

ARTICLE

Received 29 Mar 2017 | Accepted 19 Apr 2017 | Published 1 Jun 2017

DOI: 10.1038/ncomms15664

OPEN

# The microprotein Minion controls cell fusion and muscle formation

Qiao Zhang<sup>1</sup>, Ajay A. Vashisht<sup>1</sup>, Jason O'Rourke<sup>1</sup>, Stéphane Y. Corbel<sup>1</sup>, Rita Moran<sup>1</sup>, Angelica Romero<sup>1</sup>, Loren Miraglia<sup>1</sup>, Jia Zhang<sup>1</sup>, Eric Durrant<sup>1</sup>, Christian Schmedt<sup>1</sup>, Srinath C. Sampath<sup>1,2,\*</sup> & Srihari C. Sampath<sup>1,2,\*</sup>

Although recent evidence has pointed to the existence of small open reading frame (smORF)-encoded microproteins in mammals, their function remains to be determined. Skeletal muscle development requires fusion of mononuclear progenitors to form multinucleated myotubes, a critical but poorly understood process. Here we report the identification of Minion (microprotein inducer of fusion), a smORF encoding an essential skeletal muscle specific microprotein. Myogenic progenitors lacking Minion differentiate normally but fail to form syncytial myotubes, and Minion-deficient mice die perinatally and demonstrate a marked reduction in fused muscle fibres. The fusogenic activity of Minion is conserved in the human orthologue, and co-expression of Minion and the transmembrane protein Myomaker is sufficient to induce cellular fusion accompanied by rapid cytoskeletal rearrangement, even in non-muscle cells. These findings establish Minion as a novel microprotein required for muscle development, and define a two-component programme for the induction of mammalian cell fusion. Moreover, these data also significantly expand the known functions of smORF-encoded microproteins.

<sup>1</sup>Genomics Institute of the Novartis Research Foundation, 10675 John Jay Hopkins Drive, San Diego, California 92121, USA. <sup>2</sup>Division of Musculoskeletal Imaging, Department of Radiology, University of California San Diego School of Medicine, 200 West Arbor Drive, San Diego, California 92103, USA.

\* These authors jointly supervised this work. Correspondence and requests for materials should be addressed to Sn.C.S. (email: ssampath@gnf.org) or to Sh.C.S. (email: hsampath@gnf.org).

In addition to canonically defined protein-coding genes, recent studies have indicated the existence of a new class of mammalian genes<sup>1</sup>. These small open reading frames (smORFs) are transcribed and translated by usual means, but are largely unrecognized as protein-coding genes by virtue of their size, typically encoding microproteins <100 amino acids (aa) in length<sup>2</sup>. Although estimates vary widely, the human and mouse genomes are thought to contain at least several thousand of these 'hidden' protein-coding genes<sup>2</sup>. Intriguingly, of the small number of currently known mammalian microproteins, several have been identified in muscle<sup>3–7</sup>. These largely encode regulatory factors for the sarco/endoplasmic reticulum Ca<sup>2+</sup>-ATPase (SERCA), with structural similarity to known SERCA-regulatory proteins such as sarcolipin and phospholamban<sup>3,5,7</sup>. Of note however, no essential mammalian microprotein has been described.

Skeletal muscle development requires temporally regulated stem cell activation and differentiation, fusion of progenitors to form syncytial myotubes and maturation of myotubes to generate contractile myofibres. While the early and late stages of this process have been intensively studied<sup>8,9</sup>, our understanding of the mechanisms and regulatory factors controlling cell fusion remains incomplete, particularly in mammals<sup>10,11</sup>. A recent major advance was the identification of the transmembrane protein Tmem8c/Myomaker, which is necessary for myoblast fusion and sufficient for fusion of non-muscle cells to differentiating muscle. Importantly however, Myomaker expression alone cannot induce fusion of non-muscle cells with one another, suggesting the existence of one or more additional factors that are expressed in differentiating muscle cells and required to drive *de novo* cell fusion<sup>12,13</sup>.

In this study we report the discovery of a novel smORF-encoded essential microprotein which we term Minion (microprotein inducer of fusion). We demonstrate that Minion is required for skeletal muscle development, and together with Myomaker defines a minimal two-component programme for the induction of mammalian cell fusion. In addition to the implications for muscle biology, these data also significantly expand the known functions of smORF-encoded microproteins, an under-explored source of proteomic diversity.

## Results

**Identification of Minion/gm7325.** To identify novel microproteins playing key roles in skeletal muscle, we performed whole transcriptome RNA-seq analysis of uninjured and regenerating muscle. We specifically sought to identify novel transcripts demonstrating strong temporal regulation, annotated open reading frame (ORF) length of <100 codons, and a corresponding dynamic pattern of transcriptional regulation during mouse myoblast differentiation *in vitro* (Fig. 1a, left). We focused on gene regulation at day 3 post injury to exclude effects related to the immediate post-injury immune response. The predicted gene 7325 (*gm7325*) (GenBank accession number KY857877) was the only gene meeting all criteria, encoding a putative 84 aa microprotein with possible expression in embryonic stem (ES) cells but no known function<sup>14</sup>. For reasons described below, we named this gene *Minion* (microprotein inducer of fusion). The temporal pattern of *Minion* expression was distinct from that of two other smORFs, but notably was similar to that of Myomaker (Fig. 1a)<sup>12</sup>.

### Minion expression in regenerating and developing muscle.

Western blot confirmed that the *Minion* transcript is translated; Minion protein was absent in uninjured tibialis anterior (TA) muscle but strongly induced during regeneration, peaking 3–4

days following injury (Fig. 1b). Immunofluorescence analysis demonstrated Minion expression within nascent regenerating myofibres (Supplementary Fig. 1a), whereas Minion protein was not detectable in uninjured adult muscle (Supplementary Fig. 1b) nor in multiple additional non-muscle tissues (Supplementary Fig. 1c). RNA-seq analysis of early embryonic development revealed *Minion* expression which was detectable as early as somite stage 15 but greatly increased by somite stage 36, following limb and tail bud formation (Supplementary Fig. 1d). Expression of *Minion* was seen in embryonic skeletal muscle of both somitic (limb, tongue) and non-somitic (extraocular and facial muscles) origin, but importantly not in embryonic or neonatal heart muscle (Fig. 1c, Supplementary Fig. 1e). Both mRNA and protein levels of Minion increased rapidly during *in vitro* myoblast differentiation (Fig. 1d and Supplementary Fig. 2a–c).

### Minion is a conserved membrane-associated microprotein.

Although the full-length Minion protein is predicted to contain an N-terminal signal sequence and predominant alpha-helical secondary structure (Fig. 1e), overexpression and supernatant concentration demonstrated no evidence of protein secretion (Supplementary Fig. 2d). Subcellular fractionation did however confirm significant enrichment within the membrane-associated fraction containing plasma membrane, ER and Golgi, suggesting insertion into or association with a membrane compartment (Supplementary Fig. 2e).

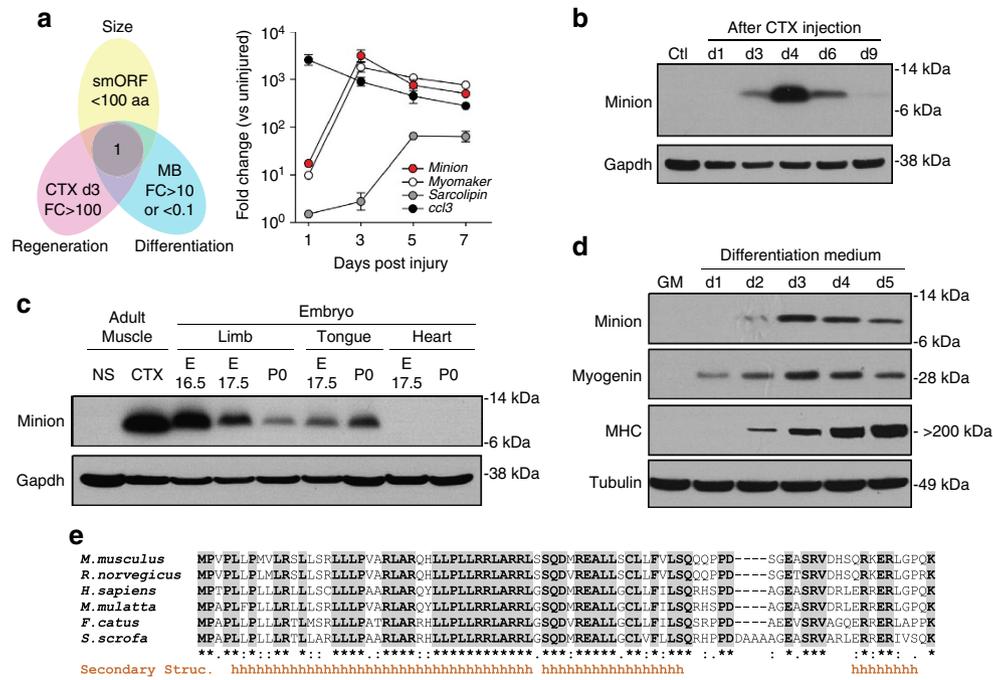
TBLASTN search revealed a putative human *MINION* homologue (*hMINION*) with an intact ORF of 84 codons (GenBank accession number KY857877) (Supplementary Fig. 3a), despite prior annotation of the transcript as a long noncoding RNA (Ensembl gene ID ENSG00000262179 and gene name *RPI-302G2.5*; NCBI gene ID 101929726 and gene name *LOC101929726*). Evolutionary conservation of the protein-coding sequence was seen across mammalian species (Fig. 1e), however no convincing sequence homologue was found in *Drosophila* or other invertebrate species. No amino acid sequence similarity was seen to sarcolipin, phospholamban, or the recently reported microprotein DWORF<sup>7</sup>.

### The *Minion* promoter contains conserved E-box elements.

We noted that Minion expression during muscle cell differentiation slightly trailed that of the basic helix–loop–helix transcription factor Myogenin (Fig. 1d), suggesting control by canonical muscle regulatory factors (MRFs, for example, MyoD and Myogenin). Indeed, analysis of the upstream regulatory regions of human and mouse *Minion* loci revealed evolutionarily conserved E-box binding sites for MRFs (Supplementary Fig. 3a). Both MyoD and Myogenin specifically bound these sites in differentiating myoblasts, as shown by ENCODE whole genome ChIP-seq (Supplementary Fig. 3b)<sup>15</sup>.

### Generation of Minion-deficient mice by CRISPR/Cas9 editing.

The spatial and temporal pattern of Minion expression together with the presence of functional MRF-binding E-boxes strongly suggested a role for Minion in skeletal muscle development. To test this, we used CRISPR/Cas9 gene editing to generate *Minion*-deficient mice carrying diverse small mutations and large deletions. For generation of each allele, two guide RNAs (gRNAs) targeting the single coding exon were coinjected into embryos, F<sub>0</sub> pups were screened for mutations (Fig. 2a, Supplementary Figs 4 and 5) and two founder lines containing either a 135 bp in-frame deletion (*Minion*<sup>Δ/Δ</sup>; Fig. 2a,b, Supplementary Fig. 4a–c) or a 155 bp frameshift deletion (Supplementary Fig. 5a,b,d) were established. Subsequent experiments are with the 135 bp deletion allele unless otherwise mentioned.



**Figure 1 | The microprotein Minion is specifically expressed during skeletal muscle development and regeneration.** (a) Left: Overlap of RNA-seq from regenerating adult mouse *tibialis anterior* (TA) muscle and differentiating C2C12 myoblasts (MB). CTX, cardiotoxin; FC, fold change compared to uninjured muscle (bottom left) or undifferentiated myoblasts (bottom right). Right: fold change of reads per kilobase per million mapped reads (RPKM) for selected genes upregulated after CTX injury. Values are normalized to uninjured muscle, representing mean  $\pm$  s.d. of fold change, three 8–10 week old mice per time point. (b) Western blot of control uninjured (Ctl) and CTX-injured regenerating adult TA muscle ( $n = 2$ , two 8–10 week old mice per time point, two technical replicates each). (c) Western blot of embryonic muscle samples ( $n = 3$ , two embryos each). Day 4 post-CTX TA or normal saline (NS) injection were positive and negative controls. E, embryonic day; P, post-natal. (d) Western blot of C2C12 myoblasts cultured under growth conditions (GM) or under differentiation conditions (DM) for the indicated number of days ( $n = 4$ ). MHC, myosin heavy chain. (b–d) Glyceraldehyde-3-phosphate dehydrogenase (Gapdh) and Tubulin served as loading controls. (e) Protein sequence alignment of mouse Minion with putative orthologues from other mammalian species (GenBank accession numbers and UniProt IDs from top to bottom are: NP\_001170939.1, XP\_017452417.1, NP\_001302423.1, EHH18375.1, M3X8W7\_FELCA and FIRQU5\_PIG.)

**Minion loss leads to defective muscle fusion and lethality.** Although heterozygous *Minion* <sup>$\Delta$ /+</sup> animals were viable and recovered at expected Mendelian ratios, live homozygous mutant *Minion* <sup>$\Delta$ / $\Delta$</sup>  animals could only be observed prenatally, and no viable neonatal or adult *Minion* <sup>$\Delta$ / $\Delta$</sup>  animals were recovered (Supplementary Figs 4c,d and 5b,e). Loss of Minion protein was confirmed in both embryonic and perinatal limb and tongue skeletal muscle from *Minion*-deficient animals (Supplementary Fig. 6). These findings are consistent with perinatal lethality in the absence of Minion.

Late stage *Minion* <sup>$\Delta$ / $\Delta$</sup>  embryos were clearly distinguishable by their decreased size and weight, reduced limb diameter, spinal curvature and atony, as well as by the dorsal and nuchal subcutaneous oedema seen at E16 and earlier stages (Supplementary Figs 5c and 7). *Minion* <sup>$\Delta$ / $\Delta$</sup>  E17.5 embryos and P0 neonatal pups demonstrated diminutive forelimb and intercostal musculature and decreased total size of muscle groups (Fig. 2c, Supplementary Fig. 7b). No gross impairment was seen at very early embryonic stages where the lack of formed muscle makes analysis difficult (Supplementary Fig. 8); in contrast, clear abnormality in skeletal muscle formation was seen at E18.5–19.5, as judged by both histology and immunofluorescence staining for the muscle cell markers myosin heavy chain (MHC) and Desmin. Whereas control tongue skeletal muscle contained abundant elongated polynucleated ( $\geq 3$  nuclei) myotubes, *Minion* <sup>$\Delta$ / $\Delta$</sup>  muscle demonstrated marked reduction in fused fibres, with accumulation of both short nascent fibres as well as unfused mononucleated cells (Fig. 2d, Supplementary Fig. 9). Similar defects were present in *Minion* <sup>$\Delta$ / $\Delta$</sup>  forelimb, facial and jaw

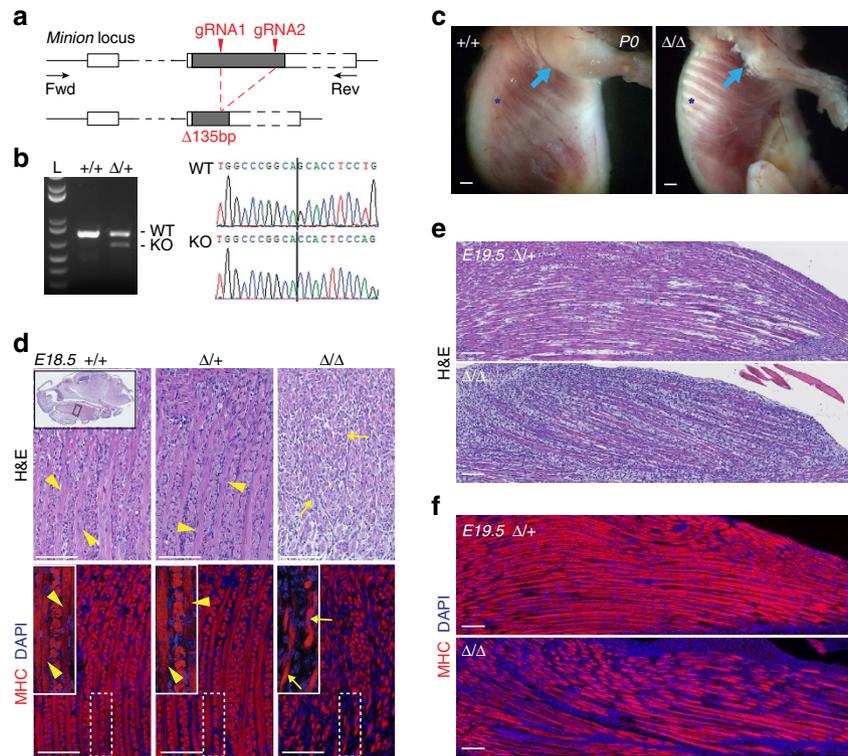
musculature, diaphragm and intercostal musculature (Figs 2e,f, 3 and 4a,b, Supplementary Fig. 8d).

**Disruption of respiratory function in *Minion* <sup>$\Delta$ / $\Delta$</sup>  mice.**

The observed defects above suggested that the perinatal lethality of *Minion* <sup>$\Delta$ / $\Delta$</sup>  animals could reflect disruption of respiratory function, a possibility we further tested by assessment of lung inflation. Late stage fetuses were delivered by caesarean section and monitored for 1 h during exposure to room air. Of note, all E18.5 *Minion* <sup>$\Delta$ / $\Delta$</sup>  embryos died soon after delivery. After 1 h, lungs were dissected and subjected to flotation testing. In keeping with the dramatic decrease in diaphragm and intercostal muscle formation, lungs from *Minion* <sup>$\Delta$ / $\Delta$</sup>  but not control animals failed to float (Fig. 4c), demonstrating absence of postnatal lung inflation.

**Minion-deficient myoblasts exhibit defective fusion *in vitro*.**

The marked reduction of polynucleated myofibres in *Minion* <sup>$\Delta$ / $\Delta$</sup>  muscle suggested that Minion might specifically function in the process of myoblast fusion. Indeed, induction of differentiation in *Minion* <sup>$\Delta$ / $\Delta$</sup>  primary embryonic myoblasts resulted in near complete failure to form polynucleated myotubes (Fig. 5a,b, Supplementary Fig. 10). Importantly, markers of myogenic commitment and terminal differentiation were induced normally in *Minion* <sup>$\Delta$ / $\Delta$</sup>  myoblasts both *in vivo* and *in vitro* (Figs 2–4, Supplementary Figs 9 and 10), suggesting that the muscle formation defect did not result from a block to progenitor differentiation *per se*.



**Figure 2 | Minion is required for skeletal muscle development.** (a) Strategy for CRISPR/Cas9 mutagenesis of the *gm7325/Minion* locus using a dual sgRNA approach. Grey box, *Minion* ORF; white box, non-coding exons; sgRNA, single gRNA; Fwd and Rev, forward and reverse genotyping primers. (b) Left: representative genotyping PCR of *Minion* wild type (+/+) and heterozygous ( $\Delta$ /+) mice carrying the 135-bp deletion depicted in a.  $n = 20$  (more than 300 total adult mice of both sex). Right: representative sequence traces. Black line indicates 5' boundary of the deletion. WT, wild-type allele; KO, knockout allele (135-bp deletion). (c) Photographs of skinned *Minion*<sup>+/+</sup> and *Minion* <sup>$\Delta$ / $\Delta$</sup>  P0 mice. Cyan arrows and Blue asterisks indicate forelimb and intercostal musculature, respectively. Three litters of embryos. (d) Histological and immunofluorescence (IF) analyses of embryonic tongue skeletal muscle from E18.5 embryos. Yellow arrowheads and yellow arrows indicate multinuclear myofibres and unfused differentiating elongating myoblasts, respectively. Top row: hematoxylin and eosin (H&E) staining of sagittal tongue sections. Inset demonstrates the originating region and orientation of the provided tongue sections. Bottom row: Immunofluorescence staining for the muscle marker MHC (red), with DNA counterstain DAPI (4',6-diamidino-2-phenylindole; blue). Insets demonstrate magnification of the boxed areas. Three litters of embryos. (e) Histological images of H&E-stained E19.5 forelimb longitudinal sections of indicated genotypes. MHC (red) and DAPI (blue) staining are shown. Two litters of embryos. (f) Immunofluorescence images of forelimb longitudinal sections for E19.5 embryos with indicated genotypes. MHC (red) and DAPI (blue) staining are shown. Two litters of embryos. (d-f) Paraffin-embedded embryos of different stages were examined. (c-f) Embryos from the same litter were compared with 1-2 embryos for each genotype in each experimental repeat. Scale bars, 1 mm (c), 100  $\mu$ m (d-f).

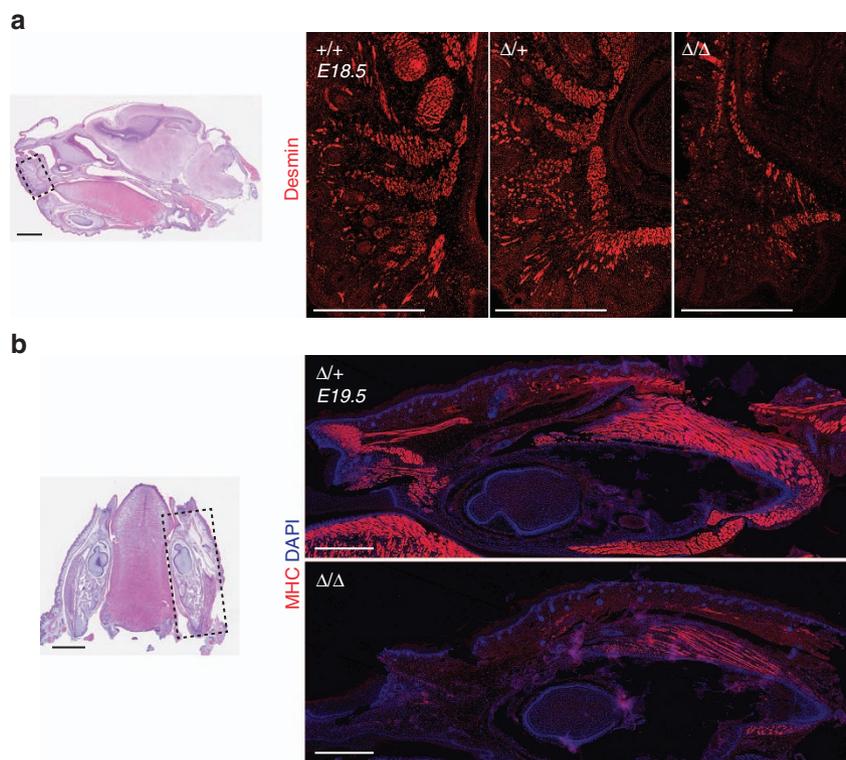
This was further confirmed using loss-of-function in both immortalized and primary mouse myoblasts via stable lentiviral transduction with shRNAs targeting the *Minion* coding sequence and 3' untranslated region (UTR) (Supplementary Fig. 11a). Near complete suppression of Minion expression was achieved using individual shRNAs (Supplementary Fig. 11b), and a combination of the two most active shRNAs resulted in undetectable protein levels in differentiating cells (*Minion*<sup>KD</sup>; Fig. 5c). Immunofluorescence staining of both wild type and *Minion*<sup>KD</sup> myoblasts demonstrated endogenous expression of Minion protein in both differentiating mononuclear myoblasts and nascent multinuclear myotubes (Supplementary Fig. 12a,b). This expression pattern was further confirmed by immunofluorescence staining of longitudinal regenerating limb muscle sections at 3 days post cardiotoxin (CTX) injection (Supplementary Fig. 12c).

Analysis of Myogenin, MyoD, Desmin and MHC expression confirmed both the absence of any molecular differentiation defect in *Minion*<sup>KD</sup> cells (Fig. 5c-e, Supplementary Fig. 13b), as well as the presence of a severe block to myoblast fusion (Fig. 5d,f,g). Interestingly, differentiating *Minion*-deficient myoblasts elongated and aligned normally despite failing to fuse (Fig. 5d, Supplementary Fig. 10), suggesting that myoblast apposition was not impaired. Similar results were obtained using

lentiviral shRNA transduction of primary, non-immortalized adult mouse myoblasts (Supplementary Fig. 13).

**Human MINION rescues the *Minion*<sup>KD</sup> fusion defect.** As the shRNAs used to target the *Minion* transcript in *Minion*<sup>KD</sup> cells recognize the 3' UTR, we tested the ability of various ORF cDNA clones to complement the *Minion*<sup>KD</sup> cell fusion defect (Supplementary Fig. 14a). Both full-length untagged and C-terminally tagged mouse Minion robustly rescued myoblast fusion (Supplementary Fig. 14b), demonstrating that the fusion defect observed in *Minion*<sup>KD</sup> cells was not the result of off-target effects. The putative human orthologue, previously annotated as a long noncoding RNA (GRCh37 genome assembly), was then tested in a similar complementation assay, demonstrating that both untagged and C-terminally epitope-tagged human MINION ORFs strongly reconstituted cell fusion in *Minion*<sup>KD</sup> cells (Fig. 5h, Supplementary Fig. 14b).

To definitively establish that these ORFs function via protein coding, single nucleotide insertions or deletions were introduced into the untagged mouse and human MINION cDNAs, respectively. These frameshift point mutants failed to complement the fusion defect (Supplementary Fig. 15),



**Figure 3 | Minion deficiency blocks non-somitic skeletal muscle formation.** (a) Immunofluorescence images of non-somitic facial musculature from sagittal sections of E18.5 embryos with the indicated genotypes. Desmin (red) staining is shown. Black box in histological image at left demonstrates the area shown at right in fluorescence images. (b) Immunofluorescence images of non-somitic jaw and facial musculature on transverse sections from E19.5 embryos with the indicated genotypes. MHC (red) and DAPI (blue) staining are shown. (a,b) Two litters of embryos were examined, and 1–2 embryos of each genotype and in the same litter were compared in each experimental repeat. Scale bars, 1 mm for H&E images and 500  $\mu$ m for immunofluorescence images.

confirming that these transcripts function not as non-coding RNAs but by encoding functional microproteins. Reconstitution of *Minion*<sup>KD</sup> cells with cDNA mimicking the 135 bp deletion allele found in the *Minion* <sup>$\Delta/\Delta$</sup>  mice likewise failed to rescue cell fusion (Supplementary Fig. 16), confirming that this represents a true loss-of-function allele. Taken together, these data demonstrate that *Minion* encodes a microprotein essential for skeletal muscle formation via a specific function in myoblast fusion.

#### Minion is required for Myomaker-induced fusion.

The requirement for Minion in cell fusion appears muscle specific, as Minion expression was not seen in other settings of physiologic cell fusion, such as in the placenta or in fusing macrophage lineage cells (Supplementary Fig. 17). This restricted expression pattern mirrors that of Myomaker, a recently described transmembrane regulator of myoblast fusion, and we therefore investigated the functional relationship between these proteins to better understand the mechanism of Minion-associated cell fusion. We observed that Myomaker expression was readily detectable in regenerating muscle and differentiating mouse myoblasts (Supplementary Fig. 18), when Minion levels are likewise high (Fig. 1b–d).

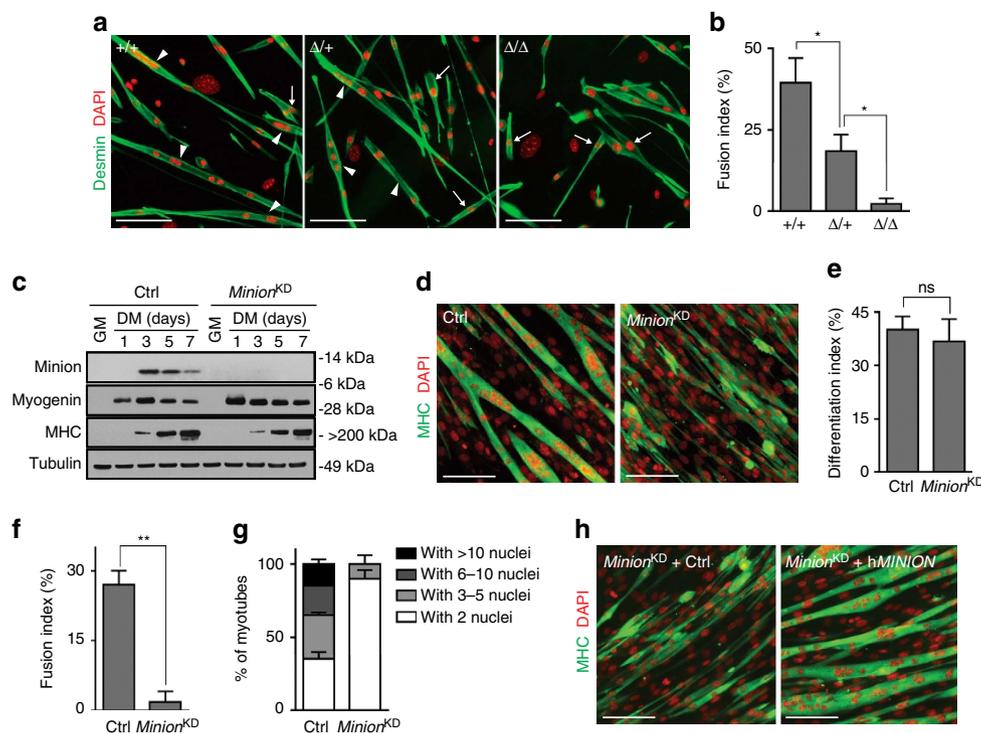
It was previously reported that Myomaker is sufficient to induce the fusion of fibroblasts to differentiating myoblasts<sup>4</sup>, and we therefore examined whether Minion exhibited a similar phenotype when expressed in fibroblasts. C3H 10T1/2 fibroblasts were infected with retroviruses encoding GFP and either control, Myomaker or Minion; GFP-positive fibroblasts were then mixed with wild-type C2C12 myoblasts and kept in differentiation medium for 3–4 days. Differentiating myoblasts and myotubes

were marked by MHC expression, and fusion of GFP-positive fibroblasts to myotubes was readily detected when Myomaker, but not Minion or control, was expressed (Fig. 6a). However, Myomaker-expressing fibroblasts failed to fuse to differentiating *Minion*<sup>KD</sup> myoblasts (Fig. 6b). Minion loss did not impair Myomaker expression (Fig. 6c), indicating that the failure of *Minion*<sup>KD</sup> myoblasts to fuse with Myomaker-expressing fibroblasts is not due to absence of Myomaker in the myoblasts. Myomaker overexpression was likewise incapable of rescuing the intrinsic fusion defect in *Minion*<sup>KD</sup> myoblasts (Fig. 6d,e), demonstrating that Minion is required for Myomaker-induced fusion between differentiating myoblasts. We therefore conclude that the cell fusion defect seen in the absence of Minion does not reflect deficiency in Myomaker expression, and that Minion is required in differentiating muscle cells for Myomaker-induced fusion.

**Minion and Myomaker together induce cell fusion.** Previous studies have demonstrated that expression of Myomaker alone fails to induce fusion between non-myogenic cells, and that at least one additional, as yet unidentified factor is required<sup>4</sup>. We likewise observed that heterologous expression of neither Myomaker nor Minion alone in 10T1/2 fibroblasts was sufficient to drive fusion of these cells with one another (Fig. 7a–c). Remarkably, however, we observed that simultaneous expression of Minion and Myomaker together drove rapid and uniform fusion of transduced fibroblasts with one another, leading to the formation of large multinuclear syncytia (Fig. 7b,c).

To further demonstrate that syncytium formation induced by Minion and Myomaker represents true cell fusion and not





**Figure 5 | Minion is specifically required for fusion of skeletal muscle progenitors.** (a) Immunofluorescence of primary embryonic myoblasts isolated from E18.5 *Minion*<sup>+/+</sup>, *Minion*<sup>Δ/+</sup> and *Minion*<sup>Δ/Δ</sup> embryos, following 3 days in DM. Desmin (green) and DAPI (red). White arrowheads: myotubes; white arrows: elongating myoblasts. *n* = 3 (five technical replicates each). (b) Fusion index of myoblasts in a, calculated as % nuclei in Desmin<sup>+</sup> myotubes (≥3 nuclei) of total nuclei in Desmin<sup>+</sup> cells. Asterisk: *P* < 0.05, unpaired two-tailed Student's *t*-test. Each value represents mean ± s.d. *n* = 4 (two 0.7 mm × 0.7 mm fields each). (c) Western blots of C2C12 myoblasts cultured in GM or DM. Cells were lentivirally infected with either control *luciferase* targeting shRNA (Ctrl) or serially with two shRNA targeting the *Minion* 3'UTR (*Minion*<sup>KD</sup>) and cultured in GM or DM for the indicated number of days. *n* = 3. (d) Immunofluorescence images of Ctrl and *Minion*<sup>KD</sup> myofibres following 5 days in DM. MHC (green) and DAPI (red). *n* = 5 (eight technical replicates each). (e) Differentiation index for d, calculated as % nuclei in MHC<sup>+</sup> cells of total nuclei. NS, not significant, unpaired two-tailed Student's *t*-test. (f) Fusion index for d, calculated as % nuclei in MHC<sup>+</sup> myotubes (≥3 nuclei) of total nuclei. Double asterisks: *P* < 0.001, unpaired two-tailed Student's *t*-test. (g) Quantification of myotubes by nuclei number for d. For e–g, each value represents mean ± s.d. *n* = 4 (two 0.7 mm × 0.7 mm fields each). (h) Immunofluorescence images of *Minion*<sup>KD</sup> cells expressing either control protein (NanoLuc) or human *MINION* orthologue, after 5 days in DM. MHC (green) and DAPI (red). *n* = 3 (six technical replicates each). Scale bars, 100 μm.

development, in which fusion requires active cytoskeletal remodelling and occurs between distinct populations of founder cells and fusion competent myoblasts<sup>16–22</sup>. Our findings are further confirmed by additional independent studies<sup>23,24</sup>.

Despite the severe fusion defect observed upon loss of Minion or Myomaker, we note a difference between Minion-deficient myoblasts and Myomaker-deficient myoblasts. Differentiating Minion-deficient myoblasts elongate and align normally (Fig. 5d, Supplementary Figs 10 and 13), while the previously reported Myomaker-deficient myoblasts do not appear to align<sup>12</sup>, suggesting that Myomaker might play a role in initial cell–cell recognition. This suggests that in the setting of Myomaker expression, Minion-deficient myoblasts are able to undergo apposition, but not fusion pore formation. These distinct roles are consistent with the lack of detectable physical interaction between Minion and Myomaker. We suggest a model in which the transmembrane protein Myomaker induces interaction of cell membranes<sup>25</sup>, whereas the microprotein Minion drives the cytoskeletal reorganization needed for fusion (Fig. 8c). The precise molecular mechanism connecting Minion to the cytoskeletal remodelling and additional pathways leading to fusion pore formation remains to be determined.

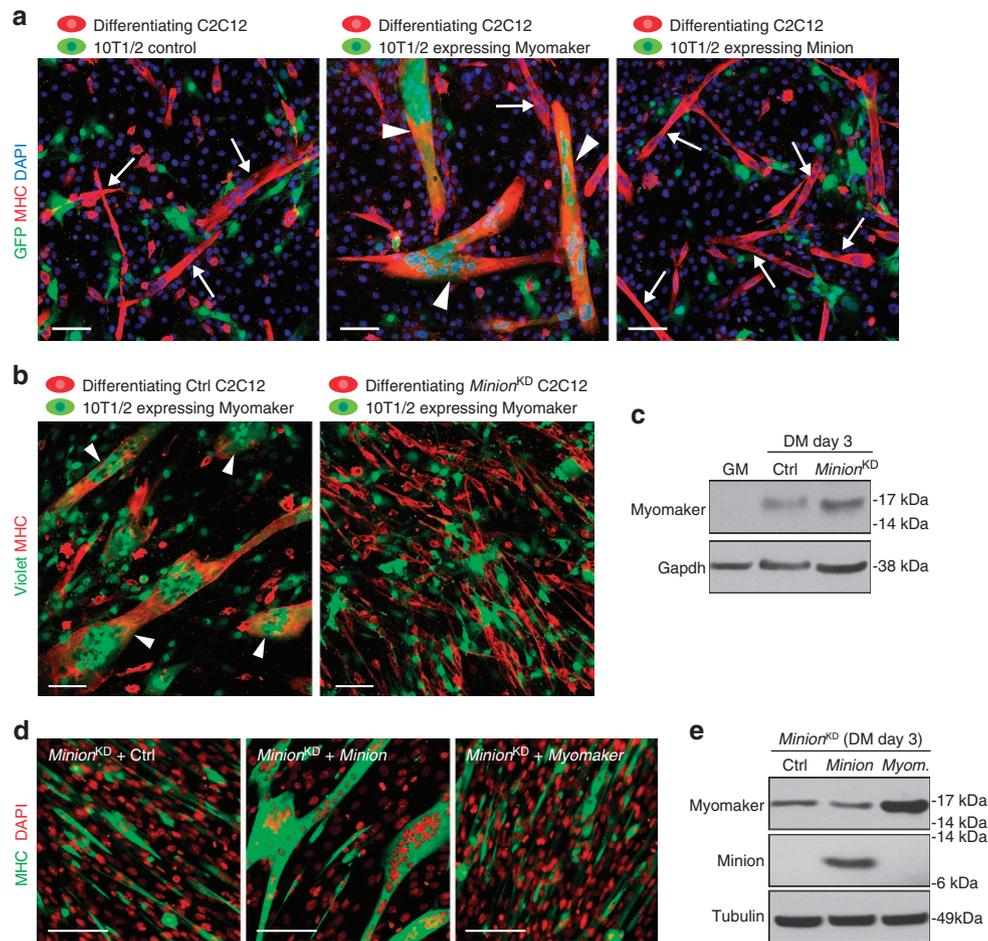
The two-factor fusion system described here opens the door to programmable and potentially targeted cell fusion, which may find therapeutic and research application in oncolytic fusion of cancer cells<sup>26</sup>, fusion of dendritic cells to cancer cells in

immunotherapy<sup>27</sup>, therapeutic fusion in regenerative medicine<sup>28,29</sup> and heterokaryon-based studies of nuclear reprogramming<sup>30</sup>. Finally and importantly, our studies constitute the first report, to our knowledge, of an essential mammalian microprotein. This represents the strongest evidence to date that this class of diminutive proteins in fact constitutes a 'microproteome' with critical and largely unexplored functions.

## Methods

**Animals.** All animal experiments were approved by the GNF IACUC and carried out in accordance with approved guidelines. C57BL/6J mice were initially purchased from the Jackson Laboratory and expanded through in-house breeding. For details regarding generation of genetically engineered mice, please see 'Generation of *Minion*-knockout mice by CRISPR/Cas9-mediated gene editing' below. For experiments with adult animals, 6–12 weeks old male C57BL/6J mice were used unless otherwise mentioned.

**Cardiotoxin injury model.** The cardiotoxin injury model is a well-established model to study mouse skeletal muscle regeneration. CTX from *Naja mossambica* (Sigma, C9759) was dissolved in normal saline (0.9% w/v of NaCl) to make a 10 μM working solution, and was aliquoted and stored at –20 °C. After anaesthesia of the mouse with isoflurane (1.5–2% in oxygen), the anterior aspect of the adult mouse hindlimb was shaved to expose the skin, and ~50 μl of CTX solution was injected into the midbelly of the TA muscle using a 0.3 ml U100 BD insulin syringe. TA muscles were collected and examined at different time points after CTX injection. Adult mice at a similar age if not from the same litter without CTX injection or with equal volume normal saline injection were used as controls, as indicated in the figures. For western blot analyses, two mice were used for each



**Figure 6 | Minion is required for Myomaker-mediated fusion.** (a) Immunofluorescence images of cell-mixing between 10T1/2 fibroblasts and wild-type C2C12 myoblasts (1:2 ratio) after 3 days in differentiation medium. Differentiating myoblasts and myotubes are marked by MHC (red). 10T1/2 fibroblasts were infected with retrovirus expressing GFP (green) and proteins of interest (left, NanoLuc control; middle, mouse Myomaker; right, mouse Minion). DAPI marks nuclei. Fibroblasts expressing Myomaker fused with wild-type differentiating myoblasts and myotubes to become large thick myotubes (white arrowheads), while fibroblasts expressing Minion failed to do so. White arrows indicate MHC-positive myotubes that are not fused to fibroblasts.  $n = 2$  (eight technical replicates each).  $0.7 \text{ mm} \times 0.7 \text{ mm}$  fields are shown. (b) Immunofluorescence images of cell mixing between 10T1/2 fibroblasts expressing Myomaker and C2C12 myoblasts after 3 days in DM.  $0.7 \text{ mm} \times 0.7 \text{ mm}$  fields at  $\times 20$  magnification are shown. Control and *Minion*<sup>KD</sup> myoblasts were used. Differentiating myoblasts and myotubes are marked by MHC (red). 10T1/2 fibroblasts expressing Myomaker were labelled with CellTrace Violet dye before mixing (pseudocoloured in green). Fibroblasts expressing Myomaker fused with differentiating control myoblasts (white arrowheads), but failed to fuse to *Minion*<sup>KD</sup> myoblasts.  $n = 2$  (eight technical replicates each). (c) Western blot of wild-type C2C12 myoblasts in GM, as well as Ctrl and *Minion*<sup>KD</sup> myoblasts in DM for 3 days.  $n = 2$ . (d) Immunofluorescence images of *Minion*<sup>KD</sup> myoblasts expressing Luciferase (Ctrl), Minion, or Myomaker, after 5 days in DM. MHC (green) and DAPI (red).  $n = 2$  (seven technical replicates each). (e) Western blot of cell lines shown in d.  $n = 2$ . Scale bars,  $100 \mu\text{m}$ .

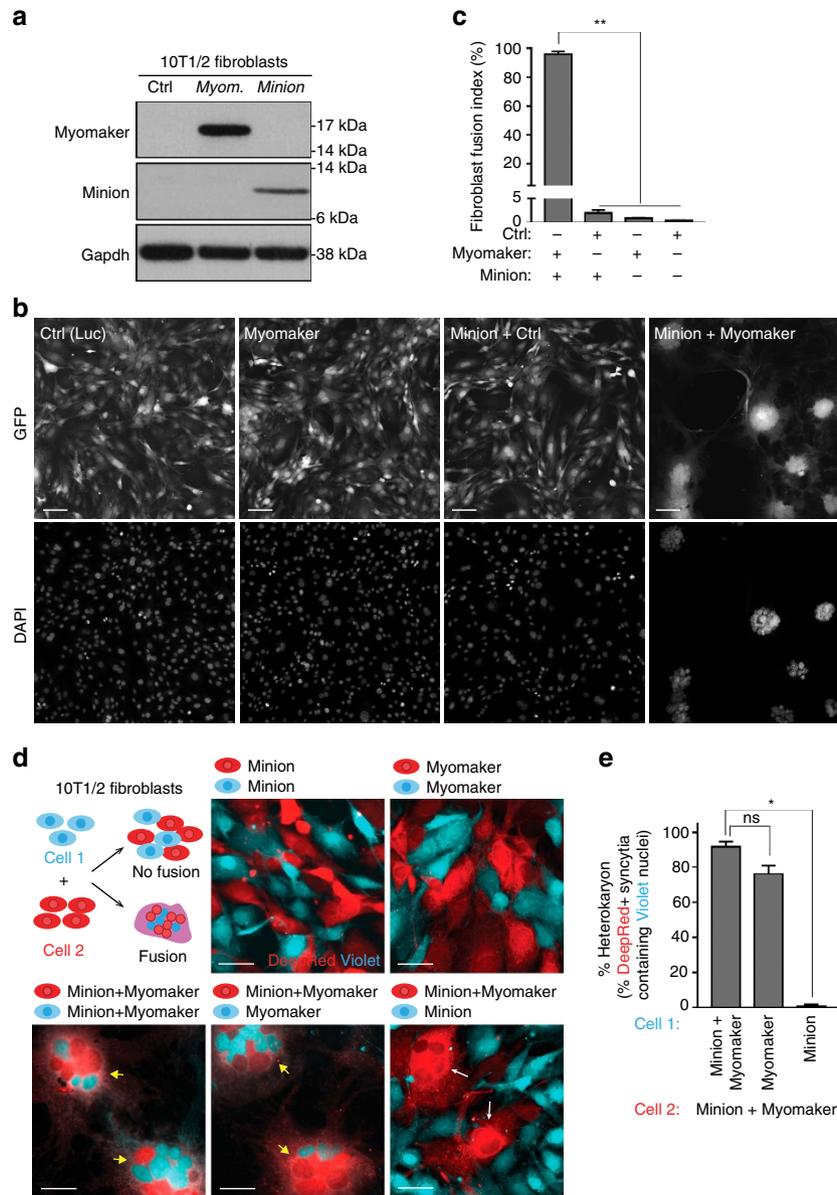
time point for each experimental repeat; for RNA sequencing experiments, three mice were used for each time point.

**Mouse skeletal muscle RNA sequencing.** Twelve 8–10 week old C57BL/6 mice were injected with CTX into the TA muscle as described above, and the TA muscles were collected with three mice for each time point. TA muscles from three 8–10 week old uninjured mice were also collected. Total RNA from each muscle sample was isolated by TRIzol Reagent (Thermo Scientific, 15596026) according to manufacturer's instructions and purified by Qiagen RNeasy columns. The RNA samples (three replicates for each time point) were submitted to the in-house Sequencing and Expression Analysis Core for quality checking, library preparation, and next-generation single-read sequencing using standard techniques. Briefly,  $1 \mu\text{g}$  of total RNA was used to make Illumina-compatible sequencing libraries, and the libraries were sequenced using 50 bp single reads on an Illumina HiSeq 1000. Reads were aligned to the mouse transcriptome (Refseq mouse transcripts as of March 2013) using BWA<sup>31</sup>. An average of 36 million reads per sample mapped to the mouse transcriptome. To analyse the raw data, reads per kilobase of transcript per million mapped reads (RPKM) were calculated for each gene at each time point, the RPKM of each gene was normalized to that of uninjured muscles, and the results were averaged to generate the fold change in expression level. The data were then analysed in two ways: (1) genes that exhibited more than 100-fold increase in CTX day 3 muscles compared to uninjured muscles were selected and (2) genes

annotated to contain an ORF of  $< 100$  codons were selected. Genes meeting both criteria were then further examined against in-house RNA-Seq data from undifferentiated and differentiated primary myoblasts and C2C12 immortalized myoblasts; *Minion* was the only novel smORF demonstrating a dynamic expression pattern, with greater than tenfold change between undifferentiated and differentiated myoblast samples.

**Developmental RNAseq analysis.** RNAseq analysis of early embryonic development was from the Deciphering the Mechanisms of Developmental Disorders program via EMBL-EBI Expression Atlas (<https://www.ebi.ac.uk/gxa>)<sup>32,33</sup>. The data set was interrogated using the following Ensembl IDs: mouse *Minion/gm7325*, ENSMUSG00000079471; mouse *myod1*, ENSMUSG0000009471).

**Generation of *Minion*-knockout mice by CRISPR/Cas9 editing.** Four week-old female C57BL/6J mice were superovulated by intraperitoneal injection of 5 IU pregnant mare's serum gonadotropin followed 47 h later by 5 IU of human chorionic gonadotropin (HCG). Female mice were mated to C57BL/6J male mice 1:1 immediately after HCG injection. The following morning, the females were checked for copulatory plugs and zygotes were collected from the oviducts of plugged females. *In vitro* transcribed Cas9 mRNA ( $100 \text{ ng} \mu\text{l}^{-1}$ ) and two gRNAs



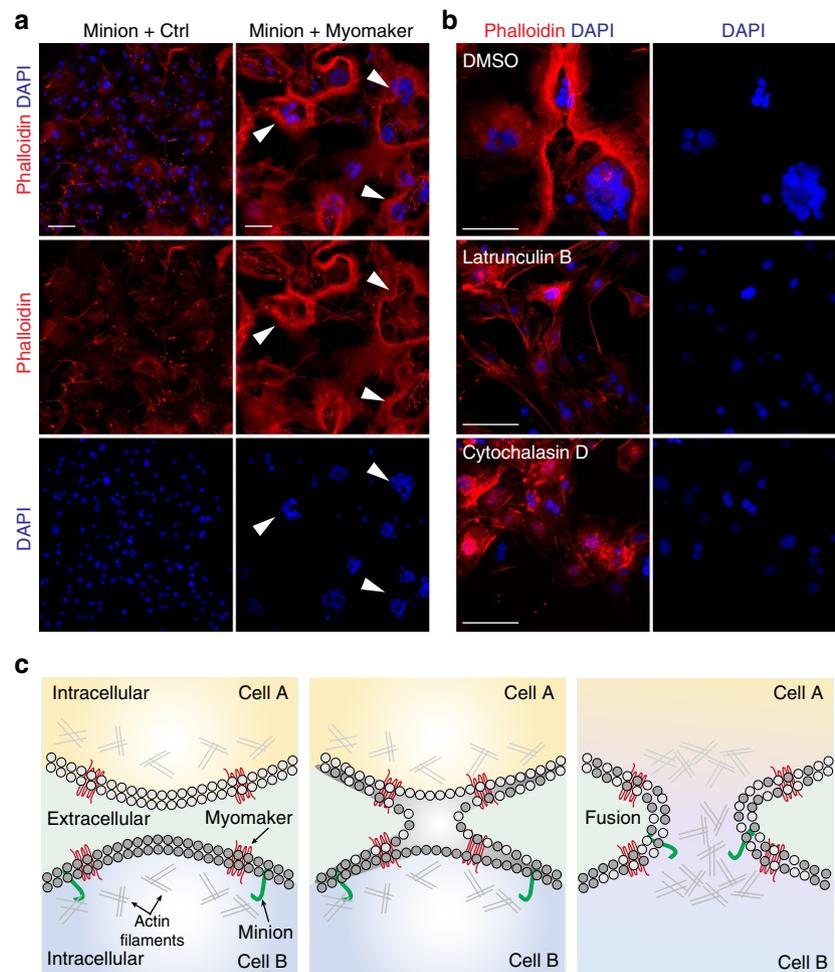
**Figure 7 | Minion and Myomaker are sufficient to induce fusion of non-muscle cells.** (a) Western blot of 10T1/2 fibroblasts expressing Luciferase (Ctrl), Myomaker or Minion.  $n = 2$ . (b) Retroviral vectors encoding Luciferase, Myomaker or Minion were transduced as indicated into 10T1/2 fibroblasts. All vectors contain IRES-GFP downstream of the gene of interest, causing infected cells to uniformly express GFP. Split-channel grayscale images for GFP and DNA are included.  $n = 3$  (eight technical replicates each). (c) Quantification of GFP<sup>+</sup> syncytia in fibroblasts expressing combinations of proteins as indicated. Fusion index was calculated as the percentage of nuclei found within GFP-positive syncytia containing  $\geq 3$  nuclei. Syncytia were scored 24 h after seeding. Each value represents mean  $\pm$  s.d.  $n = 3$  (two 1.4 mm  $\times$  1.4 mm fields each). Double asterisks:  $P < 0.001$ , unpaired two-tailed Student's  $t$ -test. (d) Fluorescence images from cell-mixing experiments using differentially labelled 10T1/2 fibroblasts. Cells were serially infected with retroviruses encoding the indicated combinations of Minion, Myomaker, or controls (label omitted for simplicity). CellTrace Violet (cyan) and CellTracker DeepRed (red) dyes were used for labelling. Yellow arrows indicate syncytia containing both DeepRed<sup>+</sup> cells and Violet<sup>+</sup> nuclei. White arrows indicate syncytia derived from DeepRed<sup>+</sup> cells only.  $n = 5$  (six technical replicates each). See Supplementary Fig. 19 for split-channel images. (e) Quantification of fusion in (d) (bottom panels), measured as percentage of DeepRed<sup>+</sup> syncytia ( $\geq 3$  nuclei) containing  $\geq 1$  Violet<sup>+</sup> nucleus. Each value represents mean  $\pm$  s.d.  $n = 4$  (six 1.4 mm  $\times$  1.4 mm fields each). NS, not significant; single asterisk:  $P < 0.05$ , unpaired two-tailed Student's  $t$ -test. Scale bars, 100  $\mu$ m (b) and 50  $\mu$ m (d).

(50 ng  $\mu$ l<sup>-1</sup>) were co-injected into the pronuclei of fertilized zygotes. Zygotes surviving the injection procedure were transferred into a single oviduct of pseudopregnant ICR recipient females (50–60 embryos/oviduct). Mice produced from injected embryos were genotyped and sequenced (see 'Assay for genome modification' below) to determine the presence of mutations within the genomic region of *Minion*. Mutant founder animals were then bred to C57BL/6J mice and offspring were analysed for germline transmission.

To generate *in vitro* transcribed Cas9 mRNA, a 10 bp spacer and the T7 promoter were added to the *Streptococcus pyogenes* Cas9 coding region by PCR amplification from a construct (pCR-Blunt II-TOPO-NLS-Cas9-NLS) made in-house, and the amplified gel-purified Cas9 PCR product was used as the

template for *in vitro* transcription using mMESSAGE mMACHINE T7 ULTRA kit (Thermo Scientific).

The gRNA sequences were designed to target mouse *gm7325/Minion* gene (NCBI Gene ID: 653016; Ensembl ID: ENSMUSG00000079471). To generate *in vitro* transcribed gRNA, two types of oligonucleotides were first synthesized (Oligonucleotide 1 and 2, see Supplementary Table 2; IDT). After oligonucleotide annealing and PCR amplification, a T7 promoter with an additional 'G' at the 5' end was added to the gRNA PCR product. The amplified gel-purified gRNA PCR product was used as the template for *in vitro* transcription using a MEGAscript T7 kit (Thermo Scientific). Both the Cas9 mRNA and the gRNAs were purified using a MEGAclear kit (Thermo Scientific) and eluted into



**Figure 8 | Minion and Myomaker-induced fusion requires cytoskeleton reorganization.** (a) Fluorescence images of 10T1/2 fibroblasts co-overexpressing Minion and Myomaker. F-actin (Alexa546-Phalloidin, red) and DAPI (blue) staining are shown. White arrowheads point to the boundaries of multinuclear cells.  $n = 2$  (six technical replicates each, five fields each). (b) Fluorescence images of 10T1/2 fibroblasts co-overexpressing Minion and Myomaker and treated for 24 h with DMSO control or the actin polymerization inhibitors latrunculin B ( $0.1 \mu\text{M}$ ) or cytochalasin D ( $0.3 \mu\text{M}$ )<sup>17</sup>.  $n = 2$  (six technical replicates each, five fields each). Scale bars,  $100 \mu\text{m}$  (a,b). (c) A proposed model for Minion and Myomaker-induced cell-cell fusion. We suggest that Minion and Myomaker have separable roles in the fusion process; Myomaker mediates pre-fusion pore events such as cell-cell recognition and/or, membrane apposition, whereas Minion mediates later fusion pore formation, at least in part via induction of cytoskeletal rearrangements.

RNase-free water. gRNA protospacer sequences targeting the mouse *Minion* ORF region are given in Supplementary Table 2. For generation of the 135 bp deletion allele, gRNAs 1 and 2 were used; for generation of the 155 bp deletion allele, gRNAs 3 and 4 were used.

**Assay for genome modification and genotyping.** To detect mutations in the mice arising from injected embryos, PCR was performed using primers flanking the targeted region (Seq-F and Seq-R) (Supplementary Table 2). The PCR products were used in a surveyor assay according to manufacturer's instructions (IDT). The PCR products were then analysed by agarose gel electrophoresis and selected products were cloned into a pCR-blunt-cloning vector and sequenced. Deep sequencing was also used to identify mutations. Mice containing mutations in the target region were bred to confirm germline transmission. Among the mouse strains with mutations, two with 135 and 155 bp deletions within the *Minion* ORF was selected for further analysis. Note that experimental results with the 155 bp deletion allele are shown in Supplementary Fig. 5, while the *Minion*<sup>Δ</sup> allele in all other figures represents the 135 bp deletion allele.

For genotyping of the subsequent progeny carrying the *Minion*<sup>Δ</sup> 135 bp deletion allele, a primer set flanking the targeted region (Fwd and Rev; see Supplementary Table 2) was used to amplify genomic DNA, generating a  $\sim 760$  bp product from the unmodified allele and a  $\sim 625$  bp product from the mutated allele respectively. For genotyping of the progeny carrying the 155 bp deletion allele, a different primer set was used to amplify genomic DNA (Fwd-2 and Rev-2; Supplementary Table 2), generating a  $\sim 555$  bp product from the unmodified allele and a  $\sim 400$  bp product from the mutant allele, respectively. The PCR products were separated by gel electrophoresis using 2% agarose (Sigma, A9539). Wild-type mice demonstrate a single band of the larger size, while homozygotes containing the deletion

demonstrate a single band of the smaller size, and heterozygous mice demonstrate both bands.

**Myoblast isolation.** For mouse embryonic myoblast isolation, pregnant C57BL/6J female mice were humanely euthanized, and embryos were rapidly but gently dissected and placed into dissection buffer containing Ham's F10 nutrient mix (Thermo Scientific, 11550043) with  $1 \times$  antibiotic-antimycotic (Thermo Scientific, 15240062). One pregnant female was euthanized for each experimental repeat, and embryos from the same litter were dissected and compared. For each embryo, the tail was kept in a numbered tube for genotyping, while all four limbs were skinned, dissected, and placed into a 2 ml numbered Eppendorf tube containing  $\sim 1.5$  ml dissection buffer including  $4 \text{ mg ml}^{-1}$  Collagenase type II (Worthington LS004176; freshly made and filtered before use). The samples were rotated on a platform rocker at 80 r.p.m. and  $37^\circ\text{C}$  for 30–45 min, until the muscles were mostly digested and only bones and soft tissues were left. In general, earlier stage embryos required shorter incubation times. Cell suspensions were checked microscopically after each step to avoid over-digestion. After allowing the unwanted tissues to settle at room temperature, the supernatant was transferred to a 50 ml conical tube. Dispase II (Thermo Scientific, 17105041) was then added at a final concentration of  $\geq 1.2 \text{ mg ml}^{-1}$  ( $> 0.6 \text{ U ml}^{-1}$ ) to both tubes: (1) for the supernatant in the 50 ml tubes, cells were incubated at  $37^\circ\text{C}$  for 20–30 min with occasional mixing; (2) for the remaining tissues in the 2 ml tubes, 1.5 ml freshly made and filtered dissection buffer with  $4 \text{ mg ml}^{-1}$  Collagenase type II and  $\geq 1.2 \text{ mg ml}^{-1}$  Dispase II, and the tubes were again rocked on the rocker at 80 r.p.m. and  $37^\circ\text{C}$  for 20–30 min to allow further digestion and dissociation.

After dissociation, the suspension in the 2 ml tube was mixed with that in the 50 ml tube. After adding  $\sim 4$ –5 volumes of wash medium (Ham's F10 and 10%

horse serum, filtered), the digested mixture was passed through a 10 ml 20 Gauge needle slowly and gently for ~four times, while scrupulously avoiding generation of bubbles. More wash medium was then added to bring the final volume to 30 ml. This suspension was filtered through a prewashed 40 µm Nylon Mesh filter on top of a new 50 ml conical tube, and the filter was rinsed with 10 ml wash medium into the same tube. All of the 50 ml tubes were then centrifuged at 125g for 5 min at room temperature, and the supernatant was transferred and spun down again at 125g for 5 min. The pellets from two centrifugations were resuspended and mixed in 2 ml myoblast isolation medium followed by an additional 20 ml of media containing a 1:1 mixture of DMEM low glucose (Gibco, 11885084) and Ham's F-10 Nutrient Mix (Gibco, 11550043); 20% (v/v) FBS; 1 × antibiotic-antimycotic; and freshly added 2.5 ng ml<sup>-1</sup> rhFGF (Promega, G5071). Medium lacking DMEM but containing the remaining items above also produced similar results.

The isolated cell mixture from each embryo was first plated into a regular 150 mm TC-treated dish for 30 min at 37 °C (preplate I) and replated into another 150 mm dish for 30 min at 37 °C (preplate II) to eliminate fibroblasts, and then the supernatant containing mostly myoblasts was transferred into two 100 mm collagen-coated dishes. Cells were examined the next day to determine necessity for passaging. Occasionally the preplate II dish also contained some amount of myoblasts, and these were kept and expanded in addition to those in the collagen dishes. 0.05% trypsin was used for dissociating the cells from dishes. After a few passages, 1 × antibiotic-antimycotic was replaced by 1 × penicillin-streptomycin (Gibco). As the fibroblast number decreases in culture, the embryonic myoblasts may start to proliferate very slowly and they should be seeded more densely to recover from the slow growth. Adult mouse myoblasts were isolated similarly but with a few modifications: the muscles were removed from bones and minced with small scissors; 15 ml conical tubes were used instead of 2 ml tubes, with twice the volume of digestion buffer; and longer digestion and dissociation times were used.

**In vitro myoblast differentiation assay.** For primary myoblasts derived from both embryos and adult mice, ~3,000 cells in 50 µl myoblast growth medium (1:1 mixture of DMEM low glucose and Ham's F-10 Nutrient Mix; 20% (v/v) FBS; freshly added 2.5 ng ml<sup>-1</sup> rhFGF) were seeded into each well of a 384-well Collagen-coated PerkinElmer CellCarrier plate (6007550) for imaging purposes. The next day, differentiation medium (DMEM high glucose (Gibco, 11995073) with 3–5% horse serum) was added to the cells (differentiation condition (DM) day 0). Differentiation medium was replaced daily. The cells were fixed at DM day 3 and day 4 for immunofluorescence staining. For C2C12 cells, ~1,500–2,000 cells were seeded into each well of a 384-well plate (DMEM high glucose with 10% FBS) and around 2 × 10<sup>5</sup> cells were seeded into each well of a six-well plate, using C2C12 growth medium (DMEM high glucose with 10% FBS). The following day, differentiation medium (DMEM high glucose with 2% horse serum) was added to the cells (DM day 0), and differentiation medium was subsequently replaced daily. The cells were collected or fixed at different time points as described.

**Histology.** For paraffin sections with embryonic samples, mouse embryos (E14.5 and later) were decapitated and the tails were collected in numbered tubes for genotyping. To enhance fixation in later-stage embryos (E17.5 and later), embryos were skinned in the area to be studied. Embryos were fixed overnight using 4% paraformaldehyde (PFA; Electron Microscopy Sciences #15714) in PBS at 4 °C with gentle rotation. Following two quick rinses with PBS, embryos were placed into 70% ethanol for dehydration and long-term storage. For the tissues/organs to be studied, the appropriate portions were cut and submitted to histology core for paraffin embedding and sectioning using routine protocols. Note that embryos from the same litter were used for each comparison.

For cryosections of adult mouse tissue, skeletal muscle samples were dissected and partially embedded in gum tragacanth (Sigma G1128; 10% w/v in PBS) on a wooden dowel, and frozen in 2-methylbutane in a glass beaker cooled on liquid nitrogen. The fresh frozen muscle samples were then sectioned at 10 µm thickness using a cryostat cooled to -20 °C. These fresh frozen muscle sections were then fixed in 1% PFA diluted in PBS at room temperature for 5 min before subsequent staining procedures. Both cryosections and paraffin sections were stained with Hematoxylin and Eosin (H&E) following routine protocols. Note that adult mice from the same litter were used for each comparison.

**Immunofluorescence staining on tissue sections.** For muscle cryosections, after fixation with 1% PFA/PBS as described above, slides were washed with PBS and permeabilized with 0.2% Triton X-100 diluted in PBS at room temperature for 10 min, and were then washed again with PBS. Sections were blocked at room temperature for 1 h using a freshly prepared and filtered solution containing 1% heat-inactivated donkey serum, 1% BSA, 0.025% Tween20 in PBS. After blocking, sections were incubated with primary antibody at 4 °C overnight, washed with PBS and then incubated with secondary antibody for 2 h at room temperature. After a 5 min wash with PBS, the sections were incubated with the nuclear stain DAPI (Molecular Probes D1306; 5 mg ml<sup>-1</sup> stock) at a 1:20,000 dilution in PBS for 5 min, and slides were mounted and sealed using ImmMount (Shandon) and glass coverslips. For paraffin sections with embryonic tissues, the deparaffinized and rehydrated slides were permeabilized with 0.2% Triton X-100 in PBS for 10 min and washed again with PBS. The sections were then blocked at room temperature

for 1 h using a freshly made and filtered solution containing 5% heat-inactivated normal goat serum in PBS. After blocking, similar procedures were performed as mentioned above for cryosections. Note that embryos from the same litter were used for each comparison.

Primary antibodies used for immunofluorescence were: Mouse anti-MHC (MY32 clone, Sigma M4276, 1:300 dilution on paraffin sections and 1:500 dilution on cryosections); Mouse anti-Desmin (D33 clone, DAKO M0760, 1:300 dilution); Sheep anti-Gm7325/Minion (R&D systems AF4580; 1:200 dilution of 1 mg ml<sup>-1</sup> stock on cryosections only); Sheep IgG was used as negative control. All secondary antibodies (Invitrogen Alexa-Fluor) were used at 1:250 dilution, and the host species was either donkey or goat. Only secondary antibodies from the same host species were used together for co-staining.

**Immunofluorescence staining with adherent cells.** For immunofluorescence staining with adherent cells, mainly C2C12 and primary myoblasts from adults and embryos, 384-well PerkinElmer CellCarrier plates were again used. Cells were fixed with 4% PFA in PBS for 8–10 min and quickly washed twice with PBS before permeabilization with 0.2% Triton X-100 in PBS for 10 min. After one wash with PBS, cells were blocked with freshly made and filtered 5% heat-inactivated normal goat serum (or 5% heat-inactivated normal donkey serum for experiments using sheep anti-Gm7325/Minion antibody) in PBS for 1 h, and were incubated with primary antibodies overnight at 4 °C. The next day, after two quick washes with PBS, cells were incubated with secondary antibodies for 1–2 h at room temperature. After three quick washes with PBS, the cells were incubated with DAPI (5 mg ml<sup>-1</sup> stock, 1:20,000 dilution in PBS) for 10 min. The 384-well plate was then imaged using either UltraVIEW confocal or ImageExpress Micro (IXM and IXC; Molecular Devices) confocal imaging systems (see the Microscopy part below).

Primary antibodies used were: Mouse anti-MHC (MY32 clone, Sigma M4276, 1:400 dilution); Mouse anti-Desmin (D33 clone, DAKO M0760, 1:300 dilution); Sheep anti-Gm7325/Minion (R&D systems AF4580; 1:200 dilution of 1 mg ml<sup>-1</sup>). All secondary antibodies (Invitrogen Alexa-Fluor) were used at 1:250 dilutions, and the host species was either donkey or goat. Only secondary antibodies from the same host species were used together for co-staining.

**Fluorescence staining with adherent cells.** For fluorescence staining of actin filaments in fibroblasts and myoblasts, the high-affinity F-actin probe Alexa Fluor 546-conjugated phalloidin (Invitrogen, A22283) was used according to manufacturer's instructions. Briefly, cells grown in 384-well plates were fixed with 4% PFA for 10 min at room temperature, washed with PBS, and permeabilized with 0.1% Triton X-100 for 5 min. After two PBS washes, the cells were blocked with PBS containing 1% BSA for 30 min and incubated with staining solution (1:80 dilution of the phalloidin methanolic stock in blocking solution) for 1 h at room temperature. After two to three quick PBS washes, the cells were incubated with DAPI (5 mg ml<sup>-1</sup> stock, 1:20,000 dilution in PBS) for 10 min. The plate was then imaged using either UltraVIEW confocal or IXM confocal imaging systems (see the Microscopy part below).

**Microscopy and imaging.** The Invitrogen EVOS FL Auto Imaging System was used for routine examination of immunofluorescence staining, GFP virus infection, and cell labelling. For imaging of histological and immunostained tissue sections on glass slides, the Hamamatsu NanoZoomer and Aperio VERSA scanners were used to obtain whole-slide images using a ×20 objective. For imaging of the immunofluorescence cell samples in 384-well plates, the IXM and IXC confocal high-content imaging systems were used with ×10 and ×20 objectives. To acquire higher resolution images for tissue sections and cell samples, the UltraVIEW VoX 3D live cell imaging system (PerkinElmer) spinning disk confocal microscope system was used with ×20, ×40 and ×60 objectives. All pictures of whole mouse embryos were taken using iPhone 5S in combination with Leica KL200 LED dissection microscope.

**Cell culture.** For culture of primary myoblasts isolated from later-stage mouse embryos and adult mice, filtered myoblast growth medium (1:1 mixture of DMEM low glucose and Ham's F-10 Nutrient Mix; 20% FBS) with freshly added 2.5 ng ml<sup>-1</sup> rhFGF was used. In general, around 2–4 × 10<sup>5</sup> cells were seeded into a 100 mm collagen-coated dish, and the cells were split once every 2–3 days at a ratio of 1:2 to 1:4, depending on proliferation speed. 0.05% trypsin was used for dissociating cells from dishes. Myoblasts typically went through a crisis period after the removal of most fibroblasts, and could be seeded more densely at this point. Primary myoblasts in culture were monitored every day with fresh medium replacement as needed. For the culture of immortalized C3H/C2C12 myoblast cells (ATCC), filtered C2C12 growth medium (DMEM high glucose with 10% FBS) was used. Approximately 1.5 × 10<sup>5</sup> cells were seeded into a 100 mm tissue culture-treated dish, and cells were split every 2 days. 0.25% trypsin was used for cell dissociation. The cells tested negative for mycoplasma contamination. For culture of the immortalized C3H/10T1/2 fibroblasts (ATCC), filtered fibroblast growth medium (DMEM high glucose with 15% FBS) was used. Approximately 1 × 10<sup>5</sup> cells were seeded into a 100 mm tissue culture-treated dish, and the cells were split once every 3 days. 0.25% trypsin was used for cell dissociation. The cells tested negative for mycoplasma contamination. For culture of the immortalized

RAW264.7 macrophage line (ATCC), filtered growth medium containing DMEM high glucose with 10% FBS was used. Around  $2\text{--}3 \times 10^6$  cells were seeded into a  $175\text{ cm}^2$  flask. The cells were split once every 2–3 days, when they were ~60–75% confluent. To ensure cell lifting and reduce cell death, 0.25% trypsin and a cell scraper were used in combination. To induce the formation of multinuclear osteoclast-like cells,  $50\text{ ng ml}^{-1}$  sRANKL (Peprotech, 174 aa) was incubated with the cells for 3 days. For the culture of C17 embryonic stem cells derived from 129 mice (kindly provided by Tom Gridley), freshly made and filtered growth medium was used, consisting of ESGRO Complete PLUS medium (Millipore, SF001-500P) with 15% FBS and three inhibitors: GSK3 $\beta$  inhibitor which comes with the medium; MEK inhibitor PD184352 ( $0.8\text{ }\mu\text{M}$  final) and FGFR inhibitor PD173074 ( $0.1\text{ }\mu\text{M}$  final). Normally the cells were co-cultured with mouse embryonic fibroblasts according to standard procedures, but for the purpose of RNA and protein isolation they were seeded onto gelatin-coated dishes without a fibroblast feeder layer. Approximately  $1 \times 10^6$  cells were seeded into each 100 mm dish. The medium was replaced every day. Cells were split once every two days at a ratio of 1:5 to 1:10, depending on experimental need. 0.05% trypsin was used for cell dissociation. All cell culture media contained 100 units per ml of penicillin and  $100\text{ }\mu\text{g ml}^{-1}$  of streptomycin, unless otherwise specified. None of the cell lines is listed in the database of commonly misidentified cell lines maintained by ICLAC.

For experiments shown in Fig. 8b, 10T1/2 cells of indicated genotypes were seeded into 384-well Perkin Elmer COC plates at 800, 1,600, 3,200 cells per well in fibroblast growth medium. After 15 h, the cells were incubated with growth media containing DMSO (0.003%), latrunculin B (100 nM) or cytochalasin D (300 nM) for 24 h before further analysis<sup>17</sup>.

**Tissue and cell lysates preparation for protein analysis.** Both embryonic and adult mouse tissue samples were weighed, snap-frozen in liquid nitrogen and stored at  $-80\text{ }^\circ\text{C}$  until use. For preparation of protein lysates, eight volumes of ice-cold lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 10% glycerol, with freshly added  $2 \times$  Halt protease inhibitor cocktail and  $1 \times$  Roche PhosSTOP phosphatase inhibitor cocktail) and one to two 3 mm tungsten carbide beads (Qiagen) were added to each sample in a 1.5 ml or 2 ml Eppendorf tube. These were then homogenized at 30 cycles per second for 3–8 min at  $4\text{ }^\circ\text{C}$  using a TissueLyser II. Detergents were then added to the lysates to a final concentration of 0.1% SDS, 0.1% sodium deoxycholate and 1% Triton X-100, and the samples were rotated at  $4\text{ }^\circ\text{C}$  for 2–4 h. Lysates were then transferred to new tubes and spun down at 15,000–21,000g for 10 min at  $4\text{ }^\circ\text{C}$ . For organs containing significant amount of lipids, the supernatant was transferred and spun down again at 15,000–21,000g for 10 min at  $4\text{ }^\circ\text{C}$ .

For cell samples, buffer from Alfa Aesar (J60423) was generally used (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 5% Glycerol, 0.1% SDS, 0.5% sodium deoxycholate and 1% Triton X-100, with the above mentioned protease and phosphatase inhibitor cocktails). Cells were quickly rinsed with DPBS and then  $\sim 300\text{ }\mu\text{l}$  ice-cold lysis buffer was added to each well of a six-well plate. After incubation on ice for 5 min, the cells were pipetted up and down and transferred to 1.5 ml Eppendorf tubes and incubated on ice for 30 min, with 1 s of vortexing every 10 min. The samples were then spun down at 15,000–21,000g for 10 min at  $4\text{ }^\circ\text{C}$ . A second buffer without ionic detergents was used for some of experiments (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 10% glycerol, 1 mM EDTA and 1% Triton X-100, with the above mentioned protease and phosphatase inhibitor cocktails), with similar results. Bio-Rad DC protein assay with BSA standard was performed to measure protein concentration of supernatants. Lysates were mixed with NuPAGE LDS Sample Buffer (NP0007) and dithiothreitol (100 mM final), and boiled at  $94\text{ }^\circ\text{C}$  for 10 min prior to SDS–polyacrylamide gel electrophoresis (PAGE).

**Subcellular fractionation analysis.** C2C12 cells incubated with differentiation medium for 4 days were used for the subcellular fractionation studies using the Qproteome Cell Compartment system (Qiagen, 37502) following the manufacturer's instructions. The cells were dissociated from the dishes first before addition of the first buffer; similar results were obtained with direct cell lysis on the plate. The cytosolic/membrane/nuclear/cytoskeletal fractions were extracted from the cells, and for each fraction, the protein lysate extracted from an equivalent number of cells was loaded for western blot (Supplementary Fig. 2e).

To examine whether Minion protein is secreted and soluble in cell conditioned media, C2C12 myoblasts with indicated genotypes were first differentiated for two days in differentiation medium containing 2% horse serum, and then incubated with serum-free differentiation medium consisting of DMEM with  $1 \times$  ITS-G (Thermo Scientific, 41400045) for 24 h. The supernatant conditioned media were collected, centrifuged at  $8\text{ }^\circ\text{C}$  to eliminate dead cells, and filtered using a  $0.45\text{ }\mu\text{m}$  vacuum filter bottle to further eliminate cell debris. The filtered supernatants were then concentrated at  $8\text{ }^\circ\text{C}$  using 3k Amicon ultracentrifugation filters for multiple rounds according to the manufacturer's instructions. The supernatants were concentrated  $\sim 400\text{--}1,000$  fold. As a comparison, whole cell extracts were prepared from the original cell pellet using the lysis buffer mentioned above, and denatured supernatants and whole cell extracts generated from equal amounts of cells were loaded for western blot analysis (Supplementary Fig. 2d).

**SDS-PAGE and western blots.** The NuPAGE Novex gel electrophoresis system was used for the separation of proteins. Approximately  $10\text{--}30\text{ }\mu\text{g}$  of cell lysate or  $30\text{--}60\text{ }\mu\text{g}$  of tissue lysate were loaded per well. NuPAGE MES SDS Running Buffer (NP0002) and 4–12% NuPAGE Novex Bis-Tris gels were used. Proteins were transferred to PVDF or nitrocellulose membranes using the iBlot transfer system (Thermo Scientific). Freshly prepared 5% milk in TBST (137 mM NaCl, 20 mM Tris, 0.1% Tween-20, pH7.6) was generally used as the blocking buffer with both PVDF and nitrocellulose membranes. However, for the detection of Minion protein using primary antibody raised in sheep, freshly made and filtered 10% donkey serum in TBST was used as the blocking buffer with PVDF membrane (Millipore Immobilon-PSQ,  $0.2\text{ }\mu\text{m}$  pore size). The information of primary and secondary antibodies used in western blots is listed in Supplementary Table 3). Two ECL substrates with different sensitivity were used as indicated. We found that the anti-human TMEM8C antibody recognized both endogenous and overexpressed Myomaker protein in both mouse primary muscle and cultured cell lysates (Figs 6 and 7a, Supplementary Fig. 18), but required extended antibody incubation and exposure times. Note: All uncropped western blots shown in main figures have been included in Supplementary Figs 22–25.

**RNA preparation and RT-qPCR.** Total RNA was isolated from cell lines using TRIzol Reagent (Thermo Scientific, 15596026) according to the manufacturer's instructions. First-strand cDNA synthesis was performed using qScript cDNA SuperMix (Quanta BioSciences) according to manufacturer's instructions. For PCR, cDNA from  $\sim 5\text{ ng}$  RNA was used in a  $12.5\text{ }\mu\text{l}$  reaction with Power SYBR Green PCR Master Mix (Thermo Scientific, 4367659). Reactions with RNA only were prepared as negative controls. An Applied Biosystems 7900HT thermocycler was used with the following primers (Supplementary Table 2): *Minion*-F and *Minion*-R; *Gapdh*-F and *Gapdh*-R. Relative quantification was performed using the comparative CT method. The CT value of *Minion* gene was normalized to that of the reference gene *gapdh* in the same sample using the formula:  $2^{\Delta\Delta\text{CT}}$ .

**Lung flotation assay.** E18.5 embryos from *Minion* <sup>$\Delta/\Delta$</sup>   $\times$  *Minion* <sup>$\Delta/\Delta$</sup>  intercrosses were quickly isolated by caesarean section from humanely sacrificed pregnant females, and were placed on dry Kimwipes. To maintain body temperature, these newborns were incubated by hand and subsequently in a  $37\text{ }^\circ\text{C}$  chamber. Pups were exposed to normal room air following delivery, and were monitored for at least 1 h. *Minion* <sup>$\Delta/\Delta$</sup>  newborns were uniformly atonic, apneic, and became cyanotic almost immediately after delivery. The majority of *Minion* <sup>$\Delta/\Delta$</sup>  and *Minion* <sup>$\Delta/\Delta$</sup>  mice exhibited normal breathing and demonstrated pink body colour indicative of adequate ventilation and perfusion. After at least 1 h of air breathing, pups were anaesthetized, weighed, tailed for genotyping, decapitated, and the lungs were dissected and placed into PBS in 15 ml conical tubes or 2 ml Eppendorf tubes for flotation assay. The lungs were then monitored for more than 15 min, after which they were scored as either floating or sinking. Approximately 60 E18.5 embryos were examined.

**Plasmids and cloning.** For the cloning of shRNA constructs, 19–21 nucleotide target sequences were selected using both BLOCK-iT RNAi Designer (Thermo Scientific) and in-house optimized algorithms. For the mouse *Minion* mRNA transcripts (GenBank Accession No. NM\_001177468.1, NM\_001177469.1 and NM\_001177470.1), four shRNA target sequences were chosen initially to target all the transcripts, two targeting the coding sequence and two targeting the 3' UTR. A control sequence was used targeting the firefly (*Photinus pyralis*) luciferase gene, which exists in the pGL3 luciferase reporter vector but which lacks similar sequence in the mouse transcriptome. For each shRNA, two 55–59 nt oligonucleotides were designed as shown in Supplementary Table 4 and synthesized (IDT). The oligonucleotides, each containing sense and antisense target sequences, a 9 nt intervening hairpin loop, and TTTG at the 5' ends with GATC at 3' ends for cohesive-end cloning, were annealed. These were then ligated with BbsI/SpeI-digested pGWL-si2/U6 vector. Subsequently using Gateway LR Clonase II Enzyme Mix, these shRNA cassettes were cloned into the vector pLentiLox3.7-GW (pLL3.7-GW), a third generation lentiviral gateway vector that expresses shRNAs under the mouse U6 promoter. A CMV-EGFP reporter cassette was included in the vector to monitor expression.

For the cloning of cDNA constructs, the coding sequences (CDS) of mouse *Minion* (84 aa isoform; CDS and protein accession numbers CCDS50119.1 and NP\_001170939.1), human *MINION* orthologue (CDS and protein accession numbers CCDS83093.1 and NP\_001302423.1), mouse *tmem8c* (*Myomaker*; CDS and protein accession numbers CCDS15823.1 and NP\_079652.1), *nanoluc* (control A), *luciferase* (control B), mouse *Minion* with 1 bp frameshift insertion, human *MINION* with 1 bp frameshift deletion, and mouse *Minion* truncation (39 aa form) were synthesized (IDT) with attB sites at both ends and a consensus Kozak sequence (5' GCCACC) before the start codon. For mouse and human *MINION* CDS, both untagged and C-terminally  $3 \times$  FLAG-1  $\times$  HA-tagged (3F1H) versions were generated. The human *MINION* CDS used here was based on prior genome assembly, whereas the most recent assembly has a C-to-T change and the resulting protein sequence has a R (Arginine) to C (Cysteine) change at the 16th residue. For the *nanoluc* CDS, a C-terminal  $1 \times$  FLAG tag was added. Using Gateway BP Clonase II Enzyme mix (Thermo Scientific, 11789020), the synthesized DNA

sequences were cloned into the pDONR221 vector and the sequence-confirmed entry vectors were subsequently cloned into the pCIGAR gateway retroviral vector using Gateway LR Clonase II Enzyme mix. The pCIGAR vector is an MSCV-based bicistronic retroviral vector modified to permit Gateway-mediated insertional recombination of transgenes immediately upstream of IRES-eGFP. In addition, the empty pCIGAR vector (MCS version) was used as another control vector (control C). After expression testing of the two 3F1H-tagged mouse and human *MINION* constructs, it was noticed that two bands were detected by western blot with anti-FLAG antibody (Supplementary Fig. 14). This reflected the presence of an extra start codon in-frame with the actual start codon, which was confirmed by sequencing to be inherited from the 3' end of the *CMV* promoter in the pCIGAR vector, giving rise to a product with an additional N-terminal 16 aa, indicated by an asterisk. As a 6-nt consensus Kozak sequence was added before the actual start codon, the intended product is the dominant protein expressed. All other cDNA vectors contained an extra T nucleotide immediately prior to the Kozak sequence, to avoid upstream translational initiation. In addition, a pLKO-TREX-ON lentiviral MCS vector (kindly provided by Feng Cong, Novartis Institutes for BioMedical Research) was used for making Minion expression vector containing C-terminally 3F1H-tagged mouse Minion, used for the AP-MS experiments.

**Lentiviral shRNA knockdown assay in myoblasts.** Lentiviral particles were produced in HEK 293T cells (ATCC) using a third generation lentiviral packaging system and FuGENE 6 transfection reagent (Promega). Fresh medium was replaced one day after transfection and the supernatant medium was collected on the following day. The medium was briefly centrifuged to remove dead cells, and neat virus was used for QC infection on 293T cells by the reverse infection method with  $8 \mu\text{g ml}^{-1}$  polybrene (overnight incubation without spin infection). Analysis of GFP expression by FACS 3 days after infection generally demonstrated a titre of  $\sim 1 \times 10^6 \text{ vp ml}^{-1}$ . Neat virus was further concentrated  $\sim 100$ -fold using a 100 kDa centrifugal filter unit (Amicon), aliquoted and stored at  $-80^\circ\text{C}$ .

To examine the knockdown efficiency of the mouse *Minion* transcript, shRNA-encoding lentiviruses were used to infect C2C12 cells. Viruses were diluted in growth medium containing  $8 \mu\text{g ml}^{-1}$  polybrene, and after a brief incubation at  $37^\circ\text{C}$  with C2C12 cells, one round of spin infection was performed at 1,100g and  $32^\circ\text{C}$  for 1–1.5 h, using either 24-well or 12-well plates. Based on an estimated viral titre of  $1 \times 10^8 \text{ vp ml}^{-1}$  after concentration (as measured on 293T cells), a virus amount equivalent to MOI30 on 293T cells was used on C2C12 cells. Fresh medium was replaced the next day and GFP<sup>+</sup> cells were sorted by FACS two days later (see gating strategy in Supplementary Fig. 26). Infection efficiency of 70–85% was routinely achieved as judged by FACS analysis, though GFP signal in undifferentiated C2C12 cells was faint by EVOS fluorescence microscopy. Sorted, GFP<sup>+</sup> C2C12 cells were recovered, expanded and seeded into six-well plates for *in vitro* differentiation assay. At DM day 4–6, cell lysates were collected and the expression of Minion protein was examined by western blot. The two shRNA constructs targeting the *Minion* 3' UTR (U1/U2) were found to reduce Minion expression most efficiently (Supplementary Fig. 11b), and for subsequent experiments in C2C12 and primary myoblasts, cells were infected with U1 and U2 shRNA viruses only. To produce the *Minion*<sup>KD</sup> C2C12 cells, the U1-infected, GFP<sup>+</sup>-sorted cells were reinfected with U2 shRNA virus and resorted for increased GFP signal (top 30–50%). Similarly, wild-type C2C12 cells were infected with the control shGL3 virus (Ctrl) in two rounds and sorted twice to generate the *Ctrl* cells. These *Minion*<sup>KD</sup> and *Ctrl* C2C12 cells were expanded and used for *in vitro* differentiation assay in 384-well and six-well plates. The *Minion*<sup>KD</sup> cells were also used as the background for *in vitro* reconstitution assay with cDNA-containing retroviruses.

Similar infection steps were followed to knockdown *Minion* in primary myoblasts derived from adult mice, except that  $6 \mu\text{g ml}^{-1}$  polybrene was used and fresh medium lacking polybrene was added by the end of the day instead of on the next day, as primary myoblasts appeared more sensitive to polybrene treatment. Due to the difficulty in expanding these infected and sorted primary cells, one round of virus infection was performed using either *Ctrl* virus or 1:1 ratio mixture of U1 and U2 viruses, and the cells were sorted by GFP signal after 2–3 days. Approximately one-third or one-half the amount of virus used on C2C12 cells was used on primary myoblasts. The sorted primary myoblasts were recovered for 3 days and then seeded into 384-well plates for *in vitro* differentiation assay.

**Retrovirus infection of C2C12 and 10T1/2 cells.** The cDNA-encoding retroviruses were made by co-transfecting 293T cells with pCIGAR retroviral vectors and pCL-Eco or pCL-10A1 packaging vectors using FuGENE 6 transfection reagent (Promega). Media was replaced one day after transfection and the supernatant medium was collected on the following day. After a brief centrifugation to eliminate dead cells, the neat virus was used for QC infection in 384-well plate by reverse infection method of NIH-3T3 cells (for pCL-Eco packaged virus) or 293T cells (for pCL-10A1 packaged virus) with  $8 \mu\text{g ml}^{-1}$  polybrene (overnight incubation without spin infection). GFP expression was analysed by FACS 3 days after infection, usually yielding a titre of  $0.3\text{--}1.2 \times 10^6 \text{ vp ml}^{-1}$ .

To infect C2C12 cells and 10T1/2 cells, neat virus was incubated with  $8 \mu\text{g ml}^{-1}$  polybrene at  $37^\circ\text{C}$  for 10 min and then added to the cells, which were then incubated for 15 min at  $37^\circ\text{C}$ . One round of spin infection was performed using 24-well or 12-well plates at 1,100g and  $32^\circ\text{C}$  for 1–1.5 h. After 4–6 h, fresh

neat virus with polybrene was added and a second round of spin infection was performed. Fresh medium was added to the infected cells to dilute the polybrene. Media was replaced the following day, and 2–3 days later, the infected cells usually exhibited very strong GFP fluorescence by EVOS microscopy. Infection efficiency using this two-round infection method was generally  $>95\%$  over a large range of viral concentrations, therefore FACS was generally not necessary. The resulting GFP signal was used to mark the boundary of cells, in addition to phase contrast imaging and nuclear markers (Supplementary Figs 7b,19 and 20); moreover, in cells that were fixed with 4% PFA but not permeabilized with detergents, the GFP signal was higher in the nuclear region, allowing unambiguous delineation of both the cell boundary and nuclei.

For each round of infection, a viral amount equivalent to MOI 3–6 on 3T3 or 293T cells was used for *Minion*<sup>KD</sup> C2C12 cells, and the infected cells were expanded and used for *in vitro* reconstitution assay in differentiation medium. For co-expression and cell-mixing experiments on wild-type 10T1/2 cells and C2C12 cells, a viral amount equivalent to MOI 2–4 on 3T3 or 293T cells was used. For each comparison, the viruses made with the same packaging plasmid and at similar MOI were used for infection. The experiments were repeated with different types of control viruses (Luciferase, NanoLuc-FLAG, empty vector). For co-expression and mixing experiments, cells were infected with one type of cDNA retrovirus first using the two-round spin infection protocol described above, expanded for several days, then reinfected with a second cDNA-encoding retrovirus using the same method. On the day following the final infection, the cells were labelled with different dyes, mixed and seeded into 384-well or 24-well plates as described below.

**Affinity purification and mass spectrometry analysis.** C2C12 myoblasts expressing both C-terminally  $3 \times \text{FLAG-1} \times \text{HA}$ -tagged mouse Minion and empty control were seeded in growth medium, then incubated in differentiation medium for 3.5 days with  $2 \mu\text{g ml}^{-1}$  Doxycycline to induce the expression of exogenous Minion protein from pLKO TREX-ON lentiviral cDNA vector. These cells were then used in one-step anti-FLAG immunoprecipitation. Cytoplasmic and membrane fractions were prepared using the ProteoExtract Native Membrane Protein Extraction Kit (Millipore, 444810). Cytosolic and membrane fractions were subjected to affinity purification using anti-FLAG M2 affinity gel (Sigma, F2426) with 1 h incubation, and the immunopurified protein complexes were eluted by the addition of FLAG peptide (Sigma, F4799). The eluate was precipitated by trichloroacetic acid and washed twice with acetone. The precipitates were dissolved in digestion buffer (8 M Urea in 100 mM Tris pH 8.5), reduced, alkylated and digested as described previously<sup>34</sup>. The digested peptide mixture was desalted using C18 pipette tips (Thermo Scientific) and fractionated online using a 75  $\mu\text{m}$  inner diameter fused silica capillary column, with a 5  $\mu\text{m}$  pulled electrospray tip and packed in-house with 16 cm of C18 reversed phase spherical silica particles of 3.0  $\mu\text{m}$ .

An Ultimate 3,000 nano LC system (Thermo Scientific) coupled to a Q Exactive mass spectrometer (Thermo Scientific) was used to deliver the linear acetonitrile gradient using buffer A (0.1% formic acid water) and buffer B (0.1% formic acid water, 100% ACN) starting from 5% buffer B to 35% over 70 min at a flow rate of  $200 \text{ nl min}^{-1}$ , followed by a 5 min ramping to 95% acetonitrile and a 5 min hold at 95% buffer B. The column was re-equilibrated with 2% buffer B for 2 min before the subsequent run. Each sample was analysed in triplicate technical replicates on LC-MS/MS.

MS/MS spectra were collected on a Q-Exactive mass spectrometer (Thermo Scientific). All MS/MS raw spectra files were processed using Proteome Discoverer 1.4 (Thermo Scientific), and the spectra from each raw file were converted to MASCOT generic files (MGF) for data set searches using licensed Mascot Daemon (Matrix Science, London, UK; version 2.4.1). The MS/MS data was searched against the Uniprot database downloaded on April 1st, 2015 (selected for *Mus musculus* with 16715 entries) including a decoy database. The carbamidomethylation of cysteines was set as a static modification with a mass of +57.02156 Da. The specificity of digestion was set for trypsin allowing three missed cleavages. The mass tolerances in MS and MS/MS mode were set to 15 p.p.m. and 0.8 Da, respectively. Peptide and protein identifications were validated using Scaffold version 4.6.2 (Proteome Software Inc., Portland, OR). Peptide and protein identifications were accepted if they were detected with  $\geq 95.0\%$  and  $\geq 99.0\%$  probabilities respectively by the Scaffold local false discovery rate algorithm, and required at least two identifiable unique peptides<sup>35</sup>. Fisher's exact test *P* value in Scaffold were calculated according to a model previously described<sup>36</sup>. The protein classes of significantly enriched proteins were identified and visualized using the Panther classification system (<http://www.pantherdb.org>)<sup>37</sup>.

**Cell labelling and mixing experiments.** The performance of a series of cell-permeant fluorescent dyes was tested on 10T1/2 fibroblasts and C2C12 myoblasts over multiple dilutions and using different labelling methods, with signal strength and pattern monitored continuously by microscopy for at least four days. CellTracker Deep Red dye (Thermo Scientific C34565, 1:250 final dilution) and CellTrace Violet dye (Thermo Scientific C34571, 1:500 final dilution) were selected for subsequent experiments, and were observed to label the cytoplasmic and nuclear regions of mononuclear cells. However, in fused multinuclear cells (Fig. 7d, Supplementary Fig. 19), the CellTrace Violet dye exhibited a strong enrichment in nuclei which had been originally labelled with the dye, and did not diffuse into

other non-violet-labelled nuclei, while the CellTracker Deep Red dye demonstrated perinuclear enrichment and was helpful in recognizing the cell boundary. These features allowed facile quantification of fusion efficiency (Fig. 7e). For experiments shown in Supplementary Fig. 20, Vybrant DiD (Thermo Scientific V22887, 1:200 dilution) was used together with Violet dye per manufacturer's instructions. DiD staining appears spotty in proliferating cells but is retained homogeneously in fused cells.

Cell labelling was performed according to manufacturer's instructions with slight modifications. Cells were trypsinized, centrifuged, washed once with PBS, transferred to a 2 ml Eppendorf tube, centrifuged, and resuspended in PBS at a concentration of  $0.8-1 \times 10^6$  cells/ml.  $2 \times$  cell labelling solution was prepared separately in PBS with fluorescent dyes and was mixed well by brief vortexing. Equal volumes of the  $2 \times$  labelling solution and the cell suspension (usually 250  $\mu$ l each) were mixed and incubated at 37 °C for 40–45 min with occasional mixing. The labelled cells were then mixed with 1 ml fresh medium, incubated at 37 °C for 5 min and centrifuged at 150g. At this point, 3 more washes of the labelled cells were performed using fresh medium, incubation at 37 °C for 10–15 min, and centrifugation at 150g. After the last wash, the cells were counted and diluted to the concentration needed for the final mixing experiment in a 384-well or 24-well plate, and were incubated at 37 °C for 30 min. These additional washes and incubation steps were used to eliminate remaining unbound dye in the cell suspension and on the cell surface, which was critical for cell-mixing experiments performed on the same day. After incubation, cells labelled with Deep Red dye were mixed with those labelled with Violet dye at a 1:1 ratio, and were seeded into 384-well, with a range of concentrations tested from 800 to 4,000 cells per well. Cells in 384-well plates were fixed at different time points (24–48 h) with 4% PFA for 10 min at room temperature, and were washed with PBS before imaging. Growth medium containing 10–15% FBS was used for most experiments in both 10T1/2 and C2C12 cells.

For cell-mixing experiments in Fig. 6a,b, cells were mixed at 1:2 ratio (one fibroblast and two myoblasts) and the cells were incubated in differentiation medium for 3–4 days before fixation. 3,000–6,000 cells per well were used in 384-well plates. For experiments in Fig. 6a, no dye was used, with 10T1/2 fibroblasts GFP positive and differentiating myoblasts MHC positive. For experiments in Supplementary Fig. 6b, all cells expressed GFP Violet dye was therefore used to label fibroblasts, with differentiating myoblasts marked by MHC. For myoblast mixing experiments in Supplementary Fig. 20b, cells were labelled, mixed, and then incubated in DM for 4 days before fixation.

**Differentiation and fusion indices.** Several indices were used to examine the fusion efficiency in C2C12 cells as well as adult and embryonic primary myoblasts during *in vitro* differentiation. The differentiation index shown in C2C12 (Fig. 5e) and adult primary myoblasts (Supplementary Fig. 13b) was calculated as the fraction of nuclei contained within all MHC<sup>+</sup> cells, including both mononuclear and multinuclear cells, as compared with the number of total nuclei within each  $\times 20$  image acquired by IXM confocal high-content imaging. At least six separate fields from independent replicate wells were quantified for each genotype in each experimental repeat. The fusion index shown in C2C12 (Fig. 5f) and adult primary myoblasts (Supplementary Fig. 13c) was calculated as the fraction of nuclei contained within MHC<sup>+</sup> myotubes which had three or more nuclei, as compared to the number of total nuclei within each  $\times 20$  image taken by IXM imaging. At least six separate fields from independent replicate wells were quantified for each genotype. The fusion index shown for embryonic primary myoblasts (Fig. 5b) was calculated as the fraction of nuclei contained within Desmin<sup>+</sup> myotubes having three or more nuclei, as compared to the number of nuclei within all Desmin<sup>+</sup> cells on each  $\times 20$  IXM image. At least six separate fields from independent replicate wells were quantified for each genotype in each experimental repeat. Since fibroblasts still existed in these early-passage primary cultures, only the total nuclei in Desmin<sup>+</sup> cells were included for quantification. Another index used to examine the fusion efficiency in C2C12 cells (Fig. 5g) and adult primary myoblasts (Supplementary Fig. 13d) was focused on myotubes. Myotubes were binned into three to four subgroups based on the number of nuclei contained within each tube: 2 nuclei, 3–5 nuclei, 6–10 nuclei or more than 10 nuclei (Fig. 5g). The fraction of each subgroup was calculated for each genotype in comparison to total myotube number. At least six separate fields from independent replicate wells were quantified for each genotype. Please find more details about the numbers in the figure legends.

For co-expression experiments in 10T1/2 cells, two indices were used to examine fusion efficiency. The first fusion index (Fig. 7c) was based on the experiment shown in Supplementary Fig. 7b, and was calculated as the percentage of nuclei within GFP<sup>+</sup> syncytia containing three or more nuclei, as compared with the number of total nuclei. At least four separate  $\times 10$  IXM images from different replicate wells were used for quantification of each genotype. To examine the ability of Minion-only or Myomaker-only expressing cells to fuse with Minion and Myomaker co-expressing cells as well as to further confirm the formation of multinuclear syncytia by fusion instead of incomplete cytokinesis, a second index was used (Fig. 7e) to calculate the fusion efficiency between Deep Red dye-labelled cells and Violet dye-labelled cells in three different combinations (Fig. 7d, Supplementary Fig. 19). In these combinations, all of the Deep Red dye-labelled cells had co-expression of Myomaker and Minion, no matter which protein was

expressed first by infection, and these cells were able to fuse to themselves, becoming Deep Red<sup>+</sup> syncytia. The Violet dye-labelled cells expressed (1) Myomaker and Minion (either Myomaker or Minion was expressed first); (2) Myomaker only (together with empty vector control); (3) Minion only (together with Luciferase control). The fraction of Deep Red<sup>+</sup> syncytia (containing three or more nuclei) containing one or more Violet dye-labelled nucleus was compared with the total number of Deep Red<sup>+</sup> syncytia on each  $\times 10$  IXM image. At least 6–12 separate fields from independent replicate wells for each experimental repeat were quantified for each combination.

**Statistical analysis.** Sample size was not predetermined statistically and no specific blinding method or randomization was applied. Each value reported represents the mean  $\pm$  s.d. of more than two-independent replicates as described for each experiment. Replicate types utilized include: independent experimental (biological) replicates; individually treated cells with viruses or compounds; mice from the same or different litters; and tests or assays run on the same sample multiple times (technical replicates). Quantitative data were analysed by unpaired two-tailed Student's *t*-test with Welch's correction and without assumption of equal standard deviations. All statistical analyses were performed using GraphPad Prism software.  $P < 0.05$  was considered statistically significant.  $P < 0.05$  was marked with an asterisk (\*) and  $P < 0.001$  was marked with double asterisks (\*\*).

**Data availability.** Source data for proteomic studies is provided in Supplementary Table 1. Raw proteomic data are also available on request. The RNA sequencing data have been submitted to the NCBI Gene Expression Omnibus (GEO) (under accession number GSE97764). The mouse *Minion* and human *MINION* coding sequences together with protein sequences have been deposited into GenBank under accession number KY857878 and KY857877.

## References

- Ma, J. *et al.* Improved identification and analysis of small open reading frame encoded polypeptides. *Anal. Chem.* **88**, 3967–3975 (2016).
- Saghatelian, A. & Couso, J. P. Discovery and characterization of smORF-encoded bioactive micropeptides. *Nat. Chem. Biol.* **11**, 909–916 (2015).
- Anderson, D. M. *et al.* A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* **160**, 595–606 (2015).
- Anderson, D. M. *et al.* Widespread control of calcium signaling by a family of SERCA-inhibiting micropeptides. *Sci. Signal.* **9**, ra119 (2016).
- Magny, E. G. *et al.* Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. *Science* **341**, 1116–1120 (2013).
- Matsumoto, A. *et al.* mTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide. *Nature* **541**, 228–232 (2017).
- Nelson, B. R. *et al.* A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science* **351**, 271–275 (2016).
- Buckingham, M. & Rigby, P. W. Gene regulatory networks and transcriptional mechanisms that control myogenesis. *Dev. Cell* **28**, 225–238 (2014).
- Comai, G. & Tajbakhsh, S. Molecular and cellular regulation of skeletal myogenesis. *Curr. Top. Dev. Biol.* **110**, 1–73 (2014).
- Hindi, S. M., Tajrishi, M. M. & Kumar, A. Signaling mechanisms in mammalian myoblast fusion. *Sci. Signal.* **6**, re2 (2013).
- Hochreiter-Hufford, A. E. *et al.* Phosphatidyserine receptor BAI1 and apoptotic cells as new promoters of myoblast fusion. *Nature* **497**, 263–267 (2013).
- Millay, D. P. *et al.* Myomaker is a membrane activator of myoblast fusion and muscle formation. *Nature* **499**, 301–305 (2013).
- Millay, D. P., Sutherland, L. B., Bassel-Duby, R. & Olson, E. N. Myomaker is essential for muscle regeneration. *Genes Dev.* **28**, 1641–1646 (2014).
- Chen, Y. M., Du, Z. W. & Yao, Z. Molecular cloning and functional analysis of ESGP, an embryonic stem cell and germ cell specific protein. *Acta Biochim. Biophys. Sin. (Shanghai)* **37**, 789–796 (2005).
- Sloan, C. A. *et al.* ENCODE data at the ENCODE portal. *Nucleic Acids Res.* **44**, D726–D732 (2016).
- Duan, R. & Gallagher, P. J. Dependence of myoblast fusion on a cortical actin wall and nonmuscle myosin IIA. *Dev. Biol.* **325**, 374–385 (2009).
- Bothe, I. & Baylies, M. K. Drosophila myogenesis. *Curr. Biol.* **26**, R786–R791 (2016).
- Chen, E. H. & Olson, E. N. Unveiling the mechanisms of cell-cell fusion. *Science* **308**, 369–373 (2005).
- Kim, J. H., Jin, P., Duan, R. & Chen, E. H. Mechanisms of myoblast fusion during muscle development. *Curr. Opin. Genet. Dev.* **32**, 162–170 (2015).
- Krauss, R. S. Evolutionary conservation in myoblast fusion. *Nat. Genet.* **39**, 704–705 (2007).
- Nowak, S. J., Nahirney, P. C., Hadjantonakis, A. K. & Baylies, M. K. Nap1-mediated actin remodeling is essential for mammalian myoblast fusion. *J. Cell Sci.* **122**, 3282–3293 (2009).

22. Shilagardi, K. *et al.* Actin-propelled invasive membrane protrusions promote fusogenic protein engagement during cell-cell fusion. *Science* **340**, 359–363 (2013).
23. Bi, P. *et al.* Control of muscle formation by the fusogenic micropeptide myomixer. *Science* **356**, 323–327 (2017).
24. Quinn, M. E. *et al.* Myomerger induces fusions of non-fusogenic cells and is required for skeletal muscle development. *Nat. Commun.* **8**, 15665 (2017).
25. Bae, G. U. *et al.* Regulation of myoblast motility and fusion by the CXCR4-associated sialomucin, CD164. *J. Biol. Chem.* **283**, 8301–8309 (2008).
26. Nakamura, T. *et al.* Antibody-targeted cell fusion. *Nat. Biotechnol.* **22**, 331–336 (2004).
27. Avigan, D., Rosenblatt, J. & Kufe, D. Dendritic/tumor fusion cells as cancer vaccines. *Semin. Oncol.* **39**, 287–295 (2012).
28. Long, M. A. & Rossi, F. M. Targeted cell fusion facilitates stable heterokaryon generation *in vitro* and *in vivo*. *PLoS ONE* **6**, e26381 (2011).
29. Dittmar, T. & Zanker, K. S. Cell fusion in health and disease. Volume II: cell fusion in disease. Introduction. *Adv. Exp. Med. Biol.* **714**, 1–3 (2011).
30. Blau, H. M. *et al.* Plasticity of the differentiated state. *Science* **230**, 758–766 (1985).
31. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
32. Mohun, T. *et al.* Deciphering the Mechanisms of Developmental Disorders (DMDD): a new programme for phenotyping embryonic lethal mice. *Dis. Model. Mech.* **6**, 562–566 (2013).
33. Petryszak, R. *et al.* Expression Atlas update—a database of gene and transcript expression from microarray- and sequencing-based functional genomics experiments. *Nucleic Acids Res.* **42**, D926–D932 (2014).
34. Wohlschlegel, J. A. Identification of SUMO-conjugated proteins and their SUMO attachment sites using proteomic mass spectrometry. *Methods Mol. Biol.* **497**, 33–49 (2009).
35. Nesvizhskii, A. I., Keller, A., Kolker, E. & Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**, 4646–4658 (2003).
36. Zhang, B. *et al.* Detecting differential and correlated protein expression in label-free shotgun proteomics. *J. Proteome Res.* **5**, 2909–2918 (2006).
37. Mi, H. *et al.* PANTHER version 11: expanded annotation data from gene ontology and reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.* **45**, D183–D189 (2017).

## Acknowledgements

We gratefully acknowledge Qiang Zhou, Cynthia Cienfuegos, and Jacqueline Avis for reverse genetics support; Brian Schwartz and Robbin Newlin for histology support; Fabio Luna, Whitney Barnes and John Walker for RNA sequencing; Vladimir Trifonov for bioinformatic assistance; Doug Quackenbush for assistance with microscopy; Minhua Qiu for image analysis; Albert Parker for advice regarding molecular biology; Daniel Mason for assistance with mass spectrometry; Seung-Hyun Woo for advice regarding

lung flotation assay; Jason Matzen, Erik Spedale and Paul Calvin for assistance with compounds; Bin Fang for software assistance; and Melanie Tucker, Mary Frazer, Diego Guzman and Richard Eddins for animal husbandry support. We thank Richard Glynne and Peter McNamara for leadership support, and David Glass, Estelle Trifilieff, Richard Glynne, Peter McNamara, John Walker and Rajesh Nair for critical reading of the manuscript.

## Author contributions

Q.Z., Sn.C.S. and Sh.C.S. designed the experiments, analysed the data and wrote the paper; Q.Z. performed most experiments; A.A.V. performed AP-MS analysis; J.O. provided experimental protocols; S.Y.C. helped with experiments and data analysis; R.M. performed embryonic histological sectioning; A.R. and L.M. prepared lentiviruses and retroviruses; J.Z. and E.D. assisted with molecular biology; C.S. participated in experimental design and data analysis and oversaw all reverse genetics.

## Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing interests:** All listed authors are employees of the Novartis Institutes for BioMedical Research.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Zhang, Q. *et al.* The microprotein Minion controls cell fusion and muscle formation. *Nat. Commun.* **8**, 15664 doi: 10.1038/ncomms15664 (2017).

**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 4.0 International License, which permits use, sharing, adaptation, 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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017