# Figures and figure supplements

Transcriptional profiling reveals extraordinary diversity among skeletal muscle tissues

**Erin E Terry *et al***

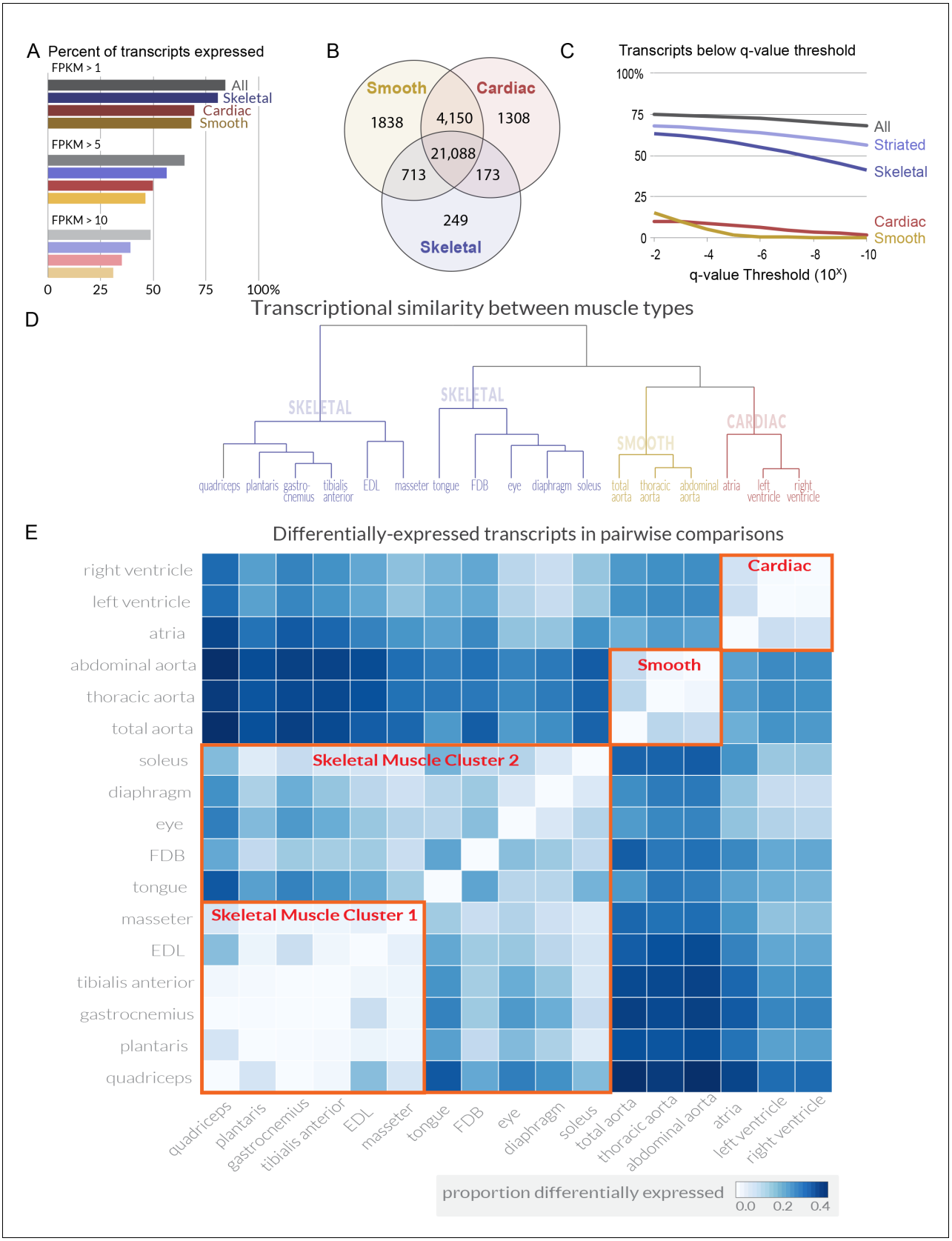



**Figure 1.** Transcriptome profiling reveals extensive gene expression differences among muscle tissues. (**A**) The percent of all transcripts detected as being expressed in different classes of tissues is shown as a bar graph. >80% of all transcripts are detectably expressed in at least one skeletal muscle

*Figure 1 continued on next page*

*Figure 1 continued*

tissue. (**B**) The number of transcripts expressed (FPKM >1) in *every* cardiac, smooth, or skeletal muscle tissue is shown as a Venn diagram. A core of ~21,000 transcripts is expressed in every contractile tissue. Please note that panel A describes transcripts expressed in at least one tissue; panel B describes transcripts expressed in every tissue. (**C**) The percent of transcripts showing differential expression between tissues is shown at different false-discovery rate (q-value) thresholds. One-way ANOVAs of all tissues, all striated, all skeletal, all cardiac, and all smooth muscles were used to calculate q-values. Striated muscle refers to skeletal plus cardiac muscles. (**D**) The overall similarity of transcriptional profiles in different tissues is displayed as a dendrogram. Notably, the three major classes of muscle (smooth, cardiac, and skeletal) cluster together as expected. (**E**) The number of differentially expressed transcripts ($q < 0.01$, fold change >2) in pairwise comparisons is shown as a heat map. Red boxes indicate clustering by similarity of (1) cardiac muscle, (2) smooth muscle, and (3) two different clusters of skeletal muscle.

DOI: https://doi.org/10.7554/eLife.34613.004

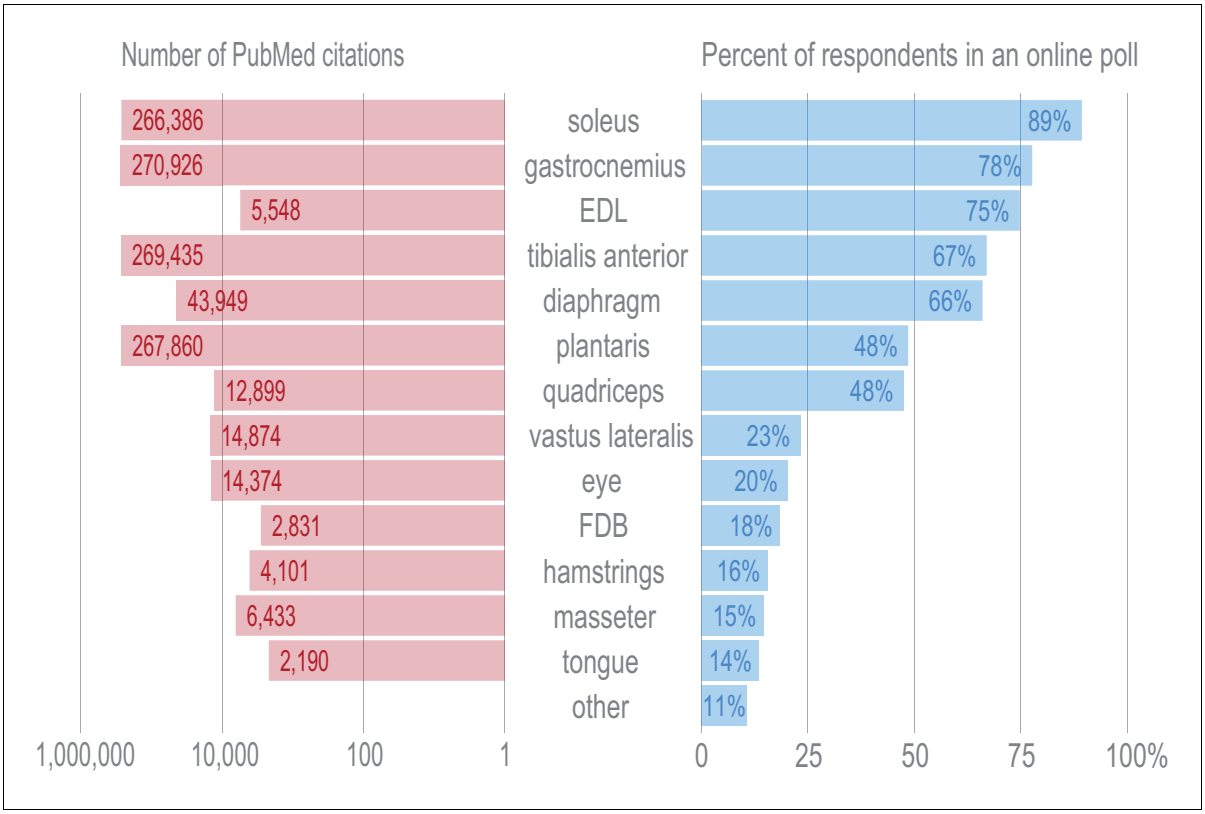



**Figure 1—figure supplement 1.** Tissues were selected for transcriptional profiling to maximize the utility of these data to the skeletal muscle field. In order to ascertain which tissues are of the greatest general interest to the skeletal muscle researchers, we distributed an online poll asking respondents to vote whether any given tissue was of interest to their laboratory (right panel). Over one hundred responses from principal investigators, graduate students, and postdocs in the field were recorded. The choice of 'other' was ranked lower than any other tissue, and write-in responses showed no evidence of broad support for any tissues beyond those included here. *Quadriceps* were selected for profiling as they were favored in the poll over the more specific *vastus lateralis* by a ~ 2:1 ratio. Masseter and tongue were chosen for transcriptional profiling over hamstrings to increase the developmental and anatomical diversity of the tissues in this data set. In order to guard against selection bias for our online poll, we queried NCBI's PubMed to identify the number of published papers on each muscle tissue (left panel). The overall distribution of PubMed citations mirrors the human poll.

DOI: https://doi.org/10.7554/eLife.34613.005

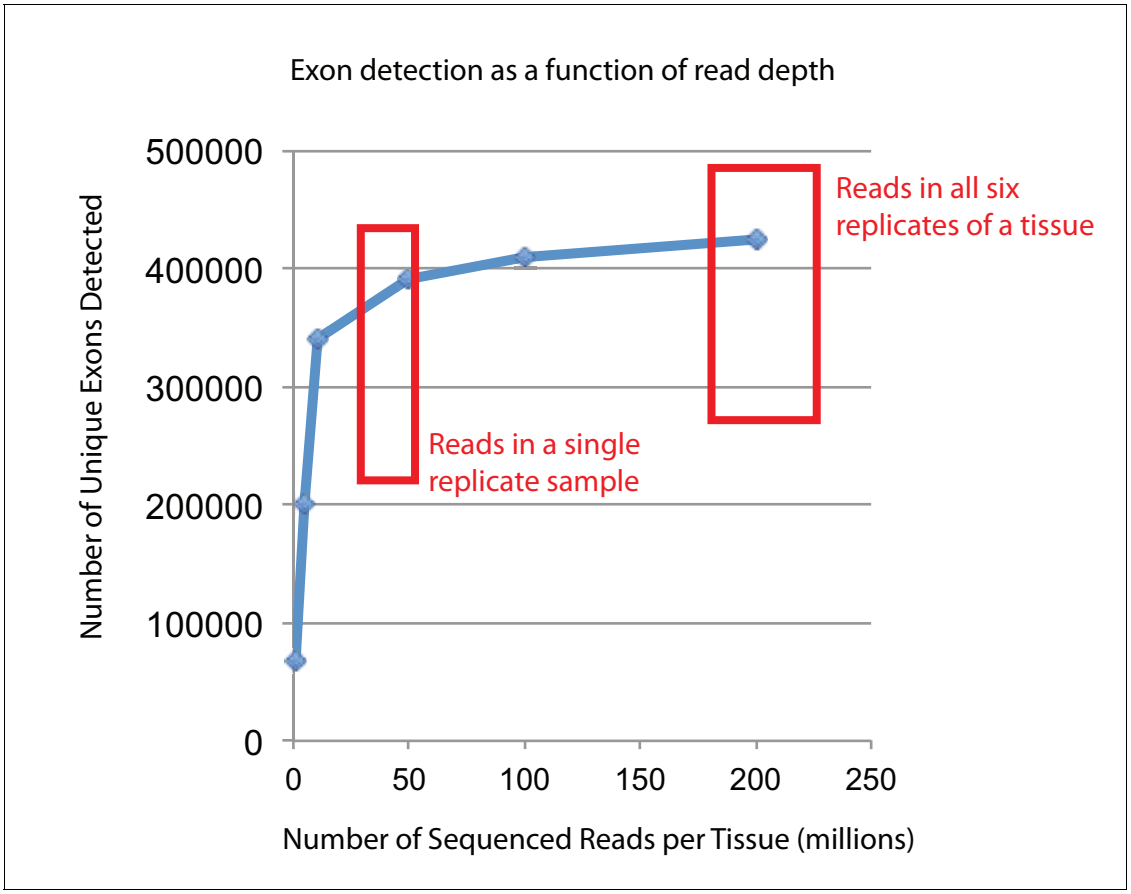



**Figure 1—figure supplement 2.** Empirical simulations show that the read depth of this study approaches saturation for detecting expressed exons. In order to verify that the RNAseq read depth was sufficient to detect most expressed transcripts, one specific tissue (mouse *FDB*) was randomly down-sampled as previously described (***Li et al., 2015***) to generate six different sequencing depths (200, 100, 50, 10, 5, and 1 million reads). These reads were aligned to the mouse genome and transcriptome using RUM (***Grant et al., 2011***), and uniquely aligned reads were used to identify putatively expressed exons. The number of unique exons detected as expressed (FPKM >0) is plotted versus the total number of reads collected. The two red boxes indicate either the approximate number of reads for any given biological replicate or the approximate number of reads per tissue (summing all six replicates). Most single replicates have been sequenced to a depth that approaches diminishing returns for exon detection, while every tissue has been sequenced to a depth that guarantees that the vast majority of truly expressed exon are detected.

DOI: https://doi.org/10.7554/eLife.34613.006

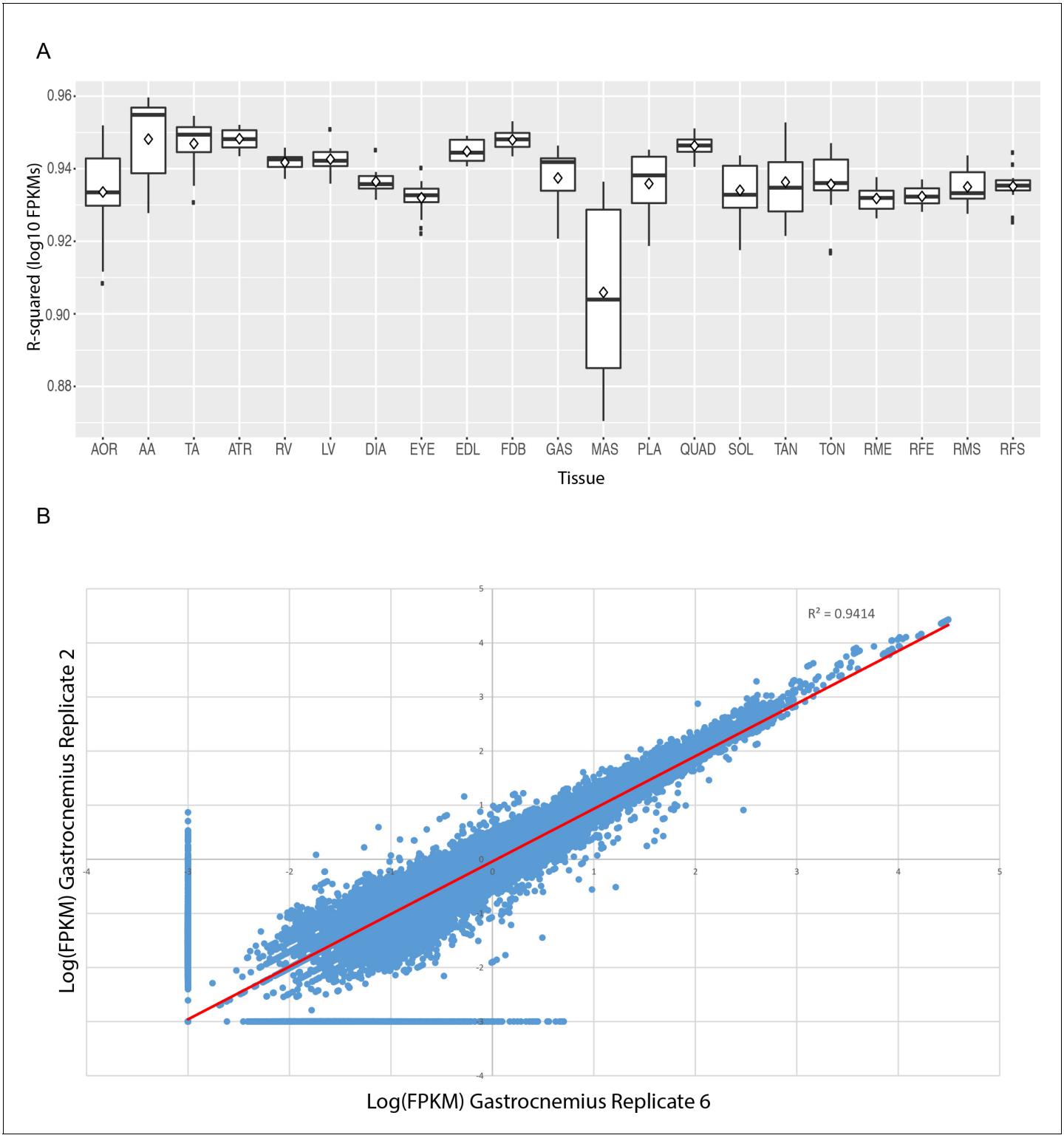



**Figure 1—figure supplement 3.** FPKM values show high reproducibility between replicate samples. (**A**) Box and whisker plots of R-squared values of pairwise comparisons between replicates of each tissue. Every median R-squared is greater than 0.90. Excluding masseter, every median R-squared is greater than 0.93. AOR = total aorta, AA = abdominal aorta, TA = thoracic aorta, ATR = atria, RV = right ventricle, LV = left ventricle, DIA = diaphragm, EYE = extraocular eye muscle, EDL = extensor digitorum longus, FDB = flexor digitorum brevis, GAS = gastrocnemius, MAS = masseter, PLA = plantaris, QUAD = quadriceps, SOL = soleus, TAN = tibialis anterior, TON = tongue, RME = male rat extensor digitorum longus, RFE = female rat extensor digitorum longus, RMS = male rat soleus, RFS = female rat *soleus*. (**B**) Scatter plot of all transcript-level FPKM values for two representative

*Figure 1—figure supplement 3 continued on next page*

*Figure 1—figure supplement 3 continued*

replicates (mouse *gastrocnemius* replicates 2 and 6). FPKM values are log10 transformed; any FPKM = 0 was plotted as an arbitrarily low expression value (FPKM = 0.001) to avoid log-transforming zero.

DOI: https://doi.org/10.7554/eLife.34613.007

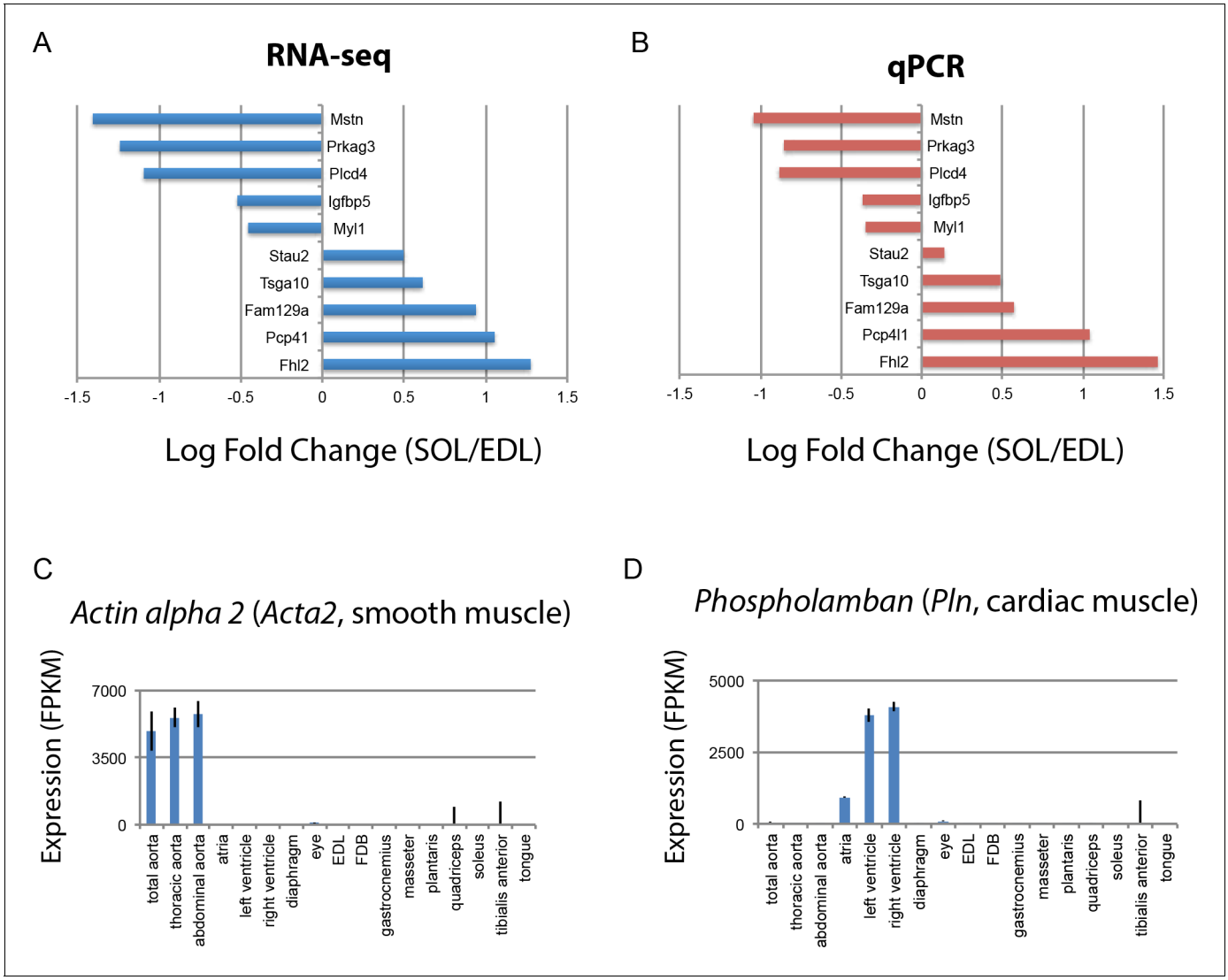



**Figure 1—figure supplement 4.** Internal controls demonstrate the reliability of differential expression analysis. 10 genes differentially expressed between mouse *EDL* and *soleus* were selected with log2 fold changes ranging from −1.5 to 1.5. Panel (**A**) shows the *EDL/soleus* log2 fold changes as a bar graph from RNAseq data. Panel (**B**) shows the log2 fold changes for the same genes, normalized to the house-keeping gene *Importin 8*, measured by qPCR on biological samples *collected independently of those used for RNAseq*. Internal controls based on known tissue-specific genes also corroborate the reliability of these data. (**C**) Expression of *Acta2*, a smooth muscle-specific gene, is plotted as a bar graph. (**D**) Expression of *Pln*, a cardiac muscle-specific gene, is plotted as a bar graph. Error bars are ±S.E.M.

DOI: https://doi.org/10.7554/eLife.34613.008

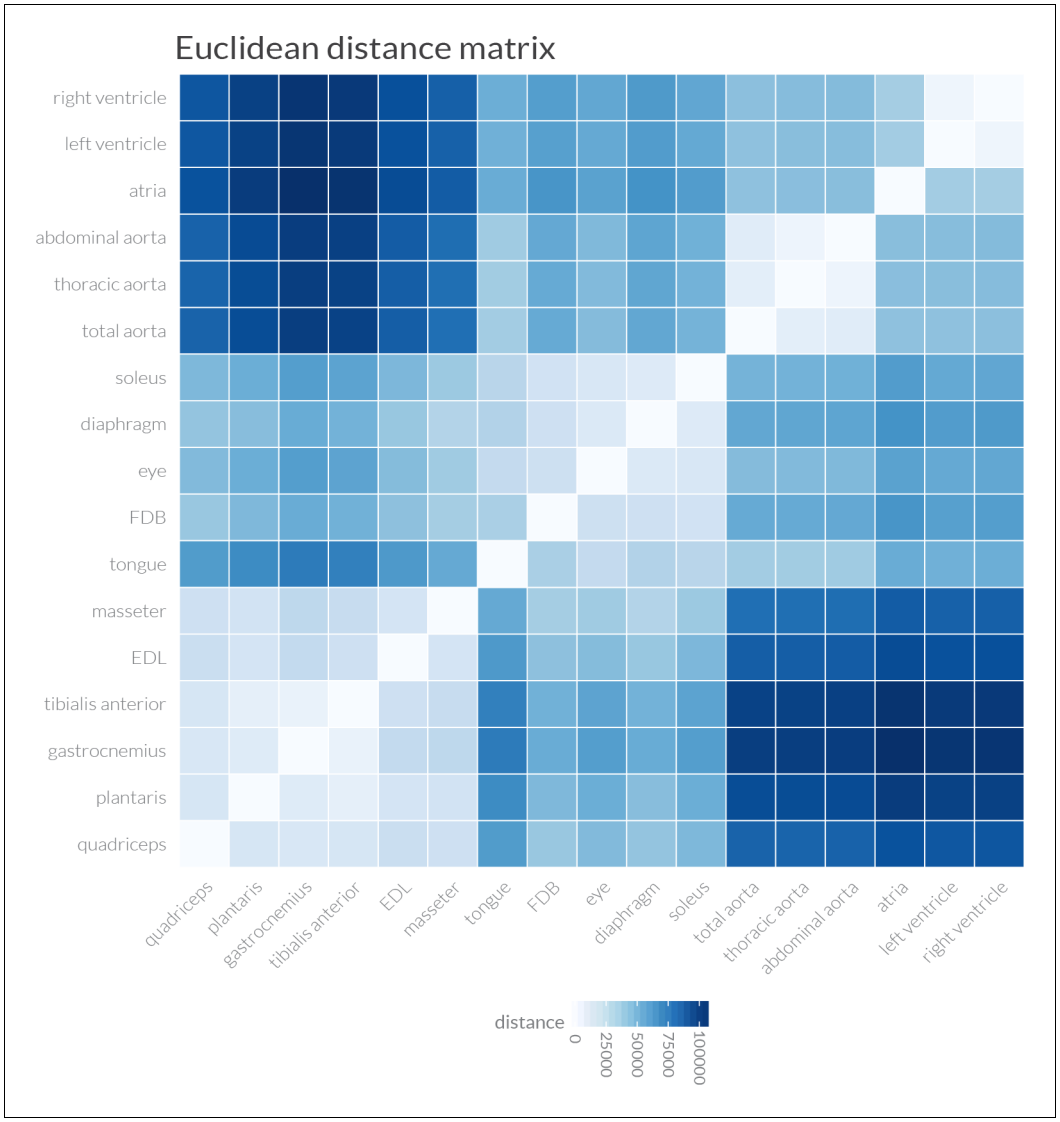



**Figure 1—figure supplement 5.** Euclidean distance measurements support the clustering of muscle transcriptomes into four distinct groups. Pairwise Euclidean distances between the entire transcriptome of every mouse tissue are shown as a heat map. Overall similarity reveals four distinct *groupings*: (**1**) cardiac muscle, (**2**) smooth muscle/aorta, (**3**) limb (except soleus) and masseter skeletal muscles, and (***Hughes et al., 2009***) other skeletal muscles, including soleus. These data were used to generate the dendrogram shown in ***Figure 1D***.
DOI: https://doi.org/10.7554/eLife.34613.009

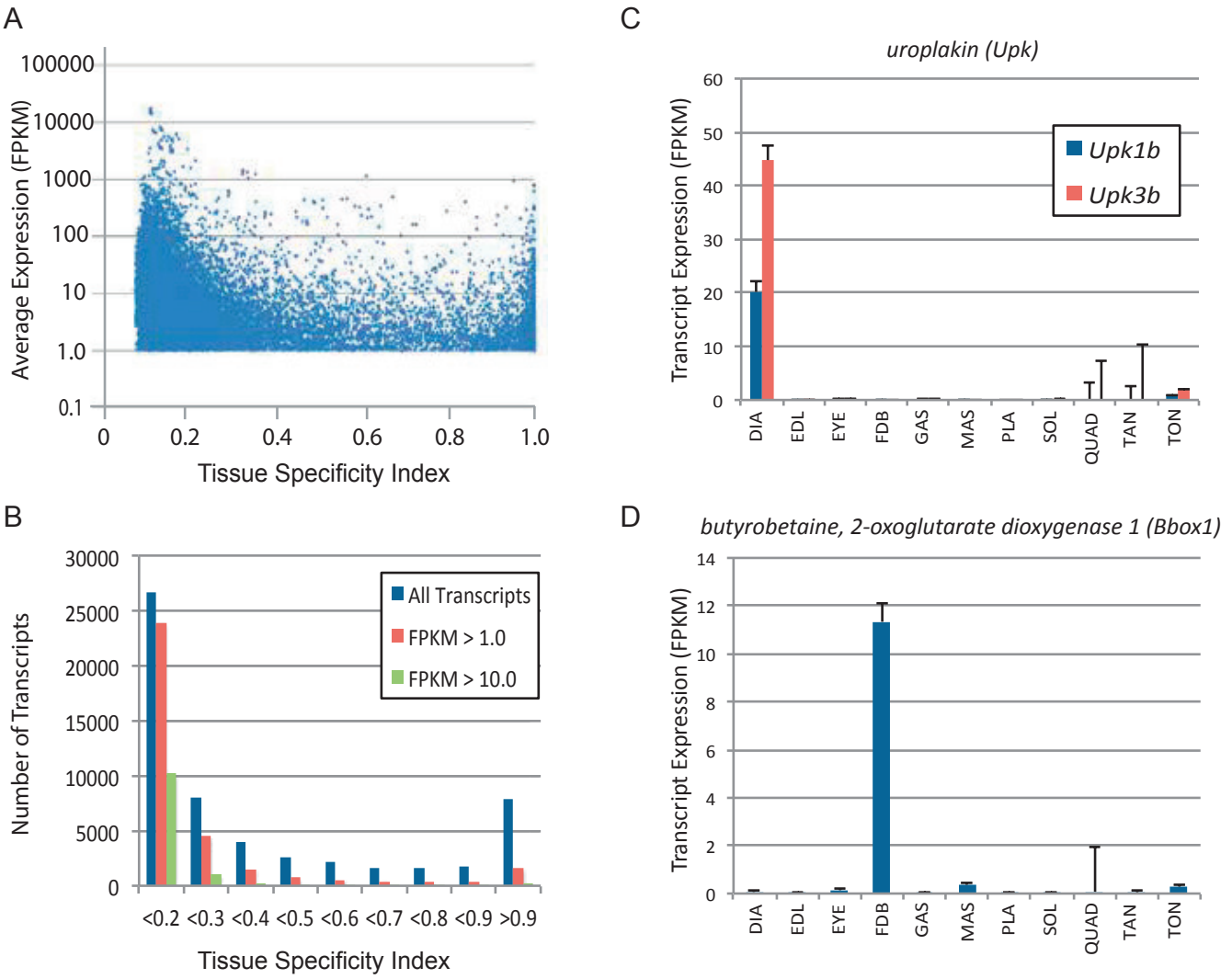



**Figure 1—figure supplement 6.** Subsets of genes are expressed specifically in different skeletal muscles. (**A**) Tissue specificity (maximum expression in *any* tissue divided by the sum total expression in *all* tissues) was calculated for all expressed transcripts in 11 different skeletal muscle tissues. The tissue specificity index ranges from 1.0 (completely specific to a single tissue) to 0.091 (evenly expressed among all 11 skeletal muscle tissues). Average expression is plotted versus tissue specificity for all transcripts expressed above an average FPKM of 1.0. The distribution of transcripts is quantified in (**B**), where the majority of expressed transcripts are expressed approximately equally across all 11 tissues (specificity index <0.2), and a smaller group of transcripts are expressed with high specificity (specificity index >0.9). Panel (**C**) shows two examples of genes (*Upk1b* and *Upk3b*) that are specifically expressed in the diaphragm. Panel (**D**) shows an example of a gene (*Bbox1*) expressed specifically in the *FDB* (Error bars are ±S.E.M.). Please see the Materials and methods for a list of antibodies tested by immunohistochemistry.

DOI: https://doi.org/10.7554/eLife.34613.010

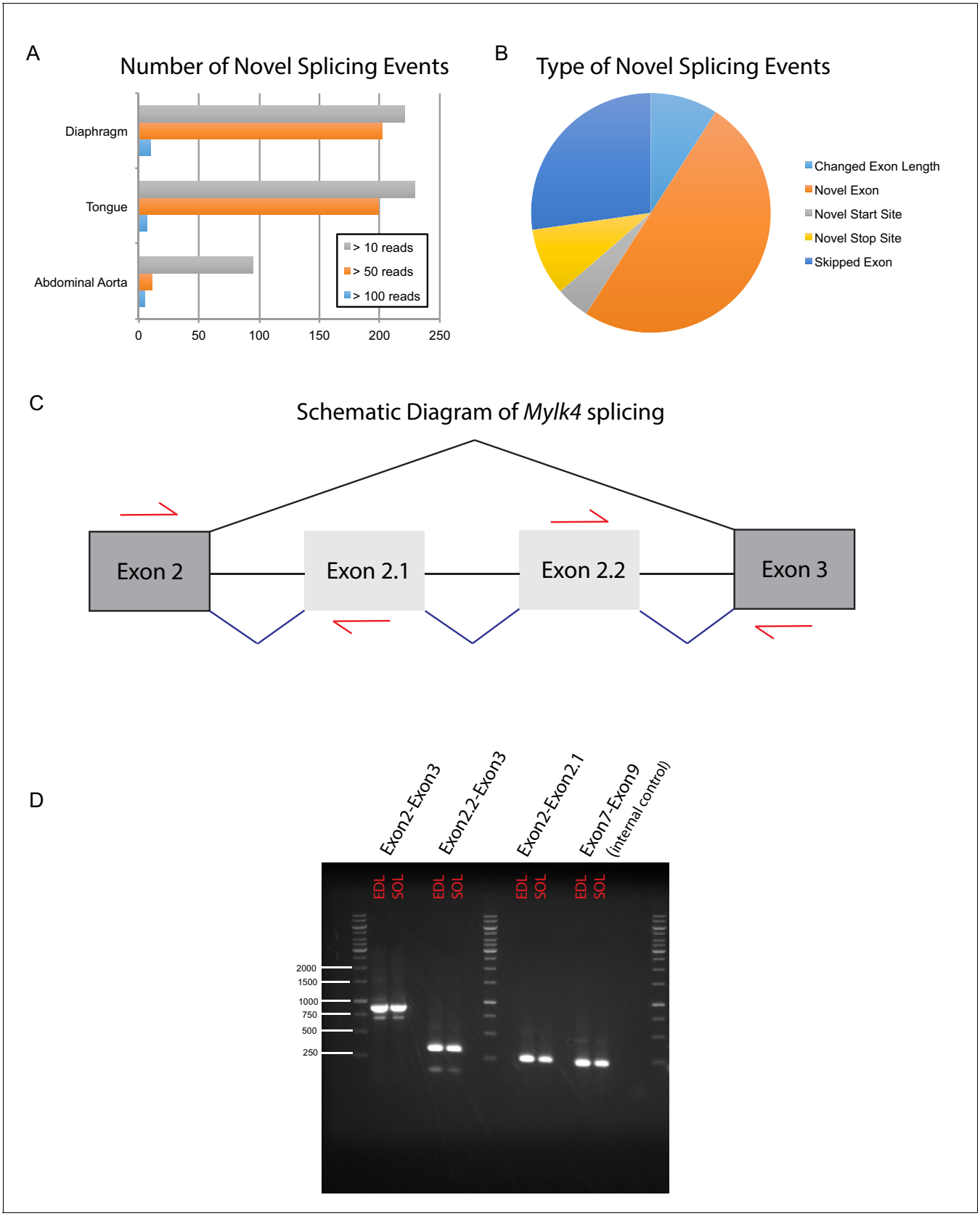


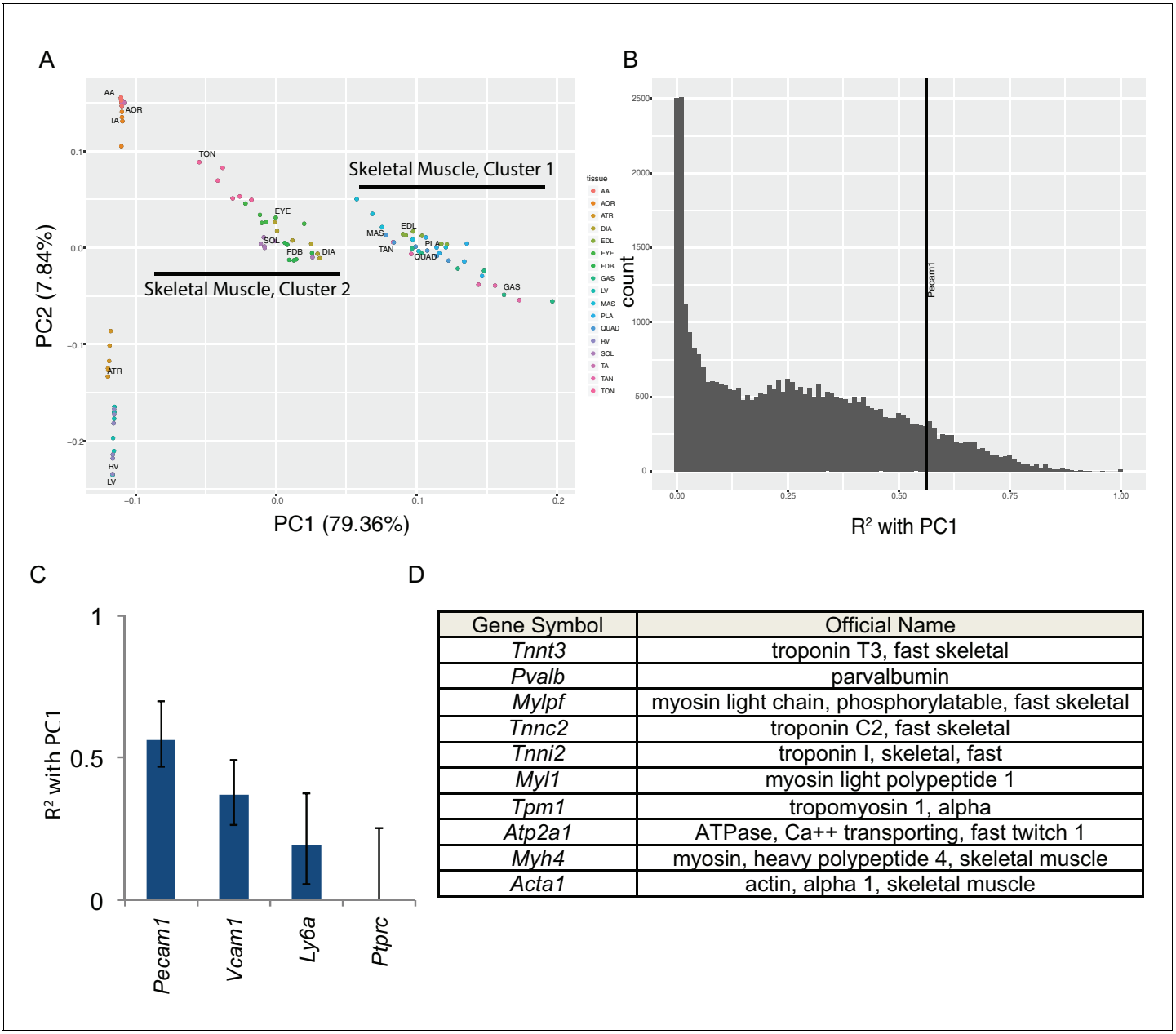

**Figure 1—figure supplement 7.** Deep sequencing the muscle transcriptome reveals numerous novel splicing events. In order to detect previously unannotated splicing events, we manually queried RUM alignments from a subset of our data, including mouse diaphragm, tongue, abdominal aorta,

*Figure 1—figure supplement 7 continued on next page*

*Figure 1—figure supplement 7 continued*

*EDL* and *soleus*. (**A**) The number of unique novel splicing events detected at different read depths in different mouse muscle tissues is plotted as a bar graph. (**B**) We manually curated the highest confidence hits (supporting reads >100) and identified five major classes of novel splicing events. Relative proportions are represented as a pie chart; the most common novel splicing event were novel exons. Panel (**C**) shows a cartoon diagram of *Myosin light chain kinase 4* (*Mylk4*) splicing between canonical exons 2 and 3 (dark grey boxes). RNAseq data predict the presence of two heretofore unannotated exons, which we provisionally term exons 2.1 and 2.2 (light grey boxes). Black lines represent canonical splicing; blue lines represent potentially novel splicing events. Red arrows indicate the location of RT-PCR primers designed to verify whether these predicted events occur in muscle tissue. (**D**) RT-PCR gel confirming that the novel splicing events occur in vivo. Input RNA was taken from male mouse *EDL* and *soleus* muscle *collected independently* of the samples used for RNAseq. Experimental lanes 1 and 2 using primers spanning canonical exons 2 and 3, show a predominant PCR product of ~880 bp in both *EDL* and *soleus*, consistent with splicing of both exons 2.1 and 2.2 into the full length transcript. This band was TOPO cloned and conventional Sanger sequencing confirmed the presence of both novel exons. A minor product of ~600 bp is also present, which most likely represents the exclusion of one of the two novel exons. Experimental lanes 3 and 4 show a PCR product of ~330 bp from both *EDL* and *soleus* using primers that span exons 2.2 to 3, confirming that this splicing event happens in vivo. Experimental lanes 5 and 6 show a PCR product of ~260 bp from both *EDL* and *soleus* using primers that span exons 2 to 2.1, confirming that this splicing event happens in vivo. Experimental lanes 7 and 8 show a PCR product spanning canonical exons 7 to 9 of *Mylk4*. These lanes were an internal control for *Mylk4* expression in our samples. Interrogation of the GENCODE database reveals that many of the novel splicing events described above, including those in *Mylk4*, are supported by sequenced cDNAs that have not yet been integrated into gene models.

DOI: https://doi.org/10.7554/eLife.34613.011




| Gene Symbol | Official Name |
|---|---|
| *Tnnt3* | troponin T3, fast skeletal |
| *Pvalb* | parvalbumin |
| *Mylpf* | myosin light chain, phosphorylatable, fast skeletal |
| *Tnnc2* | troponin C2, fast skeletal |
| *Tnni2* | troponin I, skeletal, fast |
| *Myl1* | myosin light polypeptide 1 |
| *Tpm1* | tropomyosin 1, alpha |
| *Atp2a1* | ATPase, Ca++ transporting, fast twitch 1 |
| *Myh4* | myosin, heavy polypeptide 4, skeletal muscle |
| *Acta1* | actin, alpha 1, skeletal muscle |

**Figure 2.** Principal component analysis reveals the majority of variance in these data is due to gene expression in skeletal muscle cells. (**A**) Principal components 1 and 2 are plotted on the x- and y-axes respectively for all six replicates of each muscle tissue. PC1 accounts for 79.36% of the variance and separates skeletal muscle cluster 1 from cluster 2 (compare with ***Figure 1D***). PC2 accounts for 7.84% of the variance in these data and largely separates cardiac from smooth muscle tissues. (**B**) Histogram of the $R^2$ values of each transcript's correlation with PC1. The most correlated non-muscle cell 'marker' gene, *Pecam1*, is denoted as a vertical line. (**C**) $R^2$ values with 99% confidence intervals are shown for four 'marker' genes conventionally used to identify non-skeletal muscle cells in whole muscle preparations. (**D**) The top ten genes whose expression is most highly correlated with PC1 are shown as a Table. All ten are characteristic skeletal muscle genes.

DOI: https://doi.org/10.7554/eLife.34613.012

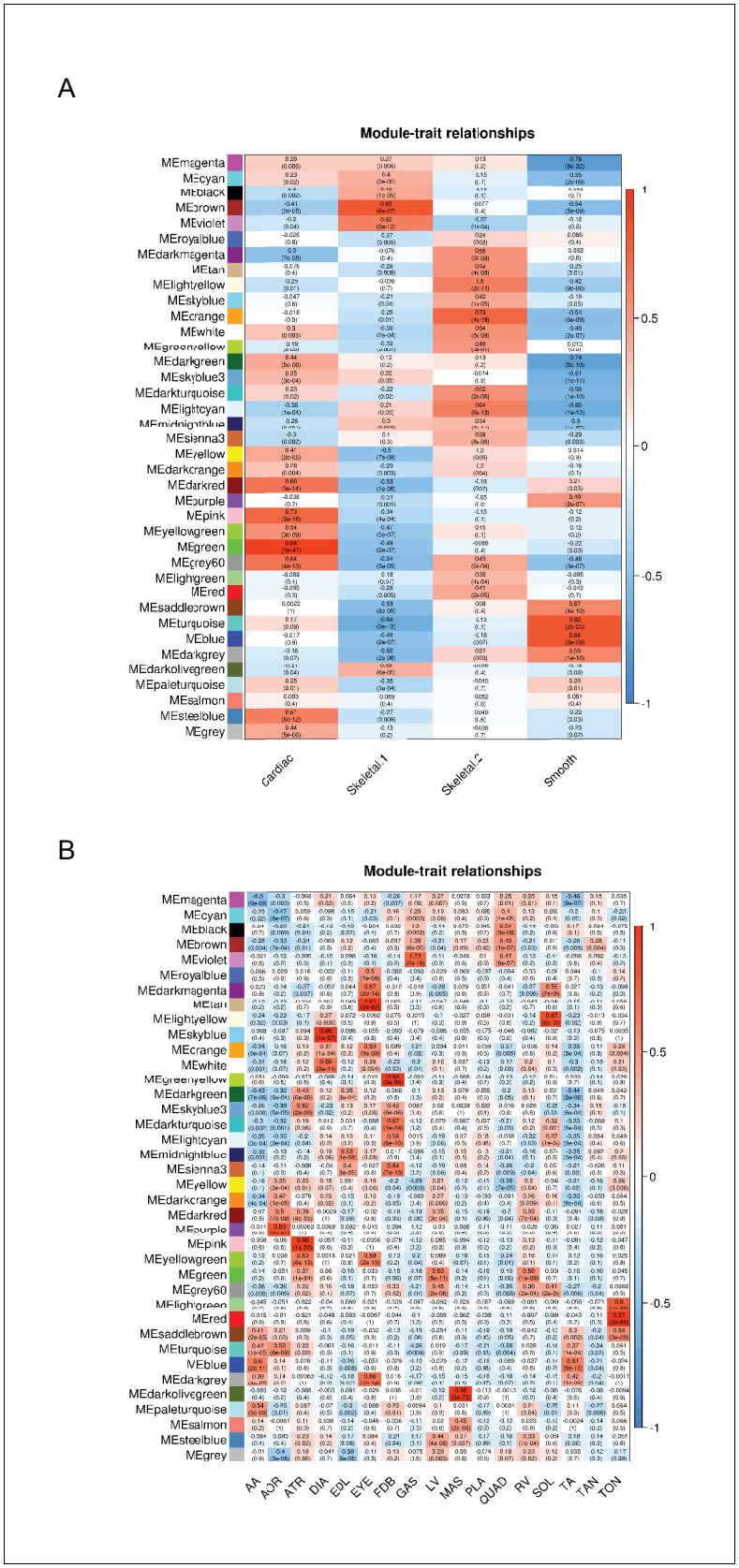



**Figure 2—figure supplement 1.** WGCNA analysis reveals co-expressed modules of genes in different muscle tissues. WGCNA analysis (***Langfelder and Horvath, 2008***; ***Zhang and Horvath, 2005***) was performed to identify

*Figure 2—figure supplement 1 continued on next page*

*Figure 2—figure supplement 1 continued*

clusters of co-expressed genes among (**A**) the four major classes of muscle identified in ***Figure 1***, and (**B**) all 17 tissues. Colored boxes on the left axis represent different co-expressed modules of genes. Each cell is color-coded using a scale from blue to red (the key is on the right axis), which represents the degree of correlation with tissue identity. The upper number in each cell represents the correlation (r-value), and the lower number represents the p-value of its correlation with any given tissue group. AA = abdominal aorta, AOR = total aorta, ATR = atria, DIA = diaphragm, EDL = extensor digitorum longus, EYE = extraocular eye muscles, FDB = flexor digitorum brevis, GAS = gastrocnemius, LV = left ventricle, MAS = masseter, PLA = plantaris, QUAD = quadriceps, RV = right ventricle, SOL = soleus, TA = thorascic aorta, TAN = tibialis anterior, TON = tongue.

DOI: https://doi.org/10.7554/eLife.34613.013

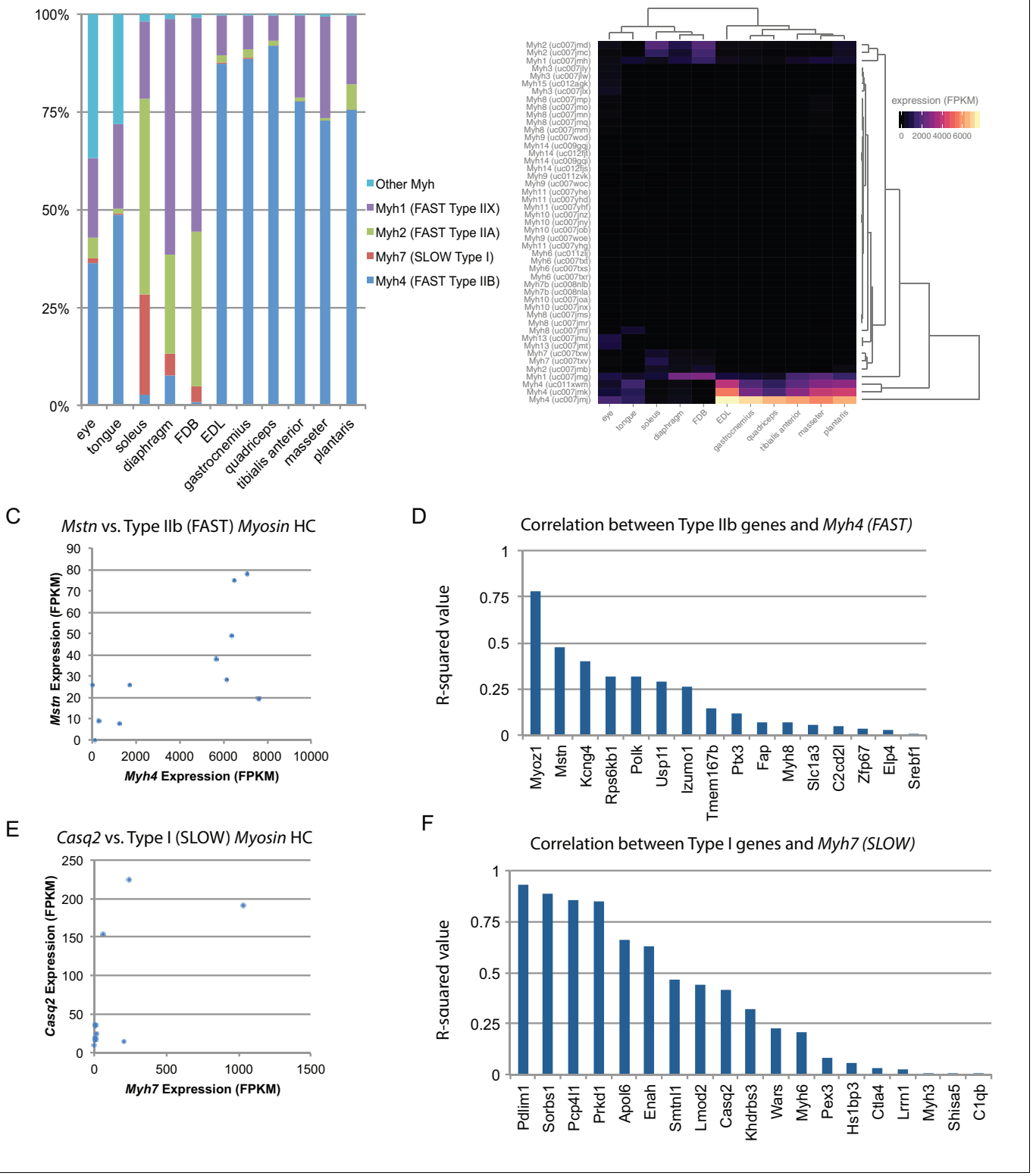



**Figure 3.** Fiber-type composition can explain some, but not all, transcriptional variability between skeletal muscle tissues. (**A**) Relative expression of *Myosin heavy chain* (*Myh*) transcripts as distributed among all 11 mouse skeletal muscle tissues is plotted as a bar graph. The y-axis describes the

*Figure 3 continued on next page*

*Figure 3 continued*

percent of reads aligning to any given *Myh* transcript relative to all *Myh*-aligning reads. (**B**) Absolute expression of *Myh* transcripts among all 11 mouse skeletal muscle tissues is represented as a heat map. Absolute levels were plotted as a heat map to illustrate the dynamic range of *Myh* expression in muscle. Tissues are clustered by overall *Myh* similarity (top dendrogram). Notably, clustering is similar to, but distinct from, clustering done on global transcriptional profiles (***Figure 1D***). (**C**) The expression of *Myostatin* (*Mstn*), a gene specifically expressed by Type IIb (fast) fibers (***Chemello et al., 2011***), is plotted versus *Myh4* expression. Each dot represents one of 11 skeletal muscle tissues ($R^2$ = 0.473). (**D**) R-squared values for correlations with *Myh4* expression are plotted as a bar graph for 16 genes known to be specific for Type IIb fibers. Median $R^2$ = 0.132; only one gene has an $R^2$ >0.5. (**E**) The expression of *Calsequestrin 2* (*Casq2*), a gene specifically expressed by Type I (slow) fibers (***Chemello et al., 2011***), is plotted versus *Myh7* epression. Each dot represents one of 11 skeletal muscle tissues ($R^2$ = 0.412). (**F**) R-squared values for correlations with *Myh7* expression are plotted as a bar graph for 19 genes known to be specific for Type I fibers. Median $R^2$ = 0.321; 7 of 19 genes (37%) tested show essentially no correlation with *Myh4* expression.

DOI: https://doi.org/10.7554/eLife.34613.015

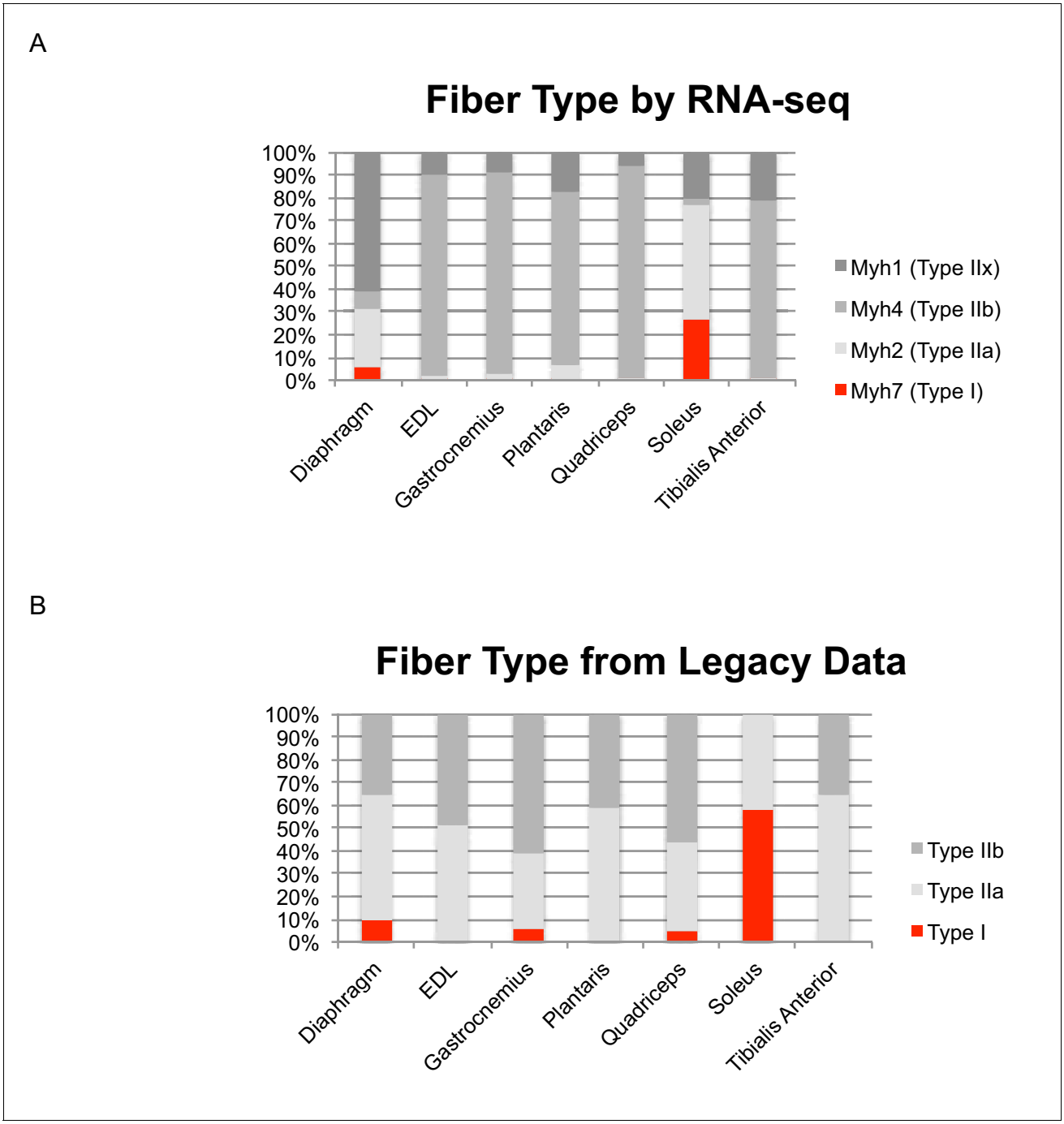



**Figure 3—figure supplement 1.** Identifying skeletal muscle fiber type by *Myh* expression agrees with legacy data. (**A**) Relative expression levels of the four principal *Myosin heavy chains* (*Myh1*, *2*, *4*, and *7*) are shown as a bar graph in select mouse muscle tissues. Fast twitch myosins are shown in grey; slow twitch myosin is shown in red. (**B**) We identified legacy studies that characterized *Myh* expression in mouse diaphragm (***Greising et al., 2015***) as well as the *EDL*, *gastrocnemius*, *plantaris*, *quadriceps*, *soleus* and *tibialis anterior* (***Greising et al., 2015***); the relative compositions of fast/slow twitch fibers is represented as a bar graph. As above, fast twitch *myosins* are shown in grey; slow twitch myosins are shown in red. We note that the previous studies cited herein: (1) measured fast/slow composition in female as opposed to male mice, which is expected to increase the relative proportion of slow twitch fibers, (2) used histology, which is limited to analyzing cross-sectional area of serial sections, rather than the entire muscle body as in the present study, 3) used enzymatic methods to distinguish different fiber types, which limits the extent to which different fast twitch fibers can be resolved, notably neglecting Type IIx fibers. With these caveats in mind, to a first approximation, RNAseq data of *Myh* expression agrees with legacy data of fiber type composition.

DOI: https://doi.org/10.7554/eLife.34613.016

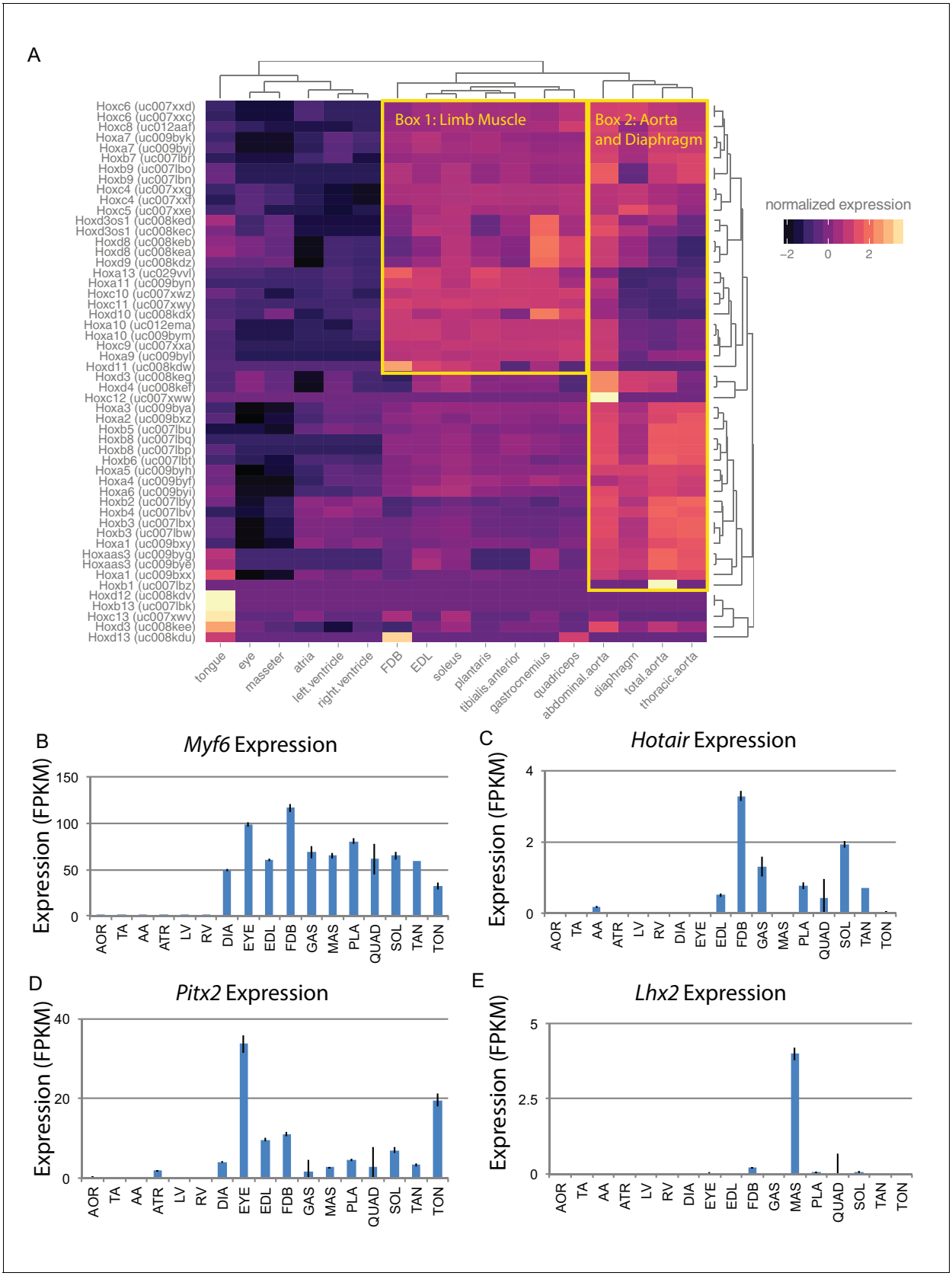



**Figure 4.** Developmental gene expression persists into adulthood in mouse skeletal muscle. (**A**) Normalized expression (Z-score by row) of all *Hox* gene transcripts is represented as a heat map. Row-normalization was chosen to display these data in a way that reveals the fine detail of all *Hox* genes,

*Figure 4 continued on next page*

*Figure 4 continued*

rather than those expressed at the highest levels. Overall similarity by *Hox* gene expression is represented as a dendrogram (top). One yellow box highlights a cluster of *Hox* genes expressed in limb skeletal muscle; a second yellow box highlights a cluster of *Hox* genes expressed in the aorta and diaphragm. (**B**) Expression of *Myf6* is shown as a bar graph. Muscle-specific expression of skeletal muscle differentiation factor persists into adulthood. (**C**) Expression of *Hotair*, a non-coding RNA involved in *Hox* gene regulation, is shown as a bar graph. *Hotair* expression is highly specific for a subset of skeletal muscle tissues involved in limb movement. (**D**) Expression of *Pitx2*, a gene involved in the development of head muscles, is shown as a bar graph. *Pitx2* expression is enriched in head and neck tissues such as the extraocular eye muscles and the tongue. (**E**) Expression of *Lhx2*, another gene involved in head and neck muscle development, is shown as a bar graph. *Lhx2* expression is highly specific for the masseter. AOR = total aorta, ATR = atria, DIA = diaphragm, EDL = extensor digitorum longus, EYE = extraocular eye muscles, FDB = flexor digitorum brevis, GAS = gastrocnemius, LV = left ventricle, MAS = masseter, PLA = plantaris, QUAD = quadriceps, RV = right ventricle, SOL = soleus, TA = thorascic aorta, TAN = tibialis anterior,TON = tongue. Error bars are ±S.E.M.

DOI: https://doi.org/10.7554/eLife.34613.017

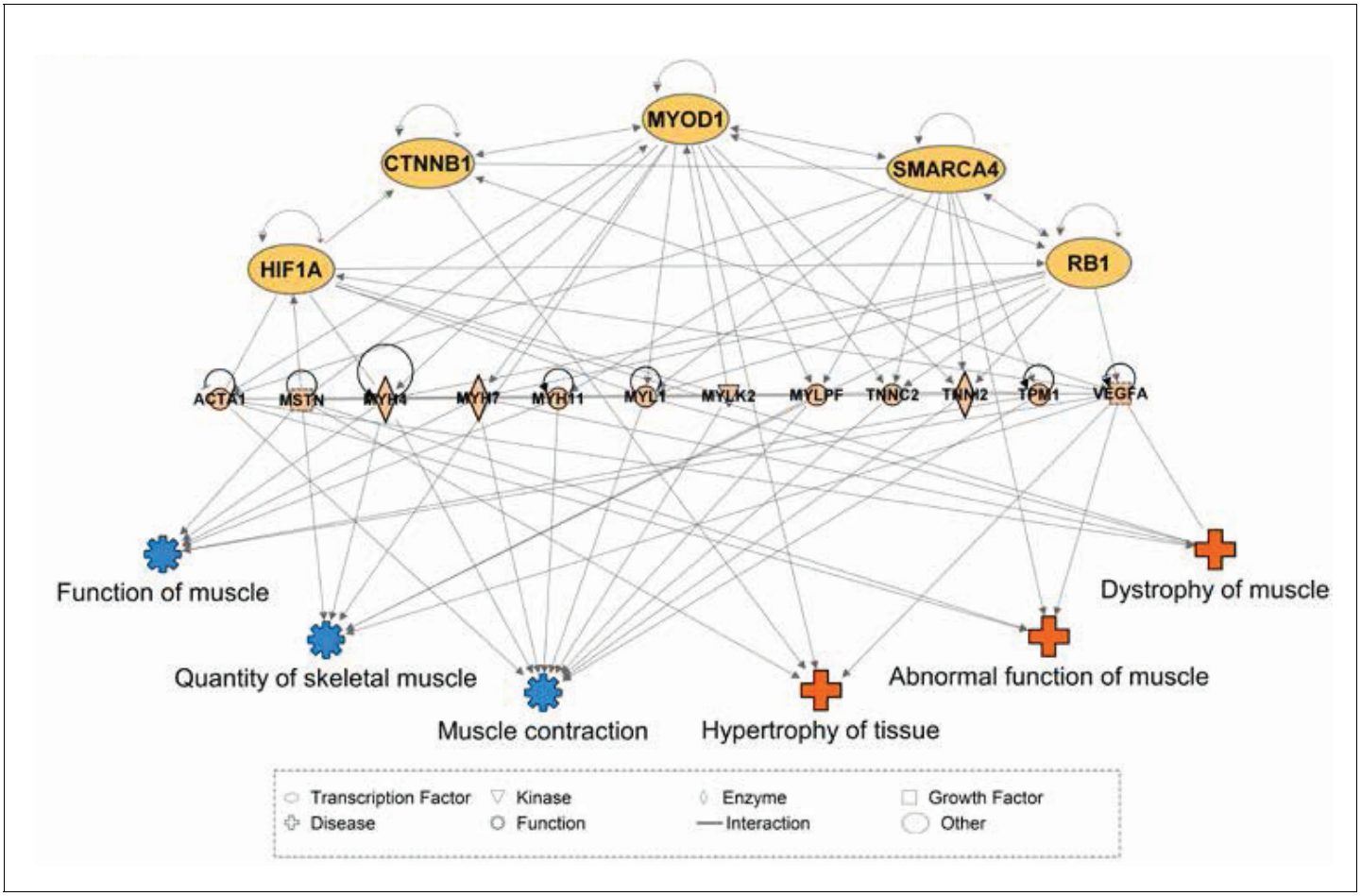



**Figure 5.** Network interactions of transcription factors predicted to be upstream of differentially expressed skeletal muscle genes. Ingenuity Pathway Analysis (IPA) was used to predict transcription factors upstream of differentially expressed muscle genes. The beige nodes (middle tier) represent muscle genes whose expression varies between muscle tissues. The orange nodes (top tier) are transcription factor genes predicted to contribute to muscle specific expression patterns. The blue and red nodes (bottom tier) represent biological functions and disease processes, respectively. Edges represent known, directional regulatory interactions. In the interest of clarity, this network has been manually trimmed to only include the most pertinent nodes. For a complete list of all predictions and regulated genes, please see ***Table 2***.

DOI: https://doi.org/10.7554/eLife.34613.018

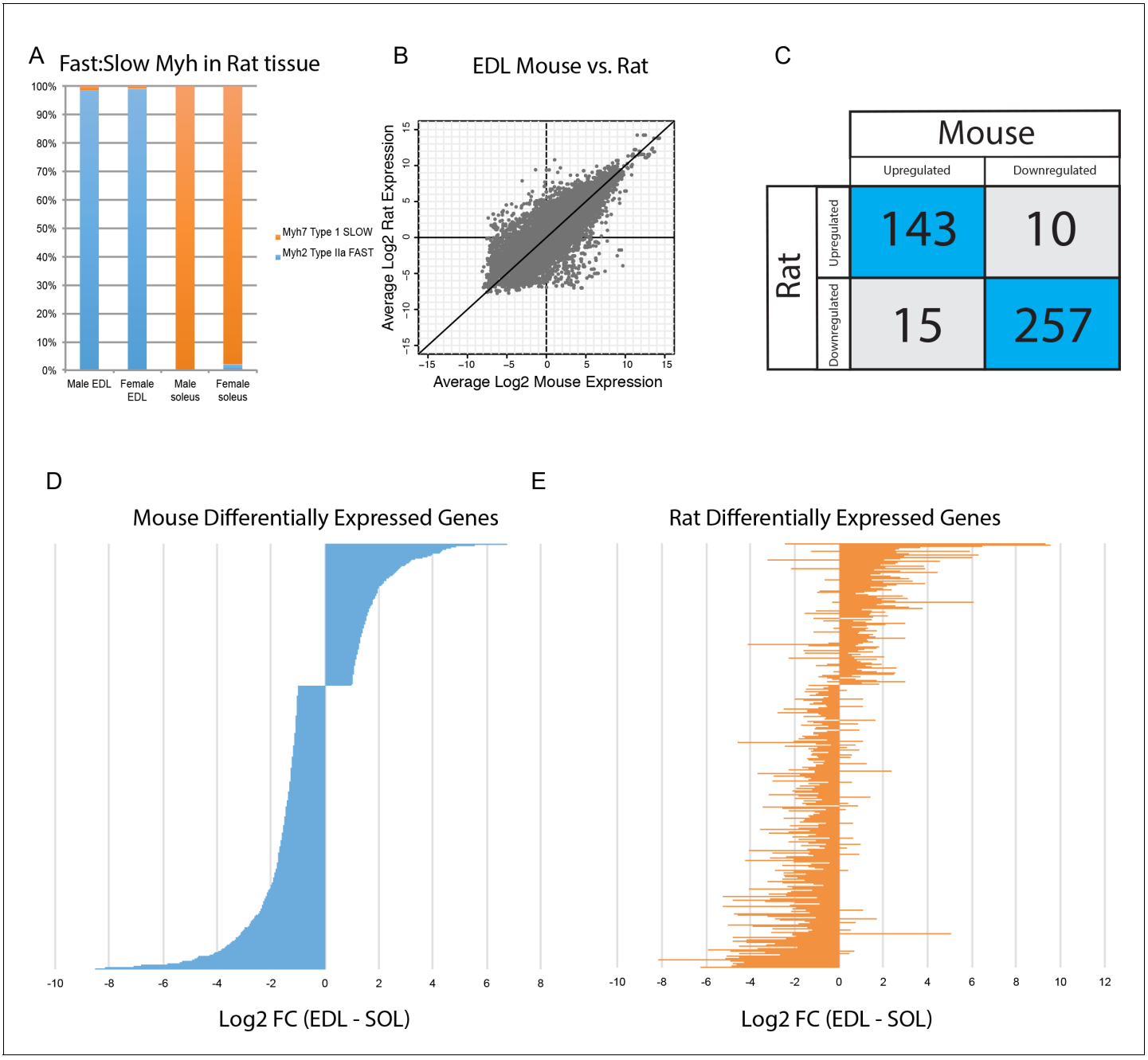



**Figure 6.** Differential gene expression is conserved between mice and rats. (**A**) The relative abundance of fast:slow *Myosin heavy chain* (*Myh*) transcripts in rat male and female *EDL* and *soleus* is shown as a bar graph. As expected, the fiber type composition of rat muscle is more homogeneous than mouse (compare with ***Figure 3B***) (**B**) Scatter plot showing the overall similarity between mouse and rat transcriptomes in *EDL* ($R^2$ = 0.637). Each dot represents a single orthologous gene shared between mice and rats. Correlation between the transcriptomes of mouse and rat *soleus* is essentially the same as for *EDL* described above ($R^2$ = 0.662). (**C**) Of the 425 genes differentially expressed between *EDL* and *soleus* in both mice and rats, the majority (94%) were differentially expressed in the same direction (i.e., up-regulated in both mice and rats or down-regulated in both mice and rats). (**D**) The fold change for all rank-ordered differentially expressed genes in mice between EDL and soleus are plotted as a bar graph (N = 691, q < 0.05, fold change >2). (**E**) The fold change (*EDL/soleus*) for the rat orthologues of the genes in (**D**) are plotted as a bar graph. The order of genes in (**D**) and (**E**) is identical. The majority of genes in rat (>90%) show differential expression in the same direction as seen in mice (i.e., up-regulated in both mice and rats or down-regulated in both mice and rats).

DOI: https://doi.org/10.7554/eLife.34613.020

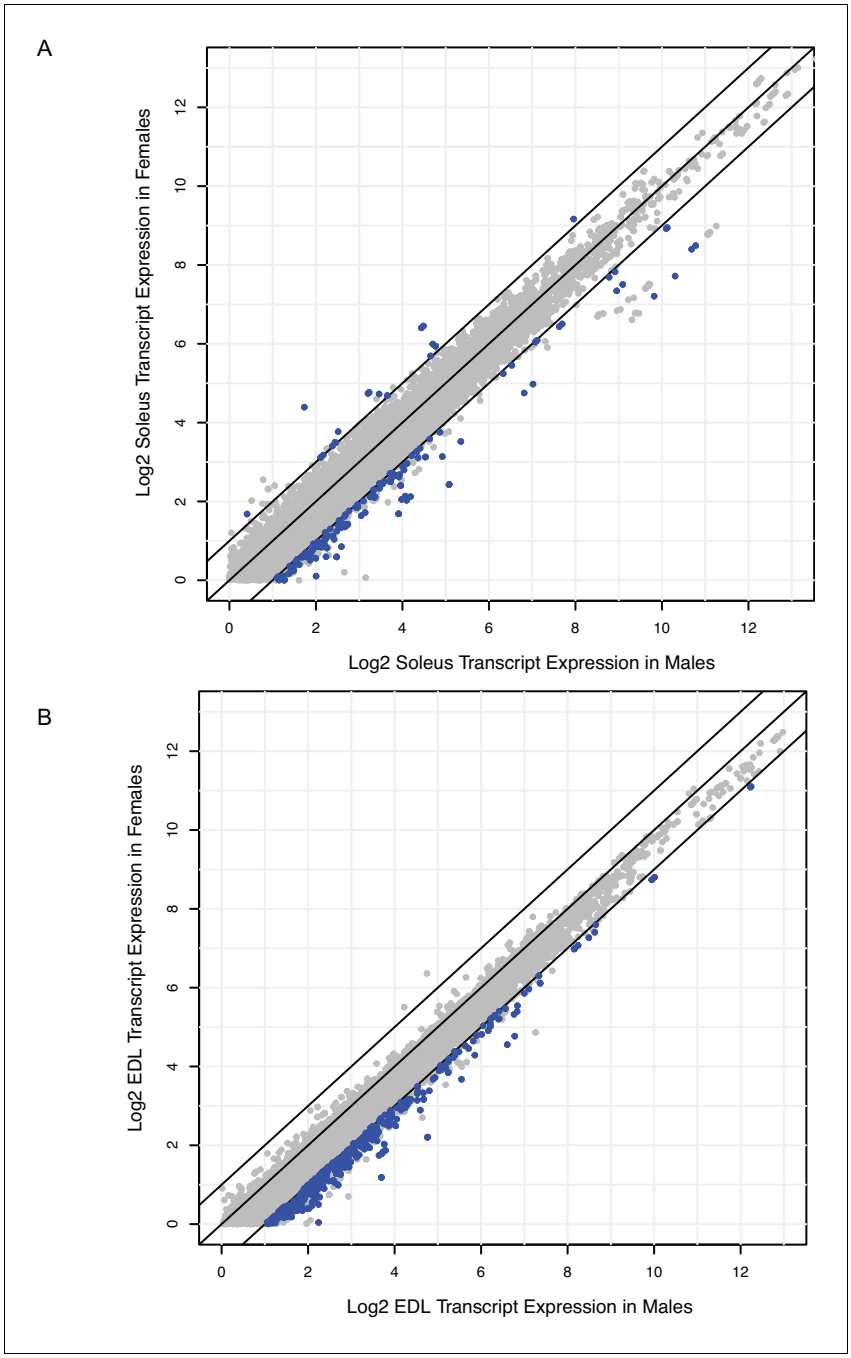



**Figure 6—figure supplement 1.** Relatively few genes are differentially expressed between males and females; most of these are up-regulated in males. (**A**) Scatter-plot showing the log2 expression (FPKM) of every transcript with average expression >1.0 in rat male *soleus* versus rat female *soleus*. Blue dots represent transcripts with q-values <0.05 and fold change >2; grey dots are transcripts that are not statistically significant. Black bars represent fold change cutoffs > 2. (**B**) Scatter-plot showing the log2 expression (FPKM) of every transcript with average expression >1.0 in rat male *EDL* versus rat female *EDL*. Blue dots represent transcripts with q-values <0.05 and fold change >2; grey dots are transcripts that are not statistically significant. Black bars represent fold change cutoffs > 2.

DOI: https://doi.org/10.7554/eLife.34613.021

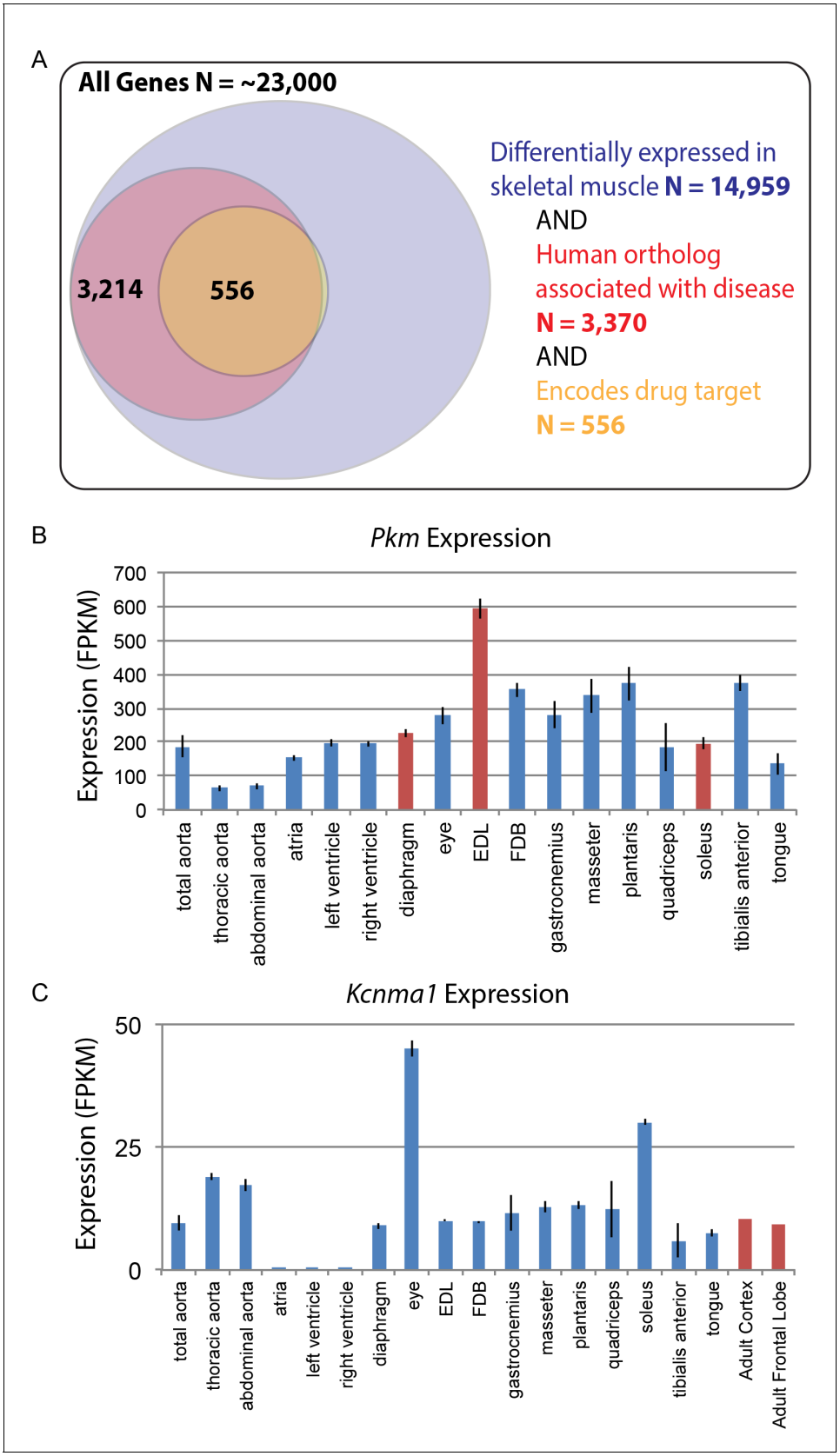



**Figure 7.** Differential gene expression in skeletal muscle may influence pathology and pharmacology. (**A**) Of roughly 23,000 genes in mice, 14,959 (~65%) are differentially expressed among skeletal muscle tissues (blue circle). Of these, 3370 (~15% of all genes) have orthologs known to influence

*Figure 7 continued on next page*

*Figure 7 continued*

human disease (red circle). Of these, 556 (>2% of all genes) encode the target of a marketed drug (orange circle). A small number of genes (N = 13) encode a drug target and are differentially expressed in skeletal muscle, but are not annotated as being associated with human disease, which explains why a sliver of the Venn diagram does not entirely overlap between red and orange circles. (**B**) The expression of *Pkm*, a key gene involved in the pathogenesis of myotonic dystrophy, is shown as a bar graph. Data points highlighted in red (*EDL*, *soleus*, and diaphragm) represent tissues that have been empirically tested for sensitivity to degeneration in a mouse model of myotonic dystrophy (***Moyer et al., 2011***). (**C**) Expression of *Kcnma1* is shown as a bar graph. *Kcnma1* encodes a potassium channel that is the molecular target of Chlorzoxazone, a drug prescribed as a muscle relaxant. Data points highlighted in red represent FPKM normalized gene expression in the adult cortex and adult frontal lobe as measured by the ENCODE consortium (***ENCODE Project Consortium, 2012***). Error bars are ±S.E.M., when available.

DOI: https://doi.org/10.7554/eLife.34613.022

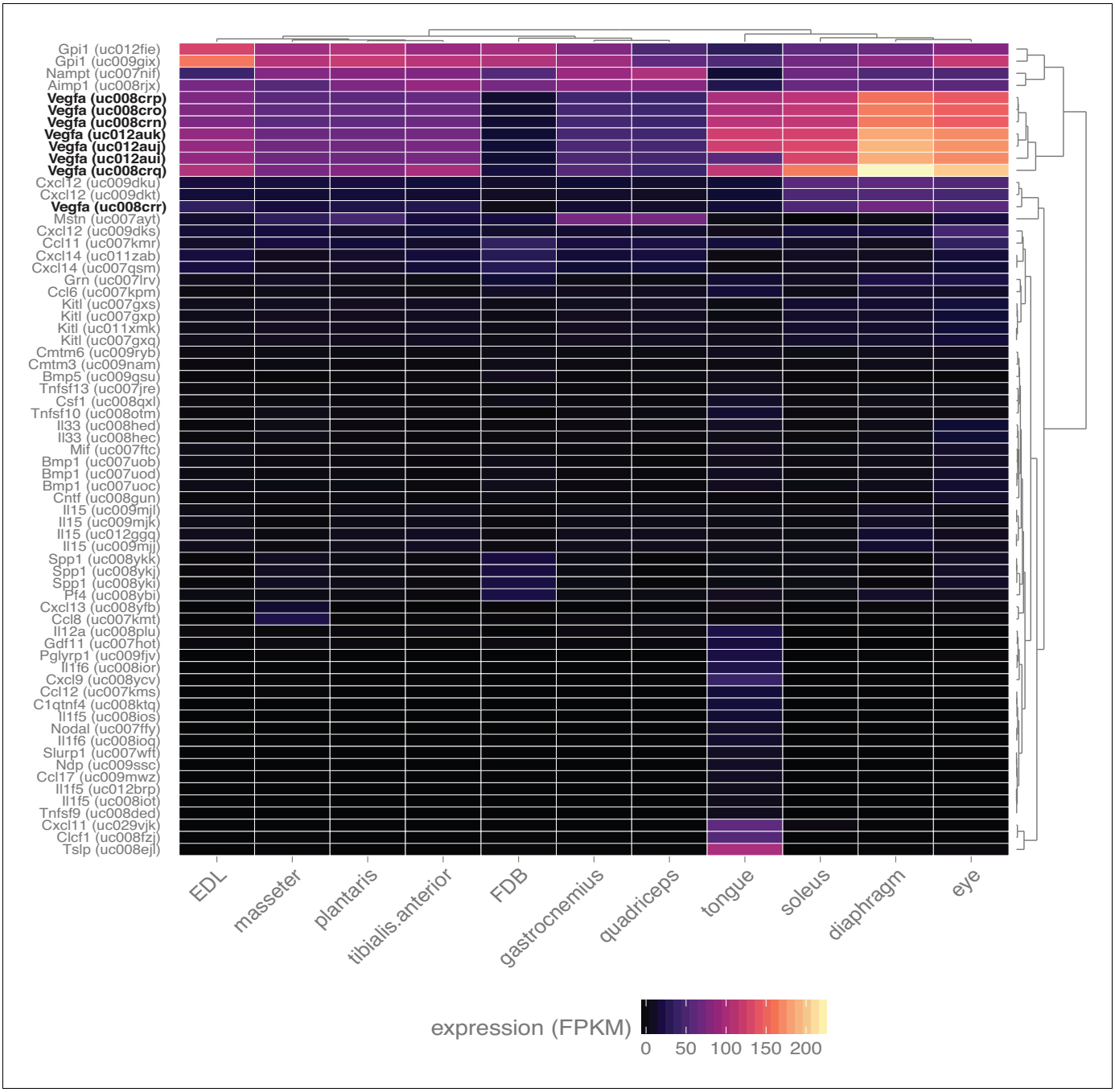



**Figure 7—figure supplement 1.** Skeletal muscle expresses numerous myokines. Heat map representation of the expression of all transcripts annotated with the gene ontology term 'cytokine activity' and filtered with a q-value threshold <0.01, minimum expression in at least one tissue >10 FPKM (N = 67 transcripts). A 10-fold higher expression threshold was chosen for this figure to enrich the gene list for highly-expressed transcripts. Dendrograms represent clustering based on similarity by tissue (top) and transcript (right side). All spliceforms of *Vegfa* (bold, black text) are expressed at notably high levels.

DOI: https://doi.org/10.7554/eLife.34613.023

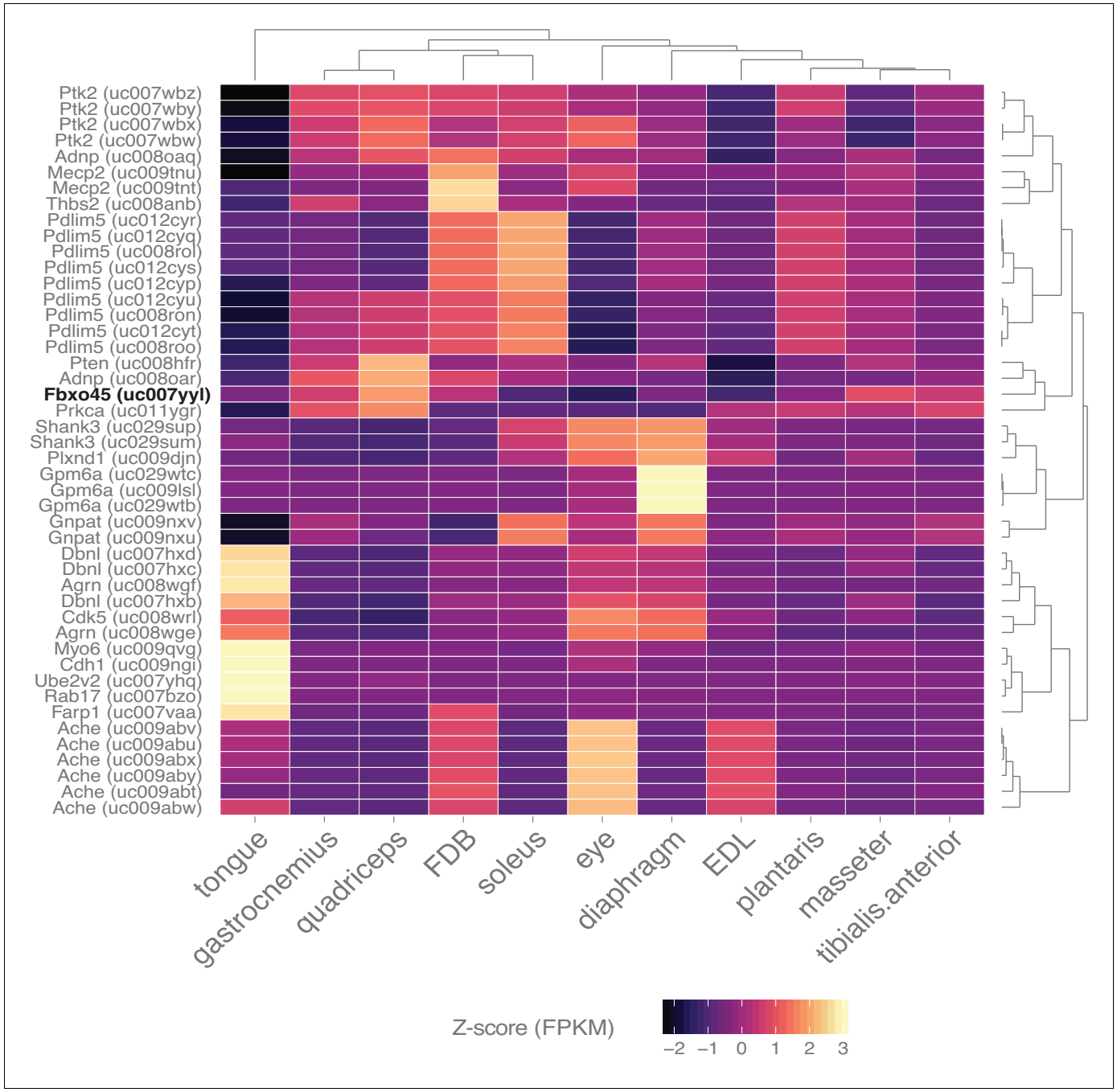



**Figure 7—figure supplement 2.** Skeletal muscle expresses many genes involved in synapse assembly. Heat map representation of the expression of all transcripts annotated with the gene ontology term 'synapse assembly' and filtered with a q-value threshold <0.01, minimum expression in at least one tissue >10 FPKM (N = 46 transcripts). A 10-fold higher expression threshold was chosen for this figure to enrich the gene list for highly-expressed transcripts. Dendrograms represent clustering based on similarity by tissue (top) and transcript (right side). *Fbxo45* (bold, black text) shows differential expression between skeletal muscles and may be involved in neuromuscular junction formation (***Saiga et al., 2009***).

DOI: https://doi.org/10.7554/eLife.34613.024