

A risk-reward examination of sample multiplexing reagents for single cell RNA-Seq

Daniel V. Brown^{a,b,*}, Casey J.A. Anttila^a, Ling Ling^a, Patrick Grave^a, Tracey M. Baldwin^a, Ryan Munnings^{a,b}, Anthony J. Farchione^{a,b}, Vanessa L. Bryant^{a,b,c}, Amelia Dunstone^a, Christine Biben^{a,b}, Samir Taoudi^{a,b}, Tom S. Weber^{a,b}, Shalin H. Naik^{a,b}, Anthony Hadla^{a,b}, Holly E. Barker^{a,b}, Cassandra J. Vandenberg^{a,b}, Genevieve Dall^{a,b}, Clare L. Scott^{a,b}, Zachery Moore^{a,b}, James R. Whittle^{a,b,d}, Saskia Freytag^{a,b}, Sarah A. Best^{a,b}, Anthony T. Papenfuss^{a,b,d}, Sam W.Z. Olechnowicz^{a,b}, Sarah E. MacRaid^a, Stephen Wilcox^a, Peter F. Hickey^{a,b}, Daniela Amann-Zalcenstein^{a,b}, Rory Bowden^{a,b}

^a The Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade VIC, Melbourne 3052, VIC, Australia

^b Department of Medical Biology, University of Melbourne, Parkville, Melbourne 3010, VIC, Australia

^c The Royal Melbourne Hospital, 300 Grattan St, Parkville, Melbourne 3010, VIC, Australia

^d Peter MacCallum Cancer Centre, 305 Grattan St, Parkville, Melbourne 3010, VIC, Australia

ARTICLE INFO

Keywords:

Single-cell
RNA-seq
Sample multiplexing
Fixed
CRISPRclean

ABSTRACT

Single-cell RNA sequencing (scRNA-Seq) has emerged as a powerful tool for understanding cellular heterogeneity and function. However the choice of sample multiplexing reagents can impact data quality and experimental outcomes. In this study, we compared various multiplexing reagents, including MULTI-Seq, Hashtag antibody, and CellPlex, across diverse sample types such as human peripheral blood mononuclear cells (PBMCs), mouse embryonic brain and patient-derived xenografts (PDXs). We found that all multiplexing reagents worked well in cell types robust to ex vivo manipulation but suffered from signal-to-noise issues in more delicate sample types. We compared multiple demultiplexing algorithms which differed in performance depending on data quality. We find that minor improvements to laboratory workflows such as titration and rapid processing are critical to optimal performance. We also compared the performance of fixed scRNA-Seq kits and highlight the advantages of the Parse Biosciences kit for fragile samples. Highly multiplexed scRNA-Seq experiments require more sequencing resources, therefore we evaluated CRISPR-based destruction of non-informative genes to enhance sequencing value. Our comprehensive analysis provides insights into the selection of appropriate sample multiplexing reagents and protocols for scRNA-Seq experiments, facilitating more accurate and cost-effective studies.

1. Introduction

Single-cell RNA sequencing (scRNA-Seq) has been powered by advancements in molecular biology, microfluidics, and high-throughput sequencing [1]. Applications of scRNA-Seq span cell atlases, pooled screens and clinical studies [2]. Single-cell approaches have expanded to include additional modalities, such as surface protein measurement, open chromatin analysis, and CRISPR perturbation [3]. As scRNA-Seq becomes more accessible, increased sample sizes and biological replicates enhance scientific rigor but necessitate larger, more complex experiments. Although the cost per cell is decreasing, overall experimental

costs remain high.

Sample multiplexing has emerged as an elegant solution to reduce costs. The concept was first demonstrated by mixing genetically distinct samples and subsequently deconvoluting them using genotypes called in the sequencing data [4]. However, in many studies the absence of natural genetic variation renders this approach unfeasible. As an alternative, various methods have been developed to deliver exogenous sample-identifying DNA barcodes. The first implementation utilized oligo-tagged antibodies targeting ubiquitous cell surface proteins [5]. Subsequent technologies have delivered DNA barcodes via lipids, concanavalin A, click chemistry, transfection, or transduction [6–10].

* Corresponding author at: The Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade VIC, Melbourne 3052, VIC, Australia.

E-mail addresses: brown.d@wehi.edu.au (D.V. Brown), bowden.r@wehi.edu.au (R. Bowden).

<https://doi.org/10.1016/j.ygeno.2024.110793>

Received 25 June 2023; Received in revised form 29 November 2023; Accepted 9 January 2024

Available online 14 January 2024

0888-7543/© 2024 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Regardless of the delivery mechanism, sample multiplexing necessitates additional upfront handling of individual samples, with the potential to perturb cell states [11] or reduce viability. Antibody- and lipid-based barcodes have become the most popular systems, for their broad applicability and ease of use. Recently, Mylka et al., directly compared these multiplexing methods, recommending different solutions for different sample types [12].

The advent of commercial fixed scRNA-Seq kits, such as the Parse Biosciences Evercode kits and 10× Genomics Flex, has effectively decoupled sample collection from processing. Both kits incorporate sample multiplexing into their molecular biology workflows, eliminating the need for specific labeling steps. The Parse Biosciences Evercode kits utilize multiwell plates, enabling sample multiplexing by dispensing each sample into a separate well. In contrast, the 10× Genomics Flex kit employs a ligation probe-based assay, where sample barcodes are embedded in the probe sequences.

We have conducted an extensive comparison of antibody, lipid-based, and fixed sample multiplexing reagents across diverse and broadly representative cell types, including human peripheral blood mononuclear cells (PBMCs), mouse embryonic brain, and ovarian carcinoma patient-derived xenografts (PDX). We compare exogenous sample labels introduced by multiplexing oligos with endogenous SNP genotypes as ground truth [5]. We also assess cell type recovery bias.

We evaluate CRISPR-based destruction of non-informative genes, an important potential adjunct in controlling the cost of larger single-cell experiments. Through upfront optimization and downstream comparative analyses, we propose guidelines for experimental design and the utilization of different protocols in various contexts.

2. Results

2.1. Comparison of multiplexing reagents in human PBMCs

To evaluate the performance of sample multiplexing reagents (Table S1) in a system where we could obtain ground truth from SNP genotypes, we analyzed PBMCs from four human donors (Fig. 1A). Each PBMC donor sample was divided into technical duplicates for the sample multiplexing labeling reaction (Fig. 1A) and captured with 10× Genomics v3.1 chemistry at a cell input of 35,000 cells, for a theoretical output of 20,000 cell-containing droplets at a 16.11% doublet rate (Satija lab calculator).

After library preparation and sequencing, we examined the count distributions for each tag and protocol (Fig. 1B). The signal-to-noise was most consistent for MULTI-Seq with hashtag antibody also performing well aside from a single labeling reaction, HTO1, which had a high background and lower signal indicating an issue with the antibody reagent rather than an error in sample handling (Fig. 1B). The HTO1 sample's poor signal to noise negatively impacted other tags. Dimension reduction of the hashtag antibody capture revealed satellite clusters with a subset of cells having a correct dominant tag but contaminated by HTO1 (Fig. S2B).

CellPlex had the lowest signal-to-noise and highest proportion of unassigned cells (p -value 0.031 and 0.004 versus hashtag and MULTI-Seq respectively) (Fig. 1C and D). Of note the doublet rate is higher for hashtag antibody and MULTI-Seq than the theoretical 16.11% expected from loading each 10× Genomics capture with 35,000 cells. We later titrated the CellPlex reagent ten-fold below the manufacturer's recommendations without a loss in signal (Fig. S3).

Examination of the relationship between oligo tag library and gene expression library size revealed multiplets had a higher library size than singlets (Fig. 1E). In contrast, unassigned cells had a similar library size to singlets, suggestive of a failure in oligo tag labeling rather than unassigned cells being enriched from empty droplets or damaged cells.

2.2. Accuracy of sample multiplexing oligos compared to SNP genotypes

We next compared the accuracy of the sample multiplexing tag assignments to ground truth SNP assignments (Fig. 2A). Given our experimental design with four unrelated donors in duplicate, the doublet rate was lower from SNP calls (12.08%) than from multiplexing tags (16.56%). The identifiable doublet rate excludes homotypic doublets.

The proportions of individual donors recovered when comparing SNPs to sample multiplexing oligos was not significantly different aside from the unassigned category (Fig. 2A). This was confirmed upon closer inspection of the multiplexing tag calls to SNP calls in an alluvial plot and heatmap (Fig. 2B and C). The major discordant droplets were non-identifiable doublets when calling multiplets based on 4 SNPs versus 8 multiplexing oligos. Having used the Cell Ranger multi algorithm for sample demultiplexing, we next compared other algorithms available in the cellhashR package [13]. The three sample multiplexing datasets generated had different characteristics with CellPlex having lower signal to noise, hashtag antibody having one poorly performing tag and MULTI-Seq being a high quality dataset (Fig. 2D).

Cell Ranger multi performed as well as other algorithms in hashtag antibody and MULTI-Seq datasets, despite a warning during runtime that these multiplexing tag oligos are not supported. In the CellPlex dataset, the Bimodal Flexible Fitting (BFF cluster) algorithm performed best (Fig. 2D). In contrast, in the high-quality MULTI-Seq dataset, BFF cluster called four-fold more false positives, assigning more multiplets to singlets. This protocol dependent variation in algorithm performance meant there was no significant difference in demultiplexing algorithms (Fig. S4E). We computed the overall classification accuracy, (OCA) the same metric used in Mylka et al., [12], (Table 1). The OCA is defined as the sum of matching assignments between demultiplexing tags and SNP assignments divided by the count of all cell-containing droplets. Consistent with the signal-to-noise metrics, MULTI-Seq performed best, followed by hashtag antibody with CellPlex the poorest performing protocol.

To evaluate the impact of sequencing depth on demultiplexing accuracy, we downsampled the multiplexing tag library sequencing data and reprocessed the data using Cell Ranger Multi. The recommended number of reads per cell by 10× Genomics is 5000 for hashtag antibodies and CellPlex. Beyond 1000 reads per cell, an increase in the number of reads per cell exerted a negligible effect on demultiplexing performance (Fig. 2E). An alternate guideline provided on the 10× Genomics website indicates 1000 usable oligo tag reads per cell, which is consistent with our findings. Particularly for high-quality datasets, a total of 5000 reads per cell for sample multiplexing oligos appears excessive.

2.3. Comparison of sample multiplexing reagents in mouse embryonic brain

PBMCs are a robust sample type that do not require dissociation and can be maintained as a single-cell suspension for a prolonged period on ice with only minor effects on viability or phenotype [14]. We next aimed to benchmark sample multiplexing reagents in mouse embryonic brain E18.5, a more challenging tissue (Fig. 3A). In contrast to PBMCs, which were processed using a low-throughput labeling protocol in 1.5 mL tubes, we utilized a high-throughput labeling protocol in a 96 well plate. We also adjusted the FACS sort step as per 10× Genomics guidelines to conduct the sort after labeling with multiplexing oligos, rather than sorting after thawing and before labeling as in the PBMC experiment.

However, in a pilot experiment, we were unable to detect any signal using mouse hashtag antibodies with this sample type (Fig. S5A). Thus, as a substitute for the hashtag antibody we evaluated a cholesterol modified custom MULTI-Seq oligo, composed of the CellPlex oligo sequence grafted onto the MULTI-Seq lipid (Fig. S5B).

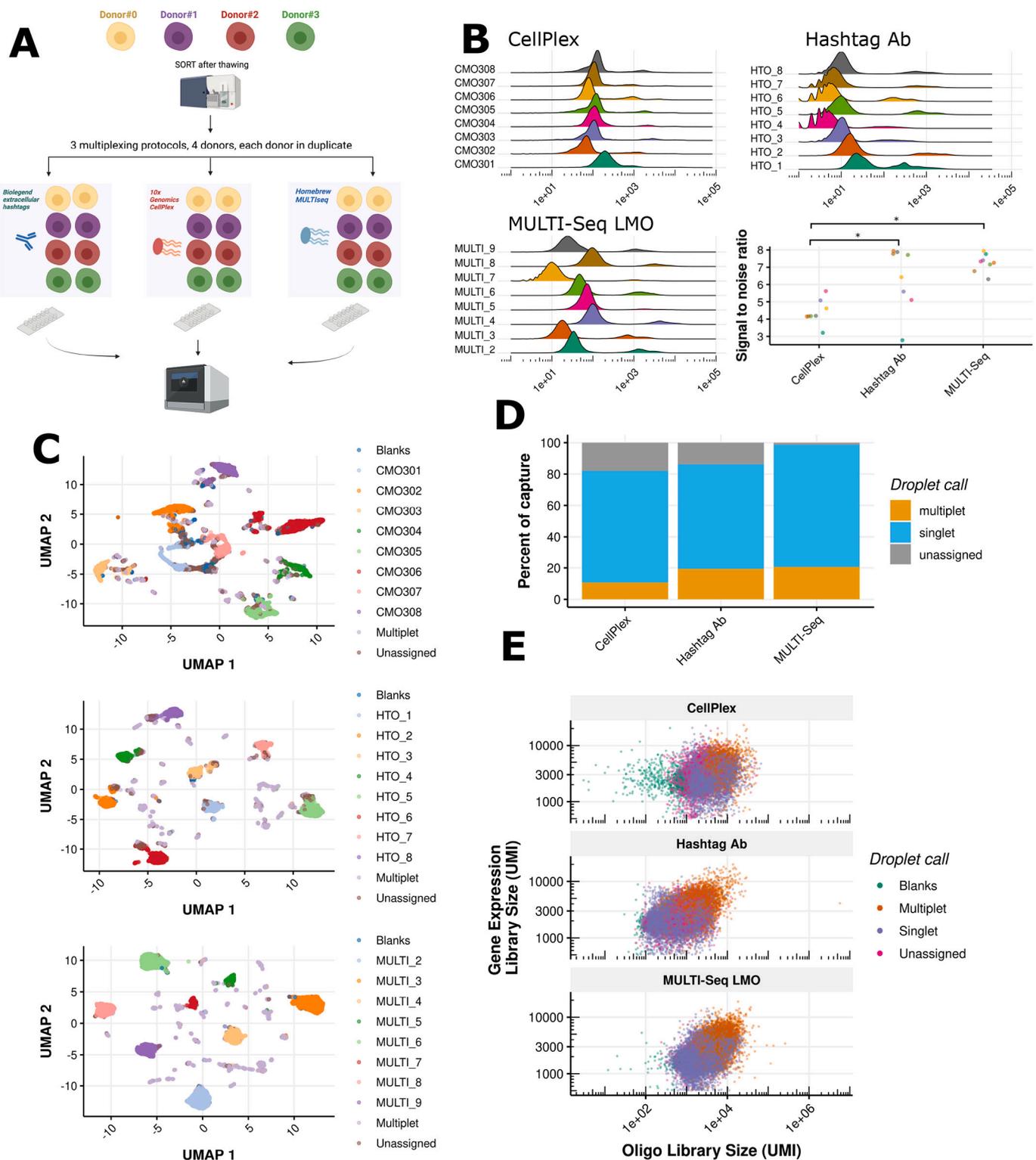


Fig. 1. Benchmarking sample multiplexing reagents in human PBMCs. (A) Experimental design. PBMCs isolated from four unrelated healthy donors were divided into technical duplicate prior to labeling. Each protocol was captured in a separate 10× Genomics v3.1 reaction. (B) Log10 transformed oligo tag counts for each multiplexing protocol with a summary of signal-to-noise. Signal-to-noise is defined as the difference between the mean background (left) and foreground (right) oligo tag counts on a log scale divided by variance. (C) UMAP dimension reduction visualization of multiplexing oligo tag counts for each protocol. Cells are colored by Cell Ranger multi call. (D) Summary of multiplexing tag calls per protocol as reported by Cell Ranger multi. Blanks and unassigned are both reported as unassigned. (E) Relationship between oligo tag and gene expression library size for each protocol tested.

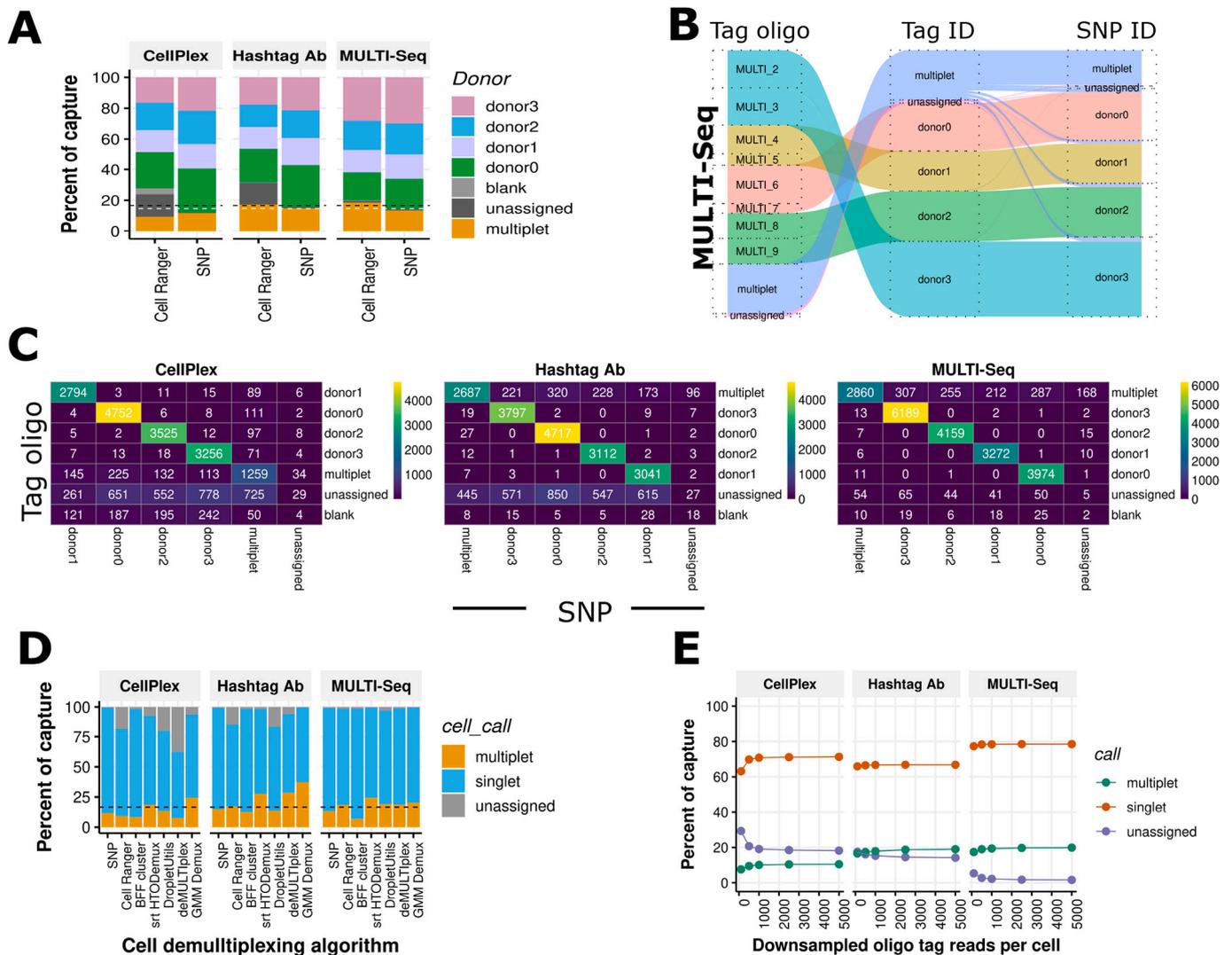


Fig. 2. Comparison of cell demultiplexing software (A) Comparison of oligo tag calls with Cell Ranger multi versus SNP calls. The white horizontal dashed line represents the theoretical doublet rate based on 4 SNP donors. The black line represents doublet rate based on eight multiplexed samples. (B) Alluvial plot for MULTI-Seq LMO comparing demultiplexing on a tag basis (left), donor basis (middle) and corresponding SNP calls (right). (C) Heatmap of the sample multiplexing oligo-snp assignment contingency table. (D) Comparison of droplet calls from multiple demultiplexing algorithms implemented in cellhashR. (E) Downsampling analysis of oligo tag libraries. Oligo tag libraries were downsampled to a fixed number of reads per cell and each dataset was demultiplexed with Cell Ranger multi.

Table 1

Overall classification accuracy. OCA is the number of matching assignments between sample multiplexing tags and SNP assignments divided by all cell-containing droplets.

Protocol	OCA	Unassigned
MULTI-Seq	0.926	0.0153
Hashtag Ab	0.804	0.145
CellPlex	0.761	0.185

Evaluation of the sequencing counts of the multiplexing tag libraries revealed a good separation of signal and background for diluted CellPlex (*p*-value 4.97e-6 and 1.72e-3 versus MULTI-Seq CMO and LMO respectively) (Fig. 3B). In contrast the two MULTI-Seq designs had poorer signal to background. During laboratory processing there was incomplete removal of supernatants due to concern over loss of cell pellets. With a cell input of 100,000 per well, pellets were invisible. The remaining dissociation media may have inhibited MULTI-Seq LMO, as it is quenched by proteins [6]. In line with this observation, MULTI-Seq CMO which is not quenched by culture media performed

comparatively better (Fig. 3C).

A further logistical issue related to shifting the sorting step until after multiplexing oligo labeling to conform to 10× Genomics supported protocols. This change necessitated a sequential sort of sample pools in the order MULTI-Seq LMO, MULTI-Seq CMO and CellPlex. Viability measurements dropped from 93% to 76% live cells between MULTI-Seq LMO and CellPlex for this reason (Fig. S5I). The number of cell-containing droplets retrieved was consistent with a drop in viability over time (Table S2).

Since CellPlex performed the best in this experiment and had the shortest time between cell sorting and single-cell capture, we later assessed if prolonged storage on ice had any effect on signal-to-noise metrics (Fig. S6). Indeed 30 min storage on ice increased background and reduced signal for the lipid based CellPlex reagent.

In the mouse embryo experiment, the MULTI-Seq reagents yielded low signal-to-noise ratios and a high proportion of unassigned droplets (Fig. 3D). As BFF cluster performed well in the low quality CellPlex PBMC dataset, we also assessed its performance in the E18.5 mouse embryo cells. Indeed, the proportion of droplets assigned to multiplets and singlet samples also increased.

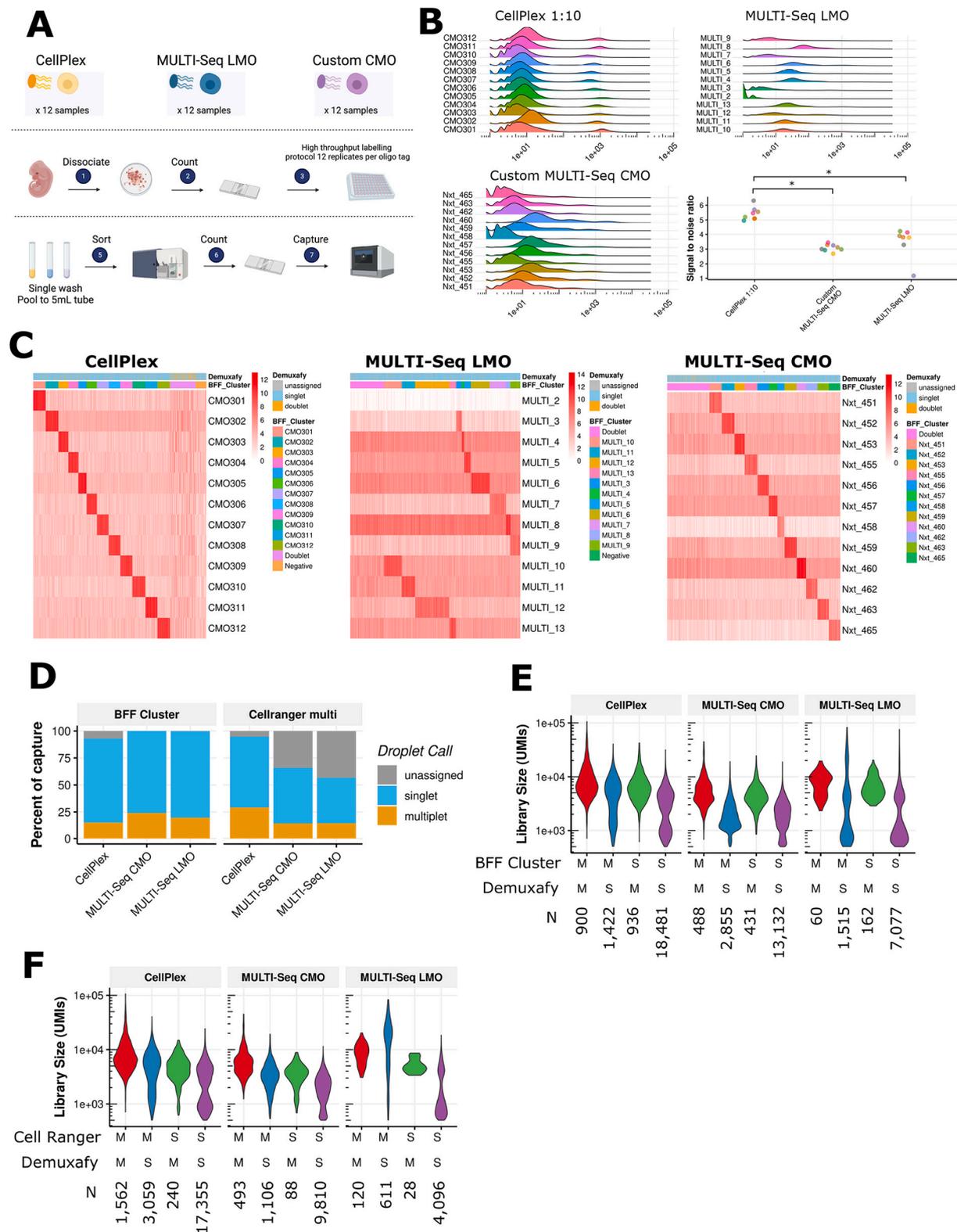


Fig. 3. Benchmarking sample multiplexing reagents in mouse embryonic brain. (A) Experimental design. Embryonic day 18.5 mouse brain from a single animal was used. The single cell suspension was split into 12 partitions for each reagent tested. A high throughput labeling protocol in 96 well plates was used. Cells were pooled prior to FACS sorting. (B) Log10 transformed oligo tag counts for each multiplexing protocol with a summary of signal-to-noise. (C) Heatmaps of oligo tag counts annotated by BFF Cluster call and Demuxify consensus call. (D) Identity of cell containing droplets from BFF Cluster and Cell Ranger multi algorithms. (E) Library size of gene expression library based on Cell Ranger demultiplexing calls. (F) Library size of gene expression library based on BFF Cluster demultiplexing calls.

As a single sample from an inbred mouse strain was used for this experiment, we were unable to use SNP genotypes as a ground truth. We therefore utilized a doublet detection software package, Demuxify, based on gene expression as an alternative source of droplet identity information. Demuxify is a wrapper program around many common algorithms [15]. Comparison of droplet calls from demultiplexing algorithms based on multiplexing tags with calls based on gene expression (Fig. 3E) showed no significant difference in gene expression library size between concordant multiplets and singlets. More singlets were recovered from the overlap of Demuxify and BFF Cluster compared to the number recovered from the overlap of Demuxify and Cell Ranger Multi

(Fig. 3F).

2.4. Comparison of sample multiplexing reagents in human tumor nuclei

Having evaluated sample multiplexing reagents in intact cells, we next compared the reagents in single-nucleus preparations. Nuclei more faithfully represent cell type composition in tissues that are difficult to dissociate enzymatically than whole-cell preparations [16,17]. Here, we used human ovarian carcinosarcoma patient-derived xenograft (PDX) tissue that was snap frozen as tissue pieces according to 10x Genomics best practices. Due to the differing growth rates of the different PDX

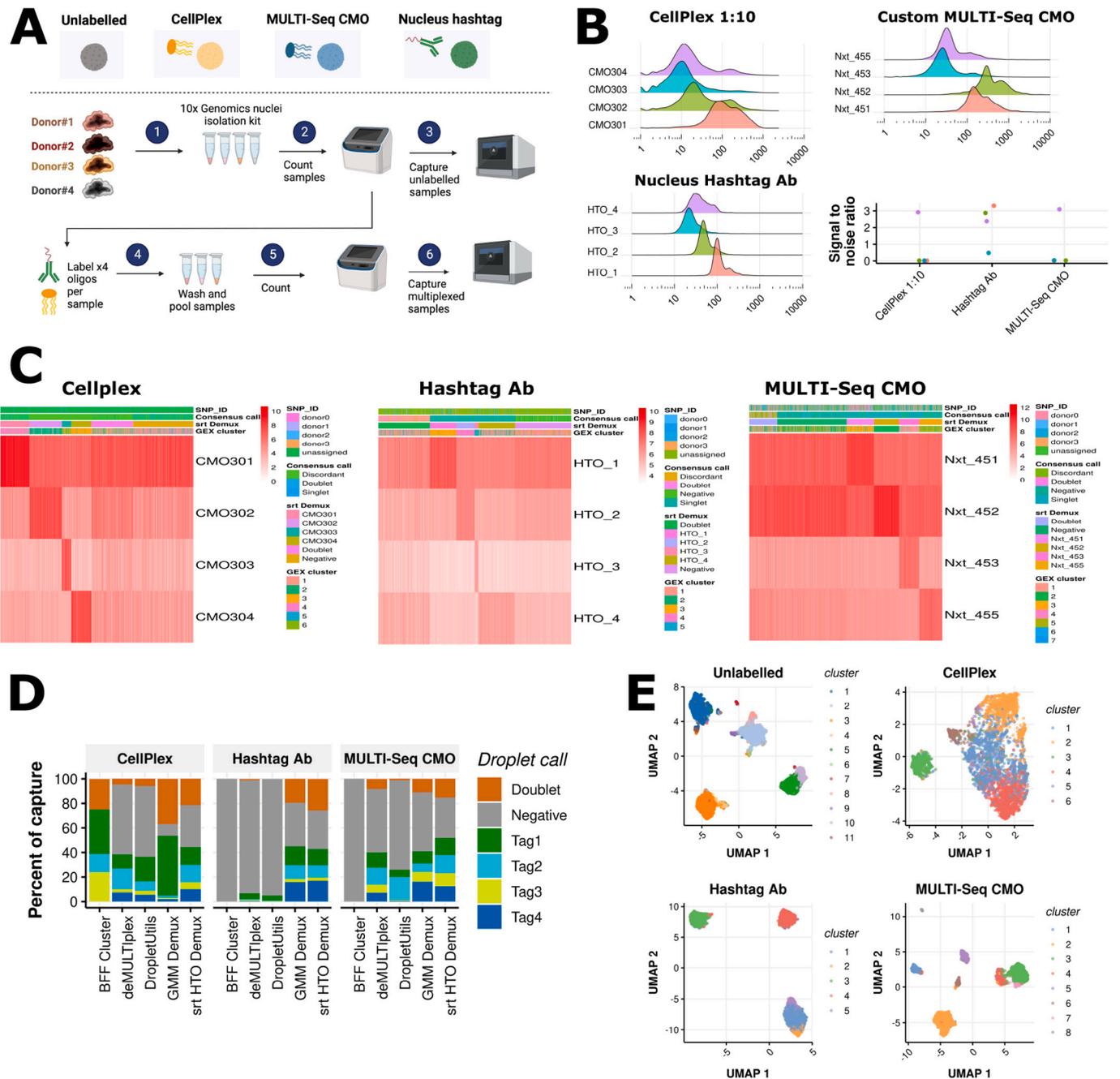


Fig. 4. Benchmarking sample multiplexing reagents in ovarian carcinosarcoma xenograft nuclei. (A) Experimental design. Nuclei were isolated from four PDX donor samples, performing an immediate microfluidics capture, followed by labeling the remaining nuclei with multiplexing oligos. (B) Log10-transformed oligo tag counts for each multiplexing protocol with a summary of signal-to-noise ratios. (C) Heatmaps of oligo tag counts annotated by cellhashR consensus call, Seurat HTODemux call, and cluster identity from parallel gene expression data. Cells (columns) are ordered by HTODemux call. (D) Comparison of droplet calls made by demultiplexing algorithms. Calls were generated by cellhashR. (E) UMAP dimension reduction visualization of gene expression data for each protocol. Cells are colored by gene expression cluster identity.

models it was impractical to process all samples fresh on the same day. We prepared nuclei with a 10× Genomics isolation kit and immediately performed a capture of unlabelled nuclei to identify the effect of prolonged sample handling required by labeling steps on the quality of the transcriptome. We subsequently labelled the remaining nuclei with three different multiplexing reagents (Fig. 4A).

The quality of the nuclei preparation immediately after isolation was good but visually deteriorated after the multiplexing oligo labeling step (Fig. S7A). The CellPlex sample experienced a clog resulting in a wetting failure and low recovery volume. While Souporecell estimated an ambient RNA content of 29.97% for droplets from the unlabelled sample, it could not generate an estimate for samples labelled with sample multiplexing oligos. In the labelled samples, the majority of cells could not be assigned to a SNP donor, likely due to a low molecular complexity and fraction of reads in cells (Fig. S7B).

The corresponding signal-to-noise ratio of the oligo tag libraries was poor with low separation from background (Fig. 4B and C). Cell ranger multi failed to assign the majority of nuclei to samples (Fig. S7C). We used the cellhashR package to compare droplet assignment methodologies; Seurat (srt) HTODemux performed the best under the expectation that samples would be in equal proportions (Fig. 4D).

Since each microfluidic capture contained cells from four PDX models, we next checked the gene expression data for separation by sample of origin after dimension reduction (Fig. 4E). The unlabelled sample showed four major clusters reflecting the SNP donors with minor satellite clusters reflecting cell doublets. This association was not as clear in the captures labelled with multiplexing oligos, likely due to the

lower quality of gene expression data (Fig. S7E).

2.5. Evaluation of fixed single-nucleus RNA-Seq in human tumor nuclei

In light of the poor performance of sample multiplexing oligos in nuclei from solid tumor samples, we evaluated fixed snRNA-Seq kits from Parse Biosciences (mini Evercode v2) and 10× Genomics (Flex v1, 4 barcodes) (Fig. 5A). Immediately after fixation the single nucleus suspension appeared free of debris and clumps (Fig. S8A). Following probe hybridization and sample pooling multiple washes are required for 10× Genomics Flex. With each centrifugation and resuspension step, the sample became increasingly more clumpy (Fig. S8B), resulting in a clogged microfluidics chip and failed capture. For the Parse Biosciences experiment, all four of the nuclei samples exhibited a degree of clumping; in 2 samples this was the minority of nuclei (Fig. S8C) but for the remaining 2 samples the majority of nuclei were clumped and omitted from processing (Fig. S8D).

To evaluate the accuracy of Parse Bioscience v2 demultiplexing, we utilized donor specific clustering based on gene expression (Fig. 5B) counting the number of cells labelled with reverse transcription barcodes that were were in each cluster (Fig. 5C). Excluding likely doublet cluster 2, 9 cells were assigned to a cluster where they were the minority donor, representing 0.28% of the experiment (Fig. 5D).

Since the 10× Genomics Flex capture was unsuccessful, we compared Parse Biosciences data to the unlabelled 10× Genomics v3.1 fresh nucleus dataset. Parse Biosciences had more reads in cells (Table S4) reflecting a more efficient use of sequencing resources.

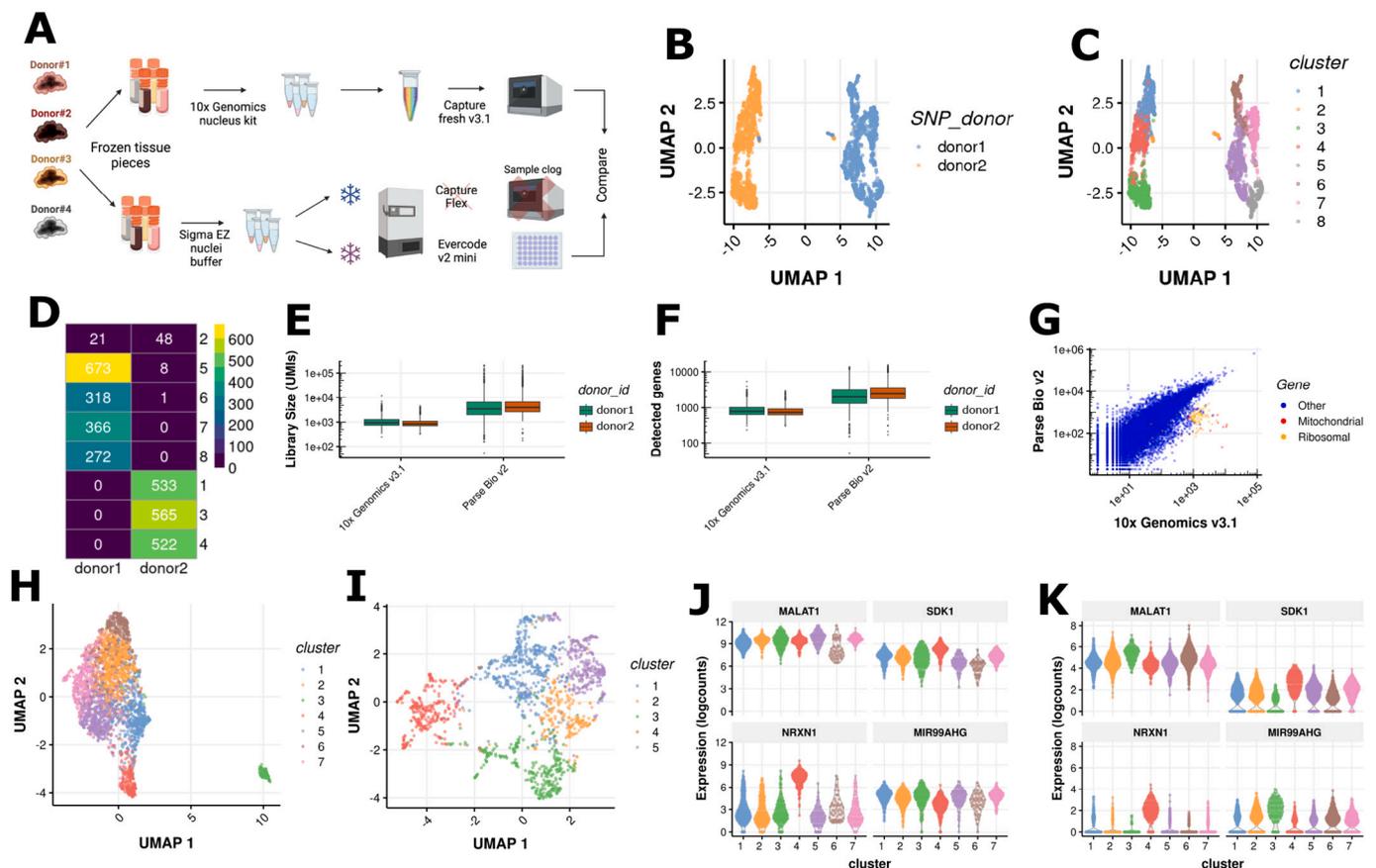


Fig. 5. Benchmarking fixed scRNA-Seq in Ovarian carcinosarcoma xenografts. (A) Experimental design. (B) UMAP of Parse Bioscience v2 gene expression data from both donors. (C) UMAP of Parse Bioscience v2 gene expression data from both donors colored by cluster. (D) For each gene expression cluster in C, the number of cells from each donor is indicated. (E) Library size comparison between scRNA-Seq protocols. Sequencing reads were downsampled to an equivalent number per cell. (F) Gene detection comparison between protocols. (G) Gene expression comparison between scRNA-Seq protocols, each dot is the sum of counts across all single cells for a gene. (H) UMAP of 10× Genomics v3.1 data for PDX donor 1. (I) UMAP of Parse Biosciences v2 data for PDX donor 1. (J) Expression plots of indicated genes per cluster in 10× Genomics v3.1 data. (K) Expression plots of indicated genes per cluster in Parse Biosciences v2 data.

Accordingly, the library sizes and number of detected genes were approximately five fold higher in Parse Biosciences when downsampling to an equivalent number of reads per library (Fig. 5E and F). Consistent with the manufacturer's specifications, the cell doublet rate was an order of magnitude lower for Parse Bioscience. There was a high concordance (Pearson correlation 0.918) in gene expression between technologies (Fig. 5G).

Importantly, the Parse Biosciences v2 dataset exhibited greater biological variation compared to the 10× Genomics v3.1 dataset. In the latter, an outlier cluster distorted the dimension reduction results (Fig. 5H). Marker gene analysis identified lncRNAs *MALAT1* and *MIR99AHG* as enriched in cluster 3 (Fig. 5J). These genes have been described to be enriched in stressed and dying cells [18]. However there was no difference in library size or mitochondrial gene percentage in this cluster (Fig. S8G). In contrast, the Parse Biosciences dataset captured biologically meaningful variation, revealing a distinct cluster expressing the *NRXN1* gene (Fig. 5K). Neurexin-1-alpha is a cell adhesion protein and may represent a more epithelial-like subpopulation within the tumor [19,20].

2.6. Evaluation of CRISPRclean destruction of abundant genes

Multiplexing tags reduce per-cell costs for cell capture and library preparation by increasing the yield of single-cell partitioning. However, superloading does not decrease the cost of sequencing. Instead, it can increase required sequencing volumes due to higher doublet rates and the need to sequence the sample multiplexing tag library. Thus methods that can reduce sequencing requirements are arguably of enhanced value for multiplexed experiments.

We evaluated the Jumpcode Genomics CRISPRclean Single Cell RNA Boost Kit in reducing the amount of uninformative sequence data by using a guide RNA library targeting unaligned reads, ribosomal, mitochondrial, and non-variable genes [21]. CRISPRclean lowered the proportion of ribosomal and mitochondrial genes in the PBMC library by over 30% (Fig. 6A).

Following CRISPRclean treatment, ribosomal and mitochondrial genes were eliminated from the list of most highly expressed genes, and *MALAT1* was considerably depleted (Fig. 6B). The enhancement in gene detection was moderate (Fig. 6C), while off-target depletion of genes was minimal (Fig. 6D). Interestingly, mitochondrially encoded ribosomal genes in the CRISPRclean sample were also degraded, potentially due to homology.

The per-cell percentage of counts for mitochondrial genes is a common quality control metric in scRNA-Seq analysis [22]. We investigated whether this metric remains reliable for removing low-quality cells in CRISPRclean-depleted samples (Fig. 6E). Although a positive correlation between untreated and depleted mitochondrial gene percentages was observed, more cells were discarded in the undepleted library (1115 vs 460). This was entirely explained by outlier-based mitochondrial gene percentage filtering (Fig. 6F).

We next confirmed that cell type recovery remained unaffected by the depletion of ribosomal and mitochondrial genes (Fig. 6G). We also examined whether CRISPRclean depletion altered the distribution of immune subsets. Following clustering and annotation, no significant difference in the composition of CRISPRclean-depleted libraries was observed (Fig. 6H).

3. Discussion

Our study compared various sample multiplexing reagents for scRNA-Seq experiments. We limited our comparison to commercially available technologies that do not require genetic manipulation of cells. For PBMCs, we found MULTI-Seq LMO to be the superior reagent. Hashtag antibodies also performed well, with the exception of a single tag oligo HTO1 in this particular experiment. Potential explanations include the formation of antibody aggregates, which is mentioned on the

frequently asked questions page of the 10× Genomics website. Reduced signal-to-noise ratios might also be attributed to incomplete removal of supernatants during wash steps.

CellPlex at the manufacturer's recommended concentration, exhibited the poorest performance due to high background. By titrating and diluting the CellPlex reagent ten-fold, the signal-to-noise ratio was improved. The additional background was likely introduced during storage of the pooled sample prior to single-cell capture, as the concentration of the CellPlex reagent is higher than for MULTI-Seq or hashtag antibody. Any passive transfer of multiplexing oligo tags by diffusion would be more pronounced for a concentrated reagent. The mechanism of tag oligo exchange between cells or between nuclei remains unclear. In our observations, it occurred in both high-quality cell lines and lower-quality dissociated tissues in a time dependent fashion. Based on 10× Genomics guidelines to maintain samples on ice after pooling, passive diffusion is the likely cause. This phenomenon could be further investigated using live cell fluorescent microscopy.

We caution against the use of sample multiplexing reagents particularly for fragile samples. Labeling of cells with multiplexing oligos necessitates additional sample incubation and washing steps. These additional manipulations can compromise the viability of fragile cell types such as mouse embryonic brain and tumor nuclei. The benefits of sample multiplexing with respect to cost and batch effect minimization should be assessed against the risk of reductions in data quality. Large-scale experiments involving upwards of 96 samples multiplexed together have only been demonstrated using cell lines [6]; However, our results suggest that it would be challenging to achieve such a scale with primary cells or tissues.

Comparing the assignments based on sample multiplexing oligos with SNP assignments confirmed the accuracy of all labeling strategies. Fewer than 2% of cells were misclassified to the wrong donor in the PBMC experiment. Nevertheless, SNP assignments operated with higher cell recovery with fewer cells being lost to an unassigned category. We recommend the use of SNPs over sample multiplexing reagents where applicable. In our algorithmic comparison, we found that all methods performed similarly when data quality is high as was the case in the PBMC experiment. This is consistent with other studies [23]. Where the algorithms diverge in terms of performance is on lower quality data. Here we found BFF Cluster could be used to rescue more cells, at the expense of an elevated false positive rate. Combining demultiplexing algorithms with orthogonal information, such as SNP genotypes or endogenous gene expression, may be a useful strategy to rescue poorly performing datasets, particularly with fragile samples where signal-to-noise is sub-optimal. It is important to note that we only tested algorithms available within the cellhashR package [13]. As new algorithms are constantly emerging [24,25], dedicated benchmarking studies focusing exclusively on bioinformatics analysis will be necessary [23]. Our primary focus was on laboratory elements that can enhance demultiplexing performance.

An advantage of the fixed scRNA-Seq kits from 10× Genomics and Parse Biosciences is that sample multiplexing is embedded in the molecular biology, with no additional sample handling required as for sample multiplexing oligos. While the Parse Biosciences kit has a lower doublet rate, the doublets in the 10× Genomics product are usable. For large experiments where gene expression is the sole readout, we expect that fixed RNA kits will become dominant. For PDX nuclei, we found that the Parse Biosciences kit was more reliable, whereas 10× Genomics Flex suffered from recurrent wetting failures. The greater propensity of nuclei for aggregation and clumping compared to intact cells may explain these failures encountered by the microfluidics-reliant 10× Genomics technology. It is also possible that the higher nuclei input may have also contributed to the formation of wetting failures. In the multiplexing configuration of 10× Genomics Flex, droplets containing heterotypic doublets are usable, therefore we input a higher number of nuclei into the microfluidics device compared to conventional 3' chemistry. This overloading feature is built into the pricing structure,

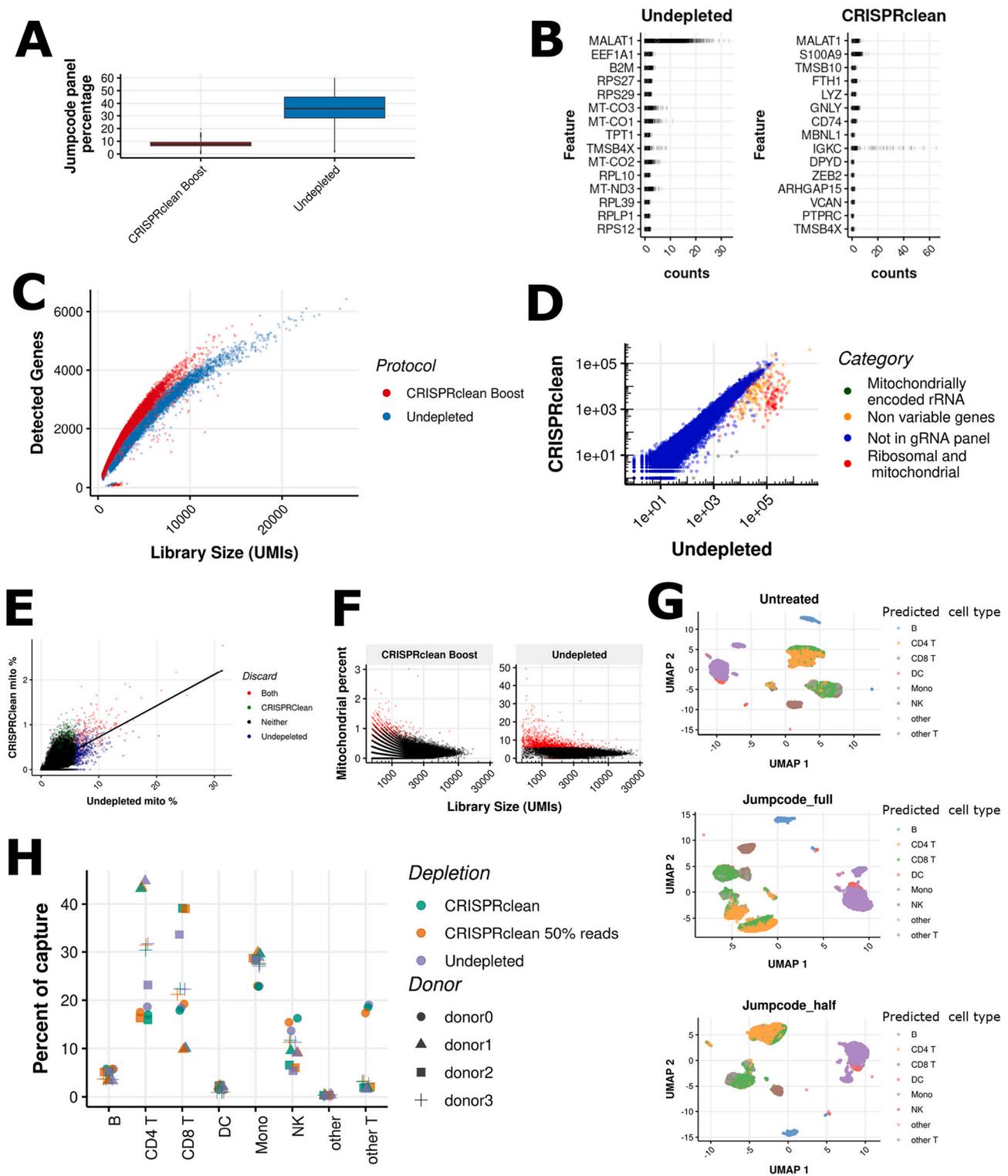


Fig. 6. Evaluation of Jumpcode CRISPRclean Single Cell RNA Boost Kit in MULTI-Seq LMO PBMCs. (A) Percentage of counts on genes in the CRISPRclean panel. The gene expression library was sequenced with and without CRISPRclean treatment. (B) Top 20 highly expressed genes in CRISPRclean treated and untreated libraries. Genes prefixed as “RP” or “MT-” represent ribosomal or mitochondrial genes, respectively. (C) Relationship between library size and number of detected genes per cell. (D) Gene expression comparison, each dot represents the sum of counts across all single cells for a gene. (E) Correlation of per-cell mitochondrial gene percentage. The trend line is a linear fit. Cells discarded by data driven thresholds are indicated. (F) Relationship between library size and mitochondrial gene percentage. Discarded cells are colored red. (G) UMAP dimension reduction based on gene expression data. The CRISPRclean library was additionally downsampled to 50% of the untreated library. (H) Summary of immune subsets recovered from untreated, CRISPRclean treated, and downsampled CRISPRclean treated gene expression data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

where Flex is more expensive than 5' and 3' kits, if overloading is not utilized.

CRISPRclean offers a promising approach to focusing sequencing resources on specific genes of interest. In our assessment of PBMCs, which represent the sample type used in the manufacturer's demonstration, we found that the technology is consistent with the manufacturer's claims. However, when applied to PDX samples, we observed a less pronounced re-focusing of sequencing effort. This discrepancy can be ascribed to the diminished expression of genes contained within the gRNA panel in PDX samples compared to PBMCs. While we conducted a comparison of cell type recovery, we did not perform an in-depth analysis on the impact of normalization, which typically relies on non-variable genes. In our data, we identified off-target depletion of genes, some of which have been previously reported [21]. The remaining off-target effects could be attributed to statistical noise and additional replicate experiments are necessary to evaluate their reproducibility. Following our assessment of CRISPRclean, a new gRNAs library has been introduced by Jumpcode, marketed as DepleteX, addressing known off-target effects and potentially enhancing the specificity of this technology.

In conclusion, our findings indicate that the choice and extent of multiplexing for scRNA-Seq should be contingent on the type of sample under investigation. Samples composed of cells that withstand *ex vivo* manipulation can accommodate a high degree of multiplexing. Conversely for delicate samples, it is more prudent to minimize multiplexing and instead invest in additional consumable costs. This approach ensures the preservation of high data quality.

4. Methods

All oligonucleotides were purchased from Integrated DNA Technologies. Sequences are provided in Table S5.

4.1. Ethical statement

PBMCs were isolated from unrelated healthy control donor samples obtained from the Volunteer Blood Donor Registry (VBDR, WEHI). Informed consent was obtained from all individual participants prior to inclusion in the study. The study was performed according to the principles of the 1964 Helsinki declaration and its later amendments and was approved by local Human Research Ethics Committee (WEHI Approved project 10/02). All experiments involving animals were performed according to the animal ethics guidelines and were approved by the WEHI Animal Ethics Committee (2019.024).

4.2. PBMC sample multiplexing labeling

We first undertook a round of optimization by flow cytometry, titrating the Total-Seq hashtag antibody to a concentration of ten-fold less than the manufacturer's recommendations (0.1 μg per reaction) (Fig. S1A). We next substituted the poly-A capture sequence of the described MULTI-Seq oligo [6] with the 10 \times Genomics feature barcode 2 sequence (Fig. S1B). Titrating the MULTI-Seq lipid modified oligos (LMO) resulted in a rapid loss of signal, therefore we used a concentration of 200 nM as reported in the original study (Fig. S1C).

CellPlex. Labeling was performed according to 10 \times Genomics demonstrated protocol CG000391 Rev. A "Cell Multiplexing Oligo Labeling for Single Cell RNA Sequencing Protocol" with an input of 250,000 cells. Library preparation was performed according to CG000388 Rev. A "Chromium Next GEM Single Cell 3' Reagent Kits v3.1 (Dual Index) with Feature Barcode technology for Cell Multiplexing".

MULTI-Seq. The MULTI-Seq protocol [6] was followed with 200 nM of each anchor-barcode complex being used in the labeling step. The poly-A capture sequence was replaced with the 10 \times Genomics feature barcode capture 2 sequence. This required the library preparation PCR to be performed with the Nextera read 1 primer instead of the TruSeq

read 1 primer. PCR conditions otherwise remained the same.

Total-Seq A Hashtag antibody. The Biologend protocol for Total-Seq A hashtag labeling was followed with the exception that ten-fold less antibody was used, 0.1 μg per labeling reaction. Each multiplexing protocol was captured on a separate 10 \times Genomics lane to avoid ambient oligo effects.

4.3. Mouse embryonic brain sample preparation

RosaERT2Cre/RosaERT2Cre mice were intercrossed with Loxcode/Loxcode mice [26,27]. Presence of a vaginal plug was used to determine day of conception. Pregnant dam was induced with 50 μg of 4 hydroxytamoxifen injected intravenously at day 7.5 of pregnancy. Embryos were collected at E18.5. After decapitation, the brain was dissected and processed for enzymatic dissociation. Embryos were dissected with the 10 \times Genomics protocol CG00055 Rev. C "Dissociation of Mouse Embryonic Neural Tissue for Single Cell RNA Sequencing" with the following modifications: Embryos were dissected in Hibernate-E Medium media (ThermoFisher A1247601), supplemented with 1% B27 (ThermoFisher 17,504,044) and 1 \times GlutaMAX (ThermoFisher 35,050,061). Benzonase Nuclease (Millipore E1014) was added to Papain (Millipore P4762) at a dilution of 1:5000 during dissociation.

4.4. Mouse E18.5 brain sample multiplexing experiment

100,000 cells were aliquotted per well across 12 wells of a round bottom 96 well plate (Falcon 353,077). To label cells at high throughput, a combination of 10 \times Genomics protocols CG000391 Rev. B "Cell Multiplexing Oligo Labeling for Single Cell RNA Sequencing Protocols with Feature Barcode technology" and CG000426 Rev. A "High Throughput Sample Preparation for Single Cell RNA Sequencing" were utilized.

CellPlex. CellPlex oligos were diluted 1:10 in PBS prior to incubation with 100,000 cells at room temperature. After 5 min, 200 μL of PBS + 1% BSA was added and samples centrifuged at 300 g for 4 min at 4 $^{\circ}\text{C}$. The supernatant was aspirated with a multichannel pipette, leaving approximately 10 μL of supernatant. A single wash of 200 μL PBS + 1% BSA was performed prior to pooling and sorting by flow cytometry. Library preparation followed the 10 \times Genomics CG000388 Rev. A protocol with no alternation of PCR cycle number.

MULTI-Seq lipid modified oligo (LMO). labeling was performed at half volume (100 μL total) to avoid overflow of the round bottom 96 well plate. 200 nM of anchor-barcode complex was used in the labeling step with 100,000 cells for 5 min on ice. Co-anchor oligo was then added for a further 5 min on ice prior to quenching with 200 μL PBS. A single wash was performed prior to flow cytometry. Library preparation followed the same workflow as for the PBMC experiment.

Custom MULTI-Seq cholesterol modified oligo (CMO). The MULTI-Seq LMO process was followed, substituting anchor and co-anchor oligos with a custom cholesterol modified oligo (CMO) containing the Nextera read 2 PCR handle (Supplementary table 5). Since the oligo was designed for compatibility with the CellPlex workflow, library preparation followed the 10 \times Genomics CG000388 Rev. A protocol without PCR cycle alterations.

10 \times Genomics capture After labeling with multiplexing tag oligos, individual samples were pooled at equal volumes without a cell count. The pooled single-cell suspension was counted and diluted to a final concentration of 812 cells per μL , aiming to load 35,000 cells into each lane and obtain 20,000 barcode-containing droplets at a 16% theoretical doublet rate. 10 \times Genomics v3.1 dual index kits were used. Each multiplexing protocol was captured on a separate 10 \times Genomics lane to avoid ambient oligo effects.

4.5. Ovarian carcinosarcoma xenograft fresh nuclei experiments

PDX models were established through transplanting fragments of

tumor tissue obtained from patients consented to the WEHI Stafford Fox Rare Cancer Program [28]. Following ethical endpoint and tumor dissection, rice-sized tissue pieces were snap-frozen on dry ice and stored at -80°C until processing. Single nucleus suspensions were generated from frozen tissue pieces using the Chromium Nuclei Isolation Kit with RNase inhibitor (PN-1000494), following user guide CG000505 Rev. A. As input tissue pieces weighed over 50 mg, the four tumor pieces from each donor were cut in half, and the nucleus preparation was performed in duplicate.

10 \times Genomics unlabelled capture. To examine the effects of extended storage time on nuclei integrity a capture was performed prior to any multiplexing labeling step, approximately 90 min before the labelled samples were captured. The single nuclei suspensions from each PDX donor were counted and pooled to a final concentration of 692 nuclei per μL to load 30,000 nuclei and obtain 17,177 barcode containing droplets at 13.82% theoretical doublet rate.

CellPlex. CellPlex oligos were diluted 1:10 in PBS prior to incubation with 250,000 cells at room temperature. After 5 min of labeling time at room temperature, 200 μL of PBS + 1% BSA was added and samples centrifuged at 300 g for 4 min at 4°C . The supernatant was aspirated with a multichannel pipette, leaving approximately 10 μL of supernatant. A single wash of 200 μL PBS + 1% BSA was performed prior to pooling and sorting by flow cytometry. Library preparation followed the 10 \times Genomics CG000388 Rev. A protocol with no alternation of PCR cycle number.

Custom MULTI-Seq custom cholesterol modified oligo (CMO). The same process for MULTI-Seq LMO was performed, except for the substitution of the anchor and co-anchor oligos for a custom cholesterol modified oligo (CMO) bearing the Nextera read 2 PCR handle. As the oligo was designed to be compatible with the CellPlex workflow, library preparation followed the 10 \times Genomics CG000388 Rev. A protocol with no alteration of PCR cycles.

TotalSeq A anti-Nuclear Pore Complex Antibody. The Biolegend Protocol for Total-Seq A hashtags was followed with 1 μg of antibody per labeling reaction (accessed 22 August 2022).

4.6. Ovarian carcinosarcoma xenograft fixed RNA experiments

To obtain sufficient nuclei to run the same suspension across both fixed kits we used EZ lysis buffer (Sigma NUC101), at the expense of greater debris. 500 μL of lysis buffer was added to approximately 50 μg frozen tissue pieces. Sample was triturated with wide bore p1000 tips until homogenized. Nuclei were incubated on ice for 5 min followed by centrifugation at 500 g for 5 min at 4°C . The supernatant was removed followed by two washes in 1 mL PBS + 1% BSA. Nuclei samples were then divided in half and fixed with manufacturer specific protocols and reagents.

Parse Biosciences Evercode version 2 on PDX nuclei. Between 150,000 and 500,000 nuclei were fixed and stored at -80°C for 6 weeks prior to processing using the Parse Biosciences Evercode WT Mini v2 kit (ECW02010) version 2.0.0 protocol. After visual inspection of the nuclei following fixation, 2 of the 4 samples exceeded the maximum clumping parameters and were omitted, leaving 2 remaining samples. Each sample was then processed in 2 wells of a version 2 mini kit following the manufacturer's guidelines. Two sublibraries of 5000 cells underwent the downstream cell lysis and PCR amplification steps.

4.7. Jumpcode CRISPRclean depletion

The MULTI-Seq LMO library from the PBMC experiment and unlabelled library from the ovarian PDX experiment were treated with Jumpcode CRISPRclean Single Cell RNA Boost Kit, (KIT1018) according to the manufacturer's instructions.

4.8. Bioinformatics analysis

All downstream analysis was performed in R version 4.2.1 [29]. The code, data and analyses used to generate these figures is available from GitHub. Each multiplexing labeling protocol evaluated was treated as a separate dataset without integration. Tabular data was manipulated with the tidyverse package [30].

Cell annotation was performed with Seurat version 4.0.6, TransferData function [31]. The reference for human PBMCs was provided with Seurat multimodal reference mapping vignette. The reference for mouse E18.5 brain was [32], subsetting cells for the E18.5 timepoint.

Doublet detection based on gene expression was performed with demuxify version 1.0.3 [15]. The majority vote from the output of DoubletFinder, scDbfFinder, scds and scrublet was used to assign multiplets and singlets. Ambient RNA estimation was performed with SoupX version 2.0 [?].

4.9. Statistical analysis

Performance comparison of oligo tag signal to noise ratios and demultiplexing algorithm calls was performed by Kruskal-Wallis test followed by Dunn test with Bonferroni correction. Differential cell type abundance analysis was performed by summarising the number of cells labelled with a given cell type annotation for each sample of origin, followed by testing for differences in the abundance of cell types between cell demultiplexing protocols with the edgeR package [33].

For differential gene expression analysis single-cells were first aggregated to pseudobulks based on CRISPRclean treatment with the aggregateAcrossCells function from the scuttle package [34]. The edgeR package was then used to compute differentially expressed genes.

AIdeclare

The initial rough draft of introduction, results and discussion sections were prepared by D.V.B without AI assistance. ChatGPT Plus (GPT-4) was subsequently used for copy editing of each paragraph with the prompt "act as a scientific copy editor and suggest improvements to my manuscript. List each change that you suggest." The edits were manually assessed before accepting or rejecting. ChatGPT Plus was then used to generate a 200 word abstract based on the results and discussion section. The output was manually reviewed and edited by D.V.B before inclusion and take they full responsibility for the content in the manuscript. Finally, ChatGPT was asked to provide five manuscript titles and keywords based on the results and discussion sections.

CRedit authorship contribution statement

Daniel V. Brown: Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Casey J.A. Anttila:** Methodology, Project administration. **Ling Ling:** Methodology. **Patrick Grave:** Methodology. **Tracey M. Baldwin:** Methodology. **Ryan Munnings:** Methodology, Resources. **Anthony J. Farchione:** Methodology, Resources. **Vanessa L. Bryant:** Resources. **Amelia Dunstone:** Software. **Christine Biben:** Investigation, Methodology, Resources. **Samir Taoudi:** Resources. **Tom S. Weber:** Methodology, Resources. **Shalin H. Naik:** Resources. **Anthony Hadla:** Conceptualization, Methodology, Resources. **Holly E. Barker:** Conceptualization, Methodology, Project administration. **Cassandra J. Vandenberg:** Conceptualization, Project administration, Resources. **Genevieve Dall:** Resources. **Clare L. Scott:** Resources. **Zachery Moore:** Methodology, Resources. **James R. Whittle:** Resources. **Saskia Freytag:** Resources. **Sarah A. Best:** Resources. **Anthony T. Papenfuss:** Resources. **Sam W.Z. Olechnowicz:** Methodology. **Sarah E. MacRaidl:** Methodology. **Stephen Wilcox:** Methodology, Project administration. **Peter F. Hickey:** Formal analysis, Investigation, Methodology, Software, Writing – review

& editing. **Daniela Amann-Zalcenstein:** Funding acquisition, Investigation. **Rory Bowden:** Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Writing – review & editing.

Declaration of competing interest

C.L.S reports non-financial support from Eisai Inc., Clovis Oncology and Beigene, grants and other support from Eisai Inc., AstraZeneca, and Sierra Oncology Inc., grants from Boehringer Ingelheim, other support from Roche and Takeda, and non-financial support and other support from MSD outside the submitted work. H.E.B reports grants from Eisai during the conduct of the study; other support from Eisai, Clovis, AstraZeneca, Sierra Oncology Inc., MSD, and Boehringer Ingelheim outside the submitted work. All other authors declare no competing interests.

Data availability

The count matrices and metadata are available as SingleCellExperiment objects at Zenodo, DOI: <https://doi.org/10.5281/zenodo.8031078>. Sequencing data is available at NCBI Bioproject PRJNA106402 for PBMCs and PRRJNA1064628 for mouse embryonic brain.

Acknowledgement

We thank all those individuals that donated the samples that enabled this study. We acknowledge the WEHI Genomics, Flow Cytometry Facility and Animal Bioservices for professional and timely service. D.V.B is supported by funding to the Advanced Genomics Facility from the Walter and Eliza Hall Institute. S.A.B. is supported by a Victorian Cancer Agency Mid-Career Research Fellowship (MCRF22003), Z.M. is supported by the Greg Lange Memorial Postdoctoral Fellowship, S.F. is supported by a National Health and Medical Research Council of Australia (NHMRC) Ideas Grant (1184421). V.L.B is supported by Sir Clive McPherson Family Fellowship and D-W Keir Fellowship. This work was financially supported in part through the authors'™ membership of the Brain Cancer Centre, support from Carrie's Beansies 4 Brain Cancer, a Priority-Driven Collaborative Cancer Research Scheme Grant funded by Cancer Australia to S.A.B. (2003127). Experimental design figures were created using [Biorender.com](https://biorender.com). This preprint is formatted using a LATEX class by HenriquesLab.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ygeno.2024.110793>.

References

- Picelli, Single-cell rna-sequencing: the future of genome biology is now, *RNA Biol.* 14 (5) (2017) 637–650.
- Jovic, X. Liang, H. Zeng, L. Lin, F. Xu, Y. Luo, Single-cell rna sequencing technologies and applications: a brief overview, *Clin. Transl. Med.* 12 (3) (2022) e694.
- J. Lee, D.Y. Hyeon, D. Hwang, Single-cell multiomics: technologies and data analysis methods, *Exp. Mol. Med.* 52 (9) (2020) 1428–1442.
- H.M. Kang, M. Subramaniam, S. Targ, M. Nguyen, L. Maliskova, E. McCarthy, E. Wan, S. Wong, L. Byrnes, C.M. Lanata, et al., Multiplexed droplet single-cell rna-sequencing using natural genetic variation, *Nat. Biotechnol.* 36 (1) (2018) 89–94.
- M. Stoeckius, S. Zheng, B. Houck-Loomis, S. Hao, B.Z. Yeung, W.M. Mauck, P. Smibert, R. Satija, Cell hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics, *Genome Biol.* 19 (1) (2018) 1–12.
- C.S. McGinnis, D.M. Patterson, J. Winkler, D.N. Conrad, M.Y. Hein, V. Srivastava, J.L. Hu, L.M. Murrow, J.S. Weissman, Z. Werb, et al., Multi-seq: sample multiplexing for single-cell rna sequencing using lipid-tagged indices, *Nat. Methods* 16 (7) (2019) 619–626.
- L. Fang, G. Li, Z. Sun, Q. Zhu, H. Cui, Y. Li, J. Zhang, W. Liang, W. Wei, Y. Hu, et al., Casb: a concanavalin a-based sample barcoding strategy for single-cell sequencing, *Mol. Syst. Biol.* 17 (4) (2021) e10060.
- J. Gehring, J. Hwee Park, S. Chen, M. Thomson, L. Pachter, Highly multiplexed single-cell rna-seq by dna oligonucleotide tagging of cellular proteins, *Nat. Biotechnol.* 38 (1) (2020) 35–38.
- D. Shin, W. Lee, J.H. Lee, D. Bang, Multiplexed single-cell rna-seq via transient barcoding for simultaneous expression profiling of various drug perturbations, *Sci. Adv.* 5 (5) (2019) eaav2249.
- C. Guo, W. Kong, K. Kamimoto, G.C. Rivera-Gonzalez, X. Yang, Y. Kirita, S. A. Morris, Celltag indexing: genetic barcode-based sample multiplexing for single-cell genomics, *Genome Biol.* 20 (1) (2019) 90.
- J. Cheng, J. Liao, X. Shao, X. Lu, X. Fan, Multiplexing methods for simultaneous large-scale transcriptomic profiling of samples at single-cell resolution, *Adv. Sci.* 8 (17) (2021) 2101229.
- V. Mylka, I. Matetovici, S. Poovathingal, J. Aerts, N. Vandamme, R. Seurinck, K. Verstaen, G. Hulselmans, S. Van den Hoecke, I. Scheyltjens, et al., Comparative analysis of antibody-and lipid-based multiplexing methods for single-cell rna-seq, *Genome Biol.* 23 (1) (2022) 1–21.
- G.J. Boggy, G. McElfresh, E. Mahyari, A.B. Ventura, S.G. Hansen, L.J. Picker, B. N. Bimber, Bff and cellhash: analysis tools for accurate demultiplexing of cell hashing data, *Bioinformatics* 38 (10) (2022) 2791–2801.
- R. Massoni-Badosa, G. Iacono, C. Moutinho, M. Kulis, N. Palau, D. Marchese, J. Rodriguez-Ubreva, E. Ballestar, G. Rodriguez-Esteban, S. Marsal, et al., Sampling time-dependent artifacts in single-cell genomics studies, *Genome Biol.* 21 (1) (2020) 1–16.
- D. Neavin, A. Senabouth, J. Lee, A. Ripoll, L. Franke, S. Prabhakar, C. Ye, D. McCarthy, M. Mele, M. Hemberg, et al., Demuxafy: Improvement in droplet assignment by integrating multiple single-cell demultiplexing and doublet detection methods, *bioRxiv*, 2022.
- N. Habib, I. Avraham-Davidi, A. Basu, T. Burks, K. Shekhar, M. Hofree, S. R. Choudhury, F. Aguet, E. Gelfand, K. Ardlie, et al., Massively parallel single-nucleus rna-seq with dronc-seq, *Nat. Methods* 14 (10) (2017) 955–958.
- E. Denisenko, B.B. Guo, M. Jones, R. Hou, L. De Kock, T. Lassmann, D. Poppe, O. Clément, R.K. Simmons, R. Lister, et al., Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus rna-seq workflows, *Genome Biol.* 21 (1) (2020) 1–25.
- A.H. Lee, L. Sun, A.Y. Mochizuki, J.G. Reynoso, J. Orpilla, F. Chow, J.C. Kienzler, R.G. Everson, D.A. Nathanson, S.J. Bensing, et al., Neoadjuvant pd-1 blockade induces t cell and cd1 activation but fails to overcome the immunosuppressive tumor associated macrophages in recurrent glioblastoma, *Nat. Commun.* 12 (1) (2021) 6938.
- S.E. Eksi, A. Chitsazan, Z. Sayar, G.V. Thomas, A.J. Fields, R.P. Kopp, P. T. Spellman, A.C. Adey, Epigenetic loss of heterogeneity from low to high grade localized prostate tumours, *Nat. Commun.* 12 (1) (2021) 7292.
- A.G. Alkhatami, A.K. Verma, M. Alfaifi, L. Kumar, M.Y. Alshahrani, A.R. Hakami, O.M. Alshehri, M. Asiri, M.M. Ali Beg, et al., Role of mirna-495 and nrnx-1 and cntn-1 mrna expression and its prognostic importance in breast cancer patients, *J. Oncol.* 2021 (2021).
- A.C. Pandey, J. Bezney, D. DeAscanis, E. Kirsch, F. Ahmed, A. Crinklaw, K. S. Choudhary, T. Mandala, J. Deason, J. Hamdi, et al., A crispr/cas9-based enhancement of high-throughput single-cell transcriptomics, *bioRxiv*, 2022, 2022–09.
- A.T. Lun, D.J. McCarthy, J.C. Marioni, A Step-by-Step Workflow for Low-Level Analysis of Single-Cell Rna-Seq Data with Bioconductor, 2016.
- G. Howitt, Y. Feng, L. Tobar, D. Vassiliadis, P. Hickey, M.A. Dawson, S. Ranganathan, S. Shanthikumar, M. Neeland, J. Maksimovic, et al., Benchmarking single-cell hashtag oligo demultiplexing methods, *NAR Genom. Bioinform.* 5 (4) (2023) lqad086.
- H.-U. Klein, demumix: demultiplexing oligonucleotide-barcoded single-cell RNA sequencing data with regression mixture models, *Bioinformatics* 39 (8) (2023) btad481.
- E. Swanson, J. Reading, L.T. Graybuck, P.J. Skene, Barware: efficient software tools for barcoded single-cell genomics, *BMC Bioinforma.* 23 (1) (2022) 1–14.
- C. Biben, T. Weber, K. Potts, J. Choi, D. Miles, A. Carmagnac, T. Sargeant, C. de Graaf, K. Fennell, A. Farley, et al., In vivo clonal tracking reveals evidence of haemangioblast and haematomesoblast contribution to yolk sac haematopoiesis, *Nat. Commun.* 14 (1) (2023) 41.
- T.S. Weber, C. Biben, D.C. Miles, S. Zhang, P. Tam, S. Taoudi, S.H. Naik, Loxcode in vivo clonal barcoding resolves mammalian epiblast contribution to fetal organs, *bioRxiv*, 2023, 2023–01.
- G.Y. Ho, E.L. Kyran, J. Bedo, M.J. Wakefield, D.P. Ennis, H.B. Mirza, C. J. Vandenberg, E. Lieschke, A. Farrell, A. Hadla, et al., Epithelial-to-mesenchymal transition supports ovarian carcinosarcoma tumorigenesis and confers sensitivity to microtubule targeting with eribulin, *Cancer Res.* 82 (23) (2022) 4457–4473.
- W. Huber, V.J. Carey, R. Gentleman, S. Anders, M. Carlson, B.S. Carvalho, H. C. Bravo, S. Davis, L. Gatto, T. Girke, et al., Orchestrating high-throughput genomic analysis with bioconductor, *Nat. Methods* 12 (2) (2015) 115–121.
- H. Wickham, M. Averick, J. Bryan, W. Chang, L.D. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, et al., Welcome to the tidyverse, *J. Open Source Softw.* 4 (43) (2019) 1686.
- Y. Hao, S. Hao, E. Andersen-Nissen, W.M. Mauck, S. Zheng, A. Butler, M.J. Lee, A. J. Wilk, C. Darby, M. Zager, et al., Integrated analysis of multimodal single-cell data, *Cell* 184 (13) (2021) 3573–3587.

- [32] D.J. Di Bella, E. Habibi, R.R. Stickels, G. Scalia, J. Brown, P. Yadollahpour, S. M. Yang, C. Abbate, T. Biancalani, E.Z. Macosko, et al., Molecular logic of cellular diversification in the mouse cerebral cortex, *Nature* 595 (7868) (2021) 554–559.
- [33] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a bioconductor package for differential expression analysis of digital gene expression data, *Bioinformatics* 26 (1) (2010) 139–140.
- [34] D.J. McCarthy, K.R. Campbell, A.T. Lun, Q.F. Wills, Scater: pre-processing, quality control, normalization and visualization of single-cell rna-seq data in r, *Bioinformatics* 33 (8) (2017) 1179–1186.