript (Fig 3b).

35

36 To identify the population of B cells within PBMCs, we used dimensional reduction and
37 clustering approaches implemented in Seurat¹⁵. We identified B cells based on
38 expression of the B cell specific marker *MS4A1*¹⁹. We mapped transcript sequences
39 obtained from B cells to the immunoglobulin (IG) sequence database, to find matches for
40 the heavy and light chain transcripts in these cells, using MiXCR 2.1.5²⁰. We visualized B

1 cells for which heavy and/or light chain transcripts were detected using t-distributed
2 Stochastic Neighbor Embedding²¹ (tSNE, Fig. 3c). The percentage of cells in which we
3 were able to detect the IG transcript sequences was directly correlated to the total number
4 of unique transcripts detected in the cells (Fig 3d). For cells with at least 1200 UMI or
5 more, we identified either a heavy or a light chain transcript in 84% of B cells, and in 42%
6 of B cells both the light and heavy chain transcripts were retrieved (Fig 3d). In contrast,
7 Drop-seq detected either heavy or light chain transcripts in only 50%, and both heavy and
8 light chain transcripts in 6% of B cells (Fig 3d).

9
10 B cells derive their repertoire diversity from the variable regions of their heavy (IGHV) and
11 light chains (IGKV, IGLV)²². We measured the frequency of variable isoforms captured
12 by DART-seq and compared it to Drop-seq. Our data clearly showed that DART-seq is
13 more efficient than Drop-seq in capturing the diverse variable isoforms population found
14 in B cells, and several isoforms were only detected by DART-seq (Fig 3e). Another
15 significant feature of DART-seq is the capability to sequence the paired variable heavy
16 and light chain transcripts in single cells. We measured clone specific paired usage for
17 the heavy variable regions (IGHV) and light variable regions (IGKV, IGLV) in
18 approximately 200 single B cells (Fig 3f). The observed trend for preferred pairings in
19 single cells was similar to published data²³.

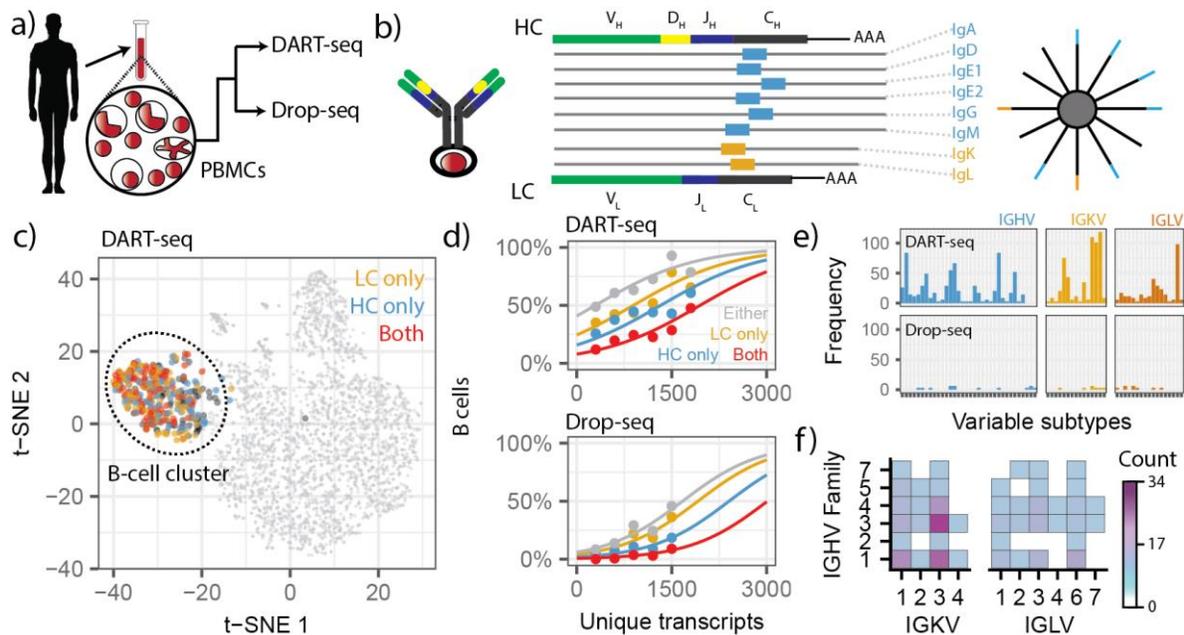


Fig. 3 DART-seq measures paired heavy and light chain B cell transcripts at single cell resolution.

(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.

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

7

8 We further applied DART-seq to measure endogenously paired, heavy and light chain
9 amplicons within the transcriptome of human B lymphocyte cells in a mixed human PBMC
10 population, while having access to full transcriptome data of all other cell types (Fig 3).
11 Determination of the paired antibody repertoire at depth can provide insights into several
12 medically and immunologically relevant issues, including vaccine design and
13 deployment²⁴⁻²⁶.

14

15

1 **METHODS**

2

3 **Primer bead synthesis.** Single-stranded DNA (ssDNA) probe sequences were designed
4 to complement regions of interest. The probes were annealed to the complementary splint
5 sequences that also carry a 10-12 bp overhang of A-repeats (Supplementary table). All
6 oligos were resuspended in Tris-EDTA (TE) buffer at a concentration of 500 μ M. Double-
7 stranded toehold adapters were created by heating equal volumes (20 μ L) of the probe
8 and splint oligos in the presence of 50 mM NaCl. The reaction mixture was heated to 95
9 $^{\circ}$ C and cooled to 14 $^{\circ}$ C at a slow rate (-0.1 $^{\circ}$ C/s). The annealed mixture of toehold probes
10 was diluted with TE buffer to obtain a final concentration of 100 μ M. Equal amounts of
11 toehold probes were mixed and the final mixture diluted to obtain the desired probe
12 concentration (2 pmoles for reovirus DART-seq design-1 and B-cell DART-seq, and 10
13 pmoles for reovirus DART-seq design-2). 16 μ L of this pooled probe mixture was
14 combined with 40 μ L of PEG-4000 (50% w/v), 40 μ L of T4 DNA ligase buffer, 72 μ L of
15 water, and 2 μ L of T4 DNA Ligase (30 U/ μ L, Thermo Fisher). Roughly 12,000 beads were
16 combined with the above ligation mix and incubated for 1 hr at 37 $^{\circ}$ C (15 second
17 alternative mixing at 1800 rpm). After ligation, enzyme activity was inhibited (65 $^{\circ}$ C for 3
18 minutes) and beads were quenched in ice water. To obtain the desired quantity of DART-
19 seq primer beads, 6-10 bead ligation reactions were performed in parallel. All reactions
20 were pooled, and beads were washed once with 250 μ L Tris-EDTA Sodium dodecyl
21 sulfate (TE-SDS) buffer, and twice with Tris-EDTA-Tween 20 (TE-TW) buffer. DART-seq
22 primer beads were stored in TE-TW at 4 $^{\circ}$ C.

23 **Cell preparation.** Murine L929 cells (L cells) in suspension culture were infected with
24 recombinant Type 3 Dearing reovirus at MOI 10. After 15 hours of infection, the cells were
25 centrifuged at 2300 rpm for 10 minutes and resuspended in PBS containing 0.01% BSA.
26 Two additional washes were followed by centrifugation at 1200 rpm for 8 min, and then
27 resuspended in the same buffer to a final concentration of 300,000 cells/mL. Human
28 PBMCs were obtained from Zen-Bio. Cells were washed three times with PBS containing
29 0.01% BSA, each wash followed by centrifugation at 1500 rpm for 5 min, and then
30 resuspended in the same buffer. The cell suspension was filtered through a 40 μ m filter
31 and resuspended to a final concentration of 120,000 cells/mL.

32 **Single cell library preparation.** Single cell library preparation was carried out as
33 described previously². Briefly, single cells were encapsulated with beads in a droplet
34 using a microfluidics device (FlowJEM, Toronto, Ontario). After cell lysis, cDNA synthesis
35 was carried out (Maxima Reverse Transcriptase, Thermo Fisher), followed by PCR (2X
36 Kapa Hotstart Ready mix, VWR, 15 cycles). cDNA libraries were tagmented and PCR
37 amplified (Nextera tagmentation kit, Illumina). Finally, libraries were pooled and
38 sequenced (Illumina Nextseq 500, 20x130 bp).

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

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

9 **Toehold ligation measurement via fluorescent hybridization.** Roughly 6000 DART-
10 seq beads were added to a mixture containing 18 μL of 5M NaCl, 2 μL of 1M Tris HCl pH
11 8.0, 1 μL of SDS, 78 μL of water, and 1 μL of 100 μM Cy5 fluorescently labeled oligo (see
12 Supplementary Table). The beads were incubated for 45 minutes at 46 °C in an Eppendorf
13 ThermoMixer C (15", at 1800 RPM). Following incubation, the beads were pooled and
14 washed with 250 μL TE-SDS, followed by 250 μL TE-TW. The beads were suspended in
15 water and imaged in the Zeiss Axio Observer Z1 in the Cy5 channel and bright field. A
16 custom Python script was used to determine the fluorescence intensity of each bead.

17 **Single cell host transcriptome profiling in viral infected cells.** We used previously
18 described bioinformatic tools to process raw sequencing reads², and the Seurat package
19 for downstream analysis¹⁵. Cells with low overall expression or a high proportion of
20 mitochondrial transcripts were removed. For clustering, we used principal component
21 analysis (PCA), followed by k-means clustering to identify distinct cell states. For meta-
22 clustering, host expression matrices from all four experiments were merged using Seurat.
23 Cells with fewer than 2000 host transcripts were excluded. k-means clustering on
24 principal components was used to identify cell clusters.

25 **Viral genotype analysis.** Sequencing reads that did not align to the host genome were
26 collected and aligned to the T3D reovirus genome²⁷ (GenBank Accession EF494435-
27 EF494445). Aligned reads were tagged with their cell barcode and sorted. The per-base
28 coverage across viral gene segments was computed (Samtools²⁸ depth). Positions where
29 the per-base coverage exceeded 50, and where a minor allele with frequency greater
30 than 10% was observed, were labeled as SNV positions. The frequency of SNVs was
31 calculated across all cells. For the combined host virus analysis, the host expression
32 matrix and virus alignment information were merged. The per-base coverage of the viral
33 genome was normalized by the number of host transcripts. Cells with fewer than 1500
34 host transcripts were excluded from the analysis.

35 **IG heavy and light chain identification.** Sequences derived from B cells (cells that are
36 part of the cluster of B cells identified in Seurat, and that have nonzero expression of the
37 *MS4A1* marker gene) were collected and aligned to a catalog of human germline V, D, J
38 and C gene sequences using MiXCR version 2.1.5²⁰. For each cell, the top scoring heavy
39 and light chain variable regions were selected for subtyping and pairing analyses (Fig 3e
40 and Fig 3f).

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

3
$$y = \frac{1}{1 + e^{-b/(x-c)}} .$$

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

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

9

10 **DATA AVAILABILITY**

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

13

14 **CODE AVAILABILITY**

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

16

17 **ACKNOWLEDGMENTS**

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

24

25 **COMPETING FINANCIAL INTERESTS**

26 The authors declare no competing financial interests.

27

28 **AUTHOR CONTRIBUTIONS**

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

32

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).