

# New cell database paints fuller picture of muscle repair

November 19 2021

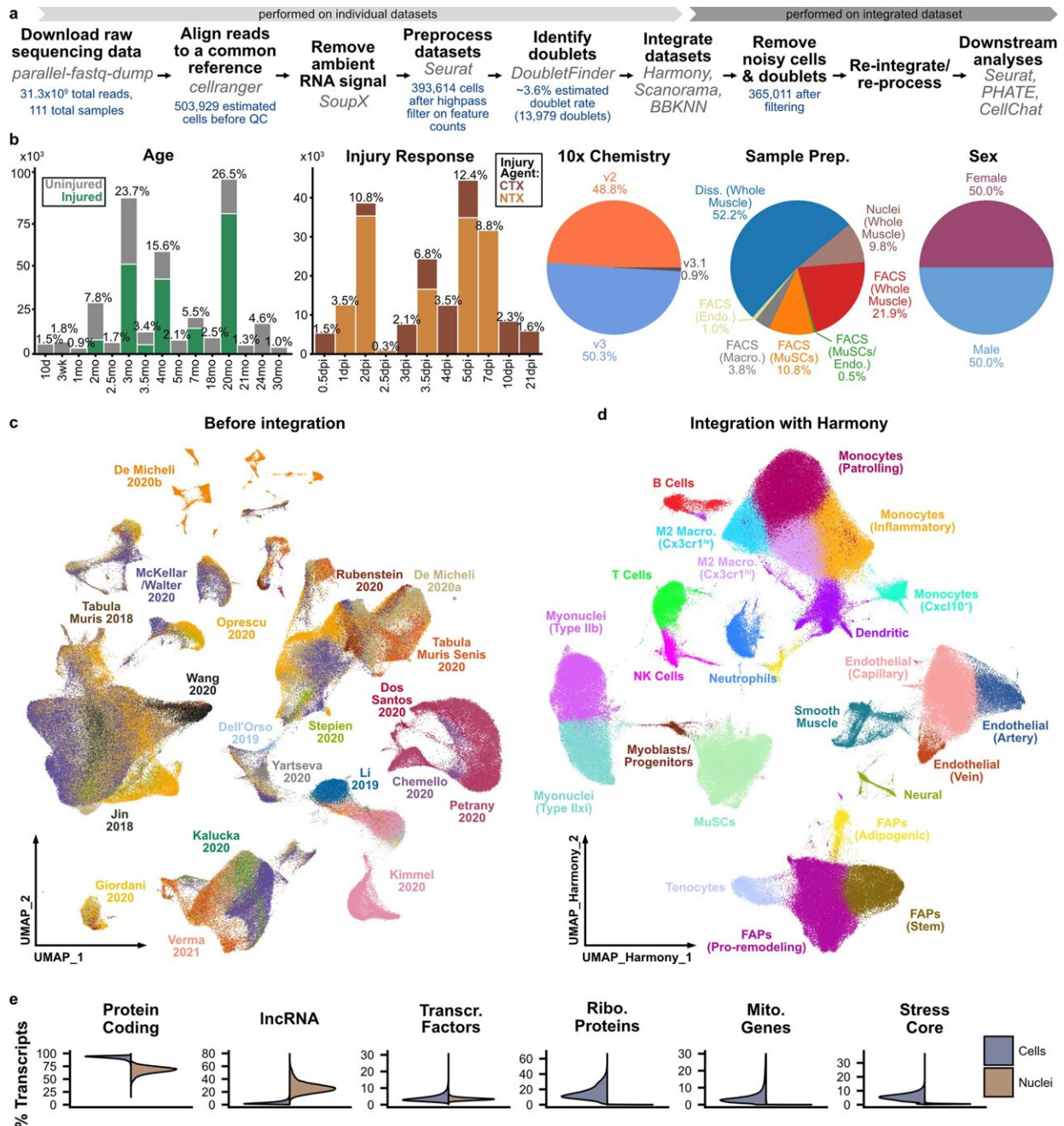


Fig. 1: Large-scale integration of 111 single-cell and single-nucleus RNAseq samples reveals cell subtypes in skeletal muscle. a Workflow used for preparation, integration, and analysis of sc/snRNAseq compendium (see “Methods”). b Overview of experimental and technical variables across compendium. The percentages shown are calculated with respect to cell number after quality control. Ages in months (mo). Injury by cardiotoxin (CTX) or notexin (NTX). Time-points in days post-injury (dpi). See also Table S1. c UMAP representation of the merged datasets after alignment, ambient RNA removal, quality control filtering and doublet removal, but before batch-correction, colored by the dataset source. d UMAP representation of integrated compendium after batch-correction with Harmony. Cells are colored by cell type, identified after Harmony integration (Fig. S3). e Differential detection of gene biotype sets between single-cell and single-nucleus datasets, including all protein-coding genes, long noncoding RNAs (lncRNAs), transcription factors, cell surface proteins, ribosomal protein subunits, mitochondrial genes, and “core” dissociation-associated stress factors. Credit: DOI: 10.1038/s42003-021-02810-x

When a muscle becomes injured, it repairs itself using a flurry of cellular activity, with stem cells splitting and differentiating into many types of specialized cells, each playing an important role in the healing process.

Biologists have struggled to study rare and transient [muscle cells](#) involved in the process, but Cornell engineers have lifted the curtain on these elusive dynamics with the launch of [scMuscle](#), one of the largest single-cell databases of its kind.

A report on the work was published Nov. 12 in the journal *Communications Biology*. Co-senior authors are Ben Cosgrove and Iwijn De Vlaminck, both associate professors of biomedical engineering in the

College of Engineering.

Recent advances in single-cell RNA sequencing allow biologists to identify tens of thousands of cells from a single tissue sample, but because muscle [stem cells](#) account for less than 1% of those cells—with their short-lived transient cell offspring being even more rare—sequencing experiments simply can't capture the complete picture of muscle regeneration.

It's a problem that Cosgrove ran into when he published a 2020 cell atlas containing 35,000 [individual cells](#) involved in the repair process. But of those cells, fewer than 200 of them were committed or fusing myogenic cells—the rare transient states that sequencing struggles to document.

"Imagine if you had a paint-by-numbers picture and you only colored in a quarter of the numbers," said Cosgrove, who co-led the development of scMuscle along with De Vlaminck and doctoral student David McKellar. "We just couldn't collect enough data ourselves to paint the whole picture of these subtle transitions as cells mature and specialize."

The Cornell team knew there were other large sequencing datasets being developed, each capturing their own share of data. So, they used advanced computational techniques to start merging collections to paint the fuller picture. They combined 88 publicly available datasets with several of their own, leading to the scMuscle database, which houses the transcriptomic data from approximately 365,000 cells involved in muscle injury over a wide range of ages and experimental conditions.

"We liken it to creating a mosaic with multiple artists. It assembles into a richer and more complicated painting," Cosgrove said. "Now we have a comprehensive picture of the very rare cell types that we know are involved in skeletal muscle repair, but weren't previously sampled."

The scMuscle database provides another important piece of information that single sequencing experiments fail to produce—spatial data that details how cells organize and interact across the tissue landscape.

"It's well known in biology that your neighbors make your identity," Cosgrove said, "and now we can identify molecular factors that are uniquely communicating between cell types and depict their spatial patterns in the injury zone."

Since soft-launching the scMuscle database in January, hundreds of researchers across the world have accessed it, searching for information such as sex-specific gene expression patterns during aging, and what gene expression signatures define different cell types involved in disease processes.

One finding reported by Cosgrove and the team answered a long-standing question about how many genes are expressed by the differentiating offspring of stem cells as they specialize in mature muscle tissue.

"It turns out the cells are really diversifying gene expression signatures and turning on all sorts of [genes](#) as the start to differentiate," Cosgrove said, "and then as soon as they begin to fuse, they hit this bottleneck and their gene expression patterns become locked in place and very restricted."

Cosgrove said the scMuscle database will continue to serve as a powerful tool for biologists and others seeking a new view of rare cellular activity in muscle regeneration, and hopes to attract funding to help with hosting and continually integrating new data into it as the field grows.

**More information:** David W. McKellar et al, Large-scale integration of single-cell transcriptomic data captures transitional progenitor states

in mouse skeletal muscle regeneration, *Communications Biology* (2021).  
[DOI: 10.1038/s42003-021-02810-x](https://doi.org/10.1038/s42003-021-02810-x)

Provided by Cornell University

Citation: New cell database paints fuller picture of muscle repair (2021, November 19) retrieved 11 May 2024 from <https://phys.org/news/2021-11-cell-database-fuller-picture-muscle.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.