Fine mapping of genes determining extrafusal fiber properties in murine soleus muscle


Published in: PHYSIOLOGICAL GENOMICS

Document Version: Peer reviewed version

Queen's University Belfast - Research Portal: Link to publication record in Queen's University Belfast Research Portal

Publisher rights
© 2016 The American Physiological Society
This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person’s rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.
Fine mapping of genes determining extrafusal fiber properties in murine soleus muscle

Carroll AM\textsuperscript{1,2}, Cheng R\textsuperscript{3}, Collie-Duguid ESR\textsuperscript{1,4}, Meharg C\textsuperscript{5}, Scholz ME\textsuperscript{1}, Fiering S\textsuperscript{6}, Fields JL\textsuperscript{6}, Palmer AA\textsuperscript{7}, Lionikas A\textsuperscript{1}.

Affiliation:
\textsuperscript{1}School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Foresterhill Aberdeen, AB25 2ZD, UK
\textsuperscript{2}The New Zealand Institute for Plant & Food Research Limited, Batchelor Road, Private Bag 11600, Palmerston North 4442, New Zealand
\textsuperscript{3}Research School of Biology, The Australian National University, 134 Linnaeus Way, Acton, ACT 2601, Australia
\textsuperscript{4}Centre for Genome Enabled Biology and Medicine, University of Aberdeen, Old Aberdeen, Aberdeen, AB24 3RY, UK
\textsuperscript{5}Institute for Global Food Security, Queen's University Belfast, 18-30 Malone Road, Belfast, BT9 5BN, UK
\textsuperscript{6}Department of Microbiology/immunology, Geisel School of Medicine at Dartmouth, Hanover, NH 03755, USA
\textsuperscript{7}Department of Psychiatry, University of California San Diego, 9500 Gilman Drive, Mail Code 0667, BRF2 3A24, La Jolla, CA 92093, USA
Abstract

Introduction. Muscle fiber cross-sectional area (CSA) and proportion of different fiber types are important determinants of muscle function and overall metabolism. Genetic variation plays a substantial role in phenotypic variation of these traits, however, the underlying genes remain poorly understood.

Aims. This study aimed to map quantitative trait loci (QTL) affecting differences in soleus muscle fiber traits between the LG/J and SM/J mouse strains.

Methods. Fiber number, CSA, and proportion of oxidative type I fibers were assessed in the soleus of 334 genotyped female and male mice of the F₃₄ generation of advanced intercross lines (AIL) derived from the LG/J and SM/J strains. To increase the QTL detection power, these data were combined with 94 soleus samples from the F₂ intercross of the same strains. Transcriptome of the soleus muscle of LG/J and SM/J females was analysed using microarray.

Results. Genome-wide association analysis mapped 4 QTL (genome-wide p<0.05) affecting the properties of muscle fibers to Chromosome 2, 3, 4 and 11. A 1.5-LOD QTL support interval ranged between 2.36 Mb and 4.67 Mb. Based on the genomic sequence information, functional and transcriptome data, candidate genes were identified for each of these QTL.

Conclusion. Combination of analyses in F₂ and F₃₄ AIL populations with transcriptome and genomic sequence data in the parental strains is an effective strategy for refining QTL and nomination of the candidate genes.

Key words: skeletal muscle, muscle fiber types, genetic variation
Introduction

Skeletal muscle plays a broad range of biological functions including locomotion, thermoregulation, respiration, postural support, protection of bones and viscera; as well as serving as a source of amino acids in times of starvation or disease. Muscle tissue in livestock also provides an essential source of dietary proteins. In humans, there is more than a 2-fold difference in muscle mass between individuals of similar age and same sex (3, 33). This is the outcome of variability in the number of muscle fibers and their size (51). These differences are of clinical relevance. Variability in muscle mass significantly impacts energy expenditure (58), influencing preponderance to obesity. In addition, individuals with lower muscle mass may be more vulnerable to impairment of these vital functions due to aging and/or disease related muscle loss. It has recently been reported that there is a positive association between muscle mass and longevity in older adults (66).

Human skeletal muscles are mainly comprised of a mixture of type I, IIA and IIX muscle fibers (62). The number of fibers, their size and varying proportions of the fiber types affect morphological and functional properties of the muscle (6). A larger diameter of the fibers and higher number of fibers typically leads to augmented muscular strength and power (25, 28). The proportion of type I muscle fibers is a factor determining success in endurance sporting events (15, 18) and overall metabolism in humans (24, 29, 44, 74). In livestock, proportion of oxidative type I fibers is associated with meat quality (65).

In humans, genetic factors account for around half of the variation in strength (19, 24, 74) and the upper limit heritability is even greater (over 0.9) for muscle mass (26). Heritability estimates of proportion of type I fibers is also high, ranging between 0.4 and 0.9, indicating that genetic factors play an important role in determining muscle fiber properties (37, 63). Effects of genetic factors on muscle fibers have also been demonstrated in mouse (20, 22,
However, the specific genes underlying these effects remain largely elusive.

Attempts at mapping the polygenic architecture of muscle fiber properties in mouse (11), pig (17, 43, 52, 55, 77), cattle (1) and carp (80) have been made. A number of QTL have been identified in these studies. However, the resolution achieved in the F2 population is not adequate for reliable nomination of the candidate genes in the majority of the QTLs of polygenic traits. The mouse soleus muscle (primarily consists of type I and IIA fiber types), closely resembles the fiber type composition of human skeletal muscles (primarily comprised of type I, IIA and IIX fiber types), and is therefore a particularly interesting experimental model. In our previous study, we mapped soleus muscle fiber traits in an F2 intercross between the LG/J and SM/J laboratory mouse strains (11). These strains differ in a number of muscular phenotypes, with the LG/J strain displaying a greater proportion of type I fibers, and a greater cross-sectional area (CSA) of type I and IIA muscle fibers. We identified in that study three significant QTLs contributing to the difference in the CSA of muscle fibers between LG/J and SM/J strains (11). Regions of conserved synteny from the identified loci were also implicated in fiber phenotypes in pig supporting the importance of these genomic regions in determining muscle fiber properties. However, the exact genes underlying their effects remain to be determined.

Integration of advanced study populations, high throughput gene expression technology and increasing availability of knockout models aid identification of the causative genes. Nomination of the genes underlying QTL effects can be facilitated by improving the mapping resolution, and by utilising genomic sequence and transcriptome information. Advanced intercross lines (AIL) have been proposed as a powerful population for mapping QTLs (16). It has been demonstrated recently that a joint F2 and AIL analysis can combine the
advantages of both mapping populations by increasing the power to detect QTLs and achieving a higher mapping resolution of various traits in mice (13, 47). Additionally, testing for differences in specific gene expression has led to several nominations of quantitative trait genes (30, 35). For validation of such candidate genes, phenotypic effects of relevant alleles can be examined in experimental populations where these alleles segregate albeit on a different genetic background. In addition, available knockout models offers particularly attractive option for validation experiments.

In the present study we aimed to fine-map QTL and nominate candidate genes affecting the CSA and proportion of oxidative type I fibers in the soleus muscle in a combined analysis of F2 and F34 AIL mice, and by cross referencing QTL data with soleus transcriptome profiles in the parental strains. Further filtering of the emerged candidates was carried out in an independent AIL and a knockout model.
Methods

Muscle Samples

This study was carried out on soleus muscles dissected from females and males of the F_{34} advanced intercross lines (AIL) of the LG/J and SM/J inbred strains. Animals were maintained as previously described (13) and sacrificed at 94 ± 4 days. All procedures were approved by the Institutional Animal Care and Use Committee of the University of Chicago.

Soleus muscle samples from F_{34} AIL mice described in our previous study (47) were subjected to histological analyses. The final sample size used in the present study was 334 F_{34} mice, 142 females and 192 males, after discarding samples of poor tissue quality. A set of 94 F2 samples (38 females and 56 males) described in our previous study (11) was also used in order to increase the QTL detection power.

In addition, we also analysed soleus muscle samples for two hypothesis driven studies aimed at testing the effects of identified candidate genes on percentage of oxidative, type I fibers. First, we examined solei samples from the Chd6 ATPase knockout (n=6), heterozygous (n=4) and wild type (n=4) females. The generation of the Chd6 mutant mice has been previously reported (40). Briefly, the genetic manipulation generated an allele with the ATPase domain of Chd6 (exon 12) flanked by loxP sites so that the action of Cre recombinase would delete this domain. The mice were mated to a germline Cre-expressing strain (Jackson lab strain 003465) to delete both exon 12 and the neomycin resistance marker used for the targeting. Subsequently breeding generated the Chd6 ATPase knockout mice utilized in the present study. Second, solei of the advanced intercross mice (generations F_{9}-F_{12}), all homozygous carries of the C57BL/6J (n=22) or DBA/2J (n=23) alleles at the region harbouring the Alad gene were selected from the tissue bank of our previous study (9).

Phenotype assessment
The soleus muscles were frozen in isopentane cooled in liquid nitrogen. Transverse sections from the belly of the muscle were cut at a thickness of 10 µm with a cryotome (Leica CM1850UV) at -20°C. The muscle sections were subjected to ATPase staining (acid pre-incubation, pH 4.47) to distinguish between fiber types (8). Microscopic images of stained sections were taken at x5 and x20 magnification.

The following phenotypes were assessed: muscle fiber number (type I and IIA) and percent of type I muscle fibers, cross-sectional area (CSA) of type I and type IIA fibers (Figure 1). Muscle fiber traits were manually analysed using ImageJ software (NIH-version 1.43). 25 measurements of each fiber type were taken using the freehand selection tool at x20 magnification to obtain a value representing the mean CSA of type I or type IIA fibers for that muscle. This was deemed as a representative sample by empirical testing as described previously (11). Total number of type I and type IIA muscle fibers were counted using the ImageJ cell counter plugin on x5 magnification images. As all fibers in mouse soleus pass through the belly of the muscle (69), this method provides an accurate estimate of the number of fibers constituting the muscle. Total number of type I fibers and total number of type IIA fibers were counted, permitting derivation of percentage of type I fibers. Over the course of the study ~200,000 muscle fibers were counted and ~6,700 fibers measured for CSA.

Statistical analyses

The GraphPad Prism version 5.0 statistical package was used (GraphPad software, La Jolla, CA). Data are presented as mean ± SD, unless otherwise stated. The CSA of type I and type IIA fibers were analysed using a two-way (sex and fiber type) paired-measures (type I and type IIA fibers) ANOVA.

Genotyping and QTL mapping

Mice were genotyped using a custom designed SNP array that included 4,610 polymorphic SNPs that were approximately evenly distributed across the genome, as described
previously (13). The genome-wide association analysis was performed in the combined population of the F_{34} and recently published F_{2} intercrosses (11) using the R package QTLRel (12). This software accounted for the complex relationships (e.g., sibling, half-sibling, cousins) among the F_{34} mice by using a mixed model, as previously described (12, 13). Due to the sex differences in muscle mass in these mice (47), and the discovery of sex specific QTL in other studies (45, 46), we included sex as an additive and interacting covariate. Threshold of significance was estimated by 1000 permutations (14). We defined the support interval for each QTL as the 1.5-LOD drop off on either side of the peak marker. This interval was expressed in physical map position (Mb) by using the nearest genotyped SNP that flanked the support interval, based on the mouse genome build GRCm38.p3.

**Transcriptome analysis**

Soleus muscle tissues from 92-day old LG/J and SM/J females (n=3 of each strain) were used. RNA was isolated using TRIlzol (Invitrogen Life Technologies, Carlsbad, CA) followed by purification and DNase digestion using RNeasy minikits (Qiagen, Venlo, Netherlands) according to the manufacturer’s instructions. Quantification of total RNA was performed on a NanoDrop spectrophotometer (Thermo Scientific) and quality tested on an Agilent Tapestation with R6K Screentapes (RIN ≥7.3). Generation of sense strand cDNA from purified total RNA (Ambion® WT expression kit, Ambion, Austin, Texas) followed by fragmentation and labelling (GeneChip WT labelling kit, Affymetrix, Santa Clara, CA) were performed according to the manufacturer’s instructions. Hybridisation, washing, staining and scanning of microarrays were carried out on Affymetrix Mouse Gene 2.0 ST microarrays according to the manufacturer’s standard protocols using a GeneChip Fluidics station 450 and GCS3000 scanner (Affymetrix®, Santa Clara, CA). Microarray data are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-5290.
Data pre-processing and quality control analysis was performed using Affymetrix®
Genechip® Expression Console v1.2. Probe cell intensity data on the Mouse Gene 2.0 ST
array (CEL files) were processed using the RMA16 algorithm (Affymetrix, Santa Clara, CA,
USA) which fits a robust linear model at probe level by employing background correction,
quantile normalisation of log2 transformed data and summarisation to probe level data (CHP
files, 41,345 probe sets).

Data was analysed for differentially expressed genes in Partek® Genomics Suite® version
6.6, build 6.15.0730 (Partek Inc., St Louis, MO) using a Mus musculus build mm10
annotation file for Mouse Gene 2.0 ST microarrays (MoGene-2_0-st-v1.na35.mm10). CEL
files (Expression Console v 1.2, Affymetrix, Santa Clara, CA) were imported to Partek
Genomics Suite v 6.6 and processed using RMA normalisation with background correction
of log2 transformed data and probe set summarisation by median polish. Differential
expression analysis between the LG/J and SM/J strains of all genes (n=41,345 transcript
clusters) was determined by 1-way ANOVA with Storey’s FDR, and q-value ≤0.05
considered significant (n=819 genes differentially expressed ≥ 1.2 fold; see Supplementary
Table 1).

To assess transcription of positional candidate genes in each strain, a hypothesis driven
analysis of differential gene expression was performed between the LG/J and SM/J strains
on all genes mapping to the support interval defined for each QTL in the GWAS described
above. Using Partek Genomics Suite v.6.6, a total of 159 genes that were represented on
the mouse Gene 2.0ST microarray, were identified in Mus musculus genome build
GRCm38, mm10 within mapping co-ordinates Chr2:158908559–162608559 (26 genes),
Chr3:33308451–35708451 (15 genes), Chr4:57605946–62913639 (77 genes) or Chr
11:27900000–31500000 (41 genes). 1-way ANOVA identified differentially expressed genes
between the LG/J and SM/J strains (P<0.05). Fold change was calculated using the
geometric mean of samples in each group.
Candidate genes

Nomination of the candidate genes was based on the following three criteria. First, we scrutinized polymorphisms in positional candidates between the LG/J and SM/J strains. The emphasis was on the indels and SNPs that would affect the coding sequence and lead to changes in amino acids. To assess whether amino acid substitution would influence the function of a protein, evolutionary conservation at the site of substitution and properties of substituted amino acids were considered using three different bioinformatics tools as described by Nikolskiy and colleagues (56). Second, we examined expression of positional candidates across a panel of over ninety mouse tissues and cell types available in BioGPS GeneAtlas MOE430, gcrma dataset (79). This analysis permits a quantitative comparison of transcript abundance of a gene between tissues. We considered that an abundant expression in skeletal muscle lineage, i.e. muscle tissue and/or C2C12 myogenic cell line, implies functional and/or structural relevance of a gene in this tissue. Third, we compared gene expression levels in the soleus muscle between the two strains as described in the previous section. Expression difference in this analysis might point at the strain-specific, genotype-dependent mechanism underlying the phenotypic difference.

Results

Phenotypic analyses

CSA. Cross section analysis of soleus muscle fibers were done on mice of both sexes from the F_{34} cohort. For muscle fiber cross-sectional area, we observed a statistically significant sex by fiber type interaction (P<0.0001). In the female F_{34} mice there was no significant difference between type I and type IIA muscle fiber areas (913 ± 229 μm², n=140; and 952 ± 242 μm², n=140 respectively; P=0.2). However, there was a significant difference within the males, with the type I muscle fiber area being smaller than IIA fiber area (1084 ± 238 μm²,
Muscle fiber area was lower in females than males for type I CSA, (P < 0.0001) and type IIA CSA (P < 0.0001).

**Percentage of type I fibers.** The number of type I fibers as a percentage of total fibers varied substantially between individuals, ranging from 30% to 67% in females, and from 26% to 59% in males (Figure 1) and was greater in females than males (46 ± 8%, n=142; and 39 ± 6%, n=189; respectively; P < 0.0001).

**Total fiber number.** No difference was observed in the total soleus fiber number between females and males (646 ± 102, n=120, and 667 ± 105, n=177, respectively; P= 0.0979).

**QTL analyses**

Muscle fiber traits approximated the normal distribution in both the F2 and F34 population (Supplementary Figure 1). We identified significant QTL (at the 1% or 5% level of genome-wide statistical significance) (39) for CSA of type I and type IIA fibers and the percentage of type I fibers. We also identified chromosome-wide significant QTL for CSA of type I and type IIA fibers, the percentage of type I fibers and total fiber number (Table 1). The size of the support interval of these QTL ranged from 0.4-40.7 Mb, with a median of 4.6 Mb.

The QTL at the genome-wide level of significance for CSA of type I and type IIA fibers on chromosome 3 was named *Mfq5*. The QTL at the genome-wide level of significance for the percentage of type I fibers on chromosome 2 and 4 were named *Mfq4* and *Mfq6*, respectively. The SM/J allele conferred a greater percentage of type I fibers at *Mfq4*, and a greater CSA at *Mfq5*. The LG/J allele conferred a greater percentage of type I fibers at *Mfq6* locus.

A significant QTL affecting CSA of type I and type IIA fibers was also detected on chromosome 11 (Figure 2) within the same region as locus *Mfq3*, previously identified in the F2 intercross of the same parental strains (11). The QTL exhibited male-specificity in both
type I and IIA fibers of the F34 mice (Figure 3). Because this QTL recapitulated properties of
the Mfq3 locus, which we also found to be male specific in the F2 population, we concluded
that the same locus has been refined in F34 and did not assign a new name for this QTL.
Earlier reported Mfq2 locus has been refined in a similar manner; a QTL on chromosome 6
affecting CSA of type I and type IIA fibers (at 1% chromosome specific threshold) was
engulfed by the support interval of Mfq2 and also replicated its increasing allele, LG/J, in
both females and males (not shown).

Gene expression analyses

We hypothesized that each identified QTL harbours one or more genetic variants that drive
phenotypic differences by means of differential gene expression. Hypothesis driven analysis
of differential expression in soleus muscle was performed between LG/J and SM/J strains for
the genes in the most robust QTLs affecting fiber CSA or % Type I fibers (Mfq3, Mfq4, Mfq5
and Mfq6). The Mouse Gene 2.0 ST expression array contains 159 genes residing within
the support intervals of these QTLs (Supplementary Table 2). Twenty genes (Table 2)
showed evidence of differential expression (ANOVA, p≤0.05), 2 of which, Alad and Hdhd3,
were significant after correction for the multiple testing problem (Storey’s FDR q≤0.05).
Compared to other tissues and cell types, expression of differentially expressed genes Mafb,
Acyp2 and Mtif2 (Table 2), is particularly enriched in skeletal muscle (BioGPS, Mouse
MOE430 gene expression data).

Genomic analyses

Positional candidates with non-synonymous polymorphisms provide a plausible genetic
cause for the phenotypic differences. Based on the genomic sequence of the LG/J and
SM/J strains (56), we identified 21 genes in the QTL regions with non-synonymous
polymorphisms predicted to affect protein function by at least one out of three algorithms
used in the analysis (Supplementary Table 3). Four of those genes (Mfq3: Mtif2, Rtn4,
Psme4; Mfq5: Dnajc19) are prioritized further because of their preferential expression in
muscle lineage (differentiated muscle and/or C2C12 myoblasts) compared to other tissues and cell types. Among those, the Mtif2 gene differs by 3 (rs26871496, rs26871494, rs29436813) and Rtn4 by 9 (rs29473364, rs29469198, rs13463765, rs29465940, rs26857726, rs26857725, rs29474377, rs26857722, rs26857721) amino acids between the two strains. At all SNPs the SM/J strains carries reference while the LG/J strain the alternative allele.

Candidate gene analyses

The Chd6 gene emerged as a differentially expressed positional candidate for the Mfq4 locus affecting percentage of type I fibers (Table 2). To test its effect we examined soleus muscles of Chd6 knockout, heterozygous and wild type littermates. This analysis however revealed that the genotype of the animals did not have a significant effect (P=0.30) on the percentage of type I fibers (Figure 4).

The Alad gene emerged as a candidate for another locus affecting proportion of type I fibers, Mfq6. In the animals of an advanced intercross between the C57BL/6J and DBA/2J strains (these strains carry one or three copies of Alad, respectively (3)), we examined if percentage of type I fibers was genotype-dependent. The analysis revealed no difference in the percentage of type I fibres between the carriers of the C57BL/6J and the DBA/2J alleles, 42 ± 7% and 42 ± 8%, respectively.
Discussion

A previous study on muscle weight in LG/J and SM/J strains identified a two-fold difference in soleus muscle size (47). We then explored the cellular and genetic mechanisms contributing to this phenomenon, finding that the difference was largely due to the CSA of muscle fibers and we mapped QTL affecting muscle fiber traits in an F$_2$ intercross between the LG/J and SM/J strains (11). The present study, which utilizes the F$_{34}$ advanced intercross, verified, refined and expanded our earlier findings.

A number of studies have previously reported the effects of Stat5a and Stat5b (36), Pgc-1$\alpha$ (42), Ky (4), myostatin (54), leptin (61), calcineurin (76), Sod1 (5), alpha-actinin-3 (50), dystrophin (7), Tbx15 (41) and IIB myosin heavy chains (2) genes on muscle fiber area in knockout or mutant models. In addition, Pgc-1$\alpha$ (75), calcineurin (76), Foxo1 (34) and myostatin (20) are reported to affect the proportion of muscle fiber types. However, the genomic positions of these genes have not been linked to muscle fiber differences between the LG/J and SM/J strains, implicating involvement of novel genes.

Muscle fiber number. The number of fibers is an important determinant of muscle size and functional properties. It is set during embryogenesis and the first post-natal week in mice (78). The number of muscle fibers in males (667 ± 105) and females (646 ± 102) of the F$_{34}$ population was comparable to that observed in the soleus of the F$_2$ population (645 ± 102 and 595 ± 107, respectively), and within the range of the fiber count observed in solei of a variety of different strains of mice ~250--900 fibers (32, 49, 57, 70, 72).

From these data it emerged that males and females are born with a similar number of fibers in soleus muscle, and that the sex difference in muscle weight (males have approximately 30% larger soleus than females) is due to the difference in fiber size. Comparison of the parental strains also revealed a similar number of fibers (11), despite the 2-fold difference in soleus weight (47), demonstrating that size rather than number of fibers determines variation in muscle weight between the LG/J and SM/J strains.
Fiber area. The CSA of muscle fibers in the LG/J strain is 49% to 90% greater than the corresponding fibers in the SM/J strain, indicating that this variable accounts for a large portion of the muscle mass difference between the strains (47).

The area of type I (1084 ± 238 μm² and 913 ± 229 μm² for males and females, respectively) and type IIA (1215 ± 294 μm² and 952 ± 242 μm², respectively) of the F34 mice was comparable to the corresponding fiber area of the F2 mice of the same lineage (11) and it is within the range reported for the type I, between 920 μm² and 1780 μm² (32, 57, 70), and type IIA fiber area, between 700 μm² and 1400 μm² (32, 70), in various inbred mouse strains.

Percentage of type I fibers. The percentage of type I fibers in male (39 ± 6%) and female (47 ± 8%) F34 mice were also within the range of previous studies, which showed the percentage of type I fibers in the soleus muscle fluctuates between ~25 and ~66% (32, 57, 70).

In the F34 mice we replicated our observation in the F2 population that the percentage of type I fibers was significantly greater in females than males. This sex difference is also observed in various human muscles where, in general, women have a higher percentage of type I muscle fibers than males (27, 53, 60, 64, 67). The phenomenon is likely to be explained, at least partly, by the effect of androgens; castration leads to a higher percentage of type I fibers in the soleus of male mice (73).

Validation and refinement of genetic architecture. In the present study, we validated and refined the genetic architecture of muscle fibers identified in an F2 intercross between the same parental strains (11). In order to increase QTL detection power, we increased sample size by combining the F34 and F2 data. The median mapping resolution of 4.6 Mb for muscle fiber QTLs was comparable with 3.7 Mb of muscle weight QTLs obtained in the same population albeit using ~1,600 fewer genetic markers than in the present analysis (47). A genome-wide significant QTL identified in the present study between 27.9 Mb and 31.4 Mb
on chromosome 11 (Table 1) overlapped with a significant QTL, Mfq3, mapped in the F2 population (11). In addition to the chromosomal location, the increasing allele of this locus (LG/J) and its male-specific effect (Figure 3) were also replicated in F34, suggesting that the same gene(s) were involved in two different populations and permitting us to refine the Mfq3 locus from 51.6 Mb to 3.57 Mb. The presence of two satellite QTL proximal of the refined Mfq3 (Table 1) suggests that the QTL observed in the F2 population (11) might have been an outcome of up to three linked loci.

The recently reported “mini-muscle” locus, mapped to 67.1–70.2 Mb on chromosome 11, affects muscle fiber area and proportion of fiber types (21-23). However, the mutation responsible for the “mini-muscle” phenotype maps to an intron of Myh4 gene located at 67.2 Mb (31), between the support intervals of two adjacent QTLs affecting fiber type between the LG/J and SM/J strains (Table 1). Together, these data suggest that a number of genes residing on chromosome 11 might be involved in the regulation of muscle fiber phenotypes.

The QTL affecting the CSA of type I and type IIA fibers on chromosome 6, albeit at 1% chromosome-wide threshold of significance (Table 1), overlapped with the Mfq2 locus found in the F2 population, characterized by the same increasing allele, LG/J. Thus, the support interval of Mfq2 could be considered to be 5.18 Mb rather than the previously reported 56.5 Mb. Importantly, the immediate proximity of the refined region (Chr 6: 110.8-116.0 Mb) to the syntenic region (Chr 6:116.0-118.0 Mb) implicated in the QTL affecting the diameter of pig IIA fibers (17) suggest that the same genes could be underlying the effects of these QTLs in mice.

A QTL affecting percentage of type I fibers (at 10% chromosome-wide threshold) on chromosome 1 (67.6to 70.8 Mb) overlapped with Mfq1 locus which influenced the CSA of type I and type IIA fiber area in the F2 population (11). However, because the CSA and percentage of type I fibers are poorly correlated traits both in the F34 (Supplementary Table
it is likely that different genes are underlying the *Mfq1* locus and the QTL identified in the F\textsubscript{34} population. Further studies are required to clarify this observation.

Transcriptome analysis

In the present study, the expressed transcriptome in soleus muscle of the parental strains was examined in order to facilitate nomination of the candidate genes within the refined QTL. We hypothesized that if the phenotypic effect of the QTL was brought about by the allele specific abundance of transcripts encoded by genes within the QTL, such genes would be differentially expressed in the transcriptome between the parental strains. Comparison of expression of the genes within the four most robust QTLs identified *Alad* and *Hdhd3* genes as potential candidates for the *Mfq6* locus, which affects the proportion of type I fibers. Transcripts of both genes are more abundant in the LG/J compared to the SM/J strain. This is consistent with our findings in the TA muscle of the same strains (48). Of these two identified candidate genes, transcripts of *Alad* are ~20 times more abundant in the mouse muscle than *Hdhd3*, regardless of strain. In addition, *Alad* may play a role during myogenesis as its expression in C2C12 myogenic cells is 5-fold higher compared to differentiated muscle (79).

Candidate genes.

The support intervals of four most robust QTLs harbor 159 genes (*Supplementary Table 2*). These regions were scrutinized further for the genes fulfilling one of the following criteria: presence of the functional variants (i.e. non-synonymous SNPs predicted to alter function of encoded protein); abundance of transcript in muscle lineage, particularly in comparison to other tissues and cell types; differential expression in the soleus of the two strains; and by comparing genomic sequence between the LG/J and SM/J strains a list of 21 genes was highlighted (*Supplementary Table 3*) with the strain-specific functional variants. Using bioinformatics, 4 genes abundantly and/or preferentially expressed in skeletal muscle compared to other types of tissues and cells were identified. Our own analysis of gene
expression in soleus muscle highlighted a set of 20 genes differentially expressed between the two strains (Table 2). Intersection of all these lists permitted us to prioritise nine candidate genes which appeared on more than one of these lists and/or for which independent and accessible validation models were available (i.e. Chd6 and Alad). Because neither the Chd6 (Figure 4) nor Alad genes were found to affect proportion of type I fibres in the way predicted by the QTL analyses, the list of prioritised candidates was reduced to 7 genes annotated in Supplementary Table 5. Three out of four QTLs contain one (Mfq6) or more candidate genes. All candidates are abundantly transcribed in muscle lineage with Psme4, Acyp2 and Mafb showing the highest level of expression in skeletal muscle compared to other tissues and cells. None of the seven candidates have been previously implicated to affect properties of skeletal muscle fibres although some of them have been implicated in cardiomyopathy or function as transcription factors (Supplementary Table 5).

Thus, genomic and gene expression analyses permitted focusing on a limited number of positional candidates in the future validation studies for establishing the causative genes.

**Conclusion**

In conclusion, we have refined the genetic architecture affecting cross sectional area of soleus muscle fibers and proportion of type I fibers in the LG/J and SM/J derived lineage. Integrating QTL mapping, genomic and transcriptome data from homologous muscle highlighted several candidate genes that may underpin muscle phenotypes critical to health and disease and worthy of follow up analyses.

Grants
The study was supported by NIH R01AR056280 grant (AL), NIH R01DA024845 (AAP) and Marie Curie International Reintegration Grant 249156 (AL).

Acknowledgements

The authors wish to acknowledge technical assistance from Diane Stewart, Jasmine C. Berry and Jake T. Brown. We acknowledge the Dartmouth Mouse Modelling Shared Resource of the Norris Cotton Cancer Center (support grant P30CA023108-27), for providing Chd6 knockout mouse samples.

Author contributions

A.L. conceived and supervised the study, A.M.C. phenotyped muscle samples, A.A.P. provided genotypes and oversaw the QTL analyses, R.C. designed the QTLRel software used in the QTL analysis, A.M.C. carried out QTL mapping, M.E.S, C.M. and E.C.D. did transcriptome analysis, S.N.F. and J.L.F. generated and provided Chd6 knockout samples, A.L. and A.M.C. wrote the manuscript with input from all co-authors.
Figure legends

Figure 1. Individual variability in proportion of oxidative fibers. Representative images of F_{34} female soleus cross-sections following myosin ATPase staining (acid pre-incubation). Dark fibers type I, pale fibers type IIA.

Figure 2. Type I fiber cross-sectional area QTL on chromosome 11. Analyses were carried out in the F_{2} intercross and in the combined F_{2} and F_{34} populations. X-axis indicates the relative position in the linkage map in centimorgan (cM). The thresholds are at the level of 0.05 genome wise significance for the F_{2} output (dotted line) and combined output (solid line).

Figure 3. Sex specificity of Mfq3 locus on cross-sectional area (CSA) of soleus type I and IIA fibers in the F_{34} intercross. Mean and SEM. Genotype at the peak marker: LG, homozygous for LG/J allele; H, heterozygous; SM, homozygous for SM/J allele.

Figure 4. Percentage of type I fibers in the soleus muscle of 4 month old Chd6 knockout (KO), heterozygous (HET) and wild-type (WT) females. There is no difference in percentage of type I muscle fibers in the soleus muscle between knockout, heterozygotes and wild-type groups (P=0.3041). Each data point is from a single mouse, horizontal lines represent group mean.
Table 1. Characteristics of muscle fiber QTL in combined analyses of the F2 and F34 intercrosses derived from the LG/J and SM/J strains.

<table>
<thead>
<tr>
<th>Chr</th>
<th>Thr**</th>
<th>Level***</th>
<th>Start Mb†</th>
<th>End Mb</th>
<th>Size Mb</th>
<th>Trait</th>
<th>Locus¥</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C</td>
<td>0.1</td>
<td>67.6</td>
<td>70.7</td>
<td>3.1</td>
<td>% Type I</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>C</td>
<td>0.1</td>
<td>193.9</td>
<td>194.3</td>
<td>0.4</td>
<td>% Type I &amp; CSA2A</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>0.1</td>
<td>92.4</td>
<td>104.8</td>
<td>12.4</td>
<td>% Type I</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>0.05</td>
<td>139.6</td>
<td>145.6</td>
<td>6.0</td>
<td>% Type I</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>G</td>
<td>0.01</td>
<td>158.8</td>
<td>162.5</td>
<td>3.7</td>
<td>% Type I</td>
<td><em>Mfq4 (SM)</em></td>
</tr>
<tr>
<td>3</td>
<td>G</td>
<td>0.05</td>
<td>33.6</td>
<td>40.0</td>
<td>6.4</td>
<td>CSA1 &amp; CSA2A</td>
<td><strong>Mfq5 (SM)</strong></td>
</tr>
<tr>
<td>4</td>
<td>G</td>
<td>0.05</td>
<td>57.7</td>
<td>62.7</td>
<td>5.0</td>
<td>% Type I</td>
<td><strong>Mfq6 (LG)</strong></td>
</tr>
<tr>
<td>4</td>
<td>C</td>
<td>0.05</td>
<td>103.9</td>
<td>106.1</td>
<td>2.2</td>
<td>% Type I</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>C</td>
<td>0.05</td>
<td>81.9</td>
<td>84.1</td>
<td>2.2</td>
<td>CSAAIA</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>C</td>
<td>0.01</td>
<td>110.8</td>
<td>116.0</td>
<td>5.2</td>
<td>CSA1 &amp; CSA2A</td>
<td><em><em>Mfq2</em> (LG)</em>*</td>
</tr>
<tr>
<td>7</td>
<td>C</td>
<td>0.05</td>
<td>138.4</td>
<td>140.0</td>
<td>1.6</td>
<td>% Type I</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>C</td>
<td>0.1</td>
<td>7.4</td>
<td>12.4</td>
<td>5.0</td>
<td>% Type I</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>C</td>
<td>0.05</td>
<td>89.0</td>
<td>92.4</td>
<td>3.4</td>
<td>TOTAL</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>C</td>
<td>0.01</td>
<td>121.9</td>
<td>128.6</td>
<td>6.7</td>
<td>TOTAL</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>G</td>
<td>0.1</td>
<td>120.7</td>
<td>121.3</td>
<td>0.6</td>
<td>% Type I</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>C</td>
<td>0.1</td>
<td>12.4</td>
<td>17.2</td>
<td>4.8</td>
<td>CSAAIA</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>C</td>
<td>0.1</td>
<td>19.1</td>
<td>23.1</td>
<td>4.0</td>
<td>CSAAIA</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>G</td>
<td>0.01</td>
<td>28.0</td>
<td>31.5</td>
<td>3.5</td>
<td>CSA1 &amp; CSA2A</td>
<td><em><em>Mfq3</em> (LG)</em>*</td>
</tr>
<tr>
<td>11</td>
<td>C</td>
<td>0.1</td>
<td>62.5</td>
<td>64.2</td>
<td>1.7</td>
<td>% Type I</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>C</td>
<td>0.1</td>
<td>70.6</td>
<td>76.2</td>
<td>5.6</td>
<td>% Type I</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>C</td>
<td>0.1</td>
<td>27.6</td>
<td>29.3</td>
<td>1.7</td>
<td>CSA1</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>C</td>
<td>0.01</td>
<td>5.3</td>
<td>9.9</td>
<td>4.6</td>
<td>% Type I</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>C</td>
<td>0.05</td>
<td>71.5</td>
<td>74.0</td>
<td>2.5</td>
<td>CSAAIA</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>C</td>
<td>0.05</td>
<td>93.6</td>
<td>102.3</td>
<td>8.7</td>
<td>CSAAIA</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>C</td>
<td>0.1</td>
<td>12.1</td>
<td>20.3</td>
<td>8.2</td>
<td>TOTAL</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>C</td>
<td>0.05</td>
<td>68.9</td>
<td>75.1</td>
<td>6.2</td>
<td>CSA1 &amp; CSA2A</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>C</td>
<td>0.01</td>
<td>11.8</td>
<td>52.5</td>
<td>40.7</td>
<td>TOTAL</td>
<td></td>
</tr>
</tbody>
</table>

* refined previously identified QTL in the LG/J and SM/J F2 intercross (47).

** C – chromosome-wide threshold, G- genome-wide threshold

*** Level of significance

¥ LG –increasing allele is LG/J, SM- increasing allele is SM/J

† Genomic positions based on GRCm38.p3.
**Table 2.** Positional candidate genes differentially expressed between LG/J and SM/J soleus muscles.

<table>
<thead>
<tr>
<th>Chr</th>
<th>QTL</th>
<th>Probe set ID</th>
<th>Gene</th>
<th>p-value*</th>
<th>Fold-Change**</th>
<th>Gene name***</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Mfq4</td>
<td>17393868</td>
<td><em>Mafb</em></td>
<td>0.033</td>
<td>-1.77</td>
<td>v-maf musculoaponeurotic fibrosarcoma oncogene family, protein B (avian)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17393910</td>
<td><em>Chd6</em></td>
<td>0.042724</td>
<td>-1.13</td>
<td>chromodomain helicase DNA binding protein 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17404652</td>
<td><em>Gm24780</em></td>
<td>0.032914</td>
<td>-1.98</td>
<td>Predicted gene Gm24780, predicted protein is B4HDV3.</td>
</tr>
<tr>
<td>3</td>
<td>Mfq5</td>
<td>17396801</td>
<td><em>Ttc14</em></td>
<td>0.033387</td>
<td>-1.14</td>
<td>tetratricopeptide repeat domain 14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17396876</td>
<td>----</td>
<td>0.0219376</td>
<td>-1.80</td>
<td>There are no assigned mRNA sequences for this probe set. The probe set lies within IncRNA Sox2ot (Sox2 overlapping transcript, non-protein coding)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17425606</td>
<td><em>Gm12526</em></td>
<td>0.046791</td>
<td>-1.17</td>
<td>predicted gene 12526</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17414380</td>
<td><em>Gm24277</em></td>
<td>0.0052644</td>
<td>-2.07</td>
<td>Gm24277 a known snRNA. The probeset also lies within an intronic region of RefSeq gene Pakap (PALM2-AKAP2), a read through transcript on chromosome 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17425701</td>
<td><em>Mir3095</em></td>
<td>0.00298235</td>
<td>-1.78</td>
<td>Mir3095 (Entrez ID 100526502; EST ENSMUST00000175552).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17426097</td>
<td><em>Mup3</em></td>
<td>0.00304</td>
<td>-1.49</td>
<td>major urinary protein 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17426126</td>
<td><em>Fkbp15</em></td>
<td>0.018038</td>
<td>-1.10</td>
<td>FK506 binding protein 15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17414545</td>
<td><em>Slc31a1</em></td>
<td>0.049799</td>
<td>1.12</td>
<td>solute carrier family 31, member 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17426166</td>
<td><em>Cdc26</em></td>
<td>0.021407</td>
<td>-1.29</td>
<td>cell division cycle 26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17426198</td>
<td><em>Hdh3</em></td>
<td>0.000869</td>
<td>1.86</td>
<td>haloacid dehalogenase-like hydrolase domain containing 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17426206</td>
<td><em>Alad</em></td>
<td>9.75E-05</td>
<td>1.91</td>
<td>aminolevulinate, delta-, dehydratase</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17414600</td>
<td><em>Rgs3</em></td>
<td>0.023096</td>
<td>1.08</td>
<td>regulator of G-protein signaling 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17248064</td>
<td><em>Mitf2</em></td>
<td>0.015517</td>
<td>-1.16</td>
<td>mitochondrial translational initiation factor 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17261285</td>
<td><em>LOC102637613</em></td>
<td>0.00432682</td>
<td>1.68</td>
<td>linc RNA [AK084560 (EST)/ Gm12092 (predicted gene)].</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17248127</td>
<td>----</td>
<td>0.00187422</td>
<td>-1.46</td>
<td>There are no assigned mRNA sequences for this transcript. The probe set lies within an intron of Sptbn1.</td>
</tr>
<tr>
<td>11</td>
<td>Mfq3</td>
<td>17261393</td>
<td><em>Acyp2</em></td>
<td>0.011777</td>
<td>1.26</td>
<td>acylphosphatase 2, muscle type</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17248196</td>
<td><em>Asb3</em></td>
<td>0.014221</td>
<td>-1.15</td>
<td>ankyrin repeat and SOCS box-containing 3</td>
</tr>
</tbody>
</table>
* ANOVA p-value for strain effect; ** Fold change uses SM/J as baseline (negative values indicate LG/J expression is down compared to SM/J, positive values LG/J expression up compared to SM/J); bold indicates that gene is predominantly and/or strongly expressed in skeletal muscle tissue (79). *** For probe sets not designed against an annotated gene, genes at the genomic loci of the Affymetrix probeset were identified in UCSC genome browser using mouse genome build GRCm38.


