

# Development of a split-toxin CRISPR screening platform to systematically identify regulators of human myoblast fusion

Received: 4 July 2025

Accepted: 2 December 2025

Published online: 15 January 2026

 Check for updates

Haifeng Zhang<sup>1,7</sup>, Renjie Shang<sup>2,7</sup>, Zheng Zhang<sup>2,7</sup>, Min Zhou<sup>1</sup>, Anne Bigot<sup>3</sup>, Yanqing Cai<sup>1</sup>, Yanran Zhao<sup>4</sup>, Yushu Wang<sup>2</sup>, Aaryahi Deshmukh<sup>2</sup>, Elena Kudryashova<sup>5</sup>, Dmitri S. Kudryashov<sup>5</sup>, Cuiyu He<sup>6</sup>, Vincent Mouly<sup>3</sup> & Pengpeng Bi<sup>1,2</sup>✉

Muscle defects are common in human developmental disorders and often cause severe functional impairment. These defects arise from intricate tissue crosstalk and rare genetic mutations, underscoring the need to systematically identify cell-autonomous mechanisms regulating human myogenesis. Here we show a rationally designed, high-throughput genetic screening platform that integrates human myoblast models, customized CRISPR libraries, and a split-toxin strategy that enables quantitative selection of fusion-defective myocytes. Leveraging this platform, our initial screen uncovers a large group of hits essential for human myoblast fusion. The majority of these hits converge into 23 protein complexes. Notably, mutations in 41 screen hits are associated with human diseases marked by abnormal skeletal-muscle morphology. Applying a new single-cell CRISPR & RNA-seq approach, we show that majority of these hits control human myoblast fusion as well as influence early-stage myogenic differentiation. This work establishes a scalable approach to identify cell-autonomous regulators of human muscle differentiation and fusion.

Human muscle development is a highly orchestrated process that begins during early embryogenesis and peaks during the fetal stage<sup>1</sup>. Myogenic progenitors, or myoblasts, originating from the paraxial mesoderm, undergo differentiation and fusion to form multinucleated myotubes under the control of muscle-specific transcriptional factors<sup>2–8</sup>, such as PAX7, MYF5, MYOD, MYOG, and MEF2C. Myoblast fusion, a hallmark of skeletal myogenesis, is controlled by bipartite membrane proteins Myomaker (MYMK) and Myomixer (MYMX, also known as Minion and Myomerger) that govern membrane coalescence<sup>9–12</sup>. As a conserved feature of vertebrate myoblast fusion, the formation of actin-propelled invasive protrusions<sup>13–16</sup> is believed to facilitate the function of membrane fusogens Myomaker and Myomixer at the fusogenic synapses. Mutations in either Myomaker or

Myomixer in patients result in fusion myopathy characterized by developmental delay, hypotonia and impaired muscle growth<sup>17–19</sup>.

As part of the broader myogenic program, the expression and activity of muscle fusogens are tightly regulated by MYOD and MYOG, along with several signaling pathways<sup>20–26</sup> including Notch, ERK1/2, CaMKII, IRE1 $\alpha$ , and TGF- $\beta$ . Post fusion, multinucleated myotubes mature into functional muscle fibers, expressing contractile proteins and establishing neural and vascular connections. By the second trimester, the architecture of human skeletal muscle is largely established, although muscle growth continues into postnatal life through fusion and myonuclear accretion<sup>27</sup>.

Persistent research using model organisms has uncovered conserved cellular and molecular mechanisms underlying skeletal

<sup>1</sup>Center for Molecular Medicine, University of Georgia, Athens, GA, USA. <sup>2</sup>Department of Genetics, University of Georgia, Athens, GA, USA. <sup>3</sup>Center for Research in Myology UMRS974, Sorbonne Université, INSERM, Myology Institute AIM, Paris, France. <sup>4</sup>Department of Cellular Biology, University of Georgia, Athens, GA, USA. <sup>5</sup>Department of Chemistry and Biochemistry, The Ohio State University, Columbus, OH, USA. <sup>6</sup>Department of Mathematics, University of Georgia, Athens, GA, USA. <sup>7</sup>These authors contributed equally: Haifeng Zhang, Renjie Shang, Zheng Zhang. ✉e-mail: [pbi@uga.edu](mailto:pbi@uga.edu)

myogenesis<sup>14, 28–33</sup>. However, despite major advances, interspecies differences in genetic interactions can lead to phenotypic variability in response to the same genetic mutations<sup>33,34</sup>. This is particularly relevant given that most model organisms are highly inbred, leading to reduced genetic diversity and potentially masking context-dependent gene functions that would be evident in the more genetically heterogeneous human population. These limitations underscore the critical importance of studying gene function directly in human systems to better understand the molecular basis of human diseases.

Recently, immortalized human myoblasts have emerged as a powerful system for investigating the molecular mechanisms of human muscle development and disease<sup>25,35,36</sup>. However, scalable genetic tools for dissecting the intricate regulatory networks involved in this process remain limited. To close this gap, we developed a CRISPR screening platform that integrates three major components. First, human myoblast lines derived from diverse donors and anatomical sites. Second, a custom-designed gRNA library (MyoCRISPR-KO<sup>lib</sup>) targeting 6896 genes selected from unbiased transcriptome analysis of human myogenesis across a panel of donors. Third and most importantly, a phenotypic readout strategy that couples cell fusion with cell viability selection using a split-toxin expression system.

We systematically evaluated the performance of each major component of the platform. First, transcriptome analysis unbiasedly validated the strong myogenic potential of the human myoblast lines. Comparative analysis showed a high degree of similarity between these in vitro-differentiated human myoblasts and human fetal muscle tissue, specifically in the activation of the gene program for muscle cell differentiation, supporting the physiological relevance of the culture system. Karyotype analysis revealed a normal diploid genome architecture in these human myoblasts. Second, for the newly generated MyoCRISPR-KO<sup>lib</sup>, we performed a proof-of-concept test in a simple myoblast fitness screen which revealed hundreds of genes essential for myoblast survival and expansion. Third, we demonstrated that the split-toxin system effectively and specifically eliminates fused myotubes while sparing fusion-defective myoblasts, thereby fulfilling its intended function.

Leveraging this platform, we conducted a preliminary genetic screen to systematically identify upstream regulators of human myoblast fusion. This screen uncovered a large group of hits, revealing protein interaction networks comprising 23 protein complexes across diverse cellular pathways. We also validated the function for the majority of these hits using myoblasts from various human donors and anatomical sites, as well as in mouse primary myoblasts.

Our cross-referencing analysis with human medical database revealed that mutations in 41 of our fusion screen hits cause human diseases presenting abnormal skeletal muscle morphology. Leveraging a new single-cell CRISPR & RNA-seq approach, we uncovered the broader roles of the fusion screen hits during human muscle differentiation. Collectively, this study not only provides a fully validated tool for quantitatively and systematically dissecting upstream regulators of human myotube formation, but also reports a group of promising candidate genes crucial for human myogenesis and congenital muscle diseases.

## Results

### Validation of cellular models for studying human muscle differentiation and fusion

We assessed the myogenic potential for a panel of human myoblast lines derived from seven donors (Supplementary Fig. 1a, b). These cells were generated previously following an established protocol<sup>35</sup>. Briefly, NCAMI<sup>+</sup> muscle precursor cells were isolated from skeletal muscle biopsies across various anatomical sites, ages, and sexes. The cells were then immortalized and clonally selected based on their high myogenic potential, exhibiting differentiation index ranging from

94.6% to 98.7% and fusion index from 90.7% to 98.0% (Supplementary Fig. 1c).

Upon differentiation, the immortalized human myoblasts robustly express skeletal myosin as well as the master regulators of muscle differentiation, MYOD and Myogenin (MYOG) (Fig. 1a and Supplementary Fig. 1d). To unbiasedly authenticate the myogenic fate of these cells, we also conducted time-course RNA-seq analysis. This revealed prompt induction for the expression of muscle differentiation markers, including *MYH1*, *MYH2*, *MYH3*, *MYH7*, *MYH8*, *MYL1*, *MYL2*, and muscle fusogens (*MYMK*, *MYMX*) for myoblasts from all donors (Fig. 1b). To assess the relevance of the cultured cells to in vivo human myogenesis, we performed gene set enrichment analysis (GSEA) by comparing myoblast transcriptomes to that of human fetal muscles<sup>1</sup> (Supplementary Fig. 2a, b). This revealed a molecular signature of cultured human myoblasts that closely mirrors human fetal muscle differentiation program in vivo (Fig. 1c, Supplementary Fig. 2c). Genes representing the core enrichment shared between the two transcriptome datasets are summarized in Supplementary Data 1.

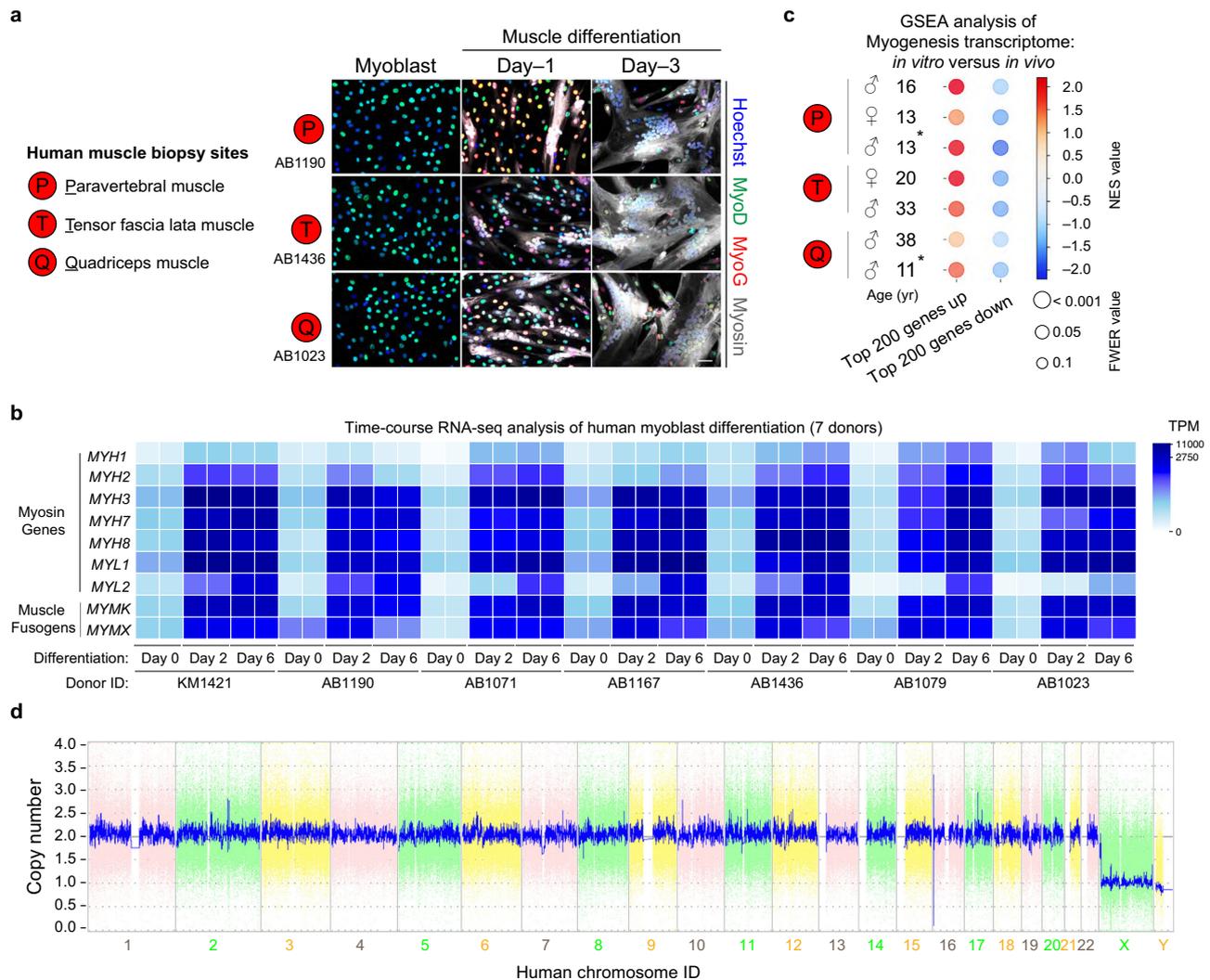
Normal diploidy in cells is essential for clear interpretation of knockout (KO) effects in gene function study, ensuring consistent gene copy number and reducing confounding from genomic instability. Thus, we analyzed the karyotype for the immortalized human myoblasts. Unlike the murine C2C12 myoblasts<sup>37</sup>, these human myoblasts maintain a normal stable diploid genome even at a relatively high passage of 20 (line AB1190, Fig. 1d). Together, these analyses establish the human myoblast lines as ideal genetic models for studying human myogenesis.

### A new gRNA library tailored for CRISPR-KO screens in human myoblasts

The scale and maneuverability of a CRISPR screening experiment are linked to the size of the CRISPR library. CRISPR screens often employ pre-made genome-wide libraries, wasting significant screening bandwidth when a large number of target genes are not even expressed in the targeted cells<sup>38</sup>. To tackle this, we designed a more cost-effective and muscle-targeted gene KO gRNA library, termed MyoCRISPR-KO<sup>lib</sup>. This lentiviral CRISPR library targets 6896 genes (Supplementary Data 2) that were consistently expressed above a relatively low threshold (TPM > 1, RNA-seq) in human myoblasts from multiple donors, either before or after myogenic induction (Fig. 2a). Although this library targets only about one-third as many protein-coding genes as genome-wide libraries, the selected genes collectively produce approximately 90% of all human protein-coding transcripts expressed at various stages of human myoblast differentiation (Fig. 2b).

The gRNAs for the selected genes were chosen from previously reported genome-wide libraries, prioritizing those with top-ranking predicted on-target and low off-target scores<sup>39,40</sup>. A pooled library of synthesized oligonucleotides containing these gRNA sequences was then cloned into a lentiviral backbone encoding a blasticidin resistance gene and optimized for high-titer virus production<sup>41</sup>. Deep sequencing of the plasmid library confirmed high cloning fidelity and uniform representation of gRNA sequences (Fig. 2c).

To assess the efficacy of the MyoCRISPR-KO<sup>lib</sup> in targeting endogenous genes, we conducted a simple myoblast fitness screen (Fig. 2d). In this assay, gRNAs targeting genes essential for cell survival or proliferation are expected to be depleted over time, whereas those targeting negative regulators of cell fitness should become enriched. Human myoblasts isolated from the paravertebral muscle of a healthy donor (line AB1190, passage 4) were transduced with the lentiviral library at a low multiplicity of infection (MOI) (MOI = 0.2), ensuring that most infected cells received a single viral integrant and expressed one unique gRNA, thereby disrupting only one gene per cell. After transduction, uninfected cells were removed by drug selection, and the remaining cells were cultured in growth medium. At multiple



**Fig. 1 | Characterization of human immortalized myoblast lines.** **a** Anatomical locations of muscle biopsies used to isolate and clonally derive human NCAM1<sup>+</sup> myoblasts (see also Supplementary Fig. 1) and immunostaining result of the myogenic markers: myosin, MyoD and MyoG. Scale bar: 50  $\mu$ m. The experiment was independently repeated three times with similar results. **b** Heatmap for RNA-seq results of selected muscle differentiation and fusion marker genes in immortalized human myoblasts isolated from seven donors. Source data are provided as a Source data file. **c** Gene set enrichment analysis (GSEA) to compare expression data for two sets of human fetal myogenesis marker genes during *in vitro* differentiation of seven immortalized human myoblast lines. The two gene sets include the top 200 up- and down-regulated genes (seen in Supplementary Data 1) in human fetal

myocytes versus fetal myoblasts from analysis of a scRNA-seq dataset (EMBL Accession: E-MTAB-8813). NES: normalized enrichment score; FWER: family-wise error rate. **d** Karyotype analysis of immortalized human myoblast line (AB1190) at passage -20. No chromosomal aberrations were found. The whole genome view displays copy numbers quantification using 1.1 million probes across somatic and sex chromosome with a resolution down to 1Mb. A value of 2 represents a normal copy number state (CN = 2). A value of 3 represents chromosomal gain (CN = 3). A value of 1 represents a chromosomal loss (CN = 1). The pink, green and yellow colors indicate the raw signal for each individual chromosome probe, while the blue signal represents the normalized probe signal which is used to identify copy number and aberrations (if any).

timepoints (T1–T5) during the culture, a subset of cells was harvested for gRNA abundance analysis, while the rest were passaged to maintain continuous culture and prolonged selection pressure.

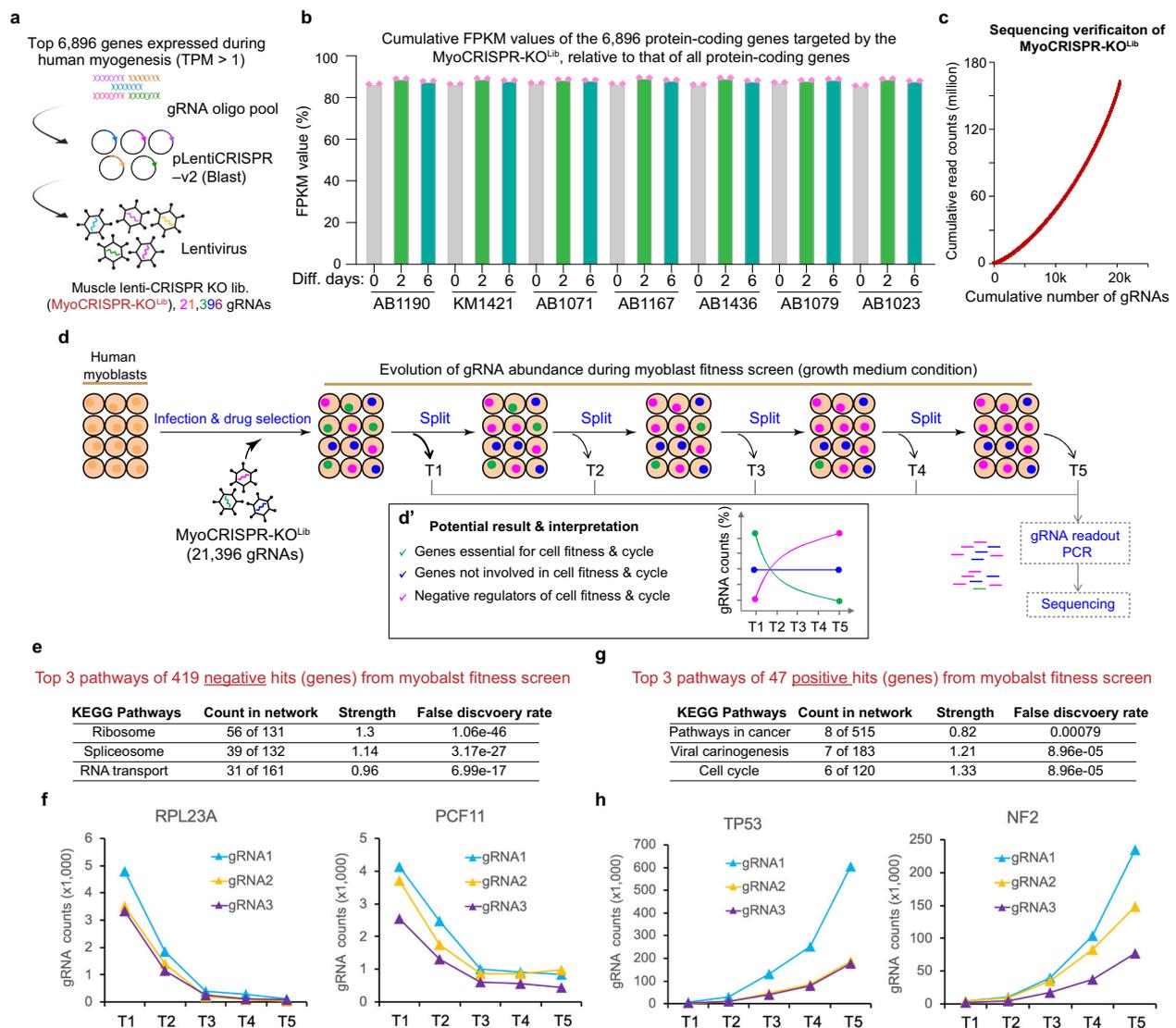
Because the lentiviral CRISPR vector integrates into the genome, each infected myoblast carries a unique gRNA, denoted by a unique color shown in Fig. 2d. At each time-point of sample collection, the identity of the gRNA present in the cells was determined by extracting genomic DNA and amplifying the gRNA cassette using primers flanking the gRNA sequence. The resulting PCR products were purified and subjected to next-generation sequencing to quantify and compare the relative abundance of each gRNA over the time-course of the fitness screen. Genes targeted by multiple gRNAs showing significant and consistent changes were classified as hits (Fig. 2d').

By this fitness screen, we were able to identify 419 genes essential for human myoblast survival or proliferation

(Supplementary Data 3). Pathway analysis revealed significant enrichment in genes involved in ribosomal structure and function (e.g., RPL23A), mRNA processing (e.g., PCF11) and RNA transport (Fig. 2e, f). Additionally, this screen also uncovered 47 negative cell cycle regulators (Supplementary Data 4), including TP53 and NF2, whose deletion led to marked enrichment of their corresponding gRNAs (Fig. 2g, h). These results validate the efficacy of the MyoCRISPR-KO<sup>Lib</sup> and our molecular and computational pipelines for gRNA detection and sequencing data analysis.

### A rationally designed phenotypic readout strategy for quantitative analysis of human myoblast fusion in a high-throughput manner

The success of any genetic screen hinges on the ability to enrich mutants that exhibit a specific and distinguishable biological



**Fig. 2 | Design and functional validation of a new gRNA library tailored for CRISPR-knockout screens in human myoblasts.** **a** Design of the muscle-targeted CRISPR knockout library (MyoCRISPR-KO<sup>lib</sup>), comprising 20,496 gene-targeting gRNAs (~3 per gene) and 900 non-targeting control gRNAs. TPM: transcripts per million. **b** Cumulative mRNA expression levels of genes included in the library, shown as a percentage of total mRNA from all protein-coding genes, based on RNA-seq analysis of seven human myoblast lines. Diff.: differentiation. **c** Cumulative read count plot showing the uniform distribution of gRNA abundances in the MyoCRISPR-KO<sup>lib</sup>, as determined by deep sequencing. **d** Overview of the myoblast

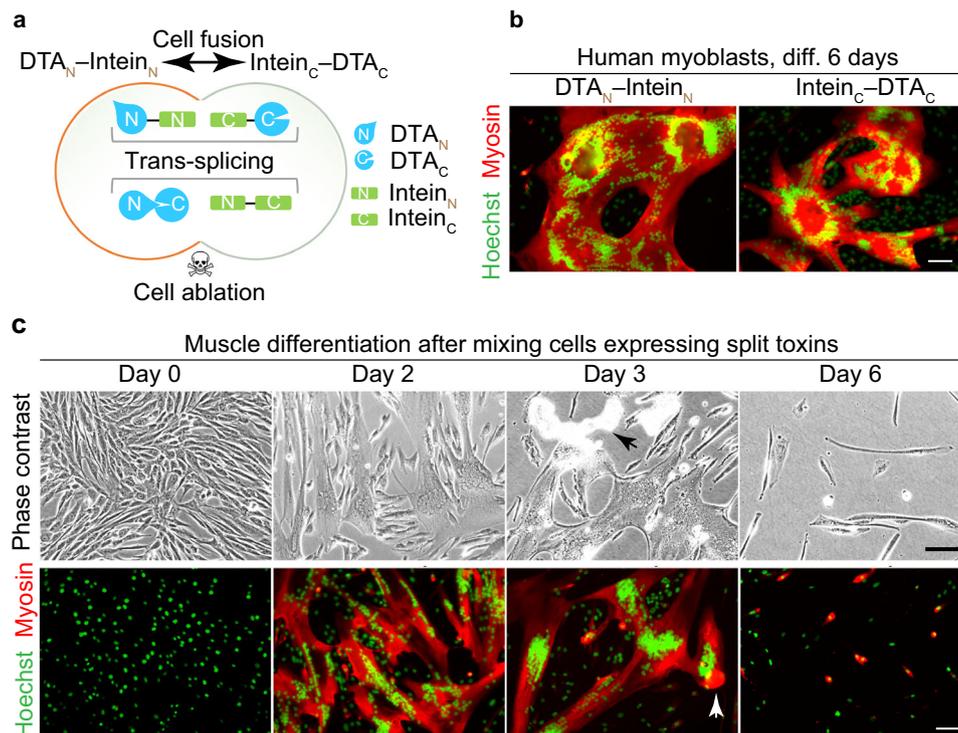
fitness screen. Each dot represents a single gRNA expression and corresponding gene perturbation. **d'**, Hypothetical results illustrate the potential impact of gene perturbation, interpreted by comparing gRNA abundance across samples from different time points. **e** Top three pathways enriched among the 419 genes (Supplementary Data 3) identified in the myoblast fitness screen. **f** gRNA quantification results for two representative hits. **g** Top three pathways enriched among the 47 negative regulators (Supplementary Data 4) identified in the myoblast fitness screen. **h** gRNA quantification results for two representative hits. Source data are provided as a Source data file.

phenotype. Myoblast fusion represents a critical step in skeletal muscle development and regeneration, making it a relevant and quantifiable cellular phenotype for studying human myogenesis. To enable selective enrichment of fusion-defective mutants, we aimed to couple myoblast fusion to cell viability.

To this end, we developed a fusion-dependent cell-killing strategy using a split toxin system. Specifically, two fragments of diphtheria toxin A (DTA) were fused to bacterial intein fragments<sup>42</sup>: DTA<sub>N</sub>-Intein<sub>N</sub> and Intein<sub>C</sub>-DTA<sub>C</sub>, and expressed separately in human myoblasts (Fig. 3a). Each construct was verified to be non-toxic individually and had no effect on myoblast viability or fusion capacity (Fig. 3b). However, when the two myoblast populations were co-cultured, cell-cell fusion allowed reconstitution of the active toxin via intein-mediated

protein splicing inside the syncytium (Fig. 3a). This led to the selective killing and detachment of multinucleated myotubes from the culture (arrow, Fig. 3c).

After validating this cell-fusion readout strategy, we integrated it into a complete screening platform to systematically identify key upstream regulators of human myoblast fusion (Fig. 4a). Briefly, human myoblasts isolated from the paravertebral muscle of a healthy donor (line ABI190, passage 4) were first transduced with retrovirus expressing DTA<sub>N</sub>-Intein<sub>N</sub>, followed by lentiviral transduction of the MyoCRISPR-KO<sup>lib</sup>. After drug selection, the gene-edited myoblasts were divided into two groups. The “killer” group was co-cultured with myoblasts from the same donor stably expressing the complementary toxin Intein<sub>C</sub>-DTA<sub>C</sub>. In this group, myoblast fusion will reconstitute



**Fig. 3 | A rationally designed phenotypic readout strategy to enrich for fusion-defective myoblasts.** **a** Schematic of intein-mediated protein trans-splicing and reconstitution of active diphtheria toxin subunit A (DTA). Fusion between the two myoblast populations, each expressing one half of the toxin, enables reconstitution of functional DTA, selectively inducing death of fused cells. **b** Myosin immunostaining of myoblasts stably expressing  $DTA_N$ -Intein $_N$  or Intein $_C$ - $DTA_C$  split toxins after myogenic induction, showing no impact on fusion in the absence of toxin

reconstitution. Diff.: differentiation. The experiment was independently repeated four times with similar results. **c** Time-lapse phase-contrast images (top) and Myosin immunostaining results (bottom) of myoblasts expressing  $DTA_N$ -Intein $_N$ , co-cultured with myoblasts expressing Intein $_C$ - $DTA_C$  split toxin, showing selective loss of multinucleated cells at end of myogenic differentiation. Arrows point to the multinucleated myotubes starting to detach. Scale bars: 100  $\mu$ m.

DTA and selectively eliminate multinucleated cells. Fusion-defective cells—due to CRISPR-induced loss-of-function mutations—do not form syncytia and would survive. In contrast, the control group was co-cultured with  $DTA_N$ -Intein $_N$  expressing myoblasts, so no active toxin could form, and both fused and unfused cells remain viable (Fig. 4a).

By design, if a gRNA targets a gene that is dispensable for myoblast fusion, cells carrying that gRNA will normally fuse and be eliminated in the killer group. In contrast, if a gRNA targets a gene that is essential for myoblast fusion, cells carrying that gRNA will survive and be enriched in the killer group (Fig. 4b). Therefore, genes targeted by multiple gRNAs that show significant enrichment in the killer group relative to the control would be classified as essential for myoblast fusion.

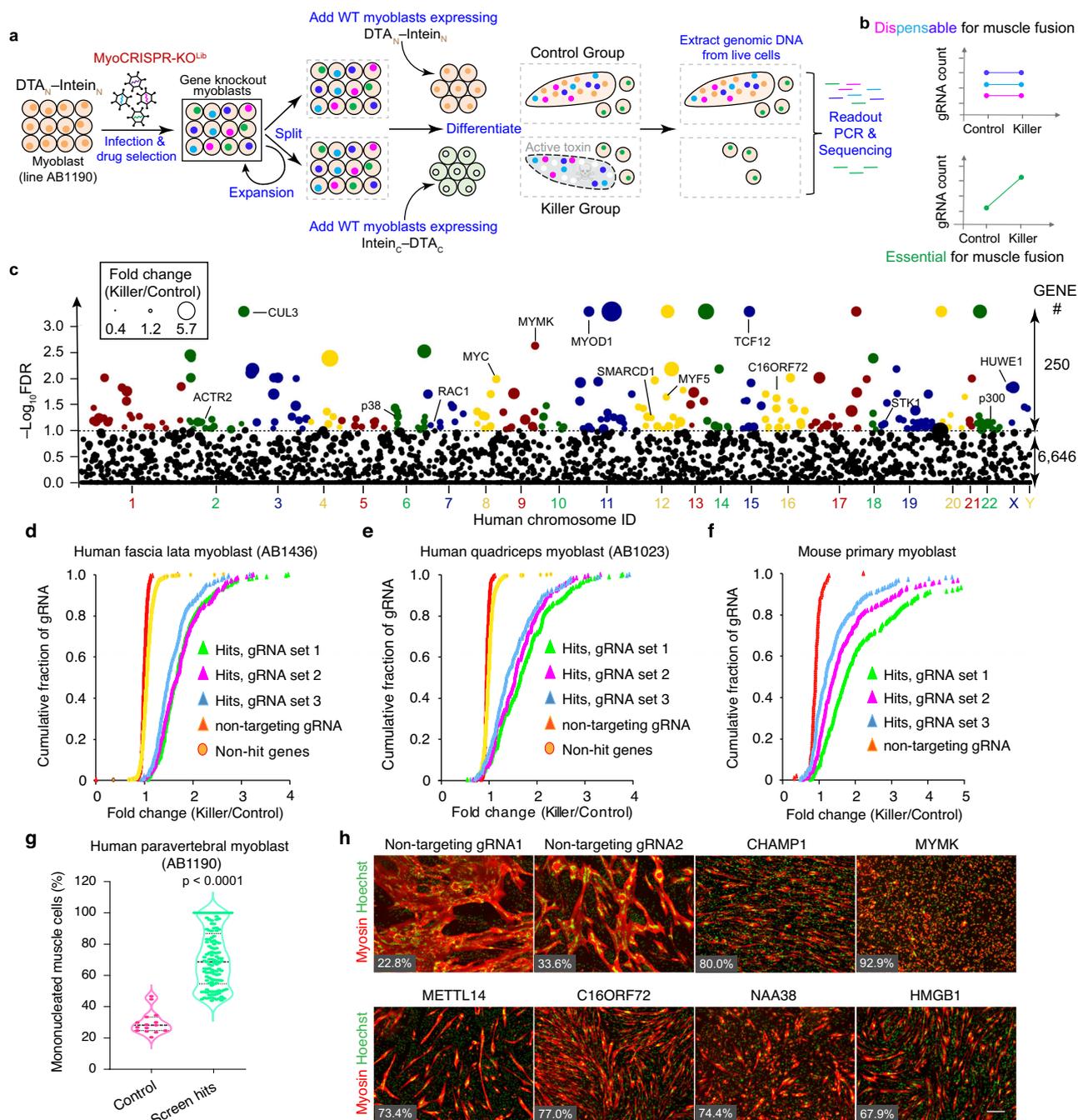
By this screen, we identified 250 hits (FDR < 0.1;  $p$  value < 0.0035) (Supplementary Data 5). Notably, fourteen hits are known myogenesis genes<sup>4,9,43–48</sup> including the muscle fusogen Myomaker (MYMK), myogenic regulators Myf5 and MyoD, and MyoD-associated inducers (CUL3, p38) and cofactors (TCF12, p300) (Fig. 4c). Most non-targeting negative control gRNAs were enriched in the control group (Supplementary Fig. 3a, b), likely due to passive statistical compression caused by the abundant enrichment of hits in the killer group. Although gRNAs for Myomixer and Myogenin, two well established regulators of myoblast fusion<sup>5,10</sup>, showed weak yet significant enrichment, they did not emerge to the top of the gene list (Supplementary Fig. 3c, d). For Myomixer, one of the two gRNAs targeted the C-terminal coding region that is dispensable for the protein function<sup>10</sup>, possibly limiting gene KO efficiency (Supplementary Fig. 3e).

To validate the screen hits, we employed two complementary approaches. First, we generated a focused CRISPR library targeting the top 250 hits and repeated the split toxin-based fusion screen in two additional human myoblast lines, derived from fascia lata (Supplementary Fig. 4a) and quadriceps (Supplementary Fig. 4b), as well as in mouse primary myoblasts (Supplementary Fig. 4c). In all cells, gRNAs targeting candidate genes showed consistent enrichment in the killer group, despite some variations in gRNA efficacy (Fig. 4d–f). Second, we individually validated 125 randomly selected hits (Supplementary Data 6) in human myoblasts without using split toxin. Disruption of these genes led to significantly increased ratios of unfused cells compared to non-targeting control gRNAs, indicating impaired fusion (Fig. 4g, h). These results demonstrate the robustness of our fusion screen across orthogonal genetic approaches and in both human and mouse myogenic models.

### High functional connectivity and clinical relevance of myoblast fusion screen hits

To explore the functional relationships among the hits identified in the myoblast fusion screen, we performed STRING analysis and constructed protein-protein interaction (PPI) networks based on experimentally validated associations<sup>49</sup>. This analysis revealed 1286 PPIs among the screen hits, significantly more than expected by chance ( $p$  value <  $1e-16$ ) when compared to a randomized protein sets matched for the size and degree of distribution.

These interactions involve 160 fusion screen hits that cluster into 23 protein complexes (Fig. 5), the majority of which have not previously been functionally linked to skeletal muscle development in any species. Notable examples include the SUMO activation complex



**Fig. 4 | Proof-of-concept test of myoblast fusion screens uncovers a large group of novel hits.** **a** Schematic of the CRISPR screen design to identify essential regulators of human myoblast fusion. Each color represents a unique gRNA and knockout of a specific gene. **b** Hypothetical results depict the potential effects of gene knockout and the interpretation of relative gRNA abundance between the killer and control groups. **c** Manhattan plot showing the chromosomal distribution of 6896 genes tested in the myoblast fusion screen. Colored dots indicate 250 hits with an false discovery rate (FDR) < 0.1 (corresponding  $p$  value < 0.0035). Names of 14 genes with known roles in myogenesis are highlighted. Circle sizes represent fold changes in gene enrichment (killer/control). Statistical analysis was performed using the MAGeCK pipeline with default settings. Exact FDR values and  $p$  values are provided in Supplementary Data 5. **d-f** Cumulative distribution of non-targeting gRNAs (negative controls) and gene-targeting gRNAs across various myoblast cell

lines. Data are from three biological replicates. For experiments involving mouse primary myoblasts, due to cell number limitations, randomly selected 1000 gRNAs for genes with an FDR > 0.9 (labeled as non-hit genes in **d,f**) in the initial screen were omitted from validation. For panels **d** and **e**, the human donor ID is provided in the bracket. For all panels, cells were differentiated for six days. **g** Percentage of mononucleated cells following full-term myogenic induction in individual gene knockout experiments. Gene list can be seen in Supplementary Data 6. \*\*\*,  $p < 0.001$ , student  $t$  test, two-sided. Two-sided student  $t$  test,  $p$  value =  $3.9E-14$ . **h** Representative myosin immunostaining results from bulk phenotype analysis. The percentage of mononucleated cells relative to the total nuclei is displayed in the bottom-left corner of each image. Scale bar: 200  $\mu$ m. The experiment was independently repeated twice with similar results. Source data are provided as a Source data file.

(SAE1, UBA2, UBE2I), ISWI chromatin remodelers (SMARCA5, RBBP4, BPTF), m6A mRNA modification machinery (METTL3, METTL14, WTAP, ZC3H13), negative cofactor 2 complex (DRI, DRAP1), chaperonin TCPI complex (CCT3, CCT5, TCPI), HOPS complex (VPS33A, VPS39), RNA stability regulators (ILF3, DHX9), N-terminal acetylation complex (NAA30, NAA35, NAA38), and dynein complex (ACTR10, DCTN1, DCTN5, CAPZB, DYNLRB1) etc.

To assess clinical relevance of our discovery, we cross-referenced the fusion screen hits with medical phenotype databases<sup>50</sup>. This revealed 41 genes whose mutations cause human diseases characterized by abnormal skeletal muscle morphology (Phenotype Ontology ID: 0011805) (Supplementary Data 7). These genes are labeled with an alpha ( $\alpha$ ) in Fig. 5 and Supplementary Fig. 5. For instance, mutations in *CHAMPI* gene, a novel fusion screen hit, cause neurodevelopmental disorders characterized by muscle developmental delay, hypotonia, muscle weakness and facial dysmorphism<sup>51–54</sup>. Currently, little is known about how *CHAMPI* functions in muscle cells and why patient mutations lead to muscle symptoms.

A parallel cross-referencing analysis of the fusion screen hits with mouse phenome data identified 48 genes whose mutations cause abnormal prenatal growth/weight/body size (Mammalian Phenotype ID: MP 0004196). These genes are labeled with a beta ( $\beta$ ) in Fig. 5 and Supplementary Fig. 5. Together, these findings underscore the in vivo functional roles and clinical significance of the genes uncovered in our myoblast fusion screen.

### Single-cell CRISPR and RNA-seq profiling of the cellular mechanisms underlying the myoblast fusion defects for the hits

Defective myoblast fusion can arise from disruptions of muscle fusion machinery (e.g., loss of Myomaker) alone or with a broader impairment in the muscle differentiation program (e.g., loss of MyoD). If the fusion defect is accompanied by a failure to initiate or progress through myogenic differentiation, this should be reflected in the transcriptomic profile of the affected cells. Therefore, we applied a split-barcoding based single-cell RNA-sequencing (scRNA-seq) technology<sup>55</sup> to unbiasedly survey transcriptomic changes for each fusion defective myoblast that survived in our fusion screens.

Toward this goal, we employed a lentiviral CROP-seq construct that expresses regular and polyadenylated gRNAs<sup>56</sup>, permitting simultaneous CRISPR gene-editing and the detection of gRNA by scRNA-seq, respectively (Supplementary Fig. 6). WT human myoblasts were transduced with newly generated lenti-CRISPR CROP-seq library containing gRNAs targeting 250 screen hits at low infection rates, ensuring single-gene perturbation per cell (Fig. 6a). These cells underwent a similar cell mixing procedure as the initial myoblast fusion screen, allowing enrichment and collection of surviving mononucleated cells in the killer group before scRNA-seq analysis. Most cells were collected after full-term myogenic induction in differentiation medium (DM), while a subset maintained in growth medium (GM) served as undifferentiated controls.

Cell clustering and expression analysis identified 10 distinct clusters (Fig. 6b'), with differentially expressed genes for each cluster summarized in Supplementary Data 8. Among them, cluster #4, cultured under GM-conditions, exhibited elevated expression of proliferation markers, including *MKI67*, *CCND1*, *BUB1*, *CENPF*, *TOP2A*, as expected. The remaining nine clusters are from DM-culture conditions. Four of these clusters (#1, #2, #5, #7) exhibited abundant expression of myogenic markers including Myomaker and muscle structural genes (e.g., *MYOG*, *MEF2C*, *MYH3*, *MYH8*, *MYH7*, *CASQ2*, *RYR1*, *TNNI1*), whereas five others (#3, #6, #8, #9, #10) lacked expression for these markers, suggesting compromised muscle differentiation (Fig. 6b').

As expected, gene ontology analysis of differentially expressed genes between the GM and DM conditions highlighted pathways associated with muscle development (Fig. 6c). From the top 1% of the differentially expressed genes, we identified 253 myoblast markers and 142 myocyte markers (Supplementary Data 9). These genes were used in an algorithm for calculating expression-based “muscle differentiation scores (MDS)” for individual cells detected in our scRNA-seq CRISPR experiment.

To increase the power of statistical analysis, we aggregated all cells expressing gRNAs targeting genes from the same protein complex or pathway. This analysis revealed that cells transduced by gRNAs targeting the *MyoD*–*TCF12*–*p300* cofactor complex had the lowest differentiation scores, while cells with gRNAs for *Myomaker* (*MYMK*) scored the highest (Fig. 6b''), d, e). These results align well with previous finding that MyoD deletion abolishes muscle differentiation and fusion, while Myomaker KO eliminates fusion without impairing myoblast differentiation<sup>9,25</sup>. Cells targeted by gRNAs from other protein complexes also showed reduced differentiation scores compared to MYMK gRNA-expressing cells, albeit to varying degrees (Fig. 6d, e, and Supplementary Fig. 7), suggesting broader defects in myogenic differentiation.

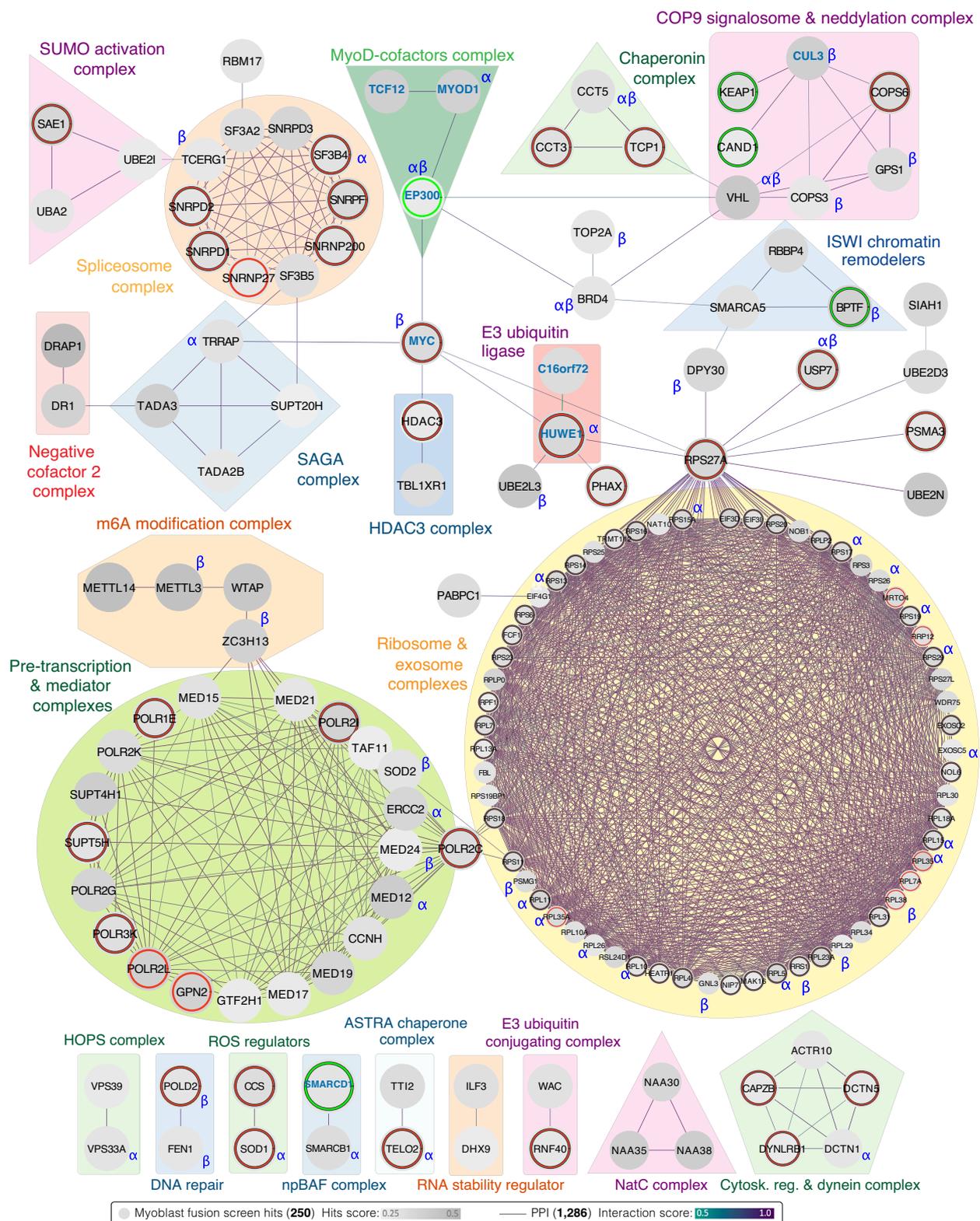
We then assessed the enrichment of gRNAs in each cell clusters. Along with Myomaker, gRNAs for pre-transcription and mediator complex, ribosome, *YPEL5* (Yippee Like 5) and *RBMX* (RNA Binding Motif Protein X-Linked) were enriched in the well-differentiated cell clusters (#1, #2, #5, #7) (Fig. 6f). In contrast, gRNAs targeting *MyoD*, its cofactors (*TCF12*, *P300*), and other complexes were enriched in the poorly differentiated clusters showing distinct gene signatures (Fig. 6f). For instance, gRNAs targeting components of m6A methyltransferase complex were significantly enriched in undifferentiated cell cluster #6, characterized by elevated expression of fibrin, type V collagen, and extracellular remodeling enzyme ADAMTS12 (Fig. 6f).

To validate findings from the scRNA-seq analysis, we focused on the m6A methyltransferase complex, which includes METTL3 and METTL14 (catalytic subunits) and WTAP (a regulatory subunit)—all of which emerged as hits in our myoblast fusion screen (Fig. 5) and were predicted by scRNA-seq data to affect muscle differentiation. Using CRISPR-induced mutagenesis in human myoblasts, we generated stable gene KO lines, each carrying biallelic frameshift mutations in one of these genes (Supplementary Fig. 8a). Indeed, immunostaining of myosin revealed pronounced defect of myoblast fusion as well as differentiation in all KO clones (Supplementary Fig. 8b). Transcriptome profiling of METTL3 deficient myoblasts further confirmed a significant reduction of *MYMK* and muscle differentiation markers (Supplementary Fig. 8c) and cell adhesion molecules (Supplementary Fig. 8d), matching the finding from our scRNA-seq experiment. These results suggest that myoblast fusion screen hits could regulate human myoblast fusion with a broader effect on early-stage myogenic differentiation.

## Discussion

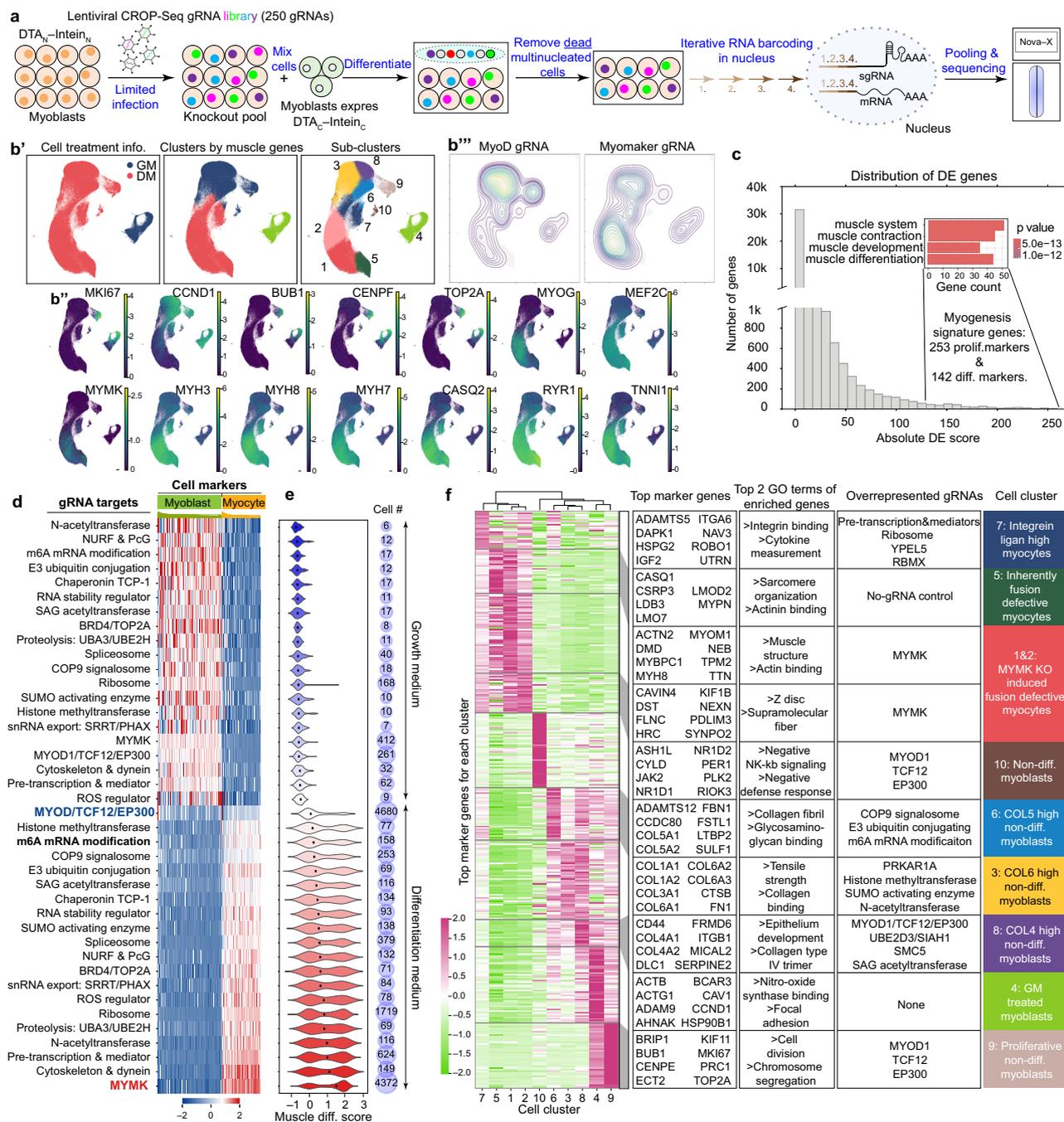
We developed and carefully validated a genetic screening platform and conducted high-coverage screens that identified a large group of previously unrecognized regulators essential for activating human myoblast fusion. CRISPR perturbation of these genes showed myoblast fusion phenotypes when validated in cells derived from various human muscle tissues of different donors, as well as in mouse primary myoblasts.

The strength of high-coverage screens lies in their ability to identify genes essential for multi-stages of tissue development, in this case, myoblast proliferation and fusion. Notably, 93 of the 236 fusion screen hits were associated with fundamental cellular functions, including genes crucial for the composition and func-



**Fig. 5 | High functional connectivity and clinical relevance of myoblast fusion screen hits.** STRING analysis identified 1286 protein-protein interactions (PPI, shown as connecting lines) among 160 myoblast fusion screen hits (represented as circles). These proteins form 23 distinct complexes, each highlighted in a different

color. Red and green circles denote knockouts that decreased or increased myoblast fitness, respectively, in our myoblast fitness screen. Gene names in blue indicate factors previously known to regulate myogenesis.



**Fig. 6 | Single-cell CRISPR & transcriptome profiling of myoblast fusion screen hits.** **a** Schematic of the experimental design. The CROP-seq, a lentiviral gRNA construct, was utilized to express regular gRNAs (for knockout) and polyadenylated gRNAs (for scRNA-seq detection) targeting 250 myoblast fusion screen hits. This enabled simultaneous scRNA-seq, CRISPR knockout & gRNA detection. **b'** Force-directed graph layout of 10 cell clusters. Marker genes for each cluster was provided in Supplementary Data 8. **b''** UMAP plot illustrating the expression of myoblast markers (MKI67, CCND1, BUB1, CENPF, TOP2A) in cluster #4 and Myomaker & muscle structural genes which are predominantly expressed in DM clusters (#1, #2, #5, #7). **b'''** Contour density plots of cells assigned to MYMK and MyoD gRNAs in UMAP space. **c** GO analysis of top differentially expressed (DE) genes in cluster #4 versus #1/2/5/7, collectively referred to as “myogenesis signature genes”

(see also Supplementary Data 9). Prolif: proliferation; Diff: muscle differentiation. Statistical analysis was performed using the clusterProfiler in R. **d** Heatmap displaying the median relative expression (column z-score) of myogenesis signature genes (columns), aggregated across cells expressing gRNAs targeting genes from the same protein complex or pathway (rows). **e** Distribution of myogenesis signature intensity across single cells (left) and the number of cells (right) for each pathway/protein complex. Black dots indicate the median. **f** Heatmap of differentially expressed marker genes in each cluster. A few top up-regulated genes for each cluster are shown. Genes upregulated in multiple clusters are assigned to the cluster with the highest fold change. To the right of the heatmap, for each cluster, are (left to right): top marker genes, top GO terms representing cell makers, overrepresented sgRNAs by odds ratio, and the assigned cluster name.

tion of spliceosome, ribosome, and pre-transcription complexes, which also emerged as significant hits in our myoblast fitness screen. Despite using the same set of gRNAs for both screens, hypomorphic rather than null mutations likely predominated in the fusion screen, allowing these cells to survive and be detected in the fusion screen. Partial disruption of these genes may have arrested the cells in a quiescent state with minimal activity, impairing myoblast fusion. Additional power of unbiased genetic screens lies in revealing networks connecting distinct pathways. For instance, ZC3H13 from the m6A RNA methylation pathway interacts with proteins in the mediator and pre-transcription complexes, which then connect to ribosome/exosome complexes via POLR2C. These connections highlight unknown interplays among cellular pathways essential for human myogenesis.

Our single-cell CRISPR & RNA-seq profiling enabled sensitive measurement of myogenic differentiation by leveraging a large panel of gene markers, providing more accurate assessments than conventional approaches relying on a few myogenic markers. Results from these experiments suggest that many of our myoblast fusion screen hits regulate myoblast fusion by controlling a broader myogenic program upstream of myoblast fusion. Using m6A regulatory complex as an example, we validated this finding through gene-KO experiment and transcriptome analysis. Indeed, genetic deletion of the key players (e.g., METTL3, METTL14, WTAP) in this complex resulted in marked defects in both myoblast differentiation and fusion. Notably, Myomaker expression was strongly reduced. A limitation for our single-cell CRISPR & RNA-seq analysis is the relatively low number of cells per gene, thus we cannot determine whether a broader effect of differentiation might also apply to singlet hits, i.e., those without known protein-protein interaction partners among the fusion screen hits. We anticipate that future functional studies for each individual hit will offer fundamental insights into the regulation of myoblast differentiation and fusion.

Despite the abundant new discoveries, our study has several limitations. First, due to the inherently high fusion index (>90%) of human myoblasts, our screen was not suited to identify negative regulators of myoblast fusion—such as components of the Notch, TGF $\beta$  and ERK signaling pathways<sup>20–24,57–59</sup>. Second, myoblast fusion only represents a narrow time window of myogenesis. Therefore, our screen cannot uncover mechanisms involved in the earlier crucial stages of human myogenesis, such as myogenic commitment<sup>3,7,60,61</sup>. We anticipate that genes identified in our screen may play multifaceted roles prior to or following myoblast fusion or in other tissues, which could influence the extent and severity of muscle symptoms in patients carrying mutations in these genes. Third, as with other genetic screens, not all known regulators of muscle fusion were recovered in our preliminary screen. This may reflect interspecies differences, functional redundancy (e.g., cadherins, talin)<sup>62–64</sup> or limitations of our CRISPR library related to gRNA number and efficiency (e.g., MYMX, MYOG)<sup>5,10</sup>. While most genes in the MyoCRISPR-KO<sup>Lib</sup> are targeted by three gRNAs, some genes (e.g., MYMX) are represented by only two gRNAs due to constraints of the source library. As expected, in a subsequent round of myoblast fusion screens targeting genes with intermediate confidence (0.1 < FDR < 0.9), the statistical significance of some weaker hits, including MYMX and MYOG, improved substantially when six gRNAs were used, as is standard in genome-wide screens. We anticipate that new screens employing larger yet muscle targeted CRISPR libraries with more gRNAs for genes missed in our preliminary screen (FDR > 0.9) will increase statistical power and thereby enhance the sensitivity of hit detection.

Collectively, our findings demonstrate the value of a rationally designed CRISPR screening platform in uncovering novel regulators of muscle development. Expanding the application of this approach will facilitate the genetic dissection of muscle disorders characterized by impaired differentiation and fusion.

## Methods

### Human myoblast isolation, culture and karyotype analysis

Human myoblasts were generated previously<sup>35</sup> and were obtained as de-identified materials. Primary cultures from muscle biopsies were co-transduced with two retroviral vectors expressing hTERT and CDK-4 cDNAs. Co-transduced cells were selected by neomycin and puromycin and then purified using magnetic beads coupled to antibodies directed against the myogenic marker CD56 (NCAM1). Individual clones with high myogenic and fusion capacity were selected (AB1190, AB1079, AB1436, AB1167, AB1023DMD11Q, KM1421, AB1071DMD13PV). Human myoblasts are available upon request from Dr. Vincent Mouly at [vincent.mouly@upmc.fr](mailto:vincent.mouly@upmc.fr). Possibly due to the immortalization and clonal expansion, these human myoblasts loss expression of PAX7. Donor information for each human myoblast line is provided in the Supplementary Fig. 1b. Human myoblasts are maintained in skeletal muscle cell basal medium (PromoCell, C-23260) supplemented with 15% fetal bovine serum (FBS; Sigma-Aldrich, F2442), 5% growth medium supplement mix (PromoCell, C-39365), GlutaMAX, and 1% gentamicin sulfate. For differentiation, myoblasts were cultured in differentiation medium comprising DMEM supplemented with 2% horse serum, 10  $\mu$ M DAPT (Cayman, 13197), and 1% penicillin/streptomycin. Karyotype analysis was performed by KaryoStat™ Assay using Genechip Probe Array containing 1.1 million probes across somatic and sex chromosomes.

### Mouse primary myoblast isolation and culture

Primary myoblasts were isolated from the hind limb skeletal muscles of 3-week old male and female mice (Jackson Laboratory, C57BL/6J). The muscle tissues were collected, minced and digested with collagenase type I and dispase B for around 40 min. The digestions were stopped with F-10 Ham's medium containing 10% FBS and centrifuged at 450 g for 5 min. Then the cells were seeded on collagen-coated dishes and cultured in growth medium containing F-10 Ham's medium, with 20% fetal bovine serum (FBS), 4 ng/mL basic fibroblast growth factor, and 1% penicillin–streptomycin at 37 °C with 5% CO<sub>2</sub>. The medium was changed every 2 days. Cells at passage 3–5 were used for the experiments.

### Retroviral expression vector cloning and retrovirus preparation

The retroviral expression vector pMXs-Puro (Cell Biolabs, RTV-012) was used to construct gene expression vectors for this study. The CHAMP1 open reading frame and its mutants were codon-optimized and synthesized by Integrated DNA Technologies, with sequences verified through Sanger sequencing or whole plasmid sequencing. For rescue experiments, sgRNA-insensitive DNA cassettes with silent mutations destroying the protospacer or PAM sequence were designed and utilized. Retrovirus production was performed by transfecting human embryonic kidney 293 cells with the retroviral plasmid using FuGENE 6 (Promega, E2692). Viral supernatant was harvested two days post-transfection, filtered, and used to infect target cells in the presence of polybrene (Sigma-Aldrich, TR1003-G). Infected cells were switched to fresh culture medium 24 h post-infection. The DTA-Intein construct structures and protein sequences are provided below.

DTA<sub>N</sub>-Intein<sub>N</sub>

MGADDVVDSSKSFVMENFSSYHGTKPGYVDSIQKGIQPKSGTQGNVDDDKWGFYSTDNKYDAAGYSVDNENPLSGKAGGVVKVYTPGLTKVLAL  
 KVDNAETIKKELGLSLTEFLMEQVGTTEEFIKRFGDASRVVLSLPAEAGSSVEYINNWEQAKALCLSYETEILTVEYGLLPVIGKIVEKRIEATV  
 YSVDNNGNIYTQVPAQWHRGEQVFEYCLEDGSLIRATKDKHKFMTVDGQMLPIDEIFERELDMRVNDLPNKAFMVTDEDIRKQEEELVQVRRK  
 LEEALMADMLA\* MBD3 coiled coil

Intein<sub>C</sub>-DTA<sub>C</sub>

p66a coiled coil  
 PEERERMIKQLKEELRLEEAKLVLLKLRQSQIQKEATAQKSGSMIKIATRKYLGKQNVYDIGVERDHFALKNGFIAASCME  
 LEINFETRGRKRGDAMYEMAQASAGNR\*

## Selection of target genes and the generation of MyoCRISPR-KO KO sub-libraries

Protein-coding genes to be targeted by MyoCRISPR-KO libraries were selected basing on the RNA-seq TPM value >1 in human myoblasts, either prior to or following myogenic differentiation. To generate the library, gRNA sequences for the selected genes were extracted from previously published CRISPR libraries<sup>39,40</sup>. gRNA sequences were fused with BsmBI restriction sites and primer binding regions were synthesized as oligo pools by GenScript (Supplementary Data 2).

To minimize batch effects and increase the maneuverability of the cloning and application of the library, the MyoCRISPR-KO<sup>lib</sup> was divided into nine sub-libraries, each containing ~2400 gRNAs—comprising 100 non-targeting control gRNAs and ~2300 gene-targeting gRNAs (Supplementary Data 1). For each sub-library, gRNAs were amplified from the master oligo pool using a specific pair of primers via PCR. The resulting PCR products were purified and subsequently cloned into the plentiCRISPR v2-Blast gRNA vector (Addgene, 83480) using Golden Gate assembly. The representation and accuracy of gRNA sequences in the CRISPR plasmid library were validated through Illumina sequencing (Admera) of gRNA readout PCR amplicons. The gRNA library can be requested through contacting the corresponding author.

## Muscle-targeted CRISPR screens in human myoblasts

Lenti-MyoCRISPR-KO KO sub-libraries were produced by transfecting HEK293 cells (Clontech, 632180) with CRISPR plasmid libraries and lentiviral packaging vectors, pMD2.G and psPAX2, using Fugene6. After 18 h, the medium was replaced with fresh DMEM containing 10% fetal bovine serum. Lentiviral supernatant was harvested 72 h post-transfection, filtered through a 0.45 μm syringe filter (ThermoFisher), and concentrated using the Lenti-X Concentrator (Takara). The viral pellet was resuspended in Opti-MEM and stored at -80 °C. The pMD2.G and psPAX2 vectors were gifts from D. Trono (Addgene, 12260 and 12259). Lentivirus titers were assessed using a blasticidin selection assay with gradient viral dilutions.

Each sub-library was used to infect 42.2 million myoblasts (line# ABI190, at passage 4 that stably expressing DTA<sub>N</sub>-Intein<sub>N</sub>) per replicate, ensuring 17,575-fold library coverage—35 times the standard CRISPR screen coverage<sup>38</sup>. The infection was controlled at a MOI of 0.5. Infected cells were selected with 10 μg/mL blasticidin for one week, then split into two groups: a control group mixed with the same split toxin DTA<sub>N</sub>-Intein<sub>N</sub> myoblasts and a killer group mixed with the myoblasts expressing complementary toxin Intein<sub>C</sub>-DTA<sub>C</sub> at a 1:1 ratio. Each group has three replicates. After 24 h, cells were switched to differentiation medium and harvested for genomic DNA extraction 5–7 days later, when large myotubes in the killer group are dead and detached from the culture dish. Genomic DNA was extracted using the MasterPure Complete DNA and RNA Purification Kit (Lucigen, MC85200) per the manufacturer's instructions. gRNA sequences were PCR-amplified using Herculase II Fusion DNA Polymerase (Agilent, 600679) with primers specific for readout PCR1. Each reaction

contained 5 μg of genomic DNA in a 50 μL volume, with cycling conditions: 98 °C for 2 min, 20 cycles of 98 °C for 30 s, 60 °C for 30 s, and 72 °C for 45 s, followed by a final extension at 72 °C for 3 min. PCR1 products were pooled and used as templates for a second round of PCR with indexing primers under similar conditions, reduced to 11 cycles. PCR products were visualized on a 1.5% TAE-agarose gel and purified using 0.9× Agencourt AMPure XP Beads (Beckman). Indexed PCR products were pooled and sequenced on an Illumina platform with a sequencing coverage >1000× of the gRNA library size for each sample. DNA sequencing data were analyzed using MAGeCK<sup>65</sup> with default settings. Enrichment of gRNAs in the killer versus control groups determined the direction of positive or negative scores. Protein interaction networks were analyzed using STRING and visualized using Cytoscape<sup>66</sup>, displaying interactions with experimentally determined scores above 0.7 (high confidence).

## Single-cell CRISPR screen and transcriptome profiling

A CRISPR-Cas9 mini-library targeting 250 candidate genes was cloned into the CROPseq-Guide-Puro which was a gift from Christoph Bock (Addgene plasmid # 86708)<sup>56</sup>. To increase the representation of positive controls, the gRNAs for MyoD and Myomaker were cloned at a higher abundance. Human myoblasts at passage 4 expressing DTA<sub>N</sub>-Intein<sub>N</sub> were infected with the lentiviral library at a low MOI (MOI = 0.1), ensuring each cell received a single gRNA. Five days post-infection, these myoblasts were mixed with Intein<sub>C</sub>-DTA<sub>C</sub>-expressing human myoblasts at a 1:1 ratio. The mixed cells were cultured in differentiation medium for six days following a 24 h co-culture period. After the depletion of large myotubes, surviving muscle cells were collected and processed for single-cell RNA sequencing using the Evercode Whole Transcriptome kit (Parse, ECW02130) per the manufacturer's protocol. Two sequencing libraries were generated: a whole transcriptome library and a CRISPR gRNA library, both sequenced on Illumina platforms. Sequencing data analysis was conducted using the ParseBiosciences-Pipeline-v1.1.2.

## Single-cell CRISPR screen analysis

The raw fastq trimming, alignment and generation of cell-gene matrices was performed using the Parse split-pipeline 1.1.2 following the manufacturer's instructions. The cell-gene matrix data was processed using the *Scanpy* package<sup>67</sup> (version 1.9.6) in *Python* 3.8.18. Briefly, the raw count data was imported from a sparse matrix format (*.mtx* file) along with corresponding gene annotation and cell metadata files. Prior to analysis, genes with missing annotations were removed from the dataset. Gene names were assigned as unique identifiers, cell barcodes and CRISPR guide RNA information were incorporated as observation names in the *AnnData* object.

Initial quality control filtering was performed to remove cells expressing fewer than 300 genes and genes detected in fewer than 5 cells. To ensure reliable transcript counts, cells were further filtered to retain only those with total transcript counts above 2000. Additional quality control was performed to filter out the cells with more than

60,000 transcripts and 10,000 detected genes to remove potential doublets. Cells with mitochondrial gene content greater than 40% were excluded to remove low-quality or dying cells. After quality control, we obtained transcriptomes for 295,467 cells with an average of 20,127 reads per cell (5.9 billion total reads). The median number of transcripts and genes detected per cell was 3755 and 2046, respectively.

The filtered data underwent total count normalization with a target sum of 10,000 counts per cell and then log-transformed using the natural logarithm to stabilize variance. Highly variable genes were identified using *Scanpy's highly\_variable\_genes* function with the following parameters: minimum mean expression: 0.015, maximum mean expression: 3.0, minimum dispersion: 0.5.

For dimensionality reduction and clustering, principal component analysis (PCA) was firstly performed using the *ARPACK* solver and retained the top 30 principal components. Then 15 nearest neighbors were used for neighborhood graph construction. After that, Uniform Manifold Approximation and Projection (*UMAP*)<sup>68</sup> was applied for non-linear dimensionality reduction with minimum distance set to 0.1 and spread parameter of 2. Finally, the *Leiden* clustering algorithm was used for cell clustering with a resolution of 0.6, resulting in ten distinct clusters<sup>69</sup>.

For the generation of the custom MDS, the clusters were manually annotated based on muscle differentiation marker genes, categorizing them into three main populations: differentiated muscle cell, non-differentiated muscle cell, and growth medium (GM) muscle cell. Differential expression analysis was then performed using the Wilcoxon rank-sum test, comparing differentiated muscle cells against GM muscle cells. Only cells from the no CRISPR guide RNA infection condition were included in this analysis. After DE analysis, top 1% of differentially expressed genes were identified as signature genes based on absolute DE scores. The genes were then separated into two categories: positive gene that enriched in differentiated muscle cells and negative genes enriched in GM muscle cells. A coefficient was then defined to each gene according to the mean expression in corresponding cell population. The coefficient was then normalized, ensuring the value in positive genes sum to 1 and negative genes sum to -1. The MDS for each single cell was then calculated as

$$MDS = w \cdot x = \sum_{i=1}^n w_i x_i \quad (1)$$

where  $x = [x_1, x_2, \dots, x_n]$  (2) represents the vector of expression levels for  $n$  genes in a single cell. Each  $x_i$  is the expression level of gene  $i$ ; and  $w = [w_1, w_2, \dots, w_n]$  (3) represents the vector of coefficients corresponding to each gene. The resulting MDS provides a quantitative measure of cellular differentiation state, with higher positive scores indicating greater differentiation.

### RNA extraction and bulk RNA sequencing

Total RNA was extracted from human myoblast cells using TRIzol reagent (Thermo Fisher Scientific, 15-596-018) according to the manufacturer's protocol. RNA quality and concentration were assessed using a TapeStation, and only samples with an RNA Integrity Number (RIN) above 9 were used for library preparation using the NEBNext Ultra II Directional RNA Library Prep Kit (E7760L) with poly(A) selection. Next-generation sequencing was performed on an Illumina platform by Admera, generating 150 bp paired-end reads with a depth of approximately 40 million reads per sample.

### Bulk RNA sequencing data analysis

Raw sequencing data was processed with initial quality control (QC) using *FastQC* (version 0.11.9: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) to assess read quality and *MultiQC*<sup>70</sup> (version 1.14) was used to aggregate and visualize quality control metrics across all samples. After QC, raw sequencing reads were processed using *fastp*<sup>71</sup>

(version 0.23.2) for adapter trimming and reads filtering with the following filtering parameters: quality score threshold: 20, maximum proportion of unknown bases (N): 10, maximum percentage of unqualified bases: 30%. A reference human *hg38* genome index was constructed using *HISAT2* (version 2.2.1) with the provided genome sequence<sup>72</sup>. Cleaned reads were aligned to the reference genome using *HISAT2*. Alignment outputs were converted to sorted BAM files using *SAMtools*<sup>73</sup> (version 1.6), and BAM files were indexed for downstream analysis. Aligned reads were converted to normalized genome browser tracks (BigWig files) using *deepTools*<sup>74</sup> (version 3.5.2) with 1 bp bin size for track visualization. Gene-level read counts were quantified using *featureCounts*<sup>75</sup> (Subread version 2.0.3) and differential expression analysis was performed using *DESeq2*<sup>76</sup> (version 1.40.2) in *R* (version 4.3.1) combined with Benjamini-Hochberg multiple testing correction. Gene Ontology (GO) enrichment analysis was conducted using *clusterProfiler*<sup>77</sup> (version 4.8.3) and the Homo sapiens annotation database (*org.Hs.eg.db*, version 3.17.0) in *R*. RNA-seq data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database under accession number: [GSE293514](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE293514).

### Gene set enrichment analysis (GSEA)

We utilized a previously published human in vivo embryonic skeletal muscle single-nuclei RNA sequencing dataset (EMBL Accession E-MTAB-8813). For data preprocessing, we first consolidate cell type annotations by combining related cell populations into muscle stem cell, myoblast, myocyte, *MYH3+* myocyte and *MYL3+* myocyte. Differential gene expression was performed comparing myocyte to myoblast populations using *Scanpy's* Wilcoxon rank-sum test. Based on the differential expression scores, top 200 upregulated genes and downregulated genes were selected and converted to *gmt* reference gene set, representatively. Transcriptome data (TPM) of human myoblast cell lines from seven donors in differentiation or growth medium condition were converted to *gct* dataset file. Multiple GSEA<sup>78</sup> were then performed to identify the correlation and similarity between in vivo embryonic muscle development and in vitro myoblast differentiation.

### Individual KO experiments in human myoblasts

sgRNAs targeting the coding regions of 125 fusion screen hits (Supplementary Data 6) were individually cloned into the LentiCRISPR v2-Blast vector and validated by Sanger sequencing. The LentiCRISPR v2 vector was generously provided by Mohan Babu (Addgene, plasmid 83480).

### Immunostaining and microscopy

Cells were fixed in 4% paraformaldehyde (PFA) in PBS for 10 minutes at room temperature, permeabilized with 0.5% Triton X-100 in PBS, and blocked with 3% bovine serum albumin (BSA) in PBS for 1 hour at room temperature. Primary antibodies (1: 500 dilution) used for immunostaining experiments are anti-MyoD (Novus Biologicals, NBPI-54153, NB100-56511, 1.0 mg/ml), anti-MYOG (DSHB, F5D), anti-myosin (DSHB, MF20), MYH3 (DSHB, F1.652). Following blocking, cells were incubated with the primary antibody overnight at 4 °C, then with an Alexa Fluor-conjugated secondary antibody (1:1000 dilution, Thermofisher Scientific, A21127, A21242, A27034, A27036, A28177). Nuclei were counterstained with Hoechst dye. Phase-contrast images were captured using a BioTek Lionheart FX automated microscope, and fluorescence images were acquired using either the BioTek system or an Olympus FLUOVIEW FV1200 confocal laser scanning microscope. Super resolution imaging was conducted with a Zeiss LSM 980 confocal microscope with Airyscan 2.

### Differentiation index and fusion index measurements

The differentiation index was calculated as the proportion of nuclei within MF20-positive cells relative to the total number of nuclei. The fusion index was determined as the proportion of nuclei within

myotubes (cells containing  $\geq 3$  nuclei) relative to the total number of myosin expressing nuclei. Both indices were derived from manual cell counts, with treatment information blinded to maintain objectivity.

### Statistics & reproducibility

All measurements were taken from distinct samples. Data were processed using GraphPad Prism 9 software. For data presented in Supplementary Data 3, 4, 5, the statistical analysis was performed using the MAGeCK pipeline with default settings. Statistical analysis was performed using MAGeCK pipeline with default settings. Error bars in all bar graphs represent standard deviation. Experiments were not randomized, and investigators were not blinded to sample allocation or outcome assessment. No statistical method was used to predetermine sample size. No data were excluded from the analyses.

### Ethical statement

All experiments involving recombinant DNA, human cells, third-generation lentiviral and retroviral vectors, and diphtheria toxin were conducted in compliance with institutional biosafety regulations and under the approval of the University of Georgia Institutional Biosafety Committee (IBC protocol no. 2023-0076). Human myoblasts were generated previously<sup>35</sup> and were obtained as de-identified materials. All animal experiments followed the ARRIVE guidelines and were approved by the University of Georgia Institutional Animal Care and Use Committee (IACUC; Animal Use Protocol A2025 01-008-Y1-A0). Mice were fed PicoLab Rodent Diet 20, irradiated (5053) with ad libitum access to chow and fresh water under standard vivarium conditions. No dietary restrictions were applied during experiments. Mice were housed in a UGA facility on a 12:12 h light/dark cycle, at 20 °C–22 °C, with 40–60% relative humidity, in accordance with UGA IACUC and the Guide for the Care and Use of Laboratory Animals.

### Reporting summary

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

### Data availability

All analyzed gene values for CRISPR screens are provided in Supplementary Data 3, 4, 5. All sequencing data for bulk RNA-seq and single-cell RNA-seq & CRISPR screens generated in this study is available in Gene Expression Omnibus (GEO) repository under the accession number [GSE293514](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE293514) and are publicly available. Source data are provided with this paper. All other data are available in the article and its Supplementary files or from the corresponding author upon request.

### Code availability

Code used for the DNA sequencing data analysis performed in this study is fully available on GitHub: [https://github.com/ZhengZhang991012/MyoCRISPRScreen\\_NatCom](https://github.com/ZhengZhang991012/MyoCRISPRScreen_NatCom).

### References

- Zhang, B. et al. A human embryonic limb cell atlas resolved in space and time. *Nature* **635**, 668–678 (2024).
- Bentzinger, C. F., Wang, Y. X. & Rudnicki, M. A. Building muscle: molecular regulation of myogenesis. *Cold Spring Harb. Perspect. Biol.* **4**, a008342 (2012).
- von Maltzahn, J., Jones, A. E., Parks, R. J. & Rudnicki, M. A. Pax7 is critical for the normal function of satellite cells in adult skeletal muscle. *Proc. Natl. Acad. Sci. USA* **110**, 16474–16479 (2013).
- Rudnicki, M. A. et al. MyoD or Myf-5 is required for the formation of skeletal muscle. *Cell* **75**, 1351–1359 (1993).
- Hasty, P. et al. Muscle deficiency and neonatal death in mice with a targeted mutation in the myogenin gene. *Nature* **364**, 501–506 (1993).
- Black, B. L. & Olson, E. N. Transcriptional control of muscle development by myocyte enhancer factor-2 (MEF2) proteins. *Annu. Rev. Cell Dev. Biol.* **14**, 167–196 (1998).
- Fung, C. W. et al. Cell fate determining molecular switches and signaling pathways in Pax7-expressing somitic mesoderm. *Cell Discov.* **8**, 61 (2022).
- Davis, R. L., Weintraub, H. & Lassar, A. B. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**, 987–1000 (1987).
- Millay, D. P. et al. Myomaker is a membrane activator of myoblast fusion and muscle formation. *Nature* **499**, 301–305 (2013).
- Bi, P. et al. Control of muscle formation by the fusogenic micropeptide myomixer. *Science* **356**, 323–327 (2017).
- Zhang, Q. et al. The microprotein Minion controls cell fusion and muscle formation. *Nat. Commun.* **8**, 15664 (2017).
- Quinn, M. E. et al. Myomerger induces fusion of non-fusogenic cells and is required for skeletal muscle development. *Nat. Commun.* **8**, 15665 (2017).
- Kim, J. H. & Chen, E. H. The fusogenic synapse at a glance. *J. Cell Sci.* **132**, jcs213124 (2019).
- Shilagardi, K. et al. Actin-propelled invasive membrane protrusions promote fusogenic protein engagement during cell-cell fusion. *Science* **340**, 359–363 (2013).
- Deng, S., Azevedo, M. & Baylies, M. Acting on identity: myoblast fusion and the formation of the syncytial muscle fiber. *Semin Cell Dev. Biol.* **72**, 45–55 (2017).
- Abmayr, S. M., Zhuang, S. & Geisbrecht, E. R. Myoblast fusion in *Drosophila*. *Methods Mol. Biol.* **475**, 75–97 (2008).
- Di Gioia, S. A. et al. A defect in myoblast fusion underlies Carey-Fineman-Ziter syndrome. *Nat. Commun.* **8**, 16077 (2017).
- Ramirez-Martinez, A. et al. Impaired activity of the fusogenic micropeptide Myomixer causes myopathy resembling Carey-Fineman-Ziter syndrome. *J. Clin. Invest.* **132**, e159002 (2022).
- Ganassi, M., Muntoni, F. & Zammit, P. S. Defining and identifying satellite cellopathies within muscular dystrophies and myopathies. *Exp. Cell Res.* **411**, 112906 (2022).
- Eigler, T. et al. ERK1/2 inhibition promotes robust myotube growth via CaMKII activation resulting in myoblast-to-myotube fusion. *Dev. Cell* **56**, 3349–3363 (2021).
- Bjornson, C. R. et al. Notch signaling is necessary to maintain quiescence in adult muscle stem cells. *Stem Cells* **30**, 232–242 (2012).
- Gioftsidi, S., Relaix, F. & Mourikis, P. The Notch signaling network in muscle stem cells during development, homeostasis, and disease. *Skelet. Muscle* **12**, 9 (2022).
- Girardi, F. et al. TGFbeta signaling curbs cell fusion and muscle regeneration. *Nat. Commun.* **12**, 750 (2021).
- Melendez, J. et al. TGFbeta signalling acts as a molecular brake of myoblast fusion. *Nat. Commun.* **12**, 749 (2021).
- Zhang, H. et al. Human myotube formation is determined by MyoD-Myomixer/Myomaker axis. *Sci. Adv.* **6**, eabc4062 (2020).
- Joshi, A. S. et al. The IRE1alpha/XBP1 signaling axis drives myoblast fusion in adult skeletal muscle. *EMBO Rep.* **25**, 3627–3650 (2024).
- Cramer, A. A. W. et al. Nuclear numbers in syncytial muscle fibers promote size but limit the development of larger myonuclear domains. *Nat. Commun.* **11**, 6287 (2020).
- Bryson-Richardson, R. J. & Currie, P. D. The genetics of vertebrate myogenesis. *Nat. Rev. Genet.* **9**, 632–646 (2008).
- Chal, J. & Pourquie, O. Making muscle: skeletal myogenesis in vivo and in vitro. *Development* **144**, 2104–2122 (2017).
- Hernandez-Hernandez, J. M., Garcia-Gonzalez, E. G., Brun, C. E. & Rudnicki, M. A. The myogenic regulatory factors, determinants of

- muscle development, cell identity and regeneration. *Semin Cell Dev. Biol.* **72**, 10–18 (2017).
31. Grimaldi, A. & Tajbakhsh, S. Diversity in cranial muscles: origins and developmental programs. *Curr. Opin. Cell Biol.* **73**, 110–116 (2021).
  32. Dobi, K. C., Schulman, V. K. & Baylies, M. K. Specification of the somatic musculature in *Drosophila*. *Wiley Interdiscip. Rev. Dev. Biol.* **4**, 357–375 (2015).
  33. Rahim, N. G., Harismendy, O., Topol, E. J. & Frazer, K. A. Genetic determinants of phenotypic diversity in humans. *Genome Biol.* **9**, 215 (2008).
  34. Sittig, L. J. et al. Genetic Background Limits Generalizability of Genotype-Phenotype Relationships. *Neuron* **91**, 1253–1259 (2016).
  35. Mamchaoui, K. et al. Immortalized pathological human myoblasts: towards a universal tool for the study of neuromuscular disorders. *Skelet. Muscle* **1**, 34 (2011).
  36. Ousterout, D. G. et al. Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy. *Nat. Commun.* **6**, 6244 (2015).
  37. Fischer, U., Ludwig, N., Raslan, A., Meier, C. & Meese, E. Gene amplification during myogenic differentiation. *Oncotarget* **7**, 6864–6877 (2016).
  38. Doench, J. G. Am I ready for CRISPR? A user's guide to genetic screens. *Nat. Rev. Genet.* **19**, 67–80 (2018).
  39. Read, A., Gao, S., Batchelor, E. & Luo, J. Flexible CRISPR library construction using parallel oligonucleotide retrieval. *Nucleic Acids Res* **45**, e101 (2017).
  40. Sanson, K. R. et al. Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat. Commun.* **9**, 5416 (2018).
  41. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
  42. Purde, V., Kudryashova, E., Heisler, D. B., Shakya, R. & Kudryashov, D. S. Intein-mediated cytoplasmic reconstitution of a split toxin enables selective cell ablation in mixed populations and tumor xenografts. *Proc. Natl. Acad. Sci. USA* **117**, 22090–22100 (2020).
  43. Papizan, J. B., Vidal, A. H., Bezprozvannaya, S., Bassel-Duby, R. & Olson, E. N. Cullin-3-RING ubiquitin ligase activity is required for striated muscle function in mice. *J. Biol. Chem.* **293**, 8802–8811 (2018).
  44. Wu, Z. et al. p38 and extracellular signal-regulated kinases regulate the myogenic program at multiple steps. *Mol. Cell Biol.* **20**, 3951–3964 (2000).
  45. Sartorelli, V., Huang, J., Hamamori, Y. & Keddes, L. Molecular mechanisms of myogenic coactivation by p300: direct interaction with the activation domain of MyoD and with the MADS box of MEF2C. *Mol. Cell Biol.* **17**, 1010–1026 (1997).
  46. Wang, S. et al. Tcf12 is required to sustain myogenic genes synergism with MyoD by remodelling the chromatin landscape. *Commun. Biol.* **5**, 1201 (2022).
  47. Puri, P. L. et al. Differential roles of p300 and PCAF acetyltransferases in muscle differentiation. *Mol. Cell* **1**, 35–45 (1997).
  48. Dilworth, F. J., Seaver, K. J., Fishburn, A. L., Htet, S. L. & Tapscott, S. J. In vitro transcription system delineates the distinct roles of the coactivators pCAF and p300 during MyoD/E47-dependent transactivation. *Proc. Natl. Acad. Sci. USA* **101**, 11593–11598 (2004).
  49. Szklarczyk, D. et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646 (2023).
  50. Putman, T. E. et al. The Monarch Initiative in 2024: an analytic platform integrating phenotypes, genes and diseases across species. *Nucleic Acids Res.* **52**, D938–D949 (2024).
  51. Tanaka, A. J. et al. De novo pathogenic variants in CHAMP1 are associated with global developmental delay, intellectual disability, and dysmorphic facial features. *Cold Spring Harb. Mol. Case Stud.* **2**, a000661 (2016).
  52. Isidor, B. et al. De Novo truncating mutations in the kinetochore-microtubules attachment gene CHAMP1 cause syndromic intellectual disability. *Hum. Mutat.* **37**, 354–358 (2016).
  53. Hempel, M. et al. De Novo mutations in CHAMP1 cause intellectual disability with severe speech impairment. *Am. J. Hum. Genet.* **97**, 493–500 (2015).
  54. Levy, T. et al. CHAMP1 disorder is associated with a complex neurobehavioral phenotype including autism, ADHD, repetitive behaviors and sensory symptoms. *Hum. Mol. Genet.* **31**, 2582–2594 (2022).
  55. Rosenberg, A. B. et al. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176–182 (2018).
  56. Datlinger, P. et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14**, 297–301 (2017).
  57. Relaix, F. et al. Perspectives on skeletal muscle stem cells. *Nat. Commun.* **12**, 692 (2021).
  58. Hung, M., Lo, H. F., Jones, G. E. L. & Krauss, R. S. The muscle stem cell niche at a glance. *J. Cell Sci.* **136**, jcs261200 (2023).
  59. Verma, M. et al. Muscle satellite cell cross-talk with a vascular niche maintains quiescence via VEGF and notch signaling. *Cell Stem Cell* **23**, 530–543 (2018).
  60. Sincennes, M. C. et al. Acetylation of PAX7 controls muscle stem cell self-renewal and differentiation potential in mice. *Nat. Commun.* **12**, 3253 (2021).
  61. Hicks, M. R. et al. Regenerating human skeletal muscle forms an emerging niche in vivo to support PAX7 cells. *Nat. Cell Biol.* **25**, 1758–1773 (2023).
  62. Conti, F. J., Monkley, S. J., Wood, M. R., Critchley, D. R. & Muller, U. Talin 1 and 2 are required for myoblast fusion, sarcomere assembly and the maintenance of myotendinous junctions. *Development* **136**, 3597–3606 (2009).
  63. Hollnagel, A., Grund, C., Franke, W. W. & Arnold, H. H. The cell adhesion molecule M-cadherin is not essential for muscle development and regeneration. *Mol. Cell Biol.* **22**, 4760–4770 (2002).
  64. Goel, A. J., Rieder, M. K., Arnold, H. H., Radice, G. L. & Krauss, R. S. Niche cadherins control the quiescence-to-activation transition in muscle stem cells. *Cell Rep.* **21**, 2236–2250 (2017).
  65. Li, W. et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
  66. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
  67. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
  68. Becht, E. et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–44 (2018).
  69. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
  70. Ewels, P., Magnusson, M., Lundin, S. & Kaller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
  71. Chen, S. F., Zhou, Y. Q., Chen, Y. R. & Gu, J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, 884–890 (2018).
  72. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
  73. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

74. Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).
75. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
76. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
77. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
78. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).

## Acknowledgements

We extend our heartfelt gratitude to trainees Stephanie Campanano, Moira Alejandra La Fuente, Lindsey Erin Hossfeld, Emilia Keys, Hannah Jean Namgoong, Xinran Zhu, Kaitlyn Burt in Bi laboratory for technical help. We thank the Myoline platform of the Myology Institute for immortalized human myoblast lines. M. Kandasamy from Biomedical Microscopy Core at UGA for imaging assistance. Funding: This work was supported by the NIH R35 award GM147209 and R21 award ARO80330 to P.B., NIGMS grant R01GM114666 to D.S.K.

## Author contributions

H.Z., and P.B. designed research. H.Z., R.S., Z.Z., M.Z., A.B., Y.C., Y.Z., Y.W., A.D., C.H., V.M., and P.B. performed research. H.Z., R.S., Z.Z., M.Z., A.B., Y.W., C.H., V.M., and P.B. analyzed data. D.S.K. and E.K. for resources. P.B. wrote the paper.

## Competing interests

P.B. is a holder of the Canada patent CA3052705A1. E.K. and D.S.K. are holders of the US patent US12037370B2. All other authors declare that they have no financial or other competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-67583-x>.

**Correspondence** and requests for materials should be addressed to Pengpeng Bi.

**Peer review information** *Nature Communications* thanks anonymous reviewer(s) 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) 2025