Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations

To the Editor: In many gene expression studies, cells are extracted by tissue dissociation and fluorescence-activated cell sorting (FACS), but the effect of these protocols on cellular transcriptomes is not well characterized and is often ignored. Here, we applied single-cell mRNA sequencing (scRNA-seq) to muscle stem cells, and we found a subpopulation that is strongly affected by the widely used dissociation protocol that we employed. One implication of this finding is that several published transcriptomics studies may need to be reinterpreted. Importantly, we detected similar subpopulations in other single-cell data sets, suggesting that cells from other tissues may be affected by this artifact as well.

Regeneration of skeletal muscles in adults depends on the activation of otherwise quiescent muscle stem cells, the satellite cells (SCs)[1]. The quiescent SC population is considered to be heterogeneous[1,2]. We sequenced single SCs that we extracted from uninjured tibialis anterior (TA) muscles of *Pax7nGFP* mice with a widely used[2–4] dissociation protocol to characterize their heterogeneity in more detail (**Supplementary Fig. 1a–e** and **Supplementary Methods**). After dissociation and FACS, we applied scRNA-seq (CEL-Seq)[5], and we identified two subpopulations in the data

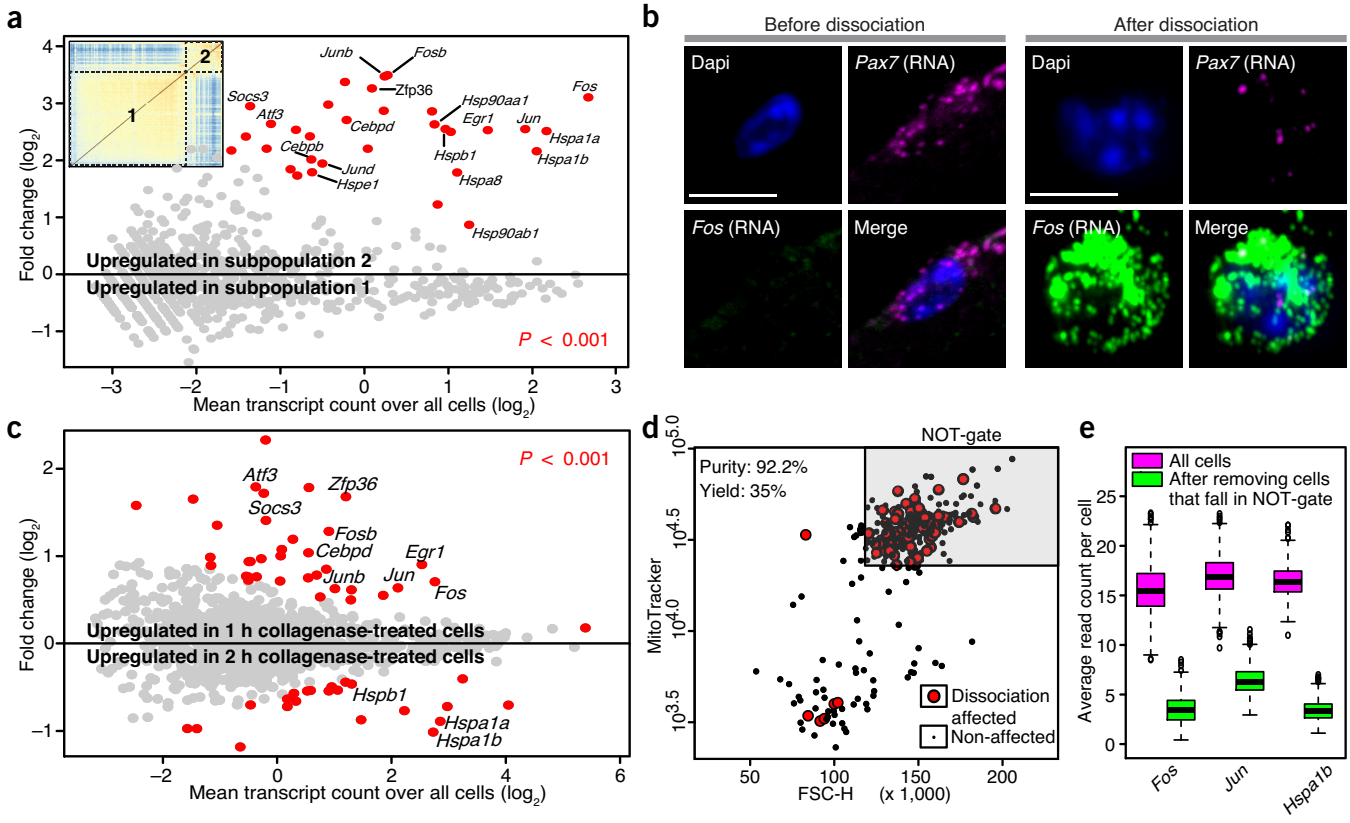

Figure 1 | Widely used tissue dissociation protocol induces transcriptional changes in a subpopulation of satellite cells. (**a**) Heatmap (inset) showing transcriptome correlations of 235 freshly isolated single-cell sequenced SCs and scatterplot showing genes that are differentially expressed between the two identified subpopulations. Significant genes are labeled in red ($P < 0.001$); P values were calculated using negative binomial distribution as previously described (**Supplementary Methods**) and were corrected for multiple testing by the Benjamini-Hochberg method; $n = 178$ and 57 cells for cluster 1 and 2, respectively. Red and blue colors in heatmap represent 1 – Pearson correlation values of 0 and 1, respectively. (**b**) Cryosection of SC in intact (all *Fos* negative; $n = 80$) and dissociated (right; *Fos* detected in 27 out of the 75 SCs) muscles that were stained for *Fos* (green) and *Pax7* (magenta) RNA using smFISH. Blue, nuclei, DAPI; scale bar, 5 μm. (**c**) Genes that are differentially expressed between 1-h and 2-h collagenase-treated SCs. P values calculated as in **a**, with $n = 272$ and 223 cells for 1-h and 2-h collagenase-treated cells, respectively. (**d**) MitoTracker and FSC-H levels of 284 MitoTracker-stained SCs. Dissociation-affected cells (red) were identified by SORT-seq; NOT-gate (gray) was designed based on a pilot study (**Supplementary Fig. 7**). (**e**) Average expression levels of *Fos*, *Jun* and *Hspa1b* in all cells (magenta) and after removing the cells that fall in the NOT-gate (green). Box plots: center line, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range; points, outliers.

(**Fig. 1a** and **Supplementary Fig. 1f,g**). The cells assigned to subpopulation 2 expressed high levels of immediate early genes (IEGs, including *Fos*, *Jun* and other activating protein 1 complex genes), *Socs3* and heat-shock proteins (HSPs) (**Fig. 1a**, **Supplementary Fig. 2** and **Supplementary Table 1**). Interestingly, these genes have been described in several satellite cell studies[3,4,6] (**Supplementary Fig. 1h**), which suggests that we identified two functionally distinct subpopulations of SCs.

To validate the existence of the two subpopulations, we performed single-molecule RNA fluorescence *in situ* hybridization (smFISH) on cryosections of *Pax7nGFP* muscles using probes designed against the subpopulation-2-specific genes *Fos* and *Socs3* (**Supplementary Table 2**). We could not detect expression of *Fos* and *Socs3* in cryosections; however, we could detect *Fos* in SCs that had undergone dissociation or both dissociation and FACS, which demonstrated that the SC isolation procedure induces *Fos* expression in a subpopulation of the SCs (**Fig. 1b** and **Supplementary Fig. 3**). Additional experiments revealed that the duration of the dissociation protocol affects the detected bulk expression levels of the genes that are unique to subpopulation 2 (**Fig. 1c**, **Supplementary Note 1**, **Supplementary Figs. 4** and **5**, and **Supplementary Tables 3** and **4**), and this confirmed that the dissociation protocol affects the transcriptome of SCs. Our observations thus suggest that subpopulation 2 might not exist *in vivo* in uninjured muscles and that, in contrast to the current consensus[1,2], the quiescent satellite cell population might be relatively homogenous *in vivo*.

Next, we developed computational and experimental strategies to remove the dissociation-affected subpopulation of SCs. The computational solution entails the *in silico* removal of dissociation-affected cells from single-cell data sets (**Supplementary Note 2**, **Supplementary Fig. 6** and **Supplementary Table 5**). The experimental solution combines indexed FACS and robot-assisted transcriptome sequencing (SORT-Seq)[7] on SCs that are stained for mitochondrial activity (**Supplementary Note 3**) in order to effectively identify and remove dissociation-affected cells during FACS (**Fig. 1d,e**, **Supplementary Note 3**, and **Supplementary Figs. 7** and **8**).

Our results show that the SC isolation procedure induces transcriptome-wide changes in a subpopulation of these cells. Even though the dissociation-affected subpopulation can be relatively small, it causes a strong contaminating signal in bulk studies because of the high expression levels of the induced IEG and HSP genes. Interestingly, the genes that are induced by dissociation are also induced by muscle injury[6], which suggests that the dissociation protocol activated some of the satellite cells (**Supplementary Note 4**). Our findings thus show that what was previously considered to be a purely quiescent subpopulation of SCs is in fact contaminated with a dissociation-affected subpopulation that might reflect activated SCs. Therefore, the results of several previous bulk studies where similar dissociation protocols have been used to study 'quiescent' SCs[2–4] warrant reinterpretation (**Supplementary Note 4**).

Since similar dissociation procedures are also used to isolate cells from other tissues, our findings may be more broadly relevant. For example, a similar IEG- and HSP-expressing subpopulation that was not validated by microscopy has been described in a recent single-cell study of mouse acinar cells[8] (**Supplementary Fig. 9a** and **Supplementary Table 6**). We also identified subpopulations with high IEG and HSP expression in other single-cell data sets from our lab, including a subpopulation of osteoblast cells in a zebrafish fin data set that is highly similar to the dissociation-affected subpopulation of satellite cells (**Supplementary Fig. 9b–f** and **Supplementary Table 7**). The overlap between our satellite cell data and other data sets suggests that dissociation protocols might induce similar problems across tissues and even across species. Taken together, our results highlight the importance of single-cell resolved experiments and validation by orthogonal methods.

Data availability statement. Sequencing data and FACS index data are deposited under accession number GSE85755. Source data for **Figure 1** is available in the online version of the paper. A **Life Sciences Reporting Summary** is available.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank M. Muraro and A. Lyubimova for experimental advice. We also thank the Hubrecht FACS facility, the Hubrecht Single Cell facility, the Hubrecht Imaging facility and the Utrecht Sequencing Facility (USF). This work was supported by a Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) TOP award (NWO-CW 714.016.001) to A.v.O. and by an European Research Council grant (ERC 220-H75001EU/HSCOrigin-309361) and the UMC Utrecht "Regenerative Medicine & Stem Cells" priority research program to C.R. *Pax7nGFP* mice were kindly provided by S. Tajbakhsh (Institut Pasteur, Paris, France).

AUTHOR CONTRIBUTIONS

S.C.v.d.B. and A.v.O. conceived and designed the project. S.C.v.d.B. and F.S. performed experiments. S.C.v.d.B., A.V. and B.S. analyzed the data. The zebrafish fin experiments were performed by J.P.-M. and analyzed by J.P.-M. and C.S.B. S.C.v.d.B. and F.S. wrote the manuscript with support from all other authors. C.R. and A.v.O. guided the project.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Susanne C van den Brink[1,2,5], Fanny Sage[1,5], Ábel Vértesy[1,2], Bastiaan Spanjaard[1–3], Josi Peterson-Maduro[1,2], Chloé S Baron[1,2], Catherine Robin[1,4] & Alexander van Oudenaarden[1,2]

[1]Hubrecht Institute–KNAW and University Medical Center Utrecht, Utrecht, the Netherlands. [2]University Medical Center Utrecht, Cancer Genomics Netherlands, Utrecht, The Netherlands. [3]Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine, Berlin-Buch, Germany. [4]University Medical Center Utrecht, Department of Cell Biology, Utrecht, The Netherlands. [5]These authors contributed equally to this work.
e-mail: a.vanoudenaarden@hubrecht.eu

1. Tierney, M.T. & Sacco, A. *Trends Cell Biol.* **26**, 434–444 (2016).
2. Rocheteau, P., Gayraud-Morel, B., Siegl-Cachedenier, I., Blasco, M.A. & Tajbakhsh, S. *Cell* **148**, 112–125 (2012).
3. Price, F.D. *et al. Nat. Med.* **20**, 1174–1181 (2014).
4. Fukada, S. *et al. Stem Cells* **25**, 2448–2459 (2007).
5. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. *Cell Rep.* **2**, 666–673 (2012).
6. Warren, G.L. *et al. J. Physiol.* **582**, 825–841 (2007).
7. Muraro, M.J. *et al. Cell Syst.* **3**, 385–394.e3 (2016).
8. Wollny, D. *et al. Dev. Cell* **39**, 289–301 (2016).

nature research

Corresponding author(s): Alexander van Oudenaarden

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

   Describe how sample size was determined.

   n.a.

2. Data exclusions

   Describe any data exclusions.

   Cells with low unique read counts (<700 for CEL-Seq single-cell datasets, <3.000 for mouse SORT-Seq datasets, <500 for zebrafish SORT-Seq dataset) were excluded from the analysis. Cutoff values were based on library complexity (Supplementary Figures 1 and 4).

3. Replication

   Describe whether the experimental findings were reliably reproduced.

   For all experiments, all replication attempts were succesful.

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   For all microscopy experiments on dissociated tissues: from each animal, one TA muscle was used as a negative control (staining on non-dissociated tissues) and the other TA muscle was used for microscopy on satellite cells that had undergone dissociation and FACS procedures.
   For time course experiment, all muscles were pooled, and the sample was randomly split into two tubes prior to collagenase incubation step (see Supplementary Methods for details).

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   n.a.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	R studio (version 3.4.0) Adobe Illustrator CC (version 2015.1.2) Adobe Photoshop CC (version 2015.1.2) ImageJ (version 1.49) Aria de DiVa software (version 8.0.1)

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	All materials are available by commercial vendors (see methods section).

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	n.a.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	n.a.
b. Describe the method of cell line authentication used.	n.a.
c. Report whether the cell lines were tested for mycoplasma contamination.	n.a.
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	n.a.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	Mus musculus, C57BL/6 and Pax7nGFP (Sambasivan et al., 2009). All Pax7nGFP mice used for microscopy and sequencing of non-MitoTracker stained cells were male; all Pax7nGFP mice used for MitoTracker staining experiments were female; all mice were between 4.7 and 7 months old at time of sacrifice.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	n.a.

nature research

Corresponding author(s): Alexander van Oudenaarden

☐ Initial submission ☐ Revised version ☒ Final submission

Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data presentation

For all flow cytometry data, confirm that:

☒ 1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ 2. The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ 3. All plots are contour plots with outliers or pseudocolor plots.

☒ 4. A numerical value for number of cells or percentage (with statistics) is provided.

Methodological details

5. Describe the sample preparation.	Satellite cells were extracted from Tibialis Anterior muscles of Pax7nGFP mice (Sambasivan et al., 2009). Sample preparation is explained in detail in Supplementary Methods.
6. Identify the instrument used for data collection.	BD FACSAria II SORP Cell Sorter (BD Biosciences)
7. Describe the software used to collect and analyze the flow cytometry data.	Aria de DiVa software (version 8.0.1)
8. Describe the abundance of the relevant cell populations within post-sort fractions.	Purity of samples was determined by re-sorting the samples. The purity for all the experiment was calculated to be between 85-95 %
9. Describe the gating strategy used.	Single cells were selected based on FSC-H/FSC-W and SSC-H/SSC-W. Live cells were selected based on Hoechst staining. Satellite cells were selected based on GFP; GFP gate was defined based on negative controls (wild type mice).

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information. ☒