nature portfolio

Peer Review File

Overloading And unpacKing (OAK) - droplet-based combinatorial indexing for ultra-high throughput single-cell multiomic profiling

Corresponding Author: Dr Spyros Darmanis

This manuscript has been previously reviewed at another journal. This document only contains reviewer comments, rebuttal and decision letters for versions considered at Nature Communications.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors did a commendable job of addressing my reviews from the previous review panel for the earlier journal submission. In particular, their comparison of sensitivity with other similar methods is helpful in placing this method in context, especially when compared to the gold standard of Chromium, the commercially available approach. Although their method does not quite achieve that level of sensitivity, the ability to analyze a significantly larger number of cells is potentially very valuable in many circumstances.

However, the issue of limited novelty remains as similar approaches for dual barcoding have been described previously. Nevertheless, their method appears superior, and I am convinced by their response that the workflow they provide is simpler and more convenient overall.

Overall, I believe that the additions they made in response to all the reviews are good, and that the innovation and value of the paper make it a good fit for Nature Communications. Therefore, I recommend accepting the manuscript for publication in its current form.

Reviewer #2

(Remarks to the Author)
This methodological work introduces a modified version of 10x Genomics protocol to achieve a ultra-high-throughput single-nucleus RNA sequencing (snRNA-Seq) and multiome analysis (snRNA-Seq + snATAC-Seq). The core principle involves overloading a 10X Genomics chip (droplets) with cells (or nuclei) to maximize the use of reagents and barcoding beads. Wet-lab and computational methods are standard and do not show any significant progress or advance. The biological results neither novel nor unexpected and largely recapitulates what is known already. The main innovation of this work relies on tweaking the 10X Genomics protocol to allow for the aliquoting of the post-RT cell nuclei suspension into multiple aliquots (e.g., 20-40 tubes) for subsequent PCR indexing and library construction. In my opinion, a tweak of an established method into a modified workflow should not be considered as an important advance. With all due respect, I do not view this work as an "important advance of significance to specialists within each field," which is one of the primary criteria set by the journal Nature Communications.

I am also skeptical about describing a modified version of an existing 10X protocol (PN-1000095 and PN 220016), as a new technology, such as OAK. It seems that my view is also shared by most recent Nature Methods editorial (doi.org/10.1038/s41592-024-02323-5), which suggested some guidelines for giving a name to a method “… a method needs to be sufficiently novel in both conceptual and technical aspects to justify a new name or acronym.” In my opinion OAK does not fulfill these requirements.

Overall, a more specialized and methodology oriented journal, such as those in the BMC series, would be a more suitable venue for this work. Regrettably, I cannot endorse this manuscript.

Nonetheless, because I went through manuscript in great details, I hope the authors will appreciate my comments and suggestions listed below.

Abstract:

- Abstract needs to clearly mention that the OAK represents a modified version of Chromium Next GEM Single Cell Multiome and scRNA-Seq system / protocol.

Introduction:

- Likewise, in the introduction part you should emphasize to the readers that OAK is an extension / modification of a well-established 10x Genomics protocol(s).

Results:

- Whenever possible be specific.

- Authors should provide the number of “aliquots” that user needs to distribute the cells after barcoding, when using 150k and 450k cells. If a user overloads 10x droplets with 450k cells per 1 lane, then after breaking emulsion and retrieving “cells” (or what is left from them) how many tubes one needs to use to distribute? How user can avoid the same 10X barcode appearing twice in the same tube? My understanding is that at >450.000 cells per 1 Chromium lane user will get over 50% of cells sharing the same 10x barcode at least twice. Please provide exact details how, after aliquoting and sequencing, do you resolve “doublets” of the same type? I understand that cell-doublets of different cell type (e.g. mouse and human) can be filtered out but how do you remove the doublets of the same cell type?

- You state that 20 aliquots are needed for 4000 cells each, but does that mean that for achieving 450k cells one will need >100 tubes?

- Related to above. Please correct me if I am wrong but if 50% of cells share the same 10x barcode then no matter how many aliquots it will be distributed into, it will still be that an aliquot with 50% of cells with the same 10x barcode.

- You should provide experimental evidence indicating how many cells that you eventually consider “singlets” are originating as true singlets (i.e. single cell in a single 10x droplet) vs doublets (two cells in a single droplet) and and mutliplets (multiple cells in a single droplet).

- Do I understand it correctly, that after overloading of the 10x chip, half of the cells (and thus reagents that were consumed to barcode them) are discarded because they clump, or cannot be distinguished as singlets?

- Line 101. What exactly constitutes primary and secondary indices? Its important. Please be precise and avoid jargons. Is it correct that 10x genomics barcode you refer as “primary index” and PCR index you refer to as secondary index?

- Line 104: Please be precise what cells exactly did you use? K-562?

- Line 105: “After sequencing a subset of cells from each experiment, we estimate that 87,864 cells were recovered…” Based on this statement it is not completely clear what “subset” actually represents, and how exactly this number (87,864 cells) have been derived? My understanding is that authors sequenced one aliquot (~4k cells) and then projected / made assumption what the total number of cell could be retrieved in total. The language needs to be more direct, to avoid dubious meanings and confusions. For example, you may say: “ After sequencing a subset of cells from each experiment, we projected that by sequencing all tubes, we could potentially recover 87,864 cells.”

- While I can understand the workflow but based on Poisson relationship I do not get occupancy rates as authors claim it to be. In another example, it is not clear why authors get such a low multiplet rate (10.6%) under conditions where each droplet contains >4 cells (lambda ≥ 4).

- Supplementary Figure 1a. These blobs do not resemble cells to me. Could you please provide additional experimental evidence beyond bright-field images? For instance, fluorescence staining of the cytoplasm, cell membrane, and organelles would be helpful to confirm that these are indeed cells.

- Please show whether OAK is biased towards long genes (i.e. transcripts that are long or unspliiced will be retained more efficiently in methanol-fixed cells).

- Looking at Figure 1C. Many cells appear to be clumping. Could you please explain in the manuscript how do you remove the adhered cells after PCR indexing?

- Line 125. You should provide sequencing saturation values at 15k reads for OAK and for 10x genomics on K562 cells.

- Line 132. I believe the statement in this line supports my earlier notion that authors do not recover cells from 10X Genomics droplets but rather what remains of the cells—namely, the cell nuclei.

I would suggest the authors be more meticulous and precise in their descriptions. How do the authors confirm that they are recovering whole cells and not just nuclei? Once encapsulated in a 10X droplet, non-crosslinked cells are lysed, their membranes disrupted and solubilized by detergents included in the 10X RT kit. Consequently, no intact cells can remain unless they are covalently crosslinked with agents like PFA or glutaraldehyde. What likely remains post-RT step is the cell nuclei, which have a proteinaceous (laminin) shell. Therefore, the claim that methanol-fixed cells are retrieved from 10X droplets after a 53°C incubation for 45 minutes cannot be accurate. What is being retrieved are nuclei.

If the authors disagree with this assessment, I would urge them to demonstrate through fluorescent labeling or other methods that it is indeed whole cells, and not merely nuclei, that are being retrieved from the droplets. This meticulousness is crucial to avoid misrepresentation and misinterpretation of the results. The authors should be aware that in bright-field microscopy, rehydrated cells and nuclei can appear similar.

- You should be more precise how many cells are being lost following RT step in 10X droplets? Are these 50% and 59% that you describe in Figure 1c? It was not very clear.

- Related to above comments. You should consider providing more details of identifying multiplet rate that comes from the same species (e.g. ≥2 K562 cells in a droplet).

- K-562 cells is relatively homogeneous and are rich in mRNA, which means that this cell line is not the best example of benchmarking the technology. User need to be certain that certain what biases OAK introduced when working with heterogeneous samples. As such I would recommend sequencing PBMCs and comparing it to 10x results

- Authors provide Figure S1l showing in vitro differentiated bronchial airway cells (Line 154) , but some expected cell types (e.g. brush, ionocytes, PNECs, etc) are missing. Its not clear why, and might indicate the problems with the methodology or the method. Again, the use of primary PBMC, fresh or frozen, would be a much more valuable content about the capabilities of the technique.

- Line 157. Authors are using hashing antibodies to stain cells and later fix them in methanol. It is not clear how it is possible that Ab-stained cells can undergo methanol fixation without losing antibodies. I find it hard to believe that Hashtag antibodies remain bound to their targets following methanol fixation.

- I find it misleading when authors sequence a small fraction of cells and then claim high cell recovery rates. For instance, while authors sequenced 8,096 cells, they assert a recovery of 44,582 cells. Although I understand the rationale behind such claims, authors should strictly adhere to the facts. There is no conclusive evidence to guarantee uniform recovery efficiency across tubes, as the library preparation for each tube can be subject to technical artifacts such as cell count discrepancies, contamination, and doublets. It is essential for the integrity of the research that these factors are acknowledged and transparently addressed in the methodology and results discussion.

- Likewise in Line 330 I find statement that OAK combined “combinatorial indexing” to be misleading. Authors simply adopt 10X genomics protocol without involving split-and-pool steps.

- Line 485. At concentration 2,400 nuclei / µL a significant degree of clumping can be expected. You may want to provide further details of the sample quality at such concentration.

- Line 496: Please be precise, how many tubes were used to transpose 150-200k nuclei? Do I understand correctly that you used 20 tubes to perform taggmentataion, and volume of each reaction is 15 µL? After taggmentation the aliquots in these tubes were pooled and loaded on 1 Chromium lane?

Line 500: Please indicate how many nuclei are resuspended in “15 µl of solution” before loading 1 channel?

Line 505-512: These are not cells, these are nuclei. Please fix.

Reviewer #3

(Remarks to the Author)
I had reviewed the manuscript for another journal. The authors have carefully and satisfactory addressed our questions. The results on PBMC are a bit worrying, but the authors have addressed this limitation in the discussion.

Version 1:

Reviewer comments:

Reviewer #2

(Remarks to the Author)

1. I believe I have a reasonable understanding of the OAK technology. My primary concern is that this work lacks conceptual novelty. Barcoding (Indexing) in droplets, barcoding (indexing) in-situ, and in plates has all been previously demonstrated, as has combining both methods (dsciATAC-seq, scifi-RNA-seq, SCITO-seq, etc). The improved sensitivity claimed is not due to the development of new biochemistry or tools by OAK, but because it utilized reagents and kit developed by 10X Genomics. The authors simply tweaked the existing 10X Genomics protocol; were rather than inactivating the RT enzyme at 85ºC for 15 minutes, they collected "cells" (or what remains of them after 42ºC for 60 minutes in the presence of lysis reagents) and then aliquoted samples into multiple tubes to perform a 2nd indexing step in PCR tubes, as depicted schematically in Figure 1.

Frankly, I find it challenging to identify conceptual or technological novelty here. With all due respect, my opinion remains unchanged: merely adapting an existing commercial reagents kit to perform a barcoding (first indexing) reaction and then proceeding to second indexing in plates (or tubes) does not constitute an advance in the field, especially considering an array of methods based on a similar concept such as dsciATAC-seq, scifi-RNA-seq, SCITO-seq, etc.

2. OK

3. OK

4. OK

5. I agree with the response and appreciate the reference to Figure S1b; it's helpful. Given that both the first indexing (in Chromium) and the second indexing (in PCR tubes, plates) are of equal importance, it would be beneficial to include both panels—Figure 1b and Figure S1b—in the main Figure. Meanwhile, Figure 1c could be relocated to the supplementary materials.

Line 558 "Multiplet rate theoretical estimation":
I agree with using the birthday problem to estimate the collision rate, but the current description is still somewhat unclear. It would be beneficial to delineate the collision rate estimation for the first indexing (in droplets) separately from the second indexing (in tubes). Could you clarify the relationships used? I agree that for the 1st indexing one can use brith problem as described by Ma et al., 2020, Cell 183, 1103–1116, with N = 150k or 450k, whereas D = 100k. This would result in collision rate being 48% or 78% (as indicated in Figure S1b).

5.1. Further Inquiry:
What formula should be applied to estimate the barcode collision rate for the second step of indexing? What values for N and D should be use? At this step, are you assuming that at the 2nd step of indexing you start with 450k cells and that each cell in a broken emulsion has a unique index 1? Expanding your Methods section and stating the assumptions (if they are made) would be helpful for potential users. Next, what relationship one should use to estimate barcode collision rate for the 2nd step indexing.

6. OK

7. It seems there might be a misunderstanding regarding the response about the mean number of cells sharing the same barcode. When you mention that the mean number of cells sharing the same 10X barcode is 1.4, it's important to clarify what this statistic signifies in the context of cell loading and barcode distribution. At a loading of 150k cells per Chromium lane with 100k droplets, approximately 48% of barcoded cells will carry a 10X barcode (the first index) that appears at least twice within the same aliquot. This scenario, which I previously referred to as cells having the same 10X barcode, suggests a significant overlap in barcode assignment due to the limited number of droplets with gelbeads compared to the number of cells.

8. OK

9. OK

10. OK

11. OK

12. OK

13. OK

14. I believe the rebuttal to my earlier critique may not fully address the core of my argument. I agree that cells preserved in methanol should indeed be referred to as cells, given they retain the plasma membrane, organelles, cytoplasm, and other cellular structures. My concern, however, is that the OAK protocol does not actually retrieve cells from the droplets following the RT reaction. Given the presence of detergents and elevated temperatures, cells will lose their plasma membranes, cytoplasm content leaving predominantly the nucleus intact, as harsher conditions (e.g., SDS, proteinase K) are required to disrupt it.

The authors counter this by stating:

Nevertheless, these "cells" are still distinct from nuclei. For example, hashing antibody-derived tags targeting the cell surface membrane are preserved in our data, as demonstrated in Fig 1h.

However, this does not necessarily validate the retrieval of cells post-RT reaction. The hashing antibody-derived tags targeting the cell surface membrane, as shown in Fig. 1h, are indeed preserved, but it's important to note that these antibodies were used prior to encapsulation when cells are still intact. Upon cell lysis in droplets, some leftovers of plasma membrane and endoplasmic reticulum are known to remain associated with cell nucleus; thus it is not very surprising that hashing-Ab information was recovered; albeit the statistics of number of reads per hashing-Ab are not shown.

15. I may not have been clear in my initial request. I was asking to evaluate the scRNA-seq reads' coverage across the transcript length of the gene body, effectively assessing transcript recovery as gene-body coverage. The purpose of this analysis is to differentiate between transcripts retained in cells—which would typically represent both short and long genes, including those crucial for physiological responses such as signaling proteins—and transcripts captured primarily in nuclei, which are often long, unspliced and carry limited biological information.

For this analysis, I suggest using the ReSQC v5.0.1 function geneBody_coverage.py to provide a detailed overview. Comparing your data with standard 10x Genomics protocols (v3 or v3.1) will clarify whether the OAK mainly captures mRNAs retained in the nucleus, potentially losing all cytoplasmic RNA. Such a finding would suggest that describing OAK as a technique for single-cell RNA-Seq is inaccurate; particularly for non-experts in scRNA-Seq techniques.

Technically, OAK is a two-step process where the second step (second indexing) is critical as it depends on the capture of mRNA predominantly from nuclei that survive the RT reaction. I recommend emphasizing this distinction in your manuscript to avoid misconceptions and misunderstandings among readers less familiar with the diverse array of RNA-Seq techniques. While these details might seem trivial to experienced users, less experienced readers might struggle to grasp all critical aspects of the method before employing it. Having said there is another concern that authors should address, that is, data quality:

After the first indexing and subsequent pooling of cells, there is a high possibility that barcoded-cDNA molecules not retained within cells will freely diffuse and nonspecifically bind to any random cell. This scenario could substantially increase the noise in the data and complicate data analysis; particularly when overloading the Chromium chips.

To better understand the impact of this phenomenon on data integrity, could you please provide some graphical representations in the supplementary materials? Specifically, seeing how clean the OAK RNA-Seq data is compared to standard 10X Genomics conditions. These visualizations will help clarify the extent of noise and potential cross-contamination between samples, which is crucial for assessing the reliability and quality of the data obtained through this method.

Two Figures such as 1) histogram of UMI density distribution (frequency vs UMI counts, refer to Figure 6.1) and „knee“ point (refer to Figure 5.9.1) might be sufficient.

Refer to for details https://biocellgen-public.svi.edu.au/mig_2019_scrnaseq-workshop/index.html).

16. OK

17. Could you explicitly detail that in the text, rather than requiring readers to infer from the figures, the differences in performance between the two techniques at the same sequencing depth (e.g. 15k reads per cell). It's important for potential users to understand that at this depth, there is nearly a threefold difference in saturation levels (23% vs. 65%) and a significant reduction in UMI recovery by half when using OAK compared to the standard methods. This observation indirectly supports the earlier assertion that OAK primarily profiles RNA derived from single nuclei rather than from cells. Providing this information explicitly in the manuscript will help readers, especially less experienced users, in understanding the method's limitations and capabilities without having to eyeball the data from the figures.

17.1 (new request) To address my concern above, please perform gene enrichment analysis and displaying gene programs that are being depleted or enriched in OAK data set vs 10x Genomics data set.

18. My concern has not been adddressed but I hope that my comments above makes now more clearly why I believe that OAK is profiling RNA of cell nucleus, and not whole cells. The fact that it cannot handle PBMCs indicates that OAK is not a single-cell RNA-Seq technique.

19. See #18.

20. OK

21. OK

22. OK

23. Great. Thanks.

24. Please review the Materials and Methods section to ensure the language is clear and consistent. As currently written, there seems to be conflicting statements around the use of hashed antibodies. The manuscript says that hashing with antibodies occurs before methanol fixation (Lines 402-409), describing the treatment of live cells that are subsequently sorted. However, in line 161 of the main text, it is stated that "cells were fixed in methanol after staining," implying a different sequence of events. Furthermore, in the rebuttal letter, the authors mention that hashed cells were used after methanol fixation, stating, "... we are not the only study that has successfully performed antibody staining followed by methanol fixation." This statement is at odds with the methods described and seems to misinterpret prior work. For example, Hwang et al., (ref #3) utilized PFA-fixed cells, not methanol-fixed cells, to profile surface proteins.

25. OK

26. OK

27. OK

28. OK

29. OK

30. OK

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

The authors did a commendable job of addressing my reviews from the previous review panel for the earlier journal submission. In particular, their comparison of sensitivity with other similar methods is helpful in placing this method in context, especially when compared to the gold standard of Chromium, the commercially available approach. Although their method does not quite achieve that level of sensitivity, the ability to analyze a significantly larger number of cells is potentially very valuable in many circumstances.

However, the issue of limited novelty remains as similar approaches for dual barcoding have been described previously. Nevertheless, their method appears superior, and I am convinced by their response that the workflow they provide is simpler and more convenient overall.

Overall, I believe that the additions they made in response to all the reviews are good, and that the innovation and value of the paper make it a good fit for Nature Communications. Therefore, I recommend accepting the manuscript for publication in its current form.

We thank the reviewer for their comment and their recommendation that the manuscript is accepted for publication.

Reviewer #2 (Remarks to the Author):

1
This methodological work introduces a modified version of 10x Genomics protocol to achieve a ultra-high-throughput single-nucleus RNA sequencing (snRNA-Seq) and multiome analysis (snRNA-Seq + snATAC-Seq). The core principle involves overloading a 10X Genomics chip (droplets) with cells (or nuclei) to maximize the use of reagents and barcoding beads. Wet-lab and computational methods are standard and do not show any significant progress or advance. The biological results neither novel nor unexpected and largely recapitulates what is known already. The main innovation of this work relies on tweaking the 10X Genomics protocol to allow for the aliquoting of the post-RT cell nuclei suspension into multiple aliquots (e.g., 20-40 tubes) for subsequent PCR indexing and library construction. In my opinion, a tweak of an established method into a modified workflow should not be considered as an important advance. With all due respect, I do not view this work as an "important advance of significance to specialists within each field," which is one of the primary criteria set by the journal Nature Communications.

I am also skeptical about describing a modified version of an existing 10X protocol (PN-1000095 and PN 220016), as a new technology, such as OAK. It seems that my view is also shared by most recent Nature Methods editorial (doi.org/10.1038/s41592-024-02323-5), which suggested some guidelines for giving a name to a method “… a method needs to be sufficiently novel in both conceptual and technical aspects to justify a new name or acronym.” In my opinion OAK does not fulfill these requirements.

Overall, a more specialized and methodology oriented journal, such as those in the BMC series, would be a more suitable venue for this work. Regrettably, I cannot endorse this manuscript.

Nonetheless, because I went through manuscript in great details, I hope the authors will appreciate my comments and suggestions listed below.

We appreciate the reviewer's thorough examination of our manuscript and the valuable comments. We would like to address their skepticism by clarifying further the key principle of OAK and how it surpasses the benchmarks for novelty and performance set by other recent publications in Nature journals.

While the reviewer correctly noted our use of 10x Genomics' Chromium platform, several comments, particularly comment #26

> "*Likewise in Line 330 I find statement that OAK combined "combinatorial indexing" to be misleading. Authors simply adopt 10X genomics protocol without involving split-and-pool steps.*"

suggest that the reviewer may not have fully recognized the combinatorial indexing nature of the OAK method. This could be due to the novel approach of using droplets, rather than a well plate, for compartmentalizing the initial indexing reactions. We will address this in greater detail in our responses to the subsequent comments. To summarize, instead of employing the droplet barcoding system in a conventional manner, a key feature of OAK is its utilization of the extensive, yet traditionally underutilized, barcoding capacity to perform the initial splitting step in a combinatorial indexing experiment. Unlike traditional combinatorial indexing, which involves multiple rounds of pipetting into micro-well plates, this new strategy leverages microfluidic systems for splitting and the chemical quality of emulsions for pooling, thereby simplifying the experimental process.

Due to the novel approach of performing split-and-pool, this method presents more than a mere modification of an established technique. Instead, it contributes to a rapidly evolving field that combines droplets and combinatorial indexing strategies in single-cell sequencing, as pioneered by dsciATAC-seq [1], scifi-RNA-seq [2], and SCITO-seq [3]. Each of these methods has introduced its own acronym and has been well received. We consider OAK a significant advancement in this area, offering higher detection sensitivity, broader compatibility with molecular modules, and a more straightforward protocol compared to other combinatorial indexing methods. Thus, the naming of our method is appropriate, beneficial, and consistent with the field.

Lastly, conceptually, 10x genomics' Chromium platform is not the only droplet system that can be integrated into the OAK strategy, as discussed in the manuscript. However, it is among the most accessible to researchers in the field, many of whom lack the resources to build their own droplet-generating platform or manufacture barcoding gel beads. Therefore, providing proof of

concept on this platform is advantageous for the adoption of this method within the wider research community.

2
Abstract:

- Abstract needs to clearly mention that the OAK represents a modified version of Chromium Next GEM Single Cell Multiome and scRNA-Seq system / protocol.

Thank you for the comment. As we described in our response to comment #1, OAK is more than a modified version of an established method. However, we acknowledge the opportunity in the abstract to emphasize a key innovation of our approach. In the revised manuscript, we have included that OAK "leverages a droplet-based barcoding system for the initial compartmentalization in combinatorial indexing".

3
Introduction:

- Likewise, in the introduction part you should emphasize to the readers that OAK is an extension / modification of a well-established 10x Genomics protocol(s).

In the revised manuscript, we have emphasized in the introduction that OAK "utilizes the Chromium microfluidic system to replace micro-well plates in the first step of split-and-pool for combinatorial indexing".

Results:

4
- Whenever possible be specific.

We sincerely appreciate your comment, and have provided more specific information in our revised manuscript.

5
- Authors should provide the number of “aliquots” that user needs to distribute the cells after barcoding, when using 150k and 450k cells. If a user overloads 10x droplets with 450k cells per 1 lane, then after breaking emulsion and retrieving “cells” (or what is left from them) how many tubes one needs to use to distribute? How user can avoid the same 10X barcode appearing twice in the same tube? My understanding is that at >450.000 cells per 1 Chromium lane user will get over 50% of cells sharing the same 10x barcode at least twice. Please provide exact

details how, after aliquoting and sequencing, do you resolve “doublets” of the same type? I understand that cell-doublets of different cell type (e.g. mouse and human) can be filtered out but how do you remove the doublets of the same cell type?

The information regarding the number of aliquots needed is provided in multiple forms and sections within the manuscript.

> First, Supplementary Figure 1b illustrates the relationship between the number of aliquots and the collision rate at two distinct cell inputs. This allows users to determine their preferred number of aliquots based on their desired collision rate. Second, the curves in Supplementary Figure 1b are derived from the closed-form solution for the expected number of collisions in the birthday paradox, as described in the Methods section. Therefore, users can also utilize the birthday paradox solution provided to determine the appropriate number of aliquots for any desired loading rate and collision rate. Thirdly, for simplicity, in the Methods section, we recommended "20 aliquots per 150,000 cells loaded" to guide users towards achieving high sensitivity and a low collision rate. Lastly, for each loading presented in the manuscript, the aliquot numbers were detailed in Supplementary Table 1 as well as in the Methods section.

We appreciate the reviewer's observation that "at >450.000 cells per 1 Chromium lane user will get over 50% of cells sharing the same 10x barcode at least twice". However, it is important to clarify that in the context of combinatorial indexing, these are not true "doublets". We realize that part of this comment, along with several subsequent comments (including comment #8, 9, 13, 16, 21, 26) may stem from misunderstanding the combinatorial indexing nature of OAK. Thus, we would like to take this opportunity to clarify the concept of combinatorial indexing and reiterate the principle of OAK.

> In combinatorial indexing [4,5], none of the barcodes are unique to any single cell. Instead, barcodes are integrated in each round of split-and-pool to form a unique barcode combination for each cell. Such barcode combination reflects the unique path each cell has taken in the process, and is used to identify each single cell.
>
> In the context of OAK, the overloading step in microfluidic droplets constitutes the first round of "split", the unpacking (emulsion breaking) step represents the "pool", and the aliquoting step makes up the final "split" of the process. Even though cells in the same droplet will share the same primary index (10x barcode), their diverging paths into different aliquots for the second indexing create different combinations of primary and secondary index for each individual cell. Thus, the droplet barcode, in combination with the aliquot barcode, is used to resolve cDNA molecules that belong to individual single cells.
>
> Consequently, true doublets are only those cells that have taken the exact same path through the split-and-pool process and share both the droplet barcode (primary index) and the aliquot barcode (second index). The occurrence of these events, usually referred to as “collisions”, is estimated by the closed-form solution for the expected

number of collisions in the birthday paradox as mentioned above, and experimentally assessed by the species-mixing experiment in Figure 1. The purpose of this experiment is not to filter out doublets after sequencing, but to assess the single-cell resolving capacity of the method. Please also note that in the manuscript, we reported both the observed multiplet rates (combinatorial barcode collision between mouse and human) as well as the overall multiplet rates. The latter extrapolates the unobservable multiplets (Human-Human and Mouse-Mouse), and is calculated by doubling the observable multiplets.

To summarize, the presence of identical 10x barcodes is an intentional feature rather than an error, and does not result in cell doublets. In methodologies utilizing combinatorial indexing, such as OAK, doublets are defined as cells sharing the same combination of barcodes. The extent of doublet occurrence in OAK has been rigorously evaluated through the barnyard experiment, which involves a mixture of mouse and human cells and is presented in Figure 1.

6

- You state that 20 aliquots are needed for 4000 cells each, but does that mean that for achieving 450k cells one will need >100 tubes?

No. With 450k cells loaded and 223,680 cells projected to be recovered, the number of aliquots required is 56 instead of >100, to meet the target. These 56 aliquots do not have to be in individually separate tubes; they can be in PCR Strip Tubes, as used in our experiments, or even on a 96-well plate if preferred. To put this number of aliquots in context, a plate-based traditional combinatorial indexing method for this number of cells would require three rounds of split-and-pool, utilizing 312 microwells (96*3+24) [5].

7

Related to above. Please correct me if I am wrong but if 50% of cells share the same 10x barcode then no matter how many aliquots it will be distributed into, it will still be that an aliquot with 50% of cells with the same 10x barcode.

We do not have 50% of cells sharing the same 10x barcode. The barcoding capacity of the Chromium platform exceeds 750,000 [6], resulting in very few cells that share a 10x barcode in the initial "split" step. As presented in Fig 1c, the mean number of cells sharing the same 10x barcode is 1.4 when 150k cells are loaded, and 4.4 when 450k cells are loaded. These correspond to 0.00093% and 0.00098%, respectively.

8
- You should provide experimental evidence indicating how many cells that you eventually consider “singlets” are originating as true singlets (i.e. single cell in a single 10x droplet) vs doublets (two cells in a single droplet) and and mutliplets (multiple cells in a single droplet).

Figure 1b presents the percentage of cell-free, single-cell, and multi-cell droplets across a spectrum of loading conditions. Fig 1c shows the mean number of cells that share the same droplet barcode in the initial "split" step. We have detailed how OAK utilizes combinatorial indexing to resolve single cells even from multi-cell droplets in our response to comment #5. In summary, with combinatorial indexing, sharing the same droplet barcode (primary index) does not make cells doublets. It is the combination of the primary and secondary indices that is used to identify individual single cells.

9
- Do I understand it correctly, that after overloading of the 10x chip, half of the cells (and thus reagents that were consumed to barcode them) are discarded because they clump, or cannot be distinguished as singlets?

No. The 50% loss is a result of both the microfluidic chip in use and the split-and-pool steps. The Chromium microfluidics chip alone contributes to a 40% loss, as indicated in the manual of the Chromium products. The remaining 10% can be attributed to cell loss in the tubes and pipette tips. Overall, OAK outperforms existing ultra-high-throughput methods that have reported recovery rates, as presented in Supplementary Fig.1c.

10
- Line 101. What exactly constitutes primary and secondary indices? Its important. Please be precise and avoid jargons. Is it correct that 10x genomics barcode you refer as “primary index” and PCR index you refer to as secondary index?

Yes, this is exactly correct. Thank you for the comment. We have edited the sentence in the revised manuscript to clarify that the primary index refers to the barcode coming from the droplets, while the second index refers to the one integrated within each aliquot. Figure 1a also conveyed this information.

11
- Line 104: Please be precise what cells exactly did you use? K-562?

Thank you for the question. In Line 111 (now Line 113 of the revised manuscript) we said that "The input cells consisted of a 1:1 mixture of a mouse (NIH/3T3) and a human (K562) cell line".

12
- Line 105: “After sequencing a subset of cells from each experiment, we estimate that 87,864 cells were recovered…” Based on this statement it is not completely clear what “subset” actually represents, and how exactly this number (87,864 cells) have been derived? My understanding is that authors sequenced one aliquot (~4k cells) and then projected / made assumption what the total number of cell could be retrieved in total. The language needs to be more direct, to avoid dubious meanings and confusions. For example, you may say: “ After sequencing a subset of cells from each experiment, we projected that by sequencing all tubes, we could potentially recover 87,864 cells.”

Thank you for the suggested edit. We have incorporated the changes into the revised manuscript.

13
- While I can understand the workflow but based on Poisson relationship I do not get occupancy rates as authors claim it to be. In another example, it is not clear why authors get such a low multiplet rate (10.6%) under conditions where each droplet contains >4 cells (lambda ≥ 4).

Based on the Poisson distribution, the average number of cells per droplet is equal to lambda. Therefore, when 150k cells are loaded with the Chromium system generating 100,000 droplets, lambda=150k/100k=1.5, which closely aligns with our observed occupancy of 1.4 as presented in Fig 1c. Similarly, when 450k cells are loaded, lambda=450k/100k=4.5, which also closely aligns with our experimental observation of 4.4, as presented in Fig 1c.

Regarding the multiplet rate, while the Poisson relationship governs the very first step of the OAK process, we would like to emphasize that the following steps are also critical components of our combinatorial indexing approach. Therefore, the multiplet rate is a result of the entire split-and-pool process, as detailed in our response to comment #5 and #8.

14
- Supplementary Figure 1a. These blobs do not resemble cells to me. Could you please provide additional experimental evidence beyond bright-field images? For instance, fluorescence staining of the cytoplasm, cell membrane, and organelles would be helpful to confirm that these are indeed cells.

Thank you very much for your question. We believe this question is related to comment #18 and reiterated in comment #19:

> "*I would suggest the authors be more meticulous and precise in their descriptions. How do the authors confirm that they are recovering whole cells and not just nuclei? Once encapsulated in a 10X droplet, non-crosslinked cells are lysed, their membranes disrupted and solubilized by detergents included in the 10X RT kit. Consequently, no*

intact cells can remain unless they are covalently crosslinked with agents like PFA or glutaraldehyde. What likely remains post-RT step is the cell nuclei, which have a proteinaceous (laminin) shell. Therefore, the claim that methanol-fixed cells are retrieved from 10X droplets after a 53°C incubation for 45 minutes cannot be accurate. What is being retrieved are nuclei.

If the authors disagree with this assessment, I would urge them to demonstrate through fluorescent labeling or other methods that it is indeed whole cells, and not merely nuclei, that are being retrieved from the droplets. This meticulousness is crucial to avoid misrepresentation and misinterpretation of the results. The authors should be aware that in bright-field microscopy, rehydrated cells and nuclei can appear similar."

All of the comments and questions in #14, #18,and #19 are excellent points, which we would like to address comprehensively.

We completely agree that due to fixation and permeabilization steps, the "cells" would have lost a substantial part of their cytoplasmic contents, including transcripts and organelles. This is evidenced by a higher intronic read ratio and a lower mitochondria read ratio presented in the manuscript, as you have precisely pointed out in comment #18. Both of these observations have also been reported by several other studies using combinatorial indexing methods [7–9].

Nevertheless, these "cells" are still distinct from nuclei. For example, hashing antibody-derived tags targeting the cell surface membrane are preserved in our data, as demonstrated in Fig 1h. The hashtag antibody clones used are specific against cell surface membrane proteins CD298 (clone LNH-94) and β2 microglobulin (clone 2M2). In addition to their use in sequencing assays [10], these clones have been demonstrated to stain the cell surface in live cell mass cytometry [11]. Moreover, other studies also have shown that cells after methanol fixation still retain signals from antibodies that target the cell surface membrane [12,13]. SCITO-seq [3] even took advantage of this, and built a single-cell combinatorial indexing method to profile cell surface proteins for these "cells". All this evidence suggests that these "cells" are not equivalent to "nuclei".

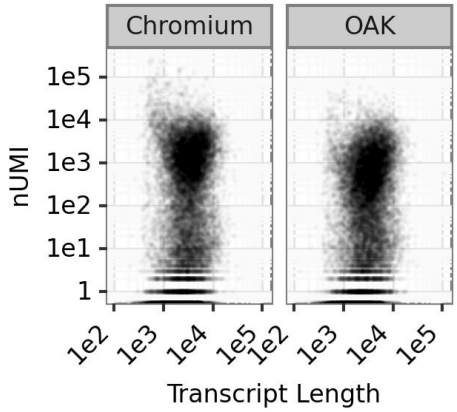
When cells instead of extracted nuclei undergo fixation and subsequent combinatorial indexing, the term "cells" is used throughout the process, and the method is deemed a single-cell sequencing method in the literature [2,4,5,8]. In contrast, when nucleus extraction is performed before fixation and combinatorial indexing, the term "nuclei" is used throughout the process, and the method is deemed a single-nucleus (sn) RNA-seq assay [5]. We support this common practice in the field, as it helps distinguish two broad categories of assays in which cells and nuclei are used as inputs, respectively. Thus, in our manuscript, we kept our terminology consistent with the literature in the field. Specifically, for the species-mixing experiment, cell-hashing experiment, and drug-treatment experiment, "cells" are the subjects of our description. In contrast, in the human retina profiling example, since nuclei were experimentally extracted, the profiling unit became the "nuclei," and we referred to the process as "snRNA-Seq and snATAC-Seq."

In summary, we agree with the reviewer that after exposure to methanol and detergents,cells can be compromised with incomplete cytoplasm, membrane, or organelles. However, they still retain cell membrane features and hence are not the same as nuclei. More importantly, in our manuscript, we refer to cells and nuclei as such to stay consistent with the common practice in the field.

15

- Please show whether OAK is biased towards long genes (i.e. transcripts that are long or unspliiced will be retained more efficiently in methanol-fixed cells).

We appreciate the reviewer's concern, and have therefore analyzed the relation between transcript detection and length as shown below. Weak correlation was observed in both standard Chromium (spearman corr = 0.20) and OAK data (spearman corr = 0.26). While the difference is minor, the divergence in Spearman correlations (0.20 vs 0.26) could originate from the fixation and permeabilization processes required in combinatorial barcoding. These processes render the data more similar to single-nucleus RNA-seq, where an overrepresentation of longer transcripts is observed. This phenomenon has been corroborated by other studies [9,14,15].



16

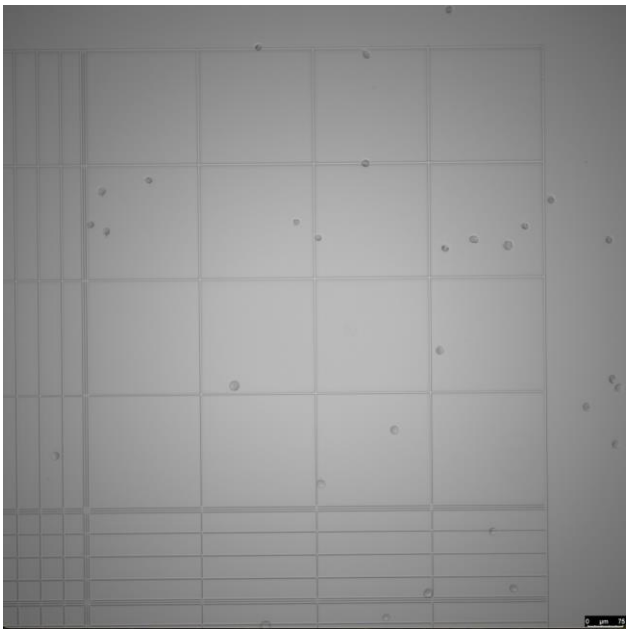
- Looking at Figure 1C. Many cells appear to be clumping. Could you please explain in the manuscript how do you remove the adhered cells after PCR indexing?

Thank you for the question. The cells in Fig 1C are not actually clumping but are multiple cells residing close to each other within droplets of about 120 µm across. This has been consistently observed in other studies that create multi-cell droplets [2,16].

While the cells might appear closely packed in the multi-cell droplets, all cells are released from the droplets during the unpacking step (emulsion breaking). At this step, microscopic images were taken (e.g., Supplementary Fig 1a; the uncropped and zoomed-out image is attached below) to ensure that cells or nuclei have been released and dissociated before splitting into

aliquots for PCR indexing. Cells are destroyed after this round of PCR indexing due to repetitive heat up to 95℃.

Last but not least, if the cells had remained adhered and thus taken the exact same path through the split-and-pool process, the doublet rate would have approached 100% in the experiment where 450,000 cells spanning two species were loaded, with each droplet on average containing 4.4 cells. Our low doublet rates presented in the manuscript further confirm that the cells initially in the same droplet were subsequently split into different aliquots during the split-and-pool process.



17
- Line 125. You should provide sequencing saturation values at 15k reads for OAK and for 10x genomics on K562 cells.

Thank you for the comment. We used Supplementary Figure 1j to provide comprehensive information on sequencing saturation across a spectrum of sequencing depths. Since sequencing saturation is calculated as 1 - (n_UMIs / n_reads), for example, at 15k reads, standard Chromium method returns a mean of 11,545 UMIs, which translates to 23% saturation. For OAK, the mean UMI number is 5,255, indicating a saturation of 65%. For context, as presented on the same graph, at the same sequencing depth, scifi-RNA-seq detects as few as 425 UMIs, translating to a saturation of 97%.

18
- Line 132. I believe the statement in this line supports my earlier notion that authors do not recover cells from 10X Genomics droplets but rather what remains of the cells—namely, the cell nuclei.

Thank you for the comment. We have addressed this comment in greater detail in our response to comment #14. In short, we agree that after exposure to methanol and detergents the cells lose some cytoplasm, membrane, and organelles. However, they still retain some cell membrane features and hence are not the same as nuclei. More importantly, in our manuscript, we refer to cells and nuclei as such to stay consistent with common practice in the field.

19
I would suggest the authors be more meticulous and precise in their descriptions. How do the authors confirm that they are recovering whole cells and not just nuclei? Once encapsulated in a 10X droplet, non-crosslinked cells are lysed, their membranes disrupted and solubilized by detergents included in the 10X RT kit. Consequently, no intact cells can remain unless they are covalently crosslinked with agents like PFA or glutaraldehyde. What likely remains post-RT step is the cell nuclei, which have a proteinaceous (laminin) shell. Therefore, the claim that methanol-fixed cells are retrieved from 10X droplets after a 53°C incubation for 45 minutes cannot be accurate. What is being retrieved are nuclei.

If the authors disagree with this assessment, I would urge them to demonstrate through fluorescent labeling or other methods that it is indeed whole cells, and not merely nuclei, that are being retrieved from the droplets. This meticulousness is crucial to avoid misrepresentation and misinterpretation of the results. The authors should be aware that in bright-field microscopy, rehydrated cells and nuclei can appear similar.

Thank you for raising this concern. We agree with your assessment, but for the reasons described in our response to comment #14, we believe it is appropriate to refer to cells and nuclei as such to stay consistent with common practice in the field.

20
- You should be more precise how many cells are being lost following RT step in 10X droplets? Are these 50% and 59% that you describe in Figure 1c? It was not very clear.

The 50% and 59% in Figure 1c summarize the overall recovery rate for the entire assay. As described in our response to comment #9, the lost cells result from both the microfluidic system and the pooling and splitting following RT. While we cannot distinguish between the losses in these two steps, the microfluidics system alone likely contributes to about 40% loss, as indicated in the manual of the Chromium products. Therefore the loss following the RT step would contribute to approximately 10%.

21

- Related to above comments. You should consider providing more details of identifying multiplet rate that comes from the same species (e.g. ≥2 K562 cells in a droplet).

Thank you for your comment. As described in our response to comment #5, in the species-mixing experiment, we reported both the observed multiplet rates (combinatorial barcode collision between mouse and human) as well as the overall multiplet rates. The latter extrapolates the unobservable multiplets (Human-Human and Mouse-Mouse), and is calculated by doubling the observable multiplets as described in Methods. We have included more information regarding this in the main text in the revised manuscript (Line 116-117).

22

- K-562 cells is relatively homogeneous and are rich in mRNA, which means that this cell line is not the best example of benchmarking the technology. User need to be certain that certain what biases OAK introduced when working with heterogeneous samples. As such I would recommend sequencing PBMCs and comparing it to 10x results

We agree that using homogeneous cell lines rich in mRNA may not reveal all qualities of single cell technologies. Therefore, in addition to benchmarking with K562 and NIH/3T3 cell lines, we have also shown a direct comparison of performance between OAK and standard Chromium on the same human retina sample in Supplementary Figure 2.

We would also like to highlight that the vast majority [2,4,5,8], if not all, of the published combinatorial indexing methods have used cell lines for benchmarking, given their more standardized nature, which facilitates data comparison across laboratories. This common practice guided our choice of benchmarking materials in our manuscript.

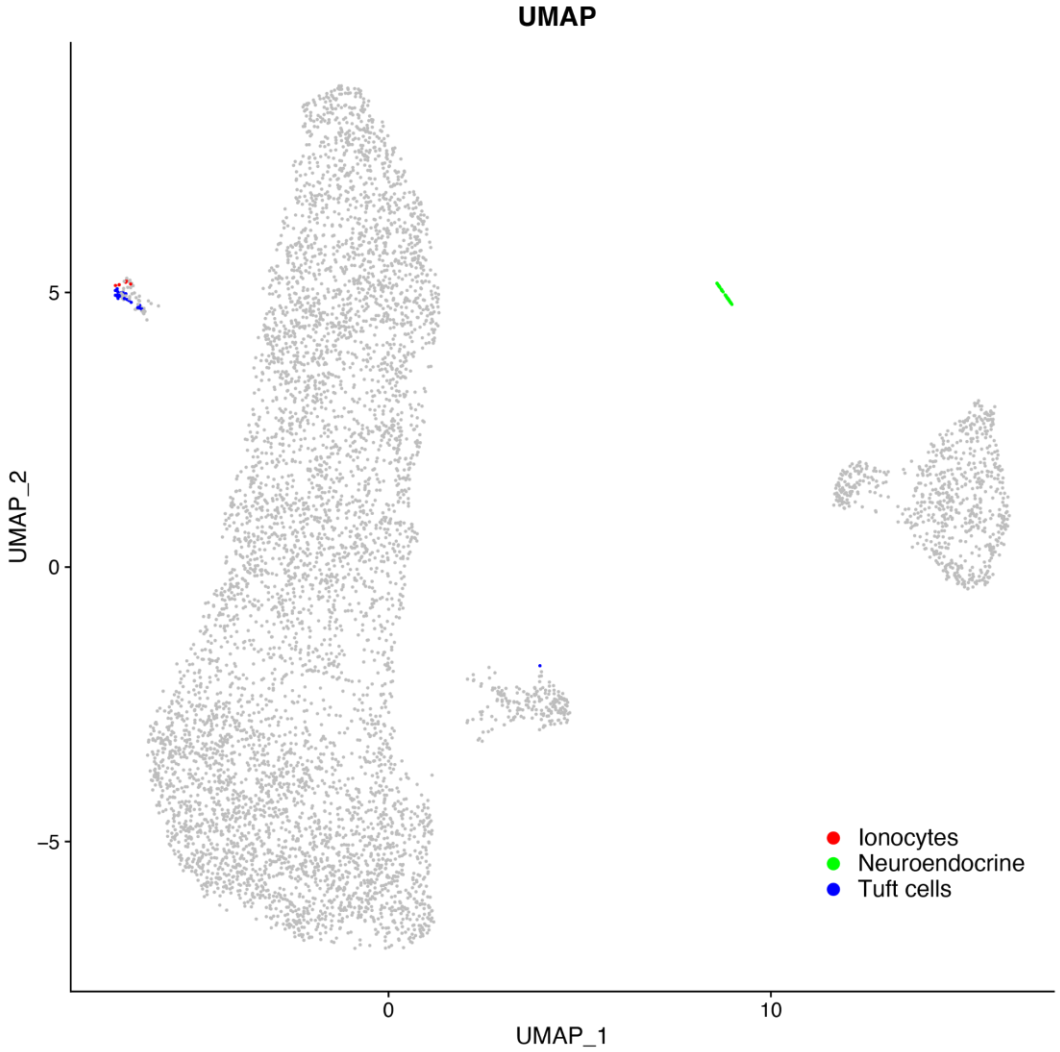
Nevertheless, we aim to understand how OAK performs across a wide variety of samples. However, as discussed in the Discussion section, the current version of OAK has limitations in generating high-complexity libraries for PBMCs. To our knowledge, none of the combinatorial indexing publications [2,4,5,8] have provided a benchmark for scRNA-seq with PBMCs. Additionally, an independent study reported that they were unable to generate high-quality PBMC data by sci-RNA-seq [7]. We recognized this as a direction for future development, as discussed in the Discussion section.

23

- Authors provide Figure S1l showing in vitro differentiated bronchial airway cells (Line 154) , but some expected cell types (e.g. brush, ionocytes, PNECs, etc) are missing. Its not clear why, and might indicate the problems with the methodology or the method. Again, the use of primary

PBMC, fresh or frozen, would be a much more valuable content about the capabilities of the technique.

Thank you for the question. While not the focus of this study, it is interesting to investigate whether the rare cells are present in the in vitro differentiation model and whether they are detectable by OAK. Based on marker gene expression, we are able to identify ionocytes, tuft cells, and pulmonary neuroendocrine cells (PNECs) as indicated in the UMAP for OAK data below. PNECs were annotated in our dataset as neuroendocrine cells in Supplementary Figure 1l. The tuft cells and ionocytes were previously labeled as "unknown" in Supplementary Figure 1l, as this label includes multiple types of rare cells that regular clustering methods do not always separate them due to their rarity and similar expression [17]. In the revised manuscript, we have relabeled this group as "rare" cells to improve clarity, as was done in previous studies.



24
- Line 157. Authors are using hashing antibodies to stain cells and later fix them in methanol. It is not clear how it is possible that Ab-stained cells can undergo methanol fixation without losing antibodies. I find it hard to believe that Hashtag antibodies remain bound to their targets following methanol fixation.

We agree that methanol treatment can lead to membrane permeabilization and protein denaturation, both of which raise concerns about the ability to recover hashing information by sequencing. This is partly why we found it necessary to perform the cell hashing experiment in conjunction with OAK to demonstrate compatibility. It is also important to note that we are not the only study that has successfully performed antibody staining followed by methanol fixation. CITE-seq has been proven compatible with downstream methanol treatment by several published works [3,12,13,18]. Among these, we found the work by Hwang et al. particularly compelling, as it demonstrated that even after methanol fixation, antibody-oligo-conjugates directed against surface proteins are retained with the cell and able to provide single-cell surface protein expression profiles.

25
- I find it misleading when authors sequence a small fraction of cells and then claim high cell recovery rates. For instance, while authors sequenced 8,096 cells, they assert a recovery of 44,582 cells. Although I understand the rationale behind such claims, authors should strictly adhere to the facts. There is no conclusive evidence to guarantee uniform recovery efficiency across tubes, as the library preparation for each tube can be subject to technical artifacts such as cell count discrepancies, contamination, and doublets. It is essential for the integrity of the research that these factors are acknowledged and transparently addressed in the methodology and results discussion.

Thank you for the comment. In the revised manuscript, we have edited the text (Line 162-164) to reflect that the total of 44,582 is a projected number based on the aliquots we sequenced.

We also appreciate that the reviewer asked for evidence to support our assumption of uniform recovery. In the retina profiling experiment, we sequenced all aliquots made (13 total), and have included the cell number from each aliquot in Supplementary Table 2 in the revised manuscript. This dataset demonstrates a tight distribution of the number of cells recovered in each aliquot, with a mean of 3130 and a standard deviation of 269.

26
- Likewise in Line 330 I find statement that OAK combined “combinatorial indexing” to be misleading. Authors simply adopt 10X genomics protocol without involving split-and-pool steps.

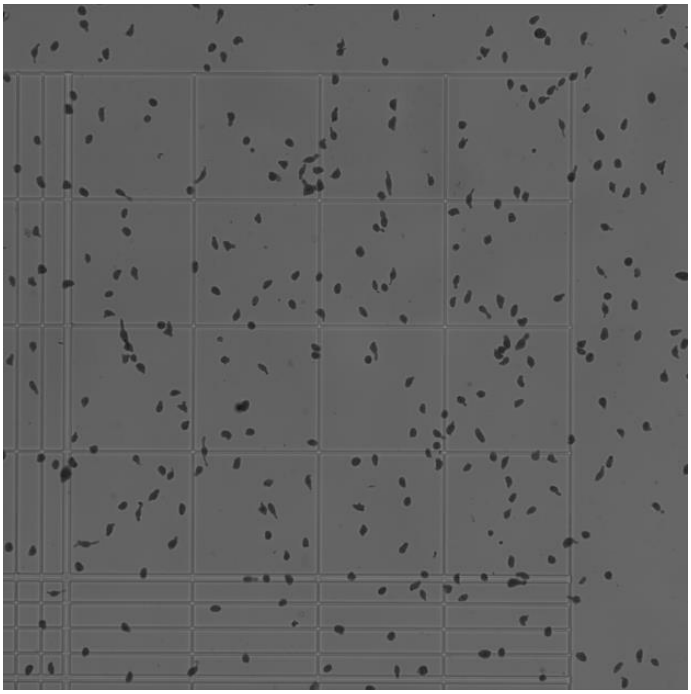
Thank you for your comment. In our response to comment #5, we have reiterated that OAK is a combinatorial indexing approach, with the first "split" performed through droplet generation, the

"pool" performed by emulsion breaking, and the second "split" performed through aliquot generation.

27
- Line 485. At concentration 2,400 nuclei / µL a significant degree of clumping can be expected. You may want to provide further details of the sample quality at such concentration.

We appreciate the concern. However, with appropriate dissociation and pipetting, this or higher concentration of input is commonly used for single-cell experiments without leading to clumping issues. For example, sci-RNA-seq [4] has targeted a final concentration of 5,000 cells or nuclei per µL for subsequent split-and-pool. Similarly, SHARE-seq [8] has recommended 2,000-4,000 cells / µL for preparation for transposition reactions. In the manual provided by 10x Genomics (https://cdn.10xgenomics.com/image/upload/v1666737555/support-documents/CG000338_ChromiumNextGEM_Multiome_ATAC_GEX_User_Guide_RevF.pdf) on page 29 the recommended nuclei stock concentration can be up to 8,060 cells / µL.

We frequently prepare samples at this or higher concentrations, and have attached below a microscopic image for a sample with >3,000 cells / µL for reference.



28
- Line 496: Please be precise, how many tubes were used to transpose 150-200k nuclei? Do I understand correctly that you used 20 tubes to perform taggmentataion, and volume of each

reaction is 15 µL? After taggmentation the aliquots in these tubes were pooled and loaded on 1 Chromium lane?

Yes, this is correct. In the revised manuscript we have rephrased this paragraph in the Methods section (Line 502-508) to improve clarity.

29
Line 500: Please indicate how many nuclei are resuspended in “15 µl of solution” before loading 1 channel?

Thank you for the question. In the revised manuscript we have rephrased Line 507-508 to clarify that all the nuclei from all the reactions were combined and resuspended in the 15 µl solution.

30
Line 505-512: These are not cells, these are nuclei. Please fix.

Thank you very much for identifying this error. We have corrected it in the revised manuscript.

Reviewer #3 (Remarks to the Author):

I had reviewed the manuscript for another journal. The authors have carefully and satisfactory addressed our questions. The results on PBMC are a bit worrying, but the authors have addressed this limitation in the discussion.

We thank the reviewer for their reply that we have carefully and satisfactorily addressed their questions.

Reference

1. Lareau, C. A. *et al.* Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. *Nat. Biotechnol.* **37**, 916–924 (2019).
2. Datlinger, P. et al. Ultra-high-throughput single-cell RNA sequencing and perturbation screening with combinatorial fluidic indexing. *Nat. Methods* **18**, 635–642 (2021).
3. Hwang, B. et al. SCITO-seq: single-cell combinatorial indexed cytometry sequencing. Nat. Methods 18, 903–911 (2021).
4. Cao, J. et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017).
5. Rosenberg, A. B. et al. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science **360**, 176–182 (2018).
6. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
7. Ding, J. et al. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. Nat. Biotechnol. **38**, 737–746 (2020).

8. Ma, S. et al. Chromatin Potential Identified by Shared Single-*Cell* Profiling of RNA and Chromatin. Cell **183**, 1103-1116.e20 (2020).
9. Simone, M. D. et al. Comparative Analysis of Commercial Single-Cell RNA Sequencing Technologies. *bioRxiv* 2024.06.18.599579 (2024) doi:10.1101/2024.06.18.599579.
10. Stoeckius, M. et al. Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19**, 224 (2018).
11. Hartmann, F. J., Simonds, E. F. & Bendall, S. C. A Universal Live Cell Barcoding-Platform for Multiplexed Human Single Cell Analysis. *Sci. Rep.* **8**, 10770 (2018).
12. Pisu, D. et al. Single cell analysis of M. tuberculosis phenotype and macrophage lineages in the infected lung. *J. Exp. Med.* **218**, e20210615 (2021).
13. Kleino, I., Nowlan, K., Kotimaa, J. & Kekäläinen, E. Optimising protein detection with fixable custom oligo- labelled antibodies for single- cell multi- omics approaches. *Biotechnol. J.* **17**, e2100213 (2022).
14. Gupta, A. et al. Characterization of transcript enrichment and detection bias in single-nucleus RNA-seq for mapping of distinct human adipocyte lineages. *Genome Res.* **32**, 242–257 (2022).
15. Pavan, R. R., Diniz, F., El-Dahr, S. & Tortelote, G. G. Gene length is a pivotal feature to explain disparities in transcript capture between single transcriptome techniques. *Front. Bioinform.* **3**, 1144266 (2023).
16. Tiemeijer, B. M. et al. Probing Single-Cell Macrophage Polarization and Heterogeneity Using Thermo-Reversible Hydrogels in Droplet-Based Microfluidics. *Front. Bioeng. Biotechnol.* **9**, 715408 (2021).
17. Sikkema, L. et al. An integrated cell atlas of the lung in health and disease. *Nat. Med.* **29**, 1563–1577 (2023).
18. Pisu, D. & Russell, D. G. Protocol for multi-modal single-cell RNA sequencing on M. tuberculosis-infected mouse lungs. *STAR Protoc.* **4**, 102102 (2023).

Reviewer #2 (Remarks to the Author):

1. I believe I have a reasonable understanding of the OAK technology. My primary concern is that this work lacks conceptual novelty. Barcoding (Indexing) in droplets, barcoding (indexing) in-situ, and in plates has all been previously demonstrated, as has combining both methods (dsciATAC-seq, scifi-RNA-seq, SCITO-seq, etc). The improved sensitivity claimed is not due to the development of new biochemistry or tools by OAK, but because it utilized reagents and kit developed by 10X Genomics. The authors simply tweaked the existing 10X Genomics protocol; were rather than inactivating the RT enzyme at 85ºC for 15 minutes, they collected "cells" (or what remains of them after 42ºC for 60 minutes in the presence of lysis reagents) and then aliquoted samples into multiple tubes to perform a 2nd indexing step in PCR tubes, as depicted schematically in Figure 1.

Frankly, I find it challenging to identify conceptual or technological novelty here. With all due respect, my opinion remains unchanged: merely adapting an existing commercial reagents kit to perform a barcoding (first indexing) reaction and then proceeding to second indexing in plates (or tubes) does not constitute an advance in the field, especially considering an array of methods based on a similar concept such as dsciATAC-seq, scifi-RNA-seq, SCITO-seq, etc.

We would like to reiterate that OAK presents significant improvements over existing combinatorial indexing technologies, including the ones mentioned by the reviewer that also utilize a droplet system (dsciATAC-seq, scifi-RNA-seq, SCITO-seq). We list below some of the improvements introduced by OAK in the field of droplet-based combinatorial indexing:

1. **Broader compatibility with molecular modules.** We have demonstrated that OAK is compatible with paired ATAC-Seq and RNA-Seq, which none of the methods referenced by the reviewer (dsciATAC-seq, scifi-RNA-seq, SCITO-seq) is. In addition, OAK is compatible with antibody labeling as well as lineage tracing applications as demonstrated in the manuscript. This broader compatibility allows for more comprehensive multi-omic profiling within a single experiment.

2. **Higher detection sensitivity compared to other combinatorial indexing methods.** We benchmarked OAK to a number of existing ultra-high throughput methods and found that OAK exhibited significantly higher detection sensitivity. Notably, even when compared to scifi-RNA-seq, another method that utilized 10x Genomics' platform, OAK showed an over ten-fold improvement (OAK's 3014 genes per cell vs. scifi-RNA-seq's 240 genes per cell). This directly counters the reviewer's statement that OAK's sensitivity is solely due to the use of reagents and kits developed by 10X Genomics. Instead, OAK provides a unique way of utilizing a droplet system to replace micro-well plates in the first step of split-and-pool, which is different from dsciATAC-seq, scifi-RNA-seq, or SCITO-seq.

3. **Simpler experimental workflow.** OAK requires the least manual pipetting among combinatorial indexing methods, and significantly reduces the experimental time required to reach the first stopping point. In addition, OAK eliminates the need for large investments in synthesizing plates of indexing oligos or assembling pre-indexed transposomes for the ATAC modality. These advantages make OAK a more accessible option by reducing both complexity and cost.

In summary and in contrast to the reviewer's opinion, we believe these advancements represent meaningful contributions to the field of single-cell RNA-seq, providing researchers with a more versatile, sensitive, and user-friendly method for ultra-high throughput single-cell analysis.

2. OK

3. OK

4. OK

5. I agree with the response and appreciate the reference to Figure S1b; it's helpful. Given that both the first indexing (in Chromium) and the second indexing (in PCR tubes, plates) are of equal importance, it would be beneficial to include both panels—Figure 1b and Figure S1b—in the main Figure. Meanwhile, Figure 1c could be relocated to the supplementary materials.

Line 558 "Multiplet rate theoretical estimation":
I agree with using the birthday problem to estimate the collision rate, but the current description is still somewhat unclear. It would be beneficial to delineate the collision rate estimation for the first indexing (in droplets) separately from the second indexing (in tubes). Could you clarify the relationships used? I agree that for the 1st indexing one can use brith problem as described by Ma et al., 2020, Cell 183, 1103–1116, with N = 150k or 450k, whereas D = 100k. This would result in collision rate being 48% or 78% (as indicated in Figure S1b).

5.1. Further Inquiry:
What formula should be applied to estimate the barcode collision rate for the second step of indexing? What values for N and D should be use? At this step, are you assuming that at the 2nd step of indexing you start with 450k cells and that each cell in a broken emulsion has a unique index 1? Expanding your Methods section and stating the assumptions (if they are made) would be helpful for potential users. Next, what relationship one should use to estimate barcode collision rate for the 2nd step indexing.

We would like to thank reviewer 2 for their feedback. Regarding the choice between Fig S1b and Fig 1c, we prioritize Fig 1c over Fig S1b in the main figure, because Fig S1b presents a theoretical estimation for the collision rate which is typical for all combinatorial indexing methods. In contrast, Fig 1c includes key experimental results on overloading and the throughput OAK achieves. We believe this arrangement leads to a more efficient use of space in the main figure.

Regarding multiplet rate theoretical estimation, although the reviewer acknowledged the suitability of using the birthday problem, it appears there may be some misunderstanding regarding its application. In Ma et al.[1], D value is calculated as 96 × 96 × 96 = 884736, which is the product of the combined barcoding space generated by all three rounds of barcoding, each occurring on a 96-well plate. Analogously, with OAK, the D value would be the combined coding possibilities from both the primary index (100k droplets) and the secondary index (12 aliquots in the case of the 150k cell loading experiment). Thus, D value should be calculated as 100k × 12, which leads to a theoretical collision rate of 5.997% (Y-axis) for 12 aliquots (X-axis) at N =150k, as indicated in Fig S1b.

Through the example provided by Ma et al. and the general approach in the combinatorial indexing field, we hope it is clear that there is no point delineating the collision rate for each round of split-and-pool, as requested in the further inquiry. The combined barcoding space effectively addresses the collision rate for the entire process, and we have experimentally assessed the agreement between theoretical and actual collision rates by performing species mixing experiments.

6. OK

7. It seems there might be a misunderstanding regarding the response about the mean number of cells sharing the same barcode. When you mention that the mean number of cells sharing the same 10X barcode is 1.4, it's important to clarify what this statistic signifies in the context of cell loading and barcode distribution. At a loading of 150k cells per Chromium lane with 100k droplets, approximately 48% of barcoded cells will carry a 10X barcode (the first index) that appears at least twice within the same aliquot. This scenario, which I previously referred to as cells having the same 10X barcode, suggests a significant overlap in barcode assignment due to the limited number of droplets with gelbeads compared to the number of cells.

We would like to thank the reviewer for the clarification. We appreciate the opportunity to better address this concern, which seems to arise from a misunderstanding of the key concepts of combinatorial indexing.

The small number of barcodes in each round of indexing relative to the massive number of cells is an intentional feature rather than a limitation. This is because combinatorial indexing leverages the combination of multiple rounds of indexing to significantly expand the barcoding space, thereby ensuring that the likelihood of two cells taking the exact same path in the random split-and-pool procedure is satisfyingly low. Both theoretical estimates and experimental evidence from previous studies, as well as this study, have demonstrated the feasibility and effectiveness of the combinatorial barcoding strategy.

To promote better understanding, we can draw an analogy to traditional plate-based combinatorial indexing methods [1–3]. In these methods, in the first splitting step 1,000-20,000 cells are distributed into each well to acquire the same first-round index. This means that 100%

of barcoded cells at this point will carry a first-round index that appears at least twice in the entire collection of cells. However, one should focus on each of the 1,000-20,000 cells for subsequent rounds of indexing to resolve, rather than concerning over the 100% of cells, because they do not all share the same first-round index.

Similarly, in the context of OAK, regardless of whether 48% or 100% of barcoded cells are non-unique in the first splitting step, they are distributed across 100k droplets, which already provide 100k distinct first-round indices. Since only the cells co-encapsulated in the same droplet will share the same first-round barcode, we should focus on these co-encapsulated cells for subsequent consideration, rather than the 48% of cells, most of which do not even share the same first-round index. The average number of co-encapsulated cells (sharing the same first-round index) is only 1.4 in this experiment, as presented in Fig 1c. The subsequent round of indexing provided the diversity (12, in this experiment) to further resolve each of these 1.4 cells into single cells. This is evidenced by the low multiplet rate assessed both theoretically and experimentally in the manuscript.

8. OK

9. OK

10. OK

11. OK

12. OK

13. OK

14. I believe the rebuttal to my earlier critique may not fully address the core of my argument. I agree that cells preserved in methanol should indeed be referred to as cells, given they retain the plasma membrane, organelles, cytoplasm, and other cellular structures. My concern, however, is that the OAK protocol does not actually retrieve cells from the droplets following the RT reaction. Given the presence of detergents and elevated temperatures, cells will lose their plasma membranes, cytoplasm content leaving predominantly the nucleus intact, as harsher conditions (e.g., SDS, proteinase K) are required to disrupt it.

The authors counter this by stating:

Nevertheless, these "cells" are still distinct from nuclei. For example, hashing antibody-derived tags targeting the cell surface membrane are preserved in our data, as demonstrated in Fig 1h.

However, this does not necessarily validate the retrieval of cells post-RT reaction. The hashing antibody-derived tags targeting the cell surface membrane, as shown in Fig. 1h, are indeed preserved, but it’s important to note that these antibodies were used prior to encapsulation

when cells are still intact. Upon cell lysis in droplets, some leftovers of plasma membrane and endoplasmic reticulum are known to remain associated with cell nucleus; thus it is not very surprising that hashing-Ab information was recovered; albeit the statistics of number of reads per hashing-Ab are not shown.

We would like to reiterate that the primary reason we refer to cells and nuclei as such in the manuscript is to stay consistent with common practice in the field, ensuring that readers are not confused.

We acknowledged in the main text that the RNA-seq data derived from OAK displayed features similar to single-nucleus RNA-seq, a phenomenon also observed in other studies that use combinatorial indexing as mentioned in the manuscript. However, it would be equally inaccurate to refer to them as nuclei. This is because as the reviewer also recognized, "leftovers of plasma membrane and endoplasmic reticulum are known to remain associated with cell nucleus". The attached plasma membrane enables the readout for surface proteins [4–6], while the attached endoplasmic reticulum enables the capture of mature cytoplasmic mRNA [7,8]. These factors make such biological units neither whole cells nor clean nuclei.

In our humble opinion, there is limited value in debating whether these "in-between" entities should be referred to as "cells" or "nuclei". Our manuscript aligns with the common practice in the single-cell field: when cells were loaded for barcoding without undergoing a nuclei extraction procedure, we referred to them as "cells". In contrast, in the human retina profiling example, since nuclei were experimentally extracted, the profiling unit became the "nuclei", and we referred to the process as single-nucleus RNA-Seq and single nucleus ATAC-Seq. We believe this approach conveys the information most clearly.

15. I may not have been clear in my initial request. I was asking to evaluate the scRNA-seq reads' coverage across the transcript length of the gene body, effectively assessing transcript recovery as gene-body coverage. The purpose of this analysis is to differentiate between transcripts retained in cells—which would typically represent both short and long genes, including those crucial for physiological responses such as signaling proteins—and transcripts captured primarily in nuclei, which are often long, unspliced and carry limited biological information.

For this analysis, I suggest using the ReSQC v5.0.1 function geneBody_coverage.py to provide a detailed overview. Comparing your data with standard 10x Genomics protocols (v3 or v3.1) will clarify whether the OAK mainly captures mRNAs retained in the nucleus, potentially losing all cytoplasmic RNA. Such a finding would suggest that describing OAK as a technique for single-cell RNA-Seq is inaccurate; particularly for non-experts in scRNA-Seq techniques.

Technically, OAK is a two-step process where the second step (second indexing) is critical as it depends on the capture of mRNA predominantly from nuclei that survive the RT reaction. I recommend emphasizing this distinction in your manuscript to avoid misconceptions and misunderstandings among readers less familiar with the diverse array of RNA-Seq techniques.

While these details might seem trivial to experienced users, less experienced readers might struggle to grasp all critical aspects of the method before employing it. Having said there is another concern that authors should address, that is, data quality:

After the first indexing and subsequent pooling of cells, there is a high possibility that barcoded-cDNA molecules not retained within cells will freely diffuse and nonspecifically bind to any random cell. This scenario could substantially increase the noise in the data and complicate data analysis; particularly when overloading the Chromium chips.

To better understand the impact of this phenomenon on data integrity, could you please provide some graphical representations in the supplementary materials? Specifically, seeing how clean the OAK RNA-Seq data is compared to standard 10X Genomics conditions. These visualizations will help clarify the extent of noise and potential cross-contamination between samples, which is crucial for assessing the reliability and quality of the data obtained through this method.

Two Figures such as 1) histogram of UMI density distribution (frequency vs UMI counts, refer to Figure 6.1) and „knee“ point (refer to Figure 5.9.1) might be sufficient.

Refer to for details https://biocellgen-public.svi.edu.au/mig_2019_scrnaseq-workshop/index.html).

The first part of this comment pertains to a similar concern raised in comment 14, suggesting that OAK primarily captures mRNAs retained in the nucleus, which according to the reviewer are "often long, unspliced and carry limited biological information". We do not agree with this statement.

> First of all, we do not find any evidence to support the claim that nucleus mRNAs "carry limited biological information". On the contrary, numerous studies have used single-nucleus RNA-seq to profile a wide variety of tissues [9–13], demonstrating a high degree of concordance with scRNA-seq in expression profiles and cell-type classification [14–17]. In some cases, single-nucleus RNA-seq even displays advantages over scRNA-seq [18,19]. Some of these studies also highlight the differences between scRNA-seq and snRNA-seq [13,15,16]. Given the rich literature readily available that dissects the similarities and distinctions of these two options of sample preparation, which is not the focus of the OAK method, we do not think it is relevant for our manuscript to replicate such studies.
>
> Secondly, unspliced mRNAs with introns have been routinely captured and included for data analysis in not only single-nucleus but also single-cell RNAseq. While it is already common to include intronic reads for combinatorial indexing methods' data analysis [2,3], recent findings have highlighted the significant presence of unspliced mRNAs even in regular single-cell data [20,21]. Including such unspliced reads improves analysis sensitivity, and hence the most frequently used single-cell analysis tool, Cellranger (10x

Genomics), has updated its settings so that intronic reads from unspliced mRNAs are included for gene counting by default [22,23].

Thus, we do not consider the capture and inclusion of unspliced mRNAs by OAK a concern. That said, we did provide Supplementary Fig 1g to disclose intronic read fraction in the OAK data, and found that our result was consistent with other combinatorial indexing methods as mentioned in the manuscript.

We found the last part of this comment also to be biased. The reviewer claimed that "barcoded-cDNA molecules not retained within cells will freely diffuse and nonspecifically bind to any random cell". Should this be true, none of the previous combinatorial indexing methods could have worked, since all of these methods include further pooling and splitting cells after the first round of cDNA barcoding. Instead, numerous studies [1–3,24] have demonstrated that using fixed cells/nuclei for combinatorial indexing achieves high quality data with satisfyingly low background noise and multiplet rate. Consistently, in our manuscript, we conducted species mixing experiments (Supplementary Fig 1d) to demonstrate the quality of OAK data. Furthermore, the cell-type specific gene expression detected from the retinal sample (Fig 2c-e) and the distinct transcription programs associated with different lineages of melanoma cells (Fig 3c-f) also suggest that OAK method is able to detect signals specific to each cell, rather than having barcoded molecules "freely diffuse and nonspecifically bind to any random cell".

16. OK

17. Could you explicitly detail that in the text, rather than requiring readers to infer from the figures, the differences in performance between the two techniques at the same sequencing depth (e.g. 15k reads per cell). It's important for potential users to understand that at this depth, there is nearly a threefold difference in saturation levels (23% vs. 65%) and a significant reduction in UMI recovery by half when using OAK compared to the standard methods. This observation indirectly supports the earlier assertion that OAK primarily profiles RNA derived from single nuclei rather than from cells. Providing this information explicitly in the manuscript will help readers, especially less experienced users, in understanding the method's limitations and capabilities without having to eyeball the data from the figures.

17.1 (new request) To address my concern above, please perform gene enrichment analysis and displaying gene programs that are being depleted or enriched in OAK data set vs 10x Genomics data set.

We did explicitly detail in the text (Ln 128-131) that at the same sequencing depth (15k per cell) OAK detected 3014 genes per cell while the standard Chromium detected 3905 genes per cell. Sequencing saturation is not a linear metric against number of genes or sequencing reads, and therefore it is most straightforward to control the sequencing depth and report the number of genes detected as we did. For readers particularly interested in sequencing saturation levels, Supplementary Figure 1j has disclosed comprehensive saturation curves across a spectrum of sequencing depths as we have explained in the previous round of revisions.

We would also like to emphasize the importance of comparing OAK to methods that offer similar ultra-high throughput capabilities. While 10x Genomics' Chromium products are known for their high sensitivity, they do not provide the same level of throughput as OAK. To present the most relevant comparisons, we have benchmarked OAK with a number of existing ultra-high throughput methods (Fig 1 f-g, Supplementary Fig 1 j-k). The results clearly demonstrate that OAK detects more genes and more UMIs than any of these methods.

That said, we acknowledge the value in exploring where OAK loses sensitivity compared to the standard Chromium method. To this end, we provided Supplementary Fig 1e and Supplementary Fig 1i, which indicate that while there is a high overall correlation between OAK and the standard Chromium data, OAK is less effective in detecting the most lowly expressed genes. We have also investigated other factors that could potentially bias gene detection across different methods and found no major biases, as detailed in our response to previous rounds of revision. The new request to focus on gene programs is highly dependent on tissue and cell types, and therefore, the conclusion may not be broadly applicable to all readers. In addition, repeatedly focusing on the sensitivity difference between OAK and 10x Genomics' products detracts from the main message of our work, which is to present a more efficient method of combinatorial indexing to achieve ultra-high throughput.

18. My concern has not been adddressed but I hope that my comments above makes now more clearly why I believe that OAK is profiling RNA of cell nucleus, and not whole cells. The fact that it cannot handle PBMCs indicates that OAK is not a single-cell RNA-Seq technique.

We have reiterated our reasoning in our response to comment 14. In addition, no method performs equally well on every tissue and cell type, and we made sure to expand on some of the limitations of OAK in the discussion section.

19. See #18.

20. OK

21. OK

22. OK

23. Great. Thanks.

24. Please review the Materials and Methods section to ensure the language is clear and consistent. As currently written, there seems to be conflicting statements around the use of hashed antibodies. The manuscript says that hashing with antibodies occurs before methanol fixation (Lines 402-409), describing the treatment of live cells that are subsequently sorted. However, in line 161 of the main text, it is stated that "cells were fixed in methanol after staining," implying a different sequence of events. Furthermore, in the rebuttal letter, the authors

mention that hashed cells were used after methanol fixation, stating, “... we are not the only study that has successfully performed antibody staining followed by methanol fixation.” This statement is at odds with the methods described and seems to misinterpret prior work. For example, Hwang et al., (ref #3) utilized PFA-fixed cells, not methanol-fixed cells, to profile surface proteins.

There is no conflicting information about the sequence of events involving hashing antibodies.

Line 402-409 states that antibody hashing occurs first, before sorting which is also before fixation.

Line 161 says methanol fixation happens after antibody staining.

In the previous rebuttal letter it was reiterated that antibody staining is followed by methanol fixation.

All of these 3 statements are consistent with the sequence of events: antibody hashing/staining happens first, and methanol fixation happens after.

Regarding prior work, we would like to bring to the reviewer's attention that the figure legend of Figure 4a in Hwang et al. explicitly mentions that cells are "fixed and permeabilized with methanol", which also happens after antibody staining as detailed in their Figure 4a. This is consistent with the sequence of events in our manuscript, and supports our methodology.

25. OK

26. OK

27. OK

28. OK

29. OK

30. OK

Reference

1. Ma, S. *et al.* Chromatin Potential Identified by Shared Single-*Cell* Profiling of RNA and Chromatin. Cell 183, 1103-1116.e20 (2020).
2. Cao, J. et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* 357, 661–667 (2017).
3. Rosenberg, A. B. et al. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science 360, 176–182 (2018).
4. Hwang, B. et al. SCITO-seq: single-cell combinatorial indexed cytometry sequencing. *Nat. Methods* 18, 903–911 (2021).
5. Pisu, D. et al. Single cell analysis of M. tuberculosis phenotype and macrophage

lineages in the infected lung. *J. Exp. Med.* 218, e20210615 (2021).
6. Kleino, I., Nowlan, K., Kotimaa, J. & Kekäläinen, E. Optimising protein detection with fixable custom oligo- labelled antibodies for single- cell multi- omics approaches. *Biotechnol. J.* 17, e2100213 (2022).
7. Drokhlyansky, E. et al. The Human and Mouse Enteric Nervous System at Single-Cell Resolution. Cell 182, 1606-1622.e23 (2020).
8. Slyper, M. et al. A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. *Nat. Med.* 26, 792–802 (2020).
9. Lake, B. B. et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. Science 352, 1586–1590 (2016).
10. Tucker, N. R. et al. Transcriptional and Cellular Diversity of the Human Heart. *Circulation* 142, 466–482 (2020).
11. Sun, W. et al. snRNA-seq reveals a subpopulation of adipocytes that regulates thermogenesis. *Nature* 587, 98–102 (2020).
12. Petrany, M. J. et al. Single-nucleus RNA-seq identifies transcriptional heterogeneity in multinucleated skeletal myofibers. *Nat. Commun.* 11, 6374 (2020).
13. Gupta, A. et al. Characterization of transcript enrichment and detection bias in single-nucleus RNA-seq for mapping of distinct human adipocyte lineages. *Genome Res.* 32, 242–257 (2022).
14. Lacar, B. et al. Nuclear RNA-seq of single neurons reveals molecular signatures of activation. Nat. Commun. 7, 11022 (2016).
15. Lake, B. B. et al. A comparative strategy for single-nucleus and single-cell transcriptomes confirms accuracy in predicted cell-type expression from nuclear RNA. *Sci. Rep.* 7, 6031 (2017).
16. Ding, J. et al. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat. Biotechnol.* 38, 737–746 (2020).
17. Wen, F., Tang, X., Xu, L. & Qu, H. Comparison of single-nucleus and single-cell transcriptomes in hepatocellular carcinoma tissue. *Mol. Med. Rep.* 26, 339 (2022).
18. Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. Nat. Methods 14, 955–958 (2017).
19. Wu, H., Kirita, Y., Donnelly, E. L. & Humphreys, B. D. Advantages of Single-Nucleus over Single-Cell RNA Sequencing of Adult Kidney: Rare Cell Types and Novel Cell States Revealed in Fibrosis. *J. Am. Soc. Nephrol.* 30, 23–32 (2019).
20. Manno, G. L. et al. RNA velocity of single cells. Nature 560, 494–498 (2018).
21. Genomics, 10x. Interpreting Intronic and Antisense Reads in 10x Genomics Single Cell Gene Expression Data. *Technical Note* (2021).
22. Genomics, 10x. Interpreting Single Cell Gene Expression Data With and Without Intronic Reads. Technical Note (2022).
23. Kim, N., Kang, H., Jo, A., Yoo, S.-A. & Lee, H.-O. Perspectives on single-nucleus RNA sequencing in different cell types and tissues. *J. Pathol. Transl. Med.* 57, 52–59 (2023).
24. Datlinger, P. et al. Ultra-high-throughput single-cell RNA sequencing and perturbation screening with combinatorial fluidic indexing. Nat. Methods 18, 635–642 (2021).