Variable cardiac α -actin (*Actc1*) expression in early adult skeletal muscle correlates with promoter methylation

Jordan K. Boutilier^{1,2#}, Rhonda L. Taylor^{1,2#}, Ramesh Ram^{1,2}, Elyshia McNamara^{1,2}, Quang Nguyen^{1,2}, Hayley Goullee^{1,2}, David Chandler³, Munish Mehta^{1,2}, Lois Balmer^{1,2}, Nigel G. Laing^{1,2}, Grant Morahan^{1,2}, Kristen J. Nowak^{1,4}

¹Harry Perkins Institute of Medical Research, QEII Medical Centre, Nedlands, Western Australia

²Centre for Medical Research, University of Western Australia, QEII Medical Centre, Nedlands, Western Australia

³Australian Genome Research Facility, Harry Perkins Institute of Medical Research, QEII Medical Centre, Nedlands, Western Australia

⁴School of Biomedical Sciences, Faculty of Health and Medical Sciences, University of Western Australia, QEII Medical Centre, Nedlands, Western Australia

These authors contributed equally to this work.

Key words: Gene expression, *Actc1*, skeletal muscle, expression quantitative trait loci (eQTL), the Collaborative Cross (CC), systems genetics.

Abbreviations: CC, the Collaborative Cross; EDL, extensor digitorum longus; TA, tibialis anterior; SNP, single nucleotide polymorphism; LOD, log₁₀ odds; ChIP, chromatin immunoprecipitation; ChART, chromatin accessibility by real time PCR; UTR, untranslated region; Chr, chromosome, AU_{log2}, arbitrary units log₂; NOD, NOD/LtJ; B6, C57BL/6J; PWK, PWK/PhJ; TSS, transcriptional start site.

Abstract

Different genes encode the α -actin isoforms that are predominantly expressed in heart and muscle. Mutations in the skeletal muscle α -actin gene (*ACTA1*) cause muscle diseases that are mostly lethal in the early postnatal period. We previously demonstrated that the disease phenotype of *ACTA1* mouse models could be rescued by transgenic over-expression of cardiac α -actin (*ACTC1*). *ACTC1* is the predominant striated α -actin isoform in the heart but is also expressed in developing skeletal muscle. To develop a translatable therapy, we investigated the genetic regulation of *Actc1* expression. Using strains from The Collaborative Cross (CC) genetic resource, we found that *Actc1* varies in expression by up to 24-fold in skeletal muscle. We defined significant expression quantitative trait loci (eQTL) associated with early adult *Actc1* expression in soleus and heart. eQTL in both heart and soleus mapped to the *Actc1* locus and replicate an eQTL mapped for *Actc1* in BXD heart and quadriceps. We built on this previous work by analysing genes within the eQTL peak regions to prioritise likely candidates for modifying *Actc1* expression. Additionally we interrogated the CC founder haplotype contributions to enable prioritisation of genetic variants for functional analyses. Methylation around the *Actc1* transcriptional start site in early adult skeletal muscle negatively correlated with *Actc1* expression in a strain-dependent manner, while other marks of regulatory potential (histone modification and chromatin accessibility) were unaltered. This study provides novel insights into the complex genetic regulation of *Actc1* expression in early adult skeletal muscles.

Author Summary

Mutations in the skeletal muscle actin gene (*ACTA1*) cause a severe muscle disease that is usually fatal within the first year of life. Some patients with mutations in the *ACTA1* gene do not have any *ACTA1* protein present in their skeletal muscles. We have previously shown that some of these patients retain expression of the foetal isoform of the gene, cardiac actin (*ACTC1*). Further, patients who have *ACTC1* expressed in their skeletal muscles have increased longevity and improved muscle function, suggesting that upregulation of *ACTC1* may be a viable therapy for *ACTA1* disease patients. Thus, we have investigated the mechanisms regulating expression of *Actc1* in a genetically diverse population of different mouse strains with the aim of identifying the regulatory controls that cause the gene to be

switched off. We identified a number of sequence variants present in *Actc1* regulatory regions and show that increased promoter methylation is inversely correlated with *Actc1* expression. We also show that expression of *ACTC1* in humans is likely to be influenced by non-coding sequence variation in regulatory regions upstream and downstream of the gene. Our findings build on previous work and identify regulatory regions of the genome that can now be investigated further.

Introduction

Actc1 is the most abundant striated α -actin isoform in the mature heart and the predominant striated α -actin in foetal skeletal muscle.^{1,2} After birth, *Actc1* protein expression in skeletal muscle is down-regulated to negligible levels, but may be expressed in diseased mature skeletal muscle (such as in muscular dystrophy) in regenerating myofibres.^{3,4} Modulation of *Actc1* expression could potentially have a therapeutic benefit for certain striated muscle diseases. For example, reduced levels of *ACTC1* protein are detected in failing heart.⁵

ACTC1 also represents a therapeutic target for disease caused by skeletal muscle α -actin (*ACTA1*) mutations. *ACTA1* mutations usually produce a severe skeletal muscle phenotype with reduced movement at birth, and death within the first year of life.⁶ There is currently no cure for *ACTA1* disease and present treatments such as mechanically assisted ventilation and feeding, only manage symptoms (reviewed in⁷). We previously showed that patients with recessive *ACTA1* mutations had a complete absence of *ACTA1* protein but retained expression of *ACTC1* in their skeletal muscles into adolescence. These patients had a less severe phenotype leading us to propose that upregulation of *ACTC1* could be a viable therapy for *ACTA1* patients.⁸ In mouse models of *ACTA1* disease, we showed that transgenic over-expression of *ACTC1* in postnatal skeletal muscle could effectively compensate for absent⁹ or mutant *Acta1*.¹⁰ Therefore, identifying genetic elements controlling endogenous *Actc1* expression would be important for possible future treatments for either cardiac or skeletal muscle diseases.

In order to inform future treatments for human skeletal muscle α -actin (*ACTA1*) disease, we investigated expression levels of cardiac α -actin (*Actc1*) in skeletal muscles from young adult

mice. To do so, we utilised the Collaborative Cross (CC), a powerful mouse genetics resource¹¹ we co-developed, to map expression quantitative trait loci (eQTL) for *Actc1*.

Mapping an eQTL involves correlating variation in gene expression levels with genetic differences, usually single nucleotide polymorphisms (SNPs).¹² Genes or regulatory regions causative for the eQTL can then possibly be manipulated for modulation of gene expression, including for therapeutic purposes. eQTL are most conveniently determined by genetic mapping in a reference population of recombinant inbred mice. One of the most widely used mammalian recombinant inbred panels is the BXD resource, derived from the parental strains C57BL/6J and DBA/2J (e.g. ^{13–19}).

Conventional recombinant inbred lines derived from two parental strains lack genetic diversity because they involve only two closely-related founder genomes.²⁰ This level of genetic diversity effectively produces blind spots that can hamper genetic mapping.²¹ Recently, Wang and colleagues argued that one way to overcome this potentially problematic linkage disequilibrium is to validate results in human cohorts and alternative mouse resources such as the CC.²²

The CC was created from eight genetically diverse founder strains (A/J, C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, NZO/HiLtJ, CAST/EiJ, PWK/PhJ and WSB/EiJ).²¹ Collectively, these founder strains capture over 90% of common genetic variation in the mouse species.²⁰ The genetic map for CC strains is denser than previous recombinant inbred panels, with strains typed at 143,259 SNP markers using the GigaMUGA array (GeneSeek). The increased genetic diversity, coupled with the high-density SNP profiles for each of the CC strains, makes this a powerful resource for mapping complex traits.²³

Wang *et al.* recently used the BXD resource to map a significant eQTL for *Actc1* in heart, where *Actc1* is highly expressed.²² In this study, we have built on the work of Wang and colleagues by confirming the *Actc1* eQTL mapped in the heart with an alternative mouse resource, the CC. We also demonstrate highly variable levels of *Actc1* expression between different CC mouse strains at six weeks of age both in a fast myofibre- and slow myofibre-predominant skeletal muscle, and provide evidence of the expression control mechanisms.

Materials and Methods

Mice

All mouse breeding and experimental procedures were approved by the Animal Ethics Committees at the Animal Resources Centre and The University of Western Australia. Mice were housed under specific pathogen-free conditions and were provided with tap water and standard rodent food *ad libitum*. All cages were kept under the same conditions, maintained at 22°C constant temperature, with a 15:9-hr light/dark cycle. CC strains were generously provided by Geniad Pty Ltd.²³

RNA preparation

For microarray analysis, extensor digitorum longus (EDL), soleus and heart muscles were collected from 6-week old male mice (n≥3 per strain), from 48, 48 and 55 strains respectively. The soleus and EDL muscles were selected as they are commonly used to represent skeletal muscle with a significant proportion of slow twitch (type I) myofibres (soleus), or fast twitch (type II) myofibres (EDL).²⁴ Tissues were dissected immediately post euthanasia, snap frozen in liquid nitrogen, and stored at -80°C until processing. RNA was extracted using an RNAeasy Fibrous Tissue kit (Qiagen). RNA quality was determined using an Agilent Bioanalyzer with an RNA integrity number of 7 used as the minimum cut-off value.²⁵ Individual RNA samples from the same tissue and strain were evenly pooled based on amount of RNA.

For qPCR analysis, soleus, EDL, tibialis anterior (TA), quadriceps and heart samples were dissected from 6 and 30 week old C57BL/6J and NOD/ShiLtJ male mice (hereafter referred to as B6 and NOD respectively). RNA and protein were extracted simultaneously using TRIzol reagent (Thermo Fisher Scientific) and RNA integrity was confirmed by gel electrophoresis.

Microarray hybridizations and data analysis

Whole genome expression profiling of 25,698 transcripts was performed using the MouseRef-8 v2 BeadChip (Illumina, USA). The BeadChip was scanned with a Bead Array Reader and raw data were exported using Illumina Genome Studio software. All microarray expression data were log2-transformed and normalised with Robust Spline Normalization (RSN), using LUMI package in R. EDL samples were processed in two batches, hence a non-parametric batch

effects removal was performed using COMBAT R module prior to combining all samples. The results for two *Actc1* probes, ILMN_2598916 and ILMN_2767216, were extracted. Probe sequences were interrogated using the Wellcome Trust Sanger Institute Mouse Genomes Project (<http://www.sanger.ac.uk/>) to ensure complementarity to all founder genomes. The ILMN_2767216 probe did not overlap any genomic variants present in founder strains. However, the ILMN_2598916 probe overlapped with three SNPs (rs13469211, rs13469210 and rs28306385) found in Cast/EiJ and NOD, Cast/EiJ and NOD, and Cast/EiJ and PWK/PhJ founder genomes respectively. Thus, only the ILMN_2767216 probe was used for analyses. All data have been made publically available at (<http://130.95.9.22/webqtl.html>). Data can be accessed by selecting 'GeneMine' under the 'Group' heading and then selecting either 'Soleus, 6wk, male, Illumina', 'EDL, 6 wk, male, Illumina', or 'Heart, 6 wk, male, Illumina' under the 'Database' heading.

Quantitative PCR

The SuperScript III First-Strand Synthesis System (Thermo Fisher Scientific) was used to synthesise cDNA from total RNA using random hexamers according to the manufacturer's protocol. The Rotor-Gene SYBR Green PCR Kit (Qiagen) was used to compose 10 µL reactions containing 1 µL cDNA and 0.8 µM each of forward and reverse primers specific to the transcripts of interest (*Actc1*, *Tbp*, *Actb*, *Eef2*; Supplementary Table 1). Thermal cycling was performed on the Rotor-Gene Q real-time PCR cyclers and data analysis was performed with the Rotor Gene Q series software (Qiagen). *Actc1* transcript abundance was normalised to the geometric mean of two endogenous control genes (*Tbp* and *Eef2*) using the delta Ct method.²⁶ To satisfy assumptions of normality, data were log transformed prior to statistical analysis.²⁶

Actc1 protein quantitation

Quantification of *Actc1* protein was performed on EDL (n=3) and soleus (n=2) muscles from 6-week old male B6, HAX2_EF (a CC strain; henceforth HAX2) and NOD mice. Multiple reaction monitoring (MRM)-mass spectrometry on EDL muscle was performed by Proteomics International as per previous methodology.²⁷ Western blot analysis was conducted using a specific *Actc1* mouse monoclonal antibody (clone Ac1-20.4.2, Sigma-Aldrich)²⁸ with *Gapdh* (clone GAPDH-71.1, Sigma Aldrich) used as an internal standard for loading. Immunostaining of frozen skeletal muscle sections also utilised the specific *Actc1* mouse monoclonal antibody,

187 along with a Zenon® Alexa Fluor® 488 mouse IgG₁ labelling kit (Thermo Fisher Scientific).²⁸

188 Muscle spindle fibres are specialised clusters of striated fibres surrounded by a connective
189 tissue capsule²⁹ which stain positively for Actc1⁹ and were identified using serial sections
190 stained by haematoxylin and eosin.

192 CC QTL mapping

193 CC strains were genotyped previously as described³⁰ using the GigaMUGA (GeneSeek) mouse
194 genotyping array from the University of North Carolina.³¹ Of the 143,259 genotyped SNPs,
195 120,521 were reliably homozygous among the eight founders, and with at least one founder
196 having the non-reference allele. The remaining SNPs were either homozygous of the
197 reference allele, heterozygous or not called in at least one of the eight founders suggesting
198 probe issues. The probe intensities of the 120,521 reliable SNPs were extracted and founder
199 haplotype probabilities were estimated using Hidden Markov Model (HMM) algorithm used
200 in DOQTL. QTL mapping was performed using online tools (www.sysgen.org/GeneMiner)
201 which interface inferred haplotype probabilities of CC strains and a mixed model,
202 implemented in the R package QTLRel.¹² The mixed model uses a kinship matrix to correct for
203 cryptic relatedness as a random effect. The genomic intervals were reduced from 120,521 to
204 7,802; the haplotype probabilities of which were calculated by taking a median of the
205 haplotype probabilities of the SNPs that lie within the boundaries of each interval. Regression
206 coefficients for the additive genetic effects of each one of eight founder haplotypes were
207 calculated at all genomic loci assessed. The test for association was a likelihood ratio test
208 (LRT).

210 We used the WebQTL mapping and trait correlation modules in the GeneNetwork
211 (<http://130.95.9.22/webqtl.html>).³² Prior to entering into the mouse gene expression
212 databases of GeneNetwork, all RSN normalised data were additionally transformed to
213 produce Z scores that have a mean of 8 and a standard deviation of 2.³³ The advantage of this
214 step is that one unit difference in the transformed data is equivalent to a 2-fold difference in
215 expression level.³³

217 The log₁₀ odds (LOD) plots were generated for the *Actc1* probe and two-LOD drop support
218 intervals for each QTL were calculated. Due to the non-uniformity of the SNP distribution, +/-

10 kb was added to the boundaries of the QTL interval. Significance thresholds of LOD scores were determined using 1,000 permutations of the expression values. Significant eQTL were selected at a genome-wide permutation P value of ≤ 0.05 . Two further thresholds were estimated for permutation P < 0.10 (approaching significance) and P < 0.63 (suggestive). Candidate genes were identified using genome sequences from the Wellcome Trust Sanger Institute Mouse Genomes Project. A *cis*-eQTL was defined as a QTL for which the maximum eQTL was co-localized within 1 Mb of the *Actc1* transcript.

Analysis of BXD data

The gene network online program (<http://www.genenetwork.org/webqtl/main.py>) was used to interrogate quadriceps data generated from 37 - 42 BXD strains (EPFL/LISP BXD CD Muscle Affy Mouse Gene 1.0 ST (Nov12) and EPFL/LISP BXD HFD Muscle Affy Mouse Gene 1.0 ST (Dec11)).^{34,35} Full experimental details can be found in the gene expression omnibus (GN Accession: GN379, GEO Series: GSE60151). Briefly, all mice were fed *ad libitum* chow diet (CD; Harlan 2018) until 8 weeks of age, and then one cohort of mice (n= 37) was switched to a high fat diet (HFD; Harlan 06414) until sacrifice at 29 weeks. All animals were fasted overnight (6pm to 9am) before sacrifice. Quadriceps muscle was harvested from 5 individuals per cohort and pooled. Gene expression analysis was performed with the Affy Mouse Gene 1.0 ST (GPL6246) array. Genome wide interval mapping was performed with 1000 permutations.

Methylation analysis by bisulphite pyrosequencing

DNA was extracted using the DNeasy blood and tissue kit (Qiagen) from soleus muscle of 6-week old male B6, HAX2 and NOD mice (n = 3 per strain). A pyrosequencing assay was designed using CHECK primers (Supplementary Table 1) and performed by the Australian Genomics Research Facility (AGRF). To quantitate CpG DNA methylation levels around the transcriptional start site primers were designed to amplify the *Actc1* promoter and 5'UTR and sequence this region in four fragments (Supplementary Table 1).

PCR analysis of the *Actc1* promoter duplication

Determination of the presence of a previously reported duplication in the 5' promoter region of *Actc1*³⁶ was performed using genomic DNA from CC strains. PCR was performed with three

primers, designed to amplify the genomic DNA of mouse chromosome 2 harbouring the duplicated or non-duplicated *Actc1* allele (Supplementary Table 1).

Functional interrogation of *Actc1* regulatory regions

A chromatin immunoprecipitation assay (ChIP) and a chromatin accessibility assay (ChART) were performed on quadriceps and TA muscle respectively (these muscles were chosen due to their size facilitating these protocols) from 6-week old B6 and NOD mice (Supplementary methods). End-point qPCR analysis was performed with primers specific for *Actc1* regulatory regions (Supplementary Table 1). Human *ACTC1* eQTL data was obtained from the IGV eQTL browser within the Genotype-Tissue Expression project (GTEx) portal (<https://www.gtexportal.org/home/>).

Statistical analyses

A log-likelihood (LR) association test was applied to test significance of *Actc1* expression of 'high' or 'low' conferring alleles at a given locus, versus all other strains. For methylation and mass spectrometry analyses, a one-way ANOVA with Tukey's post-test for multiple comparisons was performed. For differences in transcript abundance (qPCR) between strains, a Student's unpaired t-test was performed on log-transformed data. Significance was determined at * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Graphs and statistics were generated using GraphPad Prism (V6.02).

Results

***Actc1* is expressed in early adult skeletal muscle**

Microarray data were generated from soleus, EDL and heart muscles from ~50 CC strains (Figure 1). A probe complementary to the 3'untranslated region (UTR) of *Actc1* (ILMN_2767216) was used to measure *Actc1* expression levels. As expected, all CC strains had high levels of *Actc1* expression in heart (mean expression of 13.1 Log₂ arbitrary units (AU_{Log2})), with only a modest range (~4-fold) in expression across all CC strains (Figure 1A). Robust *Actc1* expression could also readily be detected in the skeletal muscle data sets, with a mean expression of 8.5 AU_{Log2} in soleus and 9.7 AU_{Log2} in EDL. The CC strains showed a large range of expression: ~24-fold in soleus and ~20-fold in EDL (Figure 1B, C).

Figure 1. *Actc1* transcript abundance varies by up to 24-fold in postnatal skeletal muscles between Collaborative Cross strains.

Microarray data from ~50 CC strains (x-axis) was normalized by transforming to the log2ratio of Cy5/Cy3 intensity such that 1 unit represented an approximate 2-fold difference in expression (y-axis). *Actc1* levels in heart (**A**), soleus (**B**) and EDL (**C**) were detected a probe complementary to the 3'UTR of *Actc1* (ILMN_Actc1_2 respectively). The ranked *Actc1* expression of NOD/ShiLtJ (white arrow), C57Bl/6J (grey arrow) and HAX2_EF (black arrow) are indicated as examples of strains with differential expression between tissues.

***Actc1* transcript and protein levels in skeletal muscles differed between early adult mice from different strains**

To validate the variation in expression of *Actc1* between strains and muscle tissues, we selected as examples of strains with differential *Actc1* expression between tissues the CC founders B6 and NOD, as well as the HAX2 strain. *Actc1* transcript abundance was measured in a range of skeletal muscles from 6-week old B6 and NOD mice (Figure 2A). The results were highly concordant with the microarray data from soleus, EDL and heart (Figure 2A). We also analysed *Actc1* expression in further hind limb skeletal muscles (quadriceps and TA) that contain myofibre type proportions similar to the EDL (e.g. a mix of fast and slow myofibres). In addition, we analysed a second time point (30 weeks) to determine if *Actc1* expression changes with age. At 6 weeks of age, *Actc1* expression was significantly different between B6 and NOD in EDL (11.7-fold), quadriceps (7.7-fold), and TA (5.6-fold), but not in soleus or heart (Figure 2A). At 30 weeks of age, expression of *Actc1* transcript in soleus, EDL, TA and quadriceps was decreased compared with six weeks but statistically significant different levels of expression between B6 and NOD were retained (Figure 2B). At 30 weeks of age *Actc1* expression between B6 and NOD was not significantly different in the heart (Figure 2B).

To determine if *Actc1* transcript production in skeletal muscles resulted in protein expression, we measured *Actc1* abundance in the soleus (n=2) and EDL (n=3) muscles of 6-week old NOD, B6 and HAX2 mice by western blot analysis (Figure 2C), multiple reaction monitoring-mass spectrometry (difference of 2.3-fold between strains, Figure 2D) and immunostaining (Figure 2E - J). The results confirmed that protein levels in both the soleus and EDL differed between strains, with HAX2 mice having the highest levels of expression in soleus while B6 mice had the highest levels of *Actc1* expression in the EDL. However, differences at the protein level were not as great as at the transcript level. Immunostaining localised *Actc1* expression to different cell types between strains (Figure 2 E – J). *Actc1* was found in a proportion of mature muscle fibres in HAX2 (soleus and EDL; Figure 2G and J respectively) and B6 (soleus and EDL; Figure 2F and I respectively) mice, as well as in muscle spindle fibres⁹ (Figure 2F, G, I, J), but only in muscle spindle fibres in NOD mice (Figure 2E, H).

Figure 2. *Actc1* protein abundance differs between C57Bl/6J, NOD/ShiLtJ and HAX2_EF in mature skeletal muscles.

Actc1 mRNA expression in 6 week (A) and 30 week (B) soleus, EDL, TA and quadriceps of NOD/ShiLtJ (NOD) and C57Bl/6J (B6) mice was determined by qPCR. *Actc1* protein abundance in EDL and soleus muscles from 6 week old NOD, B6 and HAX2_EF (HAX2) mice determined by Western blot using Gapdh as a loading control (C), multiple reaction monitoring-mass spectrometry (D) and immunohistochemistry on fresh-frozen sections (E – J). Spindle fibres are indicated with a white arrow head. Scale bars represent 50 μm. Significance of qPCR data was determined with Student's unpaired T-test, and for mass-spectrometry one-way ANOVA with Tukey's post-test for multiple comparisons was performed with significance set at **p* < 0.05, ***p* < 0.01.

Collectively, these results confirm previous reports^{1,2} that *Actc1* is downregulated to low levels in postnatal skeletal muscle. A published study on the expression of *Actc1* in murine postnatal skeletal muscle and heart showed by western blot that *Actc1* could be detected at low levels at postnatal day 7 but could not be detected by 6 months of age.¹ However, that report did not specify which strain or skeletal muscle type was used. Our data indicate that the developmental timing of *Actc1* downregulation is variable between different CC mouse strains, and in B6 mice robust protein expression can be detected in EDL muscle as late as 6 weeks after birth, a previously unpublished finding.

Significant *Actc1* eQTLs in CC heart and soleus muscles localise to chromosome 2

To identify eQTL for *Actc1* expression in heart, soleus and EDL of CC strains we performed genome-wide interval mapping using the web-based Gene Miner tool.³⁷ For the heart, a significant eQTL ($P < 0.05$) with a maximum LOD of 11.8 for *Actc1* expression at 116.302 Mbp was found on chromosome 2, with a -2 LOD drop encompassing Chr2:113.081 - 116.644 Mbp (Figure 3A, B). This peak overlaps the *Actc1* eQTL identified using the BXD resource.²² A second suggestive eQTL ($P < 0.63$) with a maximum LOD of 7.7 mapped to chromosome 9 with a -2 LOD drop spanning Chr9:31.837 - 35.264 Mbp (Figure 3A).

For soleus muscle, two eQTL with non-overlapping -2 LOD drop intervals were mapped to chromosome 2. One eQTL reached significance ($P < 0.05$) at a maximum LOD of 10 at Chr2:113.926 Mbp and produced a -2 LOD drop interval of Chr2:113.622-114.594 Mbp. The second peak had a suggestive LOD of 9.4 ($P < 0.63$) at Chr2:116.644 and a -2 LOD drop spanning Chr2:116.302-125.069 Mbp (Figure 3C, D). A third eQTL with a maximum LOD of 7.8 mapped to chromosome 5 (-2 LOD drop Chr5:149.354 - 151.148 Mbp) at suggestive significance ($P < 0.63$) (Figure 3C).

For EDL muscle, two eQTL were mapped at suggestive significance levels ($P < 0.63$) to chromosomes 9 (LOD 6.9, -2 LOD drop Chr9:74.424 - 83.225 Mbp) and 14 (LOD 6.6, -2 LOD drop Chr14:115.657 - 121.344 Mbp) (Figure 3E). It may be that multiple eQTL regulate the expression of *Actc1* in EDL muscle, and that our study was not sufficiently powered to detect multiple weak interactions at statistically significant levels. Thus, we did not further investigate *Actc1* regulatory mechanisms in EDL.

366

367

368 **Figure 3. Expression quantitative trait loci for *Actc1* expression in heart and soleus map to**
369 **the chromosome 2 region spanning 113 to 116 Mbp, containing the *Actc1* gene.**

370 Genome-wide interval mapping was performed with 48 -53 CC strains and significant (P
371 < 0.05 , red line) expression quantitative trait loci (eQTL) were identified. **A – D.** Significant
372 overlapping eQTL spanning 113 – 116 Mb on chromosome 2 were mapped for heart and
373 soleus. **E.** No significant eQTL were mapped for *Actc1* in EDL, although suggestive eQTL were
374 mapped to chromosomes 9 and 14 ($P < 0.63$). **F.** The narrow eQTL region mapped to
375 significance in soleus muscle contains 14 genes: 9 protein coding genes (black text), 2
376 validated non-coding RNA (green text), 1 predicted non-coding RNA (blue text), 1 microRNA
377 (purple text) and one pseudogene (grey text). One of the protein coding genes in this region
378 is *Actc1* itself (red arrow).

Collectively, these data indicate that *Actc1* expression in the heart and soleus from CC strains was regulated by loci residing on chromosome 2, spanning ~113 – 116 Mbp (Figure 3A - D). This narrow eQTL interval contains only 14 genes, including *Actc1* (Figure 3F, red arrow). Additional loci may also variably contribute to regulation of *Actc1* expression, but for the purposes of this study we focussed only on eQTL that reached statistical significance.

Significant *cis*-eQTL for *Actc1/ACTC1* are also identified in BXD and human skeletal muscle

In order to validate our results, we interrogated publicly available quadriceps data generated using 29-week old mice from up to 42 BXD strains.^{34,35} In that study, mice were fed either a chow diet (CD; n =42), or, from 8 weeks of age, a high fat diet (HFD; n = 37) until sacrifice. Gene expression was analysed from quadriceps RNA by microarray. We found that *Actc1* expression varied by 14- and 9.5-fold between BXD strains in the CD and HFD cohorts respectively (Figure 4A, D).

Genome-wide interval mapping in both populations revealed significant eQTL on chromosome 2 (Figure 4B, C, E, F). However the chromosome 2 eQTL for the CD and HFD cohorts did not overlap, suggesting that the *Actc1* eQTL can be modified by diet. The *Actc1* eQTL in the CD cohort produced a maximum LOD of 4.3 ($P < 0.05$) at Chr2: 134.800 Mbp and a -2 LOD drop spanning Chr2:121.5 - 136.5 Mbp (Figure 4C). While the *Actc1* eQTL in the HFD cohort produced a maximum LOD of 4.2 ($P < 0.05$) approximately 20 Mbp away, at Chr2: 112.275 Mbp and a -2 LOD drop spanning Chr2:106.5 – 114 Mbp (Figure 4F). The eQTL produced in the HFD cohort overlaps with the eQTL observed in the CC heart and soleus datasets (Figure 3A – D), the BXD heart dataset²² and contains *Actc1*.

In the CD cohort, the eQTL peak did not overlap *Actc1* and produced a large -2 LOD drop interval (121.5 – 137 Mbp) containing more than 200 genes (Supplementary Table 2). Future studies using a larger number of strains should be conducted to confirm this association and refine the genomic interval of this eQTL.

Figure 4. Expression quantitative trait loci for *Actc1* expression in the BXD quad map to multiple regions on chromosome 2 which are modified by diet.

Publicly available microarray data were generated from 29-week old BXD strains fed with either a chow diet (CD; n = 42) or high fat diet (HFD; n = 37). **A.** *Actc1* expression in quadriceps from CD-fed mice varied by 14-fold between BXD strains. Parental strains DBA2 (green arrow) and C57Bl/6J (grey arrow) are indicated. **B.** Genome wide interval mapping indicated that eQTL for *Actc1* mapped to a single significant ($P < 0.05$, LOD 4.3) peak at Chr2:134.8 Mbp. **C.** The CD eQTL on chromosome 2 did not encompass *Actc1* (purple arrow). **D.** *Actc1* expression in quadriceps from HFD-fed mice varied by 9.5-fold between BXD strains. **E.** HFD-fed mice had a significant ($P < 0.05$, LOD 4.2) eQTL for *Actc1* which mapped to Chr2: 112.275 Mbp. **F.** The HFD eQTL on chromosome 2 encompassed *Actc1* (purple arrow).

In order to determine if eQTL for *ACTC1* expression also localised to the *ACTC1* gene in human skeletal muscle, we accessed data from the GTEx portal that contains matched genotype and RNA expression data from 449 individuals across 53 tissues. A total of 56 SNPs in three tissues (skeletal muscle, tibial artery and tibial nerve) were significantly associated with *ACTC1* expression at a false discovery rate of <5% (Supplementary Table 3). In skeletal muscle, all *ACTC1*-associated SNPs were clustered between 50 – 100 kb downstream of the *ACTC1* gene (Figure 5A). Conversely, *ACTC1*-associated SNPs in tibial artery and nerve were clustered within 5 kb upstream of the *ACTC1* transcriptional start site (Figure 5A).

ACTC1-associated SNPs both upstream and downstream of the gene were closely aligned with markers of regulatory potential including transcription factor enrichment, DNase I hypersensitivity and acetylation of histone 3 lysine 27 (H3K27ac), a histone modification normally associated with active enhancers (Figure 5B). This indicates that regulatory region polymorphisms likely contribute to differential *ACTC1* expression between individual people and may also underlie the Chr2 eQTL peak in mice. Interestingly, GWAS associated SNPs linked to axial length and refractive error were also located in proximity to the putative enhancer downstream of *ACTC1* (Figure 5A, green boxes).

Taken together, results generated from BXD quadriceps and heart²² as well as in adult human skeletal muscle support the eQTL interval encompassing the *Actc1* gene mapped in the CC soleus and heart.

Figure 5. *ACTC1* eQTL in human skeletal muscle map to a putative enhancer located 70 kb downstream of the *ACTC1* gene

Human SNPs associated with *ACTC1* expression were identified using the GTEx portal. Significantly associated variants in skeletal muscle, tibial artery and tibial nerve (A) were localised to non-coding regions of the genome enriched for marks of regulatory potential (B) including H3K27 acetylation, transcription factor binding and DNase I hypersensitivity. GWAS variants linked to refractive error and axial length are also indicated (green boxes).

Differing founder haplotypes contribute to increased *Actc1* expression levels in heart and soleus

One of the features of the CC and the Gene Miner program is the ability to define the effects of each founder's haplotype on the trait being analysed. To compare the haplotype effect of the CC founder strains on *Actc1* expression levels, coefficient plots were generated for the *Actc1* eQTL for heart and soleus (Figure 6).

In the heart, CC strains with the NOD allele in the eQTL interval had significantly increased levels of *Actc1*, whilst those with the A/J haplotype had significantly lower expression (Figure 6A). In soleus muscle, the PWK/PhJ (hereafter abbreviated to PWK) allele conferred increased levels of *Actc1*, whilst there was no founder strain that significantly reduced *Actc1* levels (Figure 6B). Thus, differing founder haplotypes contribute to increased *Actc1* expression between soleus and heart muscle.

Duplication of the *Actc1* promoter does not explain the CC strain differences in *Actc1* expression in early adult skeletal muscle

It has been reported that a 9.5 kb tandem duplication event encompassing the *Actc1* promoter and the first three exons of the gene result in decreased expression of *Actc1* in the heart of BALB/c and DBA/2 mice.³⁶ Recently, analysis of *Actc1* expression in the heart of 36 BXD strains (including parental strains) confirmed the association between the *Actc1* promoter duplication and reduced *Actc1* expression in the heart.²²

To confirm the presence of this duplication in the CC strains, we performed PCR analysis using genomic DNA from a subset (n=19) of CC strains and founder lines. The duplication was present in the two CC strains that had the A/J haplotype at the *Actc1* locus (Supplementary Figure 1). These haplotype analyses also indicated that the A/J allele segregated with decreased *Actc1* expression in the heart (Figure 6A). Seven of the ten strains with the lowest levels of *Actc1* heart expression harboured the A/J allele. This supported previous reports that the duplication event contributes to decreased *Actc1* expression in the heart.^{22,36}

Figure 6. The NOD/ShiLtJ and PWK/PhJ alleles confer increased *Actc1* expression in heart and soleus respectively.

Coefficient plots for chromosome 2 from all eight founder strains were generated for heart and soleus and the association of founder haplotypes with *Actc1* expression was analysed at 114 Mbp. **A.** In the heart, the NOD/ShiLtJ (NOD) founder haplotype at the *Actc1* locus confers increased *Actc1* expression while the A/J haplotype is associated with reduced expression. **B.** For the soleus eQTL generated with the ILMN_Actc1_1 probe, the PWK/PhJ (PWK) haplotype confers increased *Actc1* expression but none of the founder haplotypes are significantly associated with reduced expression.

In contrast, the A/J allele did not have a significant effect on *Actc1* expression in the soleus (Figure 6B). Thus, the promoter duplication cannot solely explain the difference in *Actc1* expression between high and low expressing strains in **all** early adult skeletal muscle.

In the BXD quadriceps, the duplicated allele carried by the DBA/2 founder strain also segregated with decreased levels of *Actc1* expression. Conversely, the B6 allele was associated with increased expression of *Actc1* in the BXD quadriceps. **Yet the B6 allele** did not have a significant effect on *Actc1* expression in either the soleus or the heart. This could indicate that independent of the duplication, the effect of founder-strain specific variants was influenced by the fibre type proportions of the **specific** skeletal muscle. **Whilst** both the quadriceps and EDL contain a predominance of fast myofibres, **the soleus has a higher proportion of slow myofibres. Therefore, variants present in the different founder strains could have a different effect on *Actc1* expression in different fibre types. This is possibly due to the predominance of different transcription factor networks and signalling pathways driving expression of alternative gene networks (such as those for oxidative versus glycolytic pathways) in the different fibre types.**

Strains that did not contain the duplication and that had increased expression of *Actc1* in heart (NOD), soleus (HAX2 which contains the PWK allele at *Actc1*) and quadriceps (B6) were used for further investigation of the mechanisms regulating *Actc1* expression. NOD mice are also a useful reference strain for skeletal muscle since *Actc1* expression was significantly reduced in NOD mice compared to B6 and HAX2 in both soleus and EDL. **However this** reduced expression is not caused by the duplication (Figure 2).

Investigation of the genetic basis of the cis-eQTL

Prioritisation of candidate genes based on expression

In soleus muscle, the significant eQTL encompassed a narrow -2 LOD drop region (Chr2:113.081 - 114.594) of only 1.5 Mbp, which overlapped eQTL in the heart (CC and BXD) and quadriceps (BXD HFD only). This genomic interval contains only 9 protein coding genes, 2 validated ncRNA genes, 1 predicted ncRNA, 1 miRNA and 1 pseudogene (Table 1; Figure 3F). One of the protein-coding genes within this interval is *Actc1* itself (Figure 3F, red arrow). The

-2LOD drop interval in the heart extends this region to include an additional 2 protein coding genes, 2 validated non-coding RNA (ncRNA), one predicted ncRNA and 1 microRNA (Table 1). Of the genes in this *cis*-eQTL region, only 2 (*Actc1* and *C130080G10Rik*) have known or predicted functional roles in striated muscle and are expressed in mature skeletal muscle (Table 1).

Prioritisation of candidate genes based on SNP analysis

The Gene Miner program was used to prioritise variants likely to alter protein structure and/or function present in the founder strains associated with a significant change in *Actc1* expression. In the -2 LOD drop interval for the heart, only 5 variants present in the A/J strain (which confers reduced expression of *Actc1* in the heart) were associated with protein changes (such as altered splicing, synonymous and non-synonymous amino acid substitutions). These variants were located in four protein coding genes (*Meis2*, *Arhgap11a*, *Aqr* and *Spred1*) and one ncRNA (*C130080G10Rik*) (Supplementary Table 4).

In the soleus -2 LOD drop interval, 55 variants present in the PWK genome (which confers increased *Actc1* expression in soleus) were significantly associated with protein changes. These variants were present in 5 coding (*Fmn1*, *Scg5*, *Arhgap11a*, *Aqr*, *Zfp770*), 4 non-coding genes (*Gm13967*, *4930533B01Rik*, *A530058N18Rik*, *mir7002*) and one pseudogene (*Tmco5b*) (Supplementary Table 4). Of the genes containing variants with the potential to alter gene product structure or function, only *Aqr* (ubiquitous), *Zfp770* (ubiquitous) and *C130080G10Rik* are expressed in adult skeletal muscle (Table 1). Of these, *C130080G10Rik* is the only one to have a previously documented role in skeletal muscle.³⁸ Given that *cis*-eQTL are frequently driven by variants in the target gene promoter,³⁹ and in light of the fact that *ACTC1*-associated SNPs in human skeletal muscle were located in putative regulatory regions rather than in coding genes (Figure 5), we directed our attention to variants located upstream of *Actc1* and within *C130080G10Rik* (Chr2:113870000 – 113890000).

Table 1. Within the combined heart and soleus *Actc1* eQTL genomic interval, only *Actc1* and *C130080G10Rik* are expressed in postnatal skeletal muscle and have known or predicted roles in muscle function.

Gene name	Function (Genecards)	Expression	
		GTE _x (human)	Fantom5 (mouse)
<i>Tmco5b</i>	Pseudogene	Testis	Testis
<i>Fmn1</i>	Role in the formation of Adherens junction and the polymerization of linear actin cables	Ubiquitous	Widely expressed but not in heart or skeletal muscle
<i>4930533B01Rik</i>	Predicted non-coding RNA	N/A	N/A
<i>Grem1</i>	An antagonist of bone morphogenic protein that may play a role in regulating organogenesis, body patterning, and tissue differentiation	Transformed fibroblasts, oesophagus, intestine, stomach	Smooth muscle, low expression in embryonic forelimb, stomach, intestine
<i>Scg5</i>	Acts as a molecular chaperone for PCSK2/PC2, preventing its premature activation in the regulated secretory pathway	Brain, pituitary, pancreas	No expression in muscle, enriched in brain and pituitary
<i>Arhgap11a</i>	Signalling by GPCR and Rho GTPases	EBV-transformed lymphoblasts, colon, transformed fibroblasts, eosophagus, spinal cord, testis	Widely expressed including embryonic forelimb.
<i>Gjd2</i>	The channels formed by this protein allow cationic molecule exchange between human beta cells and may function in the regulation of insulin secretion	Heart, pancreas, ovary	Adrenal gland, cerebellum, eyeball, medulla oblongata, neurons, pituitary, visual cortex
<i>A530058N18Rik</i>	Unclassified non-coding RNA	N/A	Eyeball
<i>Actc1</i>	The alpha actins are found in muscle tissues and are a major constituent of the contractile apparatus	Heart, skeletal muscle	Neonatal forelimb, heart, eyeball, lung
<i>C130080G10Rik</i>	Unclassified non-coding RNA which overlaps <i>Actc1</i> and is associated with a muscle and adipose tissue enhancer ⁵¹	N/A	Neonatal heart, neonatal biceps muscle
<i>Aqr</i>	Intron-binding spliceosomal protein required to link pre-mRNA splicing and snoRNP (small nucleolar ribonucleoprotein) biogenesis	Ubiquitous	Ubiquitous
<i>Mir7002</i>	No functional information available	N/A	N/A
<i>Zfp770</i>	May be involved in transcriptional regulation	Ubiquitous	Ubiquitous
<i>Dph6</i>	Amidase that catalyses the last step of diphthamide biosynthesis using ammonium and ATP	Ubiquitous	Ubiquitous

We used the ENOCDE-CC regulatory element SNP finder embedded in GeneMiner to retrieve a list of 75 SNPs private to either the NOD or PWK genomes and located within regulatory regions defined by the FANTOM5 and ENCODE projects (Supplementary Table 5). We also obtained a list of 50 insertion/deletion mutations in this region from the Wellcome Trust Sanger Institute Mouse genome project. Of note, 16 variants were present in a region of H3K27ac enrichment between -4 and -6 kb upstream of *Actc1*, a previously defined^{40–42} enhancer region.

Increased methylation surrounding the Actc1 transcriptional start site correlates with decreased Actc1 expression

In order to determine whether variants located in regulatory regions upstream of *Actc1* might contribute to differences in transcriptional activity, we first analysed chromatin accessibility and histone modification in quadriceps and TA muscles from NOD and B6 mice. These muscles were chosen for their significantly different *Actc1* expression (Figure 2A, B). However, there were no differences in histone modification or chromatin accessibility between the NOD and B6 strains across the extended *Actc1* promoter and upstream enhancer regions (Supplementary Figure 2). This indicates that these marks of regulatory potential do not underlie the differential *Actc1* expression between these strains of mice.

We next analysed the CpG content of each regulatory region and found that only the proximal promoter, 5'UTR and first intron were enriched for CG dinucleotides. This region encompasses the *Actc1* transcriptional start site and contains 9 CG dinucleotides (Figure 7A). We performed bisulphite pyrosequencing of this region with soleus muscle from 6-week old NOD, B6 and HAX2 mice. In soleus muscle, CG dinucleotides 2, 4 and 7 – 9 were significantly differentially methylated in a manner inversely correlated with *Actc1* expression level (Figure 7B). CpG sites 2 and 4 are situated in close proximity to putative transcription factor binding sites^{43–45} while sites 7 – 9 are within the 5'UTR.

Figure 7. Methylation across the *Actc1* proximal promoter inversely correlates with *Actc1* expression.

A. Nine CG dinucleotides were identified in the *Actc1* proximal promoter and 5' untranslated region (5'UTR). **B.** Methylation of these CG dinucleotides was analysed by bisulphite pyrosequencing in soleus and EDL muscle from six week old NOD (white bars), B6 (grey bars) and HAX2_EF (black bars) mice and methylation at each site was expressed as a percentage of total nucleotides sequenced at that location. Graphed data represent mean $\pm$ SEM. Methylation between strains was analysed by one-way ANOVA with significance set at * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Discussion

Modulation of *Actc1* expression may provide therapeutic benefit for both cardiac and skeletal muscle diseases. For example, upregulation of *Actc1* in postnatal skeletal muscle has considerable treatment potential for *ACTA1* diseases, a heterogeneous group of myopathies most frequently resulting in a severe and early lethal phenotype.⁷ Therefore we wished to investigate the mechanisms controlling *Actc1* expression in skeletal muscles. To accomplish this, we set out to map eQTL for *Actc1* in early adult (6-week old) skeletal muscle using the CC mouse resource.

We interrogated *Actc1* transcript levels in soleus, EDL and cardiac tissues from 48 – 55 CC strains. One of our major findings is that in early adult mice, *Actc1* is highly expressed in skeletal muscles of certain CC strains, with a large variation in expression across strains. Indeed, in B6 mice *Actc1* is highly expressed in EDL, TA and quadriceps muscle at levels comparable to those in the B6 heart at 6 weeks of age. *Actc1* expression was evident in apparently mature skeletal muscle fibres, indicating that *Actc1* expression was not an artefact of regenerating or developing fibres. *Actc1* expression decreased at 30 weeks of age (although low levels of transcript were still apparent) indicating that the kinetics of down-regulation may be different between CC strains. By analysing differential *Actc1* levels across the CC strains at 6 weeks of age, an *Actc1* eQTL was mapped at significant ($P < 0.05$) thresholds to the same chromosome 2 region in heart and soleus. These results corroborate those of Wang *et al.* who reported a similar *Actc1* eQTL in the heart using the BXD RI model,²² and in turn are supported by BXD quadriceps data which also localise an *Actc1* eQTL to chromosome 2 when mice were fed a HFD. Thus, we confirm that *Actc1* is cis-regulated in the heart, and build upon this by showing that *Actc1* is also cis-regulated in the soleus. These data indicate that expression of *Actc1* might be regulated by different mechanisms in different fibre types (e.g. slow versus fast myofibres, discussed further below) and that this can be modulated by diet.

An important distinction between the CC and the BXD resources with regards to mapping eQTL for *Actc1*, is that one of the two founder strains (DBA/2) of the BXD contains a duplication of the *Actc1* promoter.³⁶ The previously reported tandem duplication of 9.5 kbp in the 5' region of *Actc1* contains part of the promoter and the first three exons of the gene and is thought to cause reduced *Actc1* transcript levels in the heart.³⁶ Only one of the eight

founder strains (A/J) of the CC contains this promoter duplication meaning that the effect of promoter variants on *Actc1* expression distinct from the effect of the duplication can be delineated. The duplication does influence *Actc1* levels in the mouse heart, indicated by the small but significant reduction in *Actc1* expression in the heart in CC strains harbouring the duplicated A/J allele. However, our data indicate that the 5' *Actc1* duplication is not the only mechanism leading to low expression in early adult skeletal muscle since many of the strains that have the lowest *Actc1* expression do not harbour the duplication.

We attempted to refine loci that might influence transcript abundance by analysing the effect of each founder haplotype on *Actc1* expression. In the soleus, the PWK founder allele conferred the highest levels of *Actc1* in CC strains, while in the heart the NOD allele was associated with increased *Actc1* expression (Figure 6). Abundant and variable expression of *Actc1* was also observed in the quadriceps of 29-week old BXD mice, where increased expression was associated with the B6 allele. However the B6 allele did not have a significant effect on *Actc1* expression in the CC heart or soleus.

Collectively, the haplotype data indicated that unique genetic variants underpin the modulation of *Actc1* expression for the striated muscle tissues tested. As each murine skeletal muscle is comprised of a combination of myofibre types, it is plausible that certain transcriptional regulatory networks predominate in skeletal muscles containing particular proportions of myofibre types. For example, there are multiple differences in expression of gene isoforms in slow versus fast myofibre types,⁴⁶ with some slow myofibre isoforms also expressed in the heart (e.g. beta slow/cardiac muscle myosin heavy chain 7, *Myh7*). Further, the expression of myofibre type specific genes can be modified by diet. HFD-fed mice have previously been shown to have increased levels of poly(ADP-Ribose) polymerases (*Parp1*) which is negatively correlated the expression of slow skeletal muscle troponin (*Tnni1*), and positively correlated with the fast skeletal muscle troponins (*Tnnc2* and *Tnni2*).³⁴ However, the direct mechanisms linking oxidative metabolism with altered expression of myofibre type specific genes remain to be elucidated.

Following on from the work of Wang and colleagues,²² we take the first steps towards exploring the possible control regions underlying the eQTL peak. The best candidates for

regulating *Actc1* expression were the *Actc1* promoter itself and an overlapping validated ncRNA (*C130080G10Rik*). Interestingly, *C130080G10Rik* is transcribed in a head-to-head orientation with *Actc1* itself and its expression pattern mirrors that of *Actc1*, suggesting co-regulation⁴⁷ (Figure 8). Further, the exons of *C130080G10Rik* are closely aligned with conserved and previously described enhancers that are associated with the expression of *Actc1*^{40–42} (Figure 8). These regions are also associated with DNase hypersensitivity and enrichment of activating histone marks in both heart and skeletal muscle (Figure 8), indicating significant regulatory potential. The role of this ncRNA has not been defined in detail. However, long ncRNAs have previously been implicated in the transcriptional regulation of neighbouring genes via long-range chromatin loops.^{48,49} *C130080G10Rik* is listed as a linked gene with a constitutively active muscle and adipose tissue enhancer located 40 kb upstream of *Actc1*.³⁸ A model can be envisioned whereby interaction between the ncRNA and the distal enhancer could lead to ‘leaky’ *Actc1* transcription that can be modulated by sequence variation surrounding *Actc1*. The location of human *ACTA1* eQTL 70 kb downstream of the gene also implies that chromatin looping may be an important mechanism regulating expression in human skeletal muscle.

Interestingly, some of the SNPs associated with *ACTC1* expression in human skeletal muscle are also GWAS candidate variants associated with refractive error.^{50,51} We have previously shown that *ACTC1* is expressed in extraocular muscles, which presumably explains the absence of ophthalmoplegia in *ACTA1* disease patients.²⁷ Thus, variants in the putative *ACTC1* enhancer region could lead to altered levels of *ACTC1* expression, contributing to weakness of the extraocular or pupillae muscles and ultimately refractive error.⁵¹

However, the position of long-range regulatory elements do not appear to be conserved between human and mice since there was no enrichment of enhancer-associated marks downstream of *Actc1* in murine skeletal muscle. Thus, while it is possible that the mechanisms of regulation might be conserved between humans and mice (due to conserved patterns of developmental and tissue-specific expression)^{52,53} the precise genomic position of regulatory elements are not conserved.

Figure 8. *Actc1* regulatory regions overlap a long non-coding RNA

Previously identified regulatory regions are closely aligned with features of a long-non-coding RNA transcribed in a head-to-head orientation with *Actc1* and marks of regulatory potential including DNase hypersensitivity, histone modification and enrichment or RNA polymerase II.

We analysed regulatory regions upstream of *Actc1* for differences in functional potential and the presence of sequence variation between mouse strains conferring increased and decreased *Actc1* expression (PWK, B6 and NOD strains). Which one or more of the PWK/B6 variants mediate increased transcript abundance cannot easily be determined since the effects of non-coding sequence variation are difficult to establish.⁵⁴ However, we showed that strain-dependent variation in *Actc1* promoter methylation was inversely correlated with *Actc1* expression (Figure 7). Therefore, we propose that in skeletal muscle, variation in *Actc1* expression is in part mediated by epigenetic modification at the *Actc1* start site and may be facilitated by long-distance interactions with a constitutively active skeletal muscle/adipose enhancer.³⁸ It is possible that tissue specific transcription factors in the heart, such as Gata4⁵⁵ and Nkx,⁵⁶ might drive the increased *Actc1* expression observed across strains and drive protection from methylation in this tissue.

In conclusion, we show that *Actc1* expression is highly variable in skeletal muscle of numerous CC and wild-derived mouse strains at 6 week of age. This is translated into robust protein expression in at least the B6 and HAX2 strains at this time point. We mapped a significant eQTL for *Actc1* expression in heart and soleus using the CC, which confirms and extends data generated using an alternative RI model. We also take the first steps towards interrogating the functional mechanisms underlying this eQTL peak and present a model that integrates strain-specific sequence variation with differences in transcript abundance and epigenetic modification. In future, further elucidation of the mechanisms that engender expression of *Actc1* in early adult muscle may be leveraged to drive therapeutic upregulation of ACTC1 in human ACTA1 disease. Additionally, this study validates the CC as a systems genetics resource, generating impressive trait diversity using a relatively small number of strains. Collectively, we have shown the utility of this resource for defining the expression and regulation of a clinically relevant gene.

Acknowledgements

This research was supported by a trampoline grant from the Association Française contre les Myopathies (17822, KN), an Australian Research Council Future Fellowship (FT100100734, KN), a National Health and Medical Research Council (NHMRC) Principal Research Fellowship (APP1002147, NL), an NHMRC Program Grant (1037321 GM), an NHMRC Project Grant

(1069173, GM and RR) an NHMRC EU Collaborative Grant (APP1055295, NL), and an Endeavour PhD scholarship (JB). We thank the staff at the Animal Resources Centre and Proteomics International.

We also acknowledge the Genotype-Tissue Expression (GTEx) Project which was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on 02/16/2017.

References

1. Ilkovski B, Clement S, Sewry C, North KN, Cooper ST. Defining alpha-skeletal and alpha-cardiac actin expression in human heart and skeletal muscle explains the absence of cardiac involvement in ACTA1 nemaline myopathy. *Neuromuscul Disord*. 2005; **15**: 829–35.
2. Vandekerckhove J, Bugaisky G, Buckingham M. Simultaneous expression of skeletal muscle and heart actin proteins in various striated muscle tissues and cells. A quantitative determination of the two actin isoforms. *J Biol Chem*. 1986; **261**: 1838–43.
3. Holzhausen RMH, Baumann R, Taege C, Franke WW. The cardiac isoform of α -actin in regenerating and atrophic skeletal muscle, myopathies and rhabdomyomatous tumors: an immunohistochemical study using monoclonal antibodies. *Virchows Arch*. 2006; **445**: 175–91.
4. Dennis RA, Przybyla B, Gurley C, Kortebein PM, Simpson P, Sullivan DH, et al. Aging alters gene expression of growth and remodeling factors in human skeletal muscle both at rest and in response to acute resistance exercise. *Physiol Genomics*. 2008; **32**: 393–400.
5. Dirx E, da Costa Martins PA, De Windt LJ. Regulation of fetal gene expression in heart failure. *Biochim Biophys Acta - Mol Basis Dis*. 2013; **1832**: 2414–24.
6. Goebel HH, Laing NG. Actinopathies and myosinopathies. *Brain Pathol*. 2009; **19**: 516–22.

- 762 7. Nowak KJ, Ravenscroft G, Laing NG. Skeletal muscle alpha-actin diseases
763 (actinopathies): Pathology and mechanisms. *Acta Neuropathol.* 2013; **125**: 19–32.
- 764 8. Nowak KJ, Sewry CA, Navarro C, Squier W, Reina C, Ricoy JR, et al. Nemaline
765 myopathy caused by absence of alpha-skeletal muscle actin. *Ann Neurol.* 2007; **61**:
766 175–84.
- 767 9. Nowak K, Ravenscroft G, Jackaman C, Filipovska A, Davies SM, Lim EM, et al. Rescue
768 of skeletal muscle α -actin-null mice by cardiac (fetal) α -actin. *J Cell Biol.* 2009; **185**:
769 903–15.
- 770 10. Ravenscroft G, McNamara E, Griffiths LM, Papadimitriou JM, Hardeman EC, Bakker AJ,
771 et al. Cardiac α -actin over-expression therapy in dominant ACTA1 disease. *Hum Mol*
772 *Genet.* 2013; **22**: 3987–97.
- 773 11. Threadgill DW, Churchill GA. Ten years of the collaborative cross. *Genetics.* 2012; **190**:
774 291–4.
- 775 12. Gatti DM, Harrill AH, Wright FA, Threadgill DW, Rusyn I. Replication and narrowing of
776 gene expression quantitative trait loci using inbred mice. *Mamm Genome.* 2009; **20**:
777 437–46.
- 778 13. Airey DC, Lu L, Williams RW. Genetic control of the mouse cerebellum: Identification
779 of quantitative trait loci modulating size and architecture. *J Neurosci.* 2001; **21**: 5099–
780 109.
- 781 14. Badea A, Johnson GA, Williams RW. Genetic dissection of the mouse CNS using
782 magnetic resonance microscopy. *Curr Opin Neurol.* 2009; **22**: 379–86.
- 783 15. Martin M V, Dong H, Vallera D, Lee D, Lu L, Williams RW, et al. Independent
784 quantitative trait loci influence ventral and dorsal hippocampal volume in
785 recombinant inbred strains of mice. *Genes, Brain Behav.* 2006; **5**: 614–23.
- 786 16. Di Curzio DL, Goldowitz D. The genetic basis of adrenal gland weight and structure in
787 BXD recombinant inbred mice. *Mamm Genome.* 2011; **22**: 209–34.
- 788 17. Alberts R, Chen H, Pommerenke C, Smit AB, Spijker S, Williams RW, et al. Expression
789 QTL mapping in regulatory and helper T cells from the BXD family of strains reveals
790 novel cell-specific genes, gene-gene interactions and candidate genes for auto-

791 immune disease. *BMC Genomics*. 2011; **12**.

792 18. Williams RW, Strom RC, Goldowitz D. Natural variation in neuron number in mice is
793 linked to a major quantitative trait locus on Chr 11. *J Neurosci*. 1998; **18**: 138–46.

794 19. Chesler EJ, Lu L, Shou S, Qu Y, Gu J, Wang J, et al. Complex trait analysis of gene
795 expression uncovers polygenic and pleiotropic networks that modulate nervous
796 system function. *Nat Genet*. 2005; **37**: 233–42.

797 20. Roberts A, Pardo-Manuel De Villena F, Wang W, McMillan L, Threadgill DW. The
798 polymorphism architecture of mouse genetic resources elucidated using genome-
799 wide resequencing data: Implications for QTL discovery and systems genetics. *Mamm*
800 *Genome*. 2007; **18**: 473–81.

801 21. Threadgill DW, Miller DR, Churchill GA, de Villena FP-M. The collaborative cross: A
802 recombinant inbred mouse population for the systems genetic era. *ILAR J*. 2011; **52**:
803 24–31.

804 22. Wang X, Pandey AK, Mulligan MK, Williams EG, Mozhui K, Li Z, et al. Joint mouse–
805 human phenome-wide association to test gene function and disease risk. *Nat*
806 *Commun*. 2016; **7**: 10464.

807 23. Morahan G, Balmer L, Monley D. Establishment of “The Gene Mine”: A resource for
808 rapid identification of complex trait genes. *Mamm Genome*. 2008; **19**: 390–3.

809 24. Soukup T, Zacharová G, Smerdu V. Fibre type composition of soleus and extensor
810 digitorum longus muscles in normal female inbred Lewis rats. *Acta Histochem*. 2002;
811 **104**: 399–405.

812 25. Imbeaud S, Graudens E, Boulanger V, Barlet X, Zaborski P, Eveno E, et al. Towards
813 standardization of RNA quality assessment using user-independent classifiers of
814 microcapillary electrophoresis traces. *Nucleic Acids Res*. 2005; **33**: 1–12.

815 26. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al.
816 Accurate normalization of real-time quantitative RT-PCR data by geometric averaging
817 of multiple internal control genes. *Genome Biol*. 2002; **3**.

818 27. Ravenscroft G, Colley SMJ, Walker KR, Clement S, Bringans S, Lipscombe R, et al.
819 Expression of cardiac alpha-actin spares extraocular muscles in skeletal muscle alpha-

820 actin diseases - Quantification of striated alpha-actins by MRM-mass spectrometry.
821 *Neuromuscul Disord.* 2008; **18**: 953–8.

822 28. Nowak KJ, Ravenscroft G, Jackaman C, Filipovska A, Davies SM, Lim EM, et al. Rescue
823 of skeletal muscle alpha-actin-null mice by cardiac (fetal) alpha-actin. *J Cell Biol.* 2009;
824 **185**: 903–15.

825 29. Dubowitz V, Sewry CA, Oldfors A. Spindles and Pacinian Corpuscles. *Muscle Biopsy: A*
826 *Practical Approach.* 2013. p. 33–33.

827 30. Ram R, Mehta M, Balmer L, Gatti DM, Morahan G. Rapid identification of major-effect
828 genes using the collaborative cross. *Genetics.* 2014; **198**: 75–86.

829 31. Didion JP, Buus RJ, Naghashfar Z, Threadgill DW, Morse III HC, de Villena FP-M. SNP
830 array profiling of mouse cell lines identifies their strains of origin and reveals cross-
831 contamination and widespread aneuploidy. *BMC Genomics.* 2014; **15**.

832 32. Mulligan MK, Mozhui K, Prins P, Williams RW. GeneNetwork – A Toolbox for Systems
833 Genetics. In: K S, RW W, editors. *Systems Genetics: Methods in Molecular Biology.*
834 2016. p. In Press.

835 33. Geisert EE, Lu L, Freeman-Anderson NE, Templeton JP, Nassr M, Wang X, et al. Gene
836 expression in the mouse eye: an online resource for genetics using 103 strains of
837 mice. *Mol Vis.* 2009; **15**: 1730–63.

838 34. Pirinen E, Canto C, Jo YS, Morato L, Zhang H, Menzies KJ, et al. Pharmacological
839 inhibition of poly(ADP-ribose) polymerases improves fitness and mitochondrial
840 function in skeletal muscle. *Cell Metab.* 2014; **19**: 1034–41.

841 35. Williams EG, Mouchiroud L, Frochaux M, Pandey A, Andreux PA, Deplancke B, et al.
842 An Evolutionarily Conserved Role for the Aryl Hydrocarbon Receptor in the Regulation
843 of Movement. *PLoS Genet.* 2014; **10**.

844 36. Garner I, Minty AJ, Alonso S, Barton PJ, Buckingham ME. A 5' duplication of the alpha-
845 cardiac actin gene in BALB/c mice is associated with abnormal levels of alpha-cardiac
846 and alpha-skeletal actin mRNAs in adult cardiac tissue. *EMBO J.* 1986; **5**: 2559–67.

847 37. Ram R, Morahan G. Complex Trait Analyses of The Collaborative Cross: tools and
848 databases. In: Schughart K, Williams RW, editors. *System Genetics: Methods and*

849 Protocols. 2016. p. 121–9.

850 38. Blum R, Vethantham V, Bowman C, Rudnicki M, Dynlacht BD. Genome-wide
851 identification of enhancers in skeletal muscle: The role of MyoD1. *Genes Dev.* 2012;
852 **26**: 2763–79.

853 39. Jansen RC, Nap JP. Genetical genomics: The added value from segregation. *Trends*
854 *Genet.* 2001; **17**: 388–91.

855 40. Minty A, Kedes L. Upstream regions of the human cardiac actin gene that modulate
856 its transcription in muscle cells: presence of an evolutionarily conserved repeated
857 motif. *Mol Cell Biol.* 1986; **6**: 2125–36.

858 41. Biben C, Kirschbaum BJ, Garner I, Buckingham M. Novel muscle-specific enhancer
859 sequences upstream of the cardiac actin gene. *Mol Cell Biol.* 1994; **14**: 3504–13.

860 42. Lemonnier M, Buckingham ME. Characterization of a cardiac-specific enhancer, which
861 directs alpha-cardiac actin gene transcription in the mouse adult heart. *J Biol Chem.*
862 2004; **279**: 55651–8.

863 43. Miwa T, Boxer LM, Kedes L. CArG boxes in the human cardiac alpha-actin gene are
864 core binding sites for positive trans-acting regulatory factors. *Proc Natl Acad Sci U S A.*
865 1987; **84**: 6702–6.

866 44. Gustafson TA, Miwa T, Boxer LM, Kedes L. Interaction of nuclear proteins with
867 muscle-specific regulatory sequences of the human cardiac alpha-actin promoter.
868 *Mol Cell Biol.* 1988; **8**: 4110–9.

869 45. Sartorelli V, Webster KA, Kedes L. Muscle-specific expression of the cardiac alpha-
870 actin gene requires MyoD1, CArG-box binding factor, and Sp1. *Genes Dev.* 1990; **4**:
871 1811–22.

872 46. Pette D, Staron RS. Transitions of muscle fiber phenotypic profiles. *Histochem Cell*
873 *Biol.* 2001; **115**: 359–72.

874 47. Li Y-Y, Yu H, Guo Z-M, Guo T-Q, Tu K, Li Y-X. Systematic analysis of head-to-head gene
875 organization: evolutionary conservation and potential biological relevance. *PLoS*
876 *Comput Biol.* 2006; **2**: e74.

877 48. Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, et al. A long

- noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*. 2011; **472**: 120–4.
49. Kotzin JJ, Spencer SP, McCright SJ, Kumar, DBU, Collet MA, Mowel WK, et al. The long non-coding RNA Morrbid regulates Bim and short-lived myeloid cell lifespan. *Nature*. 2016; 1–29.
 50. Verhoeven VJM, Hysi PG, Wojciechowski R, Fan Q, Guggenheim JA, Höhn R, et al. Genome-wide meta-analyses of multi-ethnic cohorts identify multiple new susceptibility loci for refractive error and myopia. *Nat Genet*. 2013; **45**: 314–8.
 51. Solouki AM, Verhoeven VJM, Duijn CM van, Verkerk AJMH, Ikram MK, Hysi PG, et al. A genome-wide association study identifies a susceptibility locus for refractive errors and myopia at 15q14. *Nat Genet*. 2014; **42**: 897–901.
 52. Chan ET, Quon GT, Chua G, Babak T, Trochesset M, Zirngibl RA, et al. Conservation of core gene expression in vertebrate tissues. *J Biol*. 2009; **8**.
 53. Fisher S, Grise EA, Vinton RM, Bessling SL, McCallion AS. Conservation of RET Regulatory Function from Human to Zebrafish Without Sequence Similarity. *Science*. 2006; **312**: 276–9.
 54. Barr CL, Misener VL. Decoding the non-coding genome: Elucidating genetic risk outside the coding genome. *Genes, Brain Behav*. 2016; **15**: 187–204.
 55. Pikkarainen S, Tokola H, Kerkelä R, Ruskoaho H. GATA transcription factors in the developing and adult heart. *Cardiovasc Res*. 2004; **63**: 196–207.
 56. Targoff KL, Colombo S, George V, Schell T, Kim S-H, Solnica-Krezel L, et al. Nkx genes are essential for maintenance of ventricular identity. *Development*. 2013; **140**: 4203–13.

Supporting Information

Supplementary Table 1. Primer sequences

Supplementary Table 2. Genes in the BXD quad 12 LOD drop region (Chr2:121.5 - 137 Mbp) of the *Actc1* eQTL

Supplementary Table 3. Human SNPs associated with expression of *ACTC1*

Supplementary Table 4. SNPs contributing to a protein change in the *Actc1* eQTL -2 LOD drop interval

Supplementary Table 5. Non-coding sequence variants in ENOCDE and FANTOM5 regulatory regions within Chr2:113870000 – 113890000 and encompassing *Actc1*

Supplementary Figure 1. Collaborative Cross stains harbouring the A/J founder haplotype at 113 to 114 Mb on chromosome 2 contain the *Actc1* 5' duplication

921 **A.** PCR primers were designed to amplify a 530 bp fragment representing the wild type
922 *Actc1* promoter or the 530 bp fragment and an 851 bp product which indicates the presence
923 of a promoter duplication. The A/J founder strain was positive for the duplication event. **B.**
924 The founding haplotypes present within Collaborative Cross strains between 113 and 114
925 Mb on chromosome 2 (x-axis) were determined and strains were sorted by *Actc1* expression
926 level in the heart (y-axis).
927

Supplementary Figure 2. Histone modification and chromatin accessibility do not differ at *Actc1* regulatory regions between B6 and NOD mice

A. A chromatin immunoprecipitation assay with antibodies for activating (H3K4me3, active transcription; H3K27ac, active enhancer) or repressing (H3K27me3, transcriptional repression) histone marks was performed with quadriceps muscle from six-week old NOD (white bars) and B6 (grey bars) mice. Promoter enrichment measured as a percentage of input DNA and normalised to total histone (H3) content was determined by qPCR amplification of known regulatory regions located upstream of the *Actc1* transcriptional start site (TSS). **B.** A chromatin accessibility assay was performed with tibialis anterior (TA) muscle from six week old NOD (white bars) and B6 (grey bars) mice. Chromatin digestion efficiency was determined by analysis of positive (open chromatin, TATA binding protein (*Tbp*)) and negative (closed chromatin, haemoglobin beta subunit (*Hbb*)) transcriptional start sites. Skeletal muscle (skeletal muscle actin (*Acta1*)) and heart myosin heavy chain 7 (*Myh7*)) positive controls were also analysed. Accessibility was determined by comparison of cut and uncut samples and normalised to the negative control locus (*Hbb*). **C.** Chromatin accessibility was subsequently determined for regulatory regions in *Actc1*.