Transcriptional profiling reveals extraordinary diversity among skeletal muscle tissues

Erin E Terry[1], Xiping Zhang[2], Christy Hoffmann[1], Laura D Hughes[1], Scott A Lewis[1], Jiajia Li[1,3], Lance Riley[2], Nicholas F Lahens[4], Ming Gong[5], Francisco Andrade[5], Karyn A Esser[2], and Michael E Hughes[1,6]

1. Division of Pulmonary and Critical Care Medicine, Washington University School of Medicine, St. Louis, MO, USA
2. Department of Physiology and Functional Genomics, University of Florida College of Medicine, Gainesville, FL, USA
3. Department of Biology, University of Missouri –St. Louis, St. Louis, MO, USA
4. Institute for Translational Medicine and Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
5. Department of Physiology, University of Kentucky School of Medicine, Lexington, KY, USA
6. To whom correspondence should be addressed: michael.hughes@wustl.edu

Keywords: Skeletal Muscle, RNA-sequencing (RNA-seq), functional genomics, gene expression, myotonic dystrophy, tissue engineering, myokine, *Myosin*, *Vegfa*

Running Title: Gene expression diversity among skeletal muscles

Summary:

Skeletal muscle comprises a family of diverse tissues with highly specialized morphology, function, and metabolism. Many acquired diseases – including HIV, COPD, cancer cachexia, critical illness myopathy, and sepsis – affect specific muscles while sparing others. Even monogenic muscular dystrophies tend to selectively affect certain muscle groups, despite their causative genetic mutations being present in all tissues. These observations suggest that factors intrinsic to muscle tissues influence their susceptibility to various disease mechanisms. Nevertheless, most studies have not addressed transcriptional diversity among skeletal muscles. Here we use unbiased RNA sequencing (RNA-seq) to profile global mRNA expression in a wide array of skeletal, smooth, and cardiac muscle tissues from mice and rats. Our data set, MuscleDB, reveals extensive transcriptional diversity, with greater than 50% of transcripts differentially expressed among skeletal muscle tissues. This diversity is only partly explained by fiber type composition and developmental history, suggesting that specialized transcriptional profiles establish the functional identity of muscle tissues. We find conservation in the transcriptional profiles across species as well as between males and females, indicating that these data may be useful in predicting gene expression in related species, such as humans. Notably, thousands of differentially expressed genes in skeletal muscle are associated with human disease, and hundreds of these genes encode targets of drugs on the market today. Related to this observation, we suggest a mechanistic explanation for how myotonic dystrophy induces weakness in the *extensor digitorum longus* (*EDL*) while sparing nearby muscles. These data may therefore provide the means by which muscle-specific sensitivity to disease may be unraveled. In addition, we detect hundreds of putative myokines that may underlie the endocrine functions of skeletal muscle. We anticipate that in conjunction with transcriptional modeling, this resource will catalyze more sophisticated tissue engineering of skeletal muscle to improve the efficacy of regenerative medicine.

Introduction:

Gene expression atlases have made enormous contributions to our understanding of genetic regulatory mechanisms. The field of functional genomics was set in motion by the completion of the Human Genome Project and the coincident development of high throughput gene expression profiling technologies. The overriding goals of this field are to understand how genes and proteins interact at a whole-genome scale and to define how these interactions change across time, space, and different disease states. The development of SymAtlas was an early, deeply influential effort to address these questions (Su et al., 2002). Custom microarrays were used to systematically profile mRNA expression in dozens of tissues and cell lines from humans and mice. Besides describing tissue-specific expression patterns, these data provided essential insights into the relationship between chromosomal structure and transcriptional regulation (Su et al., 2004).

Related approaches have had a similarly high impact on biomedical research. For example, microRNA expression throughout mammalian tissues was described in an expression atlas that has revolutionized the study of regulatory RNAs (Landgraf et al., 2007). Our lab has contributed to this growing literature with the creation of CircaDB, a database of tissue-specific mRNA rhythms in mice (Hughes et al., 2009; Zhang et al., 2014). Taken together, these projects demonstrate that publically available functional genomics data have enduring value as a resource for the research community.

Nevertheless, previous gene expression atlases have largely ignored skeletal muscle. CircaDB includes gene expression data from the heart and whole calf muscle, but it does not distinguish between their constituent tissues. Similarly, SymAtlas profiles nearly 100 different tissues, but there is only a single representative sample for either cardiac or skeletal muscle. The microRNA Atlas includes over 250 human, mouse, and rat tissues; however, skeletal muscle is entirely absent from these data. More recent human gene expression atlases have similar biases (Evangelista et al., 2015; Lindholm et al., 2014; Vissing and Schjerling, 2014). Many studies have compared muscle-specific gene expression in different tissues (Porter et al., 2001), fiber-types (Chemello et al., 2011), or disease states (Chen et al., 2000; Colantuoni et al., 2001), but on the whole, there is no systematic analysis of transcriptional diversity in skeletal muscle.

This gap in the literature is problematic since skeletal muscle comprises a remarkably diverse group of tissues. Skeletal muscle groups originate from different developmental origins, and they have characteristic morphological specializations (Merrell and Kardon, 2013; Murphy and Kardon, 2011; Noden and Francis-West, 2006). Their physiological functions are similarly diversified. For example, extraocular muscles govern precise eye movements, the diaphragm drives rhythmic breathing, and limb skeletal muscles are involved in either fast bursts of motion or sustained contractions underlying posture.

These intrinsic differences contribute to differential susceptibility of muscle groups to injury and disease. For example, there are six major classes of muscular dystrophy, and each one afflicts a characteristic pattern of skeletal and cardiac muscle tissues (Ciciliot et al., 2013; Emery, 2002). Since the causative mutations underlying congenital muscular dystrophies are germline and present in all cells, this observation indicates that there are properties intrinsic to different muscle tissues that regulate their sensitivity or resistance to different pathological mechanisms.

Acquired diseases show similar specificity in which muscle tissues they affect and which they spare. Patients with chronic obstructive pulmonary disorder (COPD) typically have pronounced myopathy in their quadriceps and dorsiflexors (Barreiro and Gea, 2016; Clark et al., 2000; Gagnon et al., 2013). Critical illness myopathy, a debilitating condition caused by mechanical ventilation and steroid treatment, causes muscle weakness in limb and respiratory muscles while sparing facial muscles (Aare et al., 2011; Hermans and Van den Berghe, 2015; Latronico et al., 2012). Cancer cachexia (Acharyya et al., 2005), HIV (Serrano et al., 2008), and sepsis (Tiao et al., 1997) cause muscle wasting, typically affecting fast twitch fibers more severely than slow twitch fibers. In fact, histologically identical muscle fibers show widely divergent responses to injury and disease depending on the muscle group in which they reside (Aravamudan et al., 2006; Ciciliot et al., 2013). Taken together, these observations strongly suggest that the intrinsic diversity of skeletal muscle has important consequences for human health and disease. However, the mechanisms through which disease susceptibility and functional specialization are established are unknown.

Here we present the first systematic examination of transcriptional programming in different skeletal muscle tissues. We find that more than 50% of transcripts are differentially expressed among skeletal muscle tissues, an observation that cannot be explained by fiber type composition or developmental history alone. We show conservation of gene expression profiles across species and sexes, suggesting that these data may reveal conserved functional elements relevant to human health. Finally, we discuss how this unique data set may be applied to the study of disease, particularly regarding muscular dystrophy and regenerative medicine.

Results:

To determine which skeletal muscle tissues are of the broadest general interest, we distributed a Google poll to leading investigators in the skeletal muscle field. Having recorded over 100 individual responses, we selected 11 mouse skeletal muscle tissues (**Table 1**) for study in order to span the functional, developmental, and anatomical diversity of skeletal muscles (**Figure S1**). To hedge against selection bias from the investigators asked to vote in the poll, we cross-correlated these results with papers indexed in NCBI's PubMed (**Figure S1**). In addition to these 11 mouse skeletal muscle tissues, we also identified representative smooth and cardiac muscle tissues for collection from mouse, and two skeletal muscle tissues (*EDL* and *soleus*) from male and female rats to permit inter-species and inter-sex comparisons (**Table 1**).

Adult mice and rats were sacrificed and whole muscle tissues were dissected to include the entire muscle body from tendon to tendon. Each sample included six biological replicates from independent animals. RNA was purified from these tissues, and RNA-seq was used to measure global gene expression (**Methods and Table S1**). On average, every muscle sample was covered by greater than 200 million aligned short nucleotide reads, for a grand total of over 4.4 billion aligned reads in the entire data set (**Table 1**). Empirical simulations indicate that this experimental design reaches saturation with respect to identification of expressed exons (**Figure S2**). Pairwise comparisons of each replicate sample further indicate a high degree of reproducibility in expression level measurements among biological replicates (median R^2 value > 0.93, **Figure S3**).

Over 80% of transcripts encoded in the genome are expressed in at least one skeletal muscle (**Figure 1A**). Comparing the transcripts expressed in all tissues identifies a core group of ~21,000 transcripts found in every skeletal, cardiac, and smooth muscle (**Figure 1B**). Presumably, these mRNAs include the minimal set of genes required for a cell to generate contractile force. Differential expression analysis shows that even at extremely stringent ($q < 10^{-6}$) statistical thresholds, 55.2% of mouse transcripts are differentially expressed among skeletal muscle tissues (**Figure 1C**). Phrased differently, over half of all transcripts are statistically different when comparing mRNA expression among the 11 mouse skeletal muscles in this study. To validate a

subset of these data, we selected 10 genes differentially expressed between *EDL* and *soleus* across a range of different fold changes and performed quantitative PCR (qPCR) on independent biological samples. All 10 replicated, and manual examination of internal controls for cardiac and smooth muscle agreed with *a priori* expectations (**Figure S4**). To explore similarity between tissues, we calculated a pairwise Euclidean distance between every tissue (**Figure S5**). From these data, we generated a dendrogram that clusters tissues based on overall transcriptome similarity (**Figure 1D**). Smooth and cardiac tissues clustered together as expected, and skeletal muscles were found in two primary clusters, presumably based on the proportion of fast twitch fibers they include (i.e. *Myosin heavy chain 4* (*Myh4*) expression). We then calculated the proportion of differentially expressed transcripts in every pairwise tissue comparison (**Figure 1E**). Some surprising observations emerged from this analysis. For example, *masseter*, a head/neck muscle, is more similar to limb muscles like *EDL* than muscles that share a similar developmental history, such as the tongue. In contrast, the *flexor digitorm brevis* (*FDB*), a muscle necessary for flexing the toes, is more similar to extraocular muscles than limb muscles. On average, 13% of transcripts are differentially expressed between any two skeletal muscles, with a maximum of 36.5% and a minimum less than 1%.

Skeletal muscle is comprised of myofibers which are typically classified into one of four types in mice based on their *Myh* expression (Haizlip et al., 2015). We used our data to quantify the relative abundance of different fiber types across skeletal muscle tissues (**Figure 2A**). These results were cross-correlated with legacy data from histological studies, showing close agreement (**Figure S6**). One hypothesis is that fiber type composition (i.e. the relative amounts of fast versus slow twitch fibers) establishes the observed diversity of gene expression profiles. To test this, we clustered skeletal muscle tissues based on similarity of *Myh* expression. The resulting dendrograms are grossly similar to those based on global gene expression (**Figure 2B**; compare with **Figure 1D**), as predominantly fast twitch muscles expressing high levels of *Myh4* (Type IIB fibers) tend to cluster together. Nevertheless, clustering within the two major skeletal muscle groups reveals important differences, such as the high similarity of diaphragm and *FDB* based on *Myh* expression, but their dissimilarity based on global gene

expression (**Figure 1E**). Similarly, the clustering of tongue and extraocular eye muscles varies considerably depending on whether *Myh* or global gene expression establishes the pairwise Euclidean distance.

To explore these observations further, we made use of a single-fiber microarray study that identified genes enriched in slow (Type I) versus fast twitch (Type IIB) skeletal muscle fibers (Chemello et al., 2011). These legacy data reveal that *Myostatin* (*Mstn*) is almost exclusively expressed in fast twitch fibers. Therefore, if fiber type composition determines gene expression patterns, we would expect close correlation between *Mstn* and *Myh4*. In actuality, there is only moderate correlation ($R^2 = 0.47$, **Figure 2C**). Expanding on this result, we find weak correlation between genes enriched in fast twitch fibers and *Myh4* (median $R^2 = 0.132$, **Figure 2D**). The low correlation between fiber type composition and gene expression signatures is also seen in slow twitch fibers (**Figure 2E and F**). Taken together, we conclude that fiber type based on *Myh* expression contributes to tissue-specific gene expression but is insufficient to establish the diversity of transcriptional patterns we observe.

Alternatively, developmental history may play an essential role in defining gene expression patterns in adult skeletal muscle. *Hox* genes are a family of transcription factors that establish anterior/posterior patterning and skeletal muscle specification during development (Krumlauf, 1994). We therefore examined the expression of *Hox* genes in muscle tissues (**Figure 3A**). Clustering of tissues based on *Hox* gene expression reveals that the head and neck muscles cluster more closely with cardiac tissues than other skeletal muscles. Similarly, the diaphragm is more similar to the aorta than limb skeletal muscle. In short, *Hox* gene expression in muscle recapitulates the developmental history of these tissues with respect to anterior/posterior axis formation. However, as these clusters are significantly different from those seen when comparing whole transcriptome expression patterns (**Figure 1D**), we conclude that *Hox* genes are insufficient to explain the mRNA diversity among adult skeletal muscle tissues.

Similar to *Hox* family members, many developmentally significant genes maintain expression in adult muscle tissues. These include the myogenic regulatory factors *Myod1*, *Myf5*, *Myog*, and *Myf6*, which are expressed in, and exclusive to, skeletal muscle (**Figure 3B**). *Hotair* is a noncoding RNA that regulates *Hox* gene expression in

development (Tsumagari et al., 2013); as expected, it is expressed in all limb muscles (**Figure 3C**). In contrast, *Pitx2* is a transcription factor with a key role in driving extraocular muscle development (Zhou et al., 2012), and *Lhx2* is a transcription factor important for *masseter* muscle development (Buckingham and Rigby, 2014). The expression of both genes is consistent with regulatory activity that continues into adulthood and may contribute to maintaining proper mRNA expression patterns (**Figure 3D and E**). We speculate that computational modeling of transcription factor networks may resolve the mechanisms underlying tissue-specific mRNA profiles, particularly since neither developmental history nor fiber type composition are sufficient to explain the diversity of the skeletal muscle transcriptome.

In fact, there are a considerable number of transcripts expressed in an entirely tissue-specific fashion in skeletal muscle. Plotting tissue specificity versus average expression reveals a bimodal distribution among all transcripts (**Figure 4A and B**). The majority of transcripts in skeletal muscle are expressed evenly across most skeletal muscle tissues, consistent with the notion that there is a minimal set of genes required to form sarcomeres and generate contractile force. Nevertheless, a smaller set of transcripts (~5%) show nearly exclusive expression in a single skeletal muscle tissue. Manual examination of these tissue-specific genes reveals that many are found in head and neck muscles that have the most divergent developmental history of tissues in this study. Nonetheless, we found examples of tissue-specific genes in the diaphragm (**Figure 4C**) and the *FDB* (**Figure 4D**). These data could therefore be used to generate transgenic mice with tissue-specific genetic manipulations, allowing follow-up studies to explore the mechanisms underlying muscle specialization.

These expression data reveal that the skeletal muscle transcriptome is considerably more complex than previously appreciated. Moreover, the unprecedented depth of sequencing coverage in muscle permits the discovery of heretofore-unannotated transcripts. Manual examination of “gapped reads” spanning two or more loci in the genome indicates that hundreds to thousands of splicing events occur in muscle that have not been previously described (**Figure S7A**). The majority of these novel splicing events result in either inclusion of novel exons or the exclusion of exons previously thought to be constitutive (**Figure S7B**). To validate this observation, we

designed oligonucleotide primers (**Figure S7C**) specific for two novel exons of *Myosin light chain kinase 4* (*Mylk4*), a gene with high expression (>200 FPKM) in many skeletal muscle tissues. PCR amplification and molecular cloning confirmed that these putative exons are bonafide components of the mature *Mylk4* transcript in two different skeletal muscle tissues, *EDL* and *soleus* (**Figure S7D**).

There are significant differences in fiber type composition between analogous muscle tissues of different mammals. For example, mouse *soleus* is a mixture of fast- and slow fibers (**Figure 2A**), while rat *soleus* is almost entirely slow twitch (**Figure 5A**). Because of this, we asked whether gene expression differences in mice are conserved across species. Expression of orthologous genes in mouse versus rat tissues has moderate to high levels of correlation ($R^2 > 0.6$, **Figure 5B**), despite the difference in fiber type composition noted above. Moreover, the vast majority of genes differentially expressed between *EDL* and *soleus* in both mice and rats changed in the same direction (**Figure 5C**). Taking this observation one step further, the fold change of all genes differentially expressed in mice *EDL* compared to *soleus* (**Figure 5D**) is largely consistent with the fold change of same genes in rat *EDL* versus *soleus* (**Figure 5E**). Sex differences did not dramatically influence differential gene expression between *EDL* and *soleus* (**Figure S8**). Fewer than 3% of transcripts were differentially expressed between male and female rat *EDL* (2.7%) and male and female rat *soleus* (1.9%). Of these differentially expressed genes, most are up-regulated in males. This observation agrees with previous studies (Roth et al., 2002) and is consistent with the possibility that androgen response elements influence sex-specific gene expression differences in skeletal muscle. Taken as a whole, these data indicate that differentially expressed genes are largely conserved between mice and rats and suggest that these data may be used to predict gene expression in related species.

Based on this potential conservation to humans, we speculate that genes associated with disease and those encoding drug targets will be of particular importance to follow-up studies. Of the roughly 23,000 genes encoded by the mouse genome, over 50% are differentially expressed among mouse skeletal muscle tissues. Of these, 3,370 differentially expressed genes have human orthologs associated with disease, and 556 of those encode molecular targets of drugs on the market today (**Figure 6A**). These

genes may contribute to the molecular mechanisms underlying differential disease susceptibility and pharmaceutical sensitivity in skeletal muscle tissues.

To give one example of differential disease susceptibility, *pyruvate kinase* (*Pkm*) Is involved in the mechanism of myotonic dystrophy (Gao and Cooper, 2013), and is dramatically up-regulated in *EDL* compared to all other muscle groups (**Figure 6B**). Myotonic dystrophy disrupts normal splicing and pathologically elevates *Pkm*, which in turn disrupts normal metabolism, decreasing oxygen consumption and increasing glucose consumption. Based on these observations, elevated expression of *Pkm* in adult tissues is a critical step in the pathology of myotonic dystrophy (Gao and Cooper, 2013). Since *Pkm* is expressed dramatically higher in *EDL* compared to all other muscle types (**Figure 6B**), we predict that *EDL* would be more sensitive to degeneration than other muscles. Although testing this hypotheses requires suitably controlled genetic and biochemical experiments, we note with great interest that in mouse models, *EDL* is considerably more susceptible to muscle weakness than either diaphragm or *soleus* (Moyer et al., 2011).

In addition to disease susceptibility, these data may help explain differential drug sensitivity in muscle. To give one example, the drug Chlorzoxazone (brand name: Lorzone) is used to treat muscle spasms. It is thought to act on the central nervous system by regulating a potassium channel encoded by the gene *Kcnma1* (Dong et al., 2006). Although *Kcnma1* is expressed at high levels in the central nervous system (ENCODE Project Consortium, 2012), it is also found at comparable levels in most skeletal muscle tissues (**Figure 6C**). The two exceptions are extraocular eye muscles and the *soleus* that have greater than three-fold higher *Kcnma1* expression than other muscle tissues. The differential abundance of Chlorozoxazone's target could impact either the efficacy of this drug or the severity of its side effects in different muscles. Given that some commonly prescribed drugs, such as glucocorticoids, cause muscle wasting in specific subsets of muscle tissues (Schakman et al., 2008), we speculate that this data set will be a valuable resource for exploring the mechanisms underlying differential drug sensitivity among skeletal muscles.

Skeletal muscles are endocrine tissues, expressing numerous secreted cues that influence the physiology and metabolism throughout the body. Termed "myokines",

these signaling molecules have important impacts on human health and disease (Lightfoot and Cooper, 2016). The total number of myokines is currently unknown. Our data contribute to this field by comprehensively defining candidate myokines for future study (**Figure S9** and **Table S2**). Dozens of genes encoding secreted proteins are differentially expressed among skeletal muscles at relatively high levels (N = 42 unique genes, FPKM > 10). One notable example is *Vegfa* (q-value ~ 10^{-30}), a gene involved in angiogenesis, cardiac disease, cancer progression, and many other normal and pathological processes (Smith et al., 2015). Average expression of its predominant isoform is in the top 98th percentile of all transcripts expressed by skeletal muscle. Moreover, the serum concentration of VEGF-A in normal healthy adults is considerably greater than serum levels of IL6, a canonical myokine (Lightfoot and Cooper, 2016). These data suggest that *Vegfa* expression in skeletal muscle has endocrine as well as paracrine functions. Moreover, maximal expression of *Vegfa* in the diaphragm is nearly 10-fold greater than its minimal expression in the FDB, indicating that there are muscle-specific mechanisms for regulating *Vegfa* levels in particular and myokine levels in general.

Regenerative medicine has made considerable progress in generating skeletal muscle from stem cells (Qazi et al., 2015). Nevertheless, engineered tissues have important deficiencies in generating sufficient force and forming appropriate neuromuscular synapses (Juhas and Bursac, 2013). Given the extensive diversity observed among skeletal muscle tissues, we speculate that the inability to form proper synaptic connections in engineered tissue may be due to the expression of inappropriate or incomplete transcriptional programs. In other words, differentiating stem cells into generic skeletal muscle may not recapitulate all the cues necessary for muscle-specific synapse formation. Therefore, we examined the differential expression of genes implicated in synapse assembly (**Figure S10**). Dozens of transcripts involved in synapse formation are expressed at relatively high levels (FPKM > 10), suggesting that they are bonafide skeletal muscle mRNAs, rather than contamination from nearby neurons. One illustrative example, *Fbxo45*, is an E3 ligase involved in synapse formation, and mouse knock-outs have disrupted neuromuscular junction (NMJ) formation in the diaphragm, resulting in early lethality (Saiga et al., 2009). Moreover, the

C. elegans ortholog, *FSN-1*, also disrupts NMJ formation, with some synapses being over-developed while others are under-developed (Liao et al., 2004). We speculate that this protein and its orthologs may play a conserved role in tissue-specific NMJ assembly.

Discussion:

General textbook discussions of skeletal muscle typically focus on developmental patterning, neuromuscular synapse physiology, or the biophysics of contractile properties (Alberts, 2014). This reflects the generally held belief that adult skeletal muscle is interesting only as a mechanical output of the nervous system. Functional genomics studies have internalized this assumption to the extent that every gene expression atlas generated to date has selected at most one skeletal muscle as representative of the entire family of tissues. As such, the null hypothesis of this study was that gene expression profiles would be largely similar among skeletal muscle tissues.

However, as more than 50% of transcripts are differentially expressed among skeletal muscles (**Figure 1C**), and 13% of transcripts are differentially expressed between any two skeletal muscle tissues on average (**Figure 1E**), the data are entirely inconsistent with the null hypothesis. These results indicate that there is no such thing as a representative skeletal muscle tissue. Instead, skeletal muscle should be viewed as a family of related tissues with a common contractile function, but widely divergent physiology, metabolism, morphology, and developmental history. Based on these antecedents, it should come as no surprise that the transcriptional programs maintaining skeletal muscle specialization in adults are highly divergent as well.

This study is the first systematic examination of transcriptome diversity in skeletal muscle. At greater than 200 million aligned short nucleotide reads per tissue and six biological replicates apiece, this data set is unprecedented in its accuracy and reproducibility (see **Figures 3, S2, S4,** and **S6**). Moreover, the depth of sequencing allows the detection of previously unannotated transcripts that may play a role in muscle physiology (see **Figures S3** and **S7**). To assist future studies, we have made our data resource freely available (https://muscledb.org).

We acknowledge that many processes besides steady-state mRNA levels regulate protein expression. Nevertheless, studies show close agreement between transcriptional and proteomic profiling even in highly dynamic settings (Robles et al., 2014). Moreover, the larger dynamic range of mRNA measurements permits a more comprehensive description of expression profiles than would be possible with existing

proteomic technology. We further acknowledge that the samples collected and analyzed in this study are bulk tissues rather than single-cell preparations, and that contaminating tissues such as vasculature and immune cells may influence some gene expression measurements. However, we see agreement between our data and the few single-cell profiling papers in the literature (Chemello et al., 2011), indicating that this potential bias is unlikely to confound the major observations herein. Furthermore, we emphasize that the majority of studies in the field use bulk tissues rather than single cell preparations. As such, our experimental design yields the greatest possible consistency with previous and future studies.

Acquired and genetic diseases show remarkable selectivity in which muscles they affect and which muscles they spare. For example, Duchenne muscular dystrophy severely affects the diaphragm and proximal limb extensors, while oculopharyngeal dystrophy causes weakness in the neck, facial, and extraocular muscles (Emery, 2002). At present, there is no satisfying explanation for how this occurs. A reasonable hypothesis is that intrinsic properties of muscle cells, such as gene expression, determine their sensitivity to different pathological mechanisms. These data are a starting point for future studies on how specialized transcriptional programs in muscles are maintained and how they ultimately influence disease. As an illustrative example, we note that the naturally elevated expression of *Pkm* in *EDL* may in part explain how this muscle is most dramatically affected in mouse models of myotonic dystrophy (**Figure 6**). Related to this, the use of engineered skeletal muscles in regenerative medicine has been limited by poor contractile properties and inadequate neuromuscular synapse formation (Juhas and Bursac, 2013). We speculate that this deficiency may be corrected if engineered tissues are programed to express appropriate synaptic assembly proteins (**Figure S10**).

Finally, skeletal muscle is an endocrine organ that regulates many normal and pathological processes, including sleep, bone health, diabetes, cancer, and cardiovascular disease (Giudice and Taylor, 2017; Iizuka et al., 2014; Karsenty and Olson, 2016). At roughly 40% of an adult human's body weight, skeletal muscle has an enormous capacity to influence other tissues through the expression of local or systemic signaling molecules. This study reveals extensive differential expression of putative

myokines with largely unexplored functional significance (**Figure S9**). As such, we predict these data will be instrumental in future studies of the endocrine mechanisms through which skeletal muscle regulates health and disease.

Methods:

Animal care and tissue collection: Adult male *C57Bl6J* mice were acquired from Jackson Laboratories at ten weeks of age. They were housed in light tight cages with a 12L:12D light schedule for four weeks with water and normal chow *ad libitum*. At 14 weeks of age, mice were sacrificed at between two and five hours after lights-on (i.e. ZT 2-5), and muscle tissues were rapidly dissected and flash frozen in liquid nitrogen for subsequent purification of total RNA. Adult male and female *Sprague Dawley* rats were obtained from Charles River (Wilmington, MA) at 12 weeks of age. Rats were housed in pairs with a 12-h:12-h light/dark cycle, and standard rat chow and water were provided ad libitum. At 14 weeks of age, rats were sacrificed at between two and five hours after lights-on (i.e. ZT 2-5), and muscle tissues were rapidly dissected and flash frozen in liquid nitrogen for subsequent purification of total RNA. All animal procedures were conducted in compliance with the guidelines of the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC) and were approved by the Institutional Animal Care and Use Committee at University of Kentucky.

RNA purification and library preparation: Between 5-20 mg of frozen tissue were manually homogenized in Trizol reagent (Invitrogen), and total RNA was purified using a standard chloroform extraction. RNA samples for gene expression analysis were mixed with an equal volume of 70% ethanol and further purified with RNEasy columns (Qiagen) using the manufacturer's protocol. RNA was purified from tissue collected from individual mice; samples from three individual mice were then pooled together in equimolar amounts for further analysis.

To assess RNA integrity, aliquots of each sample were denatured for 2 minutes at 70°C and analyzed on the Agilent 2100 Bioanalyzer using Eukaryote Total RNA Nano chips according to manufacturer's protocol. RNA integrity numbers (RINs) for all samples were above 8.0 with a median RIN of 9.2. Libraries were prepared using the Illumina Truseq Stranded mRNA LT kit using single end indexes according to manufacturer's protocol. Approximately 500 ng of total RNA was used as starting material and amplified with 13 cycles of PCR. Libraries were validated for size and

purity on the Agilent 2100 Bioanalyzer using DNA 1000 chips according to manufacturer's protocol.

RNA-sequencing and analysis: Pilot runs to verify library integrity were sequenced on an Illumina MiSeq (University of Missouri—St. Louis), and subsequent sequencing was performed on an Illumina HiSeq 2500 (University of Michigan) or HiSeq 3000 (Washington University in St. Louis). All sequenced reads were 50bp, single-end. Raw reads were aligned to the genome and transcriptome of *Mus musculus* (build mm10) or *Rattus norvegius* (build rn5) using RNA-seq Unified Mapper (RUM) (Grant et al., 2011) with the following parameters: *"--strand-specific --variable-length-reads --bowtie-nu-limit 10 --nu-limit 10*". Mouse and rat gene models for RUM alignments were based on UCSC gene models updated as of December 2014. 65%-92% of reads uniquely mapped to the genome/transcriptome, and the total number of aligned reads (including unique and non-uniquely aligned reads) was at least 94.7% for every replicate sample (**Table S1**). FPKM values for each transcript/exon/intron were calculated by RUM using strand-specific unique reads normalized to the total number of reads uniquely aligned to the nuclear genome (i.e., uniquely aligned mitochondrial reads were excluded from the denominator). R-squared values of log-transformed FPKMs between biological replicates of the same tissue were generally ~0.94 (**Figure S2**), and internal controls (e.g. **Figure S4**) were used to verify the biological validity of these measurements.

Differential expression between tissues was determined by one-way ANOVA of log-transformed FPKM values and adjusted for multiple testing using a Benjamini-Hochberg q-value (Hochberg and Benjamini, 1990). Unless otherwise noted, a q-value less than 0.01 and fold-change greater than 2.0 among transcripts expressed with an FPKM > 1.0 was deemed statistically significant. Disease and drug target associations were identified using public data sets (DrugBank v5.0.9 and DisGeNET). Mouse genes were mapped to human orthologs using the BiomaRt package for R (Durinck et al., 2009).

Validation of novel splicing events: Independent biological replicates of mouse *EDL* and *soleus* tissues were collected as described above. Total RNA was purified and

integrity was verified as described above. Reverse transcription reactions were performed with 500ng starting material using the manufacturer's protocol (TaqMan Fast Universal PCR Master Mix, Applied Biosystems). PCR reactions were performed using 1ul of the reverse transcription reaction using the manufacturer's protocol (Clontech Takara PCR kit). PCR products were visualized on a 1% agarose gel using conventional methods. The primary PCR products for *EDL* and *soleus* (**Figure S7D**) from primers Exon 2-> Exon 2.1 and Exon2 -> Exon3 were excised, purified, and TOPO cloned using the manufacturer's protocol (TOPO TA, Thermo Fisher). Cloned fragments were sequenced using conventional Sanger methods. All PCR primers were ordered from IDT; primer locations are described in **Figure S7C**. Primer sequences as follows: Exon 2 – AGGATCTCAGATTTGCTCACG, Exon 2.1 – GGATCCACTTTCCAGAATGC, Exon 2.2 – CATCTTTGCACCTGCATTC, Exon 3 – TATGGTCCAACCGTGCACTA, CtrlF – AGTGTGGGCGTCATCACAT, CtrlR – GTGGAGCTTGTGGTCTGACA.

Data Availability: All raw data are available on NCBI's Gene Expression Omnibus (accession number: GSE100505), and transcript-level expression values can also be downloaded from MuscleDB (http://muscledb.org/), a web application built using the ExpressionDB platform (Hughes et al., 2017).

Acknowledgements:

We thank Jeanne Geskes, Robert Lyons and the University of Michigan DNA sequencing core facility for assistance with next-generation sequencing. We thank John Hogenesch (Cincinnati Children's Hospital), Jeff Haspel (WashU Deptartment of Medicine), Patty Parker (UMSL) and members of the Hughes and Esser laboratories for helpful discussion and technical support throughout this project. We thank Dr. Michael Chicoine (WashU Department of Neurosurgery) for inestimable contributions without which this paper would never have been written. Work in the Gong lab is supported by NIH award NIH HL106843. Work in the Esser lab is supported by NIH award R01AR066082. Work in the Hughes Lab is supported by NIH award R21AR069266 and start-up funds from the Department of Medicine at Washington University in St. Louis. We thank the Genome Technology Access Center in the Department of Genetics at Washington University School of Medicine for help with genomic analysis. The Center is partially supported by NCI Cancer Center Support Grant #P30 CA91842 to the Siteman Cancer Center and by ICTS/CTSA Grant# UL1 TR000448 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. This publication is solely the responsibility of the authors and does not necessarily represent the official view of NCRR or NIH.

References:

Aare, S., Ochala, J., Norman, H.S., Radell, P., Eriksson, L.I., Göransson, H., Chen, Y.-W., Hoffman, E.P., and Larsson, L. (2011). Mechanisms underlying the sparing of masticatory versus limb muscle function in an experimental critical illness model. Physiol. Genomics *43*, 1334–1350.

Acharyya, S., Butchbach, M.E.R., Sahenk, Z., Wang, H., Saji, M., Carathers, M., Ringel, M.D., Skipworth, R.J.E., Fearon, K.C.H., Hollingsworth, M.A., et al. (2005). Dystrophin glycoprotein complex dysfunction: a regulatory link between muscular dystrophy and cancer cachexia. Cancer Cell *8*, 421–432.

Alberts, B. (2014). Molecular Biology of the Cell (New York, NY: Garland Science).

Aravamudan, B., Mantilla, C.B., Zhan, W.-Z., and Sieck, G.C. (2006). Denervation effects on myonuclear domain size of rat diaphragm fibers. J. Appl. Physiol. Bethesda Md 1985 *100*, 1617–1622.

Barreiro, E., and Gea, J. (2016). Molecular and biological pathways of skeletal muscle dysfunction in chronic obstructive pulmonary disease. Chron. Respir. Dis. *13*, 297–311.

Buckingham, M., and Rigby, P.W.J. (2014). Gene regulatory networks and transcriptional mechanisms that control myogenesis. Dev. Cell *28*, 225–238.

Burkholder, T.J., Fingado, B., Baron, S., and Lieber, R.L. (1994). Relationship between muscle fiber types and sizes and muscle architectural properties in the mouse hindlimb. J. Morphol. *221*, 177–190.

Chemello, F., Bean, C., Cancellara, P., Laveder, P., Reggiani, C., and Lanfranchi, G. (2011). Microgenomic analysis in skeletal muscle: expression signatures of individual fast and slow myofibers. PloS One *6*, e16807.

Chen, Y.W., Zhao, P., Borup, R., and Hoffman, E.P. (2000). Expression profiling in the muscular dystrophies: identification of novel aspects of molecular pathophysiology. J. Cell Biol. *151*, 1321–1336.

Ciciliot, S., Rossi, A.C., Dyar, K.A., Blaauw, B., and Schiaffino, S. (2013). Muscle type and fiber type specificity in muscle wasting. Int. J. Biochem. Cell Biol. *45*, 2191–2199.

Clark, C.J., Cochrane, L.M., Mackay, E., and Paton, B. (2000). Skeletal muscle strength and endurance in patients with mild COPD and the effects of weight training. Eur. Respir. J. *15*, 92–97.

Colantuoni, C., Jeon, O.H., Hyder, K., Chenchik, A., Khimani, A.H., Narayanan, V., Hoffman, E.P., Kaufmann, W.E., Naidu, S., and Pevsner, J. (2001). Gene expression profiling in postmortem Rett Syndrome brain: differential gene expression and patient classification. Neurobiol. Dis. *8*, 847–865.

Dong, D.-L., Luan, Y., Feng, T.-M., Fan, C.-L., Yue, P., Sun, Z.-J., Gu, R.-M., and Yang, B.-F. (2006). Chlorzoxazone inhibits contraction of rat thoracic aorta. Eur. J. Pharmacol. *545*, 161–166.

Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. *4*, 1184–1191.

Emery, A.E.H. (2002). The muscular dystrophies. Lancet Lond. Engl. *359*, 687–695.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Evangelista, D., Avino, M., Tripathi, K.P., and Guarracino, M.R. (2015). A Web Resource on Skeletal Muscle Transcriptome of Primates. In Computational Intelligence Methods for Bioinformatics and Biostatistics, (Springer, Cham), pp. 273–284.

Gagnon, P., Maltais, F., Bouyer, L., Ribeiro, F., Coats, V., Brouillard, C., Noël, M., Rousseau-Gagnon, M., and Saey, D. (2013). Distal leg muscle function in patients with COPD. COPD *10*, 235–242.

Gao, Z., and Cooper, T.A. (2013). Reexpression of pyruvate kinase M2 in type 1 myofibers correlates with altered glucose metabolism in myotonic dystrophy. Proc. Natl. Acad. Sci. U. S. A. *110*, 13570–13575.

Giudice, J., and Taylor, J.M. (2017). Muscle as a paracrine and endocrine organ. Curr. Opin. Pharmacol. *34*, 49–55.

Grant, G.R., Farkas, M.H., Pizarro, A.D., Lahens, N.F., Schug, J., Brunk, B.P., Stoeckert, C.J., Hogenesch, J.B., and Pierce, E.A. (2011). Comparative analysis of RNA-Seq alignment algorithms and the RNA-Seq unified mapper (RUM). Bioinforma. Oxf. Engl. *27*, 2518–2528.

Greising, S.M., Medina-Martínez, J.S., Vasdev, A.K., Sieck, G.C., and Mantilla, C.B. (2015). Analysis of muscle fiber clustering in the diaphragm muscle of sarcopenic mice. Muscle Nerve *52*, 76–82.

Haizlip, K.M., Harrison, B.C., and Leinwand, L.A. (2015). Sex-based differences in skeletal muscle kinetics and fiber-type composition. Physiol. Bethesda Md *30*, 30–39.

Herbison, G.J., Jaweed, M.M., and Ditunno, J.F. (1979). Muscle atrophy in rats following denervation, casting, inflammation, and tenotomy. Arch. Phys. Med. Rehabil. *60*, 401–404.

Hermans, G., and Van den Berghe, G. (2015). Clinical review: intensive care unit acquired weakness. Crit. Care Lond. Engl. *19*, 274.

Hochberg, Y., and Benjamini, Y. (1990). More powerful procedures for multiple significance testing. Stat. Med. *9*, 811–818.

Hughes, L.D., Lewis, S.A., and Hughes, M.E. (2017). ExpressionDB: An open source platform for distributing genome-scale datasets. PLOS ONE *12*, e0187457.

Hughes, M.E., DiTacchio, L., Hayes, K.R., Vollmers, C., Pulivarthy, S., Baggs, J.E., Panda, S., and Hogenesch, J.B. (2009). Harmonics of circadian gene transcription in mammals. PLoS Genet. *5*, e1000442.

Iizuka, K., Machida, T., and Hirafuji, M. (2014). Skeletal muscle is an endocrine organ. J. Pharmacol. Sci. *125*, 125–131.

Juhas, M., and Bursac, N. (2013). Engineering skeletal muscle repair. Curr. Opin. Biotechnol. *24*, 880–886.

Karsenty, G., and Olson, E.N. (2016). Bone and Muscle Endocrine Functions: Unexpected Paradigms of Inter-organ Communication. Cell *164*, 1248–1256.

Krumlauf, R. (1994). Hox genes in vertebrate development. Cell *78*, 191–201.

Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A.O., Landthaler, M., et al. (2007). A mammalian microRNA expression atlas based on small RNA library sequencing. Cell *129*, 1401–1414.

Latronico, N., Tomelleri, G., and Filosto, M. (2012). Critical illness myopathy. Curr. Opin. Rheumatol. *24*, 616–622.

Li, J., Grant, G.R., Hogenesch, J.B., and Hughes, M.E. (2015). Considerations for RNA-seq analysis of circadian rhythms. Methods Enzymol. *551*, 349–367.

Liao, E.H., Hung, W., Abrams, B., and Zhen, M. (2004). An SCF-like ubiquitin ligase complex that controls presynaptic differentiation. Nature *430*, 345–350.

Lightfoot, A.P., and Cooper, R.G. (2016). The role of myokines in muscle health and disease. Curr. Opin. Rheumatol. *28*, 661–666.

Lindholm, M.E., Huss, M., Solnestam, B.W., Kjellqvist, S., Lundeberg, J., and Sundberg, C.J. (2014). The human skeletal muscle transcriptome: sex differences, alternative splicing, and tissue homogeneity assessed with RNA sequencing. FASEB J. *28*, 4571–4581.

Merrell, A.J., and Kardon, G. (2013). Development of the diaphragm -- a skeletal muscle essential for mammalian respiration. FEBS J. *280*, 4026–4035.

Moyer, M., Berger, D.S., Ladd, A.N., and Van Lunteren, E. (2011). Differential susceptibility of muscles to myotonia and force impairment in a mouse model of myotonic dystrophy. Muscle Nerve *43*, 818–827.

Murphy, M., and Kardon, G. (2011). Origin of vertebrate limb muscle: the role of progenitor and myoblast populations. Curr. Top. Dev. Biol. *96*, 1–32.

Noden, D.M., and Francis-West, P. (2006). The differentiation and morphogenesis of craniofacial muscles. Dev. Dyn. Off. Publ. Am. Assoc. Anat. *235*, 1194–1218.

Porter, J.D., Khanna, S., Kaminski, H.J., Rao, J.S., Merriam, A.P., Richmonds, C.R., Leahy, P., Li, J., and Andrade, F.H. (2001). Extraocular muscle is defined by a fundamentally distinct gene expression profile. Proc. Natl. Acad. Sci. U. S. A. *98*, 12062–12067.

Qazi, T.H., Mooney, D.J., Pumberger, M., Geissler, S., and Duda, G.N. (2015). Biomaterials based strategies for skeletal muscle tissue engineering: existing technologies and future trends. Biomaterials *53*, 502–521.

Robles, M.S., Cox, J., and Mann, M. (2014). In-vivo quantitative proteomics reveals a key contribution of post-transcriptional mechanisms to the circadian regulation of liver metabolism. PLoS Genet. *10*, e1004047.

Roth, S.M., Ferrell, R.E., Peters, D.G., Metter, E.J., Hurley, B.F., and Rogers, M.A. (2002). Influence of age, sex, and strength training on human muscle gene expression determined by microarray. Physiol. Genomics *10*, 181–190.

Saiga, T., Fukuda, T., Matsumoto, M., Tada, H., Okano, H.J., Okano, H., and Nakayama, K.I. (2009). Fbxo45 forms a novel ubiquitin ligase complex and is required for neuronal development. Mol. Cell. Biol. *29*, 3529–3543.

Schakman, O., Gilson, H., and Thissen, J.P. (2008). Mechanisms of glucocorticoid-induced myopathy. J. Endocrinol. *197*, 1–10.

Serrano, A.L., Jardi, M., Suelves, M., Klotman, P.E., and Munoz-Canoves, P. (2008). HIV-1 transgenic expression in mice induces selective atrophy of fast-glycolytic skeletal muscle fibers. Front. Biosci. J. Virtual Libr. *13*, 2797–2805.

Smith, G.A., Fearnley, G.W., Harrison, M.A., Tomlinson, D.C., Wheatcroft, S.B., and Ponnambalam, S. (2015). Vascular endothelial growth factors: multitasking functionality in metabolism, health and disease. J. Inherit. Metab. Dis. *38*, 753–763.

Su, A.I., Cooke, M.P., Ching, K.A., Hakak, Y., Walker, J.R., Wiltshire, T., Orth, A.P., Vega, R.G., Sapinoso, L.M., Moqrich, A., et al. (2002). Large-scale analysis of the human and mouse transcriptomes. Proc. Natl. Acad. Sci. U. S. A. *99*, 4465–4470.

Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., et al. (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. Proc. Natl. Acad. Sci. U. S. A. *101*, 6062–6067.

Tiao, G., Lieberman, M., Fischer, J.E., and Hasselgren, P.O. (1997). Intracellular regulation of protein degradation during sepsis is different in fast- and slow-twitch muscle. Am. J. Physiol. *272*, R849-856.

Tsumagari, K., Baribault, C., Terragni, J., Chandra, S., Renshaw, C., Sun, Z., Song, L., Crawford, G.E., Pradhan, S., Lacey, M., et al. (2013). DNA methylation and differentiation: HOX genes in muscle cells. Epigenetics Chromatin *6*, 25.

Vissing, K., and Schjerling, P. (2014). Simplified data access on human skeletal muscle transcriptome responses to differentiated exercise. Sci. Data *1*, sdata201441.

Zhang, R., Lahens, N.F., Ballance, H.I., Hughes, M.E., and Hogenesch, J.B. (2014). A circadian gene expression atlas in mammals: implications for biology and medicine. Proc. Natl. Acad. Sci. *111*, 16219–16224.

Zhou, Y., Gong, B., and Kaminski, H.J. (2012). Genomic Profiling Reveals Pitx2 Controls Expression of Mature Extraocular Muscle Contraction–Related Genes. Invest. Ophthalmol. Vis. Sci. *53*, 1821–1829.

Figure and Table Legends:

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 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. (C) The percent of transcripts showing differential expression between tissues is shown at different false-discovery rate (q-value) thresholds. (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.

Figure 2: 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. (B) Absolute expression of *Myh* transcripts among all 11 mouse skeletal muscle tissues is represented as a heat map. 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.

Figure 3: Developmental gene expression persists into adulthood in mouse skeletal muscle. (A) Normalized expression (Z-score) of all *Hox* gene transcripts is represented as a heat map. 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 skeletal muscle differentiation factors (*Myod1*, *Myf5*, *Myog*, and *Myf6*) is shown as a bar graph. Skeletal muscle-specific expression of these genes 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*.

Figure 4: 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 for 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.).

Figure 5: 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 2B**) (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 450 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).

Figure 6: 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, 3,370 (~15% of all genes) have orthologs known to influence 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).

Figure S1: 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.

Figure S2: Empirical simulations show that the read depth of this study approaches saturation for detecting expressed exons. In order to verify that the RNA-seq 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.

Figure S3: 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 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.

Figure S4: 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 RNA-seq data. Panel (B) shows the log2 fold changes for the same genes measured by qPCR on biological samples *collected independently of those used for RNA-seq*. 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.

Figure S5: 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 (4) other skeletal muscles, including soleus. These data were used to generate the dendrogram shown in **Figure 1D**.

Figure S6: 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* (Burkholder et al., 1994); 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. With these caveats in mind, to a first approximation, RNA-seq data of *Myh* expression agrees with legacy data of fiber type composition.

Figure S7: 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, *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). RNA-seq 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 RNA-seq. Experimental lanes 1 and 2 using primers spanning canonical exons 2 and 3, show a predominant PCR product of ~880bp 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 ~600bp 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 ~330bp 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 ~260bp 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.

Figure S8: 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 *soleu*s 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.

Figure S9: 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). 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.

Figure S10: 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). 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).

Table 1: The biological samples collected in this study and the total number of aligned RNA-seq reads.

Table S1: Details on the total number of reads and aligned reads for every replicate sample.

Table S2: Every unique gene that is a candidate myokine. This Table is distinct from **Figure S9** in that no filtering based on differential expression was performed, and the filter for minimal expression is less stringent (i.e. FPKM > 1 instead of FPKM > 10).

Table 1:

Tissue	Type	Species	Sex	Replicates	Total Aligned Reads (millions)
Total Aorta	Smooth	Mouse	M	6	205.2
Abdominal Aorta	Smooth	Mouse	M	6	219.7
Thoracic Aorta	Smooth	Mouse	M	6	214.0
Atria	Cardiac	Mouse	M	6	169.0
Left Ventricle	Cardiac	Mouse	M	6	193.9
Right Ventricle	Cardiac	Mouse	M	6	176.2
Diaphragm	Skeletal	Mouse	M	6	159.1
EDL	Skeletal	Mouse	M	6	176.8
Extraocular Eye	Skeletal	Mouse	M	6	171.7
FBD	Skeletal	Mouse	M	6	202.5
Massater	Skeletal	Mouse	M	6	216.2
Plantaris	Skeletal	Mouse	M	6	185.1
Soleus	Skeletal	Mouse	M	6	172.2
Tongue	Skeletal	Mouse	M	6	210.4
Gastrocnemius	Skeletal	Mouse	M	6	276.1
Quadriceps	Skeletal	Mouse	M	6	275.6
Tibialis Anterior	Skeletal	Mouse	M	6	276.6
EDL	Skeletal	Rat	M	6	226.4
EDL	Skeletal	Rat	F	6	206.8
Soleus	Skeletal	Rat	M	6	261.2
Soleus	Skeletal	Rat	F	6	243.0
Average				6	211.3
Total				132	4437.7

Table S1:

Tissue	Sample	Replicate	Reads	Uniquely aligned reads	NU reads	Total Aligned
	37279	1	35765748	26731364	8603905	35335269
	37280	2	31038637	24883980	5704678	30588658
	37281	3	30353264	24233571	5596921	29830492
	39355	4	38886763	32063652	6182299	38245951
	39356	5	42485079	32388161	9532093	41870254
AOR_Total	39357	6	29914315	24681124	4615500	29296624
	41571	1	30266476	25485624	4474814	29960438
	41572	2	40931092	35464053	4930950	40395003
	41573	3	33766479	28949133	4281326	33230459
	41574	4	38666034	33380429	4861536	38241965
	41575	5	37804938	32141028	5184776	37325804
AOR_Abdominal	41576	6	41050049	35407806	5141013	40548819
	41577	1	43662119	36964662	6199829	43164491
	41578	2	34997461	29773877	4815437	34589314
	41579	3	35036310	29786186	4550291	34336477
	41580	4	42858909	35838634	6519265	42357899
	41581	5	28862699	24465485	4096783	28562268
AOR_Thoracic	41582	6	31330884	26377387	4594574	30971961
	37282	1	29777446	24230611	5055362	29285973
	37283	2	34156167	27639215	6125353	33764568
	37284	3	24278447	19236842	4767696	24004538
	39358	4	31460555	25131538	5963360	31094898
	39359	5	26360416	20967505	5074935	26042440
ATR	39360	6	25083000	19990362	4778708	24769070
	37288	1	35921978	24299022	11305260	35604282
	37289	2	33835874	22288359	11280095	33568454
	37290	3	32331818	24462711	10609175	32071886
	39361	4	35325775	25155344	9871094	35026438
	39362	5	22541615	16140501	6143658	22284159
LV	39363	6	35702202	24951477	10418772	35370249
	37285	1	24329044	16839803	7285748	24125551
	37286	2	31780005	21790508	9734000	31524508
	37287	3	36994396	25156729	11537917	36694646
	39364	4	29036772	20557089	8240098	28797187
	39365	5	27626476	19211196	8187105	27398301
RV	39366	6	27913480	19437173	8230564	27667737
	37291	1	33019480	23836474	8732772	32569246
DIA	37292	2	31629751	23388399	7885484	31273883

	37293	3	31945551	23927206	7636605	31563811
	39367	4	21311980	15246726	5789401	21036127
	39368	5	19037470	13946740	4898128	18844868
	39369	6	24059171	17556632	6259110	23815742
	37294	1	31576015	24902704	6228134	31130838
	37295	2	31451634	24783197	6244774	31027971
	37296	3	33216118	25518587	7299078	32817665
	39370	4	21401430	16499880	4625790	21125670
	39371	5	31197719	24046037	6789188	30835225
EDL	39372	6	30283998	22589060	7306367	29895427
	37297	1	31762860	23064615	8354068	31418683
	37298	2	31945869	22979679	8623587	31603266
	37299	3	28119001	20432792	7362473	27795265
	39373	4	27629144	19883997	7435738	27319736
	39374	5	29768612	20721448	8732342	29453790
EYE	39375	6	24372471	16484361	7648242	24132603
	41732	1	34935337	28433677	6119522	34553199
	41733	2	30805461	25201955	5279068	30481023
	41734	3	31796088	26115282	5339080	31454362
	41735	4	35574878	29408999	5747764	35156763
	41736	5	34474174	28449265	5646436	34095701
FDB	41737	6	37140146	30331631	6409407	36741038
	42744	1	47698800	34455257	12875450	47330707
	42745	2	39717983	28537223	10778018	39315241
	42746	3	36344724	26913940	8999850	35913790
	42747	4	31720345	24373850	6949606	31323456
	42748	5	24410038	18689701	5472331	24162032
MAS	42749	6	38477271	28325335	9795468	38120803
	MISEQ	1	17325408	12810123	4335484	17145607
	41738	2	34137807	26160356	7603385	33763741
	41739	3	31601098	23656195	7648805	31305000
	41740	4	34857760	25974447	8531031	34505478
	41741	5	33451589	24867139	8263107	33130246
PLA	41742	6	35532067	25491116	9752493	35243609
	37300	1	31211755	23594793	7267721	30862514
	37301	2	29040389	21985458	6780345	28765803
	37302	3	44546485	32296159	11882663	44178882
	39376	4	16966114	12313609	4455934	16769543
	39377	5	27390307	20596063	6560878	27156941
SOL	MISEQ	6	24667682	18746989	5685297	24432286
	42750	1	35362248	27098477	7860465	34958942
	MISEQ	2	25788691	20501956	4939573	25441529
TON	MISEQ	3	22681867	17660674	4759243	22419917

	42751	4	35246582	26663556	8179347	34842903
	42752	5	47072213	36377715	10147398	46525113
	42753	6	46723356	33908057	12306310	46214367
	22779	1	43307706	31808676	10886046	42694722
	22780	2	47299527	35755931	10985816	46741747
	22781	3	42511529	29532830	12009838	41542688
	22782	4	51314235	35201237	15402036	50603273
	22783	5	44385898	30349707	13565390	43915097
GAS	22784	6	51072010	35829864	14757854	50587718
	22785	1	44938344	31407593	13055370	44462963
	22786	2	52931439	37490718	14945743	52436461
	22787	3	46228667	32133031	13666625	45799656
	22788	4	46585936	31177483	15001039	46178522
	22789	5	42713171	28924327	13397648	42321975
QUAD	22790	6	44978803	31435740	12981242	44416982
	22773	1	49434172	36880210	11788899	48669109
	22774	2	42644431	31629918	10259282	41889200
	22775	3	43880529	33727439	9482543	43209982
	22776	4	46479997	35288566	10396487	45685053
	22777	5	51663346	33884301	17232178	51116479
TAN	22778	6	46475026	29919594	16106053	46025647
	63697	1	35917834	32305280	2967629	35272909
	63698	2	25834807	22865878	2470654	25336532
	63699	3	37387833	33958290	2859997	36818287
	63700	4	37807711	34129098	3062278	37191376
	63701	5	47677545	43118940	3806030	46924970
Rat_EDL_Male	63702	6	45519983	41389888	3423204	44813092
	63703	1	38802668	35252252	2923319	38175571
	63704	2	32971683	29993013	2476602	32469615
	63705	3	30542834	27632762	2456582	30089344
	63706	4	32764438	29637389	2624571	32261960
	63707	5	34906504	32037603	2411567	34449170
Rat_EDL_Female	63708	6	39902690	36336398	3015174	39351572
	63709	1	49036067	43522793	4598859	48121652
	63710	2	33886651	29183959	2927237	32111196
	63711	3	43321337	39114618	3477018	42591636
	63712	4	50690677	46131412	3765876	49897288
	63713	5	44198978	40178753	3322346	43501099
Rat_SOL_Male	63714	6	45713162	41315077	3639574	44954651
	63715	1	40377339	36221612	3502692	39724304
	63716	2	38801558	34942748	3218873	38161621
	63717	3	41758025	37643174	3450133	41093307
	63718	4	47229846	42704883	3804804	46509687

	63719	5	40138570	35732912	3765790	39498702
Rat_SOL_Female	63720	6	38627993	34582180	3407001	37989181

Table S2:

gene	transcript	q-value	Maximum expression	Tissue of maximal expression
Aimp1	uc008rjx	3.36E-15	94.52	tibialis anterior
Areg	uc008ybt	6.81E-18	1.18	eye
Bmp1	uc007uoc	1.54E-17	19.26	eye
Bmp10	uc012eol	2.22E-13	1.20	tongue
Bmp2	uc008mnr	1.41E-15	2.14	FDB
Bmp3	uc008ygh	5.09E-12	1.26	eye
Bmp4	uc011zjm	3.68E-20	6.15	diaphragm
Bmp5	uc009qsu	3.04E-23	11.33	FDB
Bmp6	uc007qdp	3.46E-12	6.85	eye
Bmp7	uc008odc	3.44E-05	5.44	tongue
Bmp8a	uc012dks	4.83E-11	8.24	tongue
Bmp8b	uc008uot	0.0361	4.95	tongue
C1qtnf4	uc008ktq	0.000327	24.21	tongue
Ccl11	uc007kmr	1.46E-14	39.92	eye
Ccl12	uc007kms	0.0035	23.24	tongue
Ccl17	uc009mwz	3.18E-05	13.55	tongue
Ccl2	uc007kmp	2.49E-08	3.50	FDB
Ccl20	uc007bsm	6.60E-06	8.43	tongue
Ccl22	uc009mww	8.30E-05	2.98	masseter
Ccl24	uc008zyn	1.04E-14	9.00	tongue
Ccl25	uc009ktw	2.49E-18	6.05	quadriceps
Ccl27a	uc008sju	6.46E-10	4.92	diaphragm
Ccl28	uc007sal	6.70E-13	3.61	tongue
Ccl3	uc007kpn	0.0203	17.57	tongue
Ccl5	uc007kpi	0.26	8.83	masseter
Ccl6	uc007kpm	1.70E-14	22.46	tongue
Ccl7	uc007kmq	3.21E-06	2.72	tongue
Ccl8	uc007kmt	1.24E-07	29.63	masseter
Ccl9	uc007kpj	3.61E-12	7.92	FDB
Cd40lg	uc009thb	0.082	4.92	tongue
Cklf	uc009nad	2.80E-05	2.21	tongue
Clcf1	uc008fzj	4.84E-07	59.78	tongue
Cmtm3	uc009nam	1.28E-15	10.74	eye
Cmtm4	uc009nan	5.65E-09	5.59	eye
Cmtm5	uc007txq	8.67E-05	4.73	eye
Cmtm6	uc009ryb	2.60E-18	11.21	eye
Cmtm7	uc009ryc	2.05E-10	8.85	tongue
Cmtm8	uc009ryg	1.10E-22	7.36	soleus
Cntf	uc008gun	4.19E-16	17.77	eye
Crlf1	uc009mak	1.05E-11	5.58	tongue

Csf1	uc008qxl	1.84E-19	17.13	tongue
Cx3cl1	uc009mwx	2.26E-17	2.35	eye
Cxcl10	uc008ycz	3.45E-05	1.09	masseter
Cxcl11	uc029vjk	1.25E-16	64.41	tongue
Cxcl12	uc009dku	9.39E-31	64.45	diaphragm
Cxcl13	uc008yfb	1.59E-08	19.78	masseter
Cxcl14	uc011zab	2.67E-23	34.93	FDB
Cxcl16	uc011xxt	9.67E-14	4.31	tongue
Cxcl3	uc008ybj	4.63E-17	9.57	tongue
Cxcl5	uc008ybg	1.42E-10	1.07	tongue
Cxcl9	uc008ycv	1.59E-17	42.87	tongue
Ebi3	uc008daj	0.00877	1.04	tongue
Edn1	uc007qfl	1.64E-23	4.55	eye
Fam3b	uc008adj	2.33E-15	1.90	tongue
Fgf2	uc008paq	4.20E-10	7.17	tongue
Flt3l	uc009gto	2.23E-14	7.51	eye
Gdf10	uc007tab	1.03E-14	5.56	eye
Gdf11	uc007hot	5.13E-27	27.71	tongue
Gdf15	uc009mba	1.83E-08	1.60	tongue
Gdf2	uc007tac	7.09E-15	1.42	tongue
Gdf5	uc012chg	0.000127	1.71	tongue
Gdf7	uc007mzl	0.471	6.15	tongue
Gdf9	uc007iwa	4.95E-05	8.09	tongue
Gpi1	uc009gix	7.32E-06	162.22	EDL
Gpi1	uc012fie	7.57E-07	130.74	EDL
Grem2	uc007dtf	3.61E-28	9.60	gastrocnemius
Grn	uc007lrv	3.62E-15	29.19	eye
Hmgb1	uc009apd	0.0117	7.51	quadriceps
Ifne	uc008tod	0.471	14.81	tongue
Il10	uc007cmu	0.221	55.17	tongue
Il12a	uc008plu	2.62E-30	28.94	tongue
Il12b	uc007inc	0.0154	4.45	tongue
Il15	uc009mjj	1.26E-17	20.65	diaphragm
Il16	uc009idr	1.06E-10	6.19	masseter
Il17b	uc008fck	2.22E-05	3.45	tongue
Il17c	uc012gmc	0.0113	2.97	tongue
Il18	uc009pju	4.38E-14	8.14	tongue
Il1a	uc008mhr	3.08E-10	1.85	tongue
Il1b	uc008mht	0.000855	1.69	tongue
Il1f10	uc008iou	1.74E-59	5.78	tongue
Il1f5	uc008ios	2.23E-20	21.05	tongue
Il1f6	uc008ioq	6.99E-23	18.82	tongue
Il1f6	uc008ior	2.13E-22	30.65	tongue

Il1f8	uc008ion	2.60E-26	2.57	tongue
Il1f9	uc008ioo	2.33E-10	6.91	tongue
Il20	uc011wrm	2.43E-14	9.81	eye
Il23a	uc007hmc	0.306	20.17	tongue
Il24	uc007cmo	0.723	6.83	tongue
Il25	uc007txn	0.421	22.77	tongue
Il33	uc008hec	6.13E-20	21.61	eye
Il34	uc009nla	3.42E-11	4.28	tongue
Il5	uc007iwv	5.16E-10	1.06	tongue
Il7	uc008ook	1.21E-13	3.30	tongue
Inhba	uc007pnw	6.91E-05	1.34	FDB
Kitl	uc007gxs	8.16E-09	24.13	eye
Lefty1	uc007dww	0.000193	5.14	tongue
Lefty2	uc007dwu	8.58E-05	4.74	tongue
Lta	uc008cgu	2.94E-05	1.80	masseter
Ltb	uc008cgp	0.000128	6.81	masseter
Mif	uc007ftc	5.33E-11	16.39	eye
Mstn	uc007ayt	4.34E-31	78.31	gastrocnemius
Nampt	uc007nif	1.52E-14	107.45	quadriceps
Ndp	uc009ssc	2.57E-05	14.82	tongue
Nodal	uc007ffy	5.65E-07	16.82	tongue
Nog	uc007kwh	3.73E-08	3.39	tongue
Osm	uc007hus	0.00301	1.89	tongue
Pf4	uc008ybi	4.28E-20	28.80	FDB
Pglyrp1	uc009fjv	8.66E-11	28.30	tongue
Scgb3a1	uc007iqv	0.000813	5.43	tongue
Slurp1	uc007wft	3.22E-16	14.73	tongue
Spp1	uc008yki	5.69E-12	28.13	FDB
Thpo	uc007yra	1.23E-12	5.81	soleus
Timp1	uc009stz	4.54E-07	9.82	soleus
Tnf	uc012arb	0.798	2.07	tongue
Tnfsf10	uc008otm	4.28E-19	20.48	tongue
Tnfsf13	uc007jre	5.83E-17	10.65	tongue
Tnfsf13b	uc009kun	0.0537	4.15	tongue
Tnfsf9	uc008ded	2.62E-08	10.48	tongue
Tslp	uc008ejl	3.08E-11	105.21	tongue
Vegfa	uc008crq	1.87E-25	225.26	diaphragm
Wnt5a	uc007suh	1.35E-15	6.01	tongue
Wnt7a	uc009cxz	0.723	2.49	tongue

Figure 1:

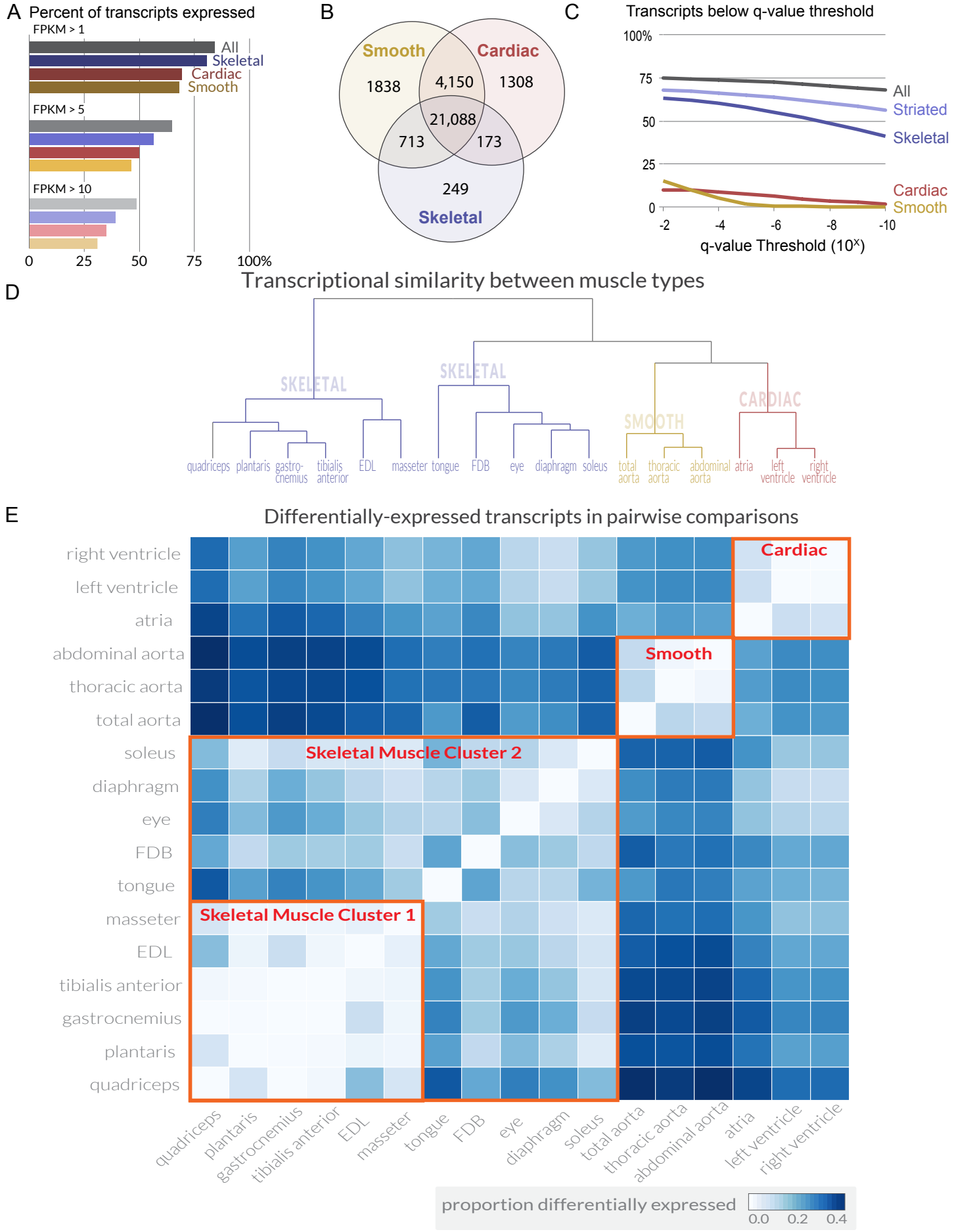

A
Percent of transcripts expressed
FPKM > 1
All
Skeletal
Cardiac
Smooth
FPKM > 5
FPKM > 10
0
25
50
75
100%
B
Smooth
Cardiac
1838
4,150
1308
21,088
713
173
249
Skeletal
C
Transcripts below q-value threshold
100%
75
50
25
0
All
Striated
Skeletal
Cardiac
Smooth
-2
-4
-6
-8
-10
q-value Threshold (10^x)
D
Transcriptional similarity between muscle types
SKELETAL
SKELETAL
SMOOTH
CARDIAC
quadriceps
plantaris
gastro-cnemius
tibialis anterior
EDL
masseter
tongue
FDB
eye
diaphragm
soleus
total aorta
thoracic aorta
abdominal aorta
atria
left ventricle
right ventricle
E
Differentially-expressed transcripts in pairwise comparisons
Cardiac
Smooth
Skeletal Muscle Cluster 2
Skeletal Muscle Cluster 1
right ventricle
left ventricle
atria
abdominal aorta
thoracic aorta
total aorta
soleus
diaphragm
eye
FDB
tongue
masseter
EDL
tibialis anterior
gastrocnemius
plantaris
quadriceps
proportion differentially expressed
0.0
0.2
0.4

Figure 2:

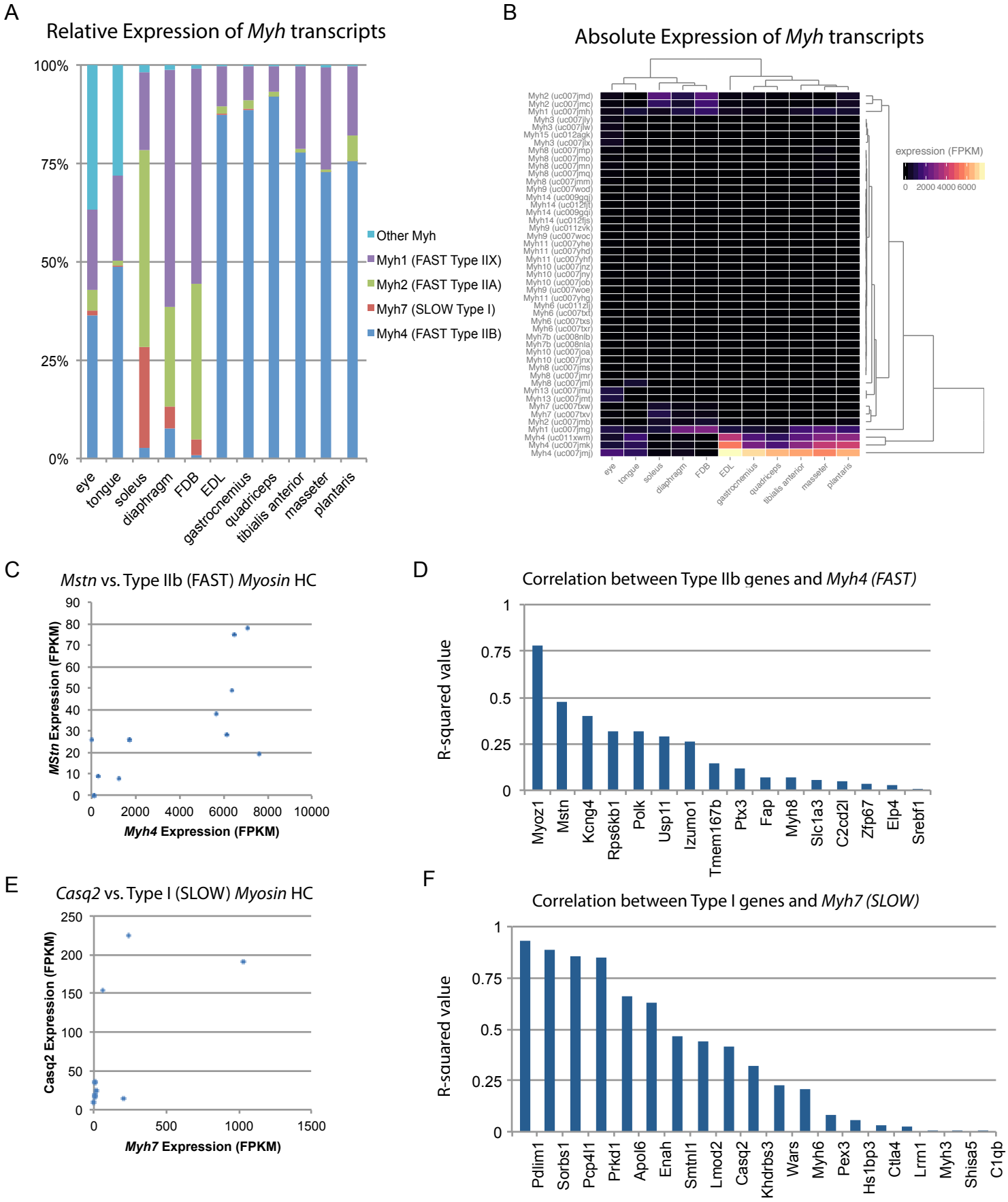
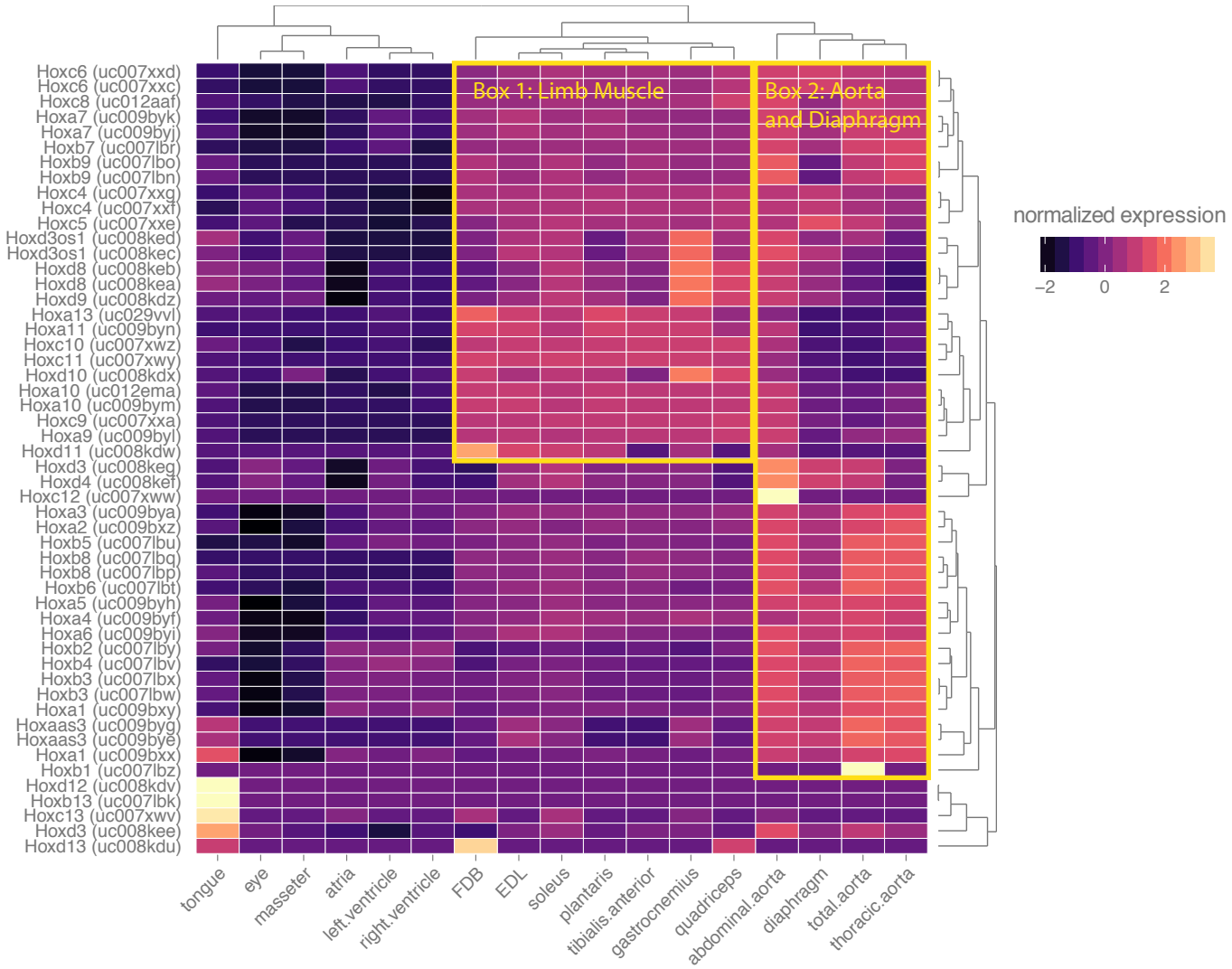
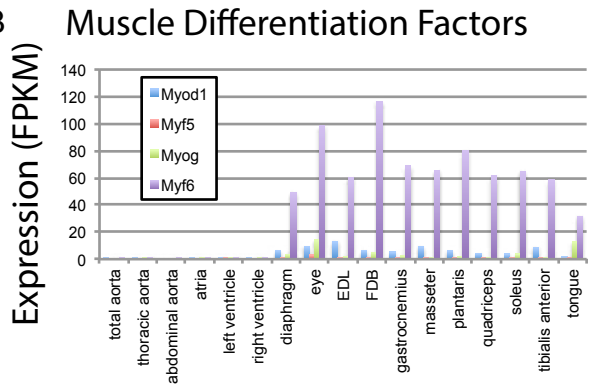

A
Relative Expression of *Myh* transcripts
100%
75%
50%
25%
0%
eye
tongue
soleus
diaphragm
FDB
EDL
gastrocnemius
quadriceps
tibialis anterior
masseter
plantaris
Other Myh
Myh1 (FAST Type IIX)
Myh2 (FAST Type IIA)
Myh7 (SLOW Type I)
Myh4 (FAST Type IIB)
B
Absolute Expression of *Myh* transcripts
expression (FPKM)
0 2000 4000 6000
C
Mstn vs. Type IIb (FAST) *Myosin* HC
MStn Expression (FPKM)
Myh4 Expression (FPKM)
D
Correlation between Type IIb genes and *Myh4 (FAST)*
R-squared value
Myoz1
Mstn
Kcng4
Rps6kb1
Polk
Usp11
Izumo1
Tmem167b
Ptx3
Fap
Myh8
Slc1a3
C2cd2l
Zfp67
Elp4
Srebf1
E
Casq2 vs. Type I (SLOW) *Myosin* HC
Casq2 Expression (FPKM)
Myh7 Expression (FPKM)
F
Correlation between Type I genes and *Myh7 (SLOW)*
R-squared value
Pdlim1
Sorbs1
Pcp4l1
Prkd1
Apol6
Enah
Smtnl1
Lmod2
Casq2
Khdrbs3
Wars
Myh6
Pex3
Hs1bp3
Ctla4
Lrm1
Myh3
Shisa5
C1qb

Figure 3:

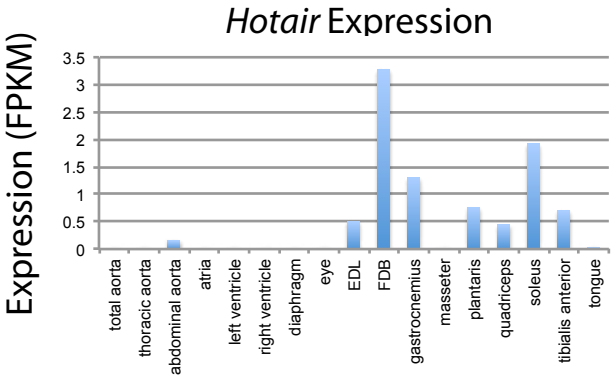
A



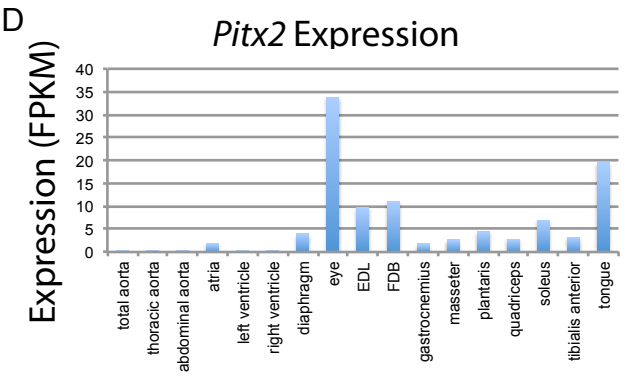
B



C



D



E

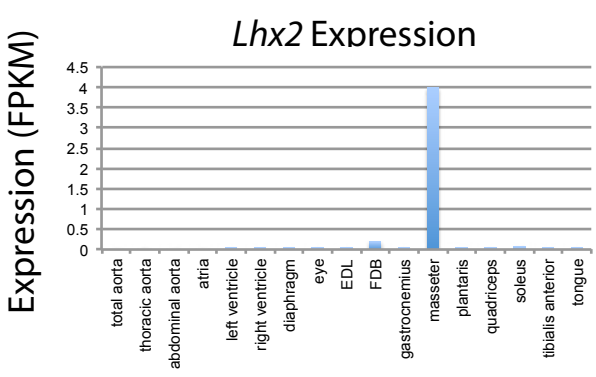


Figure 4:

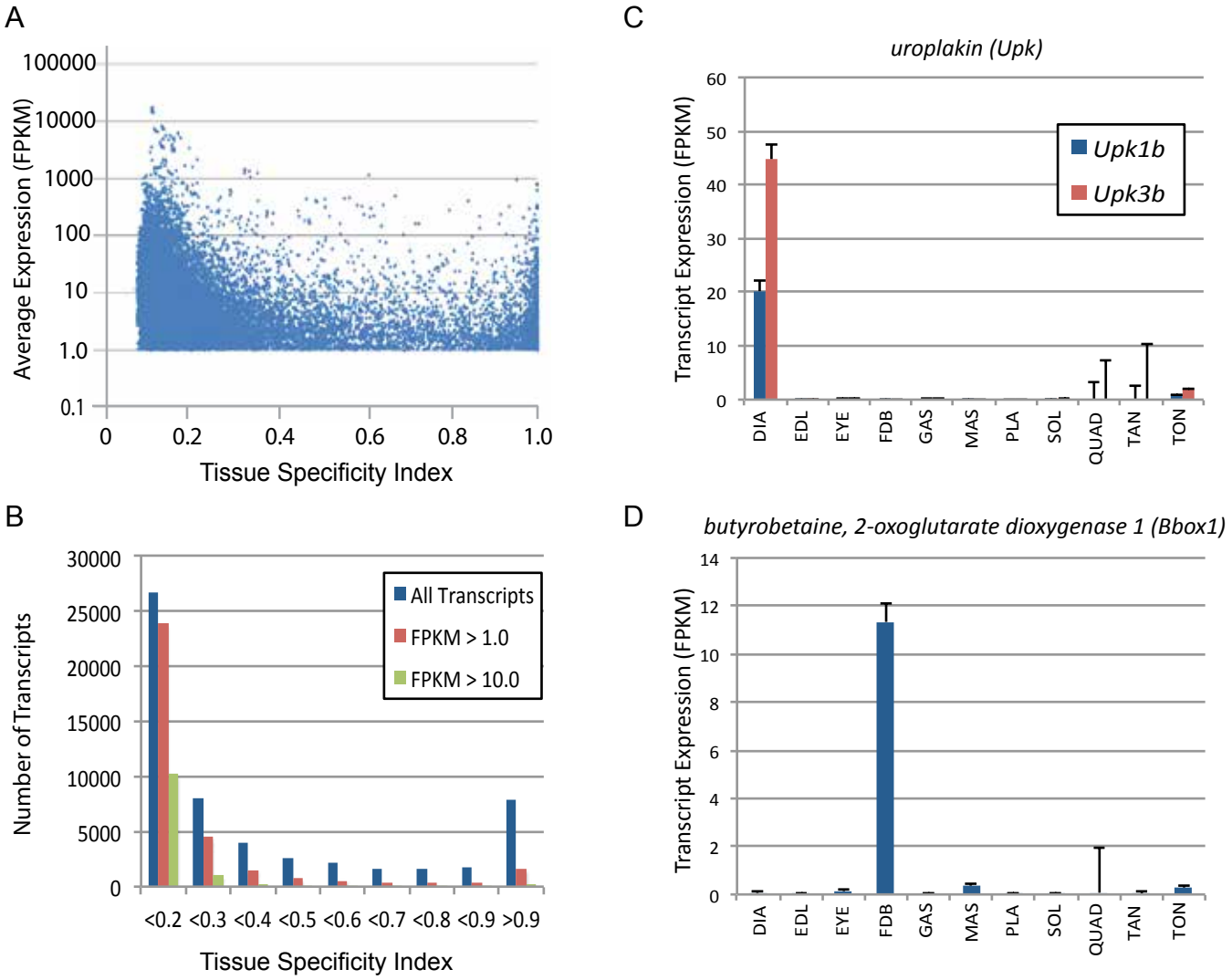
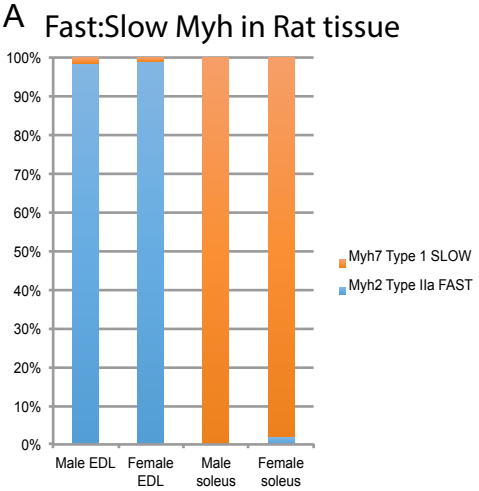
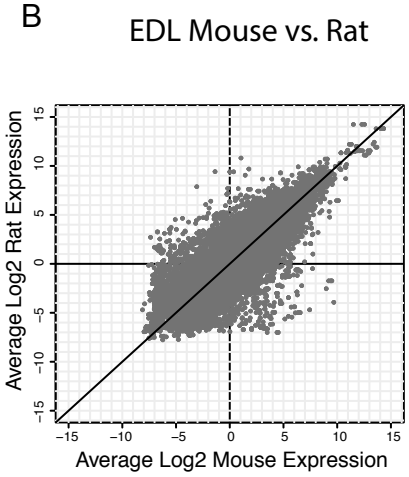

A
Average Expression (FPKM)
100000
10000
1000
100
10
1.0
0.1
0
0.2
0.4
0.6
0.8
1.0
Tissue Specificity Index
B
Number of Transcripts
30000
25000
20000
15000
10000
5000
0
All Transcripts
FPKM > 1.0
FPKM > 10.0
<0.2 <0.3 <0.4 <0.5 <0.6 <0.7 <0.8 <0.9 >0.9
Tissue Specificity Index
C
uroplakin (Upk)
Transcript Expression (FPKM)
60
50
40
30
20
10
0
Upk1b
Upk3b
DIA EDL EYE FDB GAS MAS PLA SOL QUAD TAN TON
D
butyrobetaine, 2-oxoglutarate dioxygenase 1 (Bbox1)
Transcript Expression (FPKM)
14
12
10
8
6
4
2
0
DIA EDL EYE FDB GAS MAS PLA SOL QUAD TAN TON

Figure 5:

A



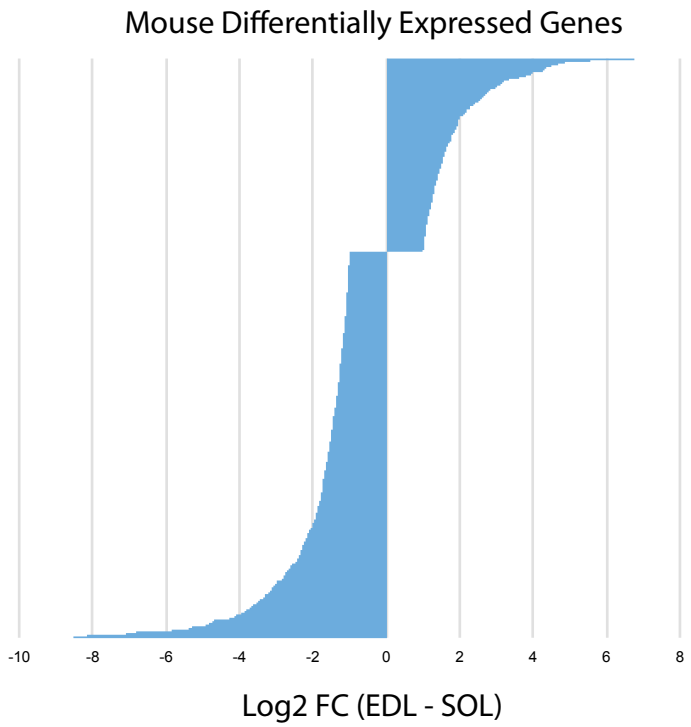
B



C

		Mouse	
		Upregulated	Downregulated
Rat	Upregulated	143	10
	Downregulated	15	257

D



E

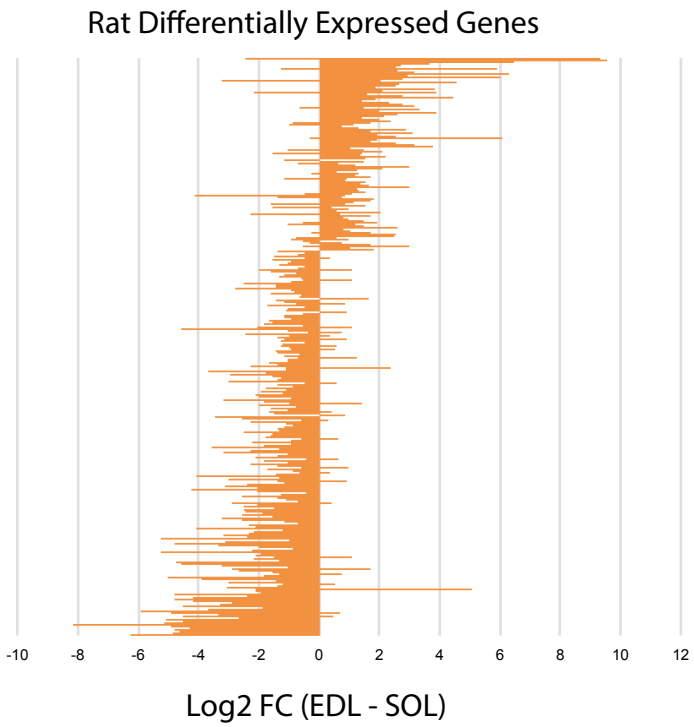
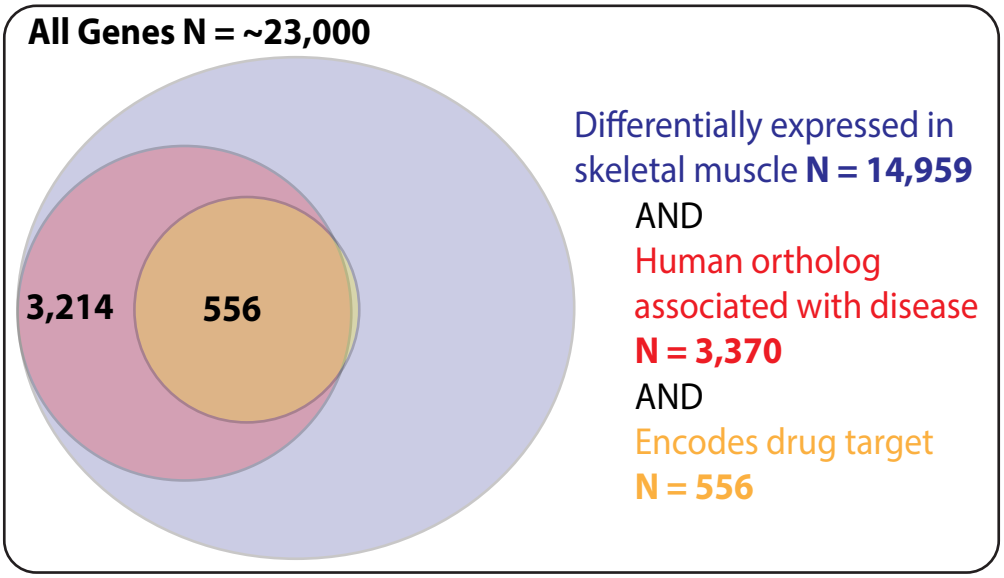
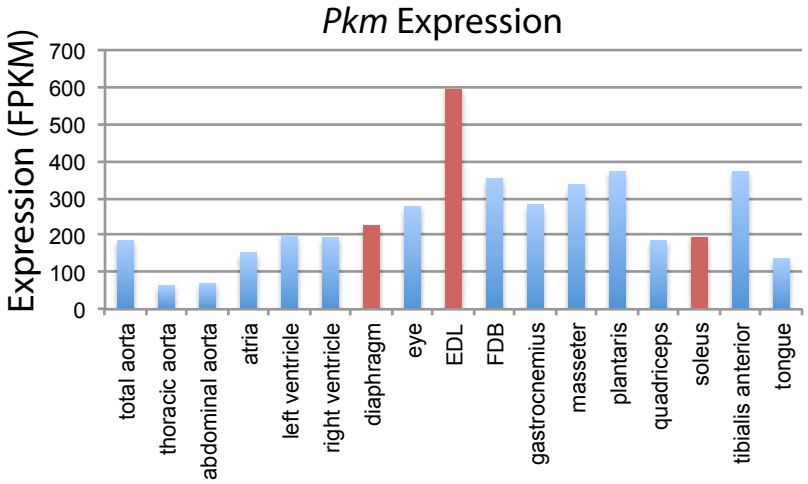


Figure 6:

A


All Genes N = ~23,000
3,214
556
Differentially expressed in skeletal muscle N = 14,959
AND
Human ortholog associated with disease
N = 3,370
AND
Encodes drug target
N = 556

B


Pkm Expression
Expression (FPKM)
700
600
500
400
300
200
100
0
total aorta
thoracic aorta
abdominal aorta
atria
left ventricle
right ventricle
diaphragm
eye
EDL
FDB
gastrocnemius
masseter
plantaris
quadriceps
soleus
tibialis anterior
tongue

C

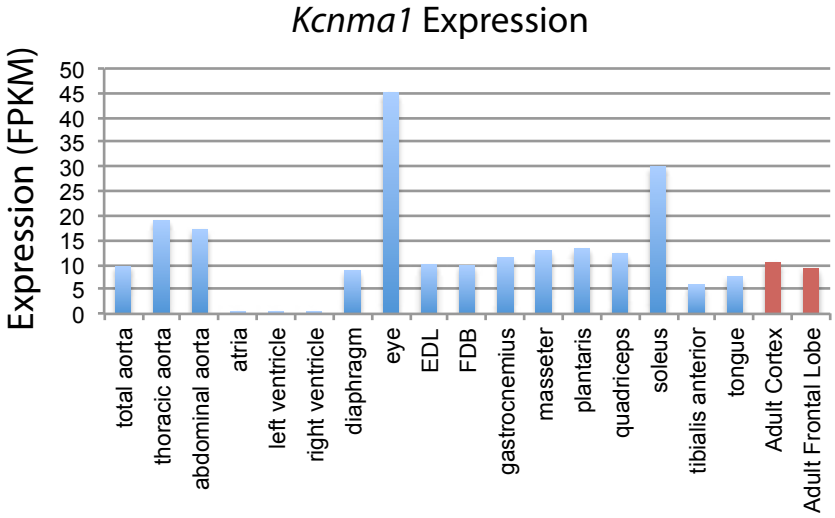

Kcnma1 Expression
Expression (FPKM)
50
45
40
35
30
25
20
15
10
5
0
total aorta
thoracic aorta
abdominal aorta
atria
left ventricle
right ventricle
diaphragm
eye
EDL
FDB
gastrocnemius
masseter
plantaris
quadriceps
soleus
tibialis anterior
tongue
Adult Cortex
Adult Frontal Lobe

Figure S1:

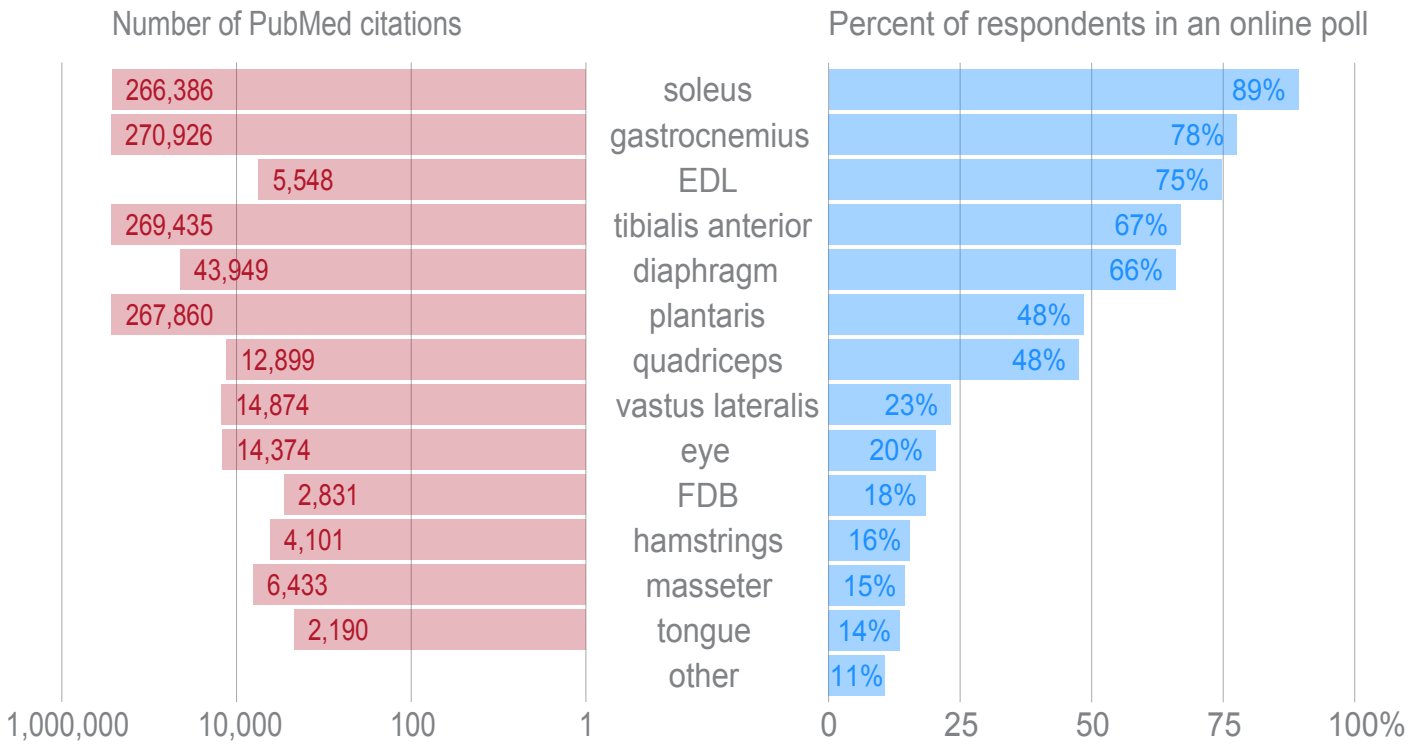

Number of PubMed citations
Percent of respondents in an online poll
266,386
270,926
5,548
269,435
43,949
267,860
12,899
14,874
14,374
2,831
4,101
6,433
2,190
soleus
gastrocnemius
EDL
tibialis anterior
diaphragm
plantaris
quadriceps
vastus lateralis
eye
FDB
hamstrings
masseter
tongue
other
89%
78%
75%
67%
66%
48%
48%
23%
20%
18%
16%
15%
14%
11%
1,000,000
10,000
100
1
0
25
50
75
100%

Figure S2:

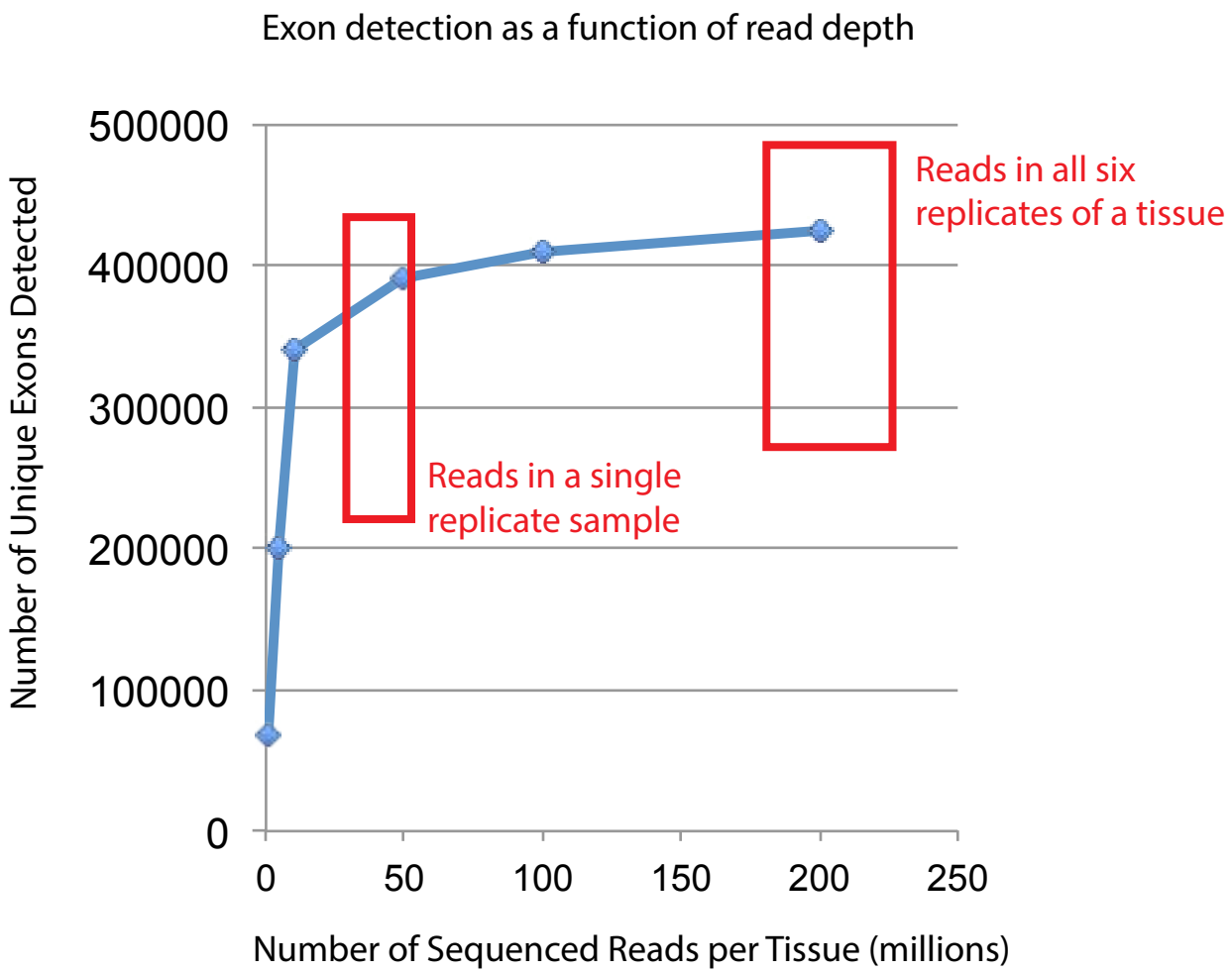
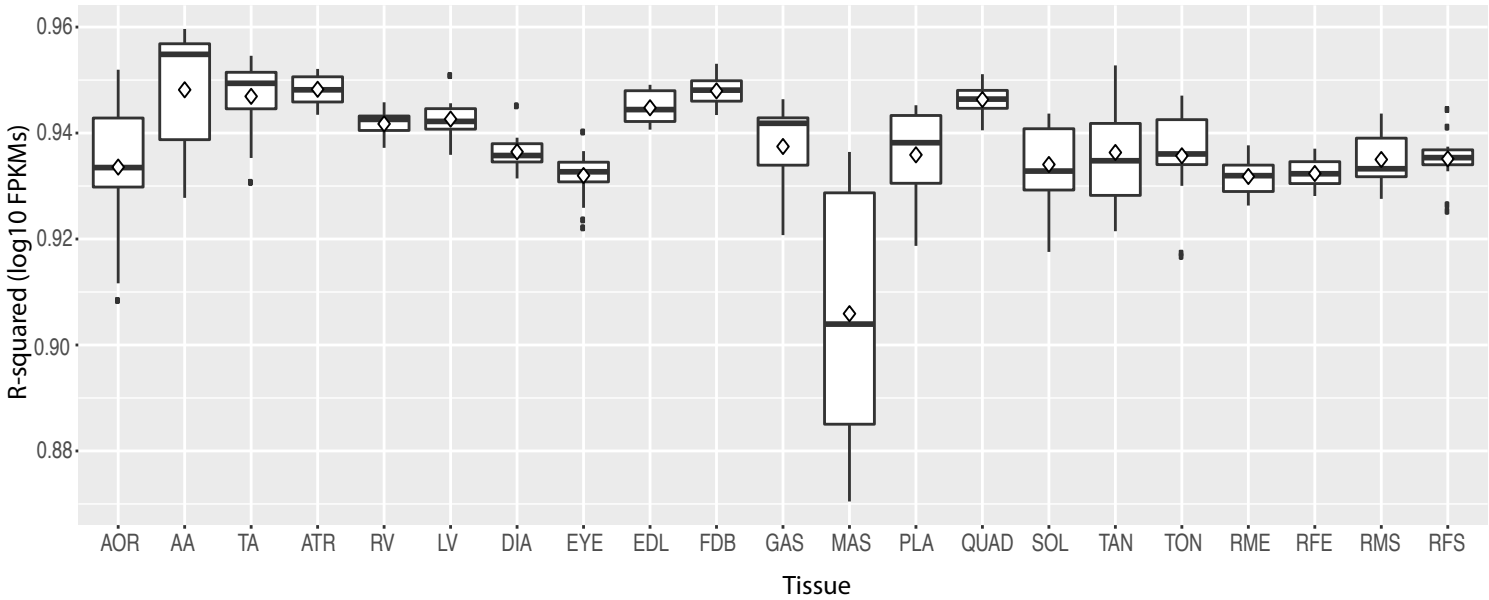

Exon detection as a function of read depth
Number of Unique Exons Detected
500000
400000
300000
200000
100000
0
0
50
100
150
200
250
Number of Sequenced Reads per Tissue (millions)
Reads in a single replicate sample
Reads in all six replicates of a tissue

Figure S3:

A


R-squared (log10 FPKMs)
0.96
0.94
0.92
0.90
0.88
AOR
AA
TA
ATR
RV
LV
DIA
EYE
EDL
FDB
GAS
MAS
PLA
QUAD
SOL
TAN
TON
RME
RFE
RMS
RFS
Tissue

B

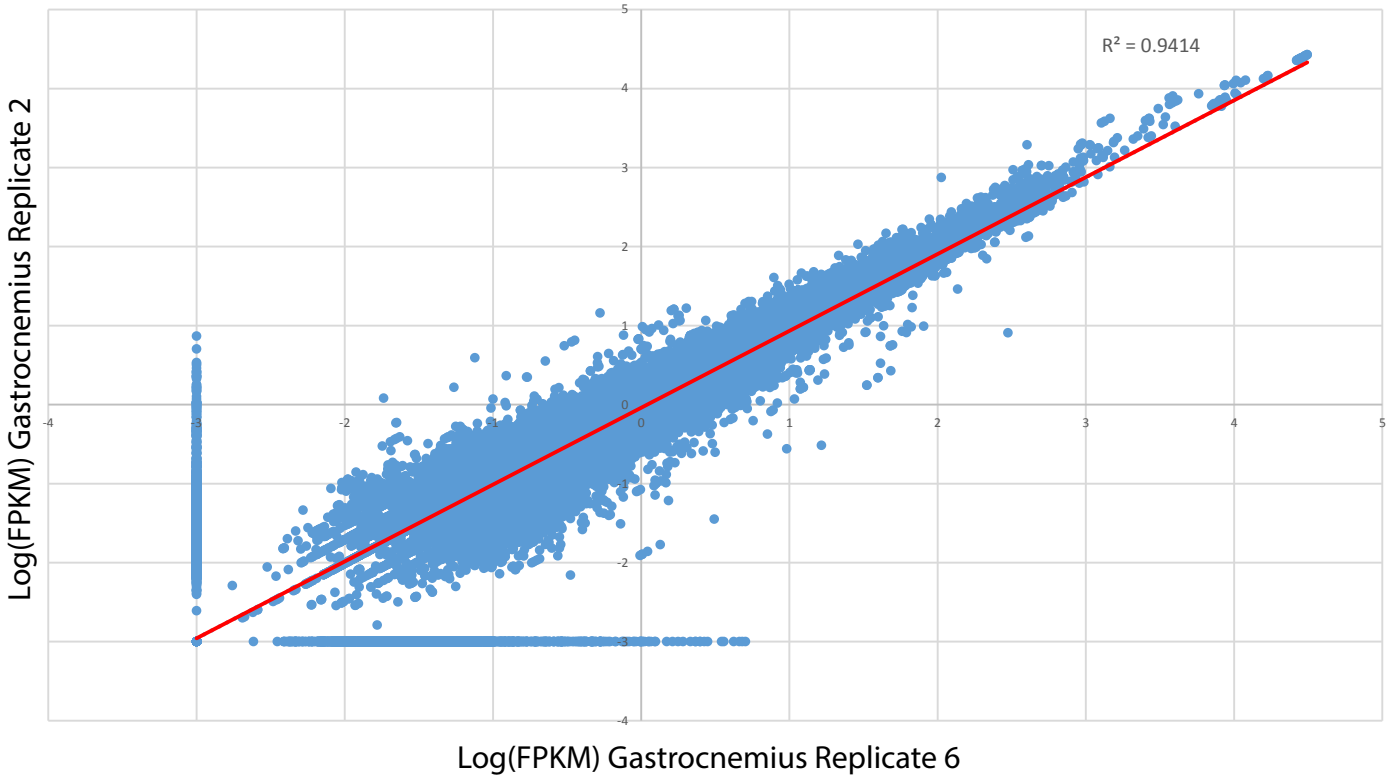

R² = 0.9414
Log(FPKM) Gastrocnemius Replicate 2
Log(FPKM) Gastrocnemius Replicate 6

Figure S4:

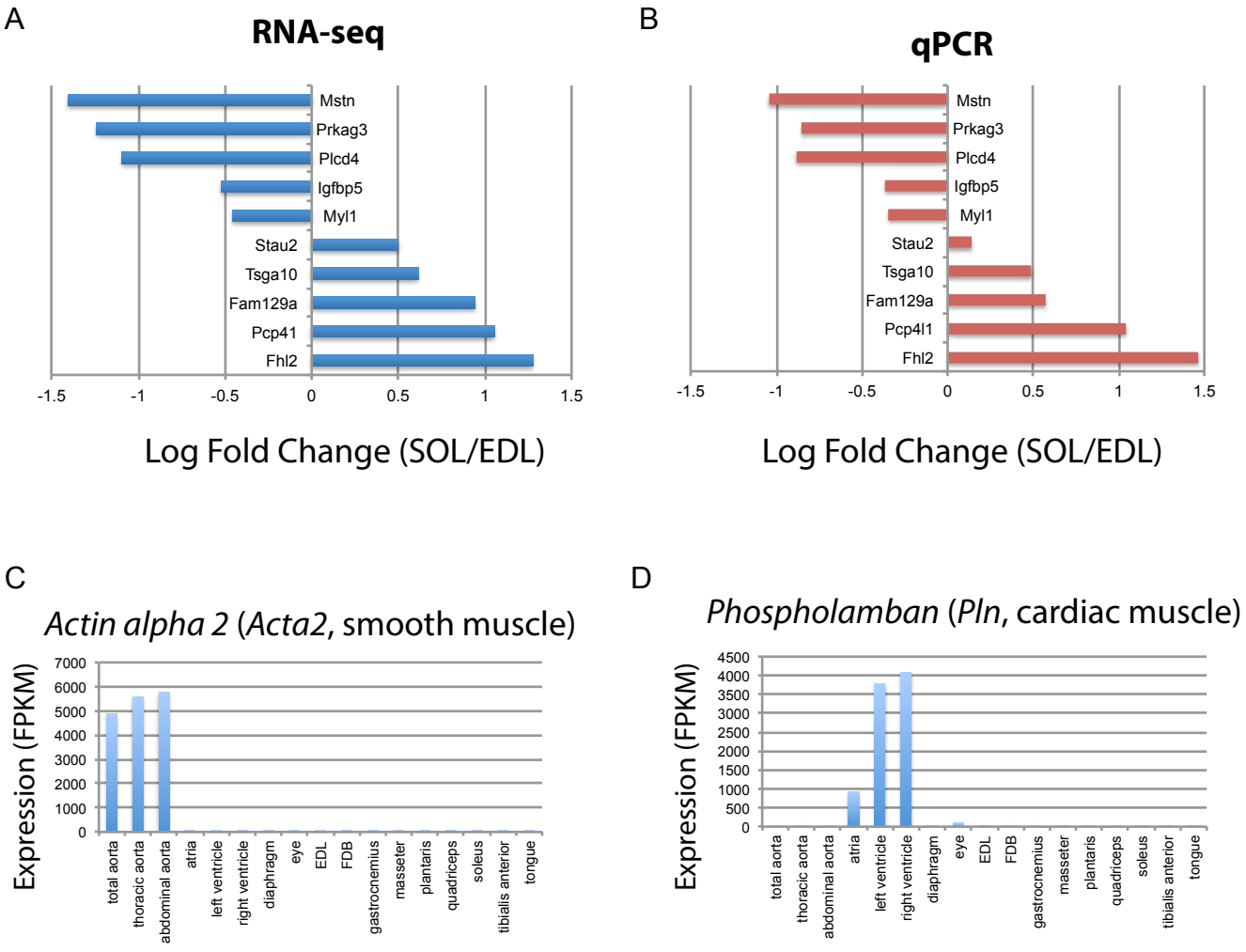

A
RNA-seq
Mstn
Prkag3
Plcd4
Igfbp5
Myl1
Stau2
Tsga10
Fam129a
Pcp41
Fhl2
-1.5
-1
-0.5
0
0.5
1
1.5
Log Fold Change (SOL/EDL)
B
qPCR
Mstn
Prkag3
Plcd4
Igfbp5
Myl1
Stau2
Tsga10
Fam129a
Pcp4l1
Fhl2
-1.5
-1
-0.5
0
0.5
1
1.5
Log Fold Change (SOL/EDL)
C
Actin alpha 2 (Acta2, smooth muscle)
Expression (FPKM)
7000
6000
5000
4000
3000
2000
1000
0
total aorta
thoracic aorta
abdominal aorta
atria
left ventricle
right ventricle
diaphragm
eye
EDL
FDB
gastrocnemius
masseter
plantaris
quadriceps
soleus
tibialis anterior
tongue
D
Phospholamban (Pln, cardiac muscle)
Expression (FPKM)
4500
4000
3500
3000
2500
2000
1500
1000
500
0
total aorta
thoracic aorta
abdominal aorta
atria
left ventricle
right ventricle
diaphragm
eye
EDL
FDB
gastrocnemius
masseter
plantaris
quadriceps
soleus
tibialis anterior
tongue

Figure S5:

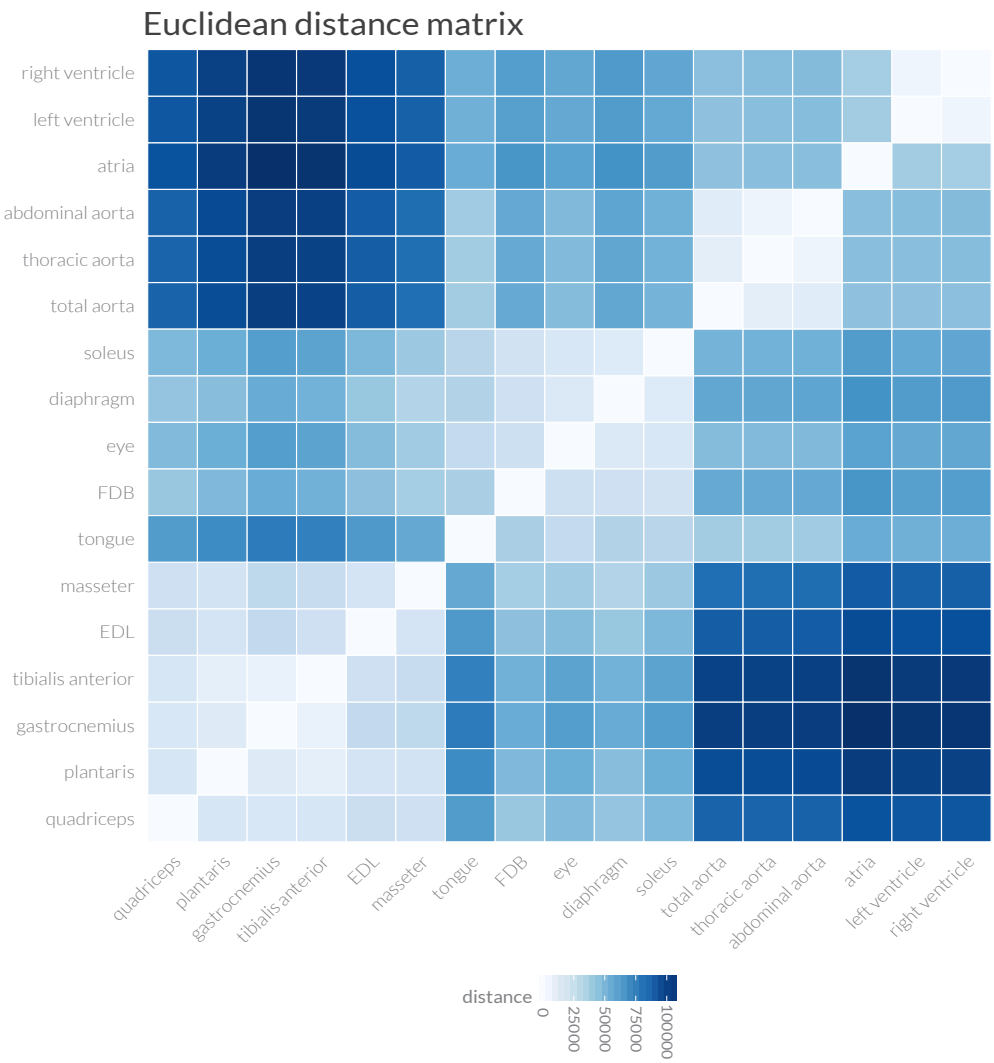
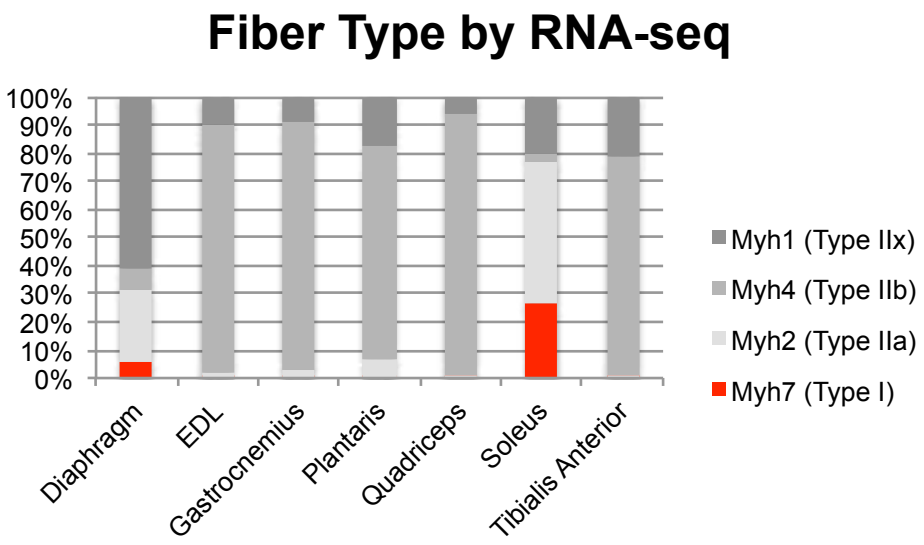

Euclidean distance matrix
right ventricle
left ventricle
atria
abdominal aorta
thoracic aorta
total aorta
soleus
diaphragm
eye
FDB
tongue
masseter
EDL
tibialis anterior
gastrocnemius
plantaris
quadriceps
quadriceps
plantaris
gastrocnemius
tibialis anterior
EDL
masseter
tongue
FDB
eye
diaphragm
soleus
total aorta
thoracic aorta
abdominal aorta
atria
left ventricle
right ventricle
distance
0
25000
50000
75000
100000

Figure S6:

A


Fiber Type by RNA-seq
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Diaphragm
EDL
Gastrocnemius
Plantaris
Quadriceps
Soleus
Tibialis Anterior
Myh1 (Type IIx)
Myh4 (Type IIb)
Myh2 (Type IIa)
Myh7 (Type I)

B

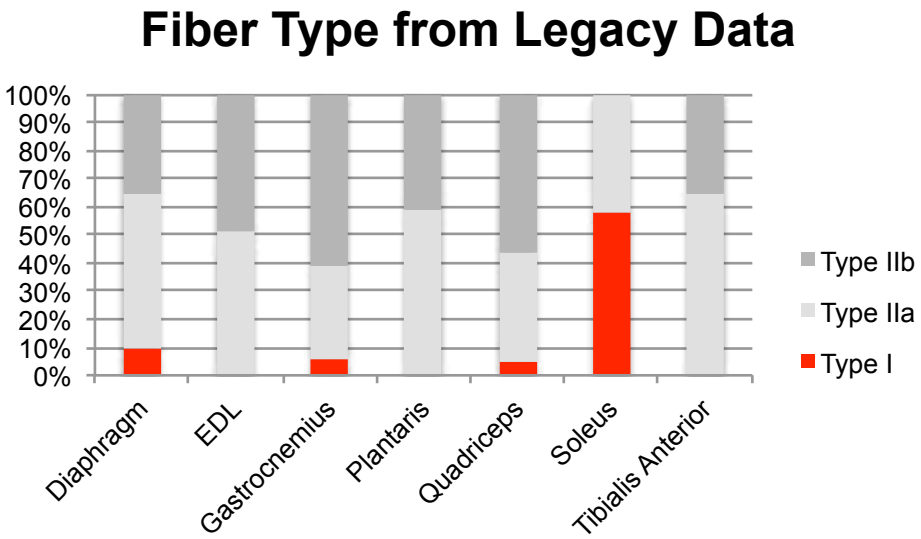
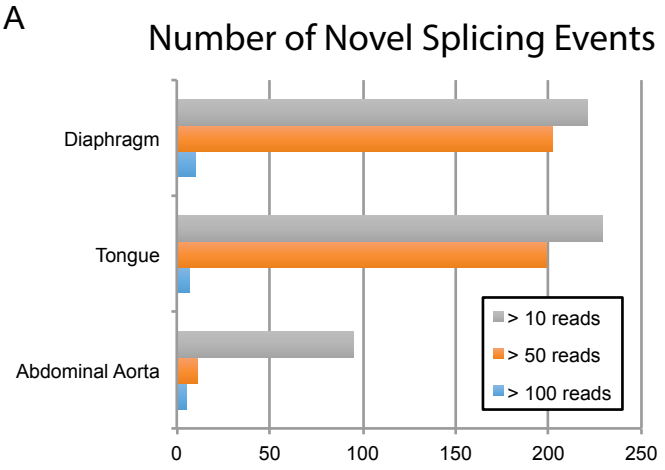
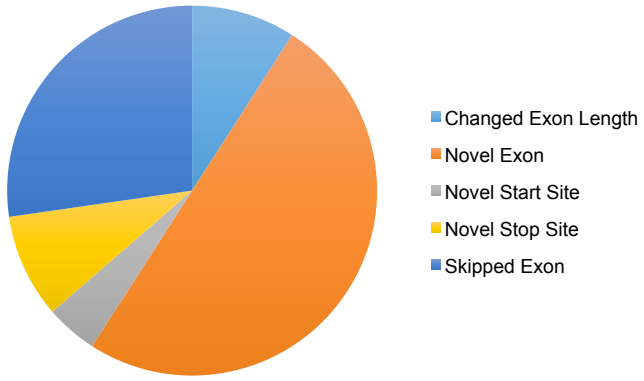

Fiber Type from Legacy Data
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Diaphragm
EDL
Gastrocnemius
Plantaris
Quadriceps
Soleus
Tibialis Anterior
Type IIb
Type IIa
Type I

Figure S7:

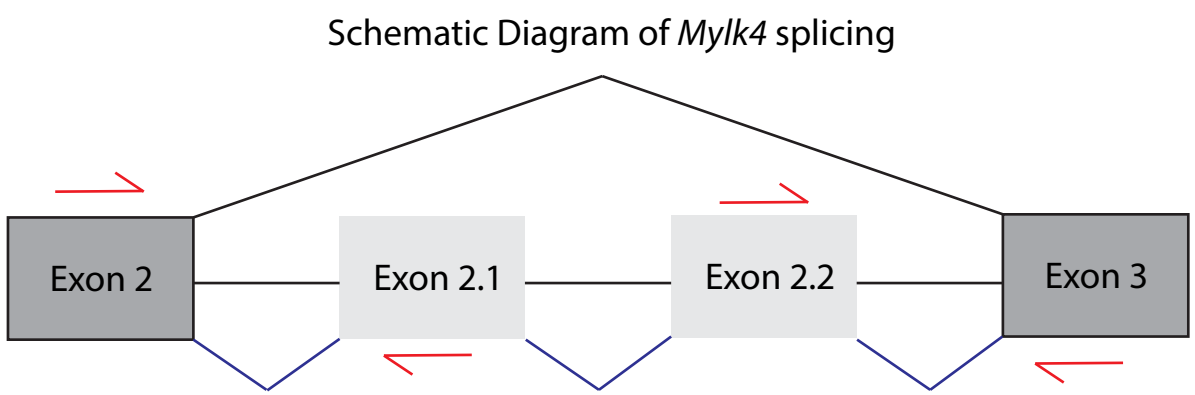
A



B



C



D

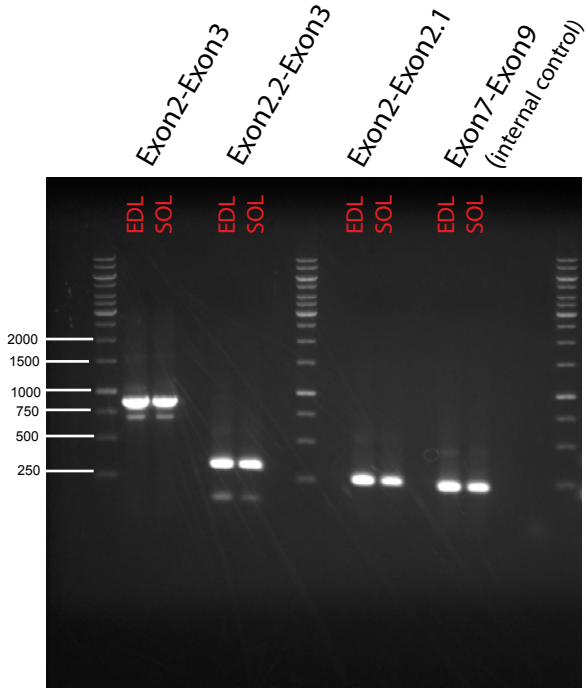
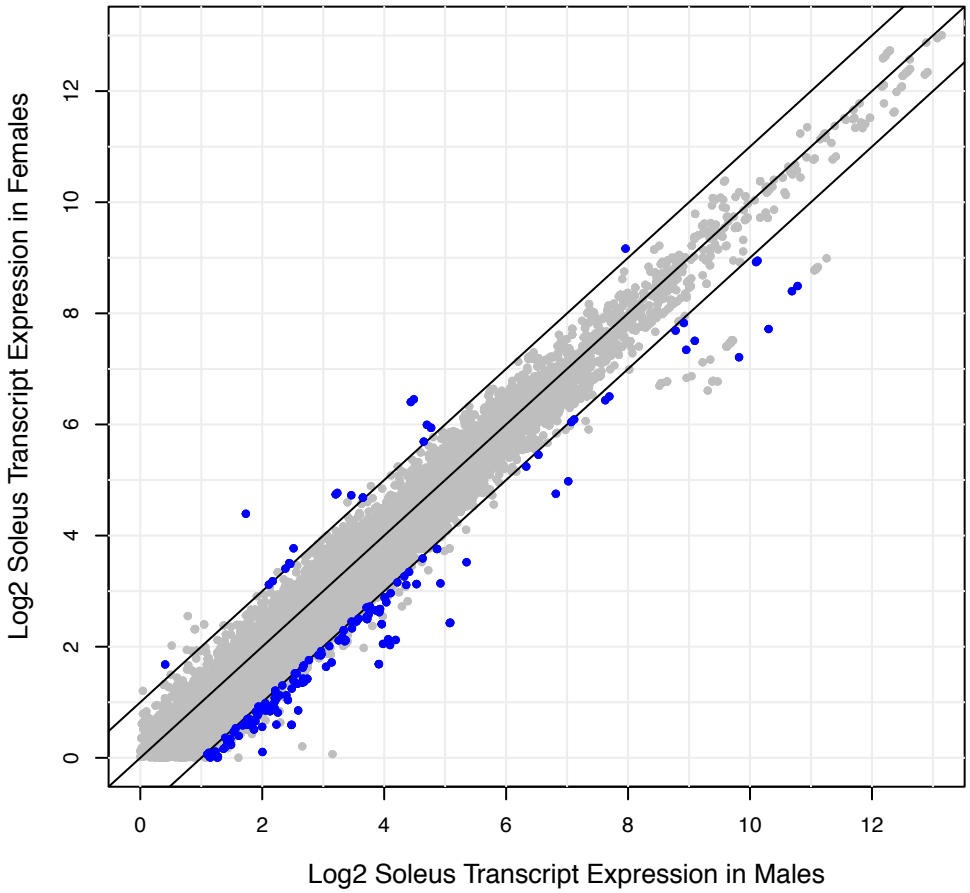


Figure S8

A


Log2 Soleus Transcript Expression in Females
Log2 Soleus Transcript Expression in Males
0
2
4
6
8
10
12

B

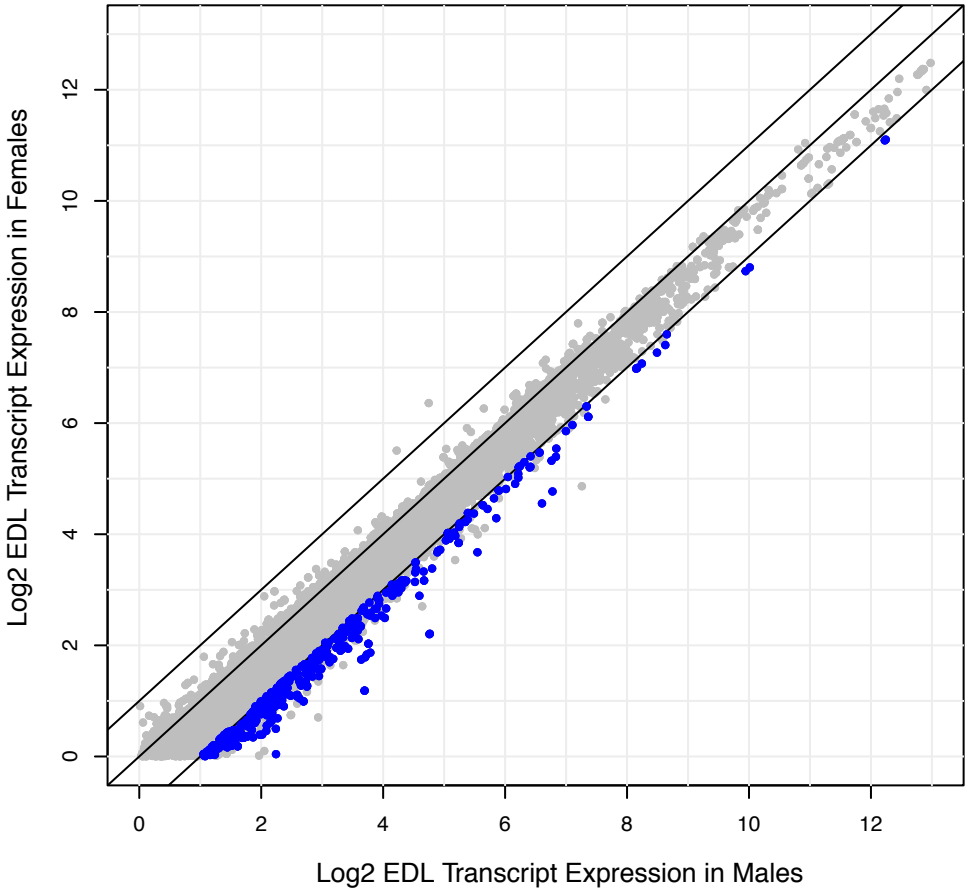

Log2 EDL Transcript Expression in Females
Log2 EDL Transcript Expression in Males
0
2
4
6
8
10
12

Figure S9:

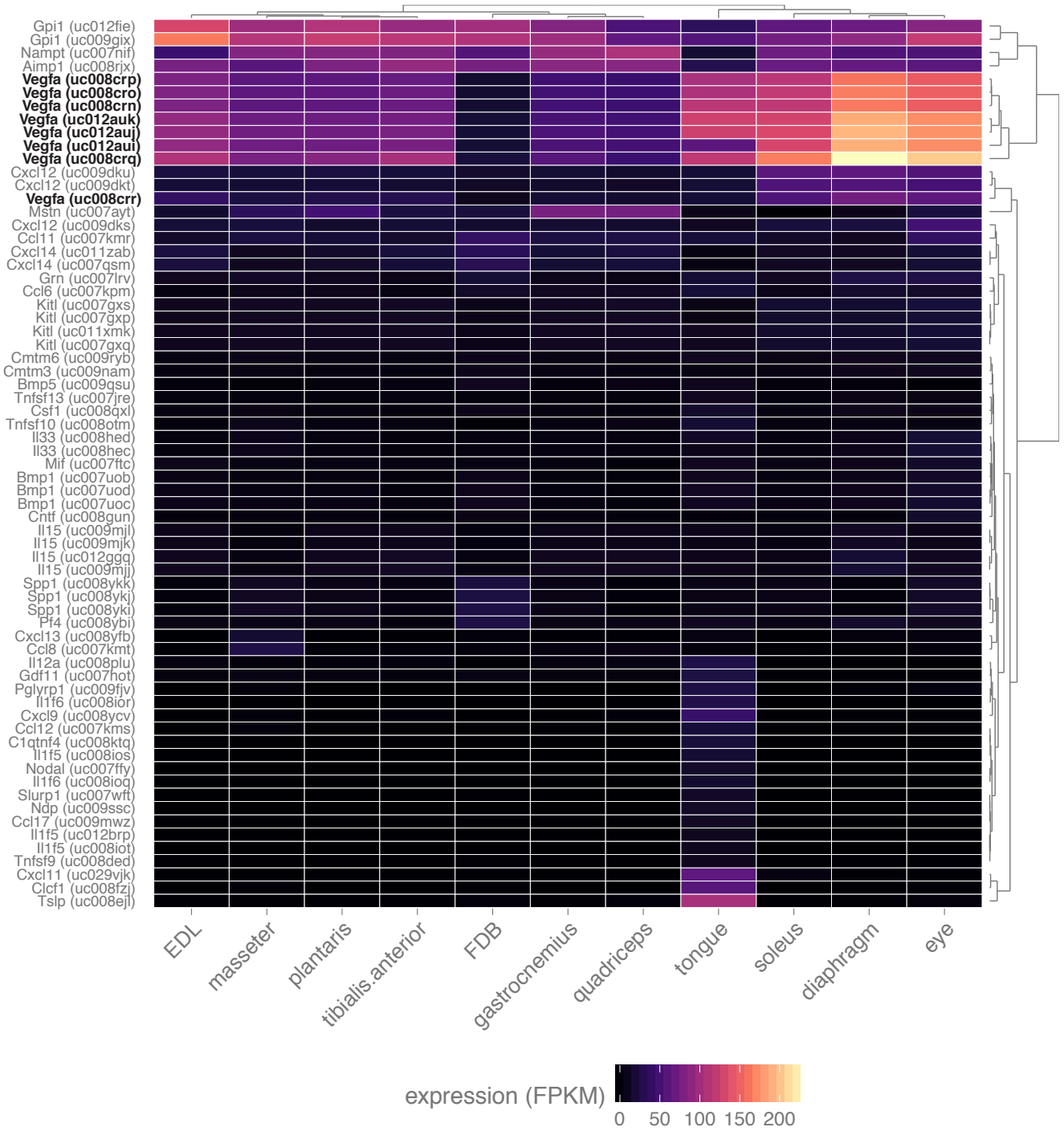
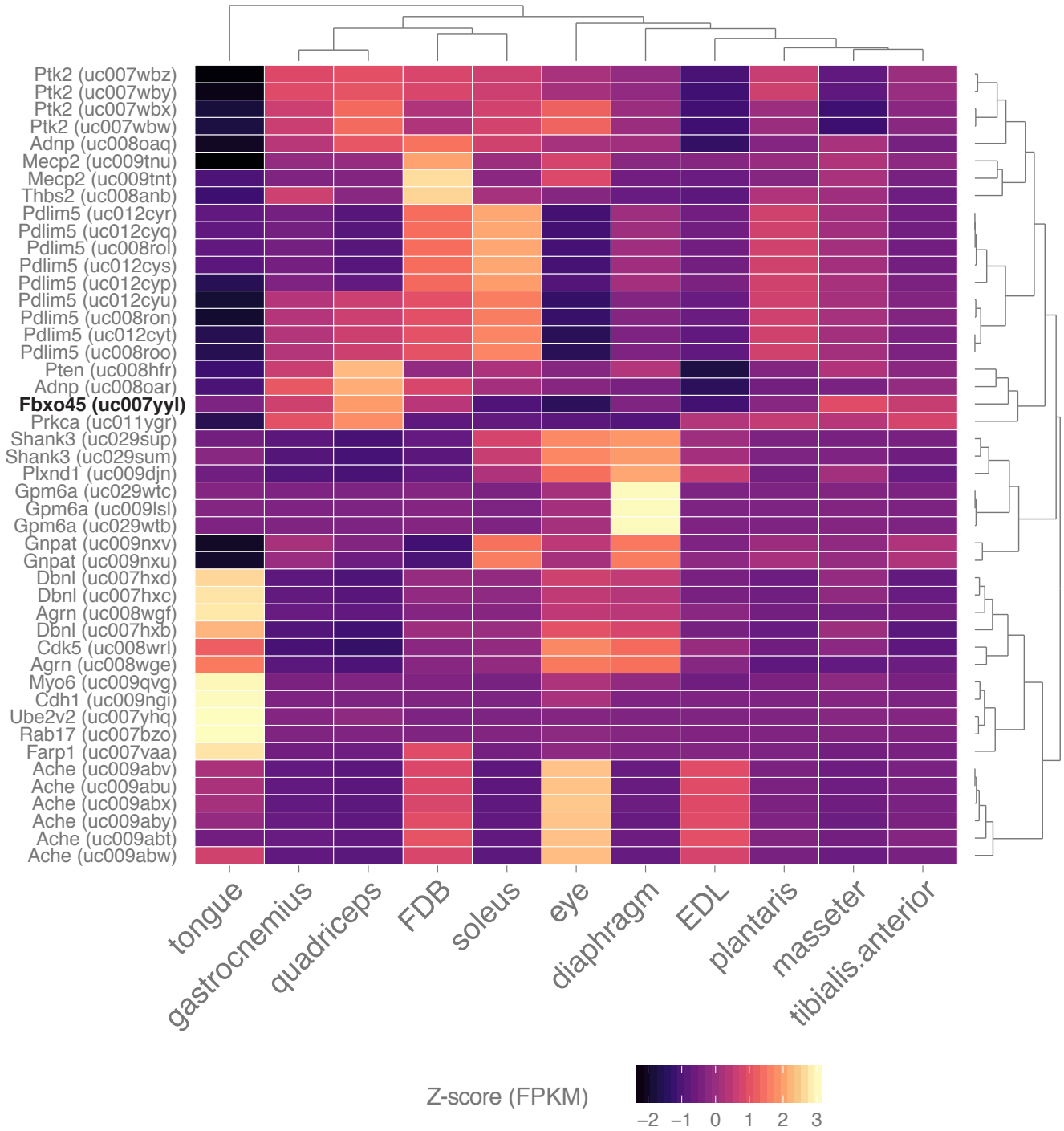

Gpi1 (uc012fie)
Gpi1 (uc009gix)
Nampt (uc007nif)
Aimp1 (uc008rjx)
Vegfa (uc008crp)
Vegfa (uc008cro)
Vegfa (uc008crn)
Vegfa (uc012auk)
Vegfa (uc012auj)
Vegfa (uc012aui)
Vegfa (uc008crq)
Cxcl12 (uc009dku)
Cxcl12 (uc009dkt)
Vegfa (uc008crr)
Mstn (uc007ayt)
Cxcl12 (uc009dks)
Ccl11 (uc007kmr)
Cxcl14 (uc011zab)
Cxcl14 (uc007qsm)
Grn (uc007lrv)
Ccl6 (uc007kpm)
Kitl (uc007gxs)
Kitl (uc007gxp)
Kitl (uc011xmk)
Kitl (uc007gxq)
Cmtm6 (uc009ryb)
Cmtm3 (uc009nam)
Bmp5 (uc009qsu)
Tnfsf13 (uc007jre)
Csf1 (uc008qxl)
Tnfsf10 (uc008otm)
Il33 (uc008hed)
Il33 (uc008hec)
Mif (uc007ftc)
Bmp1 (uc007uob)
Bmp1 (uc007uod)
Bmp1 (uc007uoc)
Cntf (uc008gun)
Il15 (uc009mjl)
Il15 (uc009mjk)
Il15 (uc012ggq)
Il15 (uc009mjj)
Spp1 (uc008ykk)
Spp1 (uc008ykj)
Spp1 (uc008yki)
Pf4 (uc008ybi)
Cxcl13 (uc008yfb)
Ccl8 (uc007kmt)
Il12a (uc008plu)
Gdf11 (uc007hot)
Pglyrp1 (uc009fjv)
Il1f6 (uc008ior)
Cxcl9 (uc008ycv)
Ccl12 (uc007kms)
C1qtnf4 (uc008ktq)
Il1f5 (uc008ios)
Nodal (uc007ffy)
Il1f6 (uc008ioq)
Slurp1 (uc007wft)
Ndp (uc009ssc)
Ccl17 (uc009mwz)
Il1f5 (uc012brp)
Il1f5 (uc008iot)
Tnfsf9 (uc008ded)
Cxcl11 (uc029vjk)
Clcf1 (uc008fzj)
Tslp (uc008ejl)
EDL
masseter
plantaris
tibialis.anterior
FDB
gastrocnemius
quadriceps
tongue
soleus
diaphragm
eye
expression (FPKM)
0
50
100
150
200

Figure S10:


Ptk2 (uc007wbz)
Ptk2 (uc007wby)
Ptk2 (uc007wbx)
Ptk2 (uc007wbw)
Adnp (uc008oaq)
Mecp2 (uc009tnu)
Mecp2 (uc009tnt)
Thbs2 (uc008anb)
Pdlim5 (uc012cyr)
Pdlim5 (uc012cyq)
Pdlim5 (uc008rol)
Pdlim5 (uc012cys)
Pdlim5 (uc012cyp)
Pdlim5 (uc012cyu)
Pdlim5 (uc008ron)
Pdlim5 (uc012cyt)
Pdlim5 (uc008roo)
Pten (uc008hfr)
Adnp (uc008oar)
Fbxo45 (uc007yyl)
Prkca (uc011ygr)
Shank3 (uc029sup)
Shank3 (uc029sum)
Plxnd1 (uc009djn)
Gpm6a (uc029wtc)
Gpm6a (uc009lsl)
Gpm6a (uc029wtb)
Gnpat (uc009nxv)
Gnpat (uc009nxu)
Dbnl (uc007hxd)
Dbnl (uc007hxc)
Agrn (uc008wgf)
Dbnl (uc007hxb)
Cdk5 (uc008wrl)
Agrn (uc008wge)
Myo6 (uc009qvg)
Cdh1 (uc009ngi)
Ube2v2 (uc007yhq)
Rab17 (uc007bzo)
Farp1 (uc007vaa)
Ache (uc009abv)
Ache (uc009abu)
Ache (uc009abx)
Ache (uc009aby)
Ache (uc009abt)
Ache (uc009abw)
tongue
gastrocnemius
quadriceps
FDB
soleus
eye
diaphragm
EDL
plantaris
masseter
tibialis.anterior
Z-score (FPKM)
−2 −1 0 1 2 3