

# Lineage tracing of nuclei in skeletal myofibers uncovers distinct transcripts and interplay between myonuclear populations

Received: 13 September 2023

Accepted: 10 October 2024

Published online: 30 October 2024



Chengyi Sun<sup>1</sup>, Casey O. Swoboda<sup>1,2</sup>, Fabian Montecino Morales<sup>1</sup>, Cristofer Calvo<sup>1</sup>, Michael J. Petrany<sup>1</sup>, Sreeja Parameswaran<sup>3</sup>, Andrew VonHandorf<sup>3</sup>, Matthew T. Weirauch<sup>2,3,4</sup>, Christoph Lepper<sup>5</sup> & Douglas P. Millay<sup>1,4</sup> ✉

Multinucleated skeletal muscle cells need to acquire additional nuclei through fusion with activated skeletal muscle stem cells when responding to both developmental and adaptive growth stimuli. A fundamental question in skeletal muscle biology has been the reason underlying this need for new nuclei in cells that already harbor hundreds of nuclei. Here we utilize nuclear RNA-sequencing approaches and develop a lineage tracing strategy capable of defining the transcriptional state of recently fused nuclei and distinguishing this state from that of pre-existing nuclei. Our findings reveal the presence of conserved markers of newly fused nuclei both during development and after a hypertrophic stimulus in the adult. However, newly fused nuclei also exhibit divergent gene expression that is determined by the myogenic environment to which they fuse. Moreover, accrual of new nuclei through fusion is required for nuclei already resident in adult myofibers to mount a normal transcriptional response to a load-inducing stimulus. We propose a model of mutual regulation in the control of skeletal muscle development and adaptations, where newly fused and pre-existing myonuclear populations influence each other to maintain optimal functional growth.

In vertebrate animals, skeletal muscle is a highly plastic tissue that can adapt its structure and function in response to diverse stimuli. Unlike most other cell types, mature muscle cells (myofibers) possess hundreds of nuclei. This multinucleation is established during embryonic and postnatal development through the fusion of progenitors that have differentiated from muscle stem cells (MuSCs or satellite cells)<sup>1–3</sup>. Adult muscle can also acquire more nuclei through the activation and fusion of a normally quiescent MuSC population<sup>4–13</sup>. Since nuclei provide templates for the biosynthetic capacity of cells, the conventional explanation for multinucleation in skeletal muscle is that there needs to be sufficient DNA content to support large cytoplasmic volumes associated with myofibers<sup>11,14–23</sup>. Indeed, myofibers rank among the

largest cells in mammals. In mice, these cells can extend up to 0.5 cm in length and occupy a volume of 7 nL, whereas human myofibers can reach lengths of up to 42 cm, with a volume of ~1000 nL<sup>3,24,25</sup>. To establish size, grow, and adapt, myofibers have two sources of transcriptional templates, either nuclei already present in the syncytium or newly fused nuclei. However, it is unclear how these nuclear sources transcriptionally coordinate muscle growth and adaptations<sup>14,26,27</sup>.

Despite previous controversy about the requirement of MuSC-mediated myonuclear accretion for muscle growth in the adult, there is now reasonable agreement that pre-existing myonuclei can increase output within limits, but additional nuclei are needed for sustained functional growth<sup>28–34</sup>. Potential models to explain the need for new

<sup>1</sup>Division of Molecular Cardiovascular Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. <sup>2</sup>Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. <sup>3</sup>Division of Human Genetics, Center for Autoimmune Genetics and Etiology (CAGE), Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. <sup>4</sup>Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA. <sup>5</sup>Department of Physiology & Cell Biology, College of Medicine, The Ohio State University, Columbus, OH, USA. ✉e-mail: [douglas.millay@cchmc.org](mailto:douglas.millay@cchmc.org)

myonuclei is that they aid in adaptations by increasing gene dosage or they make unique transcriptional contributions needed for growth that cannot be elicited from pre-existing nuclei. Assessing the role of myonuclear number and accretion in muscle development and growth has proven to be complex perhaps because studies have mainly utilized phenotypic analyses at the level of the whole muscle or individual myofibers. There is minimal information at nucleus-level resolution, but if available, it could help explain from a molecular perspective the need for multinucleation and new nuclei in the adult for adaptations. The development of next-generation sequencing (NGS) strategies such as single nucleus RNA-sequencing (snRNA-seq) could be helpful to parse apart the molecular identities of newly fused and pre-existing myonuclei<sup>35</sup>. snRNA-seq has been performed on human and mouse muscle, and transcriptional states of distinct myonuclear populations have been detected<sup>36–41</sup>. However, snRNA-seq approaches lack spatial information and knowledge about the recency of fusion, which are needed to separate contributions of newly fused and pre-existing nuclei. Lineage tracing approaches are also needed to identify recently fused nuclei, but the classical genetic Cre-LoxP system will not work because it causes indelible expression of a marker gene that can diffuse to other nuclei within the shared cytoplasm<sup>42</sup>.

Here, we report the development of a recombination-independent lineage tracing approach, which we combined with snRNA-seq and bulk nuclear RNA-sequencing (bnRNAseq) strategies to define the transcriptional contributions of newly fused and pre-existing myonuclei to muscle development and adult load-induced adaptations. We found a conserved signature for newly fused myonuclei between development and the adult, which was characterized by elevated expression of *H19* and *Igf2*. During muscle overload in the adult, we observed a distinct transcriptional signature of new myonuclei based on their fusion with previously formed myofibers or fusion to form new myofibers (de novo). We also detected a population of pre-existing mature myonuclei that activate a response to overload, associated with expression of *Atf3*, *Flnc*, and *Enah*. We found that fusion and myonuclear accrual are needed to coordinate transcription of pre-existing nuclei during load-induced adaptations and that the syncytial destination of newly fused myonuclei is also a determinant of their transcriptional profiles. Overall, by deploying multiple RNA-seq strategies and a mouse model that can distinguish newly fused from pre-existing myonuclei we have identified that these two populations of myonuclei influence each other during developmental and adaptive growth.

## Results

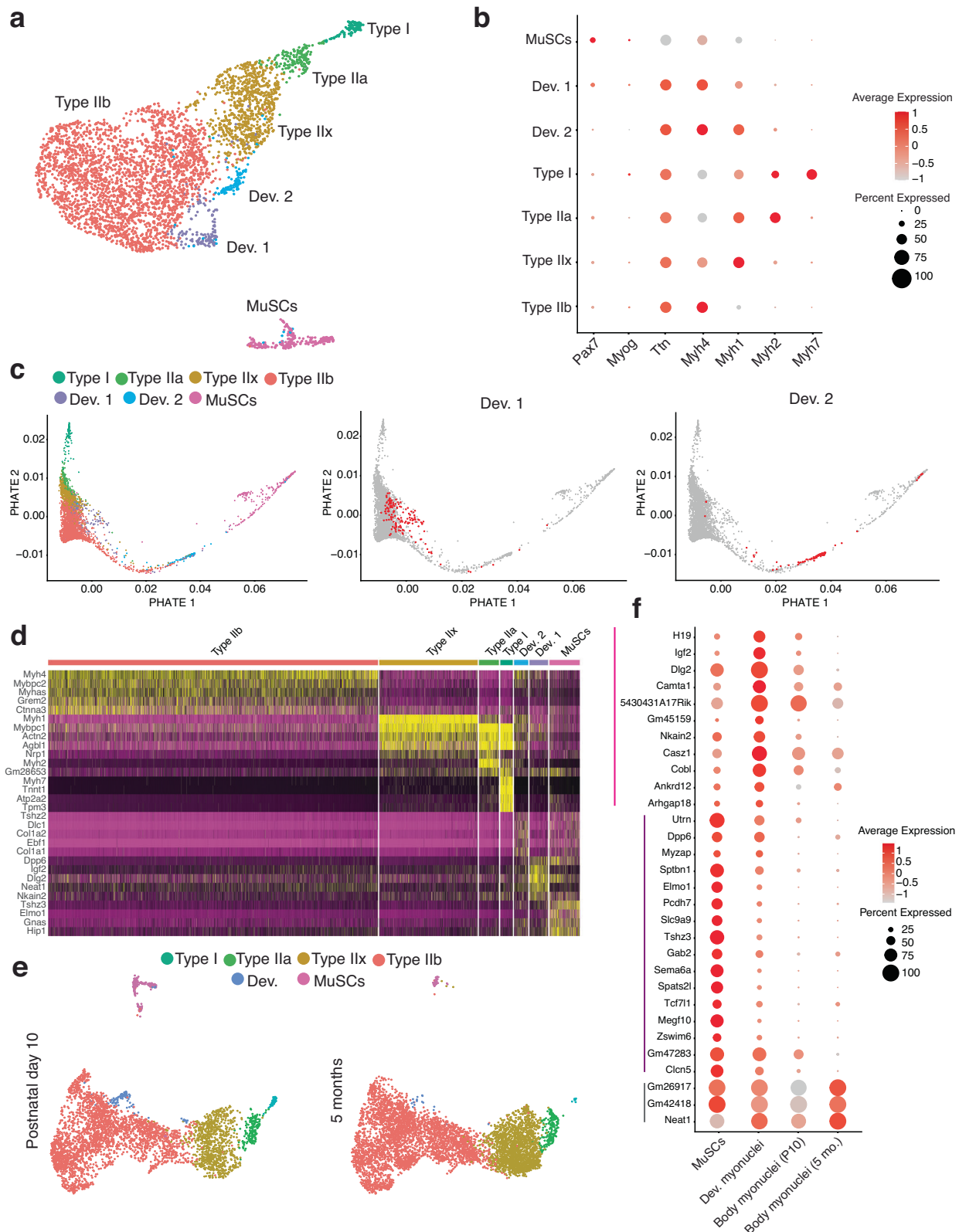
We sought to effectively and reliably determine the transcriptional identity of newly fused myonuclei within a syncytium. To this end, we first tested if a potential newly fused myonuclear population could be detected by snRNAseq data from postnatal (P) day 10, when fusion is ongoing. We previously profiled transcriptomes at this time point from all nuclear populations from tibialis anterior muscles<sup>39</sup>. Using these data, we selected only myogenic populations including MuSCs, myoblasts, and mature myonuclei to display by Uniform Manifold Approximation and Projection (UMAP) (Fig. 1a, Supplementary Data 1). MuSCs and myonuclear clusters were identified through distinct expression of canonical marker genes including *Pax7* (MuSCs), *Myog* (myoblasts), *Myh4* (Type IIb myonuclei), *Myh1* (Type IIx myonuclei), *Myh2* (Type IIa myonuclei), and *Myh7* (Type I myonuclei) (Fig. 1b). We also identified two myonuclear populations (Dev 1 and Dev 2) that possessed low levels of *Pax7* and *Myog* and some expression of *Myh* isoforms (Fig. 1b). To test whether the Dev populations may represent a transcriptional state that has progressed past differentiation but not reached full maturity as myonuclei we utilized PHATE trajectory analysis, which inferred that Dev 1 nuclei fall closest to a myonuclear identity in the lineage, while Dev 2 nuclei are further distanced from bona fide myonuclei (Fig. 1c).

Unbiased analysis of gene expression in these myonuclear clusters revealed that Dev 1 nuclei were enriched for specific genes, while Dev 2 nuclei exhibited a continuum of gene expression that aligns with all myogenic populations (Fig. 1d). We then integrated snRNAseq data from P10 muscle with snRNAseq data from mice that were 5 months of age, since there is limited fusion once muscle fully develops. Here, we found an enrichment of only one Dev myonuclear population from P10 muscle (Fig. 1e). Comparison of this population to the Dev populations found in Fig. 1a showed a similarity to Dev 1. The Dev 2 gene signature did not yield a unique cluster after integration, therefore we focused our analysis on Dev 1. Gene expression in these nuclei could be separated into one group (*H19*, *Igf2*, *Dlg2*) that were increased in Dev nuclei compared to MuSCs and decreased in other myonuclear populations at P10 and 5 months of age (Fig. 1f, pink bar). In contrast, there is also a group of genes (*Utrn*, *Dpp6*, *Myzap*) expressed in Dev nuclei but at a reduced level compared to MuSCs (Fig. 1f, purple bar). Thus, we detected a myonuclear population associated with development, and since there is ongoing fusion in muscle at this time they could represent newly fused nuclei.

## A recombination-independent tracking system labels recently fused myonuclei

To determine if the Dev myonuclei identified through snRNA-seq truly correspond to newly fused myonuclei we needed to develop a lineage tracing strategy to distinguish if they are present within the syncytium and recently fused. Traditional Cre/LoxP methods are not suitable here as they result in an indelible reporter expression that can label neighboring myonuclei in a syncytium<sup>42</sup>. To overcome these issues, we developed a non-recombination system where the transcription of the nuclear reporter is temporally regulated but produces a long-lived reporter protein. We utilized mice where a reverse tetracycline transactivator (rtTA) construct was knocked-in to the *Pax7* locus (*Pax7*<sup>rtTA</sup>), which results in transcription of rtTA in MuSCs, but rtTA requires Doxycycline (Dox) to bind to tetracycline-responsive elements (TRE) to induce transcription. *Pax7*<sup>rtTA</sup> mice were bred with mice containing a H2B-GFP expression cassette under control of a TRE (TRE<sup>H2B-GFP</sup>)<sup>43</sup>. Treatment of *Pax7*<sup>rtTA</sup>; TRE<sup>H2B-GFP</sup> mice with Dox should activate rtTA in MuSCs leading to expression of H2B-GFP mRNA and protein (Fig. 2a). As MuSCs become activated and differentiate, *Pax7* is downregulated resulting in no transcription of rtTA or H2B-GFP thereby reducing mRNA levels of the nuclear reporter prior to fusion, but H2B-GFP protein has a long half-life<sup>44</sup> and should be a sustained nuclear mark as the myogenic progenitor fuses into the syncytium (Fig. 2a).

We first assessed if MuSCs are labeled in *Pax7*<sup>rtTA</sup>; TRE<sup>H2B-GFP</sup> mice by administering Dox chow to 5 month-old mice for 5 days. Flow cytometry analysis for MuSCs (VCAM1<sup>+</sup> Lin<sup>−</sup>) confirmed that 93% were GFP<sup>+</sup> (Supplementary Fig. 1a). We also determined that 91% of the GFP<sup>+</sup> cells were MuSCs (Supplementary Fig. 1b), indicating accurate labeling of MuSCs. We next assessed if this system can label newly fused myonuclei. *Pax7*<sup>rtTA</sup>; TRE<sup>H2B-GFP</sup> mice were administered Dox for 16 days starting at postnatal day 12 (P12) (Fig. 2b) when fusion is ongoing. Even though we observed labeling of MuSCs after 5 days of Dox treatment, we chose to administer Dox for the entirety of our experimental window to capture as many GFP<sup>+</sup> nuclei as possible. Removal of Dox could result in reductions in GFP signal in MuSCs, and if these cells then activate and fuse to myofibers, it is possible they would not be captured. Isolation of myofibers from the extensor digitorum longus (EDL) muscle indicated the presence of GFP<sup>+</sup> and GFP<sup>−</sup> nuclei (Fig. 2b). FACS analysis for PCMI<sup>+</sup> myonuclei<sup>45</sup> from gastrocnemius muscles revealed 21% are GFP<sup>+</sup> but GFP<sup>+</sup> myonuclei were absent in TRE<sup>H2B-GFP</sup> mice (Supplementary Fig. 1c). We also found 25% GFP<sup>+</sup> myonuclei in the tibialis anterior muscle in cryosections, which were identified as myonuclei based on their presence inside dystrophin<sup>+</sup> myofibers (Fig. 2c). Taken together, we detected that this recombination-



independent system labels ~25% of myonuclei from P12 to P28, which is consistent with the number of nuclei added to myofibers during this period<sup>3,46</sup>, indicating that this system can accurately distinguish myonuclei based on their recency of fusion. To validate that our system does not underestimate the number of recently fused nuclei, we assessed if the H2B-GFP signal is maintained 2 weeks after labeling, which is the general time frame for our studies. *Pax7*<sup>Cre</sup>; *TRE*<sup>H2B-GFP</sup> mice

were administered Dox for 2 weeks starting at postnatal day 12, and cohort of mice were sacrificed at P28 to analyze GFP<sup>+</sup> myonuclei, whereas Dox was removed for a separate cohort of mice and they were analyzed at P42. We found a similar number of GFP<sup>+</sup> nuclei (Supplementary Fig. 1d), indicating that myonuclei are still marked by H2B-GFP over this two-week chase and that our system accurately estimates newly fused nuclei.

**Fig. 1 | Distinct myonuclear populations are associated with times of cell fusion.** **a** Uniform Manifold Approximation and Projection (UMAP) visualization represents seven color-coded myogenic nuclei clusters identified by snRNA-seq. Data were generated from tibial anterior (TA) muscles at postnatal (P) day 10. **b** Dot plot depicting canonical markers of the various myogenic states. Dot size denotes the percentage of nuclei expressing the gene. **c** PHATE trajectory illustrating myogenic populations presented in **(a)** (left), with Dev. 1 and Dev. 2 populations highlighted in the same PHATE plot on the right. **d** Heatmap of myogenic nuclei

from P10 TA muscle shows a distinct transcriptional signature. **e** Integrated dataset displays myogenic nuclei from TA muscle at P10 and 5 months of age (Split UMAP view). **f** Dot plot demonstrates altered genes in Dev. myonuclei. Dot size denotes the percentage of nuclei expressing the gene. Pink bar indicates genes that are increased in Dev nuclei compared to MuSCs and decreased in other myonuclear populations at P10 and 5 months of age. Purple bar indicates genes expressed in Dev nuclei but at a reduced level compared to MuSCs. Gray bar highlights genes that are similarly expressed across myonuclear populations.

### The transcriptional contribution of newly fused myonuclei during development and adult muscle overload

Having established that snRNA-seq data revealed a myonuclear population enriched during development, and the generation of a lineage tracing strategy that marks newly fused myonuclei, we set out to test if the Dev population is indeed newly fused and define the transcriptional contributions of newly fused and pre-existing myonuclei during development. Based on gene expression from the snRNA-seq data at P10, *H19* and *Igf2* are present in the Dev 1 myonuclear population. *H19* and *Igf2* are imprinted genes, where *H19* is maternally expressed and *Igf2* is expressed from the paternal chromosome. *H19*, a long non-coding RNA, and *Igf2*, a protein coding gene, both have known roles in myogenesis<sup>47–49</sup> and in adult myofibers<sup>50–52</sup>. We directly tested if this expression corresponds to enrichment in newly fused nuclei by performing smRNA-FISH for *H19* and *Igf2* in isolated myofibers and on cryosections from *Pax7*<sup>TA</sup>; *TRE*<sup>H2B-GFP</sup> mice that were treated with Dox from P12 to P28. We observed an enrichment of *H19* and *Igf2* around GFP<sup>+</sup> myonuclei in isolated EDL myofibers (Fig. 2d). Similarly, on cryosections from the TA we found increased *H19* and *Igf2* signal in myofibers with GFP<sup>+</sup> myonuclei (Fig. 2e). To further confirm that newly fused myonuclei contribute *H19* and *Igf2* to the syncytium, we employed a mouse model that blocks fusion after temporal deletion of Myomaker in MuSCs (*Myomk*<sup>loxP/loxP</sup>; *Pax7*<sup>CreER</sup> (*Myomk*<sup>scKO</sup>))<sup>53</sup>. Myomaker is deleted upon treatment with tamoxifen and this was performed at P6, and muscles were analyzed at P28 (Fig. 2f). As expected, *H19* and *Igf2* levels were reduced in TA myofibers where fusion was blocked at P6 (Fig. 2g). These results confirm that the Dev population identified through snRNA-seq at P10 are recently fused myonuclei and those nuclei deliver *H19* and *Igf2*.

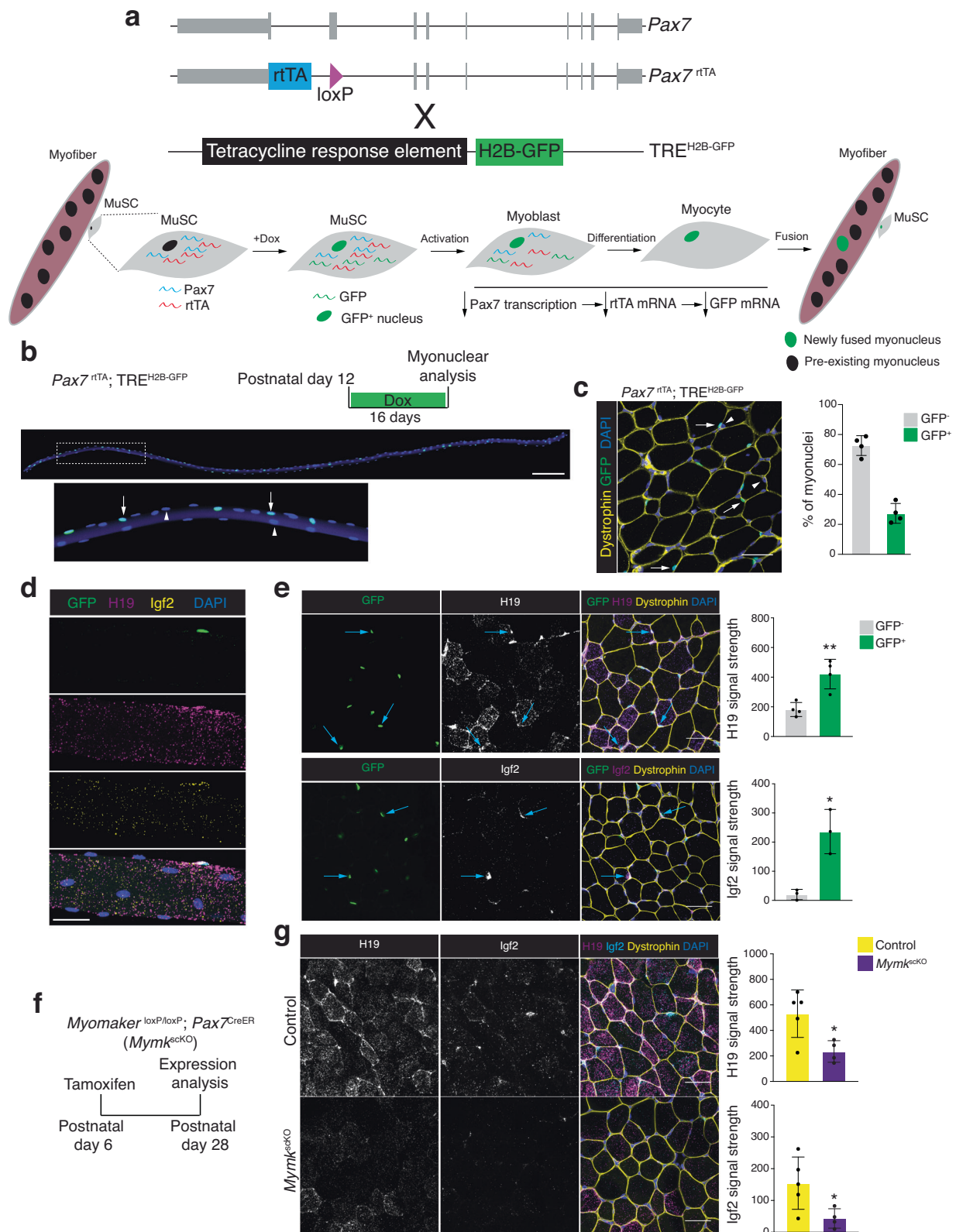
Because the models for lineage tracing of newly fused nuclei and genetic blockade of myonuclear accretion use alleles that cause *Pax7* haploinsufficiency (*Pax7*<sup>TA</sup> and *Pax7*<sup>CreER</sup>), potentially leading to an impact on MuSC function, we performed additional control experiments to validate our results. A recent report showed that *Pax7*<sup>CreER</sup> mice have normal MuSC numbers and regenerate normally in the absence of tamoxifen, but display delayed regeneration only when also treated with tamoxifen diet<sup>54</sup>. We sought to determine if there is an effect in *Pax7*<sup>CreER</sup> mice during development after a single dose of tamoxifen administered through intraperitoneal injections, which is the protocol we used to block fusion. At P35, we analyzed *H19* levels in muscle from *Myomk*<sup>loxP/+</sup>; *Pax7*<sup>CreER</sup> or *Myomk*<sup>loxP/+</sup> mice after a single dose of tamoxifen at P11 (Supplementary Fig. 2a). Here we observed similar levels of *H19* in these animals (Supplementary Fig. 2b) indicating minimal impact of *Pax7* haploinsufficiency in the developmental system. Moreover, our goal with the *Myomk*<sup>loxP/loxP</sup>; *Pax7*<sup>CreER</sup> animals is to block fusion and myonuclear accretion to test if new nuclei are a major source of *H19* and *Igf2*. If there is an impact of *Pax7* haploinsufficiency on myogenesis and fusion, it would be an added effect to Myomaker deletion, and allow even more efficient blockade of accretion. Further evidence that there is not an impact of *Pax7* haploinsufficiency is that in the *Pax7*<sup>TA</sup>; *TRE*<sup>H2B-GFP</sup> model, we observed normal rates of myonuclear accretion (~20% newly fused nuclei from P12 to P28, Fig. 2c) compared to independent reports in wild-type animals. Overall, we observed that *Pax7* haploinsufficiency did not effect MuSC function or expression of *H19* in our experimental setup during development.

However, it should be noted that *Pax7* haploinsufficiency during development has not been thoroughly investigated and use of these alleles could be a potential limitation without proper controls.

To more thoroughly characterize expression in newly fused myonuclei, we performed bulk nuclear RNA-sequencing (bnRNA-seq) on nuclei from *Pax7*<sup>TA</sup>; *TRE*<sup>H2B-GFP</sup> mice that were treated with Dox from P18 to P28. P28 was chosen for analysis because at this stage developmental growth processes are ongoing and there are both recently fused nuclei (P18 to P28) and myonuclei that are more mature (fused before P18); we thus surmised that this plan could lead to starker differences in gene expression. We separated GFP<sup>+</sup> (most recently fused) and GFP<sup>-</sup> (pre-existing) through FACS for PCMI (from TA and Gastroc pooled together), and then performed independent RNA-sequencing on those populations (Fig. 3a). Principal component analysis (PCA) revealed strong separation of GFP<sup>+</sup> and GFP<sup>-</sup> myonuclei indicating distinct gene expression signatures (Fig. 3b). *H19* and *Igf2* were enriched in GFP<sup>+</sup> myonuclei, consistent with snRNA-seq results. We detected 1877 genes more highly expressed in GFP<sup>+</sup> myonuclei and 2789 genes more highly expressed in GFP<sup>-</sup> myonuclei (Fig. 3c). In particular, *Gm15668*, *Gm3669*, *Cpne4*, and *Btg4* were significantly enriched in GFP<sup>+</sup> myonuclei based on a 0.01 adjusted *p*-value (FDR) and 2-fold change (Supplementary Data 2). In addition, we found *Myomk* enriched in GFP<sup>+</sup> nuclei, which likely remains from the fusion process and thus highlights the fidelity to distinguish expression in these nuclear populations with our system. *Pcdhgal*, *Amy2a3*, *Grem2*, *Lrrtm3*, and *Kcnn2* were the top downregulated genes in GFP<sup>+</sup> myonuclei compared to GFP<sup>-</sup> (Supplementary Data 2). Gene ontology analysis based on biological process for genes enriched in GFP<sup>+</sup> myonuclei suggested an effect on membrane processes involving adhesion, ion transport, and synapse regulation (Fig. 3d). In contrast, genes enriched in GFP<sup>-</sup> myonuclei were associated with mature muscle processes such as muscle contraction, sarcomeres, and mitochondrial metabolism (Fig. 3d). Indeed, GFP<sup>-</sup> myonuclei were enriched for *Maf* (Supplementary Data 2), a transcription factor identified as a regulator of the mature muscle program in fast myofibers<sup>40</sup>. These data indicate that nuclei that enter myofibers during postnatal developmental growth take on a role associated with plasma membrane remodeling and do not possess a strong transcriptional signature for processes associated with muscle growth such as sarcomere organization. Since there is a preponderance of evidence showing that nuclear numbers determine myofiber sizes, it is possible that pre-existing myonuclei are the main transcriptional contributors to increases in myofiber volumes through sarcomerogenesis during early postnatal growth. Newly fused nuclei could either be regulating surface area increases through membrane homeostasis or be transcriptionally unfocused on specific muscle processes because the level of volume growth needed can be driven by the pre-existing population, and newly fused nuclei are recruited for sarcomerogenesis during later maturation and growth, when volumes attained exceed the biosynthetic capability of the pre-existing myonuclear population.

To test if the transcriptional signature of newly fused nuclei detected during development is consistent regardless of the growth stimulus that leads to activation of MuSCs and their subsequent fusion with myofibers, we utilized a model of increased load in adult mice through a synergist ablation surgery (muscle overload (MOV)). The





gastrocnemius and soleus are removed leaving the plantaris muscle to bear the entire load of the hindlimb leading to myonuclear accretion and hypertrophy<sup>55</sup>. We first determined if the *Pax7*<sup>rtTA</sup>; TRE<sup>H2B-GFP</sup> lineage tracing strategy for newly fused nuclei works in adult muscle. Mice were treated with Dox before and after muscle overload (Supplementary Fig. 3a). Mice were maintained on Dox after overload to efficiently capture GFP<sup>+</sup> newly fused nuclei, and to mitigate the possibility

that MuSCs would be initially labeled but lose GFP expression prior to fusion. Analysis of plantaris myofibers showed heterogeneous GFP expression within myonuclei (Supplementary Fig. 3b). Using FACS to analyze myonuclei (PCMI<sup>+</sup>) we found 25% GFP<sup>+</sup> myonuclei (Supplementary Fig. 3c). Analysis of GFP<sup>+</sup> nuclei within dystrophin<sup>+</sup> myofibers also revealed GFP<sup>+</sup> and GFP<sup>-</sup> nuclei, where 25–30% were GFP<sup>+</sup> (Supplementary Fig. 3d). These results are in agreement with the predicted

**Fig. 2 | A recombination-independent nuclear tracking system distinguishes newly fused nuclei and allows evaluation of gene expression.** **a** Schematic of the  $Pax7^{rtTA}$ ;  $TRE^{H2B-GFP}$  alleles and Dox-inducible lineage tracing system. **b** Representative images of a single myofiber isolated from the Extensor digitorum longus (EDL) muscle at P28 after 16 days of Dox treatment. The bottom panel represents a magnified field of view of the region indicated by the box in the top panel. GFP<sup>+</sup> myonuclei (white arrows) and pre-existing GFP<sup>+</sup> myonuclei (arrowhead) are observed. Scale bar, 200  $\mu$ m. **c** Representative image showing newly fused GFP<sup>+</sup> myonuclei (white arrows) and pre-existing GFP<sup>+</sup> myonuclei (arrowhead) (left). Quantification of the percentages of GFP<sup>+</sup> myonuclei and GFP<sup>+</sup> myonuclei (right) in TA muscles. Dystrophin was used to outline myofibers. ( $n = 4$ ). Scale bar, 20  $\mu$ m. **d** Representative images of smRNA-FISH for *H19* and *Igf2* on an EDL myofiber, taken from the same mice used in (c). Scale bar, 20  $\mu$ m. **e** Representative images of

smRNA-FISH for *H19* and *Igf2* (left) on  $Pax7^{rtTA}$ ;  $TRE^{H2B-GFP}$  TA muscle at P28. Quantification of signal in GFP<sup>+</sup> and GFP<sup>+</sup> is shown on the right. Blue arrows indicate increased signal in newly fused GFP<sup>+</sup> myonuclei. ( $n = 3-4$  independent animals). Scale bar, 20  $\mu$ m. **f** Experimental design used to delete Myomaker and block myonuclear accretion during development is shown. *Myomaker*<sup>loxP/loxP</sup>; *Pax7*<sup>CreER</sup> mice were treated with tamoxifen to generate fusion-incompetent mice (*Myomaker*<sup>scKO</sup>). Control mice were *Myomaker*<sup>loxP/loxP</sup> treated with the same tamoxifen regimen. **g** Representative images of smRNA-FISH for *H19* and *Igf2* (left) from control and *Myomaker*<sup>scKO</sup> TA muscles. Quantification of *H19* and *Igf2* signal is shown on the right. ( $n = 4-5$  independent animals). Scale bar, 20  $\mu$ m. Data are presented as mean  $\pm$  SD. Statistical tests used were (e) two-tailed paired *t*-test; (g) two-tailed unpaired *t*-test; \* $p < 0.05$ , \*\* $p < 0.01$ . Source data are provided as a Source Data file.

number of new myonuclei that have been observed after muscle overload in previous studies<sup>31</sup>. To test if the system fails to capture fusion events when H2B-GFP levels are very low due to proliferation prior to fusion into the myofiber, we administered EdU to  $Pax7^{rtTA}$ ;  $TRE^{H2B-GFP}$  mice during muscle overload (Supplementary Fig. 3e). We observed that 92% of myonuclei were either EdU<sup>+</sup> GFP<sup>+</sup> or EdU<sup>+</sup> GFP<sup>+</sup>, and 8% of myonuclei were EdU<sup>+</sup> GFP<sup>+</sup> (Supplementary Fig. 3e), indicating that the  $Pax7^{rtTA}$ ;  $TRE^{H2B-GFP}$  mice captures the majority of fusion events.

To assess global gene expression in newly fused and pre-existing myonuclei after adult muscle overload we treated  $Pax7^{rtTA}$ ;  $TRE^{H2B-GFP}$  mice with Dox 5 days prior to synergist ablation, then isolated GFP<sup>+</sup> and GFP<sup>+</sup> myonuclei 1 week after surgery (Fig. 3e). bnRNA-seq was performed on myonuclei from sham animals, where minimal MuSC activation and myonuclear accrual occurs, and GFP<sup>+</sup> and GFP<sup>+</sup> populations from mice that received muscle overload. PCA of these populations showed unique signatures of both GFP<sup>+</sup> and GFP<sup>+</sup> myonuclei from muscle overload compared to sham nuclei (from  $Pax7^{rtTA}$ ;  $TRE^{H2B-GFP}$  mice treated with Dox) but we observed a high degree of similarity between the GFP<sup>+</sup> and GFP<sup>+</sup> myonuclear transcriptomes (Fig. 3f). Comparison of GFP<sup>+</sup> or GFP<sup>+</sup> myonuclei to sham myonuclei both reveal hundreds of differentially expressed genes (DEGs) (Fig. 3g). Many of these DEGs are shared, with upregulated genes in both GFP<sup>+</sup> and GFP<sup>+</sup> myonuclei including *Runx1*, *Ankrd1*, *Ankrd2* (Supplementary Data 3, Supplementary Fig. 3f), and are consistent with a response to MOV<sup>56</sup>. We also performed a paired comparison of GFP<sup>+</sup> and GFP<sup>+</sup> myonuclei from overloaded muscles, which revealed only 46 upregulated and 78 downregulated DEGs in GFP<sup>+</sup> myonuclei with FDR < 0.01 (Fig. 3h), consistent with a high degree of similarity between these nuclear populations. bnRNA-seq only detected a relatively low number of DEGs in newly fused nuclei in MOV, which is in stark contrast to the number of DEGs detected in newly fused nuclei during development, suggesting that newly fused nuclei added during development and MOV are distinct. Indeed, PCA of all of the groups profiled by bnRNA-seq revealed that GFP<sup>+</sup> nuclei added during development and MOV exhibit a distinct transcriptional profile (Fig. 3f). Overall, these data indicate that newly fused and pre-existing myonuclei exhibit distinct transcriptional profiles in certain contexts.

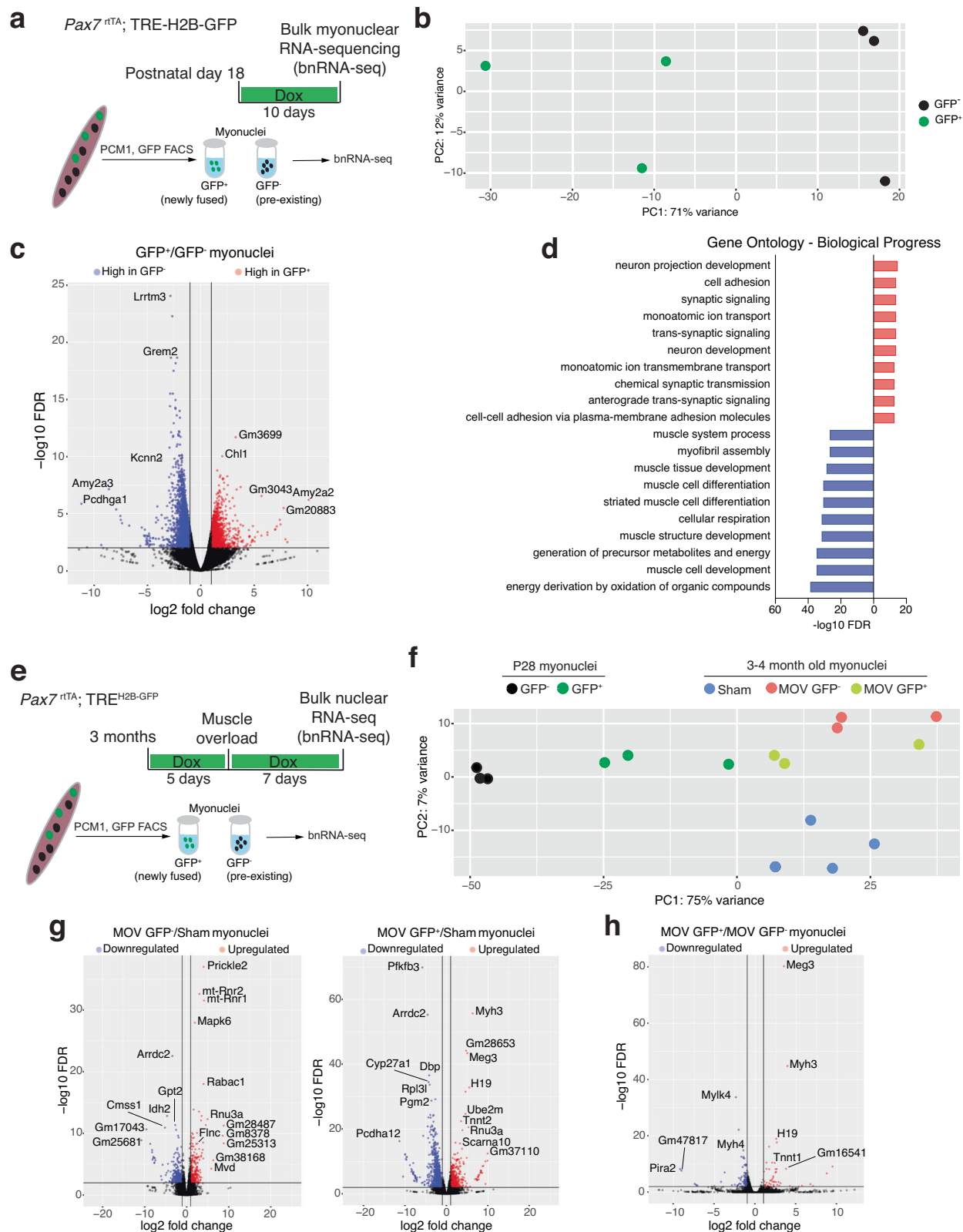
### Divergent contributions of newly fused and pre-existing myonuclei to adult muscle adaptations

While bnRNA-seq did not reveal a robust unique transcriptional signature in GFP<sup>+</sup> myonuclei after MOV in the adult, the upregulated genes in GFP<sup>+</sup> myonuclei included *H19*, *Myh3*, *Myh8*, *Tnnt2*, *Tnnt1* (Supplementary Data 3). We also found genes associated with extracellular matrix including *Postn*, *Col5a1*, *Col3a1*, *Col12a1*, *Col1a1*, and *Fbn1*. Downregulated genes in GFP<sup>+</sup> myonuclei (upregulated in GFP<sup>+</sup> myonuclei) included mature sarcomeric genes such as *Myh4*, *Myh4*, *Nrnp*, and *Xirp2* (Supplementary Data 3). We detected some DEGs in GFP<sup>+</sup> and GFP<sup>+</sup> myonuclei after MOV, but there is also a high level of variation in gene expression between mice subjected to MOV (Fig. 3f),

which could be expected given that MOV surgery can lead to a heterogeneous response. Due to this variation, we deduced that DEGs could be difficult to identify through a paired analysis of GFP<sup>+</sup> and GFP<sup>+</sup> myonuclei, and thus utilized snRNA-seq to determine gene expression in these myonuclear populations after MOV. Here we profiled all myonuclear populations 1 week after synergist ablation using the tet-on  $HSA^{rtTA}$ ;  $TRE^{H2B-GFP}$  system<sup>37</sup>, which results in GFP-labeling of all myonuclei, including newly fused and pre-existing populations (Fig. 4a). This system was chosen over the  $Pax7^{rtTA}$ ;  $TRE^{H2B-GFP}$  system to maximize the number of nuclei obtained for snRNA-seq.

We profiled 11,416 transcriptomes from libraries generated from overloaded plantaris muscle, and 11,182 transcriptomes from libraries generated from the muscle in sham control group ( $HSA^{rtTA}$ ;  $TRE^{H2B-GFP}$  mice treated with Dox). To make sure that the common artifacts were dealt with appropriately in droplet-based transcriptome analysis, we utilized CellBender for removing ambient RNA and Solo for removing doublets from individual samples before integrating the data with Seurat's SCTransform integration workflow. Sham and MOV data were integrated and myonuclear populations that are in specialized compartments including neuromuscular junction (NMJ), myotendinous junction (MTJ), and MTJ-B populations were removed<sup>37-39</sup>. We retained myonuclei not in specialized compartments and performed an additional dimensionality reduction and annotated nuclear clusters based on fiber type (myosin isoform expression) or distinct transcriptional signature (Fig. 4b). Four of these seven clusters, belonging to different myofiber types, were characterized by the expression of canonical markers (type I, *Myh7*; type IIa, *Myh2*; type IIx, *Myh1*; type IIb, *Myh4*) and present in both sham and overloaded muscles (Supplementary Fig. 4a). We also detected two populations only present in overload samples and one population that was expressed in sham but enriched in plantaris muscles after overload (Fig. 4c). Analysis of the percentage of nuclei confirmed that the emerging populations, referred to as unclassified A and B, were mainly present in overloaded muscles, while the enriched cluster, named *Atf3*<sup>+</sup> myonuclei, were present in sham and increased after overload (Fig. 4d). Distinct gene expression was detected in the unclassified A and B populations, including *H19*, *Igf2*, *Runx1*, and embryonic myosin isoforms (*Myh3* and *Myh8*) (Fig. 4e and Supplementary Fig. 3b). We also detected differences between unclassified A and B populations, where *Tnnt2* and *Dclki* are more enriched in unclassified B nuclei (Fig. 4e). The *Atf3* population was not enriched for *H19*, *Igf2*, *Myh3*, or *Myh8* but instead was enriched for *Atf3*, *Flnc*, *Enah*, and *Nrnp* (Fig. 4e, Supplementary Fig. 4b, c), indicating a similar gene signature to that of *Atf3*<sup>+</sup> sarcomere assembly myonuclei observed during postnatal muscle growth<sup>39,58</sup>. Supplementary Data 4 shows genes that mark these populations after MOV.

We next utilized the  $Pax7^{rtTA}$ ;  $TRE^{H2B-GFP}$  system to test if the overload responsive nuclear populations (unclassified A and B, and *Atf3*<sup>+</sup> myonuclei) were derived from newly fused or pre-existing nuclei.  $Pax7^{rtTA}$ ;  $TRE^{H2B-GFP}$  mice received a 5 day doxycycline regimen prior to synergist ablation surgery and were continuously treated with



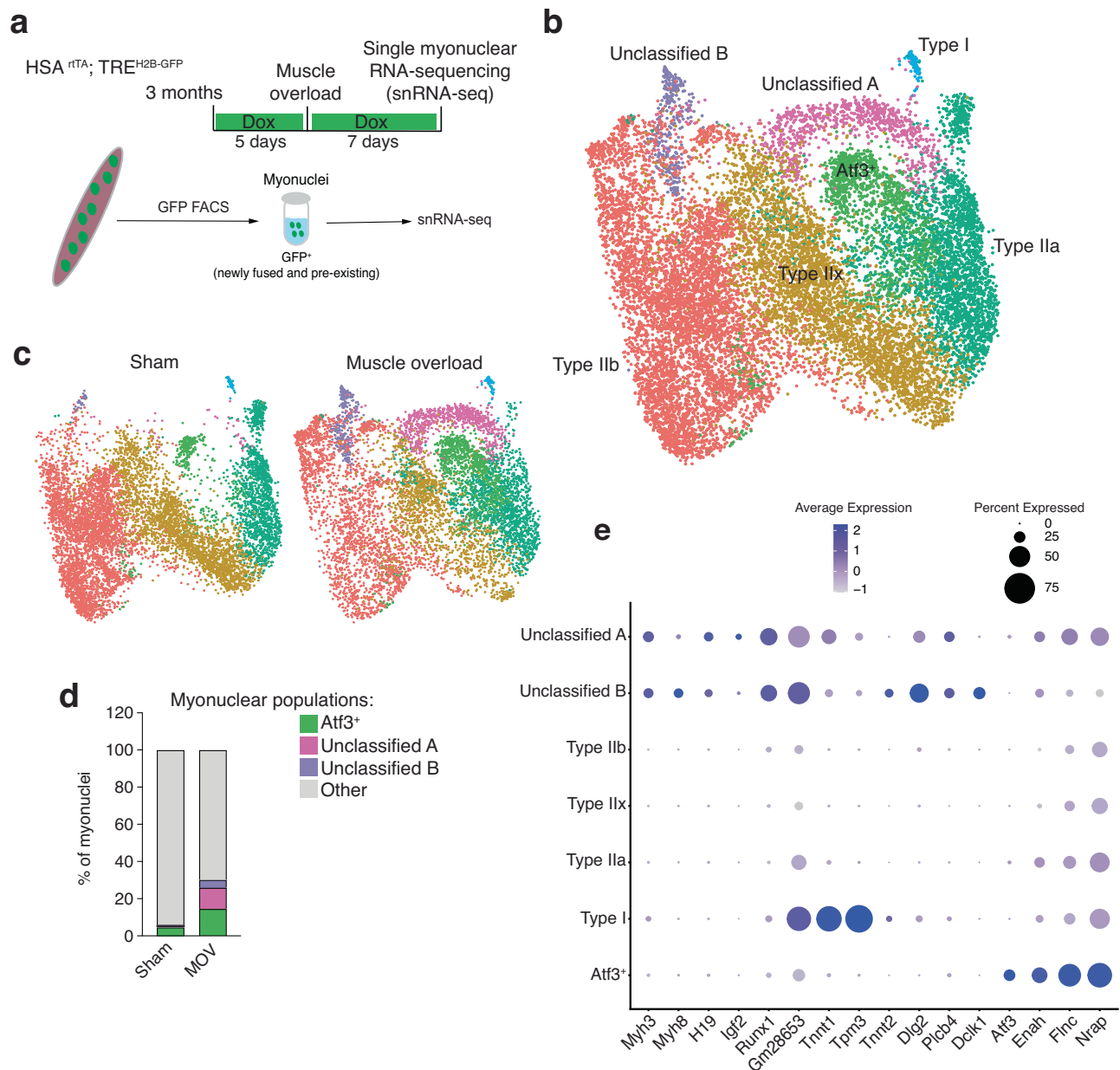
doxycycline until sacrifice 1 week after the procedure (Fig. 5a). Analysis of *H19*, *Igf2*, *Tnnt1*, *Tnnt2*, and *Atf3* transcripts was performed using smRNA-FISH on cross-sections of plantaris muscle with or without synergist ablation. Our results showed an enrichment of *H19*, *Igf2*, *Tnnt1*, *Tnnt2* surrounding GFP<sup>+</sup> myonuclei (Fig. 5a). In contrast, *Atf3* was enriched in GFP<sup>-</sup> myonuclei (Fig. 5a). Genetic blockade of fusion in *Myomk<sup>scKO</sup>* mice reduced levels of *H19*, *Igf2*, *Tnnt1* and *Tnnt2* in

myofibers (Fig. 5b). These data highlight that unclassified A and B populations in muscle overload are recently fused, and the *Atf3* population originates from pre-existing nuclei.

Our analyses of newly fused nuclei during development and after muscle overload suggest that *H19* and *Igf2* are two conserved markers whereas *Tnnt1* (marker of unclassified A and B), *Tnnt2* (marker for unclassified B) and *Myh3* (marker for both unclassified A and B) are

**Fig. 3 | bnRNA-seq analysis comparing newly fused myonuclei acquired during developmental and adult growth.** **a** Schematic diagram of the bulk nuclei RNA-seq of GFP<sup>+</sup> and GFP<sup>-</sup> myonuclei from *Pax7<sup>rtTA</sup>; TRE<sup>H2B-GFP</sup>* mice during postnatal development. TA and gastrocnemius muscles at P28 were used after 10 days of doxycycline treatment. **b** Principal component analysis of the top 500 variable genes from bnRNA-seq of newly fused myonuclei (GFP<sup>+</sup>) and pre-existing myonuclei (GFP<sup>-</sup>). **c** Volcano plot depicting DEGs between newly fused myonuclei (GFP<sup>+</sup>) and pre-existing myonuclei (GFP<sup>-</sup>). **d** Gene Ontology (GO) analysis of the DEGs from (c), showing significantly changed biological processes with a false discovery rate (FDR) < 0.01. Gene lists used for GO analysis are shown in Supplementary Data 2.

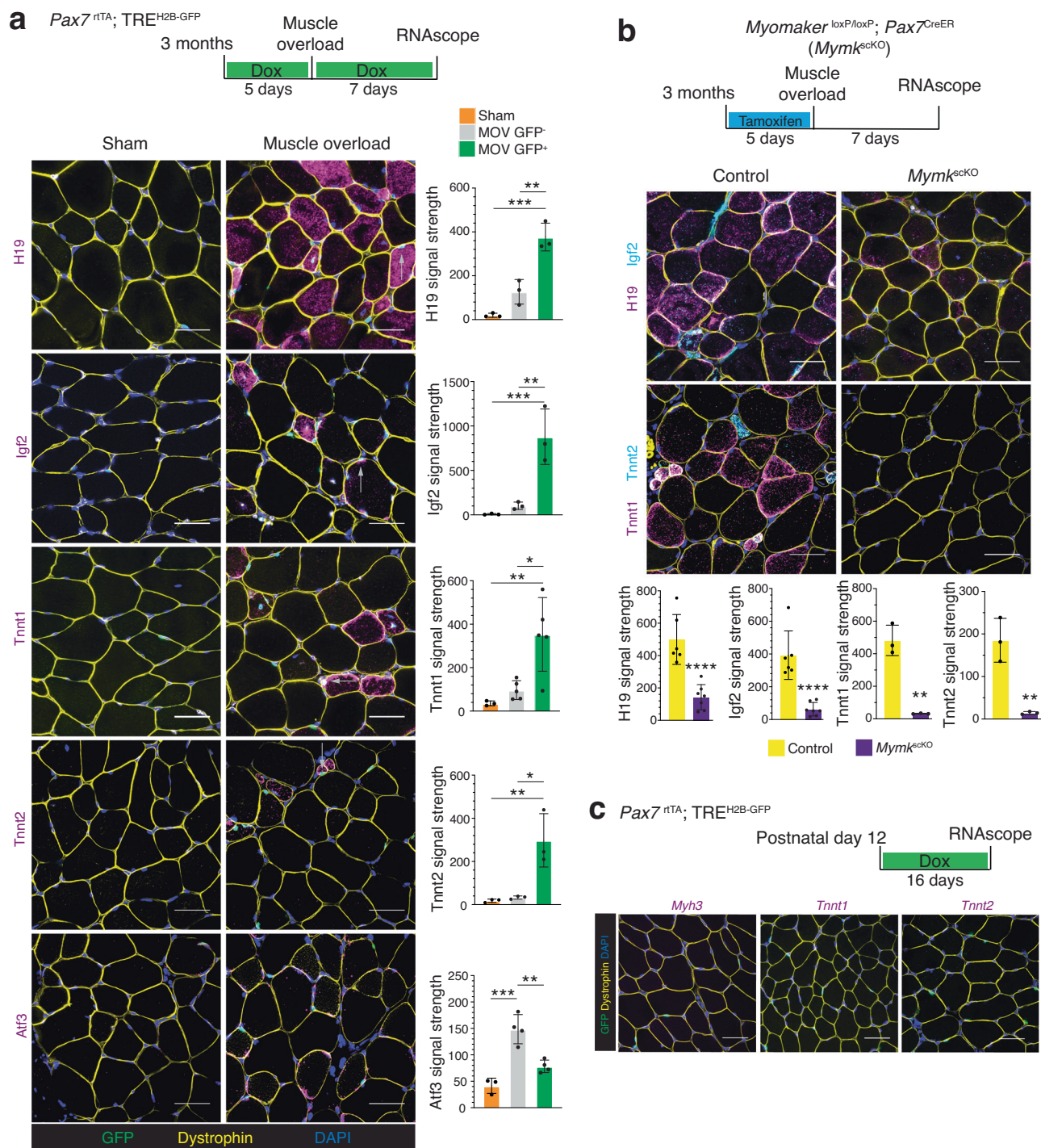
**e** Schematic diagram for bnRNA-seq of GFP<sup>+</sup> and GFP<sup>-</sup> myonuclei from *Pax7<sup>rtTA</sup>; TRE<sup>H2B-GFP</sup>* mice during adult muscle overload. **f** Principal component analysis of bnRNA-seq data from newly fused myonuclei (GFP<sup>+</sup>) and existing myonuclei (GFP<sup>-</sup>) isolated from plantaris muscle 1 week after muscle overload, compared with sham myonuclei (*Pax7<sup>rtTA</sup>; TRE<sup>H2B-GFP</sup>* mice treated with Dox as in (e)) and developmental myonuclei datasets from (b). **g** Volcano plots reveal up- and down-regulated DEGs in newly fused myonuclei (GFP<sup>+</sup>) or pre-existing myonuclei (GFP<sup>-</sup>) compared with sham myonuclei. **h** Volcano plot reveals up- and down-regulated DEGs from comparison of GFP<sup>+</sup> and GFP<sup>-</sup> myonuclei.



**Fig. 4 | Differential contributions of newly fused and pre-existing myonuclei to adult muscle adaptations.** **a** Schematic diagram of the single-nuclei RNA-seq methodology applied to *HSA<sup>rtTA</sup>; TRE<sup>H2B-GFP</sup>* mice, where all myonuclei were labeled with GFP and sorted for subsequent sequencing analysis. **b** Integrated UMAP from plantaris muscle following overload stimulation ( $n = 2$ ) or under sham condition

( $n = 1$ ) (*HSA<sup>rtTA</sup>; TRE<sup>H2B-GFP</sup>* treated with Dox as depicted in (a)). **c** Split UMAP view of myogenic nuclei shown in (b). **d** Proportion of unclassified A, B, and *Atf3*<sup>+</sup> nuclear populations. **e** Dot plot depicts selected genes differentially expressed in various myonuclear populations. Dot size represents the percentage of nuclei expressing the gene. Source data are provided as a Source Data file.





**Fig. 5 | smRNA-FISH analyses validates gene expression in newly fused and pre-existing myonuclei. a** Representative images of smRNA-FISH on sections from *Pax7<sup>rtTA</sup>; TRE<sup>H2B-GFP</sup>* muscle for selected genes enriched in newly fused (*H19*, *IgI2*, *Tnnt1*, *Tnnt2*) or pre-existing myonuclei (*Atf3*). Quantification of the signal for each gene is shown on the right. ( $n = 3$ –5 independent animals). Scale bar, 20  $\mu$ m. **b** Representative images of smRNA-FISH for the indicated genes on sections from control (*Mymk<sup>loxP/loxP</sup>* + tamoxifen) or *Mymk<sup>scKO</sup>* mice subjected to muscle overload. Quantification of the signals are shown on the bottom. ( $n = 3$ –6 independent animals). Scale bar, 20  $\mu$ m. **c** Representative images of smRNA-FISH for *Myh3*, *Tnnt1*, and *Tnnt2* on sections from *Pax7<sup>rtTA</sup>; TRE<sup>H2B-GFP</sup>* muscle during postnatal development. ( $n = 3$ ). Scale bar, 20  $\mu$ m. Data are presented as mean  $\pm$  SD. Statistical tests used were (a) one-way ANOVA with Tukey's correction for multiple comparisons; (b) two-tailed unpaired *t*-test; \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ . Source data are provided as a Source Data file.

unique for muscle overload. smRNA-FISH for *Tnnt1*, *Tnnt2* and *Myh3* on muscle sections from *Pax7<sup>rtTA</sup>; TRE<sup>H2B-GFP</sup>* mice where newly fused myonuclei were labeled from P12 to P28 revealed no expression (Fig. 5c), confirming that these genes are specific markers of newly fused nuclei during MOV.

### The syncytial muscle environment is a determinant of gene expression in newly fused myonuclei

Multiple observations described thus far suggest newly fused myonuclei exhibit differential gene expression. First, PCA of bnRNA-seq data of myonuclear populations during development and MOV

revealed separation of newly fused nuclei based on condition (Fig. 3f). Second, two newly fused nuclear populations (unclassified A and B) were detected by snRNA-seq after MOV (Figs. 4c and 5a). Third, *Tnnt1*, *Tnnt2*, and *Myh3* were expressed in newly fused nuclei during MOV but not during development (Fig. 5c). We next sought to uncover how this divergent gene expression of newly fused nuclei is controlled, which could include that the stimulus directly entrains gene expression in myogenic progenitors prior to fusion or that the syncytial environment to which the new nucleus is fusing into could impact its transcriptional trajectory. To parse apart these possibilities we took advantage of the diversity of cell fusion events during muscle overload, where myoblast-myoblast fusions generate de novo myofibers and myoblast-myofiber fusions result in addition of new nuclei to established myofibers<sup>31,34,59</sup>. In this situation, myogenic progenitors are generally subjected to the same stimulus (increased load) but are fusing into different syncytial environments. Recent evidence indicates that fusion to form de novo myofibers is rapid and thus the nuclei are fusing into an immature syncytium<sup>5</sup>. In contrast, during myoblast-myofiber fusions myogenic progenitors enter an established syncytium governed by mature myonuclei.

The presence of nuclei in the center of myofibers is a hallmark characteristic of de novo myofibers and repair of pre-existing myofibers. However, a long-standing question is whether newly fused myonuclei can be deposited in the periphery of myofibers. The data presented in Fig. 5 highlight that new nuclei added to myofibers after MOV can be peripherally located. While we observed mainly central GFP<sup>+</sup> myonuclei in small myofibers (potentially de novo), we found central and peripheral GFP<sup>+</sup> myonuclei in larger myofibers (potentially pre-existing). This model also allowed us to test if *Myh3*, a classical marker of de novo myofibers is also present in larger myofibers that acquired a new nucleus<sup>60</sup>. We detected *Myh3* mRNA in newly fused myonuclei regardless of whether they are in a central or peripheral location (Fig. 6a, b). While *Myh3* is not a distinguishing marker of central or peripheral newly fused nuclei after muscle overload, we observed distinctions in *Tnnt2* expression patterns, indicating a divergent transcriptional signature. *Tnnt2* expression, a marker of the unclassified B population, was restricted to smaller, potentially de novo myofibers after muscle overload (Fig. 5b). In contrast, *Tnnt1*, a marker of the unclassified A population, was present in both smaller and larger myofibers (Fig. 5b), similar to *Myh3*. These data suggest that the unclassified B population represent nuclei from de novo myofibers, which is further supported by data that *Tnnt2* is enriched in embryonic myonuclei compared to postnatal myonuclei<sup>40</sup>. However, our characterization of de novo and pre-existing myofibers are only based on size, thus we are unable to definitively conclude that *Tnnt2*<sup>+</sup> myofibers are de novo. We directly compared transcriptomes of unclassified A and unclassified B myonuclei to the transcriptomes of mature myonuclei that are responding to MOV. We found that unclassified B has a more distinct transcriptional signature compared to unclassified A, which exhibits a range of gene expression resembling other myonuclei (Fig. 6c). Unclassified A and B myonuclei are subjected to the same MOV stimulus as progenitors and yet exhibit distinct gene expression patterns, suggesting that the stimulus itself does not solely explain differential gene expression of newly fused myonuclei in this context, and one of the determinants is the syncytial environment itself.

### Cellular fusion coordinates gene expression of myonuclei after muscle overload

The response of myofibers to muscle overload involves activation of the expression of *Atf3*, *Flnc*, and *Enah* in resident myonuclei, accretion of additional nuclei that express *Tnnt1* into pre-existing myofibers, and generation of smaller myofibers whose nuclei express *Tnnt2*. Given this combinatorial response of newly fused and pre-existing myonuclei, that myonuclear accretion is required for proper

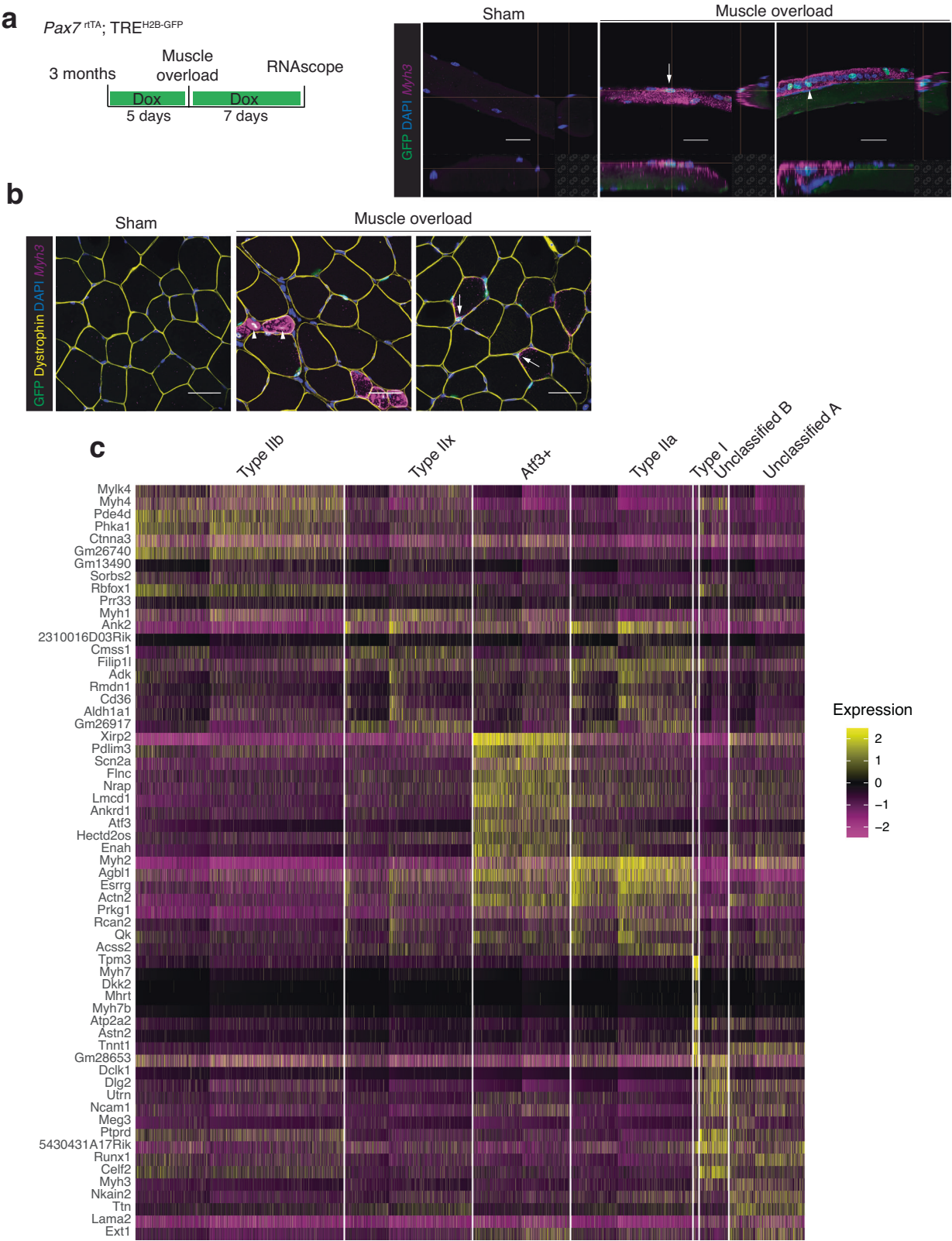
adaptations of muscle to increased load, and an apparent role of the syncytium in impacting gene expression in newly fused nuclei described here, we asked if the myonuclear accretion process globally impacted gene expression in pre-existing nuclei during adult muscle overload.

We tested this concept by analyzing the transcriptional response of pre-existing myonuclei to muscle overload in fusion-defective *Mymk*<sup>scKO</sup> mice (Fig. 7a). Using bnRNA-seq, we compared gene expression between sham myonuclei (*Pax7*<sup>Cre</sup>; TRE<sup>H2B-GFP</sup> (+Dox) mice or *Mymk*<sup>loxP/loxP</sup> (+tamoxifen) that did not receive muscle overload), GFP<sup>+</sup> and GFP<sup>-</sup> myonuclei from *Pax7*<sup>Cre</sup>; TRE<sup>H2B-GFP</sup> mice after muscle overload (referred to as control MOV), and *Mymk*<sup>scKO</sup> myonuclei after muscle overload. PCA revealed that the transcriptional signature in *Mymk*<sup>scKO</sup> myonuclei were distinguished from both sham and control MOV GFP<sup>+</sup>/GFP<sup>-</sup> myonuclei (Fig. 7b). To ascertain the effect of myonuclear accretion on pre-existing myonuclei during muscle overload, we used a scatter plot to identify differential gene expression (Fig. 7c, d) between MOV myonuclei (GFP<sup>-</sup> + *Mymk*<sup>scKO</sup>) vs. sham myonuclei (y-axis) and control GFP<sup>-</sup> MOV vs. *Mymk*<sup>scKO</sup> MOV (x-axis). Note that only GFP<sup>-</sup> myonuclei were used from control MOV because our goal is to understand the effects on pre-existing myonuclei in the absence of fusion, and we merged GFP<sup>-</sup> MOV vs. *Mymk*<sup>scKO</sup> MOV for the comparison to sham on the x-axis to identify all genes changed during MOV with and without fusion. A 0.5 log2 fold change was used for this analysis and revealed genes that are upregulated (Fig. 7c) and downregulated (Fig. 7d) after muscle overload.

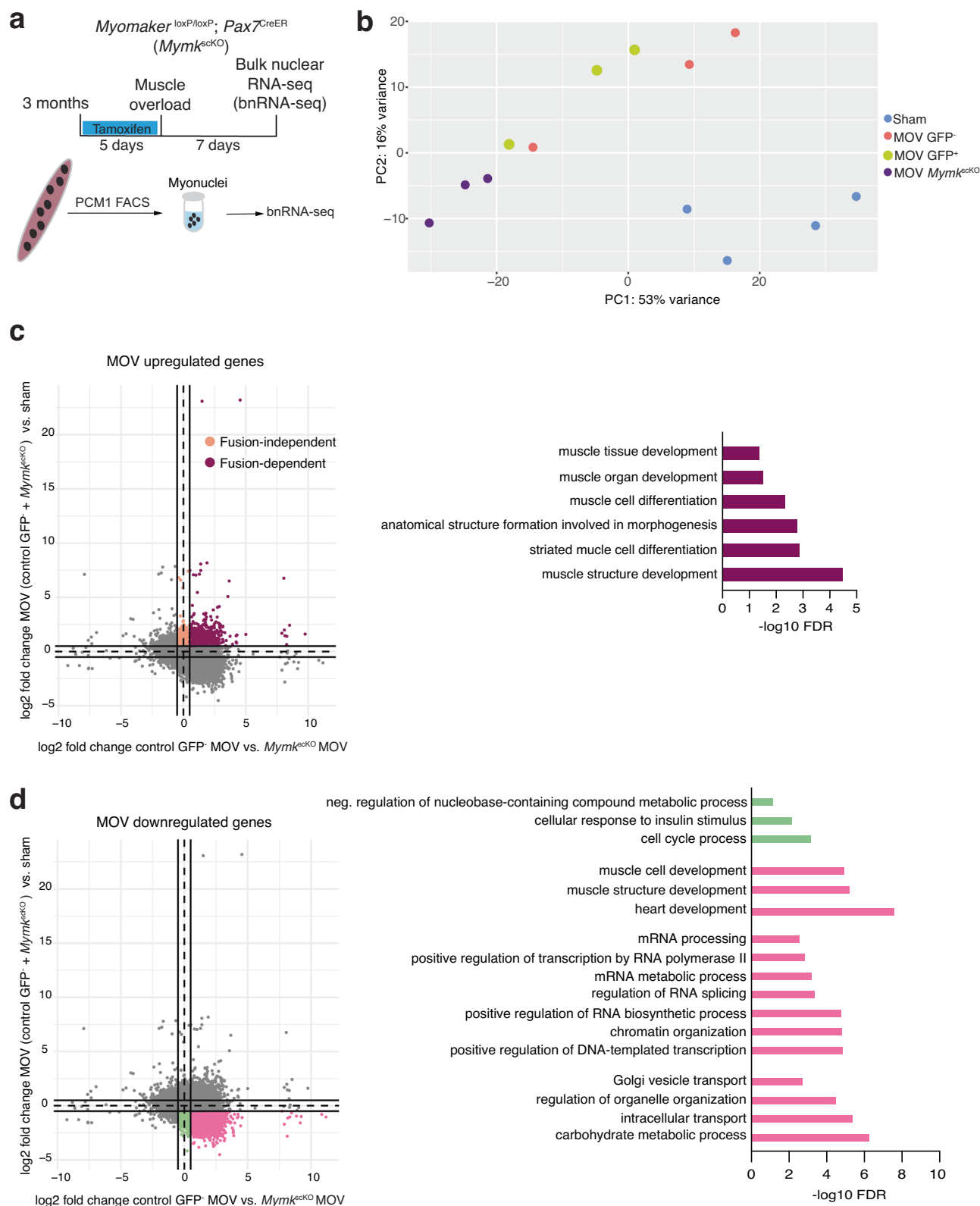
We found upregulated genes in MOV that were increased in GFP<sup>-</sup> myonuclei compared to *Mymk*<sup>scKO</sup> myonuclei (Fig. 7c, maroon dots). These genes are categorized as fusion-dependent because they are not as highly activated in overloaded muscle in the absence of fusion. Gene ontology analysis on this group of genes (0.1 FDR, 122 genes) revealed processes associated with muscle development and differentiation (Fig. 7c, Supplementary Data 5). We also found upregulated genes in MOV that exhibited no change between GFP<sup>-</sup> myonuclei and *Mymk*<sup>scKO</sup> myonuclei (Fig. 7c, peach dots), and we categorized these genes as fusion-independent. Gene ontology analysis on these genes (0.1 FDR, 751 genes, Supplementary Data 5) did not reveal significant biological processes indicating that nuclear accretion more specifically aids in the response to muscle processes within resident myonuclei during muscle overload. We did not analyze MOV upregulated genes that were reduced in GFP<sup>-</sup> myonuclei compared to *Mymk*<sup>scKO</sup> myonuclei because it is unclear if these are relevant to a normal MOV response in myofibers undergoing fusion. Overall, these data indicate that pre-existing myonuclei need myonuclear accretion for a normal transcriptional response to muscle overload.

We also uncovered downregulated genes after MOV (Fig. 7d), and 634 of these genes (0.1 FDR) did not change between GFP<sup>-</sup> and *Mymk*<sup>scKO</sup> myonuclei (green dots, Supplementary Data 5). We interpret that these genes are normally downregulated in response to MOV regardless of fusion status, and gene ontology analysis revealed three significant terms including cell cycle process, regulation of insulin stimulus and nucleobase-containing processes (Fig. 7d). We also detected an additional 780 genes (0.1 FDR), the expression of which was decreased in resident myonuclei when myonuclear accretion is blocked (Fig. 7d, pink dots, Supplementary Data 5). Gene ontology analysis highlighted muscle-related processes, multiple pathways associated mRNA regulation, and intracellular processes related to metabolism and organelles. This response is divergent from the normal downregulated response to MOV and suggests that in the absence of myonuclear accretion, myofibers not only fail to activate muscle-specific pathways needed for adaptations to muscle overload but there is also an overall reduction of genes positively associated with biosynthetic output. Thus, the fusion process not only contributes unique transcripts it also aids in the coordination of gene expression from pre-existing myonuclei in the syncytium.





**Fig. 6 | Comparative gene features of newly fused myonuclei.** **a** Representative images of smRNA-FISH for *Myh3* on plantaris myofibers from *Pax7*<sup>rtTA</sup>; TRE<sup>H2B-GFP</sup> mice undergoing overload stimulation or under sham conditions. *Myh3* transcription was detected in both peripheral (white arrow) and central (arrowhead) newly fused myonuclei. (*n* = 3–4). Scale bar, 20 μm. **b** Representative images of smRNA-FISH for *Myh3*, performed on the same mice used in (A), show that both central newly fused GFP<sup>+</sup> myonuclei (arrowhead) and peripheral newly fused GFP<sup>+</sup> (white arrow) are enriched for *Myh3* transcription. (*n* = 3–4). Scale bar, 20 μm. **c** Heatmap of genes showing distinct transcriptional signature in nuclei of myogenic populations from plantaris muscle subjected to 1 week of overload stimulation, as used in Fig. 4c (*n* = 2).



## Discussion

Although there are multiple syncytial cell types, skeletal myofibers are unique in that they are long-lived cells that not only accrue hundreds of nuclei during development but also can add new nuclei in the adult in response to various stimuli. It has been established that the number of nuclei in skeletal myofibers is a major factor for size and function, and adult myofibers require new nuclei for adaptations<sup>29–33,61</sup>, which

has led to questions about how new nuclei and pre-existing nuclei synergize to maintain a functional syncytium. The main rationale for this study was to determine if newly fused nuclei are transcriptionally distinct from the pre-existing myonuclei. If so, it would suggest that newly fused nuclei possess a transcriptional machinery independent from what is encountered as part of the syncytium. If newly fused nuclei were similar to pre-existing nuclei, it would suggest that they are



**Fig. 7 | Fusion and myonuclear accrual is needed for pre-existing myonuclei to transcriptionally respond to muscle overload.** **a** Schematic diagram of the bulk nuclei RNA-seq methodology for analyzing pre-existing myonuclei adapting to muscle overload (MOV) in the absence of myonuclear accretion. *Myomaker*<sup>loxP/loxP</sup>; *Pax7*<sup>CreER</sup> mice were treated with tamoxifen to generate fusion-incompetent mice (*Myomaker*<sup>loxP/loxP</sup>). Note that three sham mice were *Pax7*<sup>CreER</sup>; TRE<sup>H2B-GFP</sup> (+Dox) and one sham animal was *Myomaker*<sup>loxP/loxP</sup> (+tamoxifen). **b** Principal component analysis of bulk RNA-seq from *Myomaker*<sup>loxP/loxP</sup> myonuclei isolated from plantaris muscle 1 week after muscle overload, myonuclei (GFP<sup>+</sup> and GFP<sup>-</sup>) derived from control mice that underwent the same surgery, and myonuclei from sham mice. **c** Scatter plot

highlighting genes upregulated in both control muscle overload (MOV) (GFP<sup>+</sup> pre-existing myonuclei) and *Myomaker*<sup>loxP/loxP</sup> MOV myonuclei (peach dots, fusion independent) or upregulated in only GFP<sup>+</sup> pre-existing myonuclei (maroon dots, fusion dependent). Gene Ontology analysis of the genes upregulated that are fusion-dependent (maroon dots) is shown on the right. **d** Scatter plot highlighting genes downregulated in both control muscle overload (MOV) (GFP<sup>+</sup> pre-existing myonuclei) and *Myomaker*<sup>loxP/loxP</sup> MOV myonuclei (green dots) or downregulated in only *Myomaker*<sup>loxP/loxP</sup> MOV myonuclei (pink dots). Gene Ontology analysis of both sets of genes is shown on the right.

entrained to become a functioning part of the syncytium. Our results reveal that transcriptional profiles of newly fused nuclei reflect a confluence of both independence and syncytial assimilative forces, both of which are relatively under-discussed aspects of skeletal muscle biology but could represent nodal processes for the regulation of skeletal muscle growth and function.

In this study we overcome a major obstacle in understanding gene expression in myonuclear populations through the development of a lineage tracing model in the mouse that distinguishes newly fused and pre-existing myonuclei. We identify conserved markers of newly fused myonuclei, including *H19* and *Igf2*, which are present in this population during both developmental and adult accrual. Our work also establishes that newly fused myonuclei added to the syncytium during development are distinct from those added during adult load-induced muscle adaptations. We propose a model where the transcriptional trajectory of newly fused nuclei are impacted by the syncytial environment, which is controlled in part by the pre-existing nuclei. Reciprocally, myonuclear accretion is needed for pre-existing myonuclei to appropriately respond to an overload stimulus, indicating a bilateral interaction mechanism between newly fused and pre-existing myonuclei. In addition to defining gene expression of newly fused and pre-existing myonuclei in response to various stimuli, this work establishes an interaction between these nuclear populations in the myofiber syncytium.

What is the purpose of myofibers adding new nuclei during development and adult adaptations? The consequences of cellular fusion in these two scenarios should be considered separately since the physiological and functional activities of myofibers are different and our results show that new nuclei are molecularly distinct. In a previously reported model where fusion is blocked early during postnatal development, remaining myonuclei are able to transcriptionally compensate and support growth<sup>46</sup>. In contrast, during adult muscle hypertrophy, a fusion blockage results in a lack of growth, indicating adult resident myonuclei are unable to fully compensate<sup>31–33</sup>. When asking about unique transcriptional contributions of a newly fused myonucleus to a syncytia, one must first determine the bioavailability of certain factors already present within that syncytium. During development, we found that newly fused nuclei are more transcriptionally distinct compared to other nuclei within the syncytium, whereas after muscle overload new nuclei are more similar to pre-existing myonuclei.

For adult muscle adaptations, one idea is that newly fused nuclei are highly specialized and deliver specific transcripts needed for a particular aspect of the adaptation. We confirmed that myonuclear accretion is essential for muscle adaptations in the adult, but our results did not yield an obvious single factor contributed by newly fused nuclei that could explain their necessity for growth. *H19* and *Igf2* are candidate factors since they have known roles in myogenesis and skeletal muscle<sup>47,51,62–65</sup>, but their function in myofibers after fusion remains to be determined. The potential significance that *H19* and *Igf2* are co-transcribed from the same locus in an imprinted manner is unclear. New models will be needed to uncouple roles of these factors during myogenesis from potential roles in myofibers after contribution of newly fused nuclei. In newly fused nuclei after overload, we also

detect enrichment of *Tnnt* genes, which are preferentially expressed in different muscle fiber types. *Tnnt1* is expressed in slow myofibers and *Tnnt2* is expressed in cardiac muscle and embryonic skeletal myofibers. Given that *Tnnt* genes encode for Troponin T, which governs the initiation of muscle contraction, and that *Tnnt1* is associated with slow-type myofibers, we speculate that expression of this gene allows short-term compensation of contraction in de novo and pre-existing myofibers during the response to muscle overload. It should be noted that in our experiments we did not discern fusion with different fiber types, and that could be regulating transcriptional trajectory of newly fused nuclei.

An alternative interpretation is that the main contribution of new nuclei for muscle adaptations is to increase the number of DNA copies rather than to contribute unique transcripts that perform a specific function<sup>66</sup>. However, it is possible that a transcript that currently lacks functional annotation detected in newly fused nuclei does perform a critical function. While new nuclei are providing unique transcripts to the syncytium, these transcripts may not be needed for a specific response to the stimulus but instead coordinate a synergistic response to the stimulus by impacting pre-existing nuclei. Support for the concept that there is global sensing of cellular fusion in myofibers during adult muscle adaptations includes data presented here where pre-existing nuclei transcriptionally decompensate to levels below that of non-stimulated adult myonuclei when not able to accrue new nuclei during overload. These data are consistent with a report that pre-existing myonuclei exhibit divergent gene expression after depletion of MuSCs in response to an 8-week load stimulus<sup>28</sup>, and further clarify that fusion and myonuclear accrual is the cellular process that pre-existing myonuclei need to mount a functional response to increased load. Our work here highlights the presence of a global sensing mechanism in myofibers where pre-existing myonuclei transcriptionally respond to the accrual of new nuclei, which could include direct communication between the nuclei or indirectly sensing alterations in the syncytium due to lack of new nuclei. However, we cannot exclude the possibility that external factors, for instance continued mechanical stress on myofibers due to overload, contribute to the dysregulated response when myofibers are not able to accrue new nuclei.

One unanticipated conclusion from our study, which in retrospect should have been obvious given the strong reprogramming abilities of myonuclei<sup>67,68</sup>, is the extent to which the syncytial environment impacts the trajectory of newly fused nuclei. While myogenic progenitors may possess intrinsic differences based on age of the animal, muscle type, and presence of external stimuli<sup>30,69,70</sup>, leading to heterogeneous transcriptional states that could impact the molecular trajectory once it enters myofibers, our data highlight a role for the syncytium as well. The first observation indicating that the syncytial environment impacts newly fused nuclei is that the transcriptional distinctiveness of newly fused nuclei compared to other nuclei within the syncytium was most evident during development. Since there is less of a requirement for new nuclei for muscle growth at this stage given that pre-existing nuclei exhibit an ability to increase output to support larger cytoplasmic volumes, we suspect that the transcriptional differences are due to the overall state of the syncytium during

development. Indeed, myonuclei are in a more immature state in the postnatal stage<sup>40</sup>, where there is likely not a sufficient quantity of mature muscle transcripts to reprogram newly added myonuclei. Another observation supporting a role for the syncytium is that after MOV we detected transcriptional differences in myonuclei that were added to already formed myofibers compared to myonuclei in de novo myofibers. After MOV, newly fused myonuclei in de novo myofibers, which have a recently established syncytium, were enriched for *Tnnt2* and *Dclki1* transcripts that also mark embryonic myonuclei<sup>40</sup>. In contrast, newly fused myonuclei in pre-existing myofibers after MOV exhibited transcriptomes similar to those myonuclei that were already present in the syncytium prior to MOV, further suggesting a role for the syncytial environment in regulating the trajectory of newly fused nuclei. It should be noted that the stimulus itself could entrain gene expression of myogenic progenitors prior to fusion, and we are not excluding that as a regulatory mechanism. However, the presence of divergent transcription observed in nuclei added to pre-existing myofibers after MOV compared to those in new myofibers implicates the syncytial environment since these progenitors were subjected to the same load increases. Regardless of the molecular details, we propose that the syncytial environment is one determinant of gene expression in newly fused myonuclei.

In summary, a unique and exclusive function of newly fused myonuclei remains to be determined, but the availability of the lineage tracing model developed here and the subsequent identification of transcriptional markers of newly fused nuclei opens up the possibility to more deeply explore this issue. Our results unexpectedly reveal that newly fused and pre-existing myonuclei reciprocally influence each other to establish the myofiber syncytium and respond to adaptations in adult muscle.

## Methods

Our research complies with all ethical regulations. The study protocols were approved by the Biosafety Committee and IACUC at Cincinnati Children's Hospital Medical Center.

## Animals

Animal studies were conducted in accordance with guidelines from the National Institutes of Health and the Division of Veterinary Services (DVS) at Cincinnati Children's Hospital Medical Center. The IACUC protocol number for this study is #2020-0047. The transgenic mouse lines utilized in this research include *Pax7*<sup>CreER71</sup>, *Myomaker*<sup>loxP/loxP53</sup>, *Pax7*<sup>rtTA</sup>, *HSA*<sup>rtTA57</sup>, and *TRE*<sup>H2B-GFP43</sup>. All mice were maintained on a C57BL/6 genetic background. The *Myomaker*<sup>loxP/loxP</sup>, *Pax7*<sup>CreER</sup> mice (*Myrk*<sup>scKO</sup>) used in this study were previously described<sup>31</sup>. Tamoxifen (75 mg/kg body weight), dissolved in corn oil with 10% EtOH, was administered to mice through intraperitoneal injection.

The *Pax7*<sup>rtTA</sup> mice were generated by Dr. Christoph Lepper from The Ohio State University. In brief, the targeting construct for homologous recombination in C57BL/6;129sv hybrid ES cells was generated via recombineering method<sup>72</sup>. A BAC clone (C57BL/6) containing the entire *Pax7* locus was obtained from BACPAC Genomics. A chimeric gene cassette consisting of a GPI-anchored YFP fused to HaloTag® (Promega) followed by the reverse tTA (rtTA) gene (Tet-On 3 G, Takara Bio), separated via self-cleaving T2A peptide was generated via Gibson Assembly (NEB). This cassette including a 5' chimeric intron and 3' polyadenylation signal sequence replaces the sequence of *Pax7* exon 1 immediately 3' of the ATG start codon. For selection in ES cells, a male germline self-excising Ace-Cre-Neo cassette replaces *Pax7* exon 2. This *Pax7* targeting vector was used for homologous recombination in C57BL/6;129sv hybrid ES cells. Chimeric animals were produced via blastocyst injection of confirmed recombined ES cell clones. Germline transmission of the *Pax7*<sup>rtTA</sup> allele was confirmed via genomic PCR. The *HSA*<sup>rtTA</sup> mice<sup>57</sup> were provided by Dr. John McCarthy from the University of Kentucky. For *TRE*<sup>H2B-GFP</sup> expression, mice were given chow

supplemented with 0.0625% doxycycline (TestDiet) at the indicated time points.

## Synergist ablation (muscle overload)

Plantaris muscles were subjected to muscle overload by implementing bilateral synergist ablation of the soleus and gastrocnemius muscles, following the procedure described previously<sup>31</sup>. An incision on the lower limb's posterior-lateral side was made to expose the soleus and gastrocnemius muscles. The distal and proximal tendons of the soleus were cut followed by cutting of the distal gastrocnemius tendon and excision of 75% of the lateral and medial gastrocnemius. EdU (APEX-BIO) was dissolved in PBS at a concentration of 1 mg/mL and administered intraperitoneally (IP) to mice at a dose of 100 µl per mouse. Both surgery and sham groups received daily IP injections of EdU for 1 week following surgery.

## Isolation of nuclei

Following euthanasia, tibialis anterior, gastrocnemius, or plantaris muscles were dissected, minced, and placed in homogenization buffer (0.25 M sucrose and 1% BSA in Mg<sup>2+</sup>-free, Ca<sup>2+</sup>-free, RNase-free PBS). An Ultra-Turrax T25 was utilized for further homogenization of the minced tissue. The homogenate was incubated for 5 min. upon the addition of Triton-X100 (2.5% in RNase-free PBS, added at a 1:6 ratio). All samples were filtered through a 100 µm strainer, centrifuged (3000 × g for 10 min at 4 °C) to yield a crude pellet, and resuspended in sorting buffer (2% BSA/RNase-free PBS) before being filtered again via a 40 µm strainer. Nuclei were labeled with Hoechst dye and 0.2 U/µl Protector RNase inhibitor (Roche 03353399001). The labeled nuclei were purified using FACS (BD Aria, 70 µm nozzle) and gathered in sorting buffer containing Protector RNase inhibitor (0.2 U/µl).

## Immunofluorescence

Muscle sections were fixed using 1% PFA for a duration of 2 min., followed by a wash in PBS. The sections were then permeabilized in a solution of 0.2% Triton X-100/PBS for 10 min., subsequently washed in PBS, and then blocked with a mixture of 1% BSA, 1% heat-inactivated goat serum, and 0.025% Tween20/PBS for an hour at room temperature in a humidity chamber. All slides were later incubated with rabbit anti-dystrophin primary antibodies (1:100; Abcam Ab15277) overnight at 4 °C in a humidity chamber. Following the primary antibody incubation, the slides were rinsed in PBS, and then treated with goat anti-rabbit Alexa Fluor 555 secondary antibody (1:100; Invitrogen A31572) for 45 min. at room temperature in a humidity chamber. The slides were then washed with PBS and mounted using VectaShield containing DAPI (Vector Laboratories H-1200). We used the Click-iT™ Plus EdU Cell Proliferation Kit for Imaging (ThermoFisher scientific, C10640) to detect EdU positive nuclei. Upon the completion of immunostaining, images were captured using a Nikon AIR confocal microscope.

## smRNA-FISH on cryosections

Single-molecule FISH assays were conducted using the RNAscope Multiplex Fluorescent Reagent Kit v2 (ACDBio 323110), with slight modifications to the standard protocol. Fresh-frozen cross-sections of tibialis anterior (TA) or plantaris muscles were used as specified. Briefly, 12 µm-thick sections were obtained from frozen muscle samples, air-dried for 10 min. at room temperature (RT), and then immersed in cold 4% paraformaldehyde (PFA) for 20 min. The slides were dehydrated sequentially in 50%, 70%, and 100% ethanol for 5 min. each at RT. Dehydrated slides were preserved in 100% ethanol at -20 °C until needed.

All RNAscope analyses on cryosections were co-stained with dystrophin antibodies for analysis of transcriptional signal in myonuclei. To perform immunofluorescence staining, sections were blocked using a solution containing 1% bovine serum albumin, 1% heat-inactivated goat serum, and 0.025% Tween-20 in phosphate-buffered

saline for 30 min. at RT. The sections were then incubated with anti-dystrophin antibodies (1:100, Abcam Ab15277) overnight at 4 °C, followed by two washes with 1xPBS, 5 min. each. Subsequently, Alexa-Fluor secondary antibodies (1:100, Invitrogen A21206, A31572, A21244) were applied at room temperature for 1 hour, followed by two washes with 1xPBS, 5 min. each, and a post-fixation step with 4% PFA for an additional 20 min. All slides were then subjected to a 10 min. incubation in hydrogen peroxide at RT, followed by a 20 min. incubation with Protease III at RT. Between each incubation, slides were washed with 1xPBST for 5 min. Slides were then hybridized with selected probes (ACDBio, Mm-Atf3 426891-C2, Mm-H19 423751, Mm-Igf2 437671-C3, Mm-Tnnt1 466911-C2, Mm-Tnnt2 418681-C3, Mm-Myh3 498851-C2) at 40 °C for 2 hours. The remaining AMP hybridization and HRP incubation steps were performed according to the instructions provided in the ACDBio user manual.

The intensity of the smRNA-FISH signal was assessed using two distinct analytical pipelines in *Pax7<sup>CreTA</sup>; TRE<sup>H2B-GFP</sup>* or *Myomk<sup>scKO</sup>* mice. In *Pax7<sup>CreTA</sup>; TRE<sup>H2B-GFP</sup>*, the median smRNA-FISH signal intensity was determined within a specified Region of Interest (ROI), each being a 25 square micrometer circular area encompassing either newly formed myonuclei (GFP<sup>+</sup>) or pre-existing myonuclei (GFP<sup>-</sup>). 10–20 ROIs from each image were included in the analysis. For the analysis of *Myomk<sup>scKO</sup>* mice and their littermate controls, the median smRNA-FISH signal strength was measured across the entire fiber area within dystrophin-labeled myofibers. NIS Elements software (Nikon) was employed for all signal strength analyses.

### smRNA-FISH on isolated myofibers

Extensor digitorum longus (EDL) or plantaris muscles were harvested and incubated in high-glucose DMEM (HyClone Laboratories SH30022.01) containing 0.2% (EDL muscle) or 0.4% (plantaris muscle) collagenase type I (Sigma-Aldrich C010-1G) at 37 °C in a cell culture incubator. Following a 30 min. incubation, muscles were gently triturated using a wide-bore glass pipette to facilitate myofiber release before returning to the incubator. After a maximum of 1 hour of incubation, muscles were triturated until myofibers detached from the tissue. Isolated myofibers were collected, fixed in 4% PFA/PBS for 20–30 min. at room temperature, and rinsed in PBS. Myofibers were then arranged on Cell-tak adhesive (Corning 354240)-coated slides and transferred directly to 100% methanol, stored at –20 °C until needed.

For rehydration, slides were washed in a descending methanol/PBS + 0.1% Tween-20 (PBST) series (50% MeOH/50% PBST; 30% MeOH/70% PBST; 100% PBST) for 5 min each. All slides underwent a 10 min. hydrogen peroxide incubation, followed by a 20 min. Protease III incubation at room temperature. Slides were washed with 1xPBST for 5 min. between incubations. Selected probes were used to hybridize slides at 40 °C overnight. The ACDBio user manual provided guidance for the remaining AMP hybridization and HRP incubation procedures.

### Single nuclear RNA-sequencing (snRNA-seq)

Plantaris muscles from *HSA<sup>CreTA</sup>; TRE<sup>H2B-GFP</sup>* mice were used for snRNA-seq in this study. Four plantaris muscles, pooled from two mice that either received or did not receive synergist ablation, were used for nuclei isolation using the protocol described above. After filtration via a 40 µm strainer, nuclei were labeled with Hoechst dye and 0.2 U/µl Protector RNase inhibitor (Roche 03335399001). All GFP<sup>+</sup> Hoechst stained myonuclei were purified using FACS (BD Aria, 70 µm nozzle) and gathered in sorting buffer containing Protector RNase inhibitor (0.2 U/µl). The nuclei were counted with a hemocytometer, and the concentration was fine-tuned as needed to reach the optimal range for the 10X Chromium chip. The 10X Chromium system was then used to load the nuclei using the Single Cell 3' Reagent Kit v3.1, following the guidelines provided by the manufacturer. Around 12,000 nuclei were loaded for each operation. Sequencing was performed on an Illumina NovaSeq 6000 System.

### Bulk nuclear RNA-sequencing (bnRNA-seq)

To obtain sufficient material for sequencing, skeletal muscles were pooled together for this analysis. For the early development stage, 4 × Gastrocnemius and 4 × tibialis anterior muscles per replicate were used, while for adult muscle overload, 4 × Plantaris muscles per replicate were used. Nuclei were extracted using the protocol described in the section above. The washed nuclei pellets were resuspended in sorting buffer (2% BSA/RNase-free PBS) and labeled with anti-PCMI (1:500, Sigma-Aldrich HPA023374) on ice for 45 min. Nuclei were washed twice in sorting buffer and then labeled with an Alexa Fluor 647-conjugated secondary antibody (1:100, Invitrogen A21244) at 4 °C for 30 min., followed by labeling with Hoechst dye. The nuclei were washed again and resuspended in sorting buffer with 0.2 U/µl Protector RNase inhibitor (Roche 03335399001) prior to cell sorting to separate the PCMI-labeled fractions using a FACS cell sorter (BD Aria). The sorted nuclei were then pelleted at 3000 × g for 10 min. before mRNA isolation using the Direct-zol RNA Microprep kit (Zymo research R2060). All RNA samples were initially amplified using the Ovation RNA-Seq System v2 kit (Tecan Genomics), following the manufacturer's guidelines. This system amplified RNA samples, producing double-stranded cDNA. Their concentrations were gauged using the Qubit dsDNA BR assay, and the cDNA size of each was determined with an Agilent HS DNA Chip.

For all samples, libraries were constructed using the Illumina's Nextera XT DNA Sample Preparation Kit. Specifically, 1 ng of cDNA was combined with Tagment DNA Buffer. Tagmentation, which includes fragmentation and adapter tagging, was executed with the Nextera enzyme, and the samples were then incubated at 55 °C for 10 min. NT Buffer was added for neutralization. Libraries were produced via PCR with the Nextera PCR Master Mix and two Nextera Indexes, using the following program: 72 °C for 3 min., 98 °C for 30 seconds, 12 cycles of 95 °C for 10 seconds, 55 °C for 30 seconds, 72 °C for 1 min., and a final cycle 72 °C for 5 min. Concentrations of these libraries were measured using the Qubit dsDNA HS assay, and the size was determined using the Agilent HS DNA chip. After pooling the samples, the concentration was optimized to ensure at least 40 million reads per sample. 100 bp paired-end sequencing was performed using a NovaSeq SP (200 cycles) v1.5 flow cell. All libraries were sequenced on the Illumina NovaSeq 6000 system.

### FACS analysis

The labeling efficiency of *Pax7<sup>CreTA</sup>; TRE<sup>H2B-GFP</sup>* in MuSCs in vivo was analyzed by subjecting harvested gastrocnemius muscles to enzymatic digestion with collagenase and pronase. The resultant mononuclear cell suspensions were treated with antibodies against VCAM1 (Biolegend 105717) for APC detection of myogenic cells, and with CD45, CD31, Sca1, and Ter119 for negative selection (all PE-conjugated, Biolegend 103105, 102407, 108107, 116207). The labeling efficacy of *Pax7<sup>CreTA</sup>; TRE<sup>H2B-GFP</sup>* in myonuclei in vivo was assessed using nuclei extracted from either TA or plantaris muscle as specified. The identical myonuclei sorting protocol described above was applied for GFP<sup>+</sup> myonuclei percentage analysis. All FACS analysis was performed on an LSRII platform (BD Biosciences), and the creation of data plots was achieved with FlowJo v10.8.1 software.

### bnRNA-seq data processing

FastQ files were processed using the nf-core/rnaseq v3.11.1 pipeline with minor modifications<sup>73,74</sup>. Briefly, quality control was performed using FastQC. Low-quality bases and adapter sequences were then trimmed and filtered from the reads using Cutadapt v3.4<sup>75</sup> to improve the overall data quality. Ribosomal RNA sequences were subsequently removed from the aligned data using SortMeRNA v4.3.4<sup>76</sup> to eliminate potential contamination from non-target RNA species. The resulting alignments were sorted and indexed using SAMtools v1.6.1<sup>77</sup>, enabling efficient storage and retrieval of the alignment information. Reads



were aligned to the mm10 genome obtained from GENCODE (vM21) using STAR v2.7.9a<sup>78</sup>. For read quantification, we utilized Salmon<sup>79</sup>, which specializes in estimating transcript abundance from RNA-Seq data. Selected QC metrics associated with each sample are reported in Supplementary Data 6. Transcripts were annotated using a modified GTF file obtained from gencode (vM24) containing both intronic and exonic regions for genes. Reads were imported into DESeq2<sup>80</sup> using the tximport command from the tximport package<sup>81</sup>. All differential expression analysis was performed using DESeq2 under various conditions based on the experimental design and whether analysis was paired or unpaired. Detailed information on sample processing and DESeq2 analyses can be found in the supplemental scripts found in GitHub link within the Data Availability section.

### snRNA-seq data processing

Initial read alignment and quantification of FASTQ files were generated using CellRanger/5.0.1. Selected QC metrics reported by CellRanger for each sample are shown in Supplementary Data 7. For each dataset, we corrected for ambient background RNA using cellbender<sup>82</sup>. Reads from cellbender were then imported into Seurat objects<sup>83</sup>, and nuclei with <200 unique features were removed from downstream analysis. Seurat objects with the remaining nuclei then underwent doublet identification using Solo/0.2. Nuclei having higher than the 95th percentile of number of unique features, number of UMIs, and mitochondrial reads were excluded. Violin plots for these metrics for all snRNA-seq experiments used in this study are shown in Supplementary Fig. 5. An additional filtration of any nuclei containing >5% reads mapping to the mitochondrial genome were excluded from all datasets. Datasets were normalized using the SCTransform() command, setting the “vst.flavor” argument to “v2”, and the vars.to.regress argument set to “percent.mt”. Additional regression of cell cycle scores was included in the SCTransform vars.to.regress argument only for datasets derived from whole muscle sequencing (P10 and 5 months), as these datasets included non-postmitotic cells. First, each cell was given a score using the CellCycleScoring() function provided by Seurat on the normalized RNA Assay, which was normalized using the NormalizeData() function using the gene signatures for S Phase and G2M phase provided by Seurat<sup>84</sup>. These genes were converted to the mouse genome by lowercasing all but the first letter of each gene.

### snRNA-Seq data integration

For each integration, the SelectIntegrationFeatures() and PrepSCTIntegration() were used on a list containing each dataset being integrated<sup>85</sup>. The identified features output from the previous command are given to the PrepSCTIntegration() command. Then, integration anchors were generated using the FindIntegrationAnchors() command, inputting the identified integration features to the “anchor.features” parameter, and specifying the “normalization.method” parameter as “SCT”. Datasets were then integrated by supplying the anchors to the IntegrateData() command, specifying the “normalization.method” parameter as “SCT”. Dimensionality reduction was performed on all integrated datasets using the FindClusters(), FindNeighbors(), and RunUMAP() commands. In the cases of myogenic subsets, after subset, an additional round of dimensionality reduction was performed using RunPCA(), FindClusters(), FindNeighbors(), and RunUMAP() again. UMAPs were visualized using the DimPlot() function.

### PHATE trajectory

The postnatal day 10 snRNAseq dataset was subset to contain only myogenic populations from non-specialized compartments. The phate() command provided from the phateR package<sup>86</sup> was run on the normalized RNA assay, with the parameter  $t = 30$ .

### Gene expression visualization

All Heatmaps were created using the DoHeatmap() function provided by Seurat. Feature plots were generated using the FeaturePlot() function. DotPlots were created using the DotPlot() function. Input features for Heatmaps and Dotplots were either manually selected, or generated through the FindAllMarkers() function provided by Seurat. VolcanoPlots for the bnRNA-Seq data were generated using the R package ggplot2 function<sup>87</sup>, plotting log transformed p adjusted values on the y-axis and log2 fold change on the x-axis. For Fig. 7, ggplot2 was used to plot the log2 fold change of the indicated samples.

### Statistics and reproducibility

We used the following criterion for differentially expressed genes: of  $\log_2FC \geq 1$ ,  $FDR < 0.01$  (upregulated genes) and  $\log_2FC \leq -1$ ,  $FDR < 0.01$  (downregulated genes), except for Fig. 7c, d, where  $\log_2FC \geq 0.5$ ,  $FDR < 0.1$  (upregulated genes) and  $\log_2FC \leq -0.5$ ,  $FDR < 0.1$  (downregulated genes) was used. Gene ontology analysis was performed using ToppGene (<https://toppgene.cchmc.org>)<sup>88</sup>. Sample sizes are noted in the figure legends. Data were processed using GraphPad Prism 9 software. In all graphs, error bars represent standard deviation (SD). Data were compared between groups using various statistical tests, indicated in the figure legends, and based on number of groups, normality of the data, and variance of standard deviations. The criterion for statistical significance was \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ . No statistical method was used to predetermine sample size. No data were excluded from the analyses. Mice were randomly placed into experimental or control groups. Investigators were blinded for image analysis.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Raw sequencing data has been deposited in GEO under accession number GSE241035 and are publicly available. Source data are provided with this paper.

### Code availability

Data analysis is described in the Methods. Scripts are available on GitHub [<https://github.com/cswoboda/NF-Myonuclei>].

### References

- Bachman, J. F. et al. Prepubertal skeletal muscle growth requires Pax7-expressing satellite cell-derived myonuclear contribution. *Development* **145**, dev167197 (2018).
- Pawlikowski, B., Pulliam, C., Betta, N. D., Kardon, G. & Olwin, B. B. Pervasive satellite cell contribution to uninjured adult muscle fibers. *Skelet. muscle* **5**, 1–13 (2015).
- White, R. B., Biérinx, A.-S., Gnocchi, V. F. & Zammit, P. S. Dynamics of muscle fibre growth during postnatal mouse development. *BMC Dev. Biol.* **10**, 1–11 (2010).
- Bachman, J. F. & Chakkalakal, J. V. Insights into muscle stem cell dynamics during postnatal development. *FEBS J.* **289**, 2710–2722 (2022).
- Collins, B. C. et al. Cellular dynamics of skeletal muscle regeneration. *bioRxiv* <https://doi.org/10.1101/2023.05.02.538744> (2023).
- Dumont, N. A., Bentzinger, C. F., Sincennes, M. C. & Rudnicki, M. A. Satellite cells and skeletal muscle regeneration. *Compr. Physiol.* **5**, 1027–1059 (2015).
- Hawke, T. J. & Garry, D. J. Myogenic satellite cells: physiology to molecular biology. *J. Appl. Physiol.* **91**, 534–551 (2001).



8. Kang, J.-S. & Krauss, R. S. Muscle stem cells in developmental and regenerative myogenesis. *Curr. Opin. Clin. Nutr. Metab. Care* **13**, 243–248 (2010).
9. Mauro, A. Satellite cell of skeletal muscle fibers. *J. Biophys. Biochem Cytol.* **9**, 493 (1961).
10. Millay, D. P. Regulation of the myoblast fusion reaction for muscle development, regeneration, and adaptations. *Exp. cell Res.* **415**, 113134 (2022).
11. Relaix, F. & Zammit, P. S. Satellite cells are essential for skeletal muscle regeneration: the cell on the edge returns centre stage. *Development* **139**, 2845–2856 (2012).
12. Blau, H. M., Cosgrove, B. D. & Ho, A. T. The central role of muscle stem cells in regenerative failure with aging. *Nat. Med.* **21**, 854–862 (2015).
13. Porpiglia, E. & Blau, H. M. Plasticity of muscle stem cells in homeostasis and aging. *Curr. Opin. Genet. Dev.* **77**, 101999 (2022).
14. Prasad, V. & Millay, D. P. In *Seminars in Cell & Developmental Biology*. 3–10 (Elsevier).
15. Ralston, E. & Hall, Z. W. Restricted distribution of mRNA produced from a single nucleus in hybrid myotubes. *J. cell Biol.* **119**, 1063–1068 (1992).
16. Pavlath, G. K., Rich, K., Webster, S. G. & Blau, H. M. Localization of muscle gene products in nuclear domains. *Nature* **337**, 570–573 (1989).
17. Allen, D. L., Roy, R. R. & Edgerton, V. R. Myonuclear domains in muscle adaptation and disease. *Muscle Nerve* **22**, 1350–1360 (1999).
18. Bruusgaard, J., Liestøl, K., Ekmark, M., Kollstad, K. & Gundersen, K. Number and spatial distribution of nuclei in the muscle fibres of normal mice studied in vivo. *J. Physiol.* **551**, 467–478 (2003).
19. Liu, J. X. et al. Myonuclear domain size and myosin isoform expression in muscle fibres from mammals representing a 100 000-fold difference in body size. *Exp. Physiol.* **94**, 117–129 (2009).
20. Van der Meer, S., Jaspers, R. & Degens, H. Is the myonuclear domain size fixed? *J. Musculoskelet. Neuro. Interact.* **11**, 286–297 (2011).
21. Qaisar, R. & Larsson, L. What determines myonuclear domain size? *Indian J. Physiol. Pharmacol.* **58**, 1–12 (2014).
22. Bagley, J. R., Denes, L. T., McCarthy, J. J., Wang, E. T. & Murach, K. A. The myonuclear domain in adult skeletal muscle fibres: past, present, and future. *J. Physiol.* **601**, 723–741 (2023).
23. Murach, K. A., Englund, D. A., Dupont-Versteegden, E. E., McCarthy, J. J. & Peterson, C. A. Myonuclear domain flexibility challenges rigid assumptions on satellite cell contribution to skeletal muscle fiber hypertrophy. *Front. Physiol.* **9**, 635 (2018).
24. Arnold, E. M. & Delp, S. L. Fibre operating lengths of human lower limb muscles during walking. *Philos. Trans. R. Soc. B: Biol. Sci.* **366**, 1530–1539 (2011).
25. Hansson, K.-A. et al. Myonuclear content regulates cell size with similar scaling properties in mice and humans. *Nat. Commun.* **11**, 6288 (2020).
26. Murach, K. A., Dungan, C. M., Von Walden, F. & Wen, Y. Epigenetic evidence for distinct contributions of resident and acquired myonuclei during long-term exercise adaptation using timed in vivo myonuclear labeling. *Am. J. Physiol.-Cell Physiol.* **322**, C86–C93 (2022).
27. Pinheiro, H. et al. mRNA distribution in skeletal muscle is associated with mRNA size. *J. cell Sci.* **134**, jcs256388 (2021).
28. Englund, D. A. et al. Satellite cell depletion disrupts transcriptional coordination and muscle adaptation to exercise. *Function* **2**, zqaa033 (2021).
29. Englund, D. A. et al. Depletion of resident muscle stem cells negatively impacts running volume, physical function, and muscle fiber hypertrophy in response to lifelong physical activity. *Am. J. Physiol.-Cell Physiol.* **318**, C1178–C1188 (2020).
30. Fukuda, S. et al. Sustained expression of HeyL is critical for the proliferation of muscle stem cells in overloaded muscle. *Elife* **8**, e48284 (2019).
31. Goh, Q. & Millay, D. P. Requirement of myomaker-mediated stem cell fusion for skeletal muscle hypertrophy. *elife* **6**, e20007 (2017).
32. Goh, Q. et al. Myonuclear accretion is a determinant of exercise-induced remodeling in skeletal muscle. *Elife* **8**, e44876 (2019).
33. Randrianarison-Huetz, V. et al. Srf controls satellite cell fusion through the maintenance of actin architecture. *J. Cell Biol.* **217**, 685–700 (2018).
34. Egner, I. M., Bruusgaard, J. C. & Gundersen, K. Satellite cell depletion prevents fiber hypertrophy in skeletal muscle. *Development* **143**, 2898–2906 (2016).
35. Ding, J. et al. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat. Biotechnol.* **38**, 737–746 (2020).
36. Chemello, F. et al. Degenerative and regenerative pathways underlying duchenne muscular dystrophy revealed by single-nucleus RNA sequencing. *Proc. Natl Acad. Sci. USA* **117**, 29691–29701 (2020).
37. Dos Santos, M. et al. Single-nucleus RNA-seq and FISH identify coordinated transcriptional activity in mammalian myofibers. *Nat. Commun.* **11**, 5102 (2020).
38. Kim, M. et al. Single-nucleus transcriptomics reveals functional compartmentalization in syncytial skeletal muscle cells. *Nat. Commun.* **11**, 6375 (2020).
39. Petrany, M. J. et al. Single-nucleus RNA-seq identifies transcriptional heterogeneity in multinucleated skeletal myofibers. *Nat. Commun.* **11**, 6374 (2020).
40. Dos Santos, M. et al. Opposing gene regulatory programs governing myofiber development and maturation revealed at single nucleus resolution. *Nat. Commun.* **14**, 4333 (2023).
41. Kurland, J. V. et al. Aging disrupts gene expression timing during muscle regeneration. *Stem Cell Rep.* **18**, 1325–1339 (2023).
42. Masschelein, E. et al. Exercise promotes satellite cell contribution to myofibers in a load-dependent manner. *Skelet. muscle* **10**, 1–15 (2020).
43. Foudi, A. et al. Analysis of histone 2B-GFP retention reveals slowly cycling hematopoietic stem cells. *Nat. Biotechnol.* **27**, 84–90 (2009).
44. Morcos, M. N. et al. Continuous mitotic activity of primitive hematopoietic stem cells in adult mice. *J. Exp. Med.* **217**, e20191284 (2020).
45. Winje, I. et al. Specific labelling of myonuclei by an antibody against pericentriolar material 1 on skeletal muscle tissue sections. *Acta Physiol.* **223**, e13034 (2018).
46. Cramer, A. A. et al. Nuclear numbers in syncytial muscle fibers promote size but limit the development of larger myonuclear domains. *Nat. Commun.* **11**, 6287 (2020).
47. Martinet, C. et al. H19 controls reactivation of the imprinted gene network during muscle regeneration. *Development* **143**, 962–971 (2016).
48. Li, J. et al. Long non-coding RNA H19 promotes porcine satellite cell differentiation by interacting with TDP43. *Genes (Basel)* **11**, 259 (2020).
49. Katoku-Kikyo, N. et al. Per1/Per2-Igf2 axis-mediated circadian regulation of myogenic differentiation. *J. Cell Biol.* **220**, e202101057 (2021).
50. Zhang, Y. et al. The lncRNA H19 alleviates muscular dystrophy by stabilizing dystrophin. *Nat. Cell Biol.* **22**, 1332–1345 (2020).
51. Yue, Y. et al. The long noncoding RNA lnc-H19 is important for endurance exercise by maintaining slow muscle fiber types. *J. Biol. Chem.* **299**, 105281 (2023).

52. Bella, P. et al. Blockade of IGF2R improves muscle regeneration and ameliorates duchenne muscular dystrophy. *EMBO Mol. Med.* **12**, e11019 (2020).
53. Millay, D. P., Sutherland, L. B., Bassel-Duby, R. & Olson, E. N. Myo-maker is essential for muscle regeneration. *Genes Dev.* **28**, 1641–1646 (2014).
54. Mademtoglou, D., Geara, P., Mourikis, P. & Relaix, F. Pax7 haploinsufficiency impairs muscle stem cell function in Cre-recombinase mice and underscores the importance of appropriate controls. *Stem Cell Res. Ther.* **14**, 294 (2023).
55. Timson, B. F. Evaluation of animal models for the study of exercise-induced muscle enlargement. *J. Appl. Physiol.* **69**, 1935–1945 (1990).
56. Murach, K. A. et al. Multi-transcriptome analysis following an acute skeletal muscle growth stimulus yields tools for discerning global and MYC regulatory networks. *J. Biol. Chem.* **298**, 102515 (2022).
57. Iwata, M. et al. A novel tetracycline-responsive transgenic mouse strain for skeletal muscle-specific gene expression. *Skelet. muscle* **8**, 1–8 (2018).
58. Roman, W. et al. Muscle repair after physiological damage relies on nuclear migration for cellular reconstruction. *Science* **374**, 355–359 (2021).
59. Rochlin, K., Yu, S., Roy, S. & Baylies, M. K. Myoblast fusion: when it takes more to make one. *Dev. Biol.* **341**, 66–83 (2010).
60. Schiaffino, S., Rossi, A. C., Smerdu, V., Leinwand, L. A. & Reggiani, C. Developmental myosins: expression patterns and functional significance. *Skelet. Muscle* **5**, 1–14 (2015).
61. Fukada, S.-I., Higashimoto, T. & Kaneshige, A. Differences in muscle satellite cell dynamics during muscle hypertrophy and regeneration. *Skelet. Muscle* **12**, 1–10 (2022).
62. Jiao, S. et al. Differential regulation of IGF-I and IGF-II gene expression in skeletal muscle cells. *Mol. Cell Biochem* **373**, 107–113 (2013).
63. Erbay, E., Park, I. H., Nuzzi, P. D., Schoenherr, C. J. & Chen, J. IGF-II transcription in skeletal myogenesis is controlled by mTOR and nutrients. *J. Cell Biol.* **163**, 931–936 (2003).
64. Li, Y. et al. Functional significance of gain-of-function H19 lncRNA in skeletal muscle differentiation and anti-obesity effects. *Genome Med* **13**, 137 (2021).
65. Dey, B. K., Pfeifer, K. & Dutta, A. The H19 long noncoding RNA gives rise to microRNAs miR-675-3p and miR-675-5p to promote skeletal muscle differentiation and regeneration. *Genes Dev.* **28**, 491–501 (2014).
66. Hansson, K.-A. & Eftestøl, E. Scaling of nuclear numbers and their spatial arrangement in skeletal muscle cell size regulation. *Mol. Biol. Cell* **34**, pe3 (2023).
67. Blau, H. M., Chiu, C. P. & Webster, C. Cytoplasmic activation of human nuclear genes in stable heterocaryons. *Cell* **32**, 1171–1180 (1983).
68. Pomerantz, J. H., Mukherjee, S., Palermo, A. T. & Blau, H. M. Reprogramming to a muscle fate by fusion recapitulates differentiation. *J. Cell Sci.* **122**, 1045–1053 (2009).
69. Relaix, F. et al. Perspectives on skeletal muscle stem cells. *Nat. Commun.* **12**, 692 (2021).
70. Dumont, N. A., Wang, Y. X. & Rudnicki, M. A. Intrinsic and extrinsic mechanisms regulating satellite cell function. *Development* **142**, 1572–1581 (2015).
71. Lepper, C., Conway, S. J. & Fan, C.-M. Adult satellite cells and embryonic muscle progenitors have distinct genetic requirements. *Nature* **460**, 627–631 (2009).
72. Wu, S., Ying, G., Wu, Q. & Capecchi, M. R. A protocol for constructing gene targeting vectors: generating knockout mice for the cadherin family and beyond. *Nat. Protoc.* **3**, 1056–1076 (2008).
73. Ewels, P. A. et al. The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **38**, 276–278 (2020).
74. Di Tommaso, P. et al. Nextflow enables reproducible computational workflows. *Nat. Biotechnol.* **35**, 316–319 (2017).
75. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).
76. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).
77. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Giga-science* **10**, giab008 (2021).
78. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
79. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. methods* **14**, 417–419 (2017).
80. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
81. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* **4**, 1521 (2015).
82. Fleming, S. J., Marioni, J. C. & Babadi, M. CellBender remove-background: a deep generative model for unsupervised removal of background noise from scRNA-seq datasets. *BioRxiv* <https://doi.org/10.1101/791699> (2019).
83. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
84. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
85. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
86. Moon, K. R. et al. Visualizing structure and transitions in high-dimensional biological data. *Nat. Biotechnol.* **37**, 1482–1492 (2019).
87. Wickham, H. & Wickham, H. *Data analysis*. Vol. 864 (Springer, 2016).
88. Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic acids Res.* **37**, W305–W311 (2009).

## Acknowledgements

We thank members of the Millay laboratory and Vikram Prasad (Molkentin lab) for discussion. We would like to acknowledge the assistance of the Research Flow Cytometry Core, Single Cell Genomics Core, and DNA sequencing and Genotyping Core at Cincinnati Children's Hospital Medical Center. This work was mainly supported by a grant to D.P.M. from the National Institutes of Health (R01AG059605). Work in the Millay laboratory is also funded by grants to D.P.M. from Children's Hospital Research Foundation, National Institutes of Health (R01AR068286, R61AR076771). Work for this project in the Weirauch laboratory is funded by grants to M.T.W. from the National Institutes of Health (P30AR070549) and Cincinnati Children's Hospital Research Foundation (ARC award #53632). The Lepper laboratory is supported by a grant from the National Institutes of Health (R01AR078231).

## Author contributions

C.S., C.O.S., and D.P.M. conceived the project. C.S., C.O.S., F.M.M., C.C., M.J.P., S.P., and A.V. conducted experiments and analyzed the data. M.T.W., C.L., and D.P.M. supervised the project. C.S., C.O.S., and D.P.M. wrote the manuscript with input from all authors.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-53510-z>.

**Correspondence** and requests for materials should be addressed to Douglas P. Millay.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024