RESOURCE

High-resolution myogenic lineage mapping by single-cell mass cytometry

Ermelinda Porpiglia[1,2,3], Nikolay Samusik[2,4], Andrew Tri Van Ho[1,2,3], Benjamin D. Cosgrove[1,2,3,6], Thach Mai[1,2,3], Kara L. Davis[2,4,6], Astraea Jager[2,4], Garry P. Nolan[2,4], Sean C. Bendall[2,4,6], Wendy J. Fantl[2,5] and Helen M. Blau[1,2,3,7]

Muscle regeneration is a dynamic process during which cell state and identity change over time. A major roadblock has been a lack of tools to resolve a myogenic progression *in vivo*. Here we capitalize on a transformative technology, single-cell mass cytometry (CyTOF), to identify *in vivo* skeletal muscle stem cell and previously unrecognized progenitor populations that precede differentiation. We discovered two cell surface markers, CD9 and CD104, whose combined expression enabled *in vivo* identification and prospective isolation of stem and progenitor cells. Data analysis using the X-shift algorithm paired with single-cell force-directed layout visualization defined a molecular signature of the activated stem cell state (CD44+/CD98+/MyoD+) and delineated a myogenic trajectory during recovery from acute muscle injury. Our studies uncover the dynamics of skeletal muscle regeneration *in vivo* and pave the way for the elucidation of the regulatory networks that underlie cell-state transitions in muscle diseases and ageing.

Adult muscle stem cells (MuSCs) drive skeletal muscle repair and regeneration[1]. Normally quiescent during homeostasis, MuSCs become activated following muscle injury to replenish the stem cell pool, and simultaneously give rise to progeny that will differentiate and repair the damage[2]. While the role and phenotype of stem cells in muscle regeneration has been extensively studied, little is known about the myogenic progenitor stage, due to the lack of tools to capture this transient cell population *in vivo*.

In haematopoiesis, the identification and isolation of haematopoietic stem and progenitor cells by standard cell sorting methods[3] and single-cell mass cytometry[4] have enabled the discovery of the key pathways that go awry at specific stages of development[5,6], profoundly impacting the treatment of a multitude of blood diseases[3,7]. Similarly, the elucidation of myogenic progenitors *in vivo* has the potential to define the key molecular events that govern cell-state transitions during the course of regeneration, and drive the development of therapeutic strategies for muscle diseases.

To address the cellular and molecular complexity of the myogenic compartment, a major challenge in the muscle field, we applied a high-dimensional single-cell platform called mass cytometry, also known as cytometry by time of flight (CyTOF). CyTOF enables the simultaneous measurements of up to 50 parameters per single cell using antibodies conjugated to metal isotopes[4,8]. The multidimensional attribute of CyTOF allowed us to identify previously unrecognized progenitor cell populations *in vivo*, on the basis of surface marker combinations, and to determine their role during muscle regeneration. This study defines an *in vivo* developmental progression from stem to progenitor cells in skeletal muscle, providing the foundation for future studies of cellular signalling dysfunction within these myogenic populations in the context of ageing, dystrophy and other disease states. Moreover, it paves the way for future investigations of such cell populations in other systems.

RESULTS

Identification of a myogenic progression *in vivo* by single-cell mass cytometry

To discover surface markers that could uniquely distinguish between myogenic stem and progenitor cells *in vivo* in skeletal muscle, we performed a high-throughput fluorescence-based flow cytometry screen with 176 antibodies to integral membrane proteins in both MuSCs,

isolated from Pax7-ZsGreen reporter mice[9], and myoblasts, a primary culture system used to study the late stages of myogenic differentiation and fusion. Flow cytometry data analysis identified several surface markers (Fig. 1a), for which antibodies were then included in our CyTOF panel on the basis of two criteria: presence of the markers on either MuSCs or myoblasts; differential expression levels on MuSCs versus myoblasts. In addition, the screen confirmed the expression on Pax7-ZsGreen MuSCs of known markers previously used to isolate MuSCs, such as α7 integrin and CD34 (ref. 10), β1 integrin/CD29 and CXCR4/CD184 (ref. 11), and VCAM/CD106 (ref. 12) (Fig. 1a).

To identify myogenic progenitors *in vivo* within skeletal muscle, we capitalized on the CyTOF technology and assembled a panel of 23 isotope-conjugated antibodies. These reagents were used to simultaneously measure, at the single-cell level, the expression of previously unrecognized surface markers revealed in our screen, known surface proteins used to isolate MuSCs by us and others, and myogenic transcription factors known to define distinct stages of myogenesis[1] (Supplementary Table 1 and Supplementary Fig. 1a–d). Single-cell suspensions of hindlimb muscles, gastrocnemius (GA) and tibialis anterior (TA) were prepared from 8-week-old mice as described previously[10] and processed for CyTOF analysis (Fig. 1b). To distinguish stem and progenitor cell populations, we analysed the CyTOF data set with a recently developed K-nearest-neighbour density-based clustering algorithm called X-shift[13], which performed unsupervised clustering analysis of cells within the myogenic compartment, defined as Lineage[−]/α7integrin[+] cells and subsequently refined to Lineage[−]/α7integrin[+]/CD9[+] cells. With this multivariate algorithm six clusters were generated, based on the expression of known and previously unrecognized surface markers, and key myogenic transcription factors. To visualize the spatial relationships between the cell types within these X-shift clusters, 2,000 randomly sampled cells from each cluster were subjected to a force-directed layout[13–15]. The resultant map revealed that cells fell densely into three prominent clusters and were linked by sparsely populated paths (Fig. 1c). Cells within the densely populated regions were characterized by differential expression of the myogenic transcription factors Pax7, Myf5, MyoD and Myogenin (Fig. 1d), which are known to mark muscle stem cells, activated stem cells, committed and differentiated muscle cells, respectively[1]. The unique expression pattern of the myogenic transcription factors allowed us to infer developmental stage for these clusters, which we named stem cell population SC, based on high Pax7 expression, and progenitor cell populations P1 and P2, based on Myf5, MyoD and Myogenin expression (Fig. 1d). The co-expression patterns of two surface markers identified in the screen, CD9, a tetraspanin[16,17], and CD104, an integrin[18–22], distinctively defined these populations. SC and P1 were distinguished by differential expression levels of CD9 (intermediate in SC and high in P1), whereas P2 expressed high levels of both CD9 and CD104 (Fig. 1e). These *in vivo* data were supported by the results of the flow cytometry screen in which CD9 was expressed at higher levels in myoblasts than in MuSCs (Supplementary Fig. 1e,f) and CD104 was expressed exclusively by myoblasts (Fig. 1a and Supplementary Fig. 1e). Importantly, the combination of these two surface markers identified myogenic progenitors *in vivo*, a critical missing link between stem cells and differentiated muscle cells.

Data analysis by principal component analysis (PCA), a dimensionality reduction algorithm, revealed that the progenitor populations (P1 and P2) differ from the stem cell population (SC) on the basis of their protein co-expression patterns (Supplementary Fig. 1g). This corroborated the X-shift clustering analysis and led us to hypothesize that staining with antibodies against CD9 and CD104 could serve as a strategy to prospectively isolate stem (Pax7[high]/Myf5[mid]/MyoD[low]/Myogenin[low]) and progenitor (Pax7[low]/Myf5[high]/MyoD[low/high]/Myogenin[high]) cells by fluorescence-activated cell sorting (FACS) (Fig. 1f). Importantly, the three X-shift clusters could be identified by manual gating, the gold standard of flow cytometry analysis, solely on the basis of the expression of CD9 and CD104 (Fig. 2a and Supplementary Fig. 1d). Manual gating also revealed a rare population, P3 (Fig. 2a), whose frequency was too low for functional assays. Therefore, these two markers now provide a sorting strategy to prospectively isolate stem and progenitor cell populations that represent landmarks of myogenesis, as confirmed by the differential expression of the myogenic transcription factors Pax7, Myf5, MyoD and Myogenin, shown in the third dimension (Fig. 2b).

Quantification of each population, identified in the biaxial dot plot of CD9 by CD104, showed that SC comprised the highest fraction of Pax7[+] cells (Fig. 2c), with the highest expression levels (Fig. 2d). In contrast, P1 and P2 comprised the highest fraction of Myf5[+] and Myogenin[+] cells (Fig. 2c), with the highest expression levels (Fig. 2d). P2 also comprised a small fraction of MyoD[+] cells (Fig. 2b–d). These data are consistent with the identification of SC as the MuSC population and P1 and P2 as previously unrecognized progenitor populations (Fig. 1c–e). Further, they suggest a model in which as cells transition from the stem cell state (SC) to the differentiated myogenic state, they sequentially upregulate the expression of CD9 and CD104, giving rise to P1 and then P2 (Fig. 1f). In support of this model, Pax7 knockout (Pax7[−/−]) mice demonstrated a shift in the distribution of cell populations SC, P1, P2 and P3 (Fig. 2f). Compared with wild-type controls, the fraction of cells in SC in Pax7[−/−] mice was markedly diminished at the neonatal stage and progressively declined by 3 weeks of age (Fig. 2f,g). At birth, the decrease in Pax7[−/−] cells within the SC population was paralleled by an increase in the fraction of cells in P1 (Fig. 2g), and was accompanied by alteration of a key signalling pathway involved in the regulation of myogenic commitment[23–26], as shown by increased phosphorylation of MAPKAPK2 (Supplementary Fig. 2a), a direct target of p38α/β MAPK (ref. 27). In one rare Pax7[−/−] mouse that survived to three weeks of age, a majority of α7integrin[+]/CD9[+] muscle cells were in P2 and P3. The cells expressed high levels of MyoD (Fig. 2f) and Myogenin (Supplementary Fig. 2b), and exhibited increased phosphorylation of MAPKAPK2 (Supplementary Fig. 2a), indicative of premature differentiation. Collectively, our data suggest that in Pax7[−/−] muscle, stem cells are progressively lost, the stem cell state is not maintained, and committed cells are the prevalent population by three weeks of age, when the role of Pax7 is required and mutant mice start to die[28,29].

The identified progenitor populations originate from muscle stem cells and exhibit myogenic potential *in vitro* and *in vivo*

To determine whether the progenitor populations were derived from MuSCs, we cultured α7integrin[+]/CD34[+] MuSCs isolated from murine hindlimbs, and measured changes in the expression of CD9 and CD104 during commitment to differentiation. Sorted MuSCs were expanded on hydrogels for one week and then treated with
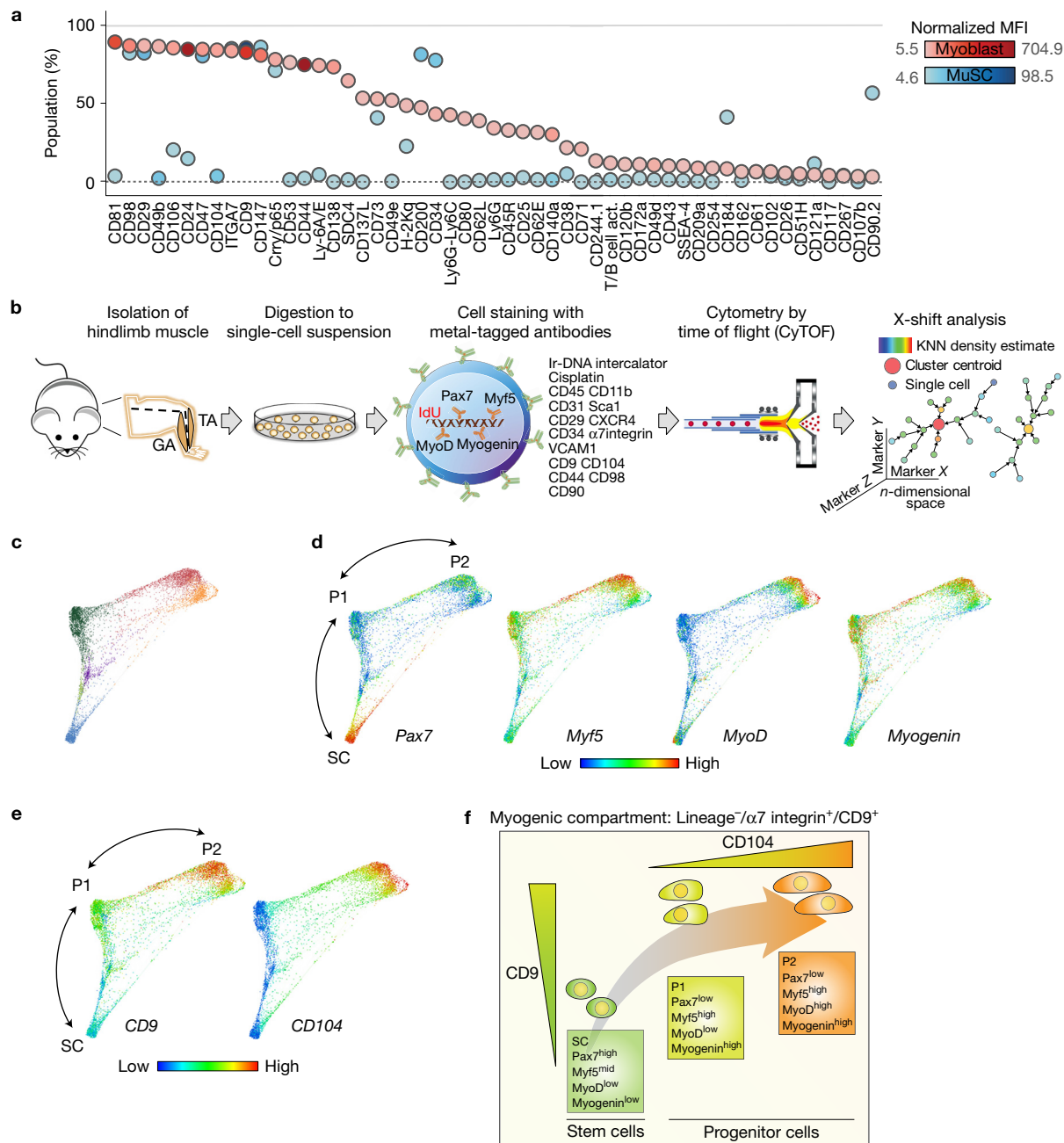
Figure 1 Identification of distinct cell surface markers that delineate a myogenic progression *in vivo*. (**a**) Cell surface marker screening panel analysis of muscle stem cells (MuSCs) and myoblasts. A single-cell suspension of hindlimb muscle isolated from Pax7-ZsGreen reporter mice, and cultured myoblasts were each stained with 176 cell surface antibodies and analysed by fluorescence-based flow cytometry. MuSCs were identified as ZsGreen+ cells. The plot shows the fraction of cells expressing each cell surface marker (*y* axis) and the level of protein expression, indicated as median fluorescence intensity (MFI) and depicted as intensity of blue (MuSCs) and red (myoblasts). T/B cell act., T/B cell activation antigen. (**b**) CyTOF mass cytometry workflow. Tibialis anterior (TA) and gastrocnemius (GA) muscles were triturated, digested to a single-cell suspension, stained with isotope-chelated antibodies and run through the CyTOF instrument. Stained cells were passed through an inductively coupled plasma, atomized, ionized, and the elemental composition was mass measured. Signals corresponding to each elemental tag were correlated to the presence of the respective isotopic marker. Data were analysed using standard flow cytometry software and the clustering algorithm X-shift. (**c**) Live/Lineage−/α7integrin+/CD9+ cells gated from murine hindlimb muscles (TA and GA) were analysed with the X-shift algorithm yielding six clusters (colour-coded in blue, purple, light green, dark green, red and orange). These clusters were visualized using single-cell force-directed layout. Up to 2,000 cells were randomly selected from each X-shift cluster, each cell was connected to 30 nearest neighbours in the phenotypic space and the graph layout was generated using the ForceAtlas2 algorithm[13–15] (representative experiment, $n=3$ mice; 4 independent experiments). (**d**) Expression level of the myogenic transcription factors Pax7, Myf5, MyoD and Myogenin was visualized in cells from the X-shift clusters shown in **c**. Developmental time was inferred and three distinct populations were identified as SC, P1 and P2 (representative experiment, $n=3$ mice). (**e**) Expression level of CD9 and CD104 was visualized in cells from the X-shift clusters shown in **c** (representative experiment, $n=3$ mice). (**f**) Model summarizing the expression pattern of the identified surface markers, CD9 and CD104, during the myogenic progression from the stem cell (SC) to the progenitor (P1, P2) state.
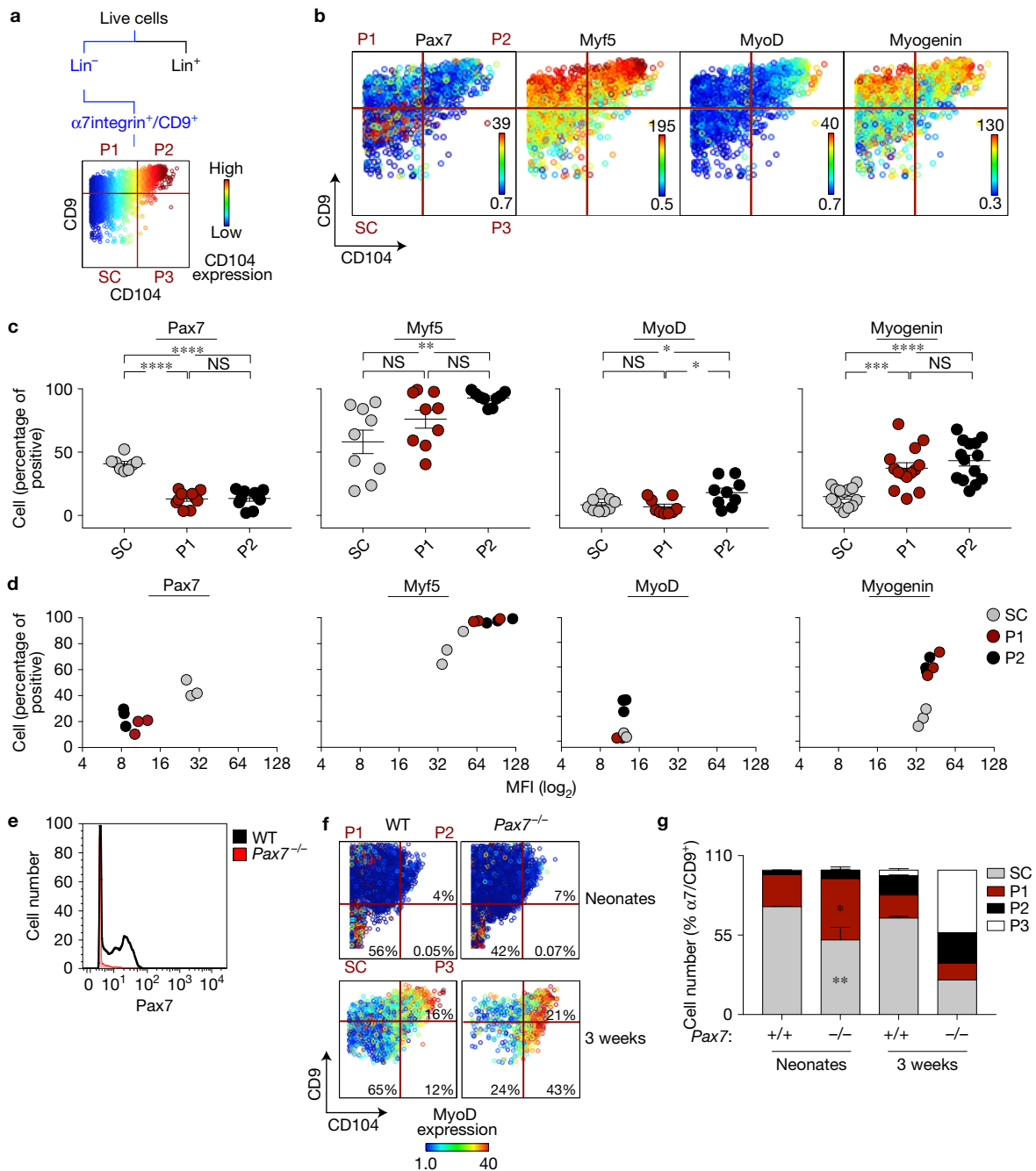
Figure 2 Unique strategy for prospective isolation of stem and progenitor cells *in vivo* in skeletal muscle. (**a**) Scheme of gating strategy for CyTOF data. Live cells are identified on the basis of lack of cisplatin binding. Live/Lineage⁻/α7integrin⁺/CD9⁺ cells are selected and a biaxial dot plot of CD9 (*y* axis) by CD104 (*x* axis), coloured by channel (CD104 expression), is shown. (**b**) Representative biaxial dot plot of CD9 by CD104 coloured by channel, showing expression of Pax7, Myf5, MyoD and Myogenin in the stem and progenitor cells populations. (**c**) Data summary shows the fraction of cells (*y* axis) within populations SC, P1 and P2, expressing Pax7, Myf5, MyoD and Myogenin ($n = 10$ mice (Pax7), $n = 9$ (Myf5 and MyoD), 3 independent experiments; $n = 14$ (Myogenin), 4 independent experiments). Line represents mean ± s.e.m. ANOVA test was performed with significance determined by Bonferroni's multiple comparisons test. (**d**) Data summary of myogenic transcription factor expression levels within populations SC, P1 and P2. Each graph shows the relationship between the percentage of positive cells (*y* axis) and the MFI within the positive population (*x* axis,

log₂) for the expression of Pax7, Myf5, MyoD and Myogenin (representative experiment, $n = 3$ mice). (**e**) Histogram overlay of Pax7 expression in muscle isolated from Pax7 knockout (*Pax7⁻/⁻*) and wild-type (WT) mice and stained with an isotope-chelated antibody against Pax7. (**f**) Representative biaxial dot plots of CD9 by CD104 coloured by channel, showing MyoD expression within populations SC, P1 and P2, in *Pax7⁻/⁻* and WT muscle isolated from neonates and 3-week-old mice. (**g**) Stacked columns indicate the relative proportion of each population within the Live/Lineage⁻/α7integrin⁺/CD9⁺ myogenic compartment in *Pax7⁻/⁻* and WT muscle isolated from neonates (mean ± s.e.m. from $n = 3$ mice, 2 independent experiments) and 3 week-old-mice ($n = 1$ *Pax7⁻/⁻*; mean ± s.e.m. from $n = 10$ WT, 2 independent experiments). Multiple *t*-tests analysis with Bonferroni correction was used to determine difference between *Pax7⁻/⁻* and WT neonates within populations SC, P1, P2. *, **, *** and **** represent statistical significance at $P < 0.05$, $P < 0.01$, $P < 0.001$ and $P < 0.0001$ respectively. NS, statistically non-significant.

differentiation media for 2, 4 and 7 days respectively (Fig. 3a). On day 2 of the assay, the majority of cells comprised stem cells (SC). Over the course of one week the fraction of cells bearing the P1 phenotype progressively increased from 1% to 27%, whereas cells bearing the P2 phenotype increased from 0.4% to 12% (Fig. 3a). These results are consistent with a muscle stem cell of origin giving rise to sequential populations of myogenic progenitor cells that could now be distinguished by differential expression of CD9 and CD104.

To determine the functional capabilities of the myogenic progenitors *in vitro* and *in vivo*, two independent assays were designed. In the first assay, individual stem and progenitor cell populations were sorted, placed into culture for a week on hydrogel and then differentiated to induce fusion. Cells from both the stem cell population (SC) and progenitor cell populations (P1 and P2) retained proliferative capacity and successfully differentiated and fused *in vitro* (Supplementary Fig. 2d). In the second assay, the regenerative potentials of the stem cell population (SC) and the progenitor cell populations (P1, P2) were each assessed *in vivo* using standard assays[10,30]. Individual cell populations were isolated from GFP/Luciferase mice and transplanted into the irradiated TA muscles of NOD/SCID mice (Fig. 3b). At 5 weeks post-transplant, the contribution of the donor cells to regenerated damaged tissue was determined by bioluminescence imaging (BLI) (Fig. 3b,c). Based on engraftment frequencies and BLI signal intensity of the engrafted transplants, the regenerative potential of the populations was highest for SC and progressively lower for P1 and P2 (Fig. 3b,c), consistent with the functional attributes of stem and two progressively more differentiated progenitor cell populations.

To visualize the MuSC (SC) to progenitor (P1, P2) transition *in vivo* we performed lineage tracing in response to acute injury, induced by notexin injection. In this scenario, Pax7CRE[ERT2]ROSA-LSL-tomato mice[31] were injected with tamoxifen for 5 consecutive days, and injected with notexin in the TA and GA muscles 10 days later. Tissue was harvested and cells were analysed by flow cytometry at day 6 post-injury (Fig. 3d). Cell populations SC, P1 and P2 were identified on the basis of differential expression of CD9 and CD104 and analysed for the expression of tomato, indicative of prior Pax7 expression. In the pre-injury state, most tomato[+] cells were in SC (Fig. 3e,f and Supplementary Fig. 2e). However, in response to injury, while the fraction of tomato[+] cells in SC remained elevated, there was a marked increase in the fraction of tomato[+] cells in P1 and P2 (Fig. 3e,f and Supplementary Fig. 2e). Together, these *in vitro* and *in vivo* functional studies demonstrate that the identified progenitor populations originate from MuSCs, have myogenic potential, and exhibit distinct regenerative capacity *in vivo*.

Trajectory of activated stem cells to progenitor cells revealed by CyTOF analysis after acute muscle injury

To capture the progression of myogenesis *in vivo* and define the cellular and molecular dynamics of regeneration, we employed CyTOF analysis of skeletal muscle during an injury time course by notexin injection. To mark cells in S-phase and follow changes in proliferation, IdU was injected 8 h before tissue processing[32] (Fig. 4a). Muscle tissue was digested to a single-cell suspension and stained with the CyTOF antibody panel described above (Fig. 1b and Supplementary Table 1). Data analysis revealed that at day 3 post-injury the cells in population SC incorporated IdU, with a fraction exhibiting a

concomitant increase in the expression of MyoD (Fig. 4b). Consistent with the increased IdU incorporation, a marker of S-phase, there was a striking increase in the proportion of SC cells at day 3 post-injury. This was accompanied by a reduction in the proportion of P1 and P2 cells (Fig. 4c,d). The sharp decline in the fraction of SC cells that incorporated IdU, from ∼52% ± 4% at day 3 post-injury to ∼9% ± 1% by day 6, and the minimal IdU incorporation into cells within P1 and P2 suggested that early after muscle injury, cells within SC are the most proliferative (Fig. 4b,e).

The CyTOF technology enabled high-resolution molecular analysis of muscle stem and progenitor cell populations during the functional response to injury. A heatmap showing protein expression levels for each population revealed that at the steady state, cell surface markers CD44, a glycoprotein implicated in myoblast migration and differentiation[33], and CD98, an amino acid transporter upregulated in human skeletal muscle in response to increased availability of essential amino acids[34], together with myogenic transcription factors Myf5, MyoD and Myogenin, were expressed at higher levels in P1 and P2 compared with SC (Fig. 5a). Following injury, the most striking increases for IdU incorporation were seen on day 3 within population SC. Additionally, CD44 and CD98 were transiently co-expressed on day 3, concomitant with the transcription factor MyoD (Fig. 5b), in about 50% of the SC population (Fig. 5c,d). All of these proteins returned to pre-injury levels by day 6 (Fig. 5b), indicating that they identify a molecular signature of the activated stem cell state. PCA analysis (Fig. 5e), corroborated by a supervised clustering analysis enforcing a two-cluster solution (Supplementary Fig. 3a), showed that at day 3 individual populations cluster away from day 0 and day 6, confirming that significant cellular and molecular alterations occur by day 3 and subside to a level approximating the uninjured state by day 6.

To gain increased resolution into the dynamic changes that occur during the 6 day course of recovery from injury, we applied the unsupervised X-shift analysis approach described above (Fig. 1c–e). High-dimensional X-shift analysis of the mass cytometry data set over time allowed us to construct a molecular map of regeneration as a continuum. A force-directed layout visualization combining cells from X-shift clusters at all time points revealed three dense regions that could be distinguished by the expression pattern of the surface markers CD9 and CD104 (Fig. 6a). Analysis of the regeneration map over time revealed that the occupancy of cells in the dense regions differed at each time point (Fig. 6b), which is consistent with the findings shown in Fig. 4c,d. Moreover, the single-cell spatial resolution enabled by X-shift analysis paired with single-cell force-directed layout data visualization uncovered dynamic changes in the expression of Pax7, MyoD and Myogenin as well as IdU incorporation, revealing distinct patterns and marking a differentiation trajectory (Fig. 6c,d). At the pre-injury state (day 0) there was no IdU incorporation into any of the stem or progenitor cell populations, consistent with the literature[35,36]. During the early phase of recovery from injury (day 3), cells in population SC transited to a different location on the map, characterized by high IdU incorporation (Fig. 6c). MyoD upregulation clearly marked a portion of these cells and its elevated expression persisted during the late phase of regeneration (day 6), when MyoD expression partially coincided with Myogenin upregulation (Fig. 6d). Myogenin expression, during the late phase of recovery from injury extended in a unique pattern along the path that progresses from SC to P2, delineating a trajectory
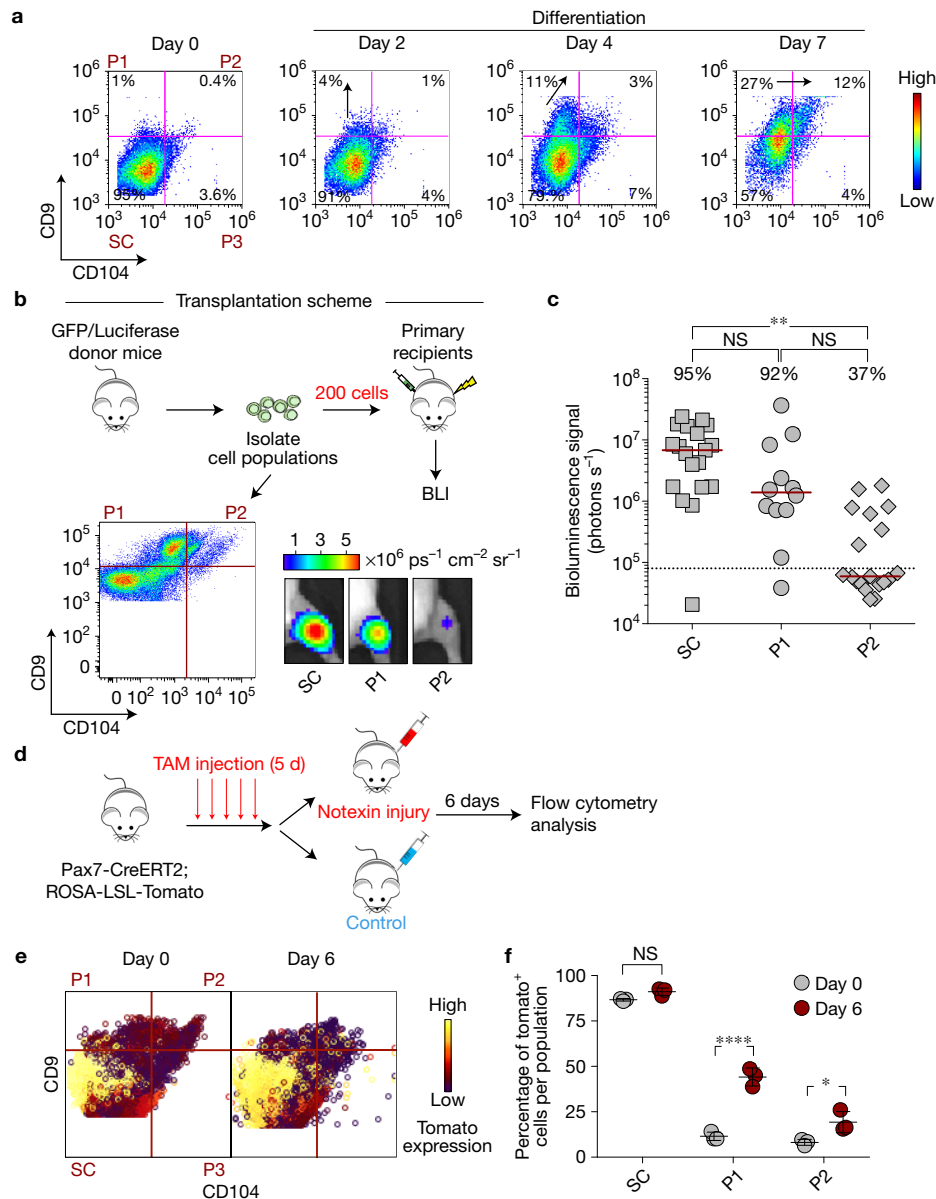
Figure 3 Progenitor cell populations originate from muscle stem cells and exhibit distinct regenerative capacity *in vivo*. (**a**) Flow cytometry analysis of sorted α7integrin$^+$/CD34$^+$ cells, cultured in growth media for 1 week on biomimetic hydrogels and then in differentiation media on collagen-coated plates for 2, 4 and 7 days respectively. Representative biaxial plots of CD9 by CD104 show the fraction of populations SC, P1, P2 and P3 at the indicated time points ($n=5$, 2 independent experiments). The arrows indicate a myogenic progression from muscle stem cells (SC) to progenitor cells (P1, P2) *in vitro* that parallels the one seen *in vivo* (Fig. 1f). (**b**) Scheme depicting the *in vivo* assay of regenerative capacity. Hindlimb muscles isolated from GFP/Luciferase mice were digested to a single-cell suspension. Cell populations SC, P1 and P2 were sorted on the basis of expression of CD9 and CD104 (dot plot, lower left panel) and transplanted (200 cells per injection) into the TA muscle of hindlimb-irradiated NOD/SCID mice. Representative BLI images at 5 weeks post-transplant are shown (lower right panel). (**c**) Scatter plot shows the percentage of transplants from each condition that engrafted above threshold (dashed line, 80,000 photons s^{-1}) into recipient tissue and the

BLI signal intensity (*y* axis). Solid line represents median BLI signal ($n=19$ mice (SC and P2), 3 independent experiments; $n=12$ (P1), 2 independent experiments). ANOVA test was performed with significance determined by Bonferroni's multiple comparisons test. (**d**) Scheme depicting lineage-tracing experiment to track progenitor populations *in vivo* following injury. TAM, tamoxifen. (**e**) TA muscle isolated from uninjured and injured mice (day 6 post-injury) was digested to a single-cell suspension and stained using fluorescently conjugated antibodies against lineage markers, α7integrin, CD9 and CD104. Representative biaxial dot plots of CD9 by CD104 coloured by channel show the expression of tdTomato in populations SC, P1 and P2 at day 0 and day 6 post-injury ($n=3$ mice per condition). (**f**) The scatter plot shows the fraction of tomato$^+$ cells in populations SC, P1 and P2 at day 0 and day 6 post-injury (mean ± s.d. from $n=3$ mice per condition). Multiple *t*-tests analysis with Bonferroni correction was used to determine the difference between the two groups (uninjured and injured) within populations SC, P1 and P2. *, ** and **** represent statistical significance at $P<0.05$, $P<0.01$ and $P<0.0001$. NS, statistically non-significant.

of myogenic differentiation *in vivo*, which is highlighted in the merged panel for days 0, 3 and 6 (Fig. 6d, lower right panel, arrows). The MyoD-negative fraction of SC showed a distinctive upregulation of

Pax7, indicating a return to quiescence (Fig. 6c,d, upper right panels, arrows). In summary, unsupervised X-shift analysis of the myogenic compartment during the time course of recovery from acute injury
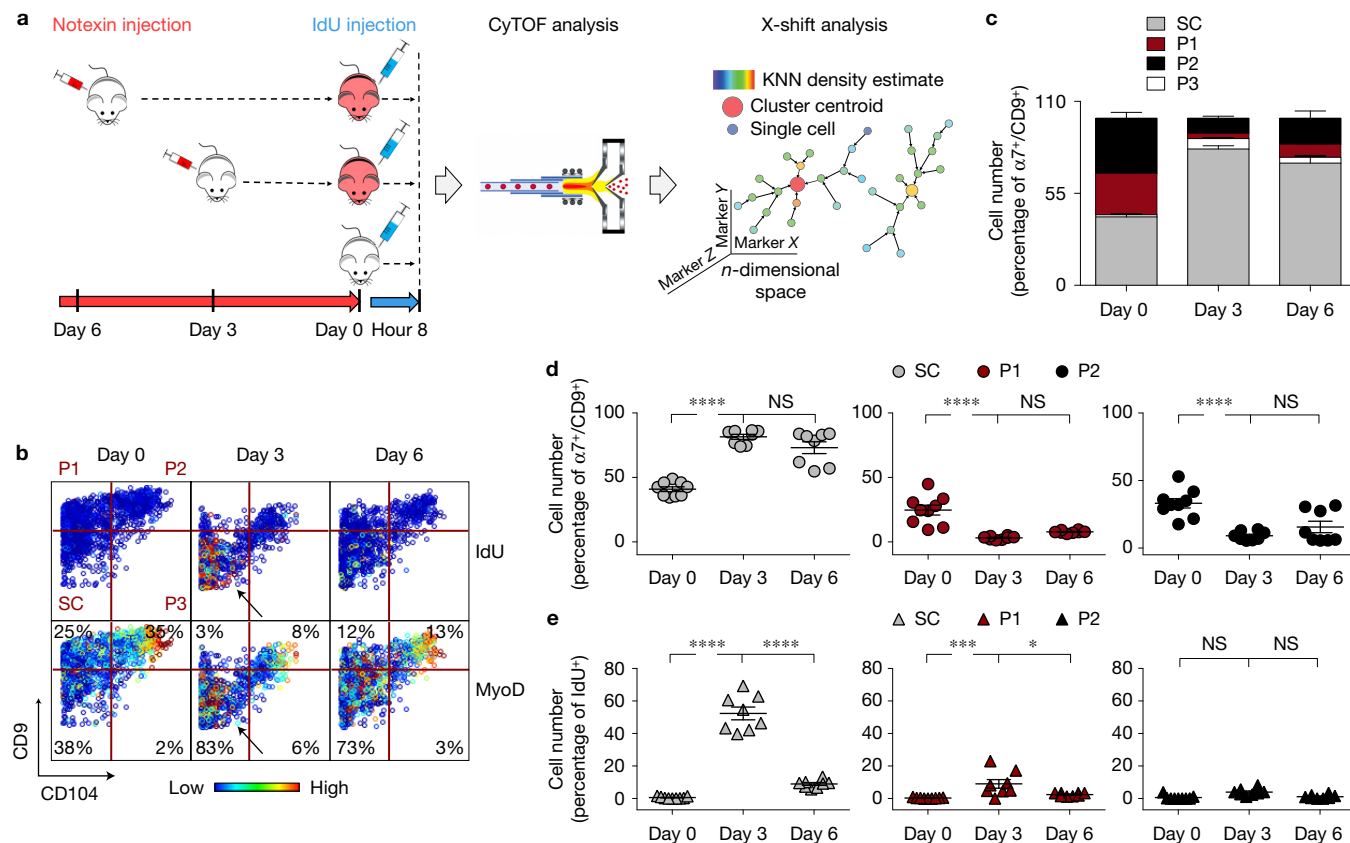
Figure 4 CyTOF analysis reveals the cellular and molecular dynamics within stem and progenitor cell populations during recovery from acute injury. (**a**) Experimental scheme depicting acute injury time course. Mice were acutely injured by notexin injection in the TA and GA muscles, 6 or 3 days prior to tissue collection and injected with IdU 8 h prior to being euthanized. Muscle tissues of 3 indicated groups were simultaneously collected at day 0, stained with isotope-chelated antibodies, run through the CyTOF instrument and analysed using the X-shift clustering algorithm. (**b**) Representative biaxial dot plot of CD9 by CD104 coloured by channel showing IdU incorporation (upper panels) and MyoD expression (lower panels) within individual populations during the injury time course (day 0, day 3, day 6). Arrows indicate increased IdU incorporation and MyoD expression in the SC population at day 3. (**c**) The cells within each population in **b** are quantified as a fraction of the $\alpha 7$ integrin$^+$/CD9$^+$ myogenic compartment during the injury time course. Stacked columns indicate the relative proportion of each population within the Live/Lineage$^-$/$\alpha 7$ integrin$^+$/CD9$^+$ population at each time point (mean $\pm$ s.e.m. from $n = 8$ mice, 3 independent experiments). (**d**) Scatter plots showing cell number (%) for each population in **b** as a function of time (day) post-injury. The line represents mean $\pm$ s.e.m. (**e**) Scatter plot showing the fraction of IdU$^+$ cells within each population in **b** during the time course of muscle injury. The line represents mean $\pm$ s.e.m. Statistical analyses were performed using two-way ANOVA test with significance level determined by Bonferroni's multiple comparisons test. *, *** and **** represent statistical significance at $P < 0.05$, $P < 0.001$ and $P < 0.0001$ respectively. NS, statistically non-significant.

enabled us to visualize the spatial relationships between MuSCs (SC) and the identified progenitor cells (P1, P2), and uncover a dynamic myogenic trajectory that delineated cell-state transitions. Together, these data suggest an *in vivo* myogenic regeneration model in which activated MuSCs (SC) give rise to progeny, some of which express high levels of MyoD, commit and differentiate into progenitors (P1, P2), while others upregulate Pax7 expression and go back to quiescence to replenish the stem cell pool. Importantly, we have identified two surface markers, CD9 and CD104, whose differential expression allowed us to capture these myogenic progenitor populations *in vivo*.

DISCUSSION

Previous work has highlighted the potential of single-cell mass cytometry to resolve the heterogeneity in the haematopoietic compartment, based on cell surface markers[4,5]. Here we pioneered the application of CyTOF to simultaneously analyse cell surface markers

and key myogenic transcription factors. This enabled us to resolve the heterogeneity of the myogenic compartment, yielding previously unknown progenitor cell populations, and delineating a myogenic progression from stem to progenitor cells *in vivo* in skeletal muscle. We showed that these progenitor cell populations originate from Pax7$^+$ stem cells and exhibit myogenic function *in vitro* and *in vivo*, by a combination of lineage tracing and functional assays of myogenesis.

High-dimensional CyTOF analysis of skeletal muscle resolved the intermediate stages of myogenesis at an unprecedented level of detail and revealed the complex relationship between stem and progenitor states during regeneration, suggesting that these states are in constant flux. For example, during recovery from acute injury, SC, the most stem-like population, is activated to give rise to subsets that are MyoD$^+$ and MyoD$^-$, the latter returning to occupy a ground state identified by high Pax7. Indeed, X-shift clustering analysis combined with single-cell force-directed layout visualization of the myogenic
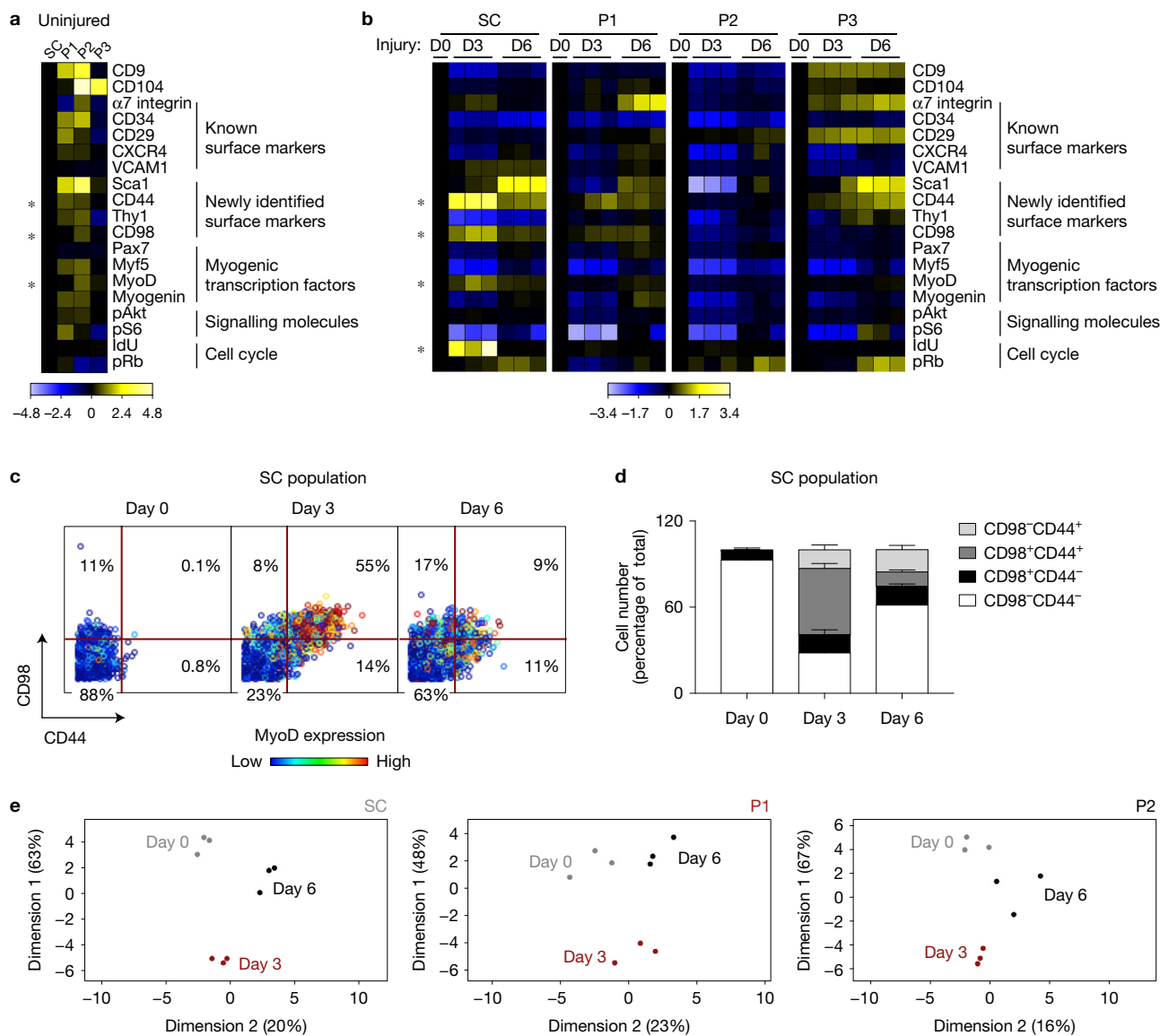
Figure 5 High-dimensional analysis of acute muscle injury identifies a molecular signature of the activated stem cell state. (**a**) Heatmap of protein expression in uninjured populations (transformed ratios compared with population SC, $n=3$ mice per condition). (**b**) Heatmap of protein expression in stem (SC) and progenitor (P1 and P2) cell populations during the time course of injury (transformed ratio compared with day 0, $n=3$ mice per condition). (**c**) Representative biaxial dot plots of CD98 (y axis) by CD44 (x axis) coloured by channel, showing MyoD expression in subsets of the SC population defined by CD98 and CD44 expression, during the injury time course (day 0, day 3, day 6) ($n=5$, 2 independent experiments). (**d**) Cells within each CD98 by CD44 subset are quantified as a fraction of the SC population during the injury time course. Stacked columns indicate the relative proportion of each subset within the SC population at each time point (mean $\pm$ s.e.m. from $n=5$, 2 independent experiments). Statistical analyses were performed using two-way ANOVA test with significance level determined by Bonferroni's multiple comparisons test. The relative increase in the fraction of cells within the CD98+/CD44+ subset from day 0 to day 3 post-injury is highly significant ($P < 0.0001$). (**e**) PCA plot of individual populations in **b** during the time course of injury. Proteins were clustered by their $\log_2$ median intensities ($n=3$ mice per condition).

compartment provided a high-resolution lens that revealed the cellular and molecular dynamics defining the response to acute muscle injury. It uncovered unanticipated heterogeneity within the SC compartment, leading to the identification of a molecular signature of the activated stem cell state, in addition to a differentiation trajectory from MuSCs (SC) to progenitor cells (P1 and P2) that could not have previously been envisioned.

Previous studies suggest that CD9, the myogenic progenitor marker identified here, has a functional role in the activation of the myogenic

program. A member of the tetraspanin family, CD9 plays a key role in signal transduction and has been implicated in various biological processes, such as cell migration, homing, adhesion and fusion[37]. In systems other than muscle, CD9 has been shown to negatively regulate ADAM10 and ADAM17, two metalloproteases involved in activation of Notch signalling[38–40], an essential pathway for MuSC maintenance[41]. Therefore, we speculate that upregulation of CD9 at the transition from MuSC (SC) to progenitor cell (P1, P2) could lead to indirect inhibition of Notch signalling via ADAMs, initiating
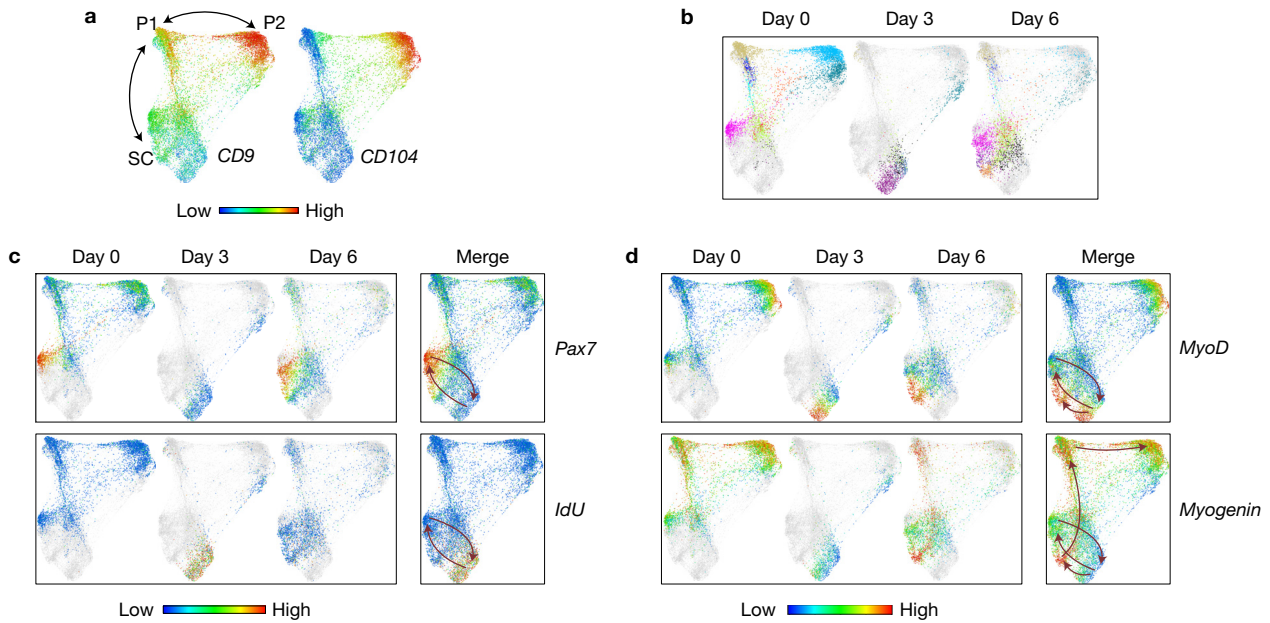
Figure 6 High-dimensional analysis of acute muscle injury uncovers cell-state transitions *in vivo*. (**a**) Live/Lineage⁻/α7integrin⁺/CD9⁺ cells gated from hindlimb muscles isolated during the course of notexin injury (day 0, day 3, day 6) were clustered with the X-shift algorithm and cells from the resultant clusters were visualized using single-cell force-directed layout as described in Fig. 1c. The colour code shows the expression level of CD9 and CD104 (representative analysis, $n = 9$ mice). (**b**) Visualization of cells from the X-shift clusters at each time point (day 0, day 3, day 6) using single-cell force-directed layout. The colour code shows X-shift clusters ($n = 3$ mice per condition). (**c**) Visualization of cells from the X-shift clusters as in **b**. The colour code shows the expression level of Pax7 (upper panels), and IdU incorporation (lower panels). Arrows indicate the trajectory of SC over time. (**d**) Visualization of cells from the X-shift clusters as in **b**. The colour code shows the expression level of MyoD (upper panels) and Myogenin (lower panels). Arrows indicate the trajectory of SC as in **c** and the progression at day 6 from SC to P1 and P2 (highlighted by Myogenin expression).

commitment and differentiation. Another tetraspanin, CD82, has been recently identified on the surface of human muscle stem and progenitor cells[42,43]. Due to its low expression range in murine muscle cells (Supplementary Fig. 3b), CD82 cannot faithfully discriminate the cell populations described here (SC, P1 and P2).

CD104, which marks progenitor population P2, is also known as integrin beta 4 and is a receptor for laminins[18]. It plays a key structural role in the hemidesmosomes of epithelial cells[44]. Like other integrins it mediates cell–matrix, cell–cell adhesion and activates signalling pathways that regulate cell growth. CD104 has been implicated in the epithelial to mesenchymal transition[20], tumour progression[21] and cellular reprogramming to pluripotency[15]. In skeletal muscle, it has been shown to mark a heterogeneous cell population, which CD104 alone could not resolve to functional homogeneous subsets[22].

The approach described here provides a tool for identifying and isolating myogenic progenitors *in vivo* and therefore offers the opportunity to uncover key factors and the downstream signalling pathways that govern cell fate decisions and regulate the balance between cell states, during muscle homeostasis and regeneration. Understanding the fundamental biology and manipulating these cell states *ex vivo* will be instrumental for overcoming barriers to successful transplantation and for optimizing muscle cell therapeutic strategies. Furthermore, high-resolution single-cell studies of the signalling pathways that modulate the transition from MuSC to committed and differentiated muscle cell may provide insight into the mechanisms underlying the development of rhabdomyosarcoma, a highly heterogeneous tumour that affects children[45]. While here we present the simultaneous measurements of transcription factors and

cell surface proteins, the complexity can be extended further. Single-cell mass cytometry has the unique attribute, over other available single-cell approaches, to study post-translational modifications and cell signalling in complex populations as they change over time.

In summary, the work presented here sets the stage for addressing questions that were previously beyond reach, such as uncovering how stem and progenitor cell interaction regulates tissue regeneration[46,47], and establishes a foundation for investigating the cellular and molecular defects that characterize ageing and muscle diseases. Such studies may reveal alterations in population distribution or cell signalling of fundamental interest and potential utility in the identification of novel therapeutic targets. These results provide a paradigm for other studies of solid tissues and their characteristic kinetic changes in the course of development. Finally, our work paves the way for high-resolution analysis of other regenerating tissues, with the potential to discover as yet unknown stem and progenitor cell populations with a role in development and disease. □

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of this paper.

Note: Supplementary Information is available in the online version of the paper

1. Chang, N. C. & Rudnicki, M. A. Satellite cells: the architects of skeletal muscle. *Curr. Top. Dev. Biol.* **107**, 161–181 (2014).
2. Blau, H. M., Cosgrove, B. D. & Ho, A. T. V. The central role of muscle stem cells in regenerative failure with aging. *Nat. Med.* **21**, 854–862 (2015).
3. Weissman, I. L. Translating stem and progenitor cell biology to the clinic: barriers and opportunities. *Science* **287**, 1442–1446 (2000).
4. Bendall, S. C. *et al.* Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* **332**, 687–696 (2011).
5. Bendall, S. C. *et al.* Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* **157**, 714–725 (2014).
6. Levine, J. H. *et al.* Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell* **162**, 184–197 (2015).
7. Druker, B. J. Translation of the Philadelphia chromosome into therapy for CML. *Blood* **112**, 4808–4817 (2008).
8. Spitzer, M. H. & Nolan, G. P. Mass cytometry: single cells, many features. *Cell* **165**, 780–791 (2016).
9. Bosnakovski, D. *et al.* Prospective isolation of skeletal muscle stem cells with a Pax7 reporter. *Stem Cells* **26**, 3194–3204 (2008).
10. Sacco, A., Doyonnas, R., Kraft, P., Vitorovic, S. & Blau, H. M. Self-renewal and expansion of single transplanted muscle stem cells. *Nature* **456**, 502–506 (2008).
11. Cerletti, M. *et al.* Highly efficient, functional engraftment of skeletal muscle stem cells in dystrophic muscles. *Cell* **134**, 37–47 (2008).
12. Liu, L., Cheung, T. H., Charville, G. W. & Rando, T. A. Isolation of skeletal muscle stem cells by fluorescence-activated cell sorting. *Nat. Protoc.* **10**, 1612–1624 (2015).
13. Samusik, N., Good, Z., Spitzer, M. H., Davis, K. L. & Nolan, G. P. Automated mapping of phenotype space with single-cell data. *Nat. Methods* **13**, 493–496 (2016).
14. Jacomy, M., Venturini, T., Heymann, S. & Bastian, M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS ONE* **9**, e98679 (2014).
15. Zunder, E. R., Lujan, E., Goltsev, Y., Wernig, M. & Nolan, G. P. A continuous molecular roadmap to iPSC reprogramming through progression analysis of single-cell mass cytometry. *Cell Stem Cell* **16**, 323–337 (2015).
16. Karlsson, G. *et al.* The tetraspanin CD9 affords high-purity capture of all murine hematopoietic stem cells. *Cell Rep.* **4**, 642–648 (2013).
17. Tachibana, I. & Hemler, M. E. Role of transmembrane 4 superfamily (TM4SF) proteins CD9 and CD81 in muscle cell fusion and myotube maintenance. *J. Cell Biol.* **146**, 893–904 (1999).
18. Clarke, A. S., Lotz, M. M. & Mercurio, A. M. A novel structural variant of the human β4 integrin cDNA. *Cell Adhes. Commun.* **2**, 1–6 (1994).
19. Su, L., Lv, X. & Miao, J. Integrin β4 in neural cells. *Neuromol. Med.* **10**, 316–321 (2008).
20. Masugi, Y. *et al.* Upregulation of integrin β4 promotes epithelial-mesenchymal transition and is a novel prognostic marker in pancreatic ductal adenocarcinoma. *Lab. Invest.* **95**, 308–319 (2015).
21. Guo, W. *et al.* β4 integrin amplifies ErbB2 signaling to promote mammary tumorigenesis. *Cell* **126**, 489–502 (2006).
22. Liadaki, K. *et al.* β4 integrin marks interstitial myogenic progenitor cells in adult murine skeletal muscle. *J. Histochem. Cytochem.* **60**, 31–44 (2012).
23. Jones, N. C. *et al.* The p38α/β MAPK functions as a molecular switch to activate the quiescent satellite cell. *J. Cell Biol.* **169**, 105–116 (2005).
24. Troy, A. *et al.* Coordination of satellite cell activation and self-renewal by par-complex-dependent asymmetric activation of p38 α/β MAPK. *Stem Cell* **11**, 541–553 (2012).
25. Cosgrove, B. D. *et al.* Rejuvenation of the muscle stem cell population restores strength to injured aged muscles. *Nat. Med.* **20**, 255–264 (2014).
26. Segalés, J., Perdiguero, E. & Muñoz-Cánoves, P. Regulation of muscle stem cell functions: a focus on the p38 MAPK signaling pathway. *Front. Cell Dev. Biol.* **4**, 91 (2016).
27. Hausburg, M. A. *et al.* Post-transcriptional regulation of satellite cell quiescence by TTP-mediated mRNA decay. *eLife* **4**, e03390 (2015).
28. Seale, P. *et al.* Pax7 is required for the specification of myogenic satellite cells. *Cell* **102**, 777–786 (2000).
29. 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).
30. Gilbert, P. M. *et al.* Substrate elasticity regulates skeletal muscle stem cell self-renewal in culture. *Science* **329**, 1078–1081 (2010).
31. 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).
32. Behbehani, G. K., Bendall, S. C., Clutter, M. R., Fantl, W. J. & Nolan, G. P. Single-cell mass cytometry adapted to measurements of the cell cycle. *Cytometry* **81A**, 552–566 (2012).
33. Mylona, E., Jones, K. A., Mills, S. T. & Pavlath, G. K. CD44 regulates myoblast migration and differentiation. *J. Cell. Physiol.* **209**, 314–321 (2006).
34. Drummond, M. J. *et al.* An increase in essential amino acid availability upregulates amino acid transporter expression in human skeletal muscle. *Am. J. Physiol. Endocrinol. Metab.* **298**, E1011–8 (2010).
35. Conboy, M. J., Karasov, A. O. & Rando, T. A. High incidence of non-random template strand segregation and asymmetric fate determination in dividing stem cells and their progeny. *PLoS Biol.* **5**, e102 (2007).
36. Rocheteau, P., Gayraud-Morel, B., Siegl-Cachedenier, I., Blasco, M. A. & Tajbakhsh, S. A subpopulation of adult skeletal muscle stem cells retains all template DNA strands after cell division. *Cell* **148**, 112–125 (2012).
37. Leung, K. T. *et al.* The tetraspanin CD9 regulates migration, adhesion, and homing of human cord blood CD34+ hematopoietic stem and progenitor cells. *Blood* **117**, 1840–1850 (2011).
38. Gutiérrez-López, M. D. *et al.* The sheddase activity of ADAM17/TACE is regulated by the tetraspanin CD9. *Cell. Mol. Life Sci.* **68**, 3275–3292 (2011).
39. Arduise, C. *et al.* Tetraspanins regulate ADAM10-mediated cleavage of TNF-α and epidermal growth factor. *J. Immunol.* **181**, 7002–7013 (2008).
40. Mizuno, S. *et al.* A disintegrin and metalloprotease 10 is indispensable for maintenance of the muscle satellite cell pool. *J. Biol. Chem.* **290**, 28456–28464 (2015).
41. Koch, U., Lehal, R. & Radtke, F. Stem cells living with a Notch. *Development* **140**, 689–704 (2013).
42. Uezumi, A. *et al.* Cell-surface protein profiling identifies distinctive markers of progenitor cells in human skeletal muscle. *Stem Cell Rep.* **7**, 263–278 (2016).
43. Alexander, M. S. *et al.* CD82 is a marker for prospective isolation of human muscle satellite cells and is linked to muscular dystrophies. *Cell Stem Cell* **19**, 800–807 (2016).
44. van der Neut, R., Krimpenfort, P., Calafat, J., Niessen, C. M. & Sonnenberg, A. Epithelial detachment due to absence of hemidesmosomes in integrin β4 null mice. *Nat. Genet.* **13**, 366–369 (1996).
45. Saab, R., Spunt, S. L. & Skapek, S. X. Myogenesis and rhabdomyosarcoma the Jekyll and Hyde of skeletal muscle. *Curr. Top. Dev. Biol.* **94**, 197–234 (2011).
46. Hsu, Y.-C., Pasolli, H. A. & Fuchs, E. Dynamics between stem cells, niche, and progeny in the hair follicle. *Cell* **144**, 92–105 (2011).
47. Hsu, Y.-C., Li, L. & Fuchs, E. Transit-amplifying cells orchestrate stem cell activity and tissue regeneration. *Cell* **157**, 935–949 (2014).

567

METHODS

Mouse lines. All animal protocols were approved by the Stanford University Administrative Panel on Laboratory Animal Care (APLAC) (protocol no. 10509 and no. 29398) and experiments were performed in compliance with the institutional guidelines of Stanford University. C57BL6 young adult mice were purchased from Jackson Laboratories and used at 2–4 months of age (median age 2 months). Mice ubiquitously expressing a green fluorescent protein (GFP) transgene and mice ubiquitously expressing a firefly luciferase (Fluc) transgene driven by the ACTB promoter (L2G85 strain) were obtained and genotyped as described previously[10]. Double-transgenic GFP/Luciferase mice were generated by breeding the above strains and validated by appropriate PCR-based strategies to validate the genotype. Cells from C57BL/6 mice were used for cell proliferation and protein expression experiments. Cells from GFP/Luciferase mice were isolated for transplantation experiments from young mice (median age 2 months). Pax7-ZsGreen transgenic mice were a kind gift from M. Kyba (University of Minnesota, Minneapolis, Minnesota, USA). Pax7 knockout mice were a kind gift from M. A. Rudnicki (University of Ottawa, Ontario, Canada). Double-transgenic *Pax7^CreERT2^;Rosa26-LSL-tdTomato* were generated by crossing *Pax7^CreERT2^* mice obtained from Jackson Laboratory (Stock no. 017763) and *Rosa26-LSL-tdTomato* obtained from Jackson Laboratory (Stock no. 007914). We validated these genotypes by appropriate PCR-based strategies. We genotyped Pax7 knockout mice using the following primer combinations: wild type forward 5′-GTGGGG TCTTCATCAACGGTC-3′; mutant forward 5′-TCGTGCTTTACGGTATCG CCG-CTCCCG-3′; reverse 5′-GGGCTTGCTGCCTCCGATAGC-3′. Double-transgenic *Pax7^CreERT2^;Rosa26-LSL-tdTomato* were genotyped using the following Tomato Primer Mix: MMW no. 314, 5′-AAG GGA GCT GCA GTG GAG TA-3′; MMW no. 315, 5′-CCG AAA ATC TGT GGG AAG TC-3′; MMW no. 316, 5′-GGC ATT AAA GCA GCG TAT CC-3′; MMW no. 317, 5′-CTG TTC CTG TAC GGC A TG G-3′. CreERT2 expression was validated using the following primers: common forward 5′-GCT GCT GTT GAT TAC CTG GC-3′; wild type reverse 5′-CTG CAC TGA GAC AGG ACC G-3′; mutant reverse 5′-CAA AAG ACG GCA ATA TGG TG -3′. The number of animals for each data set and all relevant details regarding the sample size are reported with each experiment.

Cell isolation and dissociation of muscle tissue. Tibialis anterior (TA) and gastrocnemius (GA) muscles were dissected and subjected to mechanical dissociation using the gentleMACS dissociator (Miltenyi Biotech), followed by collagenase (0.25%) and dispase (0.04 U ml^{-1}; Roche) digestion at 37 °C for 2 h. The resulting cell suspension was passed through a standard syringe needle and subsequently through a 70-μm nylon filter (BD Biosciences). No established cell lines were used in this study.

High-throughput flow cytometric surface screening. A high-throughput flow cytometric screen of 176 cell surface antibodies was performed, using the Mouse Cell Surface Marker Screening Panel (BD Biosciences). Pax7-ZsGreen reporter mice were used, as Pax7 is the most faithful marker of the entire MuSC population throughout lifespan[9]. Briefly, mouse TA and GA muscles were isolated from Pax7-ZsGreen reporter mice[9], dissected and subjected to collagenase and dispase digestion. Non-muscle tissue was removed under a dissection microscope and muscle fibres were dissociated. After digestion, the remaining cell suspension was passed through a nylon 70-μm filter. Single-cell suspension (500,000 cells per well) was added to the antibody-containing wells. Cells were stained first with the purified monoclonal antibodies, then with the respective biotinylated secondary antibodies and finally with AlexaFluor-647-conjugated streptavidin. Subsequently cells were analysed on a BD-LSRII flow cytometer using FACS Diva Software. During analysis, ZsGreen-positive cells were identified and candidate novel markers that were expressed by the MuSC population (ZsGreen-positive cells) were identified. The same assay was performed on primary myoblasts. For myoblast preparation, adult murine hindlimb muscles were dissociated using a combination of mechanical and enzymatic digestion, as in the muscle stem cell isolation, and myoblasts were derived by differential plating as previously described[48].

Isolation of muscle stem cells and novel progenitor populations. Muscle tissue was isolated and digested to a single-cell suspension as described above. Cells were incubated with biotinylated antibodies reactive to CD45, CD11b, CD31 and Sca1 (BD Biosciences) for 30 min at 4 °C and washed. Cells were then incubated with streptavidin magnetic beads (Miltenyi Biotech), streptavidin-APC-Cy7 (Invitrogen), α7 integrin-PE antibody (AbLab), and CD34-eFluor660 antibody (eBioscience) or CD9-APC antibody and CD104-FITC antibody (eBioscience). After magnetic depletion of the biotin-positive cells, the lineage-negative cells were stained with DAPI and sorted on a FACSAria cell sorter in purity mode, using FACS Diva software (BD Biosciences). To isolate MuSCs, we first gated for viable cells (DAPI negative), then for cells negative for the lineage markers (CD45, CD31, CD11b, and Sca1), finally for cells positive for both CD34 and α7 integrin, which represent the muscle stem cell fraction. To simultaneously isolate stem cells and the novel progenitor

populations using the new sorting strategy, we first gated for viable cells (DAPI negative), then for cells negative for the lineage markers (CD45, CD31, CD11b and Sca1) and positive for α7 integrin and CD9. Finally, we sorted SC–P2 on the basis of their unique expression pattern of CD9 and CD104. Flow cytometry scatter plots were generated using FlowJo v8.7 (Treestar, Ashland).

Transplantation of stem cells and novel progenitor cell populations. Sorted cell populations were transplanted immediately following FACS isolation into the TA muscle of recipient mice as previously described[10,30]. Cells from GFP/Luciferase mice were transplanted into gender-matched, hindlimb-irradiated NOD/SCID mice (Jackson Laboratories). NOD/SCID mice (2–4 months of age, median 2 months) were anaesthetized with ketamine (2.4 mg per mouse) and xylazine (240 μg per mouse) by intraperitoneal injection and irradiated by a single dose of 18 Gy administered to the hindlimbs, with the rest of the body shielded in a lead jig. Transplantation was performed within three days post-irradiation. Freshly isolated cells were counted by haemocytometer and resuspended at desired cell concentrations in PBS with 2.5% goat serum and 1 mM EDTA. Cells were transplanted by intramuscular injection into the tibialis anterior muscle in a 10 μl volume.

In vivo **IdU labelling.** Mice were weighed, anaesthetized with isoflurane and injected intraperitoneally with IdU eight hours prior to euthanizing the animal (150 mg kg^{-1} body weight per injection)[32].

Bioluminescence imaging. Bioluminescence imaging (BLI) was performed using a Xenogen-100 system, as previously described[10], at 5 weeks post-transplant. The system is comprised of a light-tight imaging chamber, a CCD (charge-coupled device) camera with a cryogenic refrigeration unit and the appropriate computer system (Living Image software, Caliper LifeSciences). Briefly, the animals were anaesthetized under isoflurane and injected intraperitoneally with a 100 μl volume of luciferin diluted in PBS (0.1 mmol kg^{-1} body weight, Caliper LifeSciences). Immediately after injection, images were acquired each minute for a total of 15 min and data were stored for subsequent analysis, using Living Image software (Caliper LifeSciences). Bioluminescence images acquired at 12 min post-luciferin injection were used for analysis. Bioluminescence signal was calculated by drawing a consistent region-of-interest (ROI) over each hindlimb and quantifying the resulting signal. A bioluminescence signal value of 80,000 photons s^{-1} represented our detection threshold, as this level was previously determined to represent the presence of one or more GFP-positive muscle fibres in transplanted tissue[10,30].

Muscle injury. Mice (8–10 weeks) were acutely injured by a single 10 μl intramuscular injection of notexin (10 μg ml^{-1}; Latoxan) into the TA muscle and two injections in the GA muscle at the indicated time points.

Lineage tracing. *Pax7^CreERT2^; Rosa26-LSL-Tomato* mice were treated with five consecutive daily intraperitoneal injections at a tamoxifen dose of 75 mg kg^{-1} body weight to activate tdTomato expression following *Pax7*-dependent Cre-mediated recombination. Ten days after the last tamoxifen injection, mice were subjected to intramuscular injection of 10 μl of notexin (Latoxan). Flow cytometry analysis was performed at 6 days post-injury.

Intracellular flow cytometry. Muscle stem cells were isolated from young mice by FACS sorting, fixed in 1.6% paraformaldehyde, permeabilized in methanol and stored at −80 °C. Thawed cells were washed two times with staining buffer (PBS with 0.5% BSA and 2.5 mM EDTA), and blocked in staining buffer for 30 min at room temperature. Cells were stained with the following primary antibodies, mouse monoclonal anti-Pax7 antibody (sc-81648, Santa Cruz Biotechnology), rabbit polyclonal anti-Myf5 antibody (Clone C20, Santa Cruz Biotechnology), mouse monoclonal anti-MyoD antibody (clone 5.8A, BD Biosciences), mouse monoclonal anti-Myogenin antibody (Clone F5D, BD Biosciences), rabbit polyclonal anti-ZsGreen antibody (632474, Clontech) for 1 h at room temperature. Subsequently, cells were washed and incubated in the presence of secondary donkey anti rabbit-AlexaFluor-488 and goat anti-mouse IgG1-AlexaFluor-647 antibodies (Jackson ImmunoResearch Laboratories) for 30 min at room temperature. Finally, cells were washed twice with staining buffer and analysed on a BD-FACSCalibur flow cytometer using CellQuest software (BD Biosciences).

Antibody conjugation with metal isotopes. Purified antibodies were conjugated to the indicated metals for mass cytometry analysis using the MaxPAR antibody conjugation kit (DVS Sciences) according to the manufacturer's instruction. Following labelling, antibodies were diluted in Candor PBS Antibody Stabilization solution (Candor Bioscience GmbH) to 0.2 mg ml^{-1} and stored long term at 4 °C. Each antibody clone and lot was titrated to optimal staining concentrations using murine myoblasts, as well as murine muscle and spleen cell suspensions. Antibodies against the intracellular myogenic transcription factors were first optimized by flow

cytometry performed on murine muscle stem cells sorted from Pax7-ZsGreen mice and murine myoblasts. A list of all mass cytometry antibodies used for analysis can be found in Supplementary Table 1.

Mass cytometry staining. Dead cells were stained using cisplatin (WR International, Cat. no. 89150-634) as previously described[49]. In particular, cells were resuspended in 1 ml serum-free DMEM at 2×10^6 cells ml^{-1} and cisplatin was added at a final concentration of 25 µM for 1 min at room temperature. The reaction was quenched with 3 ml of DMEM/15% FBS. Samples were centrifuged at 456g for 10 min and cell pellets were resuspended in cell staining media (CSM: PBS with 0.5% bovine serum albumin (BSA) and 0.02% sodium azide) and fixed with 1.6% paraformaldehyde (PFA) for 10 min on ice. Cells were washed with CSM and stained with antibodies against surface markers included in the mass cytometry panel for 1 h at room temperature. Cells were washed twice with CSM and permeabilized with methanol for 10 min on ice. Cells were washed twice with CSM and stained with antibodies against intracellular markers included in the mass cytometry panel for 1 h at room temperature. Cells were washed twice with CSM and stained with 1 ml of 1:5,000 191/193Ir DNA intercalator[50] (DVS Sciences) diluted in PBS with 1.6% PFA for 20 min at room temperature.

Barcoding. Barcoding was performed on about 1 million fixed cells, as described previously[51,52]. In particular, fixed cells were washed once in CSM, once in PBS, and finally in the permeabilization solution (PBS with 0.02% saponin). Cells were then resuspended in the residual permeabilization solution and incubated with the barcoding dyes for 15 min at room temperature. After barcoding, cells were washed twice with CSM, pooled in a single tube and stained with isotope-chelated antibodies. Following data collection, individual samples were recovered using the single-cell deconvolution algorithm[52].

Mass cytometry measurement. Cells were acquired on the CyTOF TM mass cytometer (DVS Sciences) at an event rate of approximately 500 cells per second as previously described[4]. The instrument was run in high-resolution mode (Mass resolution ∼700) with internally calibrated dual-count detection. Noise reduction and cell extraction parameters were: cell length 10–65, lower convolution threshold 10. Samples were normalized using beta beads[53].

X-Shift analysis and graphic display of single-cell mass cytometry data. Events were clustered on the basis of a combination of surface markers and myogenic transcription factors (α7 integrin, CD34, CD9, CD104, Pax7, Myf5, Pax3, MyoD, Myogenin) using the X-Shift algorithm and clusters were validated[13]. To visualize the spatial relationships between the cell types within these X-shift clusters, 2,000 randomly sampled cells from each cluster were subjected to a force-directed layout[13,15]. All conditions from a single experiment were processed simultaneously so that the resulting map would capture all populations present in the entire data set.

Principal component analysis. Principal component analysis (PCA) plots were generated using the PCA function in the FactoMineR package in R. Dendrograms were fitted using the hclust function in R. The distance matrix was calculated through the dist function using Euclidean parameters.

Statistics and reproducibility. All animal experiments were randomized. No statistical method was used to predetermine sample size. The investigators were not blinded to allocation during experiments and outcome assessment. Transplantation experiments were performed with at least $n = 12$ mice, two independent experiments. CyTOF analysis of muscle tissue during homeostasis and acute injury was performed with at least $n = 8$ mice, 3 independent experiments, unless otherwise noted. The line indicated in the scatter plots represents the median (Fig. 3c), mean $\pm$ s.e.m. (Figs 2c and 4d,e) or mean $\pm$ s.d. (Fig. 3f). Bar graphs represent mean $\pm$ s.e.m. ANOVA or multiple t-test analysis was performed for experiments with more than 3 groups, as indicated in the legend. Bonferroni *post hoc* analysis was performed to determine significant differences between groups. Statistical significance was determined by $P < 0.05$. *, **, *** and **** represent statistical significance at $P < 0.05$, $P < 0.01$, $P < 0.001$ and $P < 0.0001$ respectively. NS represents statistically non-significant. Source data of all representative experiments are provided in Supplementary Table 2.

Code availability. The computational code used in the study can be obtained at https://github.com/nolanlab/vortex/releases.

Data availability. CyTOF data that support the findings of this study have been deposited in Flowrepository.org under accession code FR-FCM-ZY3E, FR-FCM-ZY3F, FR-FCM-ZY3G, FR-FCM-ZY3H, FR-FCM-ZY3J, FR-FCM-ZY3K, FR-FCM-ZY3W. Source data for Figs 1–6 and Supplementary Figs 1–3 have been provided in Supplementary Table 2. All other data supporting the findings of this study are available from the corresponding author on reasonable request.

48. Rando, T. A. & Blau, H. M. Primary mouse myoblast purification, characterization, and transplantation for cell-mediated gene therapy. *J. Cell Biol.* **125,** 1275–1287 (1994).
49. Fienberg, H. G., Simonds, E. F., Fantl, W. J., Nolan, G. P. & Bodenmiller, B. A platinum-based covalent viability reagent for single-cell mass cytometry. *Cytometry* **81A,** 467–475 (2012).
50. Ornatsky, O. I. *et al.* Study of cell antigens and intracellular DNA by identification of element-containing labels and metallointercalators using inductively coupled plasma mass spectrometry. *Anal. Chem.* **80,** 2539–2547 (2008).
51. Behbehani, G. K. *et al.* Transient partial permeabilization with saponin enables cellular barcoding prior to surface marker staining. *Cytometry A* **85,** 1011–1019 (2014).
52. Zunder, E. R. *et al.* Palladium-based mass tag cell barcoding with a doublet-filtering scheme and single-cell deconvolution algorithm. *Nat. Protoc.* **10,** 316–333 (2015).
53. Finck, R. *et al.* Normalization of mass cytometry data with bead standards. *Cytometry A* **83,** 483–494 (2013).
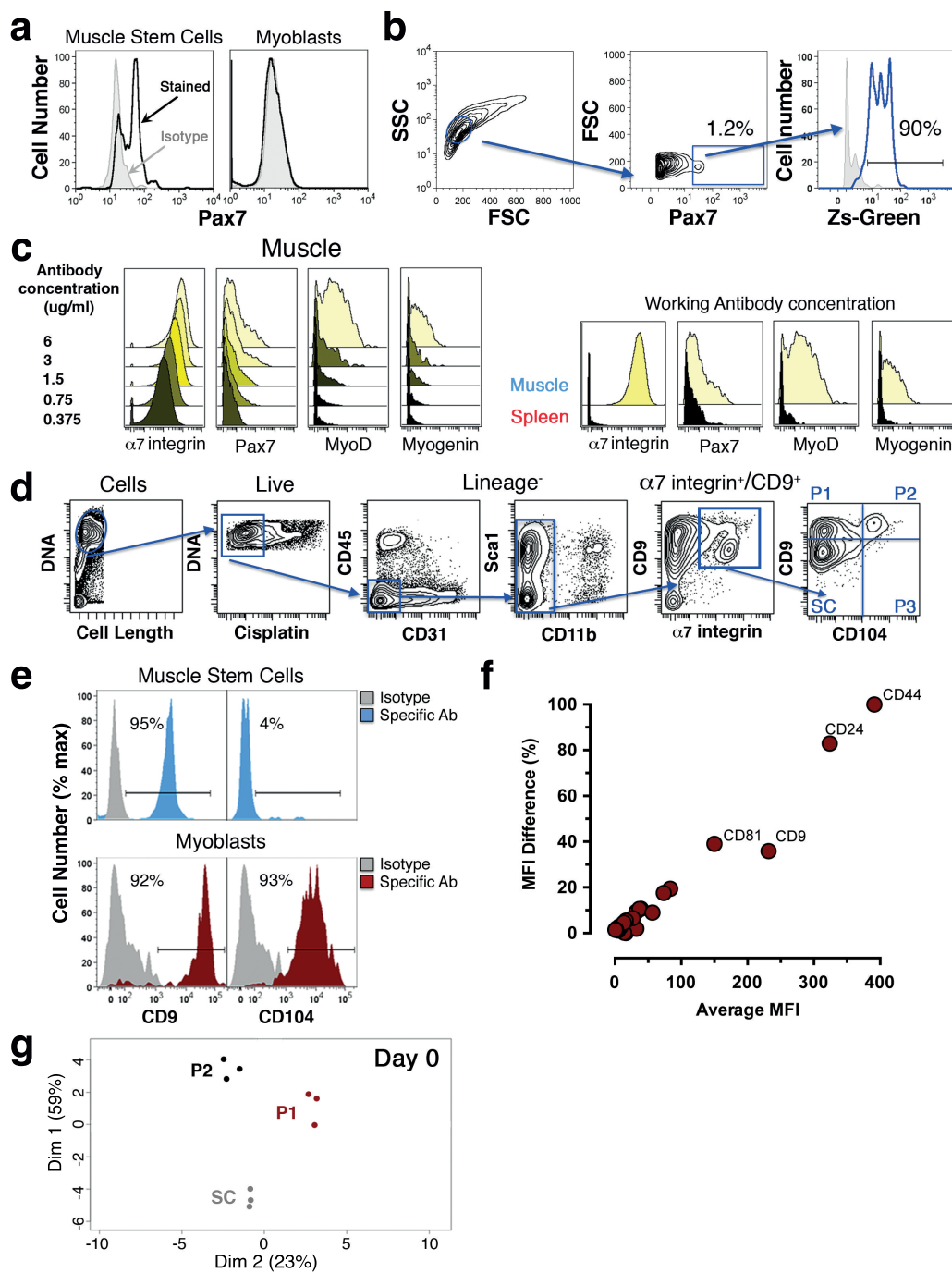
Supplementary Figure 1 (related to Figure 1). CyTOF analysis of skeletal muscle tissue. **(a)** Intracellular staining of Pax7 by flow cytometry in sorted muscle stem cells (left panel) and myoblasts (right panel). **(b)** Muscle cells isolated from Pax-Zs green reporter mice were fixed and permeabilized for intracellular staining. Cells were simultaneously stained with antibodies against Zs-Green and Pax7. Cells that were positive for Pax7 were gated and the fraction of Zs-Green+ cells was quantified to be 90%. **(c)** CyTOF antibody titration. Isotope-chelated anti-mouse antibodies against the surface marker of muscle stem cells, α7 integrin, and intracellular myogenic transcription factors Pax7, MyoD, Myogenin have been titrated using positive (muscle) and negative (spleen) controls to optimize signal, achieve saturation and minimize background. **(d)** Gating strategy on CyTOF samples as described

in Fig. 2a. Individual contour plots are shown. **(e)** Histogram plot of CD9 (left panels) and CD104 (right panels) expression in MuSCs (upper panels) and myoblasts (lower panels) compared to the respective isotype control. **(f)** Screening data were analyzed using the Bland-Altman method to measure significant differences in signal intensity (also known as Median Fluorescence Intensity (MFI)) of individual markers in myoblasts compared to MuSCs. The percentage difference from the average MFI (100 x (Myoblast MFI - MuSC MFI)/ Average MFI) is plotted (y axis) as a function of the average MFI ((Myoblast MFI + MuSC MFI)/2) (x axis). **(g)** PCA plot of Live/ Lineage-/α7 integrin+/CD9+ cells by population in uninjured (day 0) samples. Protein expression levels were clustered by their $\log_2$ median intensities (representative experiment, n=3 mice).

p-MAPKAPK2

low   **high**

Myogenin
Expression

low   **high**

Pax7
Expression

low   **high**
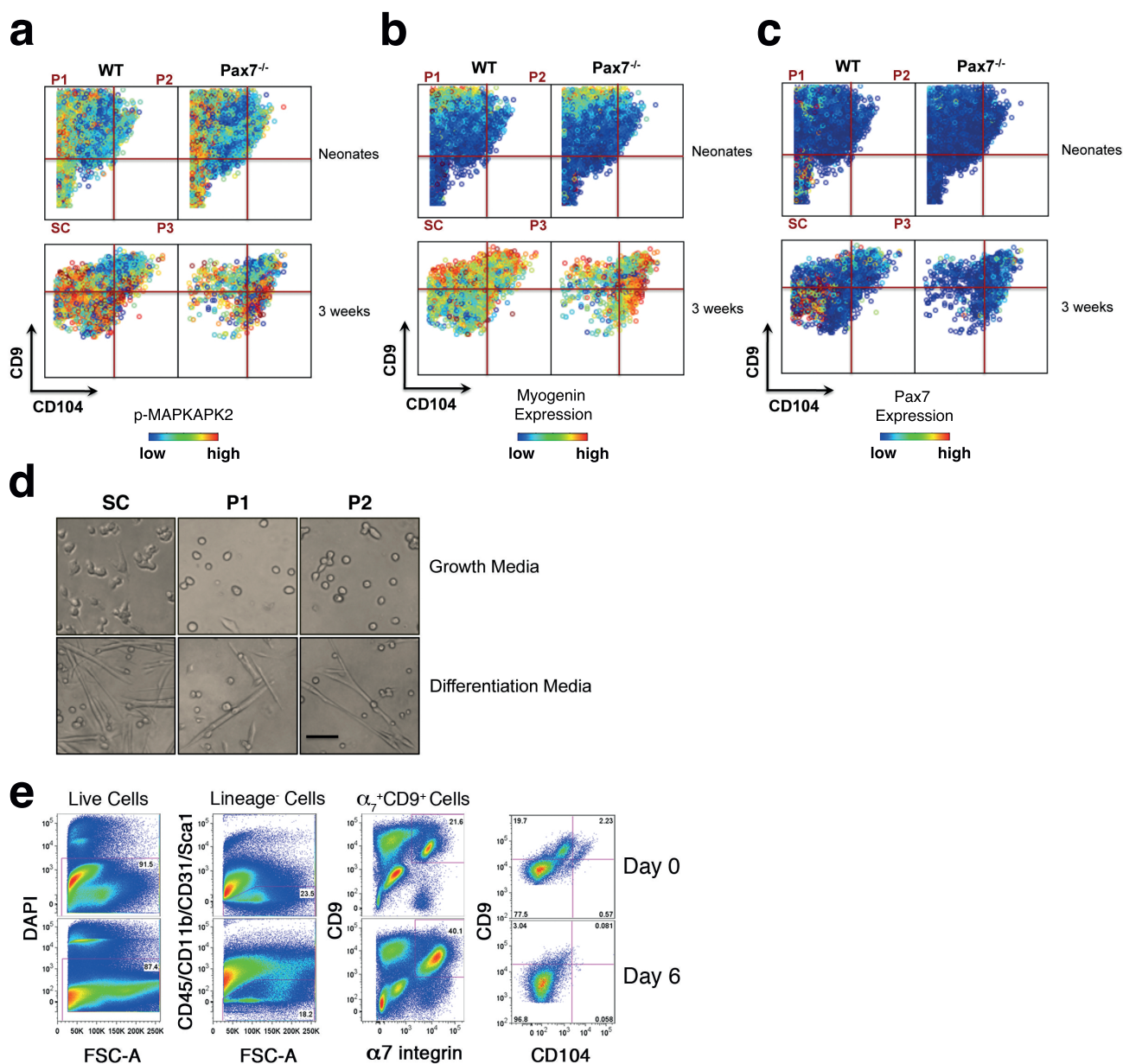
Supplementary Figure 2 (related to Figure 2 and 3). Functional characterization of the newly identified progenitor population in skeletal muscle. **(a)** Representative biaxial dot plots of CD9 (y axis) by CD104 (x axis) colored by channel, showing MAPKAPK2 phosphorylation in populations SC, P1, P2 and P3 in Pax7$^{-/-}$ muscle (right) and WT control (left), isolated from neonates (upper panels) (n=3 mice, 2 independent experiments) and 3 weeks old mice (lower panels) (n=1 Pax7$^{-/-}$; mean ± SEM from n=10 WT, 2 independent experiments). **(b)** Representative biaxial dot plots of CD9 by CD104 as in (a) colored by Myogenin expression. **(c)** Representative biaxial dot plots of CD9 by CD104 as in (a) colored by Pax7 expression. **(d)** Individual populations were sorted by FACS and cultured in differentiation media for one week. Images were acquired with an AxioPlan2 epifluorescent microscope (Carl Zeiss) with ORCA-ER digital camera (Hamamatsu Photonics). Each population was differentiated to yield fusion competent cells (n=4, 2 independent experiments). Scale bar, 50 μm. **(e)** Representative images showing the gating strategy on samples analyzed by flow cytometry at day 0 (upper panels) and day 6 (lower panels) (Figure 3d-f). Live cells are identified based on lack of DAPI staining. Lineage$^+$ cells (CD45$^+$/CD11b$^+$/CD31$^+$/Sca1$^+$) are excluded from the analysis and myogenic cells are enriched by gating on the α7integrin$^+$/CD9$^+$ fraction. A biaxial plot of CD9 (y axis) by CD104 (x axis) (far right) shows populations SC, P1 and P2 (representative images, n= 3 mice per condition).

a

b
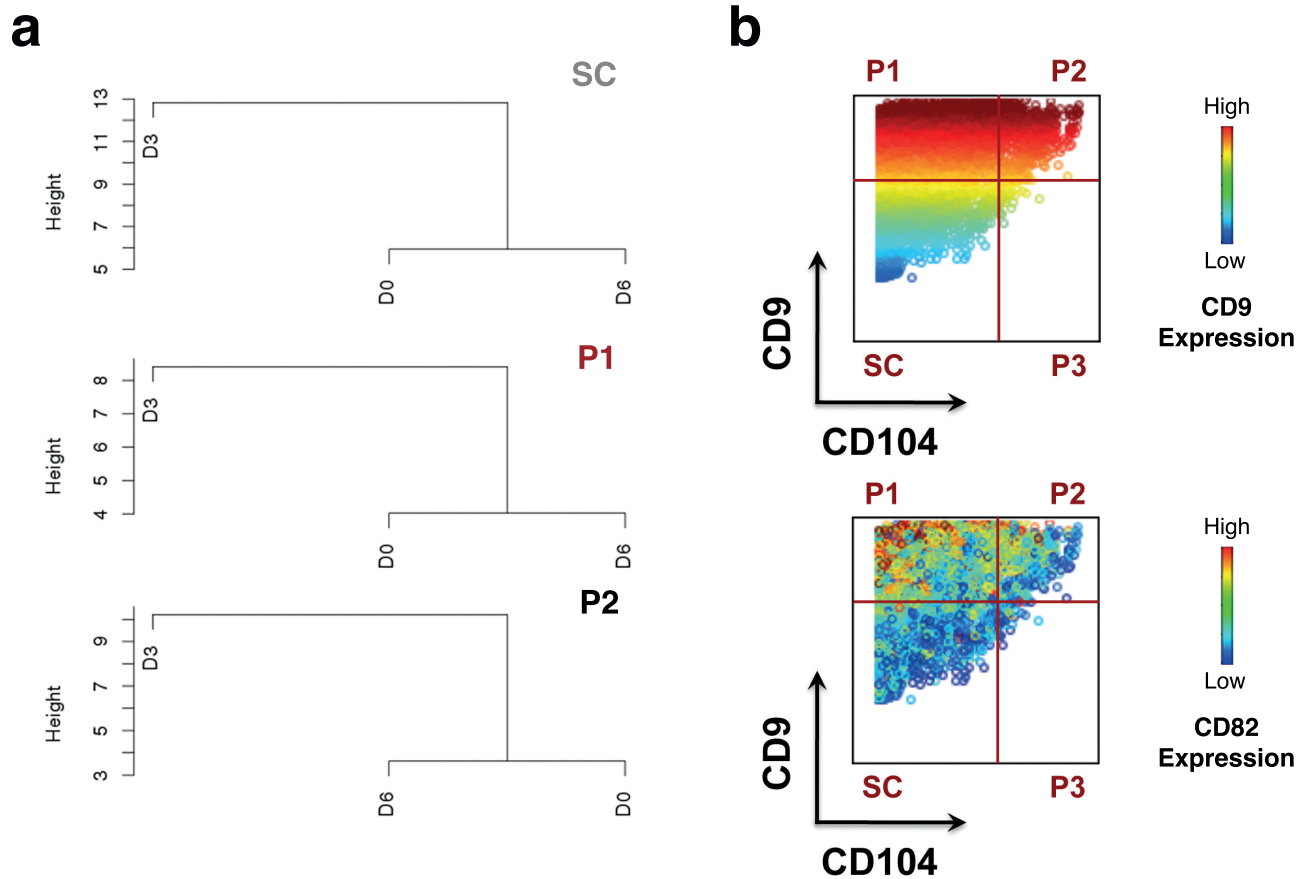
Supplementary Figure 3 (related to Figure 5). Molecular characterization of stem and progenitor cells during acute muscle injury identifies cell state transitions. **(a)** Supervised clustering analysis enforcing a 2-cluster solution in the stem (SC) and progenitor (P1, P2) cell populations during the time course of recovery from acute injury (Day 0 = D0; Day 3 = D3; Day 6 = D6), (representative experiment, n=3 mice per condition). Dendrograms were fitted using hclust function in R. The distance matrix was calculated through the dist function using euclidean parameters. **(b)** Representative biaxial dot plot of CD9 by CD104 colored by channel, showing expression of CD9 (upper panel) and CD82 (lower panel), in adult mice (n= 6 mice, 2 independent experiments). CD82 is highly expressed only in a small subset of populations P1 and P2.

Supplementary Table Legends

Supplementary Table 1 (related to Figure 1). Markers measured by Single-Cell Mass Cytometry. Single-cell mass cytometry antibodies and reagents panel, with description of clone number and supplier.

Supplementary Table 2 Statistics Source Data. Source data and statistics of all experimental replicates relative to representative images presented in the main figures are compiled.