New muscle identity code components identified by TRAP control shape and size of muscle subsets in *Drosophila*

Benjamin Bertin, Yoan Renaud, Teresa Jagla, Guillaume Lavergne, Cristiana Dondi, Jean-Philippe Da Ponte, Guillaume Junion* and Krzysztof Jagla*

GReD - INSERM U1103, CNRS UMR6293, Université of Clermont Auvergne,
28, place Henri-Dunant, 63000 Clermont-Ferrand, France

* Corresponding authors
christophe.jagla@uca.fr
guillaume.junion@uca.fr

Keywords: *Drosophila*, muscle diversification, TRAP, identity code, *Gelsolin*, *dCryAB*

Abstract

How the diversification of cell types is regulated during development to generate stereotyped tissue patterns remains a challenging question. Here we applied cell type specific translational profiling (TRAP) approach to identify gene expression signatures underlying the diversification of two muscle subsets in *Drosophila.* Targeting the Slou-positive muscle population, we identified *Ptx1* as a new component of the ventral muscle identity code. We also generated temporal transition profiles of TRAP-enriched genes and found that a set of genes encoding regulators of the actin cytoskeleton fitted a cluster specific for lateral transverse (LT) muscles. Two of them, *dCryAB* and *Gelsolin (Gel),* encoding actin-binding sHsp and actin-severing protein respectively, displayed LT muscle prevailing expression positively regulated by LT identity factors Lms and Apterous. In *dCryAB*-mutant embryos, LT muscle shapes had an irregular appearance, whereas *Gel*-devoid LTs occasionally developed muscle fibre splitting. Importantly, split LT muscles in *Gel* mutants were enlarged and bore a greater number of myonuclei, a phenotype phenocopied by overexpression of the *duf* fusion gene. Muscle fibre splitting is a hallmark of diseased muscle in Duchenne and other muscular dystrophies, but how muscles get split is unknown. Our analyses yield evidence that an excessive myoblast fusion could result in splitting, and suggest a deregulation of fusion machinery in dystrophic muscles. We identify new muscle identity gene *Ptx1* and a new class of identity realisator genes including *Gel*, whose function links muscle identity in *Drosophila* to dystrophic muscle splitting phenotype in humans.

Significance statement

Key features that make two cells differ from each other are shape and size. Here, by applying translational profiling of muscle subsets in *Drosophila*, we identify new muscle identity realisator genes *dCryAB* and *Gel* encoding actin-binding sHsp and actin-severing protein, respectively. They both regulate growth-related properties of a subset of embryonic muscles, the LT muscles. Intriguingly, *Gel*-devoid LTs undergo muscle fiber splitting, a phenotype also observed in Duchenne muscular dystrophy patients. LT muscle splitting in *Gel* mutants is due to an excessive myoblast fusion resulting in increased muscle size, offering clues to still unknown mechanisms of pathological muscle splitting in humans.

Introduction

Diversification of cell types is a fundamental process during development of multicellular organisms and is essential for building functional organs. Skeletal muscle, the major component of the animal body, diversifies into muscle types with different individual properties supporting distinct functions, such as locomotion or breathing. However, understanding the genetic and molecular mechanisms that confer these individual characteristics to each muscle during development remains a major challenge. The muscle network in *Drosophila* embryos offers an attractive system for studying cell diversification. It is composed of 30 muscle fibers per abdominal hemisegment, which are easy to detect and to follow during development. Despite common characteristics such as formation by myoblast fusion and capacity to contract, each muscle displays a specific size, orientation, number of nuclei, attachment and innervation (1). However, how these biological processes are regulated at the muscle-specific level remains elusive, although it is generally accepted (2,3) that the muscle founder cells (FCs) that are at the origin of muscle fibers harbour all the information required for individual muscle identity. This information is thought to be provided by the combinatorial expression of genes encoding transcription factors called the identity genes or iTFs (4,5). To date, twenty iTFs that control muscle identity have been discovered in *Drosophila*. They are all expressed in subsets of muscle progenitors or founders and share one important feature: their loss of function leads to the loss or aberrant formation of a muscle in which they are expressed (1). Some identity genes such as *ladybird (lb)* (6) *or lateral muscles scarcer (lms)* (7) show a restricted expression in lateral muscles only, whereas others like *slouch (slou)* or *Krüppel (Kr)* display a broader expression pattern in ventral, lateral and dorsal domains (8,9). Among recently identified iTFs, *optomotor-blind-related gene (Org-1)* is required for the specification of SBM, LO1 and VT1 muscles (10), *araucan/caupolican* for lateral muscles (11) and *tailup* for four dorsal muscles (12). Interestingly, all these factors play conserved roles during cranial musculature specification in vertebrates.

Knowledge gained on iTF gene function in *Drosophila* could help to understand how diversification processes are regulated in general. One important aspect that remains to be investigated is the gene network that operates downstream of iTFs to set individual muscle properties. Whole genome approaches based on ChIP experiments have shown that iTFs directly regulate not only other identity genes, but also downstream regulators of muscle identity, the "realisator genes", which control cell size, shape, attachment or innervation (13,14). *Eve*, *Lb* and *Slou* control the size of individual muscles by regulating the number of fusion events. They achieve this by setting expression levels of three genes (*mp20, Pax* and *mspo*) that regulate actin dynamics and cell adhesion in a muscle-specific manner (15,16). However, our knowledge about the genetic control of cell diversification during development is essentially based on functional analyses of individual genes and on genomic analyses on whole embryos at a given developmental point.

To analyse diversification processes further we optimised translating ribosome affinity purification (TRAP) (17) to small subsets of FCs and developing muscle precursors (18). This method identifies messenger RNA being translated in a cell-specific manner. We performed TRAP at three time windows targeting two restricted muscle subsets corresponding to Slou- and Lms-positive cells and the entire muscle cell population expressing Dumbfounded (Duf). TRAP-purified mRNA profiling followed by bioinformatics analyses identified new candidate genes with potential roles in setting the identity of Lms-

and Slou-positive muscles, and among them a new conserved iTF, Ptx1 contributing to the identity code of ventral muscles. Temporal transition profiles were also generated for each muscle subset, revealing a cluster of genes encoding actin binding/interacting proteins enriched in the Lms-muscle subset. Among them *Gelsolin (Gel)* and *dCryAB* displayed LT muscle dominant expression, and when knocked out via CRISPRs led to uncoordinated LT muscle growth, irregular muscle shapes and in the case of *Gel,* muscle splitting. The split muscle phenotype occurs in dystrophic muscles (refs) and as we demonstrate here could be due to a modified muscle identity program resulting in excessive myoblast fusion.

Our genome-wide transcriptional profiling of muscle subsets identified new players in muscle identity including a conserved iTF, Ptx1 and a new class of realisator genes controlling growth-related LT muscle properties.

Results and Discussion

Translational profiling of rare populations of embryonic muscle cells using TRAP.

TRAP was first developed to isolate polysome-associated mRNA from a subset of neurons in mice (17) and was later adapted to other model organisms including *Xenopus* (19), zebrafish (20) and *Drosophila* (18, 21). It is based on the polysome capture of the GFP-tagged ribosomes with their associated mRNAs (Fig. 1A). Tagging of ribosomes in the cell type of interest occurs *in vivo* by incorporation into the large ribosomal subunit of the RpL10A fused to GFP. Here we made use of the GAL4/UAS system to drive the expression of RpL10A-GFP in two sub-populations of muscle cells, DT1, LO1, VO1, VO2, VT1 and VA3 using Slou-GAL4, and LT1, LT2, LT3 and LT4 using Lms-GAL4 driver (Fig. 1B) or in all embryonic muscles with Duf-GAL4. For each muscle subset, translational profiling was performed on embryos collected from three developmental time windows T1: 7–10 h AEL, T2: 10–13 h AEL and T3: 13–16 h AEL covering the main muscle development steps. Gene expression profiling with total embryonic RNA (input fractions) as a reference was used to identify differential gene expression and perform spatial and temporal gene clustering (Fig. 1C and Fig. S1). Discrete clusters of genes upregulated in all time frames in Lms-only or in Slou-only populations were detected (Fig. 1C and Fig. S1). A significant portion of upregulated genes (FC > 2, $p < 0.05$) turned out to be common to the restricted muscle populations (Slou- and Lms-positive) and the global (Duf-positive) population (Fig. 1D). The percentage of Slou-specific transcripts remained constant at different time points, while the proportion of Lms-specific transcripts increased (Fig. 1D). This difference could be because Slou-expressing muscles are more heterogeneous than Lms-positive ones that harbor an LT muscle-specific translation program. To assess the specificity of TRAP-based muscle targeting we analysed GO terms and found that the upregulated genes fitted muscle-related GOs including muscle cell differentiation and myofibril assembly (Fig. 1E). By contrast, GO categories associated with the list of downregulated genes were not related to muscle developmental processes. Finally, to test muscle type-specific gene expression, we generated Volcano plots in Slou- and Lms- TRAP datasets (Fig. 1F) and observed that genes with previously characterised expression patterns in muscle subsets displayed expected up- or down-regulation. For example, in Lms-positive muscles *ap* and *lms* gene transcripts are enriched, whereas *slou* and *org1* transcripts, specific for Slou-positive muscles, are depleted (Fig. 1F). We also tested whether TRAP would detect “low expression” genes. To do so we crossed our lists of enriched and depleted transcripts with modEncode datasets (22) and found that more than 30% of upregulated TRAP-ed genes corresponded to “low expression” modEncode category,

whereas most depleted transcripts fitted the "high expression" modEncode category (Fig. S2).

TRAP-based translational profiling of muscle subsets was thus specific for the targeted muscle populations and sensitive enough to detect low transcript levels.

Ptx1, a new muscle identity gene identified by its differential expression in Slou-positive muscle subsets

Muscle diversification is tightly regulated by the combinatorial expression of transcription factors known as muscle identity code (1,5,16). To identify new components of muscle identity code we generated a full list of enriched genes (by pooling three time windows) for Slou and Lms muscle population (Fig. S3A and Table S1). Gene clustering according to level of enrichment was then performed, revealing four distinct gene clusters, two with genes exclusively upregulated in Slou or in Lms populations (orange and blue in the heatmap) and two others with genes displaying differential expressions in these two populations (Fig. S3A). Several Slou- and Lms-enriched genes code for the evolutionarily conserved DNA or RNA binding proteins (Fig. S3B). Those enriched in Slou-positive muscles include *Ptx1* (Fig. 2A) expressed in a subpopulation of ventral muscle precursors (23) (Fig. 2B-D), but with still unknown muscle identity functions. Interestingly, the mouse *Ptx1* counterpart is known to control hindlimb myogenesis (24) suggesting it could be a part of conserved muscle diversification programs. To test whether *Drosophila Ptx1* could contribute to muscle identity code we first analysed whether Slou and Ptx1 were coexpressed. At embryonic stages 14 and 15, a transient Ptx1 expression was detected in Slou-positive VA3 muscle precursor (Fig. 2C). Starting from early stage 13, both *Ptx1* transcripts and Ptx1 protein were prominently expressed in ventral VL4 and VO6 muscle precursors (23) (Fig. 2B-D). The main muscle phenotypes observed in *Ptx1* null mutant embryos generated by CRISPR mutagenesis (Fig. 2E) were VO6 defects (29% of segments) and VO6 loss (22% of segments) (Fig. 2F,G). Though with a lower penetrance, defects in VA3 were also detected (13% of segments) (Fig. 2F,G). In general, if affected, the VA3 muscles did not recognise their ventralmost insertion site shared with the VO4 muscle from the anterior segment (Fig. 2F). By contrast, loss of *Ptx1* had no influence on VL4. Thus *Ptx1* contributes to the specification and identity acquisition of VO6 and to the setting of the growth-related properties of VA3. Overall, our data indicate that *Ptx1* is part of the identity code for ventral muscles in *Drosophila* and a new conserved player in the diversification of muscle types.

Temporal transition profiling of TRAP datasets identifies _dCryAB_ and _Gelsolin_ expressed specifically in LT muscles

Temporal transition profiling identifies clusters of genes showing similar dynamics of expression patterns, thus potentially under common upstream regulatory cues. We considered that this approach could be applied to identify novel muscle identity realisator genes acting downstream of iTFs (15). The generated temporal transition heatmap for Slou- and Lms-positive muscles (see Materials and Methods for details) revealed clusters of genes with several gene expression behaviours (Fig. 3A Fig. S4).

We observed "up-up", "down-down", "up-down" and "down-up" transition profiles, either common to both muscle subsets or specific to one of them (Fig. 3A). Here we focused on gene clusters showing two contrasting transition behaviours, "down-down" (green) and "up-up" (red). To the red cluster belongs genes whose expression progressively increases from

early to late developmental time points. These genes could potentially contribute to all major identity acquisition steps including muscle-specific fusion, growth, attachment and innervation. By contrast, the green cluster is expected to group genes whose expression needs to be turned down while muscle differentiation progresses. The red cluster is larger and more heterogenous than the green cluster and comprises genes with behaviour that could be different for Slou- versus Lms- muscle populations. To characterise genes from these two clusters we first performed comparative GO analyses using Cluster Profiler (see also Cluster Profiler GO analysis of all transition profiles – Fig. S5). Red cluster genes showed a significant enrichment of several GO "molecular function" terms related to muscle development, whereas green cluster genes were characterised by enrichment of GOs associated with "DNA binding" and "chromatin binding" GOs, neither of which are specific to myogenesis (Fig. 3B). Among the genes progressively down-regulated in both muscle subsets (Fig. S4 for the list of genes) is *twist* and the Notch pathway-involved genes *bib*, *phyl*, *E(spl)m4-BFM* and *E(spl)mgamma-HLH* whose down-regulation is required for myogenic differentiation to occur. Examples of green cluster genes thus indicate that temporal transition profiling could help to identify genes acting in the same pathway. As stated, the cluster 2 genes are more heterogenous. Among them, conserved genes *CG34417/Smoothelin, flr/WDR1, CG18135/Gpcpd1, sals/SCAF1, Actn/ACTN2, Gel/GSN, TpnC41C/CALM1, dCryAB/l(2)efl/CRYAB* encoding actin-binding proteins display individual transition sub-profiles that are different in Lms- and Slou-muscle populations (Fig. 3C). We found this sub-cluster of particular interest as it contained genes with similar biological functions and differential Lms- versus Slou- profiles. Interestingly, *Gel* and *dCryAB* with "up-up" transition profile in the Lms subpopulation and "down-moderately up" in Slou muscles (Fig. 3C) are both preferentially expressed in Lms-positive LT muscles (Fig. 3D). *Drosophila Gel* belongs to the conserved gelsolin/villin family of actin interactors (25) with actin depolymerisation activity (26, 27), whereas *dCryAB* codes for a small heat shock protein (sHSP) carrying an actin-binding domain and known to exert chaperon functions (28). However, embryonic expression and functions of *Gel* and *dCryAB* have not yet been analysed. Here we show that in body wall muscles, *Gel* transcripts appear at late stage 14 exclusively in four LT muscle precursors (Fig. 3D) and at late embryonic stages are also detected in DT1 (Fig. S6A). Moreover, *Gel* is prominently expressed in the developing fat body and in the visceral musculature (Fig. S6B). *dCryAB* transcripts accumulate preferentially in LTs and in DT1 starting from early embryonic stage 15 (Fig. 3D), but at a lower level could also be detected in an additional set of muscles including LO1, SBM, VA2, VL1, VT1 and VO4-6 (Fig. 3D). These transcript distributions correlate well with Gel and dCryAB protein expression both showing granular cytoplasmic patterns with accumulation at LT muscle extremities and some dCryAB protein spots aligned with LT muscle cell membrane (Fig. S7). LT muscle specification is under the control of identity gene *Msh* and downstream iTFs, Apterous and Lms (7). As *lms* is exclusively expressed in LTs and in *lms* mutant embryos LTs are only partially affected (7) we wondered whether *lms* is regulating *Gel* and *dCryAB* expression in LTs. In an *lms* loss-of-function context *Gel* and *dCryAB* expression was strongly reduced in LTs and absent when probed in double *ap/lms* mutant embryos (Fig. S8), indicating that *Gel* and *dCryAB* act downstream of LT iTFs Lms and Ap.

***Gelsolin* and *dCryAB* control acquisition of growth-related LT muscle properties**

To test *Gel* and *dCryAB* functions we generated null alleles for both genes using CRISPR-Cas9 mutagenesis (Fig. 4A, B). *dCryAB* loss-of-function turned out to be homozygous

lethal whereas *Gel* null mutants survived until adult stage. Compared to wild-type (Fig. 4C), embryos devoid of *dCryAB* (Fig. 4D) showed irregular, asynchronous growth of LTs (28% segments), dissociation between LT1 and LT2 and/or LT2 and LT3 (32% of segments) and less frequently LT loss (6% of segments) or forked LT muscle extremities (8% of segments) (see Fig. 4F,G for quantification). Similar irregularity in LT growth (32% of segments) together with dissociation phenotypes (14% of segments) are observed in *Gel* mutants (Fig. 4D, F, G for quantifications). In contrast to *dCryAB*, we did not, in *Gel* mutant embryos, observe any loss of LTs or any forked termini (Fig. 4F,G). However, *Gel*-devoid LTs displayed muscle fibre splitting (17% of segments), a phenotype not observed in the *dCryAB* loss-of-function context (Fig. 4D-G).

In third instar larvae, dCryAB is expressed in all body wall muscles in a striated sarcomeric pattern and as previously shown by proteomics analyses (28), it interacts with filamin/cheerio and with actin to ensure sarcomere stability. Hence in growing embryonic LTs dCryAB might play an analogous, stabilising role in actin dynamics, controlling coordinated bi-polar LT extension. Consistent with a *dCryAB* role in muscle-growth tuning, muscles with forked extremities have been previously observed in larvae with muscle-driven *dCryAB* knockdown (28).

Like for *dCryAB, Gel* loss of function results in irregularly growing LTs (Fig. 4E). However, in a portion of *Gel* mutant embryos we also observed an additional phenotype, LT muscle splitting (Fig. 4E). This is a rather low penetrance phenotype considering the percentage of segments (17%) affected, and it most probably corresponds to a rare 5 LTs per segment phenotype that we previously reported in *lms* and *ap* LT-identity mutant embryos (7 see Fig. 5E,G).

It is well known (29, 30), that during pupal stages, splitting of persisting larval muscles precedes the formation of dorsal longitudinal muscles (DLMs), a subset of flight muscles. In this particular case, splitting requires interactions of larval muscles with imaginal myoblasts that fuse with them to produce DLMs (29) and the function of myogenic master gene Twist (30). However, how muscle splitting is regulated at cell and molecular levels remains unknown. Because split muscle fibres are also marks of muscular dystrophies, including Duchenne muscular dystrophy (31, 32), characterising LT muscle splitting observed in *Gel* mutants could provide new clues for understanding this process.

Splitting of LT muscles in *Gel* mutants results from excessive myoblast fusion

To follow LT muscle morphogenesis in *Gel* mutant embryos we recombined *Lms-Gal4;UAS-lifeActGFP (Lms>LAGFP)* LT sensor line (Fig. 5A) with *Gel 9.3.* We first confirmed that the LT splitting phenotype could be detected both on fixed (Fig. 5B) and in live preparations (Fig. S9) in *Gel* mutant embryos using *Lms>LAGFP* (Fig. 5B), and observed that split LTs were enlarged compared to non-split neighbours. Taking into account our previous findings that muscle size depends on the number of fusion events (15, 33) we wondered whether LT splitting could be associated with increased fusion. The number of Mef2-positive LT-nuclei counted in LT1-LT4 at embryonic stage 16 was significantly higher in *Gel9.3* segments with splitting events, ranging from 23 to 30 (Fig. 5B, E) compared to 18 to 21 nuclei in control *Lms>LAGFP* LTs (Fig. 5A, E). We tested whether LT-targeted increase in fusion could result in splitting and found that by overexpressing fusion gene *Dumbfounded* (*Duf*) in LTs we could mimic the *Gel* muscle splitting phenotype (Fig. 5C, E). We also found that *Lms>LAGFP* LTs had similar numbers of myonuclei to LTs in wild-type (w^{1118}) (Fig. 5E). Thus muscle splitting observed in *Gel* loss of function embryos results from excessive fusion,

with late fusion events that could be detected associated with LTs showing split phenotype (Fig. 5D). Importantly, split LTs are separated by muscle cell membrane (Fig. 5 F,G) and their extremities accumulate bPS-integrin (Fig. 5H), indicating that LT identity information ensuring choice of attachment sites is transmitted to the additional LT fibre generated by splitting.
These findings indicate the existence of a muscle identity-dependent size control system that causes muscle fibres to split when the number of fusion events is excessive with respect to the identity-predicted fusion program. *Gel* appears to be part of such a size control system acting to limit the number of fusion events according to LT muscle identity information.

Our genome-wide translational profiling of muscle subsets identified spatial and temporal gene expression signatures underlying diversification of muscle types. Among genes whose transcripts are enriched in Slou-positive muscles we identified a new iTF, Ptx1, that is evolutionarily conserved, and that in *Drosophila* embryos contributes to the identity code of ventral muscles and in particular VO6 and VA3. Temporal transition profiles in Lms- and Slou-muscle subsets defined clusters of genes with similar expression behaviour over time, among which two new muscle identity realisator genes *Gel* and *dCryAB* acting as LT muscle growth tuners. *Gel* encodes actin-severing protein and *dCryAB* a small heat shock protein (28) with potential actin polymerisation inhibiting activity (34). They both act downstream of LT muscle identity genes *ap* and *Lms* and ensure coordinated, polarised LT muscle growth. In addition, *Gel* plays a role in LT muscle size control by limiting the number of fusion events. Consequently, supernumerary myonuclei are present in *Gel*-devoid LTs which develop split muscle fibres. Our data therefore suggest that muscle splitting observed in muscular dystrophies including DMD (31) could be due to aberrant, excessive myoblast fusion in regenerating dystrophic muscles, consistent with centrally located nuclei observed in split DMD muscle fibres (32). How *Gel* acts to limit the number of fusion events in LTs remains to be investigated, but by depolymerising cortical F-actin, Gel might reduce the tension required for fusion (35) and thereby trigger fusion arrest.

Acknowledgements

This work was supported by AFM-Téléthon (MyoNeurAlp Strategic Program), Agence Nationale de la Recherche (Tefor Infrastructure Grant) and Fondation pour la Recherche Médicale (Equipe FRM Award). We thank the Genecore EMBL Heidelberg staff for their technical assistance.

Materials and methods

Fly Stocks

All *D. melanogaster* stocks and crosses were grown on standard medium at 25°C. The following strains were used: Ptx1-LacZ (gift of G. Vorbrüggen, Max-Planck-Institut, Germany), ap^{UGO35} (gift of J. Botas, Baylor, Houston, USA), lms^{s95} (gift of D. Müller, Universität Erlangen-Nürnberg, Germany), UAS-Lifeact-GFP (Bloomington, 35544) and *UAS-Dumbfounded* (gift of M. Ruiz-Gomez, Spanish National Research Council, Spain). To generate the TRAP lines, the *UAS-RpL10a-EGFP* line was crossed with the *Slou-Gal4* (gift of M. Frasch, Universität Erlangen-Nürnberg, Germany), *Lms-Gal4* (Janelia Farm collection, 46861) and *Duf-Gal4* (gift of K. Vijayraghavan, TIFR, India) driver lines to specifically target polysomes in the respective muscle populations.

Generation of *Ptx1*, *gel and dCryAB* knock-out lines by CRISPR-Cas9

The different gRNAs were designed using CRISPR optimal target finder (Gratz et al., 2014). To generate *gel* guide we used 5'-GTCGAGACCTCGACCGATGAGGC-3' and 5'-AAACGCCTCATCGGTCGAGGTCTC-3' pair of primers. For *dCryAB* the following pair of primers was designed: 5'- CTTCGACCGAAGTGCTGGTCCAAG-3' and 5'-AAAC CTTGGACCAGCACTTCGGTC-3'. For *Ptx1* 5'-TGCAGGTCAGGCCACCCGGTCCCA-3' and 5'- AAACTGGGACCGGGTGGCCTGACC-3'. Primers were annealed, digested by Bbs1 and cloned into pCFD3 plasmid (plasmid #49410, Addgene) for *gel* and into pCFD5 (plasmid #73914) for *Ptx1*. For *dCryAB* a HDR-based CRISPR technology was applied. First, two guides targeting 5'-CTTGGACCAGCACTTCGGTC-3' and 5'-GGAGGACAACGCCAAGAAGG-3' sequences located in 5' and in 3' regions of *dCryAB* gene, respectively were designed and cloned by Gibson assembly into pCFD4 plasmid (#49411). The homology arms HA1 of 1045bp and HA2 of 1058bp were amplified using the following pairs of primers:
Forward HA1: 5'-ATATCACCTGCATATTCGCAGCGACGTCATCTCTTTCGTCTG-3'
Reverse HA1: 5'-ATATCACCTGCATATCTACAAGAGGCGCGAGGTGCGCATTG-3'
Forward HA2 : 5'-ATATGCTCTTCATATAGGTGGAGACCTCCACCGCC-3'
Reverse HA2 : 5'-ATATGCTCTTCAGACTTCGTCAGGTTCGGTTACTCCG-3' into AarI and SapI cloning sites of donor pHD-DsRed plasmid (#51434). Plasmids were injected in Nos-Cas9 embryos (BestGene).

To establish KO lines, molecular characterization of target loci was performed as described (36). Briefly, genomic DNA was extracted from individual larvae by crushing them in 20 µL QuickExtarct Solution (Cambio) and releasing the DNA in a thermomixer according to the supplier's instructions. We used 1 µL (previously diluted five times) of the supernatant in 25-µL PCR reactions. PCR products were then sequenced by Sanger. Indels can be observed as regions with double peaks in heterozygous flies, corresponding to wild type and mutated allele respectively. In the case of *dCryAB*, homologous recombination events were recovered by selecting flies with red eye fluorescence.

RNA extraction and RT-qPCR

mRNA was extracted from Lms, Slou and Duf muscle populations using TRIzol reagent (Invitrogen) following the manufacturer's instructions. RNA quality and quantity were then assessed using Agilent RNA 6000 Pico kit on Agilent 2100 Bioanalyzer (Agilent Technologies).

Microarray analysis
Agilent 8x60K probe (60-mer) gene expression microarrays were used. We assessed the quality of triplicates using Pearson's correlation test. Correlations for all conditions were 85% or higher. The microarray data were quantile-quantile normalised. Gene expression data from Slou-, Lms- and Duf-positive cells were compared to the whole embryo datasets to generate lists of genes differentially expressed, fold change ≥ 2, $p < 0.05$. GO Princeton software was used to assign GO classification. We then compared Lms- and Slou-positive cells to Duf-expressing cells to make two lists of differentially regulated muscle-specific genes, fold change ≥ 2, $p < 0.05$. We computed and compared GO biological processes from these two lists using an R package cluster profiler.

Spatial heatmap. From lists generated by the comparison of Slou- and Lms-positive cells to the Duf population, we compared fold change of every selected gene of both populations over time and kept only those that were significantly different (ratio > 1.5). We distinguished two states for each population: (i) genes only up-regulated in one population and (ii) genes up-regulated in the two populations but displaying a significant difference between one population and the other. A heatmap was generated by bringing together these genes from the different time points and by keeping gene proportions for each case. We did not consider genes present in different states over time.

Temporal transition heatmap. Translational temporal profiles from Slou and Lms at the three time points were converted to "transition values" defined as log ratios between T2 and T1, T3 and T2. Transition values were considered as three discrete classes: upregulated (>1), stable (between −1 and 1), and downregulated (<1). Thus expression profiles from Lms and Slou muscles, which contain three temporal windows, were converted into vectors of two transitions (Tr1, Tr2), which allow the determination of correct gene behaviour. For example, the profile "red-red" groups genes whose RNA level increases between T2 and T1 (Tr1), and then continues to increase between T3 and T2 (Tr2).

***In situ* hybridisation and immunostaining**.
Embryos were dechorionated and fixed in 4% paraformaldehyde/heptane for all immunohistochemistry. Fluorescent *in situ* hybridisation with a TSA amplification system (Perkin-Elmer) and immunohistochemistry was as described previously (Junion et al., 2007). To generate RNA probes for *gel* (primers used: 5'-5'AATCGACTCCGTGGTGACTC-3' and 5'-GGGAGGCCAAAGATGAGCTGTC-3'), for *Ptx1* (5'-CTCGGTGAACTGCTTCAACA-3' and 5'-AACCGAGATCGGAGTGGTC-3') the corresponding DNA sequences were cloned by PCR in pCR II topo vector. The corresponding anti-sense RNAs were transcribed *in vitro* using T7 or SP6 RNA polymerase. For *dCryAB*, Gold collection clone *GH01960* was used to generate RNA probes. For fluorescent staining, the following antibodies were used: rabbit anti-β3 tubulin (1:5000; R. Renkawitz-Pohl, Philipps University, Marburg, Germany), rabbit anti-Ptx1 (1:500, G. Vorbrüggen, Max-Planck-Institut, Germany) guinea pig anti-Stripe (1:1000; T. Volk, Weizmann Institute of Science, Israel), rat anti-Actin (1:300, MAC 237; Babraham Bioscience Technologies), rabbit anti-Mef2 (Nguyen HT, 1:2000) anti-GFP (1:1000 Developmental Studies Hybridoma Bank (DSHB)), mouse anti-integrin βPS (1:200 DSHB) and mouse anti-connectin (1:200 DSHB). Cy3, Cy5, and 488 conjugated secondary antibodies were used (1:300; Jackson Immuno-Research). Embryos were mounted in anti-fade Fluoromount-G reagent (Southern Biotech). Labelled embryos were analysed using an SP8 (Leica) equipped

with HyD detector, with a 40X objective. Images were processed with ImageJ.

TRAP experiment

RPL10aGFP-tagged embryos were collected, and messenger RNAs from the different muscle populations isolated as described (18).

Live imaging

Lms-Gal4; UAS-lifeAct double transgenic line has been generated and used for time lapse imaging of LT muscle formation in *gel* mutant context. Image acquisition was performed on manually aligned living embryos at 21°C using inverted Leica SP8 confocal microscope. Time interval between acquisions was set up to 3 min and the acquisition time was 3 to 4 hr. Movies were generated and analysed using Imaris software (Bitplane).

References

1. Baylies MK, Bate M, Gomez MR (1998) Myogenesis: A View from Drosophila. *Cell* 93(6):921–927.
2. Rushton E, Drysdale R, Abmayr SM, Michelson AM, Bate M (1995) Mutations in a novel gene, myoblast city, provide evidence in support of the founder cell hypothesis for Drosophila muscle development. *Development* 121:1979–1988.
3. Landgraf M, Bossing T, Technau GM, Bate M (1997) The Origin, Location, and Projections of the Embryonic Abdominal Motorneurons of *Drosophila*. *J Neurosci* 17(24):9642–9655.
4. Carmena A, Gisselbrecht S, Harrison J, Jimenez F, Michelson AM (1998) Combinatorial signaling codes for the progressive determination of cell fates in the Drosophila embryonic mesoderm. *Genes Dev* 12(24):3910–3922.
5. Tixier V, Bataillé L, Jagla K (2010) Diversification of muscle types: Recent insights from Drosophila. *Exp Cell Res* 316(18):3019–3027.
6. Jagla T, et al. (1998) ladybird determines cell fate decisions during diversification of Drosophila somatic muscles. *Development* 125:3699–3708.
7. Müller D, et al. (2010) Regulation and Functions of the lms Homeobox Gene during Development of Embryonic Lateral Transverse Muscles and Direct Flight Muscles in Drosophila. *PLoS ONE* 5(12):e14323.
8. D'Alessio M, Frasch M (1996) msh may play a conserved role in dorsoventral patterning of the neuroectoderm and mesoderm. *Mech Dev* 58(1–2):217–231.
9. Knirr S, Azpiazu N, Frasch M (1999) The role of the NK-homeobox gene slouch (S59) in somatic muscle patterning. *Development* 126:4525–4535.
10. Schaub C, Nagaso H, Jin H, Frasch M (2012) Org-1, the Drosophila ortholog of Tbx1, is a direct activator of known identity genes during muscle specification. *Development* 139(5):1001–1012.
11. Carrasco-Rando M, et al. (2011) Drosophila Araucan and Caupolican Integrate Intrinsic and Signalling Inputs for the Acquisition by Muscle Progenitors of the Lateral Transverse Fate. *PLoS Genet* 7(7):e1002186.
12. Boukhatmi H, et al. (2012) Tup/Islet1 integrates time and position to specify muscle identity in Drosophila. *Development* 139(19):3572–3582.
13. Busser BW, et al. (2012) Molecular mechanism underlying the regulatory specificity of a Drosophila homeodomain protein that specifies myoblast identity. *Development* 139(6):1164–1174.
14. Junion G, et al. (2007) Genome-wide view of cell fate specification: ladybird acts at multiple levels during diversification of muscle and heart precursors. *Genes Dev* 21(23):3163–3180.
15. Bataillé L, Delon I, Da Ponte JP, Brown NH, Jagla K (2010) Downstream of Identity Genes: Muscle-Type-Specific Regulation of the Fusion Process. *Dev Cell* 19(2):317–328.
16. de Joussineau C, Bataillé L, Jagla T, Jagla K (2012) Diversification of Muscle Types in Drosophila. *Current Topics in Developmental Biology* (Elsevier), pp 277–301.
17. Heiman M, et al. (2008) A Translational Profiling Approach for the Molecular Characterization of CNS Cell Types. *Cell* 135(4):738–748.
18. Bertin B, Renaud Y, Aradhya R, Jagla K, Junion G (2015) TRAP-rc, Translating Ribosome Affinity Purification from Rare Cell Populations of Drosophila Embryos. *J Vis Exp* (103):52985.

19. Watson FL, et al. (2012) Cell type-specific translational profiling in the *Xenopus laevis* retina. *Dev Dyn* 241(12):1960–1972.
20. Tryon RC, Pisat N, Johnson SL, Dougherty JD (2013) Development of translating ribosome affinity purification for zebrafish: Zebrafish Translating Ribosome Purification. *genesis* 51(3):187–192.
21. Thomas A, et al. (2012) A Versatile Method for Cell-Specific Profiling of Translated mRNAs in Drosophila. *PLoS ONE* 7(7):e40276.
22. The modENCODE Consortium, et al. (2010) Identification of Functional Elements and Regulatory Circuits by Drosophila modENCODE. *Science* 330(6012):1787–1797.
23. Vorbrüggen G, et al. (1997) Embryonic expression and characterization of a Ptx1 homolog in Drosophila. *Mech Dev* 68(1–2):139–147.
24. Wang JS, Infante CR, Park S, Menke DB (2018) PITX1 promotes chondrogenesis and myogenesis in mouse hindlimbs through conserved regulatory targets. *Dev Biol* 434(1):186–195.
25. Nag S, Larsson M, Robinson RC, Burtnick LD (2013) Gelsolin: The tail of a molecular gymnast: Gelsolin Superfamily Proteins. *Cytoskeleton* 70(7):360–384.
26. Chellaiah M, et al. (2000) Gelsolin Deficiency Blocks Podosome Assembly and Produces Increased Bone Mass and Strength. *J Cell Biol* 148(4):665–678.
27. Lu M, Witke W, Kwiatkowski DJ, Kosik KS (1997) Delayed Retraction of Filopodia in Gelsolin Null Mice. *J Cell Biol* 138(6):1279–1287.
28. Wojtowicz I, et al. (2015) Drosophila small heat shock protein CryAB ensures structural integrity of developing muscles, and proper muscle and heart performance. *Development* 142(5):994–1005.
29. Roy S, VijayRaghavan K (1998) Patterning Muscles Using Organizers: Larval Muscle Templates and Adult Myoblasts Actively Interact to Pattern the Dorsal Longitudinal Flight Muscles of *Drosophila*. *J Cell Biol* 141(5):1135–1145.
30. Cripps RM, Olson EN (1998) Twist Is Required for Muscle Template Splitting during AdultDrosophilaMyogenesis. *Dev Biol* 203(1):106–115.
31. Isaacs ER, Bradley WG, Henderson G (1973) Longitudinal fibre splitting in muscular dystrophy: a serial cinematographic study. *J Neurol Neurosurg Psychiatry* 36(5):813–819.
32. Chan S, Head SI (2011) The role of branched fibres in the pathogenesis of Duchenne muscular dystrophy: Fibre branching in muscular dystrophy. *Exp Physiol* 96(6):564–571.
33. Tixier V, et al. (2013) Glycolysis supports embryonic muscle growth by promoting myoblast fusion. *Proc Natl Acad Sci* 110(47):18982–18987.
34. Mounier N (2002) Actin cytoskeleton and small heat shock proteins: how do they interact? *Cell Stress*:10.
35. Kim JH, et al. (2015) Mechanical Tension Drives Cell Membrane Fusion. *Dev Cell* 32(5):561–573.
36. Port F, Chen H-M, Lee T, Bullock SL (2014) Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. *Proc Natl Acad Sci* 111(29):E2967–E2976.

Figure Legends

Figure 1. Translational profiling of embryonic muscle subsets during acquisition of individual muscle properties using TRAP approach. A. Schematic representation of cell-specific polysome capture during TRAP. B. TRAP-targeted muscle populations revealed by anti-GFP immunostaining of stage 16 embryos showing expression of RpL10a-eGFP in six muscles per hemisegment (Slou>RpL10a-eGFP) and in four lateral muscles (Lms>RpL10a-eGFP) C. Heatmap of differentially expressed genes at the three developmental time windows (T1, T2, T3) in Lms and Slou positive muscle cells. **D**. Bar plots showing proportions of genes whose expression is commonly up (FC>2) in both individual muscle subsets and in all muscles (dark red and dark blue) versus genes up regulated exclusively in Slou (light red) and Lms (light blue) populations. **E**. Pie chart showing distribution of GO terms (biological processes) for genes commonly up-regulated (orange) or down-regulated (blue) in Slou and Lms positive muscles and at all time windows. F. Vulcano plots showing distribution of up and down regulated genes in Slou and Lms positive muscle subsets at the T3 time point. Notice that genes known to be specifically expressed in slou-population (slou, org-1) and lms-population (lms) are consistently up regulated.

Figure 2. Spatial comparison of Lms and Slou up-regulated genes identifies *Ptx1* as a new player of ventral muscle indentity. A. heatmap of genes encoding regulatory proteins with potential role in specification of muscle identity (DNA, RNA and chromatin binding proteins) isolated from the cluster of genes up regulated in Slou population (cluster 3 in Figure S3). **B**. Ptx1 transcript (upper panel) and protein (lower panel) localisation in stage 13 embryos. Arrows denote Ptx1 expression in a subpopulation of developing ventral muscles. **C.** Mapping of Ptx1 muscle expression. Ptx1 is co-expressed with Slou (Slou-GFP) in VA3 muscle precursor and as demonstrated in Ptx1-lacZ line is also prominently expressed in VL4 and VO6 (right lower panel) . **D.** Scheme illustrating expression of Ptx1 (red) versus slou (light green). Commonly expressing both genes VA3 appears yellow. **E.** Coding sequence and genomic organisation of wt *Ptx1* locus and its CRISPR mutant. gRNA induced cleavage is indicated by red arrow and generated STOP codon by the red line. **F**. Ventral view of 3 segments of stage 16 *wild type* and *Ptx1* mutant embryos showing an example of VO6 and VA3 phenotypes. In Wt panel arrows point to the contact zone between VA3 and VO4 muscles and arrowheads to three ventrally extending VO4, VO5 and VO6 muscles. Notice that in *Ptx1* mutant embryos VA3 could extend aberrantly and looses contact with VO4 (arrows) and also instead of three only two VO muscles are detected (arrowheads). **G**. Quantification of observed in Ptx1 mutants phenotypes. Fisher's exact test was applied to calculate significance. (** $p<0.01$; *** $p<0.001$). Scale bars , 100μm in B and 50μm C and F..

Figure 3. Comparison of temporal transition profiles of TRAP-enriched genes identifies cluster of actin regulators enriched in Lms muscle cells. A. Heatmap showing the transition profiles of gene expression in Lms and Slou population. Tr1 represents gene expression transition from the time window T2 to T1 and Tr2 the transition from T3 to T2. Cluster 1 (green) and cluster 2 (red) with genes showing transition profiles « down-down » and « up-up », respectively are indicated **B**. GO comparison of genes belonging to Cluster 1 and Cluster 2. An R package "Cluster profiler" has been applied to identify differential GO terms. **C**. A zoomed view of transition profiles heatmap restricted to genes from cluster 2 that

belong to "actin binding" GO category. Notice that this subset of genes shows differential transition profiles in Lms- versus Slou-positive muscle cells. Respective Tr1 and Tr2 fold changes are indicated on the heatmap. **D**. In situ hybridisation experiment showing that *gel* and *dCryAB* that are part of "actin binding" GO class are predominantly expressed in LT muscles. Lateral views of stage 15 embryos are shown. Anti-Actin or anti-β3 tubulin antibodies are used to reveal muscle pattern. Schemes of muscles in an abdominal segment are shown with *gel* and *dCryAB* expression indicated by a color cod: red – high expression and light pink – low expression levels. Scale bars – 50 μm.

Figure 4. Aberrant growth related LT muscle properties in *Gel* and *dCryAB* mutant embryos.

A, B. Schematic representations of CRISPR mutagenesis of *dCryAB* (A) and *gel* (B) loci. gRNAs targeting sites are indicated by red arrows and positions of resulting stop codons by red lines. Homology directed recombination (HDR) with replacement of *dCryAB* coding sequence by a 3P3-dsRed casette was applied to generate $dCryAB^{HR}$ null mutant line whereas 2bp insertion in 5' region of *gel* locus via non-homologous end joining (NHEJ) repair generated *gel9.3* mutants. **C-E**. Representative lateral views of muscles from late stage 15 embryos stained with anti-Actin antibody to reveal muscle shapes. Upper panels show two wild type abdominal hemisegments (C) and five hemisegments from $dCryAB^{HR}$ (D) and *gel9.3* (E) mutant embryos. Lower panels are the zoomed views of LT muscles. Yellow arrows and yellow arrowheads point to dorsal and ventral LT termini, respectively. They are aligned in wild type (C) but appear at different levels in two neighboring LTs in $dCryAB^{RH}$ (D) and in *gel9.3* mutants (E) suggesting irregular LT growth. Double-headed arrows show extents of contact between LT3 and LT4 muscles at the end of stage 15, which are often reduced in both $dCryAB^{RH}$ (D) and in *gel9.3* (E) mutant contexts indicating partial dissociation of growing LTs. Double-headed arrowhead points to forked LT end, a phenotype observed in $dCryAB^{RH}$ embryos and white arrows denote observed in *gel* mutants LT muscle splitting (one of LTs appear duplicated). **F.** Qunatification of LT muscle phenotypes in $dCryAB^{RH}$ and in *gel9.3* embryos presented as percentage of segments affected (n – number of abdominal segments examined). Fisher's exact test was applied to determine whether occurance of different LT phenotypes in $dCryAB^{RH}$ and in *gel9.3* mutants is significantly different from that observed in *wild type* embryos. * - $p<0.05$; ** $p<0.01$, *** $p<0.001$. Scale bars, 50 μm.

Figure 5. LT muscle splitting in *gel* mutant embryos is due to an excessive myoblast fusion.

A-C. LT muscle shapes at late stage 15 revealed by anti-GFP staining in Lms>LifeActin-GFP transgenic LT sensor line (A), in *gel* mutant context (B) and after LT targeted overexpression of fusion gene *Duf* (C). Myonuclei are visualised using anti-Mef2 and detected in both splitted fibers. Arrows indicate LT splitting events that is observed in *gel* mutants (B) and could be induced by increasing the number of fusion events via overexpression of Duf (C). The possibility that LT splitting in *gel* mutants is due to an excessive fusion is supported by late fusion events observed in *gel;Lms>LAGFP* embryos (**D**, arrow) and confirmed by quantifications of Mef2 positive LT1-LT4 nuclei (**E**) in *gel* mutant segments with splitted muscles compared to wild type and Lms>LAGFP contexts. Notice that *Duf* overexpression mimics *gel* LT splitting phenotype which correlates with increased number of myonuclei. Myonuclei were counted in one of thirteen wild type abdominal segments and in thirteen *Lms>LAGFP>Duf* abdominal segments with splitting events and in a larger number of abdominal segments n=43 and n=32 from *Lms>LAGFP* and *gel;Lms>LAGFP* embryos ,

respectively. Anova multiple comparison test was applied to assess statistical significance. *** $p<0.001$. **F**, **G.** Connectin stained LT fiber interphases in wild type in *gel* mutant context. Notice an increased numbers of interphases (arrowheads) in segment with splitted LT muscles in *gel* mutant context (G) compared to non-splited LT muscles in *Lms>LAGFP* cotext (F). **H**. Splitting in *gel* mutants is associated with accumulation of βPS integrin (arrows) at termini of splitted LT fibers. Scale bars, 50 μm.

Figure 1

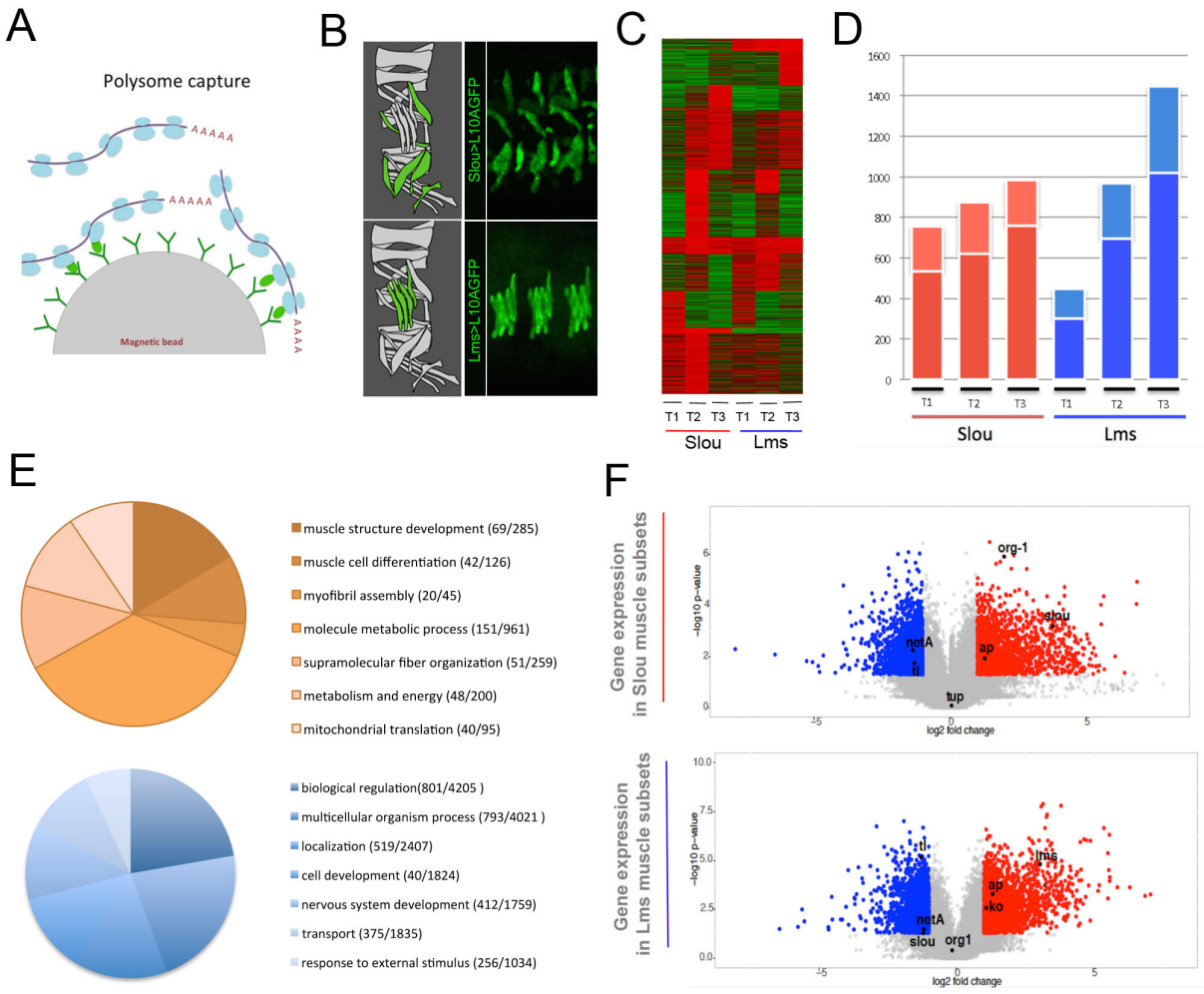

A
Polysome capture
AAAAA
AAAAA
Magnetic bead
AAAA
B
Slou>L10AGFP
Lms>L10AGFP
C
T1 T2 T3 T1 T2 T3
Slou
Lms
D
1600
1400
1200
1000
800
600
400
200
0
T1 T2 T3 T1 T2 T3
Slou
Lms
E
muscle structure development (69/285)
muscle cell differentiation (42/126)
myofibril assembly (20/45)
molecule metabolic process (151/961)
supramolecular fiber organization (51/259)
metabolism and energy (48/200)
mitochondrial translation (40/95)
biological regulation(801/4205)
multicellular organism process (793/4021)
localization (519/2407)
cell development (40/1824)
nervous system development (412/1759)
transport (375/1835)
response to external stimulus (256/1034)
F
Gene expression in Slou muscle subsets
−log10 p-value
log2 fold change
org-1
slou
netA
ap
tl
tup
Gene expression in Lms muscle subsets
−log10 p-value
log2 fold change
tl
lms
ap
ko
netA
slou
org1

Figure 2

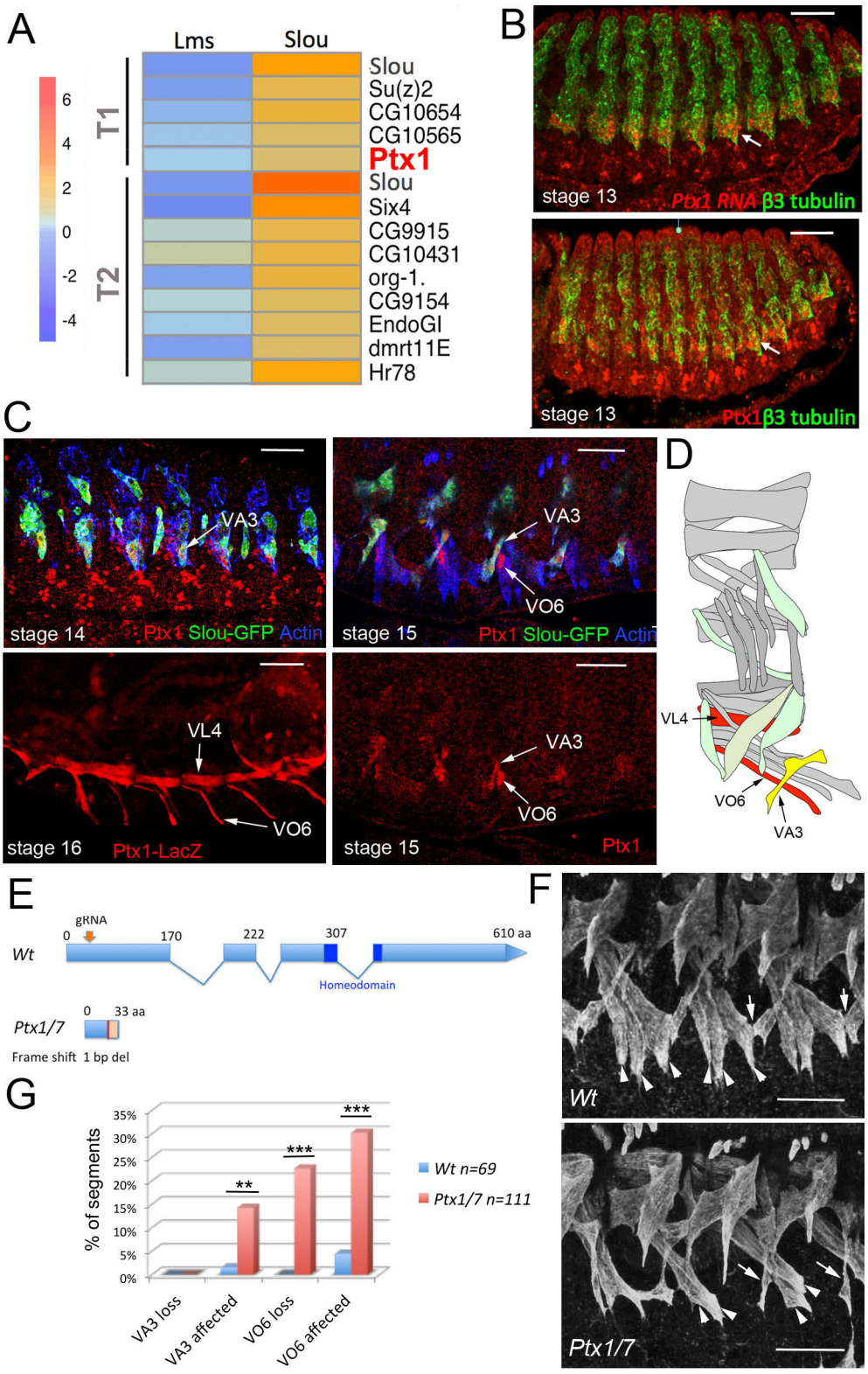

A
Lms
Slou
T1
T2
Slou
Su(z)2
CG10654
CG10565
Ptx1
Slou
Six4
CG9915
CG10431
org-1.
CG9154
EndoGI
dmrt11E
Hr78
B
stage 13
Ptx1 RNA β3 tubulin
stage 13
Ptx1 β3 tubulin
C
VA3
stage 14
Ptx1 Slou-GFP Actin
VA3
VO6
stage 15
Ptx1 Slou-GFP Actin
VL4
VO6
stage 16
Ptx1-LacZ
VA3
VO6
stage 15
Ptx1
D
VL4
VO6
VA3
E
gRNA
Wt
0
170
222
307
610 aa
Homeodomain
Ptx1/7
0
33 aa
Frame shift 1 bp del
F
Wt
Ptx1/7
G
% of segments
35%
30%
25%
20%
15%
10%
5%
0%
**
***
***
Wt n=69
Ptx1/7 n=111
VA3 loss
VA3 affected
VO6 loss
VO6 affected

Figure 3

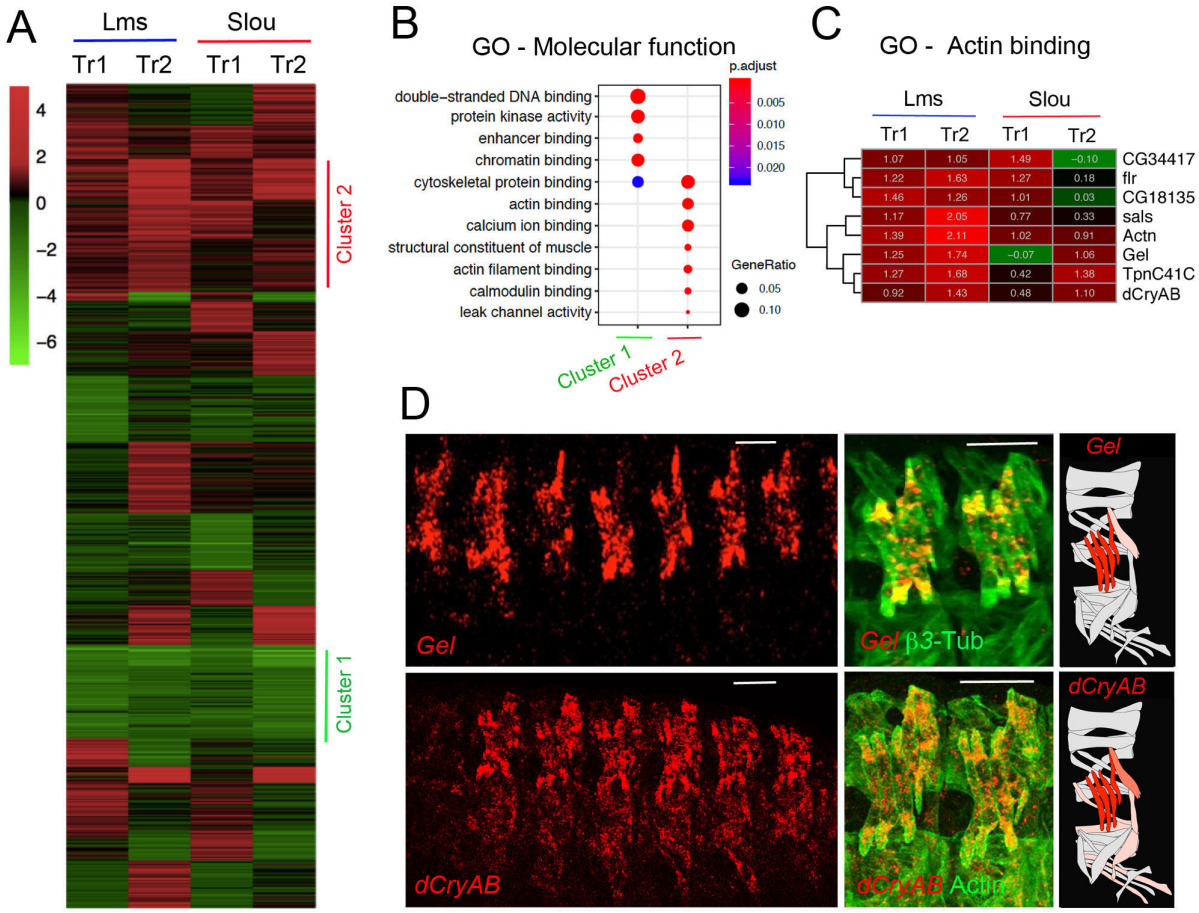

A
Lms
Slou
Tr1
Tr2
Tr1
Tr2
4
2
0
−2
−4
−6
Cluster 2
Cluster 1
B
GO - Molecular function
p.adjust
0.005
0.010
0.015
0.020
double−stranded DNA binding
protein kinase activity
enhancer binding
chromatin binding
cytoskeletal protein binding
actin binding
calcium ion binding
structural constituent of muscle
actin filament binding
calmodulin binding
leak channel activity
GeneRatio
0.05
0.10
Cluster 1
Cluster 2
C
GO - Actin binding
Lms
Slou
Tr1
Tr2
Tr1
Tr2
1.07 1.05 1.49 −0.10 CG34417
1.22 1.63 1.27 0.18 flr
1.46 1.26 1.01 0.03 CG18135
1.17 2.05 0.77 0.33 sals
1.39 2.11 1.02 0.91 Actn
1.25 1.74 −0.07 1.06 Gel
1.27 1.68 0.42 1.38 TpnC41C
0.92 1.43 0.48 1.10 dCryAB
D
Gel
Gel β3-Tub
Gel
dCryAB
dCryAB Actin
dCryAB

Figure 4

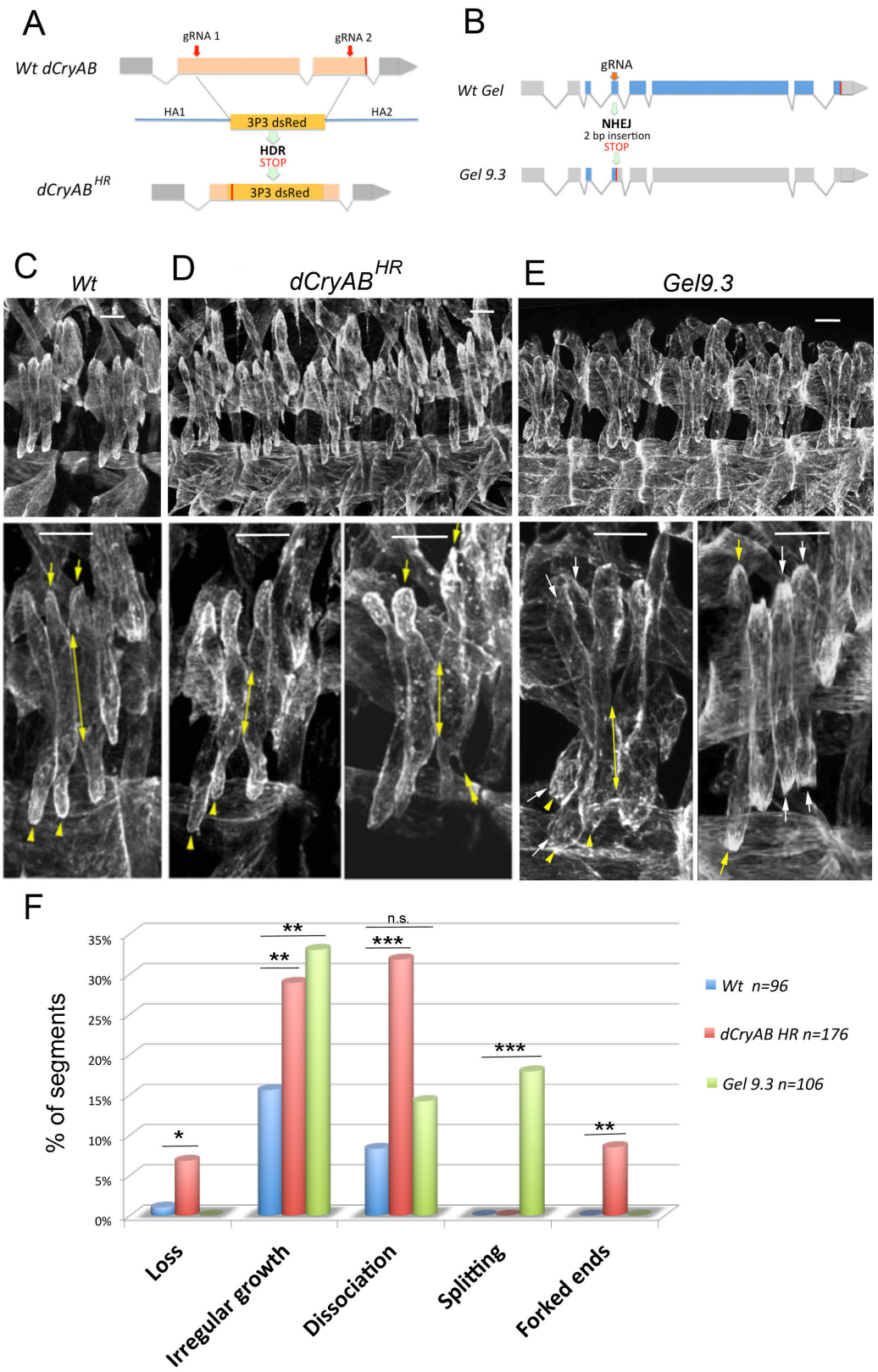
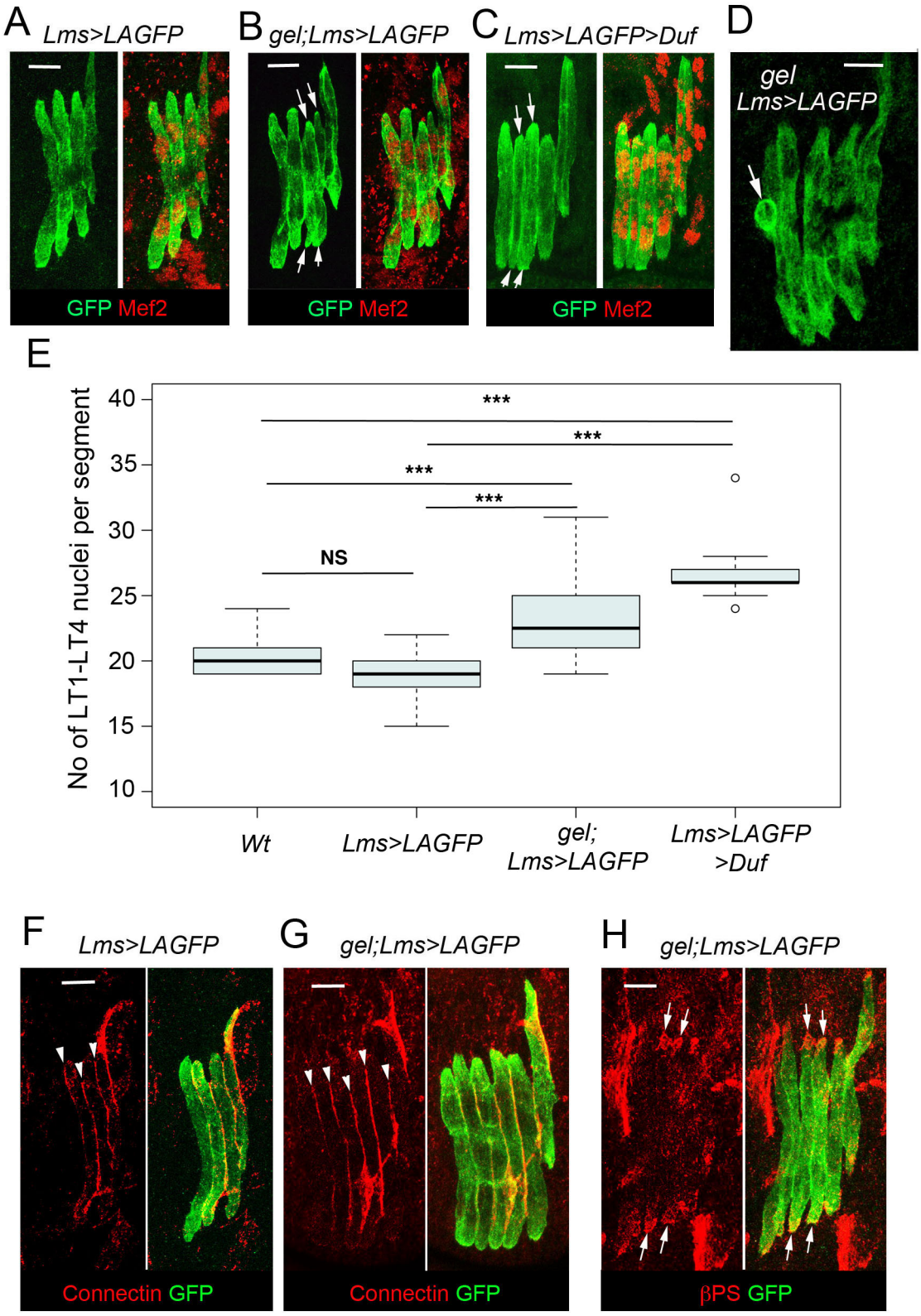

A
gRNA 1
gRNA 2
Wt dCryAB
HA1
3P3 dsRed
HA2
HDR
STOP
dCryAB HR
3P3 dsRed
B
gRNA
Wt Gel
NHEJ
2 bp insertion
STOP
Gel 9.3
C
Wt
D
dCryAB HR
E
Gel9.3
F
% of segments
35%
30%
25%
20%
15%
10%
5%
0%
*
**
**
n.s.
***
***
**
Loss
Irregular growth
Dissociation
Splitting
Forked ends
Wt n=96
dCryAB HR n=176
Gel 9.3 n=106

Figure 5


A Lms>LAGFP
B gel;Lms>LAGFP
C Lms>LAGFP>Duf
D
gel
Lms>LAGFP
GFP Mef2
GFP Mef2
GFP Mef2
E
No of LT1-LT4 nuclei per segment
40
35
30
25
20
15
10
***
***
***
***
NS
Wt
Lms>LAGFP
gel;
Lms>LAGFP
Lms>LAGFP
>Duf
F Lms>LAGFP
G gel;Lms>LAGFP
H gel;Lms>LAGFP
Connectin GFP
Connectin GFP
βPS GFP