

# Genome-wide association of multiple complex traits in outbred mice by ultra-low-coverage sequencing

Jérôme Nicod<sup>1</sup>, Robert W Davies<sup>1</sup>, Na Cai<sup>1</sup>, Carl Hassett<sup>2</sup>, Leo Goodstadt<sup>1</sup>, Cormac Cosgrove<sup>3</sup>, Benjamin K Yee<sup>4</sup>, Vikte Lionikaite<sup>5</sup>, Rebecca E McIntyre<sup>6</sup>, Carol Ann Remme<sup>7</sup>, Elisabeth M Lodder<sup>7</sup>, Jennifer S Gregory<sup>5</sup>, Tertius Hough<sup>2</sup>, Russell Joynson<sup>2</sup>, Hayley Phelps<sup>2</sup>, Barbara Nell<sup>2</sup>, Clare Rowe<sup>2</sup>, Joe Wood<sup>2</sup>, Alison Walling<sup>2</sup>, Nasrin Bopp<sup>1</sup>, Amarjit Bhomra<sup>1</sup>, Polinka Hernandez-Pliego<sup>1</sup>, Jacques Callebert<sup>8</sup>, Richard M Aspden<sup>5</sup>, Nick P Talbot<sup>9</sup>, Peter A Robbins<sup>9</sup>, Mark Harrison<sup>2</sup>, Martin Fray<sup>2</sup>, Jean-Marie Launay<sup>8</sup>, Yigal M Pinto<sup>7</sup>, David A Blizzard<sup>10</sup>, Connie R Bezzina<sup>7</sup>, David J Adams<sup>6</sup>, Paul Franken<sup>11</sup>, Tom Weaver<sup>2</sup>, Sara Wells<sup>2</sup>, Steve D M Brown<sup>12</sup>, Paul K Potter<sup>12</sup>, Paul Klenerman<sup>3</sup>, Arimantas Lionikas<sup>5</sup>, Richard Mott<sup>1,13</sup> & Jonathan Flint<sup>1,14</sup>

**Two bottlenecks impeding the genetic analysis of complex traits in rodents are access to mapping populations able to deliver gene-level mapping resolution and the need for population-specific genotyping arrays and haplotype reference panels. Here we combine low-coverage ( $0.15\times$ ) sequencing with a new method to impute the ancestral haplotype space in 1,887 commercially available outbred mice. We mapped 156 unique quantitative trait loci for 92 phenotypes at a 5% false discovery rate.** Gene-level mapping resolution was achieved at about one-fifth of the loci, implicating *Unc13c* and *Pgc1a* at loci for the quality of sleep, *Adarb2* for home cage activity, *Rtn2* for intensity of reaction to startle, *Bmp2* for wound healing, *Il15* and *Id2* for several T cell measures and *Prkca* for bone mineral content. These findings have implications for diverse areas of mammalian biology and demonstrate how genome-wide association studies can be extended via low-coverage sequencing to species with highly recombinant outbred populations.

Genome-wide association studies (GWAS) have delivered new insights into the biology and genetic architecture of complex traits, but so far they have found application primarily in human genetics<sup>1,2</sup> and in plant species where naturally occurring inbred lines exist<sup>3,4</sup>. Two obstacles stand in the way of their routine application in other species: access to a population able to deliver gene-level mapping resolution and the deployment of a genotyping technology able to capture at least the majority of sequence variants that contribute to phenotypic variation, in the absence of haplotype reference panels of the kind routinely employed in human populations to impute sequence variants.

In this study, we exploit the properties of commercially available outbred mice for GWAS in the Crl:CFW(SW)-US\_P08 (CFW) stock. In comparison to other mouse mapping populations, commercial outbred mice are maintained at relatively large effective population sizes and are descended from a relatively small number of founders, with mean minor allele frequencies (MAFs) and linkage disequilibrium (LD) resembling those found in genetically isolated human populations<sup>5</sup>. In comparison

to a human GWAS, comparatively fewer markers are needed to tag the genome, thus requiring a lower significance threshold and a smaller sample size.

GWAS methodology typically uses arrays to genotype known SNPs and represents each individual's genome as a haplotype mosaic of a reference panel of more densely typed or sequenced individuals (such as the 1000 Genomes Project<sup>6</sup>), to impute genotypes at the majority of segregating sites in a population<sup>7</sup>. However, in common with other populations that have not previously been subjected to GWAS, commercial outbred mice lack accurate catalogs of sequence variants, allele frequencies and haplotypes, thus excluding the application of standard GWAS approaches.

We show here how low-coverage sequencing overcomes these limitations. We apply a method that models each chromosome as a mosaic of unknown ancestral haplotypes that are jointly estimated as part of the analysis. Using this approach, we map the genetic basis of multiple phenotypes in almost 2,000 mice, in some cases at nearly single-gene resolution.

<sup>1</sup>Wellcome Trust Centre for Human Genetics, Oxford, UK. <sup>2</sup>Mary Lyon Centre, MRC Harwell, Harwell Science and Innovation Campus, Harwell, UK. <sup>3</sup>Peter Medawar Building for Pathogen Research, Nuffield Department of Medicine, University of Oxford, Oxford, UK. <sup>4</sup>Department of Rehabilitation Sciences, Hong Kong Polytechnic University, Hong Kong, China. <sup>5</sup>School of Medicine, Medical Sciences and Nutrition, College of Life Sciences and Medicine, University of Aberdeen, Aberdeen, UK. <sup>6</sup>Wellcome Trust Sanger Institute, Hinxton, UK. <sup>7</sup>Heart Center, Department of Clinical and Experimental Cardiology, Academic Medical Center, Amsterdam, the Netherlands. <sup>8</sup>Department of Biochemistry, AP-HP, Hôpital Lariboisière, INSERM U942, Paris, France. <sup>9</sup>Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK. <sup>10</sup>Department of Biobehavioral Health, Pennsylvania State University, University Park, Pennsylvania, USA. <sup>11</sup>Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland. <sup>12</sup>Mammalian Genetics Unit, MRC Harwell, Harwell Science and Innovation Campus, Harwell, UK. <sup>13</sup>UCL Genetics Institute, University College London, London, UK. <sup>14</sup>Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, California, USA. Correspondence should be addressed to J.F. (jflint@mednet.ucla.edu) or R.M. (r.mott@ucl.ac.uk).

Received 8 December 2015; accepted 24 May 2016; published online 4 July 2016; doi:10.1038/ng.3595

## RESULTS

## Phenotypes

We subjected 2,049 unrelated adult outbred CFW mice from Charles River Laboratories<sup>5</sup> to a 4-week phenotyping pipeline (Online Methods and **Supplementary Fig. 1**). We obtained measures for 200 phenotypes from 18 assays (Online Methods). Data are available on a mean of 1,578 mice (range of 905–1,968) per phenotype. We assigned each measure to one of the following three heuristic categories: behavior, physiological or tissue measures. Physiological measures include those taken when the mice were alive such as body weight and cardiac function, whereas the tissue measures comprise those obtained after dissection such as blood clinical chemistry and neurogenesis. The phenotypes are listed in **Supplementary Table 1**. We tested the effect of all potential covariates on the variance in each measure to regress them for the genetic analysis. The strongest effect was batch, affecting 190 measures with a mean effect of 15%.

## Genotypes

To capture all common variants in the CFW mice, we employed a two-stage genotyping strategy using low-coverage sequencing that makes use of, but does not require, previous knowledge of segregating sites. We first generated a list of candidate variant sites using the Genome Analysis Toolkit (GATK)<sup>8</sup> and then imputed genotype probabilities at these sites.

We obtained a mean coverage of  $0.15\times$  per animal for 2,073 mice (range of  $0.06\text{--}0.51\times$ ). We identified 7,073,398 SNPs in the ~370× pile-up of all sequence data that segregated in our sample and were either polymorphic in laboratory strains sequenced in the Mouse Genomes Project (MGP) or passed GATK variant quality score recalibration (VQSR) (Online Methods). We then imputed genotype dosages at these sites using our reference-panel-free method, STITCH (Online Methods and ref. 9). Following stringent quality control after imputation, we retained 5,766,828 high-quality imputed SNPs for subsequent analysis. Accuracy at these sites was very high: the mean SNP-wise correlation ( $r^2$ ) with 25,000 sites polymorphic on a genotyping microarray<sup>10</sup> using 44 samples was 0.974 before and 0.981 after quality control. We annotated the high-quality imputed SNPs using the mouse reference mm10 assembly and identified 11,931 SNP positions in protein-coding sequence causing amino acid changes in 3,938 individual genes (non-synonymous substitutions) and 25,669 SNPs that did not cause such changes (synonymous substitutions). **Supplementary Table 2** categorizes the variants by chromosome, and

**Figure 1** Sequence diversity of the CFW population. (a) Distribution of heterozygosity in 100-kb windows across the genome. (b) Histogram of genome-wide heterozygosity. (c) Example of new and total SNP density for a region of chromosome 19. Results are representative of those seen across the genome. (d) MAF density for populations of wild Indian mice ( $n = 10$ ; 44.9 million whole-genome sequencing SNPs), CFW mice ( $n = 2,073$ ; 5.7 million imputed SNPs) and HS mice ( $n = 1,904$ ; 11,000 SNPs from a genotyping array). Known CFW variation refers to variants also segregating among 14 sequenced classical inbred strains. (e) The extent of LD in CFW and HS mice. Values are mean LD  $r^2$  values for all pairs of SNPs binned by distance.

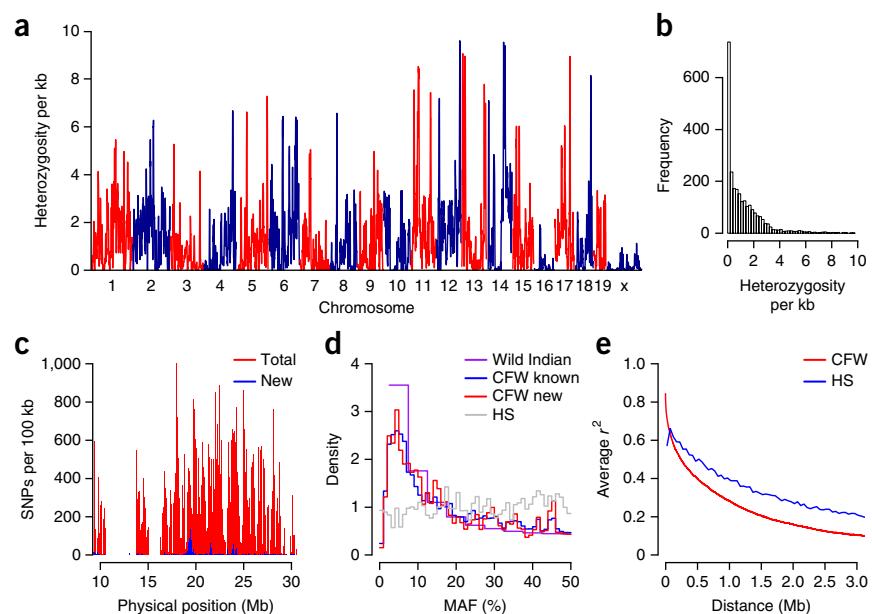
**Supplementary Table 3** lists the numbers of variants obtained at each stage of the variant calling and imputation process.

## Genetic architecture

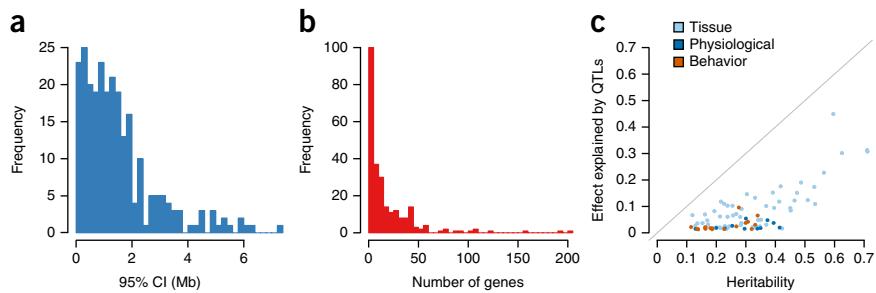
Inspection of the 5.7 million variants segregating in CFW mice identified several notable characteristics. This total number of variants is about one-third fewer than the number segregating in heterogeneous stocks derived from classical laboratory inbred strains<sup>11–13</sup> but is far less than the 45 million segregating in the recently created Collaborative Cross (CC) and Diversity Outbred (DO) populations using wild-derived strains from different subspecies of mice<sup>12</sup>. Of the 5.7 million imputed variants, 97.6% were found in 36 sequenced inbred strains in the Sanger Mouse Genomes database, release 1505. The FVB/NJ strain alone contributed 38% of the CFW alleles (**Supplementary Table 4**) and, in combination with the progenitors of Heterogeneous Stock (HS)<sup>14</sup> mice, accounted for 76% of the alleles. Wild-derived strains (LEWES/EiJ, ZALENDE/EiJ, WSB/EiJ, CAST/EiJ, MOLF/EiJ, PWK/PhJ and SPRET/EiJ) only accounted for about 5% of alternative alleles absent from other sequenced strains<sup>12,15</sup>. The new and known variants had very similar MAF distributions across the genome (**Fig. 1c,d**).

The distribution of variants across the genome was highly non-uniform (**Fig. 1a**). Chromosome 16 had only 20% of the variants found on chromosome 15, despite these chromosomes being almost the same size (**Supplementary Table 2**). This non-uniform distribution likely reflects an extreme bottleneck in the founding of the CFW line, a view supported by the fact that only four ancestral haplotypes were required for the imputation procedure to work effectively. Further, just two haplotypes modeled the majority of samples: across chromosome 19, on average, 87.1% of samples were represented by only two haplotypes. Rates of heterozygosity were low (**Fig. 1b**), with 22% of the genome close to fixation (**Fig. 1a,c**). The average MAF was 0.19. The decay of LD with increasing distance is shown in **Figure 1e** (providing an indication of the expected mapping resolution obtainable with CFW mice). Average pairwise LD  $r^2$  values fell to 0.28 at 1 Mb, to 0.16 at 2 Mb, and to 0.10 at 3 Mb.

We then identified a subset of 359,559 SNPs that tagged all other SNPs with MAF  $>0.1\%$  at LD  $r^2 > 0.98$  (Online Methods). This subset was used for subsequent analyses except where stated otherwise. To investigate population structure and unequal relatedness between



**Figure 2** Mapping resolution and effect size of QTLs. (a,b) Frequency distributions of the size (a) and number of genes present (b) for the 95% confidence interval (CI) of 255 QTLs. (c) The sum of the variance explained by QTLs plotted against heritability for 92 measures where heritability could be estimated and at least one QTL was detected. The color of the dots corresponds to the type of measure: behavior, physiological (body weight, respiratory, electrocardiography) or tissue (any measure obtained after dissection).



animals, we estimated identity by descent (IBD) from allele sharing by tagging SNPs. The proportion of the genome with IBD = 1 is plotted against the proportion of the genome with IBD = 0 in **Supplementary Figure 2**. For GWAS, we removed 135 animals with greater relatedness than second-degree relatives and 4 outliers identified from principal-component analysis (PCA) on a genetic relatedness matrix (GRM) (Online Methods). The population structure of the remaining 1,934 animals was further assessed by performing another PCA on a GRM from only these mice. The relationships between the first five principal components show no evidence of population structure (**Supplementary Fig. 3**).

### Genome-wide association

Genotypes and phenotypes were available for 1,887 mice. We performed GWAS by testing association between the 359,559 tagging SNPs and all phenotypes, using a mixed model. We transformed each phenotype by regression on relevant covariates (Online Methods) and quantile normalized the residuals. To test for association with SNPs on a given chromosome, we used a GRM based on the tagging SNPs on the other chromosomes<sup>16,17</sup> to increase power<sup>18</sup>. We calculated a genome-wide false discovery rate (FDR) separately for each phenotype to determine empirical trait-specific genome-wide significance thresholds (Online Methods).

At an FDR of 5%, we identified 255 quantitative trait loci (QTLs) for 92 of 200 phenotypes (46%) (**Supplementary Table 5**). Quantile-quantile plots for a representative selection of phenotypes are given in **Supplementary Figure 4**. It should be noted that, owing to the large number of SNPs used (in this case not pruned for LD) and the fact that LD extends over longer distances than in human populations, deviation from the expected values extended over a larger range of *P* values than is commonly seen in human association studies.

Statistical power is expected to increase with MAF, and in our QTLs the MAF of significantly associated SNPs (range of 1.7–50%, median of 31%) was higher than expected (in comparison to all 5.7 million SNPs) (Mann–Whitney *U* test,  $P = 1.95 \times 10^{-28}$ ): at 133 QTLs (52%) MAF was >30% and at only 11 QTLs (4%) was MAF <5%.

To aid in gene identification, we estimated the 95% confidence interval of every QTL using simulations based on the logarithm of odds (LOD) drop concept<sup>19</sup>. To do so, using the imputed dosages around each QTL, we simulated causal SNPs that matched the observed effect size of each QTL. A local scan of the region using the mixed model but with a simulated phenotype was performed, and the location and log *P* value for the top SNP were recorded. From 1,000 simulations, we derived an empirical distribution of the drop ( $\Delta$ ) in log *P* value between the most highly associated SNP and the causal SNP ( $\Delta$  is zero when the top and causal SNPs coincide). After ranking the simulations at a given QTL by increasing value of  $\Delta$ , the LOD drop  $\Delta(f)$  of the  $f\%$  confidence interval was estimated as the maximum seen among the simulations with the lowest  $f\%$ . The genomic interval

spanning the LOD drop  $\Delta(f)$  determined the confidence interval of the QTL in the real phenotype data<sup>19</sup>. Across all QTLs, the widths of the  $f = 95\%$  confidence interval ranged from 0.01 to 7.33 Mb with a mean at 1.50 Mb; 43% of the intervals were less than 1 Mb in width (**Fig. 2a**). On average, each QTL covered 19 protein-coding genes (range of 0–205) with a median of 9 genes. The distribution of the number of genes in a QTL is shown in **Figure 2b**.

### Heritability and variance attributable to QTLs

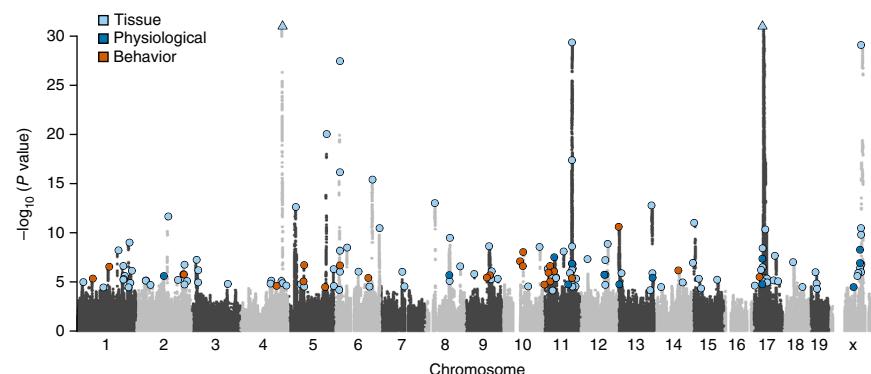
SNP-based heritability estimates exceeded 0 (at  $P < 0.05$ ) for 152 of the 200 phenotypes, with a mean value of 26.3% (range of 9.1–71.1%) (**Supplementary Table 1**). To assess how much of the heritability can be explained by detected QTLs (FDR < 5%), we first estimated the effect size of each QTL by performing analysis of variance (ANOVA) at the most significantly associated SNP and then summed the variance explained by all QTLs associated with each phenotype. On average, 21.1% of the heritability estimated for each trait with significant heritability could be explained in this way (**Fig. 2c**). This indicates that missing heritability affects the CFW population, although to a lesser degree than in most human GWAS.

Traits with higher heritabilities yielded proportionally more QTLs: the mean heritability of traits for which at least one QTL was identified was 30.6%, as compared to 20.6% for traits without QTLs, a highly significant difference (*t* test,  $P = 8.9 \times 10^{-8}$ ). Mean heritabilities differed among the three categories of phenotypes: 14.5% for behavior, 18.2% for physiological and 24.2% for tissue phenotypes. We also noted the same pattern for the median locus effect size of the three categories: 1.37% for behavioral QTLs, 1.5% for physiological QTLs and 2.8% for tissue QTLs (**Fig. 2c**).

### Distribution of QTLs reflects genome-wide diversity

Many of the loci detected overlapped and were associated with closely related phenotypic measures. Examples include the two QTLs for HDL and total cholesterol mapping over the *Apoa2* gene on chromosome 1 (ref. 20) and the eight different bone mineral content measures mapping over *Slc4a2* on chromosome 5 (ref. 21). To avoid redundancy in our analysis, we considered that if two overlapping QTLs (where the top SNP of the first QTL lies inside the 95% confidence interval of the second QTL) were associated with measures of the same biological function they were representing a single locus. Using this approach, we identified a reduced set of 156 unique loci, each associated with 1 to 12 measures. We report these 156 unique QTLs in **Supplementary Table 6**. A ‘porcupine’ plot shows the superimposed Manhattan plots for all the measures where at least one QTL was detected and highlights the 156 unique loci (**Fig. 3**). Some regions of the genome were devoid of any QTLs, reflecting the uneven genomic distribution of sequence variants, with a prime example being the lack of any QTLs detected on chromosome 16 (**Fig. 1a**). The presence of clusters of QTLs, notably on chromosomes 6,

**Figure 3** Summary porcupine plot for 92 phenotypes. Genome-wide representation of all unique QTLs ( $n = 156$ , FDR < 5%) identified in this study. Light and dark gray dots show associations for the 92 measures where at least one QTL was detected at tagging SNP positions ( $n = 359,559$ ). The most significant SNP at each QTL is marked by a colored dot, with the color corresponding to the type of measure. The y axis shows the  $-\log_{10} P$  values for association of the imputed allele dosages with tested measures and is truncated at  $-\log_{10} P = 32$ . The positions of the two strongest QTLs with  $-\log_{10} P$  values of 133 (chromosome 4) and 76 (chromosome 17) are marked by triangles.



11, 17 and X, is also highlighted in **Figure 3**. The chromosome 17 locus overlapped the major histocompatibility complex (MHC) region, a naturally highly polymorphic region in wild populations that remains highly variable in CFW mice.

### Identification of candidate genes in high-resolution QTLs

We focused on QTLs containing small numbers of genes because these loci provide a starting point for functional investigations. Of the 156 unique loci identified in this study, 56 contained three or fewer genes (36%) and 25 contained a single gene in the 95% confidence interval (6 QTLs did not overlap any gene).

The 25 QTLs containing a single gene are listed in **Table 1**. The table categorizes QTLs into three classes, according to previous evidence supporting the candidacy of the gene at the locus. The first class includes QTLs with genes whose candidacy is supported by the phenotypes of knockouts: *Met*, *Fli1* and *Grm7*. The locus on chromosome 6 containing the *Met* gene contributed to weight measures for all five muscles assessed (**Fig. 4a**). *Met* encodes a hepatocyte growth factor receptor and has a known function in embryonic development<sup>22,23</sup> and regeneration<sup>24</sup> of adult limb skeletal muscle. *Fli1* modulates B cell development<sup>25</sup>, and mice lacking *Grm7* are more active when placed in a new environment<sup>26</sup>. The second class comprises six loci with genes that are strongly corroborated by previously published evidence. The genes include the bone morphogenetic protein *Bmp2* gene at a locus for wound healing<sup>27</sup>; *Pgc1a* (encoding PGC-1 $\alpha$ ) at a locus for sleep fragmentation, which is involved in regulating inhibitory neurotransmission in the cerebral cortex in association with cortical hyperexcitability<sup>28</sup>; a gene encoding protein kinase C $\alpha$ , *Prkca*, which promotes osteoblastic cell proliferation, at a locus for bone mineral content<sup>29,30</sup>; *Pbx1* (encoding pre-B cell leukemia homeobox 1) at a locus influencing the natural killer (NK) cell population<sup>31</sup>; and an interleukin (*Il15*; **Fig. 4c**)<sup>32</sup> and the transcription factor *Id2* (ref. 33) at two independent loci affecting several T cell measures. The remaining 16 QTLs in the third class contain single genes not previously associated with the trait, including five that concern behavior.

Notably, our mapping results implicate *Unc13c* in the quality of sleep (**Fig. 4b**). UNC13C is involved in synaptic transmission<sup>34</sup> but to our knowledge has never previously been associated with sleep. However, there is evidence for differential expression of the human ortholog in individuals with poor sleep quality<sup>35</sup>. Basal home cage activity was associated with *Adarb2*, a brain-specific adenosine deaminase acting on RNA<sup>36,37</sup>. *Rtn2*, a member of the rhotekin family predominantly expressed in lymphoid cells<sup>38</sup>, influenced the intensity of reaction to startle (**Fig. 4d**). Copy number variants (CNVs) for the human ortholog of *Rtn2* have been implicated in attention-deficit and hyperactivity disorder<sup>39</sup>.

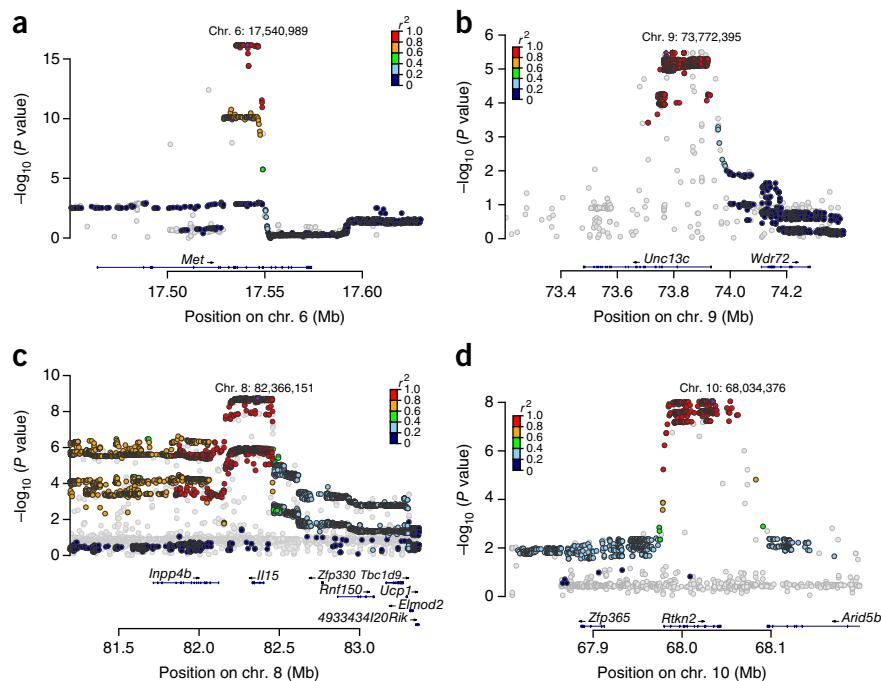
### DISCUSSION

Genome-wide association mapping for complex traits has been used extensively in human populations but less commonly in other

**Table 1** QTLs mapping to a single gene

Phenotype	Chr.	Position (Mb)	$-\log P$	Gene	Reference
<b>Knockout mouse recapitulates the phenotype</b>					
Weight of soleus muscle (g)	6	17.5	16.2	<i>Met</i>	22–24
Total distance traveled in Elevated Plus Maze (cm)	6	110.2	5.6	<i>Grm7</i>	26
CD45 $^+$ CD3 $^-$ CD19 $^+$ cells (%)	9	32.6	5.8	<i>Fli1</i>	25
<b>Association supported by literature</b>					
CD45 $^+$ CD3 $^-$ DX5 $^+$ cells (%)	1	168.2	4.7	<i>Pbx1</i>	31
Wound healing	2	134.2	5.5	<i>Bmp2</i>	27
Number of long (>1-min) sleep episodes	5	51.8	6.8	<i>Ppargc1a</i>	28
Ratio of CD3 $^+$ CD4 $^+$ to CD3 $^+$ CD8 $^+$ cells	8	82.4	8.7	<i>Il15</i>	32
Bone mineral content	11	108.2	4.6	<i>Prkca</i>	29,30
CD3 $^+$ CD8 $^+$ cells (%)	12	25.5	5.4	<i>Id2</i>	33
<b>No previous evidence</b>					
Length of tibia (mm)	5	51.7	4.5	<i>Ppargc1a</i>	
Startle pulse reactivity	6	17.5	6.7	<i>Met</i>	
Calcium (mmol/l)	6	17.5	8.3	<i>Met</i>	
Total cholesterol (mmol/l)	6	17.5	6.1	<i>Met</i>	
Total protein (g/l)	6	17.5	27.7	<i>Met</i>	
CD45 $^+$ CD3 $^-$ CD19 $^+$ cells (%)	7	72.2	6.2	<i>Mctp2</i>	
Number of long (>1-min) sleep episodes	9	73.8	5.5	<i>Unc13c</i>	
Startle pulse reactivity	10	68.0	8.0	<i>Rtn2</i>	
Weight of tibialis anterior muscle (g)	11	17.6	6.3	<i>Etaa1</i>	
Length of tibia (mm)	12	83.6	7.1	<i>Zfyve1</i>	
Basal activity	13	7.3	10.6	<i>Adarb2</i>	
Respiratory rate during hypoxic ventilatory decline	13	118.0	5.9	<i>Hcn1</i>	
Total distance traveled in Elevated Plus Maze (cm)	14	82.1	6.2	<i>Pcdh17</i>	
Measure of the size of tibia	15	26.6	5.3	<i>Fbxl7</i>	
Eosinophils (%)	17	70.4	5.2	<i>Dlgap1</i>	
Eosinophils (%)	X	155.6	6.0	<i>Ptchd1</i>	

**Figure 4** Single-gene-resolution mapping at four loci using the entire set of SNPs (7.1 million). (a) Weight of soleus muscle on chromosome 6 ( $n = 1,832$ ). (b) Measure of the number of long sleep episodes on chromosome 9 ( $n = 1,577$ ). (c) Ratio of CD3<sup>+</sup>CD4<sup>+</sup> to CD3<sup>+</sup>CD8<sup>+</sup> cells on chromosome 8 ( $n = 1,324$ ). (d) Intensity of reaction to startle on chromosome 10 ( $n = 1,740$ ). The strongest associated SNP for each region is marked with a purple diamond, and the other SNPs that passed post-imputation quality control (IMPUTE2-style INFO scores  $>0.4$  and Hardy-Weinberg equilibrium  $P > 1 \times 10^{-6}$ ) are colored according to LD  $r^2$  with the strongest SNP. The gray dots represent SNPs that failed post-imputation quality control and therefore were not used for the analysis.



organisms. We have shown here that mapping using commercially available outbred mice can identify individual genes involved in complex traits, some of which cannot easily be assayed in human subjects. Our results raise issues about the nature of mouse resources for mapping complex traits and about the biological insights that can thereby be attained.

Several resources have been developed to provide GWAS tools to rodent genetics. These resources fall into two broad categories: (i) genetic reference populations, consisting of preexisting inbred strains (Hybrid Mouse Diversity Panel, HMDP<sup>13</sup>) or recombinant inbred strains (BxD<sup>40</sup> and CC<sup>41</sup>), and (ii) populations descended after multiple generations of pseudo-random breeding from inbred strains (DO<sup>42</sup> and HS<sup>14</sup>). Each resource differs in its usefulness for GWAS, and no single population is ideal<sup>1</sup>.

Commercially available outbred mice are an alternative resource with a number of advantages, and the CFW stock has already been used to map QTLs for skull shape<sup>43</sup>. In comparison to the HMDP and HS mice, there was minimal evidence for population structure in CFW mice, and standard GWAS methods developed for human populations can be applied. LD decays quickly enough to provide gene-level mapping resolution at about one-fifth of loci, and although the resolution is still lower than in human populations it is better than in other mouse resources. The size of QTLs varied considerably, with the largest ones extending over several megabases, but half contained fewer than ten genes, providing a relatively small list to investigate the biology at these loci.

In comparison to other rodent mapping resources, our results also indicate that the CFW population delivers fewer loci for fewer phenotypes. We mapped loci for 92 of the 200 traits included in our phenotyping pipeline, yielding a mean number of 1.3 QTLs per trait, in 1,887 mice. One possible explanation for the low yield of QTLs is that the amount of genetic variation present in the CFW stock is relatively limited. Indeed, almost one-quarter of the CFW genome is virtually devoid of variants. For comparison, the 5.7 million variants in CFW mice is less than the 7.2 million segregating in the rat heterogeneous stock<sup>11</sup>. However, a more important determinant for QTL detection in CFW mice is likely to be allele frequencies ( $p$ ), which are on average lower in CFW than in heterogeneous stocks. Because the variance explained by a QTL is proportional to  $p(1-p)$ , effect sizes—and hence power—are systematically smaller in CFW mice. Indeed, the median effect size is 1.6%, which, although dwarfing the effects found

to underlie human quantitative traits, is still less than half that found in the rat heterogeneous stock (median estimate of 5%)<sup>11</sup>.

The inclusion of a large number of behavioral measures in our pipeline also contributed to the relatively low QTL detection rate. Measures for almost one-third of the traits (63/200) were collected from behavioral tests, yet the QTLs that mapped with these measures accounted for less than 14% of the total. These phenotypes typically had lower heritabilities, with fully one-quarter (16) having no significant genetic contribution. Note that these non-significant estimates (as well as those for non-behavioral phenotypes) do not necessarily mean that the traits are not heritable: the standard errors on these estimates are large (Supplementary Table 1), such that no heritability less than 10% can be reliably estimated. Those loci we did detect had relatively lower effect sizes (the mean for behavioral QTLs was 1.37, as compared to 1.5 for physiological QTLs and 2.8 for tissue QTLs).

The heritability of the behavioral measures might also have been affected by the fact that mice were repeatedly tested over a 4-week period. Most behaviors are sensitive to repeated handling and exposure to different types of new stimuli, as will happen during the extensive phenotypic battery deployed here. Habituation to these exposures makes it harder to detect alleles that affect baseline differences in behavior, especially anxiety-like behaviors for which three different assays were conducted over a relatively short time frame. A more focused assessment of a specific behavioral phenotype under tightly controlled environmental conditions could have yielded higher heritabilities for some traits.

These observations lead to two conclusions. First, finding more QTLs in the CFW population will require thousands of mice. The power to detect QTLs in the CFW population as a function of effect size and sample size is given in Supplementary Table 7. For a typical QTL corresponding to the median effect size (1.6%) and sample size (1,732) in the current study, power is about 80% at a genome-wide significance level of 10%. Power falls off for smaller effect sizes: a QTL with an effect of 0.5% is detectable with 6.6% power when using 1,732 animals; increasing the sample size to 4,000 increases the power to 51%, and with 6,000 animals it is 85%. However, ‘winner’s curse’ means that the true effects are likely to be lower than reported here,

and given that our QTLs explain only 20% of the heritability it is reasonable to assume that the majority of loci will have effect sizes less than 1%. Second, additional loci can be found using different stocks. Not all commercial outbred mice populations are the same, as we previously documented in a survey of 66 stocks in which mean heterozygosity varied from 0.5 to 45% and mean MAFs ranged from 0.003 to 0.5 (ref. 5). The use of complementary populations will make additional alleles open to discovery.

To our knowledge, our study is the first to use extremely low-coverage sequence to generate accurate genotypes without a reference panel. This strategy, and the associated STITCH algorithm<sup>9</sup>, is generally applicable to any population and any species for which there is no information about segregating variation or haplotypes. It is competitive with arrays in terms of cost, although the choice of optimal strategy will depend on the reagents available for the population in question. An advantage of sequencing over array-based genotyping is that it does not require previous information about which variants segregate in a population; nor does it require a preexisting catalog of variants or previous knowledge of the likely founders of the population. The only requirement is a high-quality reference genome.

One unexpected finding was that only 25,000 SNPs on the standard MegaMUGA mouse genotyping array are polymorphic in CFW mice; many of the QTLs we mapped would likely have been overlooked by genotyping with this array. The CFW population seems to be descended from four ancestral haplotypes, indicating that this population was likely bottlenecked to two founding individuals. Our population is effectively biallelic at most loci, and there was little to be gained by considering haplotype-based tests of association (data not shown).

We could also test associations at candidate variants responsible for the effect. For example, the *Met* gene on chromosome 6 is associated with muscle phenotypes and our sequence data identified two encoded missense variants, p.Ile851Met and p.Arg968Cys. The first variant is common among mouse strains and is not known to alter gene function. The second variant, confirmed by Sanger sequencing, is specific to the SWR/J strain<sup>44</sup>. The human counterpart for this variant (p.Arg988Cys) has been identified in two small-cell lung cancer cell lines and increases constitutive tyrosine phosphorylation activity *in vitro*<sup>45</sup>. The p.Arg968Cys missense variant is associated with the five muscle weight phenotypes assessed, but the direction of the effect of the alternative allele is positive for extensor digitorum longus (EDL) and gastrocnemius and negative for the other muscles. This disparity reflects differences in muscle fiber composition (soleus is dominated by type 1 and 2A fibers, whereas EDL is enriched in 2X and 2B fibers<sup>46</sup>), suggesting that p.Arg968Cys affects these fibers differently or shifts the composition in all muscles.

We found single genes at 16 loci where no previous evidence existed for their involvement (**Table 1**). Notably, the loci include ones for phenotypes that could not easily be assayed in human subjects, such as response to hypoxia and sleep phenotypes. More than 50 QTLs contain documented candidate genes (**Supplementary Table 5**): *Slc4a2*, which leads to osteopetrosis when disrupted in mice<sup>21</sup>, is present at a QTL affecting bone mineral content; *Apoa2* and *Scarb1*, both known to affect blood lipid homeostasis<sup>20,47</sup>, are detected at two distinct QTLs for cholesterol levels; and *Gdnf*, a gene required for the neuronal colonization of the pancreas, is detected at a locus for pancreatic amylase<sup>48</sup>. These examples demonstrate that the narrow QTLs detected in CFW mice can lead to the identification of genes affecting the measured traits, emphasizing the potential of our

results as a resource to identify new genes in QTLs without documented candidates.

**URLs.** Results from this project and the data used for analysis are maintained in an open access database and are available at <http://outbredmice.org/>. STITCH is available from <http://www.stats.ox.ac.uk/~myers/>.

## METHODS

Methods and any associated references are available in the [online version of the paper](#).

**Accession codes.** Sequencing data have been deposited in the European Nucleotide Archive (ENA) under accession [ERP001040](#).

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*

## ACKNOWLEDGMENTS

The authors wish to acknowledge excellent technical assistance from A. Kurioka, L. Swadling, C. de Lara, J. Ussher, R. Townsend, S. Lionikaite, A.S. Lionikiene, R. Wolswinkel and I. van der Made. We would like to thank T.M. Keane and A.G. Doran for their help in annotating variants and adding the FVB/NJ strain to the MGP. We thank the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics and the Wellcome Trust Sanger Institute for the generation of the sequencing data. This work was funded by Wellcome Trust grant 090532/Z/09/Z (J.F.). Primary phenotyping of the mice was supported by the Mary Lyon Centre and Mammalian Genetics Unit (Medical Research Council, UK Hub grant G0900747 91070 and Medical Research Council, UK grant MC U142684172). D.A.B. acknowledges support from NIH R01AR056280. The sleep work was supported by the state of Vaud (Switzerland) and the Swiss National Science Foundation (SNF 14694 and 136201 to P.F.). The ECG work was supported by the Netherlands CardioVascular Research Initiative (Dutch Heart Foundation, Dutch Federation of University Medical Centres, Netherlands Organization for Health Research and Development and the Royal Netherlands Academy of Sciences) PREDICT project, InterUniversity Cardiology Institute of the Netherlands (ICIN; 061.02; C.A.R. and C.R.B.). N.C. is supported by the Agency of Science, Technology and Research (A\*STAR) Graduate Academy. R.W.D. is supported by a grant from the Wellcome Trust (097308/Z/11/Z).

## AUTHOR CONTRIBUTIONS

J.N. and J.F. designed the study and experiments. J.N., B.K.Y. and N.C. processed data. C.C., R.E.M., N.B., A.B., P.H.-P., C.H., R.J., H.P., B.N., C.R. and T.H. phenotyped the mice and generated data. J.W. and A.W. developed bespoke Laboratory Information Management System (LIMS) and bioinformatics solutions for data collection. M.H. and M.F. managed importation and isolation procedures of mice into the Mary Lyon Centre. S.W., T.W. and S.D.M.B. provided infrastructure and staff, and established the phenotyping within the MLC. P.K.P. and J.N. managed the project. V.L., J.S.G. and R.M.A. quantified bone size and mineral content. D.A.B. and A.L. acquired skeletal muscle phenotypes. C.A.R., E.M.L., Y.M.P. and C.R.B. supervised cardiac data acquisition and analyzed the cardiac data. J.C. and J.-M.L. quantified serotonin. J.N., C.C., R.E.M., P.F., B.K.Y., D.J.A., P.K., N.P.T., P.A.R. and A.L. analyzed the phenotypic data. N.C. and L.G. processed the sequencing data. R.W.D. and L.G. performed genotype imputation. R.M., J.N., N.C. and J.F. performed the genetic analysis. J.N., R.W.D., N.C., R.M. and J.F. wrote the manuscript with input from co-authors.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Flint, J. & Eskin, E. Genome-wide association studies in mice. *Nat. Rev. Genet.* **13**, 807–817 (2012).
- Visscher, P.M., Brown, M.A., McCarthy, M.I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).
- Atwell, S. *et al.* Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465**, 627–631 (2010).
- Huang, X. *et al.* Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* **42**, 961–967 (2010).

5. Yalcin, B. *et al.* Commercially available outbred mice for genome-wide association studies. *PLoS Genet.* **6**, e1001085 (2010).
6. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
7. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* **11**, 499–511 (2010).
8. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
9. Davies, R.W., Flint, J., Myers, S. & Mott, R. Rapid genotype imputation from sequence without reference panels. *Nat. Genet.* <http://dx.doi.org/10.1038/ng.3594> (2016).
10. Yang, H. *et al.* A customized and versatile high-density genotyping array for the mouse. *Nat. Methods* **6**, 663–666 (2009).
11. Baud, A. *et al.* Combined sequence-based and genetic mapping analysis of complex traits in outbred rats. *Nat. Genet.* **45**, 767–775 (2013).
12. Keane, T.M. *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289–294 (2011).
13. Bennett, B.J. *et al.* A high-resolution association mapping panel for the dissection of complex traits in mice. *Genome Res.* **20**, 281–290 (2010).
14. Valdar, W. *et al.* Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nat. Genet.* **38**, 879–887 (2006).
15. Wong, K. *et al.* Sequencing and characterization of the FVB/NJ mouse genome. *Genome Biol.* **13**, R72 (2012).
16. Listgarten, J. *et al.* Improved linear mixed models for genome-wide association studies. *Nat. Methods* **9**, 525–526 (2012).
17. Yang, J., Zaitlen, N.A., Goddard, M.E., Visscher, P.M. & Price, A.L. Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* **46**, 100–106 (2014).
18. Cheng, R., Parker, C.C., Abney, M. & Palmer, A.A. Practical considerations regarding the use of genotype and pedigree data to model relatedness in the context of genome-wide association studies. *G3 (Bethesda)* **3**, 1861–1867 (2013).
19. Manichaikul, A., Dupuis, J., Sen, S. & Broman, K.W. Poor performance of bootstrap confidence intervals for the location of a quantitative trait locus. *Genetics* **174**, 481–489 (2006).
20. Weng, W. & Breslow, J.L. Dramatically decreased high density lipoprotein cholesterol, increased remnant clearance, and insulin hypersensitivity in apolipoprotein A-II knockout mice suggest a complex role for apolipoprotein A-II in atherosclerosis susceptibility. *Proc. Natl. Acad. Sci. USA* **93**, 14788–14794 (1996).
21. Coury, F. *et al.* SLC4A2-mediated Cl<sup>-</sup>/HCO<sub>3</sub><sup>-</sup> exchange activity is essential for calpain-dependent regulation of the actin cytoskeleton in osteoclasts. *Proc. Natl. Acad. Sci. USA* **110**, 2163–2168 (2013).
22. Bladt, F., Riethmacher, D., Isenmann, S., Aguzzi, A. & Birchmeier, C. Essential role for the c-met receptor in the migration of myogenic precursor cells into the limb bud. *Nature* **376**, 768–771 (1995).
23. Dietrich, S. *et al.* The role of SF/HGF and c-Met in the development of skeletal muscle. *Development* **126**, 1621–1629 (1999).
24. Webster, M.T. & Fan, C.M. c-MET regulates myoblast motility and myocyte fusion during adult skeletal muscle regeneration. *PLoS One* **8**, e81755 (2013).
25. Zhang, X.K. *et al.* The transcription factor Fli-1 modulates marginal zone and follicular B cell development in mice. *J. Immunol.* **181**, 1644–1654 (2008).
26. Cryan, J.F. *et al.* Antidepressant and anxiolytic-like effects in mice lacking the group III metabotropic glutamate receptor mGluR7. *Eur. J. Neurosci.* **17**, 2409–2417 (2003).
27. Duprez, D.M., Coltey, M., Amthor, H., Brickell, P.M. & Tickle, C. Bone morphogenetic protein-2 (BMP-2) inhibits muscle development and promotes cartilage formation in chick limb bud cultures. *Dev. Biol.* **174**, 448–452 (1996).
28. Dougherty, S.E. *et al.* Mice lacking the transcriptional coactivator PGC-1α exhibit alterations in inhibitory synaptic transmission in the motor cortex. *Neuroscience* **271**, 137–148 (2014).
29. Nakura, A., Higuchi, C., Yoshida, K. & Yoshikawa, H. PKCα suppresses osteoblastic differentiation. *Bone* **48**, 476–484 (2011).
30. Galea, G.L. *et al.* Protein kinase Cα (PKCα) regulates bone architecture and osteoblast activity. *J. Biol. Chem.* **289**, 25509–25522 (2014).
31. Sanyal, M. *et al.* B-cell development fails in the absence of the *Pbx1* proto-oncogene. *Blood* **109**, 4191–4199 (2007).
32. Kennedy, M.K. *et al.* Reversible defects in natural killer and memory CD8 T cell lineages in interleukin 15-deficient mice. *J. Exp. Med.* **191**, 771–780 (2000).
33. Cannarile, M.A. *et al.* Transcriptional regulator Id2 mediates CD8+ T cell immunity. *Nat. Immunol.* **7**, 1317–1325 (2006).
34. Chen, Z., Cooper, B., Kalla, S., Varoqueaux, F. & Young, S.M. Jr. The Munc13 proteins differentially regulate readily releasable pool dynamics and calcium-dependent recovery at a central synapse. *J. Neurosci.* **33**, 8336–8351 (2013).
35. Reddy, S.Y. *et al.* Sleep quality, *BDNF* genotype and gene expression in individuals with chronic abdominal pain. *BMC Med. Genomics* **7**, 61 (2014).
36. Melcher, T. *et al.* RED2, a brain-specific member of the RNA-specific adenosine deaminase family. *J. Biol. Chem.* **271**, 31795–31798 (1996).
37. Mittaz, L., Antonarakis, S.E., Higuchi, M. & Scott, H.S. Localization of a novel human RNA-editing deaminase (hRED2 or ADARB2) to chromosome 10p15. *Hum. Genet.* **100**, 398–400 (1997).
38. Collier, F.M. *et al.* Identification and characterization of a lymphocytic Rho-GTPase effector: rhotekin-2. *Biochem. Biophys. Res. Commun.* **324**, 1360–1369 (2004).
39. Ramos-Quiroga, J.A. *et al.* Genome-wide copy number variation analysis in adult attention-deficit and hyperactivity disorder. *J. Psychiatr. Res.* **49**, 60–67 (2014).
40. Peirce, J.L., Lu, L., Gu, J., Silver, L.M. & Williams, R.W. A new set of BXD recombinant inbred lines from advanced intercross populations in mice. *BMC Genet.* **5**, 7 (2004).
41. Churchill, G.A. *et al.* The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat. Genet.* **36**, 1133–1137 (2004).
42. Svenson, K.L. *et al.* High-resolution genetic mapping using the Mouse Diversity Outbred population. *Genetics* **190**, 437–447 (2012).
43. Pallares, L.F. *et al.* Mapping of craniofacial traits in outbred mice identifies major developmental genes involved in shape determination. *PLoS Genet.* **11**, e1005607 (2015).
44. Zaffaroni, D. *et al.* Met proto-oncogene juxtamembrane rare variations in mouse and humans: differential effects of Arg and Cys alleles on mouse lung tumorigenesis. *Oncogene* **24**, 1084–1090 (2005).
45. Ma, P.C. *et al.* c-MET mutational analysis in small cell lung cancer: novel juxtamembrane domain mutations regulating cytoskeletal functions. *Cancer Res.* **63**, 6272–6281 (2003).
46. Bloomberg, D. & Quadrilatero, J. Rapid determination of myosin heavy chain expression in rat, mouse, and human skeletal muscle using multicolor immunofluorescence analysis. *PLoS One* **7**, e35273 (2012).
47. Varban, M.L. *et al.* Targeted mutation reveals a central role for SR-BI in hepatic selective uptake of high density lipoprotein cholesterol. *Proc. Natl. Acad. Sci. USA* **95**, 4619–4624 (1998).
48. Muñoz-Bravo, J.L. *et al.* GDNF is required for neural colonization of the pancreas. *Development* **140**, 3669–3679 (2013).

## ONLINE METHODS

**Study animals and phenotyping.** A total of 2,117 outbred mice (Crl:CFW(SW)-US\_P08 (CFW); 1,065 males and 1,052 females) were purchased from Charles River Laboratories, at 4–7 weeks of age over a period of 2 years. Mice were selected from the breeding colony so as to avoid the selection of siblings and half-siblings. Monthly shipments of approximately 130 mice were delivered, maintained and tested at the MRC Harwell following local regulations. Mice of the same age in each shipment were treated as a batch (approximately 30 mice, range of 7 to 36 mice, half males and half females; the total number of batches for the entire study was 69), and each mouse was randomly assigned a testing order. Mice were housed in IVC cages (three per cage) on an *ad libitum* diet for the duration of the study. At 16 weeks of age, 2,049 mice started a 4-week phenotyping pipeline in which we collected behavioral and physiological data (**Supplementary Fig. 1**; a full description of the phenotypes measured is provided in the **Supplementary Note**). Tests were performed on mice from a single batch on the same day, following the assigned testing order. Sequencing of the mice was performed after completion of the study, with experimenters blinded to the genotype of the mice during testing. Power calculations to estimate the sample size for the mapping experiment assumed that effect sizes were similar to those identified in a previous analysis of outbred stocks<sup>5</sup>. Every effort was made to minimize suffering by considerate housing and husbandry. All phenotyping procedures were examined for potential refinements. All animal work was carried out in accordance with UK Home Office regulations. The project was reviewed by the ethics committee at MRC Harwell: Animal Welfare and Ethical Review Board, approval license PPL 30/2653.

**Preprocessing of phenotype data.** Analysis of the phenotypic data was performed using R statistical analysis software<sup>49</sup>. Outliers, defined as observations more than 3 s.d. from the mean, were excluded. The effect of covariates such as sex and batch on quantitative phenotypes was assessed with analysis of variance (ANOVA), and covariates explaining more than 1% of the variance at  $P < 0.05$  were included in a multiple linear regression model from which residual measures were obtained. Batch—defined here as mice of the same age in each individual shipment—was treated as a random effect. All tests with covariates and models used to generate the residuals for genetic mapping are shown in **Supplementary Table 1**. We then quantile normalized the residuals to minimize the effects of non-normality.

**Sequencing.** Genomic DNA was extracted from the tissue samples of 2,028 mice that began the pipeline using Nucleon BACC resin (Hologic) following the manufacturer's instructions. DNA was obtained from an additional 45 mice from the same population for which no phenotypic measures were available, producing a total of 2,073 samples for analysis. Each individual DNA sample was sonicated and barcoded with an in-house unique 8-mer oligonucleotide<sup>50</sup>. Groups of 95 barcoded DNA samples were pooled, and paired-end 100-bp sequencing was performed on one lane of a HiSeq sequencer, generating read groups of ~30 Gb of sequence per lane per pool.

**Alignment to mm10 reference and preprocessing of sequence data.** BWA version 0.5.6 (ref. 51) was used to align the reads from each read group to the mouse mm10 reference genome. The BWA alignments were refined with Stampy v1.0.21 (ref. 52) and converted into the BAM format by SAMtools v0.1.18-dev<sup>53</sup>. Library PCR duplicates were removed with SAMtools, and sequence reads were processed following a previously described pipeline<sup>12,15</sup>. All BAM files were processed through the Indel Realignment and Base Quality Score Recalibration steps of the best practices<sup>54</sup> recommended by GATK<sup>8</sup>. All preprocessing used GATK v2.4-9-g532efad. The option --rf BadCigar was applied to filter out reads that (i) had hard or soft clips in the middle of the CIGAR string, (ii) started or ended in deletions, (iii) were fully hard or soft clipped, or (iv) had consecutive indels in them. The option --rf BadMate was applied to filter out reads whose mate mapped to a different contig. Previously discovered indels from all mouse strains in MGP<sup>12,55</sup> were used as intervals for Indel Realignment in addition to those discovered in the 2,073 mice, and SNPs from the *Mus musculus domesticus* strains in MGP were used as known sites masked for Base Quality Score Recalibration.

**Variant calling from low-coverage sequencing data.** Variant calling was performed on all 2,073 BAM files using the GATK UnifiedGenotyper with thresholds -stand\_call\_conf 30 and -stand\_emit\_conf 30, as well as the following options for building variant quality recalibration tables: -A QualByDepth -A HaplotypeScore -A BaseQualityRankSumTest -A ReadPosRankSumTest -A MappingQualityRankSumTest -A RMSMappingQuality -A DepthOfCoverage -A FisherStrand -A HardyWeinberg -A HomopolymerRun.

Raw VCF files from the variant calling step for all chromosomes except the Y chromosome were pooled together for VQSR using the GATK VariantRecalibrator under SNP mode. Training, known and true sets for building the positive model are the SNPs that segregate among the classical laboratory strains of the MGP (2011 release REL-1211) on all chromosomes except the Y chromosome. Transversion (TsTv) ratios and recalibration tables were generated at 14 sensitivities (100.0, 99.9, 99.0, 97.0, 95.0, 90.0, 85.0, 80.0, 75.0, 70.0, 65.0, 60.0, 55.0 and 50.0) to training sets for runs of VQSR using different sets of annotations. A final set of annotations for VQSR and sensitivities to known sites was chosen to maximize TsTv at both known and novel sites to reduce the rate of false positive calls. A sensitivity of 97% to known sites was selected for a total of 8,597,879 SNPs (6,430,809 known and 2,177,070 new sites; TsTv ratio of 2.13 at known sites and 1.56 at new sites). We then further removed sites that were fixed alternative allele variants (and hence non-polymorphic in our study) or were multiallelic, leaving 7,073,398 biallelic SNPs (5,701,865 known and 1,371,533 new sites; TsTv ratio of 2.13 at known sites and 1.56 at new sites). The annotations used for VQSR were HaplotypeScore, BaseQualityRankSumTest, ReadPosRankSumTest, Mapping QualityRankSumTest, RMSMappingQuality, DepthOfCoverage, FisherStrand, HardyWeinberg and HomopolymerRun.

We used the 7 million biallelic SNPs in the mouse cohort for imputation, applying the method described below. To ensure the quality of the imputed SNPs used for downstream genetic analysis, we first extracted SNPs imputed with high certainty using IMPUTE2-style INFO scores. We observed by inspecting allele distributions that an INFO score greater than 0.4 indicated markers where the three genotype classes were clearly separable. Thus, we included only sites that met this criterion. We also discarded sites where more than 10% of mice had a maximum genotype probability smaller than 0.9, and on autosomal chromosomes we discarded sites where the  $P$  value for violation of Hardy–Weinberg equilibrium was smaller than  $1 \times 10^{-6}$ . This filtering resulted in a final set of 5.76 million SNPs that we used for genetic mapping. Lastly, we used the most current release of the Sanger mouse genomes database (2016, REL-1505, comprising 36 genomes, almost twice the original number) to refine the set of new SNPs. The number of new sites among the 5.76 million SNPs dropped from 799,133 (13.8%) to 152,671 (2.6%) (**Supplementary Table 3**). However, the TsTv ratios for the new SNPs remained little changed, at 1.74 and 1.73, respectively.

**Imputation.** We developed a new imputation algorithm, STITCH, described in a separate publication<sup>9</sup>. This employs a hidden Markov model (HMM) that extends the population genetic methods of Li and Stephens<sup>56</sup> and, more specifically, the fastPHASE algorithm of Scheet and Stephens<sup>57</sup>. We assume that the CFW population was founded with  $K$  unknown ancestral haplotypes and that the chromosomes of each sequenced CFW mouse are mosaics of the founder haplotypes. After some experimentation with different values of  $K$ , we found that  $K = 4$  was optimal (that is, the population was modeled as being founded from two individuals).

Simulating under the model (hidden ancestral states and sequencing reads) consists of (i) choosing initial state probabilities ( $\pi_k$ ) from one of the  $k$  haplotypes, (ii) choosing where to recombine between ancestral haplotypes assuming  $G = 100$  generations since the population's founding and a genetic distance between SNPs  $t$  and  $t+1$  ( $\sigma_t$ ), (iii) choosing the ancestry of each segment with respect to the frequency of each founder haplotype at that location ( $\alpha_{t,k}$ ), and (iv) sampling read locations, base qualities, underlying unobserved bases and observed sequenced bases, on the basis of the relative probability that ancestral haplotype  $k$  emits a reference or ancestral base at SNP  $t$  ( $\theta_{t,k}$ ). Together, these represent the parameters of the model  $\lambda = (\pi, \sigma, \alpha, \theta)$ .

To generate the probabilistic genotype of an individual CFW outbred mouse, we first calculate the probability of observing a particular sequencing read given membership in ancestral haplotype  $k$  as follows. We first remove

SNPs with low base quality (<17) and SNPs in reads with low mapping quality (<17). For an individual read  $R_r$  indexed by  $r$ , let  $J_r$  be the number of SNPs in the read and  $P(s_{r,j}|g_i=i) = \phi_{r,j}^i$  be the base-quality-scaled emission probability of sequencing read  $s_{r,j}$  given true underlying genotype  $i$ . Let SNP  $j$  in read  $R_r$  correspond to SNP  $u_{r,j}$ . We assume that the probability of a recombination in a read is low, so we assign each read as having been emitted from a central SNP  $t = c_r$ . Therefore, the probability of read  $R_r$  given that it comes from ancestral haplotype  $k$  is

$$P(R_r|q_t=k, \lambda) = \prod_{j=1}^{J_r} \left( \theta_{u_{r,j},k} \phi_{r,j}^1 + (1 - \theta_{u_{r,j},k}) \phi_{r,j}^0 \right)$$

and the probability of observing all reads at SNP  $t$  in a diploid sample given diploid hidden state at SNP  $t$  of  $q_t = (k_1, k_2)$  is

$$P(O_t|q_t=(k_1, k_2), \lambda) = \prod_{r:c_r=t} \left( \frac{1}{2} P(R_r|q_t=k_1, \lambda) + \frac{1}{2} P(R_r|q_t=k_2, \lambda) \right)$$

The full-chromosome diploid probability is then calculated using the initial, recombination and transition probabilities in the normal manner.

We ran the method for 40 expectation–maximization iterations, where in each iteration, during the expectation step, state probabilities are calculated for each mouse using the current parameters of the model and, in the maximization step, new initial, transition, recombination and emission parameters are estimated on the basis of state probabilities. Upon completion, haplotype and genotype probabilities, as well as dosages, are calculated. For example, the dosage of the number of alternative alleles is  $1 \times P(G = (\text{ref}, \text{alt})|O, \lambda) + 2 \times P(G = (\text{alt}, \text{alt})|O, \lambda)$  for a given mouse and SNP site.

**Selection of tagging SNPs.** We then identified a subset of 359,559 (353,697 autosomal) tagging SNPs with MAF >0.1% and LD  $r^2 < 0.98$  using PLINK (v1.07). Genotypes at these sites were called on the basis of maximum genotype probability from imputation. Genotypes were only called when the maximum genotype probability for a site was greater than 0.9; mice with a maximum genotype probability of less than 0.9 at a particular site would have a missing genotype at the site.

**Sample selection based on estimation of identity by descent between samples.** Pairwise IBD was estimated by calculation of pairwise identity by state (IBS) using PLINK (v1.07) at the tagging SNPs located on the autosomal chromosomes. Mice were excluded from further analysis if they had estimated PIHAT values of greater than 0.5 with at least one other mouse, percentage IBS = 1 of greater than 0.75 with at least one other mouse, or percentage of IBS = 0 of less than 0.25 with at least one other mouse. 135 mice were excluded by the above criteria.

**Sample selection based on principal-component analysis.** Linkage-disequilibrium-adjusted kinship (LDK; version 5.9)<sup>58</sup> was used to estimate local LD by calculation of local pairwise correlations between SNPs and generate weightings for each SNP in the calculation of a GRM adjusted for local LD. The GRM was generated using hard-called genotypes at tagging SNPs with MAF >5% from all autosomes. PCA was performed on the GRM to derive the top 20 principal components. PC2 separated out four mice from the rest; these four mice were excluded from further analysis.

**Estimation of whole-genome SNP-based heritability.** LDAK (version 5.9) was used to generate a new GRM using hard-called genotypes at the same tagging SNPs with MAF >5% in the mice remaining in the analysis. Restricted maximum likelihood (REML) was used to estimate heritability ( $h^2$ ) for each of the 200 phenotypes measured.

**QTL mapping.** We mapped QTLs at the tagging SNPs using purpose-written software in R. For each phenotype  $k$ , we used the quantile-normalized residuals  $y_k$  for QTL mapping and heritability analysis. Although we found little evidence of unequal degrees of relatedness among the CFW mice, as a precaution, we used mixed models to control for cryptic relatedness and to avoid false positive QTL calls. We first used the imputed dosages for tagging SNPs on

the autosomal chromosomes to compute genome-wide kinship matrices ( $K$ ). Thus, if  $a_{ip}$  is the imputed reference allele dosage of SNP  $p$  in individual  $i$ , then the genetic relationship  $K_{ij}$  between individuals  $i$  and  $j$  is defined to be the Pearson correlation coefficient of the vectors  $a_{ip}$  and  $a_{jp}$  across all autosomal tagging SNPs  $p$ . Element  $(i,j)$  of the population-wide GRM  $K$  is  $K_{ij}$ . We also computed leave-one-out kinship matrices  $K_c$  for each chromosome  $c$ , using all tagging SNPs not on chromosome  $c$ .

We modified the standard mixed-model formulation for mapping QTLs by computing separate mixed models for each chromosome, to ameliorate the reduction in statistical significance of a locus caused by the same information being present in the kinship matrix. To test association between phenotype  $k$  and tagging SNP  $p$  resident on chromosome  $c$ , we estimated the phenotypic covariance matrix  $V_{kc} = \sigma_{gkc}^2 K_c + \sigma_{ekc}^2 I$ , where the genetic and environmental variance components  $\sigma_{gkc}^2$  and  $\sigma_{ekc}^2$  are estimated as above, and factorized it into its square root using the eigen decomposition

$$V_{kc} = E'_{kc} \Lambda_{kc} E_{kc} = \left( E'_{kc} \Lambda_{kc}^{1/2} E_{kc} \right)^2 = A_{kc}^2$$

where  $E_{kc}$  is the orthogonal matrix of eigenvectors and  $\Lambda_{kc}$  is the diagonal matrix of eigenvalues of  $V_{kc}$ . Then, we fitted the transformed mixed model

$$z_{kc} = A_{kc}^{-1} y_k = \mu + \alpha (A_{kc}^{-1} a_p) + e$$

where  $\mu$  and  $\alpha$  are parameters to be estimated and the error vector  $e$  is uncorrelated so the model can be fit efficiently by computing the correlation coefficient of  $z_{kc}, A_{kc}^{-1} a_p$ .

Nominal statistical significance at a locus was measured as the log  $P$  value (the negative  $\log_{10}$  version of the  $P$  value from ANOVA comparing the fit of the allele model to the null model). We defined a candidate QTL as any locus where  $\log P$  was at a local maximum in comparison to values derived from tests at neighboring loci and no other locus within 3 Mb had a larger  $\log P$  value.

We estimated separate genome-wide significance thresholds for each phenotype, aiming to control the per-phenotype FDR. We generated  $Q = 100$  permutations of each transformed phenotype vector  $z_{kc}$ , keeping the transformed allele dosages fixed, and refitted the model. This is efficient because most of the computational effort in fitting a mixed model is reusable when fitting permuted phenotypes. We found candidate QTLs in the permuted data in the same way as in the real data and estimated the per-phenotype FDR of a QTL as

$$\text{FDR}_k(x) = \frac{P_k(x)}{QN_k(x)}$$

where  $N_k(x)$  and  $P_k(x)$  are the numbers of QTLs with  $\log P > x$  observed for phenotype  $k$  in the unpermuted and permuted data, respectively, and  $Q = 100$  is the number of permutations. Custom R code for QTL mapping, written specifically for this project, is available from R.M.

**Fine-mapping.** Once a QTL had been mapped using the tagging SNPs and exceeded the FDR threshold, association was recalculated with all imputed SNPs (from the set of 5.7 million) in a 20-Mb window centered on the peak using the same mixed model.

**Confidence interval estimation.** Confidence intervals were estimated by simulation. First, at each QTL, a residual phenotype was constructed by removing the effect of the top SNP at the QTL from the phenotype vector used in the QTL mapping above. This ablated the QTL while maintaining genetic contributions from elsewhere in the genome. Next, 1,000 SNPs were selected at random, subject to the constraint that they be within 2.5 Mb of the top SNP and polymorphic in the subset of individuals phenotyped for the trait (where the 95% confidence interval estimate was 2.0 Mb or greater, we repeated the analysis using SNPs up to 10 Mb from the top SNP). A causal variant was simulated at the SNP, with effect size matching that of the top SNP, taking into account allele frequency, and its trait value was added to the residual phenotype. A local scan of the region using the same mixed model but with the simulated phenotype was performed, and the location and  $\log P$  value of the top SNP were recorded. Across the 1,000 simulations, we estimated the distribution of the drop ( $\Delta$ ) in  $\log P$  between the simulated top SNP and the

simulated causal SNP (this was zero when the top and causal SNPs coincided). We used the fraction of simulations  $f(\Delta)$  within  $\Delta$  to determine confidence intervals for the original phenotype data. Thus, we identified the range of SNPs within 2.5 Mb of the top SNP and with a log  $P$  drop less than  $\Delta$  to define the 100 $f(\Delta)\%$  confidence interval for the QTL. We did this using both the tagging SNPs and the fine-mapping SNPs.

**Power calculation.** Because we applied an FDR approach to call QTLs, we did not use a log  $P$  threshold that would be required to determine power. However, to estimate power and the effects of sample size and effect size, we determined approximate genome-wide thresholds on the basis of permutations of the mixed-model transformed phenotypes  $z = A^{-1}y$ , keeping the genotypes fixed to preserve LD structure. For each of the 200 phenotypes, we performed 100 permutations and computed the genome-wide maximum log  $P$  value across the 359,559 tagging SNPs to define genome-wide thresholds  $T(p)$  at  $p = 0.5, 0.1$  and  $0.05$  levels of significance (for example, the threshold  $T(p)$  is such that in a fraction  $p$  of simulations the genome-wide maximum log  $P$  value exceeds  $T(p)$ ). Thresholds varied slightly between phenotypes, so we used the thresholds obtained by pooling all 20,000 simulations to estimate power for sample sizes  $N = 1,000, 1,732, 2,000$  and  $4,000$  and apparent effect sizes  $v = 0.01, 0.016$  and  $0.02$ . ( $N = 1,732$  and  $v = 0.016$  were the median sample size and effect size, respectively, in the current study.) Power  $\pi(N, v, T)$  to detect a QTL with effect size  $v$  and sample size  $N$  at genome-wide log  $P$  threshold  $T$  was computed as  $\pi(N, v, T) = \Pr(\chi < w(T) | \chi \sim \chi^2_{1, Nv})$ , where  $\chi^2_{1, Nv}$

is the non-central  $\chi^2$  distribution on 1 degree of freedom with non-centrality parameter  $Nv$  and  $w(T)$  is the quantile of a standard  $\chi^2$  distribution corresponding to  $\log P T$ , that is,  $\Pr(\chi > w(T) | \chi \sim \chi^2_{1, 0}) = 10^{-T}$ .

49. R Core Team. *R: A Language and Environment for Statistical Computing* 3.1.3 edn (R Foundation for Statistical Computing, 2015).
50. Lamble, S. *et al.* Improved workflows for high throughput library preparation using the transposome-based Nextera system. *BMC Biotechnol.* **13**, 104 (2013).
51. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
52. Lunter, G. & Goodson, M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* **21**, 936–939 (2011).
53. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
54. DePristo, M.A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
55. Yalcin, B. *et al.* Sequence-based characterization of structural variation in the mouse genome. *Nature* **477**, 326–329 (2011).
56. Li, N. & Stephens, M. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* **165**, 2213–2233 (2003).
57. Scheet, P. & Stephens, M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* **78**, 629–644 (2006).
58. Speed, D., Hemani, G., Johnson, M.R. & Balding, D.J. Improved heritability estimation from genome-wide SNPs. *Am. J. Hum. Genet.* **91**, 1011–1021 (2012).