


RESEARCH

Open Access



Comparative analysis of antibody- and lipid-based multiplexing methods for single-cell RNA-seq

Viacheslav Mylka^{1,2}, Irina Matetovici^{1,3}, Suresh Poovathingal³, Jeroen Aerts^{1,3}, Niels Vandamme^{2,4}, Ruth Seurinck^{2,4}, Kevin Verstaen^{2,4}, Gert Hulselmans^{3,5}, Silvie Van den Hoecke¹, Isabelle Scheyltjens^{6,7}, Kiavash Movahedi^{6,7}, Hans Wils⁸, Joke Reumers⁸, Jeroen Van Houdt⁸, Stein Aerts^{3,5*†} and Yvan Saeys^{2,4*†} 

*Correspondence: stein.aerts@kuleuven.vib.be; yvan.saeys@ugent.be
†Stein Aerts and Yvan Saeys authors contributed equally to this work.

²Data Mining and Modelling for Biomedicine, VIB Center for Inflammation Research, Ghent, Belgium³VIB Center for Brain & Disease Research, Leuven, Belgium
Full list of author information is available at the end of the article

Abstract

Background: Multiplexing of samples in single-cell RNA-seq studies allows a significant reduction of the experimental costs, straightforward identification of doublets, increased cell throughput, and reduction of sample-specific batch effects. Recently published multiplexing techniques using oligo-conjugated antibodies or -lipids allow barcoding sample-specific cells, a process called “hashing.”

Results: Here, we compare the hashing performance of TotalSeq-A and -C antibodies, custom synthesized lipids and MULTI-seq lipid hashes in four cell lines, both for single-cell RNA-seq and single-nucleus RNA-seq. We also compare TotalSeq-B antibodies with CellPlex reagents (10x Genomics) on human PBMCs and TotalSeq-B with different lipids on primary mouse tissues. Hashing efficiency was evaluated using the intrinsic genetic variation of the cell lines and mouse strains. Antibody hashing was further evaluated on clinical samples using PBMCs from healthy and SARS-CoV-2 infected patients, where we demonstrate a more affordable approach for large single-cell sequencing clinical studies, while simultaneously reducing batch effects.

Conclusions: Benchmarking of different hashing strategies and computational pipelines indicates that correct demultiplexing can be achieved with both lipid- and antibody-hashed human cells and nuclei, with MULTISeqDemux as the preferred demultiplexing function and antibody-based hashing as the most efficient protocol on cells. On nuclei datasets, lipid hashing delivers the best results. Lipid hashing also outperforms antibodies on cells isolated from mouse brain. However, antibodies demonstrate better results on tissues like spleen or lung.

Keywords: Hashing, scRNA-seq, MULTI-seq, CITE-seq, Sample multiplexing



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Recent advances in single-cell and single-nucleus RNA sequencing (scRNA-seq and snRNA-seq) have had an unprecedented impact on our understanding of heterogeneous cell populations [1–6]. Current scRNA-seq experiments make it possible to routinely assay many thousands of cells at once, with recent datasets reporting hundreds of thousands to millions of cells [2, 6, 7]. In standard single-cell workflows, individual samples need to be processed in parallel, which limits the throughput, increases reagent costs and has the potential to introduce batch effects. Recently, several approaches for multiplexing have been described, including the use of pre-existing genetic diversity [8] or by introducing sample-specific barcodes using oligo-labeled antibodies [9], oligo-labeled lipid anchors [10], chemical labeling with oligos [11], or genetic cell labeling [12]. Multiplexing samples by labeling cells or nuclei with sample-specific barcodes before pooling and single-cell compartmentalization, a technique called “hashing,” allows for accurate detection of two (doublets) or more (multiplets) cells originating from different samples but captured in the same compartment, which inevitably occurs in standard single-cell workflows. Therefore, implementing a barcoding multiplexing paradigm allows users to drastically increase the number of cells or nuclei loaded per reaction, which consequently decreases per-cell library preparation cost.

The development of oligo-labeled antibodies directed against cell surface proteins for sample multiplexing, is a direct evolution from the Abseq [13], REAP-seq [14], and CITE-seq [15] protocols. One of the most widely used methods to date for detection of the cell epitome is by using the TotalSeq antibodies from Biolegend in combination with the scRNA-seq technologies from 10x Genomics. There are several types of TotalSeq antibodies to be used for cell labeling, including TotalSeq-A antibodies that contain a poly-A sequence mimicking a natural mRNA. These are designed to work with any sequencing platform that relies on poly-dT oligonucleotides as the mRNA capture method, while TotalSeq-B and TotalSeq-C antibodies contain a capture sequences that are compatible with the 10x Genomics 3' scRNA-seq (v3 or v3.1) and 5' scRNA-seq workflows, respectively. For human samples, the pre-mixed TotalSeq hashtag reagents recognize cell surface markers CD298 and β 2-microglobulin. The success of using antibodies for hashing depends on the ubiquitous expression of these target antigens, which can be problematic for some samples or species [16, 17], limiting the sample-agnostic, universal application of this method. An elegant way to overcome this limitation is the use of lipid anchors that are antigen independent and insert universally into the cell or nucleus membrane, irrespective of sample type [10]. Both antibody-based and lipid-based methods are simple, straightforward and generally applicable to a wide range of single cell applications and platforms, while genetic cell labeling and chemical labeling with oligonucleotides can be more challenging. It is still unclear which method is most accurate in separating samples based on the inserted hashtags. In terms of labour intensity both hashing methods are comparable.

In this study, we compared antibody-based and lipid-based sample barcoding methods by multiplexing four distinct human cancer cell lines. By exploiting the intrinsic genetic variations of these cell lines, demultiplexing by genetic diversity serves as a “ground truth” and allows determining the hashing accuracy of each method.

Single-cell suspensions are most commonly prepared from fresh tissues, which is a major roadblock for analysing clinical samples, archived materials and tissues such as the brain, for which cells cannot be readily extracted [5, 18, 19]. To overcome these limitations, single-nuclei can be extracted and analysed similarly to a standard single-cell workflow [20]. Therefore, we included nuclei samples in our comparison of hashing methods. We also evaluated antibody and lipid hashing on different mouse tissues. Finally, we evaluated hashing accuracy of human PBMCs using TotalSeq-A, -B antibodies and CellPlex reagents, which can be very relevant for single-cell sequencing in clinical studies.

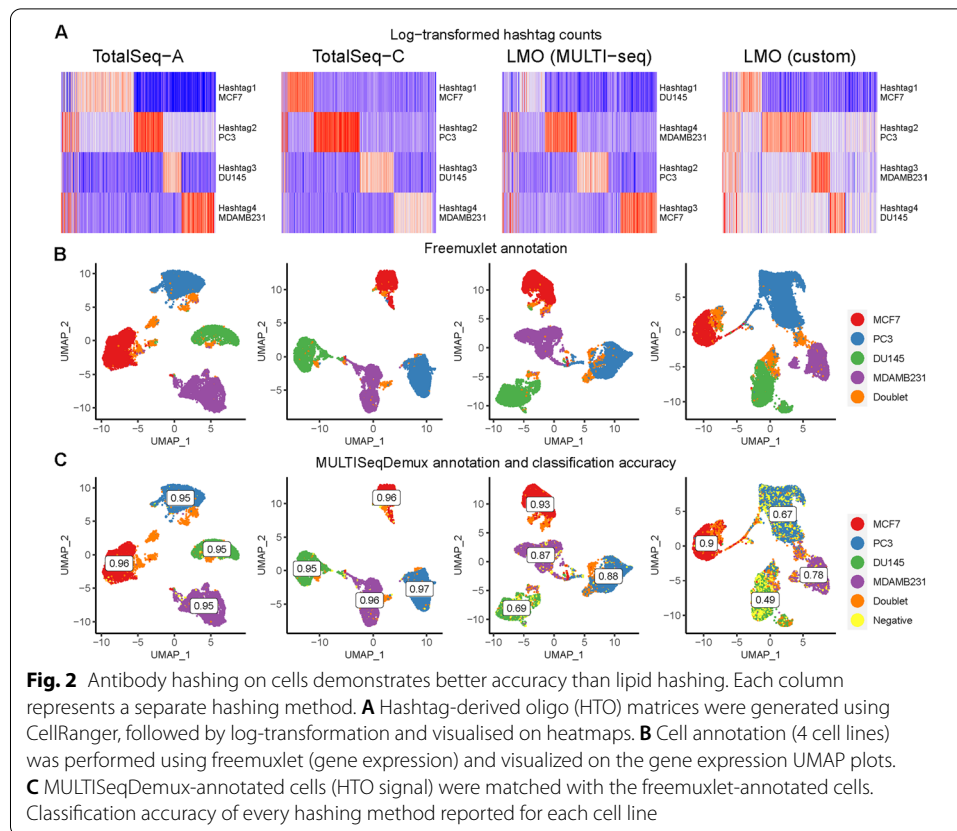
Results

Cell hashing using antibodies and lipids accurately demultiplexes majority of cells

We first compared antibody-based and lipid-based hashing techniques by pooling four hashed cancer cell lines (MCF7, PC3, DU145, and MDA-MB-231) into a single 10x Chromium run followed by a single analysis pipeline (Fig. 1) for each type of hashing. The hashing accuracy or Overall Classification Accuracy (OCA) was calculated as the overlap between cell line annotation using Seurat (MULTISeqDemux and HTODemux functions) or GMM-Demux [21], and freemuxlet as a reference annotation as follows: the OCA equals the number of all matching singlets (e.g., a given cell i annotated as one cell line/strain by both Seurat and freemuxlet), plus a number of matching non-singlets (mostly multiplets), divided by the number of all cells (Fig. 1).

Importantly, the tested demultiplexing functions from Seurat use antibody- or lipid-derived hashtag oligo (HTO) expression data for cell line annotation, while freemuxlet uses transcriptome data and a Bayes Factors approach to evaluate the likelihood of a hashed droplet being a doublet and estimation of its cell line origin. Freemuxlet exploits a similar algorithm as demuxlet [8], but does not require externally genotyped data (popscl package).

One of the first characteristics evaluated for all types of hashing techniques in our study was the level of detection of each HTO. When comparing the unnormalized, log-transformed HTO counts of each antibody- or lipid-derived hashtag oligo on a heatmap, we observed a distinct barcode for each of the four cell line clusters. Each cell cluster expresses an intrinsic gene marker relevant to each cell line (Additional file 1: Fig. S1). These findings are corroborated when each cluster is identified based on inherent genetic variation using freemuxlet (Fig. 2B), clearly separating the four cell lines and doublets. We also evaluated sample multiplexing on the same cells using 5' scRNA-seq chemistry and compatible TotalSeq-C hashing antibodies (Fig. 2, 2nd column). Cells demultiplexed using MULTISeqDemux had a very high concordance (OCA 0.96 for TotalSeq-A and TotalSeq-C hashing) with freemuxlet-based annotation (Figs. 1 and 2C). The custom lipid (LMO) and MULTI-Seq lipid hashing techniques demonstrated 0.68 and 0.84 classification accuracy, respectively (Figs. 1 and 2C). The MULTISeqDemux function annotated 1.1%, 1.8%, 21%, and 7% of cells in TotalSeq-A, TotalSeq-C, custom lipid (LMO), and MULTI-Seq lipid experiments respectively, as “Negatives” (Fig. 2C), which have a background expression for each hashtag according to the algorithm of the MULTISeqDemux function [10]. We observed ~2 times less unique molecular identifiers (UMIs) and ~3 times less genes per cell in “Negatives” compared



to singlets in the case of antibody hashing (no statistical tests performed) (Additional file 1: Fig. S2). Although a number of hashtag reads per cell is normalized across all the compared conditions here, it is noticeable that lower classification accuracy when using lipids compared to TotalSeq antibodies coincides with higher number of cells in the lipid hashing samples (Fig. 1). Nevertheless, the repeated TotalSeq-A experiment with an increased number of cells (17611) still demonstrated better classification accuracy (0.91) even when compared to the MULTI-Seq lipid hashing (0.84, 16,827 cells) (Fig. 1). Remarkably, the number of “Negatives” was relatively higher in DU145 cells hashed with both types of lipid hashes (custom and MULTI-seq) (Fig. 2C), which is in line with the relatively low hashtag level of detection in this cell line (Fig. 2A). This observation might point to a potentially lower affinity of the tested lipids with DU145 cells. When we compared doublet annotation by freemuxlet and MULTISeqDemux, we noticed that 2.4%, 6.3%, 37.9%, and 19.1% of freemuxlet-annotated doublets in TotalSeq-A, TotalSeq-C, custom lipid (LMO), and MULTI-Seq lipid experiments respectively, were recognized as singlets by the MULTISeqDemux function (Additional file 1: Table S1). We observed some variation across different hashing techniques in doublet annotation by different doublet detection tools including also scrublet and doubletFinder (Additional file 1: Fig. S3, S4, S5).

When comparing the correct assignment of cells by the MULTISeqDemux function with their attributed genotypes, the percentage of hashtag swapping (fraction of each hashtag found among singlets in non-labeled cell lines) varied considerably between

technologies with the lowest average value of $\sim 0,1\%$ mislabeled cells in both TotalSeq-A and TotalSeq-C hashing experiments, 0.89% mislabeled cells in the MULTI-Seq LMO hashing, and 2.68% in the custom LMO hashing experiment (Additional file 1: Table S2).

Regarding the performance of the demultiplexing functions from Seurat, we noticed that the MULTISeqDemux function (autoTresh =T) overall correctly deconvolves more cells than the HTODemux function (with default parameters) or GMM-Demux (Fig. 1).

Additional to the comparison of different hashing techniques on cells, we also estimated the level of expression of CD298 and $\beta 2$ -microglobulin in different human cells (mainly cancer cell lines) using flow cytometry. Both antigens are targeted by human TotalSeq -A and -C hashing antibodies, and for their detection, we used flow cytometry antibodies with the same clones. All the tested 17 human cell types including PBMCs, HEK293A, and THP-1 cells express both antigens, potentially enabling antibody-based hashing on these cells (Additional file 1: Fig. S7). To conclude, the Total-seq -A and -C hashing antibodies can provide good hashing capabilities for a range of human sample types.

Lipid (CMO)-based hashing outperforms antibody hashing on nuclei

For each of the four cell lines, we extracted the nuclei to compare the antibody-based and lipid-based hashing efficiency. Nuclei hashing using lipids (cholesterol-modified oligos or CMO) and TotalSeq-A antibodies demonstrated overall lower signal-to-noise ratio of hashtag expression compared to TotalSeq antibody cell hashing (Fig. 2A vs Fig. 3A). Nevertheless, the nuclei hashed with cholesterol oligos followed by demultiplexing using MULTISeqDemux demonstrated 0.84 OCA (concordance with the reference annotation by freemuxlet) or 0.74 when using the HTODemux function (Fig. 1). TotalSeq-A-based nuclei hashing was less accurate (0.51 for MULTISeqDemux and 0.54 for HTODemux function), mainly due to a high number of nuclei recognized as “Negatives” (cells with a background signal for each hashtag) by the demultiplexing functions used for this benchmarking study (Fig. 3C). It was especially prominent in the MCF7 cell cluster (Fig. 3C, 2nd column). However, it was possible to improve the hashing efficiency by adjusting parameters of the demultiplexing functions. For example, changing the “positive.quantile” parameter in the HTODemux function from a default value of 0.99 to 0.9, increased the classification accuracy of the TotalSeq antibody nuclei dataset by more than 10% (from 0.54 to 0.65) (Additional file 1: Fig. 6). We also observed that a large number of negatives in the MCF7 cluster (Fig. 3C) can be explained by an issue with the “hashtag 1” antibodies, since when other cells (DU145) were stained with antibodies from the same vial in a repeated experiment, many DU145 cells were also recognized as “Negatives” by Seurat (Fig. 3B, C, 3rd column). Importantly, CMO nuclei hashing still outperformed the TotalSeq A nuclei hashing (0.84 % vs 0.65 %), when hashtag 1 (DU145) was excluded from the hashing efficiency calculation and compared to CMO nuclei hashing on three cell lines.

Interestingly, the tested hashing techniques on nuclei were characterised by an opposite pattern compared to cell hashing with respect to the background expression for each hashtag (reflected in the number of detected “Negatives”). Accordingly, the MULTISeqDemux function annotated more droplets (33%) as “Negatives” in antibody nuclei hashing compared to the CMO nuclei hashing experiment (8.4% droplets detected as

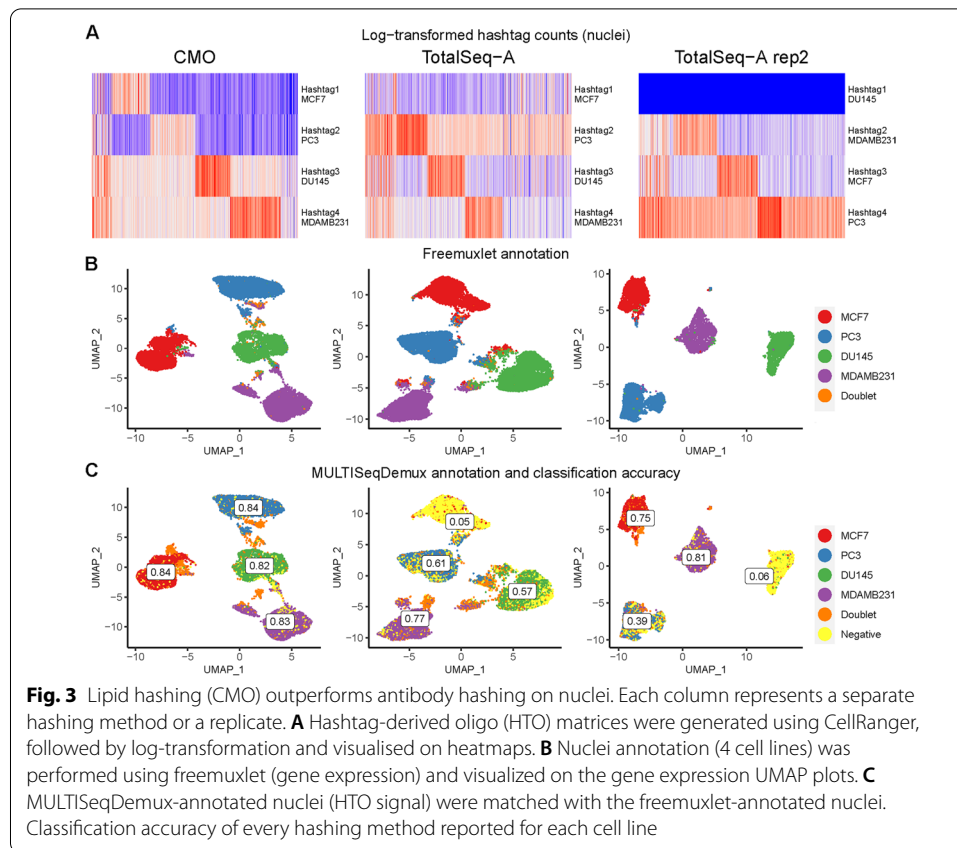


Fig. 3 Lipid hashing (CMO) outperforms antibody hashing on nuclei. Each column represents a separate hashing method or a replicate. **A** Hashtag-derived oligo (HTO) matrices were generated using Cell Ranger, followed by log-transformation and visualised on heatmaps. **B** Nuclei annotation (4 cell lines) was performed using freemuxlet (gene expression) and visualized on the gene expression UMAP plots. **C** MULTISeqDemux-annotated nuclei (HTO signal) were matched with the freemuxlet-annotated nuclei. Classification accuracy of every hashing method reported for each cell line

negatives) (Fig. 3C, 2nd column vs 3C, 1st column). This contrasts with the hashing of cells, where antibody hashing delivered less “Negatives” compared to the lipid hashing (Fig. 2C vs Fig. 3C).

The mislabeling hashtag rate for the nuclei samples varied from 0.89% of swapped labels among singlets for CMO technique to ~ 4% for TotalSeq-A nuclei hashing (Additional file 1: Table S2).

We also wanted to compare gene expression-related metrics in samples from the hashing experiments of this study and other non-hashed samples on the single cell line normalized to the same sequencing depth. We did not observe major differences in the number of genes and UMIs per nucleus when comparing MCF7 nuclei from hashing (antibody and lipid) and non-hashed experiments (Additional file 1: Fig. S9). The same conclusion was valid for all the tested hashing techniques on MCF7 cells (Additional file 1: Fig. S10).

In general, opposite to the cell hashing, we observed better hashing results on nuclei when using lipid-based hashing (CMO) compared to the antibody (TotalSeq-A) nuclei hashing.

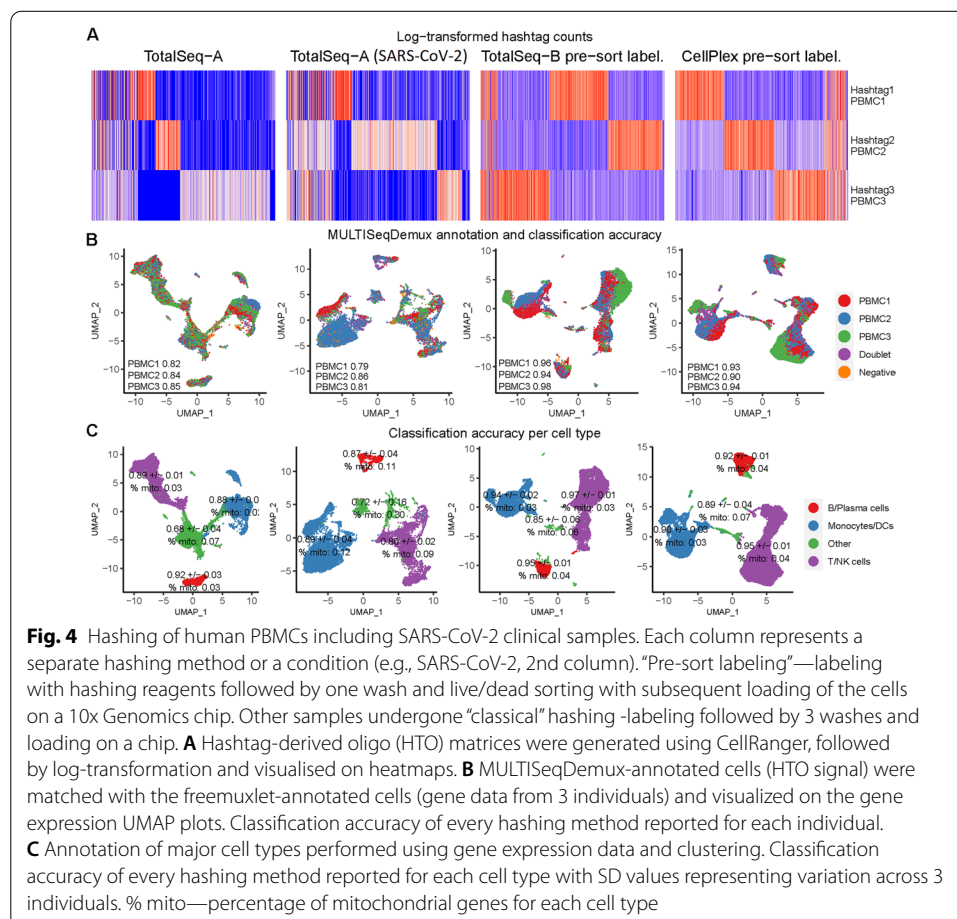
Antibody hashing can be successfully applied on clinical PBMC samples

In the context of the COVID-19 virus pandemic, we also tested TotalSeq-A antibody hashing on PBMCs from healthy individuals and SARS-CoV-2 patients in the frame of a COVID-19 clinical study (NCT04326920). All samples additionally contained a large

277 antibody CITE-seq (TotalSeq-A antibodies) panel, thus making this PBMC hashing evaluation even more relevant for other multiomics scRNA-seq experiments. The overall classification accuracy (OCA) for the pool of PBMCs from 3 healthy patients according to MULTISeqDemux (autoTresh = T) function was 0.82. For the pool of 3 SARS-CoV-2 patients, it was 0.84. (Fig. 1). Cell clustering analysis demonstrated a presence of all major cell types (T/NK cells, B cells, DCs/monocytes etc.) across 3 individuals in both pools, with no big variation of hashing accuracy across the main cell types (Fig. 4C, 1st and 2nd columns). Remarkably, cell clusters grouped as “Other”, that contain other cell types including potentially dying T-, B-, NK- cells or monocytes (indicated by a higher percentage of mitochondrial genes), had lower hashing accuracy (Fig. 4C).

The MULTISeqDemux function annotated 5.2% and 6.3% cells as “Negatives” in antibody hashing of healthy and diseased samples, respectively (Fig. 4B). According to freemuxlet, the hashtag mislabeling rate was 4.76% of swapped singlet labels for healthy sample and 1.53% for SARS-CoV-2 sample (Additional file 1: Table S2).

Next, we also evaluated the TotalSeq-B and CellPlex (a multiplexing solution from 10x Genomics) hashing on human PBMCs. Since these 2 samples were not part of the abovementioned COVID-19 study, we opted here for another labeling protocol (so called “pre-sort labeling”); labeling with antibodies or lipids followed by only one wash



and subsequent live/dead cell sorting (FACS). This resulted in a very high OCA: 0.96 for TotalSeq-B antibodies and 0.93 for CellPlex hashes (Fig. 4, 3rd and 4th columns). It is also important to mention that the CellPlex cDNA libraries are only compatible with the 10x Genomics dual sample indexing (i5 and i7). Therefore, this type of sample indexing was used for the TotalSeq-B and CellPlex hashing here, in contrast to all other cDNA libraries in our study where single sample 10x Genomics indexing (i7) was used.

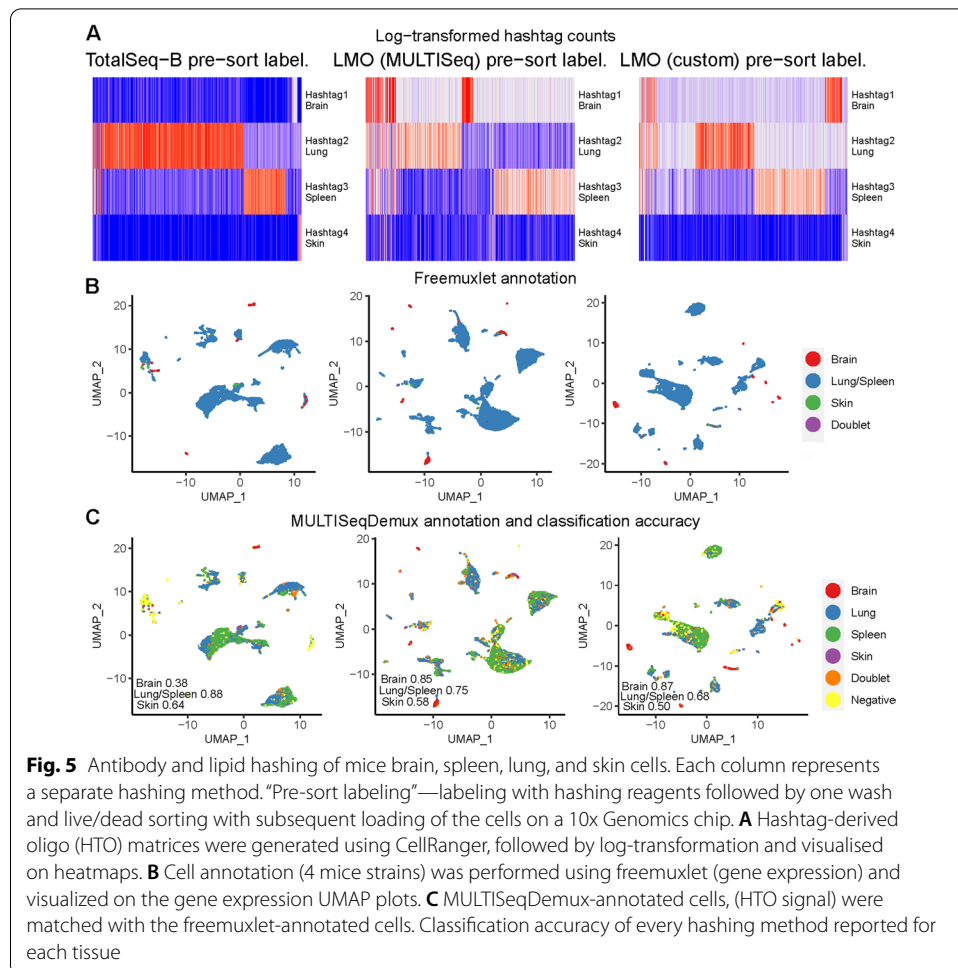
Antibody and lipid hashing of primary mouse tissues

scRNA-seq experiments on primary tissues are usually less straightforward when compared to cell lines or PBMCs due to the preceding tissue dissociation steps. In our study we evaluated antibody and lipid hashing of several mouse tissues derived from different mice strains using multiple dissociation protocols and labeling strategies.

First, we evaluated TotalSeq-B antibodies, custom and MULTISeq LMO hashing on the brain (adapted Miltenyi Biotech Neural Tissue Dissociation kit), spleen, lung, and skin cells where we applied “pre-sort labeling.” We observed that the OCA for brain cells was higher when using lipid hashing: 0.87 when using custom LMOs, 0.85 with MULTISeq lipids, and only 0.38 with antibodies. The antibody hashing resulted in many “Negatives” assigned to mostly non-microglia/macrophage cells (*C1qa*-negative cells) (data not shown).

We could isolate only a few skin cells in this experiment which demonstrated 0.64, 0.5 and 0.58 OCA for antibody, custom LMO and MULTISeq lipid hashing, respectively. We also learned from this experiment that freemuxlet is not able to separate cells originating from closely related mice strains: C57BL6J and C57BL6N. Thus, we report here a common OCA for the spleen and lung. Nevertheless, we could see that antibody hashing (0.88 OCA) outperforms lipid hashing (0.68 OCA for custom and 0.75 for MULTISeq) on the lung and spleen cells (Fig. 5B). We then confirmed this result for antibody hashing in another experiment where the spleen and lung cells were isolated from distinct mice strains: 129SVEvs7 and C57BL6J. The spleen hashing resulted in OCA 0.92 and lung antibody hashing - 0.84. Here we also evaluated a rather classical “post-sort labeling” hashing strategy where sorted live cells were labeled with antibodies followed by 2 washes (OCA 0.92 and 0.8 for the spleen and lung, respectively). Unfortunately, the brain cells were contaminated with blood during isolation in this experiment. As a result, the MULTISeqDemux function annotated basically all brain cells as “Negatives” in the “pre-sort labeling” condition (Additional file 1: Fig. S11, 1st column). In the “post-sort labeling” sample, the OCA for brain cells was only 0.30 (Additional file 1: Fig. S11, 2nd column), which is even lower than in the “pre-sort” labeling condition of the previous experiment where it was 0.38 for antibody brain cell hashing (Fig. 5).

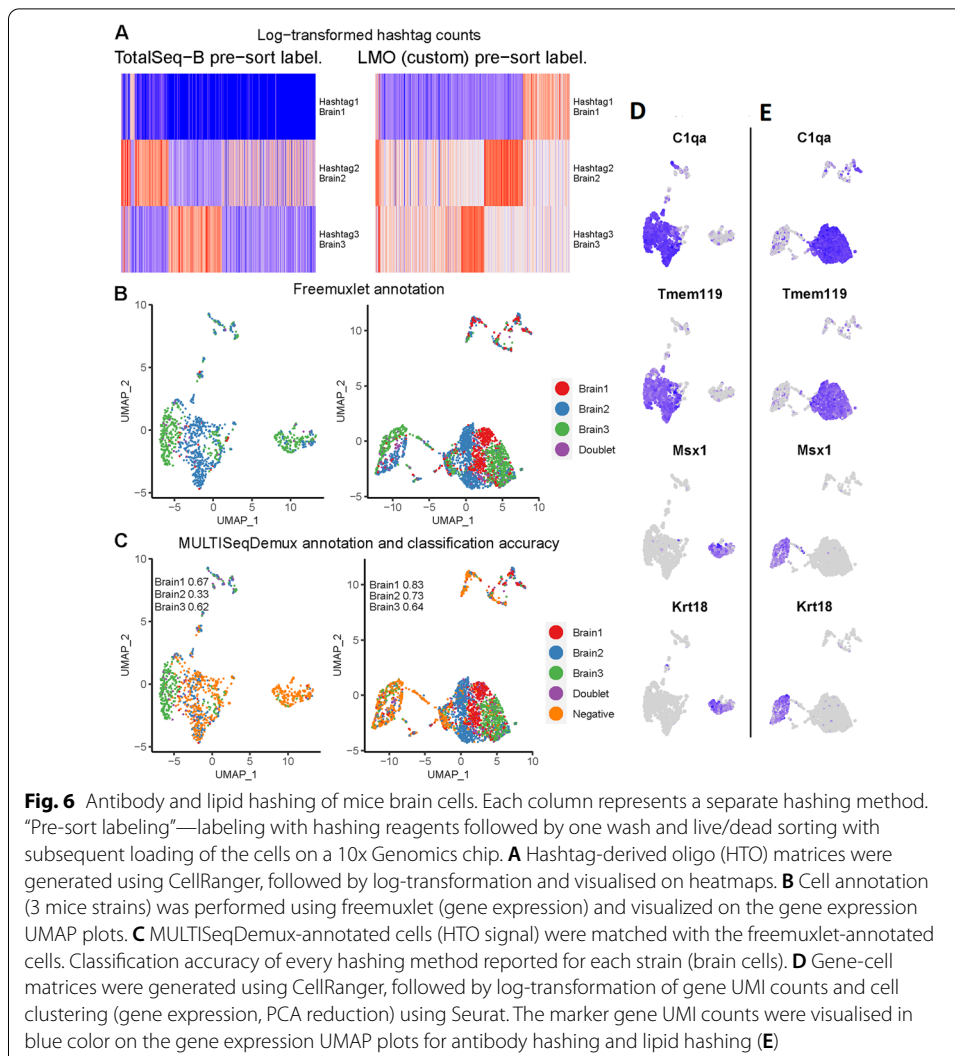
Next, we decided to evaluate antibody and lipid hashing on brain cells using another isolation protocol which also includes transcatheter perfusion to prevent blood contamination [4]. In this case, lipid hashing (custom lipids) also demonstrated better hashing accuracy than antibodies (OCA 0.73 vs 0.48) (Fig. 6). Here, we also applied hashing before live/dead cell sorting (“pre-sort labeling” protocol). Regarding the cells that could not be assigned to any hashtag (“Negatives”), even in the lipid hashing sample, these were the cells mostly belonging to cell clusters that express markers of ependymal cells (*Msx1*, *Krt181*, etc) [22]. The cells which were mostly successfully demultiplexed



when using lipid hashing are microglia/macrophages (*Tmem119/C1qa* positive cells) [4] (Fig. 6D, E).

Discussion

Correct sample demultiplexing after hashing depends on several factors: the hashing strategy itself (antibodies versus lipids), which subsequently can depend on the cell surface antigen repertoire in case of antibody hashing or biochemical/biophysical properties of membranes in case of lipid hashing. Also, this accuracy depends on the demultiplexing methods that can be subsequently influenced by other parameters (cell number, cells or nuclei, sequencing depth, etc.). In order to accurately determine the efficiency of demultiplexing, we hashed different cell lines that carry intrinsic genetic variations, which serve as a ground truth control. Importantly, for one of the hashing datasets, we compared two reference demultiplexing tools: freemuxlet which was used in this study, and the genotyping data-dependent tool demuxlet [8]. The cell deconvolution of the “TotalSeq-A cells rep1” sample using HTODemux demultiplexing function demonstrated a 0.94 OCA (overall classification accuracy) when compared with freemuxlet, and 0.95 when compared with demuxlet as a reference. Our comparison



demonstrates that even without external genotyping data (SNPs), required for demuxlet, it is possible to deconvolve different cells using a similar algorithm (freemuxlet). Although it is possible to demultiplex samples using only freemuxlet when pooling cells from different patients or from distinct mouse strains, the tool fails to separate strains such as C57BL6N and C57BL6J for example. We also noticed that “doublet vs singlet” classification by freemuxlet was not always consistent with the annotation by DoubletFinder [23] or Scrublet, the commonly used tools for computational doublet prediction [24]. For some data sets freemuxlet was underestimating the number of doublets. Additionally, the doublets annotated by HTO- and MULTISeqDemux were also not always aligned with the output of DoubletFinder and Scrublet. These findings are in accordance with the observation of a relatively high disagreement between DoubletFinder, Scrublet, and demuxlet [25]. This can be partially explained by the reliance on different data modalities that can lead to a certain bias in detection of heterotypic or homotypic doublets. For instance, the functions incorporated in Seurat (e.g., HTO-Demux,

MULTISeqDemux) predict doublets based on the hashtag signal (in particular combinations of different hashtags), while other tools (e.g., DoubletFinder) use the gene expression data and can detect doublets formed by transcriptionally distinct cells [9, 23]. Thus, it is worthwhile to consult both types of methods (gene expression- and HTO-based) while keeping in mind a theoretical multiplet rate (~0.8% for every 1000 captured cells on a 10x Genomics chip, CG000206 Rev D). Overall, this limitation in correct doublet prediction might diminish demultiplexing accuracy and requires further optimization. The impaired doublet annotation due to increased multiplet rate together with a failure to assign hashtags to a number of cells (“negatives”) might explain the lower hashing efficiency in the samples with higher number of cells loaded/captured on chip. For example, the captured cell number in the repeated TotalSeq-A experiment (rep2) was increased by 48%, while the number of doublets (detected by freemuxlet) and negatives (assigned by MULTISeqDemux) increased by 125% and 325%, respectively. An important parameter “signal-to-noise ratio” [22] was kept in this comparison on the same level by labeling equal number of cells with the same quantity of hashing antibodies.

It is important to mention that our comparison of different hashing strategies was applied to cells from filtered gene-cell matrices as produced by Cell Ranger v3 without further outlier filtering (e.g., based on mitochondrial gene expression). We could see in our PBMC dataset that in terms of cell calling, this filtered matrix contains almost the same number of cells as a “raw” matrix after filtering out all droplets with less than 200 UMIs (gene expression) and all droplets expressing one gene in less than 3 droplets. We observed that the tested demultiplexing functions from Seurat assigned some cells with relatively low number of genes and UMIs to the group “negatives”. Hence, not surprisingly, the hashing efficiency can be improved by filtering out the cell outliers. For example, correct PBMC annotation (healthy control patients) was improved from 0.84 to 0.89 after filtering out 2660 from 14635 cells using the “scater” Bioconductor package (identifying outliers based on the number of genes and UMIs per cell). On the other hand, a stringent filtering might cause a loss of biologically relevant cell subtypes. Thus, optimization of the sample demultiplexing should be performed in parallel with a deeper biological sample analysis. For evaluation of TotalSeq-B antibodies and CellPlex on PBMCs, we opted for a different protocol where labeled cells were sorted (live/dead) before loading on a chip. This “pre-sort labeling” might have contributed to a higher classification accuracy compared to the classical hashing protocol (labeling without sorting). It is in line with another observation that in PBMC clusters with a higher percentage of mitochondrial genes, hashing accuracy was lower.

We can conclude that the type of hashing strategy must be chosen based on the hashtag antigen expression, since it is known that some cells might not express both antigens targeted by available TotalSeq hashing antibodies. For instance, CD45 and MHC I antigens (targeted by mouse hashing antibodies from BioLegend) are not or lowly expressed in mouse C3 and B16-BL6 melanoma cells (Additional file 1: Fig. S8). However, a relatively lower affinity of certain cells towards different types of lipid hashes can be also observed. Nevertheless, lipid-based hashing might be a preferred strategy for multiplexing of samples with a low or unknown expression of hashing antigens (CD45 and MHC I), provided the sample demultiplexing on a computational level is well optimized. That is what

we see in a rather complex tissue such as mouse brain, where lipid hashing outperforms antibody hashing. The effect of tissue dissociation on epitope representation can also contribute to this observation in case of antibody hashing. Nevertheless, if working only with for instance brain macrophages (i.e., microglia), the antibody hashing can still be used. The antibody hashing however seems to be a preferred strategy when working with tissues like the spleen or lung where tissue dissociation is more straightforward.

In the present study on human cancer cell lines, the antibody-based hashing when targeting cells overall performed better than the lipid-based hashing. When labeling nuclei from the same cell lines, cholesterol-based hashes (CMO) demonstrated better hashing efficiency compared to the antibody-based labeling (TotalSeq-A). This result can be partially explained by the potential detrimental effect of the lysis buffer (Nuclei EZ Lysis Buffer) used for nuclei isolation on the nuclei surface proteins. The comparison of TotalSeq-A-based cell hashing with TotalSeq-C cell hashing compatible with 3' and 5' gene 10x Genomics sequencing, respectively, demonstrates similarly good hashing efficiency. When comparing TotalSeq-B hashing antibodies with CellPlex (hashing lipids from 10x Genomics) on PBMCs, both methods demonstrate very high hashing accuracy—0.96 for antibodies and 0.93 for CellPlex. Alternatively, other antibody-based hashing methods are also available on the market, e.g., hashing antibodies from BD Biosciences compatible with the BD Rhapsody platform for scRNA-seq experiments.

With respect to analysis algorithms, we observed that the MULTISeqDemux function with `autoTresh = T` parameter originally tailored for lipid-based hashing [10] performs even better than HTODemux function and only slightly better than the GMM-Demux function on a vast majority of tested data sets (Additional file 1: Fig. S12).

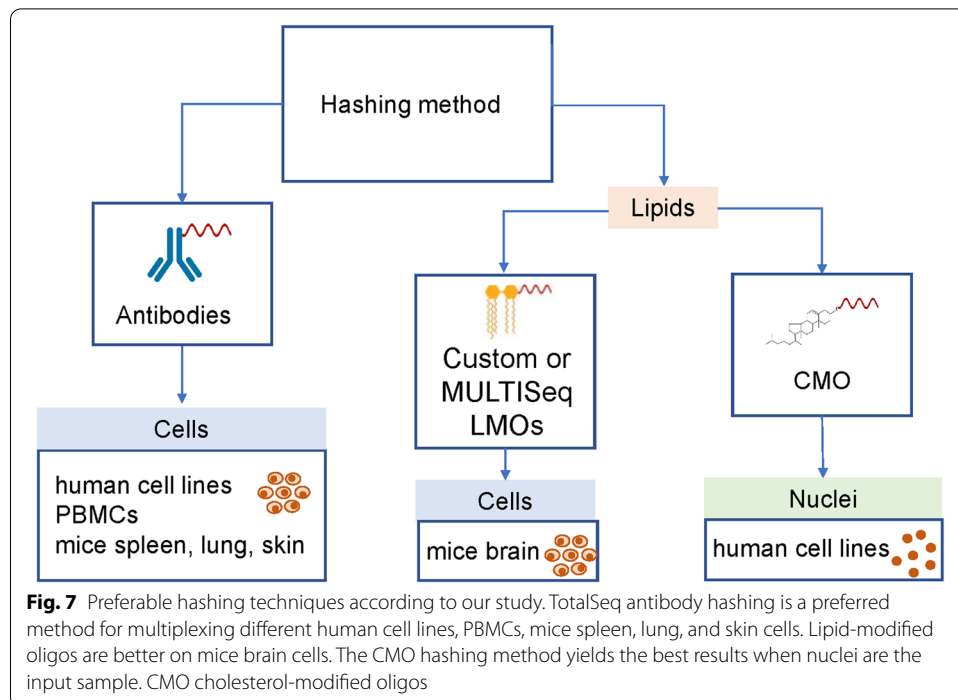
Conclusions

Overall, we can conclude that both lipid- and antibody-based hashing strategies can be successfully applied on human cells, nuclei, and on cells from mouse tissues to reduce costs and batch effects of scRNA-seq experiments. Our comparison indicates that MULTISeqDemux is the preferred demultiplexing function, antibody-based hashing is the most efficient protocol on human cell lines, PBMCs, and lipid hashing delivers the best results on nuclei. On mice tissues, this depends on the tissue, e.g., lipids work better on the brain and antibodies better on the lung and spleen (Fig. 7). We also demonstrate that antibody hashing in combination with large CITE-seq panels can be successfully applied on PBMCs from healthy individuals and SARS-CoV-2 patients.

Methods

Cell culture

MCF7, PC3, DU145, and MDA-MB-231 cells were ordered from ATCC via LGC Standards and maintained according to standard procedures in RPMI-1640 (Gibco, #21875034), F-12K (LGCstandards, ATCC 30-2004), EMEM (LGCstandards, ATCC 30-2003), and DMEM: F-12 (LGCstandards, ATCC 30-2006) medium, respectively, supplemented with 10% fetal bovine serum (Gibco, #10082147) and 1% penicillin/



streptomycin (100 U/ml and 100 µg/ml, respectively, Gibco, #15140122) at 37°C with 5% CO₂. RPMI-1640 medium used to culture the MCF7 cells was additionally supplemented with 0.01 mg/ml human recombinant insulin (Sigma Aldrich, #I3536). The cells had undergone regular tests for mycoplasma contamination during the study.

Single-cell dissociation

To prepare single-cell solutions of the cultured cell lines, the culture medium was removed, and the cells were washed with 1X PBS. Afterwards, the cells were trypsinized (0.05% trypsin-EDTA, Gibco, # 25300054) and pelleted at 200 x g for 5min. The cells were resuspended in 2 ml of culture medium, gently mixed with 8 ml of 1X PBS and put on ice. Afterwards, the cells were pelleted at 200 x g for 5min on 4°C and resuspended in 1 ml 1X PBS with 0.04% BSA. Cells were passed through a 40-µm cell strainer (Corning, # CLS431750-50EA) and counted with a LUNA FL counter, using 1 µl of acridine orange and propidium iodide dye (F23001, Westburg).

Single nuclei dissociation

Single nuclei suspensions were prepared using a modified nuclei isolation protocol from 10x Genomics (demonstrated protocol: isolation of nuclei from single-cell suspensions, #CG000124, revD). We replaced the original lysis buffer with Nuclei EZ Prep buffer (Sigma Aldrich, # NUC101-1KT) with a total lysis time of 5 min and proceeded according to the referenced protocol. Loss of cell viability as a proxy for nuclei extraction quality and the number of nuclei were assessed with a LUNA FL counter, using 1 µl of acridine orange and propidium iodide dye (F23001, Westburg).

Mice primary tissue processing

Brain

For the mice (8 weeks old BALB/cJ) brain cell isolation, the Miltenyi Biotec Neural Tissue Dissociation kit (130-092-628) was used without perfusion according to manufacturer's guideline.

Alternatively the following brain cell isolation protocol was used as described [4]: 8-week-old mice (BALB/cJ, C57BL/6N and DBA/2J) were deeply anesthetized and perfused transcardially with 20 ml of ice-cold phosphate-buffered saline (PBS). Mice were decapitated, and the brains were divided in two hemispheres and collected in 1 mL ice-cold Roswell Park Memorial Institute (RPMI) 1640 medium (Gibco). The brains were cut into small pieces (1–2 mm) after which enzyme mix was added (30 U ml⁻¹ DNase I (Roche), 10 U ml⁻¹ collagenase type I (Worthington), and 400 U ml⁻¹ collagenase type IV (Worthington) diluted in 1X Hanks' buffered salt solution (Gibco)). The brains were incubated for 20 min at 37°C and resuspended with a 1-mL pipet every 10 min. The solutions of three hemispheres per condition were pooled and filtered twice over a 100-µm nylon filter and centrifuged. The pellet was resuspended in 5 ml 70% standard isotonic percoll (SIP, GE Healthcare) diluted in 1X Hanks' buffered salt solution and carefully overlaid with 5 ml of 37% SIP, followed by a 5-ml layer of 30% SIP, forming a three-layered density gradient (centrifuged at 800g, 4°C, 30 min without acceleration/braking). The 70/37% interphase containing immune cells was collected, centrifuged, pooled per condition, and resuspended in fluorescent activated cell sorter (FACS) buffer (2 mM EDTA (Duchefa) and 2% heat-inactivated fetal calf serum (Gibco) dissolved in 1X Hanks' buffered salt solution). The duration of the entire dissociation process was 4 h.

Lung, spleen, and skin

Eight-week-old mice (C57BL6/J and C57BL6/N in the 1st experiment, C57BL6/J and 129SVEvs7 in the 2nd experiment) were used. Single-cell suspensions were prepared by digestion in 100µg/mL Liberase TL (Cat 05401020001, Roche, Germany) and 100µg/mL DNase I (11284932001, Roche, Germany) for 30 min at 37°C before passing through a 100-µm cell strainer. The spleens were rinsed with ice-cold PBS and mashed into small pieces using a plunger of a 10-mL syringe on top of a sterile 100-µm cell strainer mesh in a Petri dish containing 2-ml ice-cold RPMI. Red blood cells in the lung and spleen single-cell suspensions were lysed using ammonium chloride lysis buffer (10 mM KHCO₃, 155 mM NH₄Cl, 0,1 mM EDTA in MilliQ water). Single-cell suspensions were washed with 1x dPBS before incubation with antibodies. The skin cells were isolated as described here [26].

Antibody hashing

Cell and nuclei labeling were performed according to an adapted BioLegend cell hashing protocol (TotalSeqTM-A Antibodies and Cell Hashing with 10x Single Cell 3' Reagent Kit v3 3.1 Protocol, Biolegend). Briefly, cells were then incubated for 10 min with Fc receptor block (PN 422301, TruStain FcX, BioLegend) to block nonspecific antibody binding. Subsequently, cells were incubated with mixtures of barcoded antibodies for 30 min at 4°C. In the TotalSeq-A experiment, MCF7, PC3, DU145, and MDA-MB-231 cells were labeled with 0.1 µg TotalSeq-A0251, -A0252, -A0253, and -A0254, respectively. For

TotalSeq-C hashing, MCF7, PC3, DU145, and MDA-MB-231 cells were labeled with 0.1 µg TotalSeq-C0251, -C0252, -C0253, and -C0255, respectively. For both TotalSeq-A and TotalSeq-C experiments, 200 000 cells of each cell line were labeled with the antibodies in 100 µl PBS containing 1% BSA. After labeling, cells were washed 3 times by resuspension in 1.4 ml PBS containing 1% BSA, followed by centrifugation (200g 5 min at 4°C). After the final wash, cells were resuspended in PBS containing 0.04% BSA and equal amounts of labeled MCF7, PC3, DU145, and MDA-MB-231 cells were pooled before loading onto the 10x Chromium Single-Cell B Chip (PN# 2000060) in case of TotalSeq-A hashing or onto the 10x Chromium chip A (PN# 2000019), because of the TotalSeq-C antibody compatibility with 5' 10x chemistry (Chromium Single Cell Immune Profiling).

Nuclei were labeled with oligo-tagged anti-nucleoporin antibodies (#Mab414, BioLegend) (Sup. Table 1). MCF7, PC3, DU145, and MDA-MB-231 nuclei (840 000 of each line) were labeled with 0.5 µg TotalSeq-A0456, -A0457, -A0458, and -A0459, respectively, in 2% BSA solution with 0.02% Tween-20, 10mM Tris, 146mM NaCl, 1mM CaCl₂, and 21mM MgCl₂ [27]. After the last wash, nuclei were resuspended in the same labeling buffer as above with lower concentration of BSA (1%) and pooled prior to loading on chip B from 10x Genomics.

The DNA library preparation was performed according to the manufacturers guidelines: CG000186 Rev A protocol for 5' 10x Genomics chemistry and CG000185 Rev B protocol for 3' 10x Genomics (v3 chemistry).

Lipid hashing

Based on the MULTI-seq work, Lipid Modified Oligonucleotides (LMO) were used for cell hashing, and Cholesterol Modified Oligos (CMOs) were used for nuclei hashing. Labeling of nuclei was done with CMO because of the inefficacy of the LMOs to label nuclei in the presence of BSA. BSA specifically sequesters LMO and quenches LMO labeling [10].

In this work, we have evaluated and compared the performance of commercially available oligos (custom LMOs purchased from Integrated DNA Technologies) with the LMOs used in the MULTI-seq work [10]. Due to the commercial unavailability, LMOs with lower carbon chain length (C18; Stearyl modification) were used instead of the ligand-modified oligonucleotides (C22), described in the MULTI-seq manuscript.

The MULTI-seq lipid anchor and co-anchor were a kind gift from Christopher S. McGinnis at UCSF. The labeling of MCF7, PC3, DU145, and MDA-MB-231 cells was performed as previously described [10]. Briefly, for each cell line, 500k dissociated cells were washed and resuspended in PBS and incubated with 200 nM 1:1 molar ratio of the anchor and unique barcode oligonucleotides for 5 min on ice. Afterwards, each sample was incubated with 180nM of co-anchor in PBS on ice for 5min, followed by addition of 1% BSA in PBS to reduce off-target labeling when pooling. Cells were washed two times in 1% BSA in PBS after which all samples were pooled, filtered through a 40-µm cell strainer (Corning, # CLS431750-50EA), counted and loaded onto a single reaction of a 10x Chromium chip B.

The anchor and co-anchor of the commercial LMOs are available through Integrated DNA technologies. Also, the anchor and co-anchor of the CMOs were conjugated to cholesterol at the 3' or 5' ends via a triethylene glycol (TEG) linker and are also

commercially available from Integrated DNA Technologies (Additional file 1: Table S4). The labeling workflow for the commercially available LMOs and CMOs are identical to the MULTI-seq protocol [10].

For the mice samples, tissues from different strains were dissociated and 0.4–0.6 M cells were labeled with antibodies (0.5 µg) or lipids (200 nM) as described above. Then cells were washed once (pre-sort labeling protocol) and live cells sorted (BD Aria II, 85 µl nozzle, DAPI staining) for the subsequent cell pooling and 10x Genomics runs. Alternatively, cells were labeled (same amount of antibodies or lipids as above) after sorting (0.3–0.4M cells) and washed 2–3 times before pooling and loading on 10x chips (post-sort labeling).

Sequencing

The generated libraries were pooled targeting 85% mRNA and 15% hashtag oligo (HTO) and paired-end sequenced on individual lanes on a HiSeq4000 or NovaSeq 6000 (Illumina) instruments, using the following read lengths: 28 bp Read1, 8 bp I7 Index, and 91 bp Read2.

PBMC sample preparation and single-cell sequencing

PBMCs from 3 healthy control individuals and 3 SARS-CoV-2 patients (in a framework of the clinical study NCT04326920) were isolated as follows. Whole blood separation was performed by bringing whole blood, diluted with PBS 7.2 (ThermoFisher Scientific, # 20012027), in a Leucosep™ tube, (Greiner Bio-One, # 227290), prefilled with 15 mL Lymphoprep™ (Stemcell technologies, # 07851), and followed by a centrifugation step of 30 min at 400g (acceleration 5, brake 3). After isolation, the PBMCs were washed twice with PBS 7.2 and centrifuged at 350g during 10 min at 4°C. Isolated PBMCs were counted, cryopreserved in 1mL FCS/10%DMSO, and stored in liquid nitrogen. After the thawing as described in the PBMC sample preparation protocol from 10x Genomics (CG00039, Rev C), cells were labeled according to the adapted CITE-seq protocol [15]. In brief, 1×10^6 PBMCs were spun down and resuspended in 25 µl CITEseq antibody cocktail (TotalSeq™-A Antibodies, BioLegend) containing among others one hashing antibody per each sample (0.025µg). After a 30-min incubation on ice, cells were washed 3 times, pooled together (3 healthy donors resulting in sample one and 3 SARS-CoV-2 patients resulting in sample two), and resuspended in PBS+0.04% BSA at a final concentration of 1000 cells/µl. Cells were counted using automated counter LUNA FL™ (Logos Biosystems). Cellular suspensions (target recovery of 10000 cells) were loaded on a 10x Genomics NextGEM Single-Cell Instrument, chip G (# 2000177) to generate single-cell Gel Bead-in-EMulsion (GEMs). Single-cell RNA-Seq libraries were prepared using GemCode Single-Cell V3.1 (NextGEM) 3'Gel Bead and Library Kit (10x Genomics) according to the manufacturer's instructions. Sequencing libraries were sequenced with Illumina NovaSeq S4 flow cell with custom sequencing metrics (single-indexed sequencing run, 28/8/0/98 cycles for R1/i7/i5/R2).

For the CellPlex vs TotalSeq-B evaluation, commercial frozen PBMCs from 3 different healthy donors (SER-PBMC-F, tebu-bio) were used according to the 10x Genomics PBMC sample preparation protocol CG00039 Rev C. 0.5×10^6 PBMCs for CellPlex were labeled according to the 10x Genomics CG00391 Rev A protocol. The same number

of PBMCs were incubated for 10 min with Fc receptor block (PN 422301, TruStain FcX, BioLegend) and then incubated with 0.2 µg of TotalSeq-B0258 or -B0259, or -B0260 for 30 min at 4°C. The cells were washed once and live-sorted on BD Aria II instrument with 85 µm nozzle (DAPI staining). Afterwards, cells were collected and processed according to 10x Genomics CG000390 Rev B protocol. Generated dual index cDNA libraries were sequenced with 10x Genomics dual indexing metrics: 28/10/10/90 cycles for R1/i7/i5/R2 on NovaSeq instrument.

Single-cell data processing

Mapping

The raw sequencing data were demultiplexed and further converted into a single-cell level gene counts matrix with Cell Ranger 3.1.0 with default parameters (<https://github.com/10xGenomics/cellranger>). Only the number of expected cells was adapted in accordance with the number of targeted (captured) cells in each experiment. The mRNA reads were mapped to the human reference genome (assembly hg38 build 95) with the reads allowed to map to both exonic and intronic regions. For every type of hashing strategy, a specific Feature Reference File was generated and used as input for Cell Ranger. Filtered count matrices were used for downstream analysis. CellPlex and TotalSeq-B PBMC samples from dual index libraries were mapped with cellranger version 6.1.1. Mouse samples were mapped to GRCm38 genome.

Freemuxlet-based demultiplexing

The deconvolution of the four cell lines was determined by a genotype-free tool freemuxlet which is an extension of demuxlet [8]. Both tools are from the popsicle (commit 7b141e3) software available at <https://github.com/statgen/popsicle/>. To reduce computation time, the input bam file was sorted to contain only reads that (i) overlap with the SNPs in the VCF file and (ii) have a cell barcode listed in the cell barcode list. Code is available at https://github.com/aertslab/popsicle_helper_tools. The filtered bam file was further used with the popsicle dsc-pileup tool to pileup the reads and corresponding base quality for each overlapping SNPs and each barcode. The reference vcf file was assembled from the 1000 Genomes Project GRCh38 genetic variants data [28] adapted to the index used for mapping and the variant allele frequency > 0.9 and <0.1. were discarded. Next, freemuxlet with default parameters was used to determine sample identity and identify doublets. For the mouse samples VCF file from the “Mouse Genomes Project” was used (<https://www.sanger.ac.uk/data/mouse-genomes-project/>).

Sample demultiplexing and clustering

To assign hashing tags for each cell, two strategies implemented in Seurat 3.1.4 package [29, 30] in R version 3.6.0 were evaluated on the filtered feature-barcode matrix generated by CellRanger: (i) HTODemux function with default parameters and (ii) incorporated into Seurat MULTISEQDemux function [10] with argument autoThresh = TRUE. Subsequent clustering and visualization of the data sets was performed in R using Seurat version 3.1.4 functions as described in the “Demultiplexing with hashtag oligos (HTOs)” vignette available at (https://satijalab.org/seurat/v3.1/hashing_vignette.html). The hashing accuracy was calculated as the overlap between cells annotated

using MULTISeqDemux, HTODemux functions, or GMM-Demux (<https://github.com/CHPGenetics/GMM-Demux>) and freemuxlet as a ground truth.

Doublet identification

Doublet cell detection was performed using DoubletFinder (v2.0.1) [23] with the 10% as expected doublet rate and the following parameters: PCs=1:25, pN=0.25, pK= the highest value on BCMvn plot (find.pK function); Scrublet (v 0.1.3) and freemuxlet with default parameters, and Seurat (MULTISeqDemux (autoTresh=T) and HTODemux (default parameters) functions). The functions incorporated in Seurat predict doublets based on hashtag signal, while other tools use transcriptome data for this.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-022-02628-8>.

Additional file 1. Combination of all supplementary figures (Fig S1 –S12) and tables (table S1-S4).

Additional file 2. Reviewer History.

Acknowledgements

We gratefully acknowledge the help of Gert Van Isterdael with flow cytometry experiments and the analysis, Christopher McGinnis from UCSF for providing the MULTI-seq reagents, Mark Fiers for providing with pre-mRNA genome, Kris Davie for providing us with the vcf files, Jana Roels for help with running CellRanger/freemuxlet, and Junbin Qian for optimising and running demuxlet.

Review history

The review history is available as Additional File 2.

Peer review information

Stephanie McClelland and Barbara Cheifet were the primary editors of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Authors' contributions

V.M. performed the experiments, did the analysis, and wrote the manuscript. J.A. performed the experiments and wrote the manuscript. I.M. did the analysis and wrote the manuscript. S.P., N.V., and I.S. performed the experiments. K.V. and G.H. did the analysis. Y.S., S.A., S.VH., S.P., N.V., H.W., J.R., J.VH., N.V., R.S., and K.M. supervised the work and reviewed the manuscript. All authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

The datasets (raw and processed) supporting the conclusions of this article are available in the European Nucleotide Archive (ENA; <https://www.ebi.ac.uk/arrayexpress/experiments/>) repository, under accession numbers E-MTAB-9964 [31] and E-MTAB-11401 [32]. The SARS-CoV-2 PBMC datasets used in this work currently are not publicly available due to an ongoing clinical study but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The study NCT04326920 was conducted according to EU legislation regarding ethical regulations and involving written patient consents. All experiments were performed in accordance with the ethical committee of the Faculty of Science, UGent, and VIB.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹VIB Tech Watch, VIB Headquarters, Ghent, Belgium. ²Data Mining and Modelling for Biomedicine, VIB Center for Inflammation Research, Ghent, Belgium. ³VIB Center for Brain & Disease Research, Leuven, Belgium. ⁴Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium. ⁵Department of Human Genetics, KU Leuven, Leuven, Belgium. ⁶Myeloid Cell Immunology Lab, VIB Center for Inflammation Research, Brussels, Belgium.

⁷Laboratory for Molecular and Cellular Therapy, Vrije Universiteit Brussel, Brussels, Belgium. ⁸Discovery Sciences, Janssen Research & Development, Pharmaceutical Companies of Johnson & Johnson, Beerse, Belgium.

Received: 24 December 2020 Accepted: 8 February 2022

Published online: 16 February 2022

References

- Aizarani N, Saviano A, Sagar ML, Durand S, Herman JS, et al. A human liver cell atlas reveals heterogeneity and epithelial progenitors. *Nature*. 2019;572:199–204.
- Davie K, Janssens J, Koldere D, De Waegeneer M, Pech U, Kreft L, et al. A single-cell transcriptome atlas of the aging drosophila brain. *Cell*. 2018;174:982–998.e20.
- Han X, Wang R, Zhou Y, Fei L, Sun H, Lai S, et al. Mapping the mouse cell atlas by Microwell-Seq. *Cel*. 2018;172:1091–1107.e17.
- Van Hove H, Martens L, Scheyltjens I, De Vlaminck K, Pombo Antunes AR, De Prijck S, et al. A single-cell atlas of mouse brain macrophages reveals unique transcriptional identities shaped by ontogeny and tissue environment. *Nat Neurosci*. 2019;22:1021–35.
- Lake BB, Ai R, Kaeser GE, Salathia NS, Yung YC, Liu R, et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science*. 2016;352:1586–90.
- Schaum N, Karkaniias J, Neff NF, May AP, Quake SR, Wyss-Coray T, et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature*. 2018;562:367–72.
- Park J-E, Botting RA, Domínguez Conde C, Popescu D-M, Lavaert M, Kunz DJ, et al. A cell atlas of human thymic development defines T cell repertoire formation. *Science*. 2020;367:eaay3224.
- Kang HM, Subramaniam M, Targ S, Nguyen M, Maliskova L, McCarthy E, et al. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat Biotechnol*. 2018;36:89–94.
- Stoeckius M, Zheng S, Houck-Loomis B, Hao S, Yeung BZ, Mauck WM, et al. Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol*. 2018;19:224. <https://doi.org/10.1186/s13059-018-1603-1>.
- McGinnis CS, Patterson DM, Winkler J, Conrad DN, Hein MY, Srivastava V, et al. MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat Methods*. 2019;16:619–26. <https://doi.org/10.1038/s41592-019-0433-8>.
- Gehring J, Park JH, Chen S, Thomson M, Pachter L. Highly multiplexed single-cell RNA-seq for defining cell population and transcriptional spaces. *bioRxiv*. 2018;23:10–1.
- Guo C, Kong W, Kamimoto K, Rivera-Gonzalez GC, Yang X, Kirita Y, et al. CellTag indexing: genetic barcode-based sample multiplexing for single-cell genomics. *Genome Biol*. 2019;20:90. <https://doi.org/10.1186/s13059-019-1699-y>.
- Shahi P, Kim SC, Haliburton JR, Gartner ZJ, Abate AR. Abseq: ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding. *Sci Rep*. 2017;7:1–12. <https://doi.org/10.1038/srep44447>.
- Peterson VM, Zhang KX, Kumar N, Wong J, Li L, Wilson DC, et al. Multiplexed quantification of proteins and transcripts in single cells. *Nat Biotechnol*. 2017;35:936–9.
- Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods*. 2017;14:865–8.
- Garrido F. MHC class-I loss and cancer immune escape; 2019.
- Hermiston ML, Xu Z, Weiss A. CD45: a critical regulator of signaling thresholds in immune cells. *Annu Rev Immunol*. 2003;21:107–37.
- Grindberg RV, Yee-Greenbaum JL, McConnell MJ, Novotny M, O'Shaughnessy AL, Lambert GM, et al. RNA-sequencing from single nuclei. *Proc Natl Acad Sci*. 2013;110:19802–7 Available from: <http://www.pnas.org/content/110/49/19802.abstract>.
- Habib N, Li Y, Heidenreich M, Swiech L. Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. *Science*. 2016;353:925–8. <https://doi.org/10.1126/science.aad7038>.
- Habib N, Avraham-David I, Basu A, Burks T, Shekhar K, Hofree M, et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat Methods*. 2017;14:955–8.
- Xin H, Lian Q, Jiang Y, Luo J, Wang X, Erb C, et al. GMM-Demux: sample demultiplexing, multiplet detection, experiment planning, and novel cell-type verification in single cell sequencing. *Genome Biol*. 2020;21:188. <https://doi.org/10.1186/s13059-020-02084-2>.
- Buus TB, Herrera A, Ivanova E, Mimitou E, Cheng A, Herati RS, et al. Improving oligo-conjugated antibody signal in multimodal single-cell analysis. *Elife*. 2021;10:e61973. <https://doi.org/10.7554/eLife.61973>.
- McGinnis CS, Murrow LM, Gartner ZJ, McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst*. 2019;8:329–37 e4.
- Wolock SL, Lopez R, Klein AM, Wolock SL, Lopez R, Klein AM. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst*. 2019;8:281–91 e9.
- DePasquale EAK, Schnell DJ, Van Camp P-J, Valiente-Alandí Í, Blaxall BC, Grimes HL, et al. DoubletDecon: deconvoluting doublets from single-cell RNA-sequencing data. *Cell Rep*. 2019;29:1718–1727.e8.
- Kashem SW, Kaplan DH. Isolation of murine skin resident and migratory dendritic cells via enzymatic digestion. *Curr Protoc Immunol*. 2018;121:e45. <https://doi.org/10.1002/cpim.45>.
- Gaublomme JT, Li B, McCabe C, Knecht A, Yang Y, Drokhyansky E, et al. Nuclei multiplexing with barcoded antibodies for single-nucleus genomics. *Nat Commun*. 2019;10:1–8.
- Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, et al. A global reference for human genetic variation. *Nature*. 2015;526:68–74. <https://doi.org/10.1038/nature15393>.

29. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol.* 2018;36:411–20.
30. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive integration of single-cell data. *Cell.* 2019;177:1888–1902.e21. <https://doi.org/10.1016/j.cell.2019.05.031>.
31. Mylka V, Matetovici I, Poovathingal S, Aerts J, Vandamme N, Seurinck R, Verstaen K, et al. Hashing (sample multiplexing) of MCF7, PC3, DU145 and MDA-MB-231 cells and nuclei using different hashing methods. *ArrayExpress.* 2021.E-MTAB-9964. <https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-9964/>.
32. Mylka V, Matetovici I, Poovathingal S, Aerts J, Vandamme N, Seurinck R, Verstaen K, et al. Comparison of different hashing methods on human PBMCs and primary mice tissues. *ArrayExpress.* 2022. E-MTAB-11401. <https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-11401/>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

