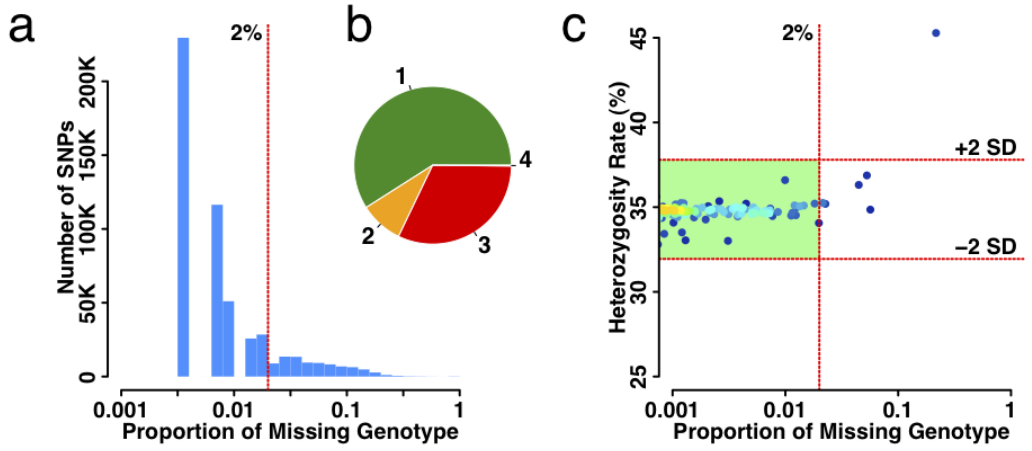

Figure S1. QC for pre-imputation genotypic data. Related to Figure 1. QC filters were applied at the SNP level first, followed by the samples level. **(a)** Histogram of missing data rates per SNP. The vertical red bar marks the 2% missing data rate, our threshold for SNP QC. **(b)** QC filters were sequentially applied to all SNPs probed by the genotyping array, yielding a total of 59% of the probed SNPs used for association studies (area 1, green). About 9% of SNPs were removed from the analysis because they featured a missing rate >2% (area 2, orange), a further 32% were removed because their minor allele frequencies were below 5% (area 3, red), and 0.1% displayed peculiar Hardy-Weinberg out-of-equilibrium (area 4, black). The pie chart shows the relative weights of the filters (data from Table S1a). **(c)** QC filters for samples were sequentially applied: samples with more than 2% missing genotypes were removed from the analysis, as well as any samples that deviated in their heterozygosity rate by more than two standard deviations. The green area highlights the samples that pass sample QC filters, marked by horizontal and vertical red bars.

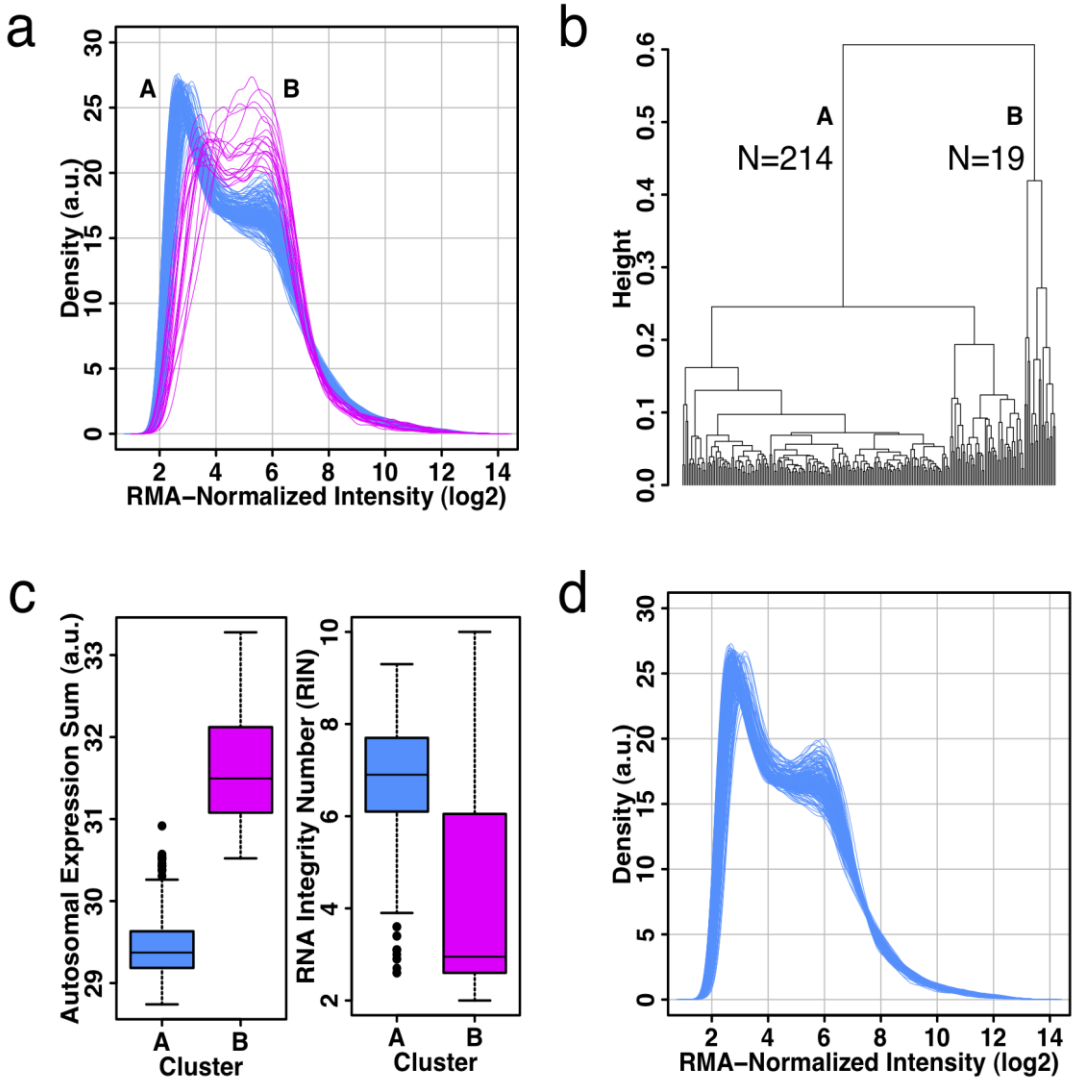

Figure S2. RNA QC via gene expression profiling. Related to Figure 1. **(a)** Gene expression profiles after an initial round of normalization. Samples are colored to highlight the effect of RNA degradation; in the profile B (magenta), the curves are shifted to the right compared to the profiles in A (blue), giving rise to an overall increase of gene expression for all genes in profile B. **(b)** A clustering of the gene expression profiles readily highlights the two distributions of profiles A and B. **(c)** Distribution of gene expression sum from autosomal chromosomes (P-value 2.3e-12) and RNA integrity number (RIN) within each cluster (P-value 3.1e-4). A lower RIN number is associated with lower RNA quality. **(d)** Gene expression profiles after the second round of RMA normalization, in which only the N=214 similar profiles have been processed altogether.

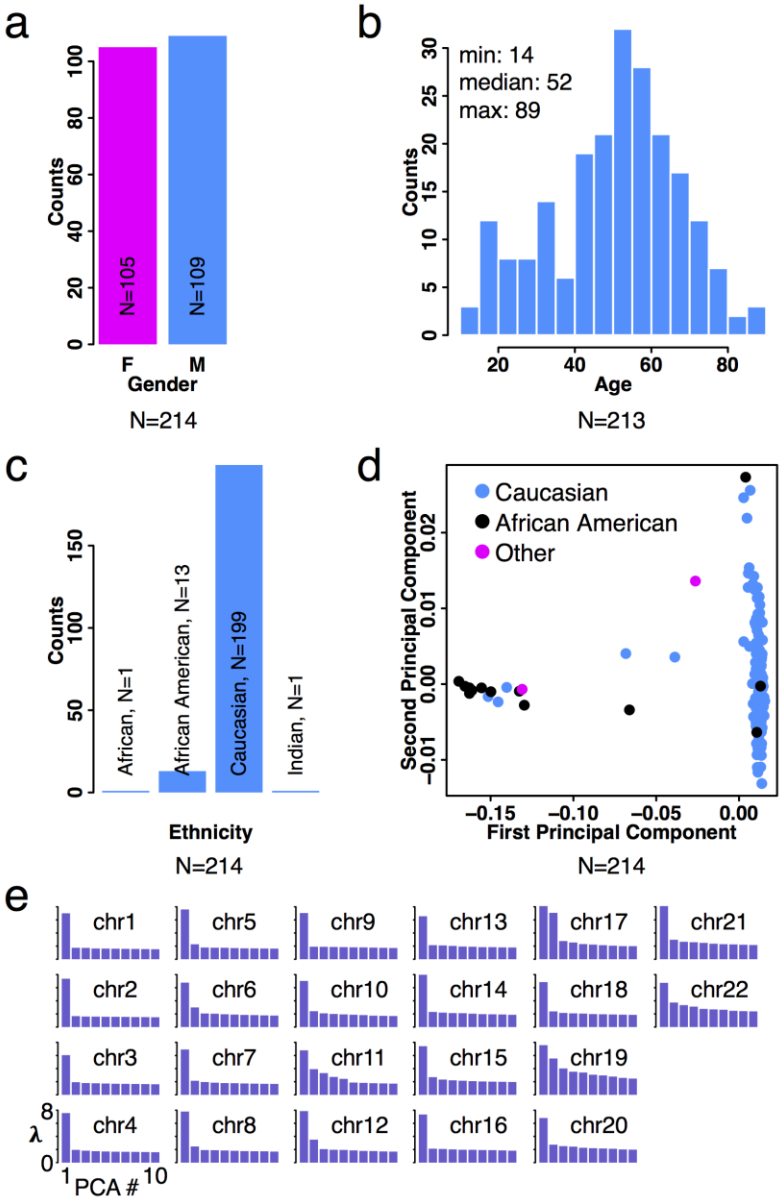

Figure S3. The DRG cohort comprises N=214 individuals with matching GWAS and RNA expression array datasets. Related to Figure 1. **(a)** Gender distribution. **(b)** Age distribution. One individual has an unknown age. **(c)** Reported ethnicity distribution. **(d)** Multi-dimensional scaling (MDS) plot, showing the first two racial principal components. **(e)** Scree plots showing eigenvalues for first ten genetic PCAs.

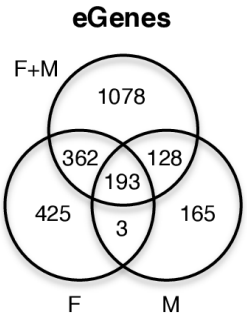

Figure S4. Three-way Venn diagram showing the extent of shared DRG eGenes across the whole cohort (M+F), the female-only cohort (F) and the male-only cohort (M). Related to Figure 1.

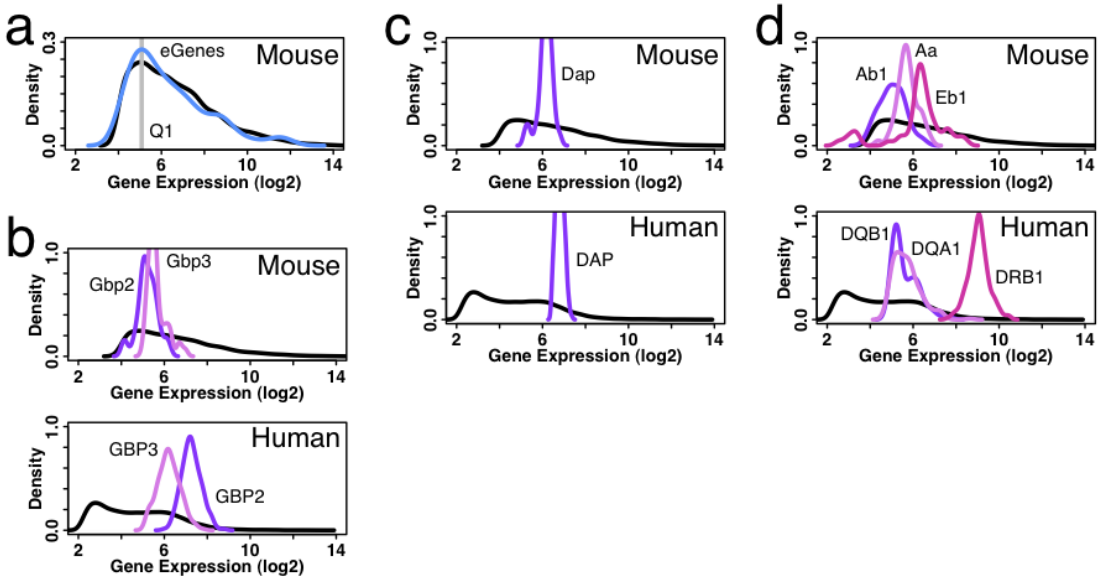

Figure S5. Comparative analysis of mouse and human DRG gene expression. Related to Figure 1. Mouse DRG expression data are from GEO set GSE65997. The black line indicates overall gene expression. Color coding indicates expression of orthologous genes between human and mouse. **(a)** Comparison of expression profile between all genes expressed in mouse DRG (black line) and mouse genes orthologous to human DRG eGenes (blue line). About 77% of these genes show expression levels above the 1st quartile (Q1). **(b)** Expression profiles of mouse and human *GBP* genes. **(c)** Expression profiles of mouse and human *DAP* gene. **(d)** Expression profiles of mouse H2 and human *HLA MHC* class II genes.

Supplemental Experimental Procedures

Description of the DRG cohort

The DRG cohort comprises samples that were collected post-mortem from brain-dead subjects following asystole. They consist of snap-frozen dorsal root ganglia (DRGs) sourced from bilateral L4 and L5. Samples are courtesy of the CORE/CORID repository at the University of Pittsburgh. 315 samples were originally collected; 214 with matched genotype/phenotype qualified for inclusion in this study.

Sample collection

Human Subjects: L4 and L5 DRGs were collected from organ donors with the consent of first-tier family members for the use of their loved one's tissue for research purposes. Detailed demographic and medical history data were collected for each donor. All procedures were approved by the University of Pittsburgh Committee for Oversight of Research and Clinical Training Involving Decedents, and the Center for Organ Recovery and Education, Pittsburgh, PA (http://www.core.org).

DRG Collection: Following the collection of tissue needed for donation purposes, L4 and L5 DRGs were accessed via a ventral approach. Briefly, the lumbosacral trunk was found running medially to the psoas major. Blunt dissection was used to follow the spinal nerves of the L4 and L5 ganglia to their respective foramen in the vertebral column. An oscillating autopsy saw (Mopec, Oak Park MI) was used to cut through the vertebral bodies above and below the L4 and L5 ganglia, respectively. Cuts were then made through the vertebral pedicals, with care to keep the saw blade angled above the ganglia. A Virchow skull breaker was then used to lift the ventral surface of the spinal column off en bloc, exposing the ganglia and cauda equina. Ganglia were carefully freed of connective tissue, the central and peripheral processes cut, and the ganglia placed in ice-cold collection media of the following composition: 129.5 mM NaCl, 5 mM KCl, 1.2 mM MgCl2, 30 mM HEPES, 2.5 mM glucose and 0.23% phosphoric acid. The pH of the collection media was adjusted to 7.35 with NaOH and filter sterilized for storage at 4 ^{o}C for no more than two weeks. With this approach, it was possible to recover all four ganglia in well under 15 min. The time between cross clamp and the start of the DRG isolation was usually between 45 min and 1 h. Two out of four ganglia were then snap-frozen for DNA and RNA extraction.

DNA & RNA preparation

Genomic DNA was isolated from frozen and homogenized DRG tissue using a QIAamp DNA Mini kit (Qiagen, Austin, TX) following the manufacturer's guidelines. DNA samples were stored at -20 °C until analysis.

RNA was isolated from human DRG samples frozen in TRIzol® reagent, thawed on ice and homogenized with TissueLyser (Qiagen, Austin, TX) using 3 mm sterile aluminum beads at 20.1 Hz for 3 min and 2 min in succession. After centrifuging at 12,000 rpm for 10 min at 4 °C to remove debris, the supernatant was passed 5–6 times through a 0.1 mm syringe-needle. The RNA was then phase separated using chloroform, precipitated with an equal volume of isopropanol, and washed with 75% ethanol following the manufacturer's instructions. Total RNA was suspended in 10 µL of 0.1% v/v diethylpyrocarbonate (DEPC)-treated water and stored at -80 °C.

DNA and RNA concentrations and purity were measured spectrophotometrically using a Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Lafayette, CO). Purity was evaluated as A260/280 and A260/230 ratios. High-quality RNA was confirmed using agarose gel electrophoresis and with an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). Integrity was assessed by running 100–200 ng of RNA sample in each lane of a 6000 Nano LabChip, using the Agilent Bioanalyzer 2100 system to obtain the RNA integrity values (RIN).

Genotyping and expression analysis

DNA genotyping: The Infinium OmniExpressExome-8 v1.2 Array requires 200 ng DNA, which was first denatured with 0.1N NaOH and then whole-genome amplified by incubation at 37 °C for 20 h in a proprietary amplification reaction mix. Amplified DNA was fragmented using vendor-supplied reagents (Illumina Inc., San Diego, CA) by incubation for 1 h at 37 °C. The fragmented DNA sample was precipitated and resuspended in hybridization buffer. DNA samples were denatured, applied to the Infinium arrays and hybridized for 16-24 h with rocking at 48 °C. Unhybridized and non-specifically hybridized DNA was washed away, the Beadchip surface was treated to facilitate the primer extension reaction, and single base extension was performed on bound primers with labeled nucleotides. Bound DNA sample was washed away and the array surface sealed to protect the dyes from atmospheric degradation. The genotyping was performed on Illumina's Infinium Human Omni Express Exome-8 v1.2 chip (~1M probes), at the University of Pittsburgh bioinformatics analysis core. The final array was scanned using an Illumina iScan, and the data analyzed using Illumina Genome Studio 2011.1 with Genotyping module 1.9.4 (Illumina Inc., San Diego, CA).

Total RNA levels from DRG cells were measured with an Affymetrix Human Transcriptome Array 2.0 (~70K gene-level probes, ~900K isoform-level probes), at the University of Pittsburgh bioinformatics analysis core. Microarray chips were scanned with the Affymetrix GeneChip® 3000 G7 Instrument System. There are number of advantages of the HTA_2 array: it boasts an advantageous sensitivity-to-cost ratio and fine exon mapping; it features on average 10 probes per probeset, making it more resilient to probe polymorphisms; and, it probes for an additional 13,525 transcripts compared to the last Affymetrix Array version. Thirty-four samples with a RIN lower than 2.5 were discarded, as they showed severe signs of RNA degradation (Schroeder et al., 2006). Normalization of the microarray data was done using R statistical software (R Core Team, 2014), with the help of the oligo (Carvalho and Irizarry, 2010) and the Platform Design Info for the Affymetrix HTA-2_0 (MacDonald, 2015). The Robust Multi-Array Average (RMA) algorithm was chosen for normalization (Irizarry et al., 2003), which, as a default post-processing step, performs quantile normalization of the probe intensities.

Sample quality control

Quality assessment and control (QC) was performed as suggested by Zondervan and colleagues (Anderson et al., 2010). We performed QC on a per-marker basis first, followed by QC on a per-individual basis. Data QC and management were done using PLINK computer software (Purcell et al., 2007) versions v1.07 64-bit (10 Aug 2009) and v1.90b3f 64-bit (2 Mar 2015).

At baseline a total of 964,058 markers were considered (Figure S1). QC filters were applied for genotyping rate $\geq$ 98%, minor allele frequency $\geq$ 5%, and deviation from Hardy-Weinberg equilibrium P-value $\leq 10^{-5}$. About 9% of markers displayed missing rates > 2% (Figure S1a, Figure S1b, Table S1a), 32% were at < 5% minor allele frequency (Figure S1b, Table S1a), and about 1% of the markers were out of Hardy-Weinberg equilibrium; all were removed from the analysis (Figure S1b, Table S1a). Thus, a total of 568,849 markers passed QC, which represents 59% of the originally genotyped markers (Figure S1b, Table S1a).

Three hundred and fifteen DNA samples were available at baseline, including internal replicates. Twenty-two DNA samples were removed from the analysis because they did not have matched RNA samples. Forty-three RNA samples were removed because of poor RNA quality assessed either by normalization of gene expression (N=19), or by RIN (N=34). Their corresponding DNA samples were also removed from the analysis. Using established QC thresholds, 8 samples were removed from the analysis because of missing genotyping rate > 2%, and 5 samples were removed because of their departure from mean heterozygosity rate by more than two standard deviations (Figure S1c). Identity by state (IBS) detected all 13 internal replicates (proportion Identity-by-Descent > 0.99), and one other pair of first-degree relatives (proportion Identity-by-Descent of 0.5075). For replicates, we retained from the pair the sample that displayed the highest genotyping rate. Finally, 3 samples were removed from the analysis because of discordant gender assignment between gender estimated from the genotype (i.e., from X chromosome

homozygosity rates), and that estimated from RNA expression pattern based on the top 10 genes that display the most marked female/male differences.

Next, QC for RNA expression levels was applied (Figure S2). Gene expression histograms demonstrated that some samples exhibited much lower gene expression intensity profiles than the majority of the samples (Figure S2a). An average-merge clustering of the expression intensity profiles identified two main clusters (Figure S2b). Comparing the total gene expression intensity between the two clusters showed that cluster B samples had significantly higher gene expression sum than samples in cluster A (Figure S2c, left) (Mann–Whitney U test P-value 2.3×10^{-12}). We hypothesized that RNA degradation could cause such a difference, as shorter RNA fragments lose specificity for microarray probe hybridization. Comparing the RIN numbers for samples in clusters A versus B (Figure S2c, right) we opted to remove the samples in cluster B from the analysis (Mann–Whitney U test P-value 3.1×10^{-4}). After removing the samples of cluster B, we re-ran RMA-normalization using only samples in cluster A. By default, the RMA algorithm performs quantile normalization on probe intensities. All remaining samples displayed similar RNA gene expression profiles (Figure S2d). Of these, 14, 24 and 44 samples have RIN numbers below 4, 5 and 6, respectively. Although these samples seemed to have low RIN numbers, they passed our two rounds of QC, generated robust hybridization results, and clustered with high RIN number samples, and thus we kept these samples in the analysis to increase statistical power. It has been showed previously that RIN number shows good correlation with RNA quality even though it is not an absolute indicator of it (Brisco and Morley, 2012). After normalization of the arrays, only the probes with RMA-normalized intensities ≥ 4 were used. After applying QC steps as described above, we analyzed 214 RNA samples tested for 36.5K gene-level and 386K exon-level probes. In the eQTL analysis we used sample average expression as a covariate.

Although we did collect information about painful diseases such as diabetes or osteoarthritis, based on the medical record it was impractical to collect any details on pain intensity, location, duration or qualities. Therefore, we did not stratify our cohort based on the presence of the diagnosis of any chronic condition that has variable rates of painful components, mixed pathophysiology and/or non-controlled reliability.

Imputation

We leveraged the 1,000 Genomes Project data (Abecasis et al., 2012) to infer allelic content at genomic loci not probed by the genotyping array. Before pre-phasing we re-introduced the SNPs that were discarded based on minor allele frequency, as even though they display low minor allelic content they have been measured by the genotyping array, and thus could prove important to resolve ambiguities in haplotype assignment.

Before the phasing step, we kept the site featuring the highest genotyping rate among all genotyping probes for the same site, and we fixed strandedness based on a published resource (http://www.well.ox.ac.uk/~wrayner/strand/), with the corresponding hg19 coordinates of our genotyping array data: HumanOmniExpressExome-8v1-2_A-b37-strand.zip.

Phasing was performed using the SHAPEIT computer program version v2.r790 (Delaneau et al., 2012; O'Connell et al., 2014) with phase 3 data of the 1,000 Genomes Project (https://mathgen.stats.ox.ac.uk/impute/). We first used SHAPEIT in the "check" mode to uncover problematic SNPs (strand flips, alleles mismatch, monomorphic sites, etc.), then in the "phase" mode while excluding these SNPs.

An imputation was performed using the impute2 computer program version 2.3.2 (Howie et al., 2012; Marchini et al., 2007), also with the 1,000 Genomes Project phase 3 data, and the output of SHAPEIT. The whole genome was imputed using 1-Mb windows. Imputation statistics can be found in Table S1b. Imputed SNPs tables were re-imported back to PLINK, with a final QC run checking for a minimum 98% genotyping rate, a minimum 5% minor allele frequency, and minimum P-value for Hardy-Weinberg distribution of 10^{-5} using: --geno 0.02 --maf 0.05 --hwe 0.00001.

Successful imputation of the genotype lead to a total of 81.65M SNPs. SNPs with $r^2 \geq 0.8$ were grouped for a total of 0.49M independent positions, while SNPs with $r^2 \geq 0.5$ were grouped for a total of 0.18M independent positions. For eQTL discovery, we tested all 4.88M SNPs for association with mRNA gene levels, but only reported the SNP with the lowest association P value within $r^2 \geq 0.5$ a given gene.

Genome-wide association study

Gene- and exon-level probe RMA-normalized intensities of at least 4 were required for inclusion in the analysis, leaving 36,552 gene-level (52% of probesets) and 386,507 exon-level (42% of probesets) expression data points. We considered an eQTL to be *cis*-acting when the distance separating the eQTL and the transcription start site (TSS) of the associated gene was ≤ 1M nucleotides (GTEx_Consortium, 2015; Ramasamy et al., 2014); eQTLs were characterized as *trans*-acting otherwise. For exons, *cis*-acting eQTLs were those found within 5K nucleotides of the exon-intron boundaries (GTEx_Consortium, 2015), and *trans*-acting otherwise. All association tests were performed using the Matrix eQTL computerized method (Shabalin, 2012), with knowledge of *cis*- and *trans*- boundaries. All statistical tests were performed with age, gender, average sample expression and the genotype's first two racial eigenvectors per chromosome as covariates. Using the first two racial eigenvectors as covariates controlled for genetic racial effects and allowed us to use a racially mixed population.

Correction for multiple testing was achieved using the Benjamini-Hochberg procedure, considering false discovery rates (FDR) of 5% for *cis*-acting, and 1% for *trans*-acting eQTLs. Data management was facilitated with the use of PLINK (Purcell et al., 2007). Descriptive statistics for eQTL associations are compiled in Table S1c. Estimates for the number of performed statistical tests while accounting for linkage-disequilibrium (LD) were obtained through the size of the Cartesian product of the number of independent SNPs (with $r2 \geq 0.5$ dependence) and number of gene- or exon-level probes. The estimated number of independent statistical tests was 4.7M for gene-level and 1M for exon-level *cis*-acting eQTLs, and 6.6M for gene-level and 70M for exon-level *trans*-acting eQTLs. P-value thresholds for statistical significance were obtained via the "fdr" setting of the "p.adjust" function in the R statistical package (R Core Team, 2014), and provided the following p-values: 2.98×10^{-5} for gene-level and 8.68×10^{-4} for exon-level *cis*-acting eQTLs at FDR of 5%, and 4.56×10^{-9} for gene-level and 1.71×10^{-9} for exon-level *trans*-acting eQTLs at FDR 1%. Given the limited statistical power it was difficult to assess which one of the SNPs in a haploblock is the functional SNP acting as an eQTL. Here, to simplify the analysis, from the pool of SNPs in a strong LD ($r2 \geq 0.5$), we retained the SNP that showed the lowest P value of association with the gene/exon expression level.

A total of 214 samples had matched genotype/phenotype data that passed established GWAS-level QC and assessment (Figure S3). These 214 samples consisted of 105 females and 109 males (Figure S3a), with a median age of 52 years old (age was not reported for one individual) (Figure S3b), and represented >90% of Caucasian-ancestry participants (Figure S3c). With few exceptions, declared race matched that of a 2-D projection of a multi-dimensional scaling (MDS) plot (Figure S3d). Factor analysis indicated that for most chromosomes the first PCA carries almost all the load, and thus our usage of the first two racial PCAs is sufficient and appropriate (Figure S3e).

eQTL QC

As QC for our eQTLs dataset, we compared our list of DRG eGenes with a list of 33 "master" eGenes with strong genetic influences across tissues observed in >70% of all extant eQTL datasets (Zhang et al., 2014). These genes represent housekeeping or master *cis*-eGenes to use as positive controls in eQTL studies. Twenty-three of our DRG eGenes overlapped with the 33 master eGenes ($P<0.05$ per gene, binomial test P-value for an overlap of 23 genes is $P=2.1 \times 10^{-4}$, Table S3), which is in line with expected master eGene frequency in any given tissue, providing confidence in the reliability of our identified eQTL signals. Interestingly, glutathione S-transferase mu 3 (*GSTM3*) is one of the DRG housekeeping eGenes,

and is a main contributor to glutathione derivative biosynthetic processes, consistent with our pathway eGene analysis.

Background SNP distribution

eQTLs were compared against a background SNP distribution in which matched sets of random SNPs provide expected counts. Set matching ensured that the list of random SNPs had similar minor allelic frequencies and gene densities to those observed. For each of our list of eQTLs (*cis*-acting, gene-|exon-level) we used as background 10 randomly generated matched sets. This was done using SNPsnap web server (Pers et al., 2015), which provided a stringent control in generating lists of SNPs with similar genomic features as the original.

To estimate the statistical significance of the distribution of 55% of cis-eQTLs (N=1055) around 100 Kb from the TSS (Figure 2a), we employed statistical simulation. We performed 70,000 permutations in which we randomly chose SNPs in a fashion proportional to that found for *cis* eQTLs (i.e. 169 SNPs on chr1, 131 on chr2, etc.) while keeping fixed the pool of eGenes. These simulations allowed us to estimate the average (and standard deviation) number of random SNPs within 100 Kb from any TSS of the eGenes, with an implied supposition that the random SNPs are eQTLs for the closest eGenes: 343 (+/- 33). The estimated P-value for the observed 1055 eQTLs within 100 Kb to eGene TSS is 3.1×10^{-102} (Z-score 21.4).

Identification of functional genomic regions

Characterization of genetic locations of identified *cis*-eQTLs was performed using the Combined Annotation Dependent Depletion (CADD) database (Kircher et al., 2014), which offers a rich set of annotated features for SNPs of interest. It contains 11 different genetic locations and associated consequences: 3'UTR, 5'UTR, downstream, intergenic, intronic, non-synonymous, synonymous, regulatory, splice site, upstream and non-coding change. We compiled and compared counts for each consequence category for DRG *cis*-eQTLs.

Polymorphism-in-probe

The presence of polymorphisms in the hybridization probes can lead to false positives and false negatives for eQTL discovery using microarray technology (Alberts et al., 2007; Benovoy et al., 2008; Ramasamy et al., 2013). Here, we utilized Affymetrix's second generation and most recent microarray: the Human Transcriptome Array 2.0. The array is designed with a median of ten 25-mer probes per exon. The previous array generation used four probes per exon. Hence, the HTA 2.0 array is much more resilient in face of the polymorphism-in-probe problem.

Sex-specific DRG eQTLs

We characterized the sex-specific composition of eQTLs in DRG (Figure S4, Tables S1e-f). We found that women feature about twice as many eGenes than men (983 vs. 489), even though our cohort is sex-balanced (Table S1b). Importantly, however, most sex-specific eGenes also overlap with those of the entire cohort: 55% and 65% for women and men, respectively.

DRG eQTLs for SNPs associated with pain phenotypes

Human Pain Genetics Database

We made use of the Human Pain Genetics Lab database (URL: http://diatchenko.lab.mcgill.ca/hpgdb/) that offers a hand-curated literature survey of genetic results related to human pain phenotypes. Association tests were performed in the original DRG dataset to uncover any gene/exon whose expression levels would be associated with the allelic content of these SNPs. The statistical threshold for significance was set using the Benjamini-Hochberg correction for multiple testing, with a false discovery rate of 5%.

The OPPERA study

The Orofacial Pain: Prospective Evaluation and Risk Assessment (OPPERA) (Maixner et al., 2011) is a large-scale study that aims to identify the psychological and physiological risk factors, clinical characteristics, and associated genetic mechanisms that influence the development of TMD and other chronic pain conditions (Maixner et al., 2011). Recently, a GWAS was completed for the OPPERA cohort (dbGaP Study Accession: phs000762.v1.p1). Individuals were recruited from four demographically diverse U.S. locations (Buffalo, NY, Florida, Baltimore, MD and North Carolina), for a total of 3104 unique samples that passed standardized genotyping QC. Samples were genotyped using the HumanOmni2.5Exome-8v1A and imputed using IMPUTE2, for a total of 34M SNPs. Over 200 pain-related phenotypes were collected within the study design. For the current analysis, two chronic pain conditions were selected as variables in a logistic regression model: chronic TMD cases and chronic LBP cases. For experimental pain modalities, results from quantitative sensory testing—including heat pain threshold and tolerance (temperatures ranging from 32 to 52 °C), mechanical pain using pinprick at 512 kPa, and pressure pain threshold on the epicondyle muscles—were selected. The motivation for selection of these pain modalities springs from the fact that nerve endings of DRG neurons have a variety of sensory receptors that are activated by many stimulus modalities. Quantitative sensory testing in the OPPERA cohort covered three of these sensory modalities: heat pain, mechanical cutaneous pain and pressure pain. These modalities were judged important since signal transduction is relayed through A-delta and C fibers whose cell bodies are situated in the DRG. Age, gender, recruitment site, and the first three eigenvectors for race were used as covariates in association testing. A quantile-quantile (QQ) plot of the distribution of observed $-\log_{10}$(P-values) from tests of association was generated (Figure 6, left column). We then mapped identified DRG eQTLs within the OPPERA dataset, and generated a QQ plot of associations of DRG eQTL with each phenotype. Enrichment of DRG eQTLs in the GWAS was assessed using a Kolmogorov-Smirnov test between the P-values of the GWAS and P-values of the GWAS for SNPs that are eQTLs in the DRG. Tissue-specific eQTL contributions to pain phenotypes were measured from the QQ plot; since this plot tracks the observed P-value versus that what is expected from a null distribution, the ratio of observed to expected is a good measure of "impact" or "contribution" (Figure 6, right column, QQ bar plot). The contribution of eQTLs from a tissue is arbitrarily defined as the sum of log2 of observed to expected ratios for the lowest 50 P-valued eQTLs from that tissue. Given that typically only a handful of SNPs are associated strongly to a phenotype of interest, and that different tissues may display different number of eQTLs, we feel that considering the top SNPs is a fair approach to quantify tissue contribution. Over-representation of eQTLs in the top 1, 10 and 100% of GWAS hits can also stand as a proxy for tissue-specific eQTL contribution to phenotype (Figure 6, right column, GWAS curves). The over-representation or enrichment is computed as the log2 of observed q/n ratio to the expected Q/N ratio, where q is the observed number of tissue-specific eQTLs in the top n GWAS hits, out of a total Q tissue-specific eQTLs in the total N GWAS hits. The horizontal grey line shows the threshold of no enrichment (i.e. O ~ E, or log2(1)=0).

To identify DRG eQTLs that would potentially interpret association results with pain phenotypes, we searched for alignment of the DRG eQTL signals with that of GWAS associations. We first tried SMR (Zhu et al., 2016) to no avail, likely due to limited sample size of both GWAS and eQTL studies. We then tried COLOC (Giambartolomei et al., 2014) also to no avail, as it seems to deal poorly with unbalanced association strength between the GWAS and eQTL studies. Selection of SNPs that are simultaneously

strongly associated to both the GWAS and eQTL analyses was done in the following way. First, we transformed P-values into ranks by sorting SNPs by increased P-value then assigning rank 1 to the lowest P-value, rank 2 to the second lowest P-value, etc. Then, the ranks were also transformed in such a way that rank 1 corresponded to a score of 1 and the lowest rank corresponded to a score of 0. In other words, the higher the score, the lower the rank, and the stronger the SNP is associated in the GWAS. The advantage of this score-based system is that it allows for comparison of P-values of varying orders of magnitude: eQTL GWASs often display very small P-values up to 10^{-80}, whereas phenotypic GWASs seldom go below $5x10^{-8}$. To quantitatively assess simultaneous association of a SNP with a phenotypic GWAS and eQTL, we computed a combined score for each given SNP from the scores of the SNP in its respective associations: score = score_GWAS * score_eQTL (Table S6). Ideally, a perfect SNP would have a combined score of 1, stemming from the fact that it would have a perfect score in both the phenotypic GWAS and the eQTL study; 1 = 1*1. In this way, we can estimate which DRG eQTLs for which genes may contribute most to a phenotype of interest. Furthermore, the score-based system can be used as a substitute ordinate coordinate for Manhattan plots (Figure 7). Statistical significance for the difference between two distributions of data (P-values, CADD scores, etc.) was assessed using a 1-sided Mann-Whitney U-test, unless otherwise specified.

Replication in the UK BioBank

We used the UK BioBank (UKBB) to replicate our findings for the MHC II locus. In brief, the UKBB is a prospective study of 150,000 people aged 40–65 recruited in the United Kingdom. For replication of OPPERA LBP results, we used UKBB's self-reported chronic back pain phenotype. Subjects who reported back pain at 3 months (CBP; field 3571.0) were defined as cases. The remaining subjects were defined as controls, except for those reporting back pain at 1 month (field 6159.0) but not at 3 months (27,083 cases / 112,319 controls). For replication of OPPERA TMD results, we used UKBB's self-reported chronic facial pain. Subjects who reported facial pain at 3 months (CFP; field 4067.0) were defined as cases. The remaining subjects were defined as controls, except those reporting: facial pain at 1 month (field 6159.0) but not at 3 months, and headaches at 3 months (field 3799.0) (1,326 cases / 136,271 controls). The analysis was limited to Caucasians. Covariates were age, gender, and genotyping platform (UK BiLeve and UKB Axiom). We considered SNPs with minor allele frequencies of at least 1%. P-values for logistic case/control association were generated by snptest (Marchini and Howie, 2010) (version 2.5.2) and collected for the MHC II locus. SNPs in each phenotype (CBP or CFP) were partitioned in two sets; those that show statistically significant scores in OPPERA's discovery cohort (SS; FDR 1%), and those that do not (NS). The distributions of P-values for SS and NS were plotted as empirical cumulative distribution functions, and the Kolmogorov-Smirnov test was used to assess if the SS set is enriched in the replication UKBB cohort with lower P-values than that of the NS set.

Comparison with mouse data

We used mouse microarray gene expression data from the GEO set GSE65997 (Wieskopf et al., 2015) (Figure S5), and from RNA-Seq gene expression dataset of single cells classified as neuronal or non-neuronal (Usoskin et al., 2015). Since these datasets feature many samples, we considered the expression of a gene corresponding to the 3rd quartile across all samples (i.e., lowest expression level among the 25% most expressed genes). A gene was declared "expressed" if the 3rd quartile expression level was found to be above "floor" expression; for microarray, we choose as the floor the 1st quartile of overall expression across all genes, whereas for RNA-Seq zero was chosen as the floor.

Animal behavioral pain model

All procedures were approved by the Downtown Animal Care Committee at McGill University, and conformed to ethical guidelines of the Canadian Council on Animal Care. Eight homozygous 10-week-old mice (4 male; 4 female) with a deletion targeting the region of the *H2* locus encoding all genes for MHC class II in mice (Madsen et al., 1999) (*MHCII*$^{-/-}$) were purchased from The Jackson Laboratory (stock #003584; B6.129S2-H2$^{dlAb1\text{-}Ea}$/J) where they have been backcrossed to C56BL/6J for 13 generations. Age-matched C57BL/6J mice (3 male; 4 female) were also purchased from the same vendor. All mice were

housed in standard polycarbonate cages in groups of three or four same-sex, same-genotype mice in a temperature-controlled (20° ± 1°C) environment (14:10-hour light/dark cycle; lights on at 07:00 h), with tap water and food (Harlan Teklad 8604) available *ad libitum*. Experimenters were blinded to genotype throughout the experiment.

All experiments took place during the light cycle, no earlier than 09:00 h and no later than 16:00 h. Mice were placed in custom-made Plexiglas cubicles (5 × 8.5 × 6 cm) on a perforated metal floor, and were habituated for at least 1 h before testing. The up-down method of Dixon was used to estimate 50% withdrawal thresholds to the application of calibrated von Frey nylon monofilaments (Chaplan et al., 1994) (Stoelting Touch Test). Filaments were applied to the plantar surface of the hind paw for 3 s and responses were recorded. Two consecutive measures were taken on each hind paw at each time point, and averaged. Although data were collected on both hind paws, only data from the hind paw injected with CFA (see below) are presented, as no significant genotype effects on the contralateral paw were observed.

All mice received 20-µl unilateral injections of 50% complete Freund's adjuvant (CFA, Sigma) into the plantar surface of the left hind paw on day 0. Mice were tested for mechanical sensitivity using von Frey fibers before at baseline (BL; one day before CFA injection) and 3, 7, 10, 14 and 21 days' post-injection to quantify mechanical allodynia.

Statistical analyses were conducted using an α level of 0.05. Data were analyzed by repeated-measures two-way ANOVA followed by a posthoc Sidak's test with reported *P* values adjusted for multiple comparisons, and graphed with GraphPad Prism 6.0.

Statistical Tests

In this section, we'll describe how statistical tests were performed and which data were used to arrive at statistical conclusions.

Table S1c, "eQTL discovery":
Estimated total number of independent statistical tests considering LD as well as FDR thresholds for statistical significance is reported and summarized in Table S1c.

Figure 1e, Table S2, "Global characterization of DRG eQTLs":
The Pathway Studio web program was used to perform Gene Set Enrichment Analyses. Pathways were deemed statistically significant only if they passed correction for multiple testing using the Bonferroni method. The number of statistical tests performed was equal to the number of pathways returned by Pathway Studio.

Figure 1e, Table S2, "Global characterization of DRG eQTLs":
The total number of master eGenes is 33, and thus we used the Bonferroni threshold for a statistical significance of $0.05 / 33 \sim 1.5 \times 10^{-3}$. Any one of these master eGenes found associated with a P-value lower than 1.5×10^{-3} in our DRG eQTL cohort was considered a replication of that eGene in the DRG tissue. We found a total of 23 eGenes replicated in the DRG out of a possible 33. To estimate the likelihood of this finding, we considered that there are 13,704 genes in our study with P-value for association lower than 1.5×10^{-3} out of a total of 36,552 genes. Then, the P-value of obtaining 23 successes out of 33 trials in a background probability of success of 13,704/36,552 is given by a binomial test for enrichment: binom.test(23,33, p=13704/36552)$p.value, or 2.1×10^{-4}.

Figure 2c, "Functional characterization of the eQTLs in DRGs":
A total of 8 categories of gene regions showed observed counts greater or equal to 10, and thus the Bonferroni threshold of statistical significance was $0.05 / 8 \sim 6.25 \times 10^{-3}$. Any binomial test for enrichment for a P-value lower than this was considered statistically significant. Total observed count "t" was 1029. Background probabilities "p" were from SNPsnap results. Expected count were considered to be t*p, enrichment ratio = observed/expected = s/(t*p). A summary of binomial tests is presented here:

Category	Binom.test(s,t,p)$p.value	P-value	Signif?
3PRIME_UTR	s=19,t=1029,p=0.008947593	3.94×10^{-3}	Y
NON_SYNONYMOUS	s=14,t=1029,p=0.004260758	1.86×10^{-4}	Y
DOWNSTREAM	s=68,t=1029,p=0.134043460	3.58×10^{-12}	Y
NONCODING_CHANGE	s=14,t=1029,p=0.019429058	2.13×10^{-1}	N
INTERGENIC	s=238,t=1029,p=0.165402642	5.56×10^{-8}	Y
REGULATORY	s=268,t=1029,p=0.087430763	1.04×10^{-59}	Y
INTRONIC	s=309,t=1029,p=0.419855134	2.74×10^{-15}	Y
UPSTREAM	s=99,t=1029,p=0.160630592	2.43×10^{-9}	Y

Figure 2d, "Functional characterization of the eQTLs in DRGs":
A total of 11 categories showed observed counts greater or equal to 10, and thus the Bonferroni threshold of statistical significance was $0.05 / 11 \sim 4.55 \times 10^{-3}$. Total observed count "t" was 6612. Background probabilities "p" were from SNPsnap results. Expected count was considered t*p, enrichment ratio = observed/expected = s/(t*p). A summary of binomial tests is presented here:

Category	Binom.test(s,t,p)$p.value	P-value	Signif?
3PRIME_UTR	s=377,t=6612,p=0.008676434	3.20×10^{-175}	Y
INTRONIC	s=2682,t=6612,p=0.474584595	1.80×10^{-29}	Y
SPLICE_SITE	s=38,t=6612,p=0.001145944	2.99×10^{-15}	Y
5PRIME_UTR	s=122,t=6612,p=0.003765245	2.04×10^{-44}	Y
NON_SYNONYMOUS	s=382,t=6612,p=0.003274126	~0 (very small)	Y
SYNONYMOUS	s=377,t=6612,p=0.003519686	6.18×10^{-309}	Y
DOWNSTREAM	s=758,t=6612,p=0.136694770	1.04×10^{-7}	Y
NONCODING_CHANGE	s=93,t=6612,p=0.014078743	1.00	N
UPSTREAM	s=680,t=6612,p=0.144552673	8.37×10^{-24}	Y
INTERGENIC	s=70,t=6612,p=0.126626831	1.07×10^{-280}	Y
REGULATORY	s=1033,t=6612,p=0.083080953	4.49×10^{-84}	Y

Figures 3a, 3c, "Functional characterization of eGenes in DRGs":

We first performed an analysis at gene-level. From a survey of the microarray definition files, we found 17,345 coding genes and 19,207 non-coding genes for a grand total of 36,552 genes genome-wide whose average expression is above 4, a nominal expression value. We observed 947 coding eGenes and 814 non-coding eGenes, for a grand total of 1,761 eGenes. Thus, we obtained P-value from a binomial test for enrichment as follow:
* Coding: binom.test(947, 1761, p=(17345 / 36552))$p.value = 1.14×10^{-7}
* Non-coding: binom.test(814, 1761, p=(19207 / 36552))$p.value = 1.14×10^{-7}
Both values were found to be statistically significant since they were lower than Bonferroni threshold of statistical significance for two tests: 0.05 / 2 = 0.025.
The non-coding genes breakup is summarized in the following table. We observed a total of 779 non-coding eGenes out of a genome-wide total of 18,067 genes in 7 non-coding categories. Since we tracked 7 categories, the Bonferroni threshold of statistical significance was 0.05 / 7 ~ 7.1×10^{-3}. Genome-wide counts for each category are also shown, and make up the background "p" for expected probabilities. Expected count was t*p, enrichment ratio = observed/expected = s/(t*p). A summary of binomial tests is presented here:

Category	Binom.test(s,t,p)$p.value	P-value	Signif?
lincRNA	s=316,t=779,p=6959/18067	2.39×10^{-1}	N
noncoding	s=174,t=779,p=3352/18067	7.50×10^{-3}	N
antisense	s=117,t=779,p=2187/18067	1.55×10^{-2}	N
pseudogene	s=102,t=779,p=3040/18067	4.69×10^{-3}	Y
piRNA	s=25,t=779,p=1140/18067	1.57×10^{-4}	Y
snoRNA	s=23,t=779,p=552/18067	1	N
miRNA	s=22,t=779,p=837/18067	1.33×10^{-2}	N

Figures 3b, 3d, "Functional characterization of eGenes in DRGs":

At exon-level, we observed a total of 4,568 eExons in coding genes and 1,332 eExons in non-coding genes. After eExons from the genes already associated as eGenes were removed, we obtain 3,640 eExons in coding genes and 832 eExons in non-coding genes. Next, we asked which gene region the eExons belong to. For background estimates we sampled randomly 10,000 exons from the microarray definition files and tracked down the gene region they belong to (8,857 in the following 8 tracked regions). Then we compared with 5,214 strongly associated eExons; the Bonferroni threshold for statistical significance was 0.05 / 8 ~ 6.25×10^{-3}. Expected count was t*p, enrichment ratio = observed/expected = s/(t*p). A summary of binomial tests is presented here:

Category	Binom.test(s,t,p)$p.value	P-value	Signif?
utr_5p	s=115,t=5214,p=144/8857	1.47×10^{-3}	Y
cds_beg	s=230,t=5214,p=349/8857	8.11×10^{-2}	N
utr_3p	s=535,t=5214,p=544/8857	5.12×10^{-30}	Y
cds_end	s=535,t=5214,p=739/8857	1.16×10^{-6}	Y
tss	s=398,t=5214,p=855/8857	3.98×10^{-7}	Y
ncrna	s=554,t=5214,p=944/8857	9.64×10^{-1}	N
intron	s=848,t=5214,p=1432/8857	8.51×10^{-1}	N
cds	s=1999,t=5214,p=3850/8857	6.03×10^{-14}	Y

Supplemental References

Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56-65.

Alberts, R., Terpstra, P., Li, Y., Breitling, R., Nap, J.P., and Jansen, R.C. (2007). Sequence polymorphisms cause many false cis eQTLs. PLoS One *2*, e622.

Anderson, C.A., Pettersson, F.H., Clarke, G.M., Cardon, L.R., Morris, A.P., and Zondervan, K.T. (2010). Data quality control in genetic case-control association studies. Nat Protoc *5*, 1564-1573.

Benovoy, D., Kwan, T., and Majewski, J. (2008). Effect of polymorphisms within probe-target sequences on olignonucleotide microarray experiments. Nucleic Acids Res *36*, 4417-4423.

Brisco, M.J., and Morley, A.A. (2012). Quantification of RNA integrity and its use for measurement of transcript number. Nucleic Acids Res *40*, e144.

Carvalho, B.S., and Irizarry, R.A. (2010). A framework for oligonucleotide microarray preprocessing. Bioinformatics *26*, 2363-2367.

Chaplan, S.R., Bach, F.W., Pogrel, J.W., Chung, J.M., and Yaksh, T.L. (1994). Quantitative assessment of tactile allodynia in the rat paw. Journal of neuroscience methods *53*, 55-63.

Delaneau, O., Marchini, J., and Zagury, J.F. (2012). A linear complexity phasing method for thousands of genomes. Nat Methods *9*, 179-181.

Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet *10*, e1004383.

GTEx_Consortium (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science *348*, 648-660.

Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., and Abecasis, G.R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet *44*, 955-959.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics *4*, 249-264.

Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet *46*, 310-315.

MacDonald, J.W. (2015). pd.hta.2.0: Platform Design Info for Affymetrix HTA-2_0.

Madsen, L., Labrecque, N., Engberg, J., Dierich, A., Svejgaard, A., Benoist, C., Mathis, D., and Fugger, L. (1999). Mice lacking all conventional MHC class II genes. Proceedings of the National Academy of Sciences of the United States of America *96*, 10338-10343.

Maixner, W., Diatchenko, L., Dubner, R., Fillingim, R.B., Greenspan, J.D., Knott, C., Ohrbach, R., Weir, B., and Slade, G.D. (2011). Orofacial pain prospective evaluation and risk assessment study--the OPPERA study. J Pain *12*, T4-11 e11-12.

Marchini, J., and Howie, B. (2010). Genotype imputation for genome-wide association studies. Nat Rev Genet *11*, 499-511.

Marchini, J., Howie, B., Myers, S., McVean, G., and Donnelly, P. (2007). A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet *39*, 906-913.

O'Connell, J., Gurdasani, D., Delaneau, O., Pirastu, N., Ulivi, S., Cocca, M., Traglia, M., Huang, J., Huffman, J.E., Rudan, I., *et al.* (2014). A general approach for haplotype phasing across the full spectrum of relatedness. PLoS Genet *10*, e1004234.

Pers, T.H., Timshel, P., and Hirschhorn, J.N. (2015). SNPsnap: a Web-based tool for identification and annotation of matched SNPs. Bioinformatics *31*, 418-420.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., *et al.* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet *81*, 559-575.

R Core Team (2014). R: A language and environment for statistical computing.

Ramasamy, A., Trabzuni, D., Gibbs, J.R., Dillman, A., Hernandez, D.G., Arepalli, S., Walker, R., Smith, C., Ilori, G.P., Shabalin, A.A., *et al.* (2013). Resolving the polymorphism-in-probe problem is critical for correct interpretation of expression QTL studies. Nucleic Acids Res *41*, e88.

Ramasamy, A., Trabzuni, D., Guelfi, S., Varghese, V., Smith, C., Walker, R., De, T., Coin, L., de Silva, R., Cookson, M.R., *et al.* (2014). Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat Neurosci *17*, 1418-1428.
Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M., and Ragg, T. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol Biol *7*, 3.
Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics *28*, 1353-1358.
Usoskin, D., Furlan, A., Islam, S., Abdo, H., Lonnerberg, P., Lou, D., Hjerling-Leffler, J., Haeggstrom, J., Kharchenko, O., Kharchenko, P.V., *et al.* (2015). Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. Nat Neurosci *18*, 145-153.
Wieskopf, J.S., Mathur, J., Limapichat, W., Post, M.R., Al-Qazzaz, M., Sorge, R.E., Martin, L.J., Zaykin, D.V., Smith, S.B., Freitas, K., *et al.* (2015). The nicotinic alpha6 subunit gene determines variability in chronic pain sensitivity via cross-inhibition of P2X2/3 receptors. Sci Transl Med *7*, 287ra272.
Zhang, X., Gierman, H.J., Levy, D., Plump, A., Dobrin, R., Goring, H.H., Curran, J.E., Johnson, M.P., Blangero, J., Kim, S.K., *et al.* (2014). Synthesis of 53 tissue and cell line expression QTL datasets reveals master eQTLs. BMC Genomics *15*, 532.
Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M., *et al.* (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet *48*, 481-487.