Integrative QTL analysis of gene expression and chromatin accessibility identifies multi-tissue patterns of genetic regulation

Gregory R. Keele[1,2,8◕], Bryan C. Quach[1,2,9◕], Jennifer W. Israel[2], Grace A. Chappell[7], Lauren Lewis[7], Alexias Safi[5,6], Jeremy M. Simon[2], Paul Cotney[2], Gregory E. Crawford[5,6], William Valdar[2,4,‡*], Ivan Rusyn[7,‡*] Terrence S. Furey[2,3,4,‡*]

1 Curriculum in Bioinformatics and Computational Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States of America
2 Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States of America
3 Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States of America
4 Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States of America
5 Department of Pediatrics, Duke University, Durham, NC, United States of America
6 Center for Genomic and Computational Biology, Duke University, Durham, NC, United States of America
7 Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX, United States of America
8 The Jackson Laboratory, Bar Harbor, ME, United States of America
9 Center for Omics Discovery and Epidemiology, Research Triangle Institute (RTI) International, Research Triangle Park, NC, United States of America

◕These authors contributed equally to this work.
‡These authors also contributed equally to this work.
* william.valdar@unc.edu; irusyn@tamu.edu; tsfurey@email.unc.edu

Abstract

Gene transcription profiles across tissues are largely defined by the activity of regulatory elements, most of which correspond to regions of accessible chromatin. Regulatory element activity is in turn modulated by genetic variation, resulting in variable transcription rates across individuals. The interplay of these factors, however, is poorly understood. Here we characterize expression and chromatin state dynamics across three tissues—liver, lung, and kidney—in 47 strains of the Collaborative Cross (CC) mouse population, paying particular attention to the regulation of these dynamics by expression quantitative trait loci (eQTL) and chromatin QTL (cQTL). QTL whose allelic effects were common across tissues were detected for 1,101 genes and 133 chromatin regions. Also detected were tissue-specific QTL, for both genes and chromatin, including a local eQTL for *Pik3c2g* detected in all three tissues but with distinct allelic effects. Leveraging the overlapping measurement of gene expression and chromatin accessibility on the same mice from multiple tissues, we used mediation analysis to identify chromatin and gene expression intermediates of eQTL effects. Based on QTL and mediation analyses over multiple tissues, we propose a causal model for the distal genetic regulation of *Akr1e1*, a gene involved in glycogen metabolism, through a zinc finger protein and chromatin intermediates. This analysis demonstrates the

complexity of transcriptional and chromatin dynamics and their regulation over multiple tissues, as well as the value of the CC and related genetic resource populations for identifying specific regulatory mechanisms in biological data.

Author summary

Genetic variation can drive alterations in gene expression levels and chromatin accessibility, the latter of which defines gene regulatory elements genome-wide. The same genetic variants may associate with both molecular events, and these may be connected within the same causal path: a variant that reduces promoter region chromatin accessibility, potentially by affecting transcription factor binding, will lower expression of that gene. Moreover, these causal regulatory paths can differ between tissues depending on tissue-specific function and cellular activity. We identify cross-tissue and tissue-specific genetic regulators of gene expression and chromatin accessibility in liver, lung and kidney tissues using a panel of genetically diverse inbred mouse strains. Further, we identify a number of candidate causal mediators of the genetic regulation of gene expression, including a zinc finger protein that helps silence the *Akr1e1* gene. Our analyses are consistent with chromatin accessibility playing a role in the regulation of transcription. Our study demonstrates the power of genetically diverse mouse populations, such as the Collaborative Cross, for large-scale studies of genetic drivers of gene regulation that underlie complex phenotypes, as well as identifying causal intermediates that drive variable activity of specific genes and pathways.

Introduction

Determining the mechanisms by which genetic variants drive molecular, cellular, and physiological phenotypes has proved to be challenging [1]. These mechanisms can be informed by genome-wide experiments that provide data on variations in molecular and cellular states in genotyped individuals. Most examples of such data, though, are largely observational, due in part to constraints of specific populations (*e.g.*, humans), the limitations of existing experimental technologies, and the challenge of coordinating large numbers of experiments with multiple levels of data [2]. One approach to shed light on these dynamics is to pair complementary datasets from the same individuals and perform statistical mediation analysis (*e.g.*, [3, 4]), which has increasingly been used in genomics [5]. These analyses can identify putative causal relationships rather than correlational interactions, providing meaningful and actionable targets in terms of downstream applications in areas such as medicine and agriculture.

In human data, co-occurence of QTL across various multi-omic data has been used to assess potentially related and connected biological processes; examples include gene expression with chromatin accessibility [6] or regulatory elements [7], and ribosome occupancy with protein abundances [8, 9]. More formal integration through statistical mediation analyses have also been used to investigate relationships between levels of human biological data, such as distal-eQTL with local gene expression [10], and eQTL with regulatory elements [11–13].

Though genetic association studies of human populations have been highly successful [14], animal models allow for more deliberate control of confounding sources of variation, including experimental conditions and population structure, and thereby provide a potentially powerful basis for detecting associations and even causal linkages. To aid in this effort, genetically-diverse mouse population resources have been established, including the Collaborative Cross (CC) [15–17] and the Diversity Outbred

(DO) mice [18]. The CC and the DO are derived from the same eight founder strains: (short names in parentheses): A/J (AJ), C57BL/6J (B6), 129S1/SvImJ (129), NOD/ShiLtJ (NOD), NZO/H1LtJ (NZO), CAST/EiJ (CAST), PWK/PhJ (PWK), and WSB/EiJ (WSB). The CC are recombinant inbred strains and therefore replicable across and within studies; the DO are largely heterozygous, outbred animals, bred with a random mating strategy that seeks to maximize diversity.

These mouse resources have previously been employed in mediation analyses across genomic assays. A genome-wide mediation approach in 192 DO mice was used to link transcriptional and post-translational regulation of protein levels [19]; the CC were used to confirm results by showing that estimates of founder haplotype effects from each of the related populations corresponded. More recently, mediation analysis was used to connect chromatin accessibility with gene expression in embryonic stem cells derived from DO mice [20].

Here we use a sample composed of a single male mouse from 47 CC strains to investigate tissue-specific dynamics between gene expression and chromatin accessibility, as determined by Assay for Transposase Accessible Chromatin sequencing (ATAC-seq), in lung, liver, and kidney tissues. We detect QTL underlying gene expression and chromatin accessibility variation across the strains and assess support for mediation of the effect of eQTL through chromatin accessibility using a novel implementation of previous methods used in the DO [19]. Additionally, we detect gene mediators of distal-eQTL. These findings demonstrate the experimental power of the CC resource for integrative analysis of multi-omic data to determine genetically-driven phenotype variation, despite limited sample size, and provide support for continued use of the CC in larger experiments going forward.

Results

Differential gene expression and chromatin accessibility

Gene expression and chromatin accessibility cluster by tissue

Gene expression and chromatin accessibility were measured with RNA-seq and ATAC-seq assays, respectively, from whole lung, liver, and kidney tissues in one male mouse from each of 47 CC strains (Fig 1). As the function of each tissue is quite distinct, we expected that these data would reflect tissue-specific differences. Principal components analysis (PCA) of each of the gene expression and chromatin accessibility profiles showed that the samples clearly clustered by tissue (S1 Fig).

Fig 1. Diagram of the experiment and analyses. RNA-seq and ATAC-seq were performed using liver, lung, and kidney tissues of males from 47 CC strains. Each CC strain was derived from an inbreeding funnel, and thus represents a recombinant inbred mosaic of the initial eight founder haplotypes. Differential analyses followed by pathway enrichment analyses were performed to identify biological pathways enriched in differentially expressed genes and accessible chromatin regions. QTL and mediation analyses were performed to identify regions that causally regulate gene expression and chromatin accessibility.

Differentially expressed genes strongly correspond with accessible chromatin regions

Differential expression (DE) and differentially accessible region (DAR) analysis were performed between the three tissues (S1 Table), and revealed between 3,564 - 5,709 DE

genes and 28,048 - 40,797 DARs (FWER < 0.1). For both expression and chromatin accessibility, liver and kidney tissues were the most similar, while lung and liver were the most distinct, also seen in the PCA plots (S1 Fig). Pathway analyses showed many between-tissue differences related to metabolic and immune-related pathways (FWER < 0.1), reflecting the distinct demands of each tissue. Energy metabolism pathways were more active in liver and kidney and immune-related pathways were more pronounced in lung, as expected. We compared the concordance between DE genes and DARs genome-wide and observed that most DE gene promoters do not show significant differences in chromatin accessibility (S2 Fig). In cases with significant variability in accessibility at the promoter of a DE gene, though, the vast majority agree in direction (*i.e.*, higher expression with greater accessibility).

QTL detection

Gene expression

The impact of genetic variation on gene expression was evaluated by eQTL mapping. This was done at three levels of stringency and emphasis: 1) at the level of the local region of a gene, defined as within 10Mb of the gene transcription start site (TSS), and hereafter termed Analysis L; 2) at the level of the chromosome on which the gene is located (Analysis C); and 3) at level of the genome (Analysis G) (details in **Methods**). After filtering out lowly expressed genes, the number of genes examined in eQTL mapping was 8401 for liver, 11357 for lung, and 10092 for kidney (UpSet plot [58] in S3 FigA).

Analysis L detected local-eQTL for 19.8% of genes tested in liver, 16.6% in lung, and 20.8% in kidney (S2 Table). Most genes with local-eQTL were observed in only one tissue (S4 FigA). Analysis C, which was more stringent, additionally detected intra-chromosomal distal-eQTL, while Analysis G, the most stringent, additionally detected inter-chromosomal distal-eQTL (S2 Table). Genomic locations of eQTL detected for each tissue, excluding the intra-chromosomal distal-eQTL detected by Analysis C, are shown in Fig 2A. See S5 Fig and S3 Table for eQTL counts with FDR < 0.2.

Fig 2. Detected QTL are largely local for both gene expression (A) and chromatin accessibility (B). QTL detected through all analyses, excluding intra-chromosomal distal-QTL detected through Analysis C, are included. The y-axis represents the genomic position of the gene or chromatin site, and the x-axis represents the genomic position of the QTL. Local-QTL appear as dots along the diagonal and distal-QTL as off-diagonal dots.

Chromatin accessibility

To determine genetic effects on chromatin structure, genomic regions were divided into ∼300 base pair windows and analyses similar to eQTL were used to detect chromatin accessibility QTL (cQTL). The number of chromatin regions tested were 11448 in liver, 24426 in lung, and 17918 in kidney. The overlap in chromatin windows tested across tissues is described in S3 FigB. Overall, there were substantially fewer cQTL detected compared to eQTL for all tissues (Fig 2B; S4 Table, S5 Table). As with eQTL, cQTL were more likely to be local than distal (66 - 94.1% local-cQTL for Analysis G; 75 - 90% local-cQTL for Analysis C).

Local-QTL have stronger effects than distal-QTL

For QTL detected on the same chromosome as the gene or chromatin region (intra-chromosomal), the strongest associations were observed within 10Mb of the gene TSS or chromatin window midpoint (S6 Fig). Intra-chromosomal distal-QTL had reduced statistical significance, more consistent with inter-chromosomal distal-QTL. This dynamic between local- and distal-QTL is also observed when using the QTL effect size as a summary of strength, which is likely biased upward due to the Beavis effect in 47 strains [42], shown in S7 Fig.

QTL driven by extreme effects of CAST and PWK

For all QTL we estimated the effects of the underlying founder haplotypes. Consistent with previous studies (*e.g.*, [38]), the CAST and PWK haplotypes had higher magnitude effects compared to the classical inbred strains. This pattern was observed for both local- and distal-eQTL (S8 FigA) and local-cQTL; the numbers of detected distal-cQTL were too low to produce clear trends (S8 FigB).

QTL paired across tissues

For a given trait, QTL from different tissues were paired based on co-localizing to approximately the same genomic region. For local-QTL, both had to be within the local window, defined as 10Mb around the gene TSS, resulting in a maximum distance of 20Mb between QTL. For distal-eQTL, both had to be within 10Mb of each other. For local-eQTL, we detected 761 (liver/lung), 1206 (liver/kidney), and 1025 (lung/kidney) pairs. For distal-eQTL, we detected 61 (liver/lung), 120 (liver/kidney), and 59 (lung/kidney) pairs. For cQTL, the vast majority of pairs were local, with 55 (liver/lung), 56 (liver/kidney), and 142 (lung/kidney) pairs found. Only 4 distal-cQTL pairs were observed, all between lung and kidney. The effect sizes of paired QTL vary across tissues, though for any given tissue pairing, the effect sizes were significantly correlated (least significant p-value $=4.6 \times 10^{-8}$; S9 Fig).

Multi-tissue QTL pairs share similar haplotype effects across tissues

Testing correlation between the founder haplotype effects for pairs of QTL (FDR < 0.1; S10 Fig) revealed significant positive correlations (FDR < 0.1) for 345 (45.3%; liver/lung), 623 (51.7%; liver/kidney), and 498 (48.6%; lung/kidney) local-eQTL pairs and 21 (31.8%; liver/lung), 34 (28.8%; liver/kidney), and 16 (27.1%; lung/kidney) distal-eQTL pairs. Highly correlated QTL pairs were very close in genomic distance to each other (S11 Fig), suggesting that the underlying causal variants are the same or linked. For local-cQTL, 47 (85.5%; liver/lung), 48 (85.7%; liver/kidney), and 118 (83.1%; lung/kidney) significantly positively correlated pairs were identified. The correlations between distal-cQTLs were not formally tested because there were only four pairs; however, three of the four pairs had correlations greater than 0.5. No significantly negatively correlated QTL pairs were detected, after accounting for multiple testing (FDR < 0.1).

Cox7c: Consistent haplotype effects for multi-tissue local-eQTL

Cytochrome c oxidase subunit 7C (*Cox7c*) is an example of a gene that possessed local-eQTL with highly correlated effects in all three tissues (Fig 3A). The local-eQTL consistently drove higher expression when the CAST haplotype was present, intermediate expression with 129, NOD, and NZO haplotypes, and lower expression with A/J, B6, PWK, and WSB haplotypes. Though the haplotype effects on relative

expression levels within each tissue were consistent, we noted that the expression levels was significantly higher in liver compared to both lung ($q = 5.41 \times 10^{-18}$) and kidney ($q = 6.25 \times 10^{-19}$). Ubiquitin C (*Ubc*) is another example of a gene with consistent local-eQTL detected in all three tissues (S12 Fig).

Fig 3. Examples of genes with local-eQTL observed in all three tissues. *Cox7c* possesses local-eQTL with highly correlated haplotype effects across all three tissues, supportive of shared causal origin (A). *Slc44a3* has a more complicated pattern of local-eQTL founder effects across the tissues, with correlated effects shared between lung and kidney, and transgressive effects in the liver eQTL by comparison, consistent with distinct casual variants comparing liver to lung and kidney. (B). For both genes, the expression data are plotted with bars representing the interquartile ranges of likely founder haplotype pair (diplotype). Differential expression between tissues is highlighted. The haplotype association for each tissue is also included near the gene TSS with variant association overlaid. The most statistically rigorous method that detected the QTL (Analysis L, C, or G) is also included. The red tick represents the gene TSS, the black tick represents the variant association peak, and the colored tick represents the haplotype association peak. Founder effects estimated as constrained best linear unbiased predictions (BLUPs) are also included with their pairwise correlations. Estimated effects were generally consistent with the expression data.

Slc44a3 and *Pik3c2g*: Tissue-specific local-eQTL

Uncorrelated QTL pairs potentially represent distinct tissue-specific QTL where the same or different genetic variants influence gene expression or chromatin accessibility in a tissue-specific manner. For example, the solute carrier family 44, member 3 (*Slc44a3*) gene has correlated local-eQTL effects in lung and kidney, but unique effects in liver (Fig 3B). Notably, the effects in liver are anti-correlated with the effects in lung and kidney, suggesting the liver eQTL could be transgressive [59] to the eQTL in lung and kidney, whereby the effects of the founder haplotypes are reversed. For *Slc44a3*, CAST, PWK, and WSB haplotypes result in higher expression in liver, but lower expression in lung and kidney. The local-eQTL for *Slc44a3* were more similar in location in lung and kidney, whereas the liver eQTL was more distal to the gene TSS. Overall, the expression data, estimated effects, and patterns of association are consistent with lung and kidney sharing a causal local-eQTL that is distinct from the one in liver.

The CC founder strains all possess contributions from three subspecies of *M. musculus*: *domesticus* (*dom*), *castaneus*, (*cast*), and *musculus* (*mus*) [60]. Allele-specific gene expression in mice descended from the CC founders have been found to often follow patterns that matched the subspecies inheritance at the gene regions [61]. We found that phosphatidylinositol-4-phosphate 3-kinase catalytic subunit type 2 gamma (*Pik3c2g*), a gene of interest for diabetes-related traits [62], had tissue-specific local-eQTL in all three tissues that closely matched the subspecies inheritance at their specific genomic coordinates. The local-eQTL in all three tissues were all pair-wise uncorrelated (Fig 4). Further, expression of *Pik3c2g* varied at statistically significant levels for liver versus lung ($q = 3.12 \times 10^{-13}$) and lung versus kidney ($q = 3.55 \times 10^{-6}$). In lung, the CAST haplotype (*cast* subspecies lineage) resulted in higher expression, consistent with a *cast* versus *dom*/*mus* allelic series. In kidney, the B6, 129, NZO, and WSB haplotypes (*dom* subspecies) resulted in higher expression, whereas A/J and NOD (*mus*), CAST (*cast*), and PWK (*dom*) haplotypes showed almost no expression, mostly consistent with a *dom* versus *cast*/*mus* allelic series. The PWK founder appears inconsistent, but we note that the QTL peak was in a small *dom* haplotype block interspersed within a broader *mus* region. The expression level in kidney of *Pik3c2g* in

PWK was low, similar to A/J and NOD, suggesting that the causal variant is located in the nearby region, where all three have *mus* inheritance and, notably, where the peak variant association occurred. In liver, the B6 haplotype resulted in lower expression compared to the other haplotypes. Interestingly, the liver eQTL was in a region that contained a recombination event between *dom* and *mus*, only present in B6, which may explain the unique expression pattern.

Fig 4. *Pik3c2g* possesses tissue-specific local-eQTL. Local-eQTL for *Pik3c2g* were detected in all three tissues in the 3Mb region surrounding its TSS. The genomes of the CC founders can be simplified in terms of contributions from three subspecies lineages of *M. musculus*: *dom* (blue), *cast* (yellow), and *mus* (red). The effects of each local-eQTL matched the subspecies contributions near the eQTL coordinates, with low expressing subspecies alleles colored teal and high expressing alleles colored salmon, consistent with local-eQTL for *Pik3c2g* being distinct and tissue-specific. The gene expression data are represented as interquartile range bars, categorized based on most likely diplotype at the eQTL for each CC strain. Haplotype and variants associations are included for each tissue, with the red tick representing the *Pik3c2g* TSS, black ticks representing variant association peak, and colored ticks representing the haplotype association peak. The most rigorous procedure to detect each QTL (Analysis L, C, or G) is reported. Founder effects, estimated as constrained BLUPs, were consistent with the expression data, and uncorrelated across the tissues.

Akr1e1 and *Per2*: Consistent haplotype effects for multi-tissue distal-eQTL

Founder haplotype effects that correlate across tissues can provide stronger confirmation of distal-QTL, even those with marginal significance in any single tissue. For example, the Aldo-keto reductase family 1, member E1 (*Akr1e1*; Fig 5A) gene is located on chromosome 13 with no local-eQTL detected in any tissue. Distal-eQTL in all tissues were detected for *Akr1e1* that localized to the same region on chromosome 4. The founder haplotype effects of the distal-eQTL were all highly correlated, with A/J, B6, 129, and CAST haplotypes corresponding to higher *Akr1e1* expression. Across tissues, expression was significantly higher in liver and kidney compared to lung ($q = 1.63 \times 10^{-19}$ and $q = 4.40 \times 10^{-34}$, respectively). Another example is the Period circadian clock 2 (*Per2*; Fig 5B) gene, which possessed intra-chromosomal distal-eQTL approximately 100Mb away from the TSS that were detected in all three tissues at a lenient chromosome-wide significance (Analysis C; FDR < 0.2). The founder haplotype effects were significantly correlated among the tissues, characterized by high expression with the NOD and PWK haplotypes present. Together, these findings provide strong validation for the distal-eQTL, which would commonly not be detected based on tissue-stratified analyses. Other unique patterns of distal genetic regulation were detected, such as for Ring finger protein 13 gene (*Rnf13*) with tissue-specific distal-eQTL, described in S13 Fig.

Mediation of eQTL by chromatin

Measurement of gene expression and chromatin accessibility data in the same mice and tissues enabled the use of mediation analysis to elucidate the relationships between genotype, chromatin accessibility, and gene expression. We considered two possible models for the mediation of eQTL effects and assessed the evidence for each in our data. The first model assessed evidence for proximal chromatin state acting as a mediator of local-eQTL (Fig 6A). We used an approach adapted from studies in the DO [19] to

Fig 5. Examples of genes with distal-eQTL effect patterns across tissues. *Akr1e1* has highly significant distal-eQTL detected on chromosome 4 in all three tissues with correlated founder effects (A). *Per2* has intra-chromosomal distal-eQTL leniently detected 100Mb away from the TSS, also with highly correlated founder effects across the tissues, providing further support that the distal-eQTL are not false positives (B). Expression data are represented as interquartile ranges for most likely diplotype, with differential expression noted when significant. Haplotype associations for each tissue distal-eQTL combination are shown, with the most rigorous statistical procedure for detection reported (Analysis L, C, or G). Red ticks signify the gene TSS and black ticks represent that eQTL peak. Fit founder effects, estimated as BLUPs, are included, along with the pairwise correlations of the eQTL.

detect mediation through chromatin accessibility of local-eQTL detected through Analysis L (see S3 Appendix for greater detail).

Fig 6. Simple mediation models for the genetic regulation of gene expression. Mediation of the local-eQTL through chromatin accessibility in the region of the gene (A) is consistent with genetic variation influencing the accessibility of gene j to the transcriptional machinery. Mediation of distal-eQTL through the transcription of genes local to the QTL (B) could be explained by transcription factor activity.

Ccdc137 and *Hdhd3*: Local-eQTL driven by local chromatin accessibility

Across the three tissues, we found from 13-42 local-eQTL with evidence of mediation through proximal accessible chromatin regions at genome-wide significance, and 35-106 at chromosome-wide significance (S6 Table). The coiled-coil domain containing 137 gene (*Ccdc137*) is a strong example of this type of mediation. Fig 7 shows the genome scans that identify both the local-eQTL for *Ccdc137* and the cQTL near it. When conditioning on the chromatin accessibility, we saw a sharp reduction in significance of the eQTL. We noted that the significance of the cQTL is stronger than the eQTL, as expected by the proposed mediator model.

Fig 7. *Ccdc137* local-eQTL is mediated by proximal chromatin accessibility. *Ccdc137* expression and the chromatin accessibility in the proximal region were highly correlated ($r = 0.71$). Genome scans for *Ccdc137* expression (yellow), nearby chromatin accessibility (blue), and chromatin mediation of the *Ccdc137* local-eQTL (red) in lung tissue. The local-eQTL and local-cQTL for the chromatin region at the TSS of *Ccdc137* (red tick) are over-lapping, and have highly correlated haplotype effects ($r = 0.98$). The steep drop in the statistical association with expression, represented as logP, at the chromatin site in the mediation scan supports chromatin mediation of *Ccdc137* expression, depicted as a simple graph in the top-right. The QTL and mediation signals were detected at genome-wide significance.

Co-localization of eQTL and cQTL is not sufficient for mediation. For both *Hdhd3* in liver (S14 FigA) and the acyl-Coenzyme A binding domain containing 4 gene (*Acbd4*) in kidney (S14 FigB), there were local-eQTL with a co-localizing cQTL. Comparing the statistical associations at the genomic locations of the eQTL and cQTL as well as the founder haplotype effects for both, we found better correspondence for *Hdhd3* compared to *Acbd4*. As with *Ccdc137*, a strong mediation signal was detected for *Hdhd3*, indicated by the decreased eQTL association when conditioning on the chromatin state.

Mediation was not detected for *Abcd4*, though, consistent with different causal origins of expression and chromatin accessibility.

Mediation of eQTL by expression

We also tested for distal-eQTL, detected through Analysis G, that were modulated by the expression of a nearby gene [56], and diagrammed in Fig 6B.

Zinc finger protein intermediates detected for *Ccnyl1*

Eight genes were identified with mediated distal-eQTL (S7 Table). For example, the distal-eQTL of cyclin Y-like 1 (*Ccnyl1*) was mediated by the expression of zinc finger protein 979 (*Zfp979*), a putative transcription factor (S15 Fig). The founder haplotype effects at the distal-eQTL of *Ccnyl1* were highly correlated with the local-eQTL effects for *Zfp979*, though with a reduction in overall strength, in accordance with the proposed mediation model.

Mediation of eQTL by both expression and chromatin

The QTL and mediation results for all three tissues were depicted as circos plots [63] in Fig 8. We saw that neither the local-QTL nor chromatin mediators were evenly distributed across the genome but rather tended to aggregate in pockets. In particular, a high concentration of cQTL and chromatin mediation were detected in all three tissues along chromosome 17, which corresponds to the immune-related major histocompatibility (MHC) region in mouse. However, most of the chromatin-mediated genes in this region are not histocompatibility genes (Files S28 - S30). Patterns of distal-QTL and gene mediation appear to be tissue-specific, though consistent regions are observed across the tissues, such as the *Idd9* region on chromosome 4 described in [64], which contains multiple zinc finger proteins (ZFPs) and regulates genes such as *Ccnyl1* and *Akr1e1*, described in greater detail below.

Fig 8. Summaries of QTL and mediation analyses. Circos plots of eQTL (yellow), cQTL (blue), and mediation (red) in lung, liver, and kidney. The two outer rings of dots represent local-eQTL and local-cQTL detected by Analysis L at chromosome-wide significance, with red lines between connecting genes and chromatin sites for which chromatin mediation was detected. The inner circle contains connections representing distal-eQTL, distal-cQTL, and gene-gene mediation from Analysis G. Thick lines represent QTL and mediators with permutation-based p-value (permP) $< 1 \times 10^{-5}$. The detected signals are primarily local, which also tend to be stronger than the observed distal signals. Fewer QTL and mediators are detected in liver tissue. Genes highlighted within the text, such as *Akr1e1*, are indicated at their genomic coordinates.

Genetic regulation of *Akr1e1* expression includes zinc finger protein and chromatin intermediates

As detailed above, *Akr1e1* possessed a strong distal-eQTL in all three tissues located on chromosome 4 in the region of 142.5 - 148.6Mb with significantly correlated haplotype effects. Mediation analysis suggested this effect is mediated in lung through activity of Zinc finger protein 985 (ZFP985), whose gene is also located on chromosome 4 at 147.6Mb (S16 Fig).

ZFP985 possesses an N-terminal Krüppel-associated box (KRAB) domain, representing a well-characterized class of transcriptional regulators in vertebrates [65],

that recruit histone deacetylases and methyltransferases, inducing a heterochromatic state associated with regulatory silencing. *Akr1e1* and *Zfp985* expression were negatively correlated ($r = -0.69$), consistent with ZFP985 inhibiting *Akr1e1* expression. This same distal-eQTL and mediator relationship for *Akr1e1* was observed in kidney tissue from 193 DO mice (S17 Fig). It was previously postulated that *Akr1e1* is distally regulated by the Reduced expression 2 gene (*Rex2*), also a ZFP that contains a KRAB domain and resides in the same region of chromosome 4 [66]. Our distal-eQTL overlap their *Idd9.2* regulatory region, defined as spanning 145.5 - 148.57Mb, mapped with NOD mice congenic with C57BL/10 (B10) [64]. The B10 haplotype was found to be protective against development of diabetes, characteristic of NOD mice. AKR1e1 is involved in glycogen metabolism and the *Idd9* region harbors immune-related genes, suggesting genes regulated by elements in the *Idd9* region may be diabetes-related. Consistent with these studies, CC strains with NOD inheritance at the distal-eQTL had low expression of *Akr1e1*, observed in all tissues. Genetic variation from B10 is not present in the CC, but the closely related B6 founder had high expression of *Akr1e1* like B10. NZO, PWK, and WSB haplotypes resulted in low *Akr1e1* expression, similar to NOD, while A/J, 129, and CAST haplotypes joined B6 as driving high expression (Fig 9). Lowly expressed *Zfp985*, resulting in low levels of ZFP985, is a strong candidate for driving these effects on *Akr1e1*.

Fig 9. Mediation model for *Akr1e1* distal-eQTL. The genetic regulation of *Akr1e1* expression is reconstructed based on relationships observed across the three tissues. Distal-eQTL were detected in all tissues at similar levels of significance. A local-eQTL for *Zfp985* that is proximal to the *Akr1e1* distal-eQTL was observed in lung, and *Zfp985* expression was detected as an anti-correlated mediator of the distal-eQTL, consistent with ZFP985 suppressing *Akr1e1* expression. The chromatin site proximal to the *Akr1e1* TSS has a distal-cQTL detected in kidney. Chromatin accessibility at the site was found to be a significant mediator of *Akr1e1* expression. Combining associations across tissues supports a biological model whereby ZFP985, whose gene is expressed in mice with NOD, NZO, PWK, and WSB haplotypes, silences *Akr1e1* through KRAB domain-induced heterochromatin. QTL and mediation genome scans are included, along with sequence phenotypes as interquartile ranges categorized according to most likely diplotype, and modeled founder effects fit as BLUPs. The relative magnitudes of the QTL effect sizes and mediation significance are consistent with the proposed model, with *Zfp985* local-eQTL > distal-cQTL > *Akr1e1* distal-eQTL and chromatin mediation > mediation through *Zfp985* expression.

Further elucidation of the genetic regulation of *Akr1e1* was observed in kidney tissue, where a strong distal-cQTL for an accessible chromatin site corresponding to the promoter of *Akr1e1* was detected in the *Idd9.2* region on chromosome 4. Mediation analysis strongly supported chromatin accessibility in this region mediating the distal-eQTL ($\text{permP}^m = 2.18 \times 10^{-13}$). The founder effects for the distal-cQTL were highly correlated with the distal-eQTL effects ($r = 0.92$). The relative magnitudes of the QTL effect sizes and mediation p-values (S18 Fig) support a causal model whereby increased *Zfp985* expression reduces expression of *Akr1e1* by altering chromatin accessibility in the *Akr1e1* promoter region (Fig 9).

Discussion

In this study we have performed QTL and mediation analyses of gene expression and chromatin accessibility data in liver, lung, and kidney tissue samples from 47 strains of the CC. We used the formal statistical testing of the correlation coefficient between

founder haplotype effects of overlapping QTL to identify QTL that are likely functionally active in multiple tissues as well as likely distinct tissue-specific QTL, as is the case with the *Pik3c2g* gene. We detected extensive evidence of chromatin mediation of local-eQTL as well as gene expression mediators underlying distal-eQTL. One unique example is the elucidation of the genetic regulation of *Akr1e1* expression, a gene that plays a role in glycogen metabolism, involving distal inhibition by a zinc finger protein mediator that reduces chromatin accessibility at the TSS. These findings highlight the ability of integrative QTL approaches, including mediation analysis, to identify interesting biological findings, including the ability to identify functional candidates for further downstream analysis.

Joint QTL analysis in multiple tissues

Multi-tissue QTL analyses are increasingly being used in both humans, such as within the GTEx project [67], and in mice (*e.g.*, [68]). We believe our use of formal statistical tests of the correlation coefficient between the founder haplotype effects of overlapping QTL is novel in defining co-localizing QTL across tissues. This method allows us to identify loci likely representing tissue-specific QTL with unique haplotype effects patterns, as demonstrated for the gene *Pik3c2g* (Fig 4). This approach can also detect consistent haplotype effects, as with the gene *Per2* (Fig 5B), which possessed only marginally significant distal-eQTL in the three tissues, suggesting jointly mapping QTL across all tissues increases power. Formal joint analysis approaches have been proposed, largely implemented for detecting SNP associations, including meta-analysis on summary statistics (*e.g.*, [69, 70]) and fully joint analysis, including Bayesian hierarchical models [71] and mixed models [72]. Extending these to haplotype-based analysis in MPPs poses some challenges, including how to best generalize methods to more complex genetic models and for the CC with a limited number of unique genomes. When multiple levels of molecular traits are measured, joint analyses may be incorporated into the mediation framework to improve detection power.

Correlated haplotype effects suggest subtle multi-allelic QTL

Haplotype-based association in MPP allows for the detection of multi-allelic QTL [61], such as potentially observed within the kidney at the local-eQTL of *Pik3c2g*, where mice with B6 contributions in the region have an intermediate level of expression. The correlation coefficient between the founder haplotype effects for QTL pairs provides an interesting summary, generally not possible in the simpler bi-allelic setting commonly used in variant association analysis. The extent of the correlation between the founder effects for QTL pairs of certain genes, such as *Cox7c* (Fig 3A) suggests that these QTL are at least subtly multi-allelic. Correlated founder effects are consistent with genetic regulation, even local to the gene TSS, being potentially complex with strain-specific modifiers.

Correlated haplotype effects with differential expression across tissues

We detected numerous eQTL pairs with correlated haplotype effects across tissues but with significantly different magnitudes of overall expression; such as for *Cox7c*, *Ubc*, *Slc44a3*, and *Akr1e1*. We propose two potential explanations for this unique co-occurrence. First, whole tissue are heterogeneous in cellular composition. These genes may be primarily expressed in a common cell type that varies significantly in its proportion in each tissue. The differential expression may simply reflect variation in composition. Single cell experiments would be able to confirm or refute this. Second,

each tissue may have additional unique regulatory elements that further modulate expression levels. In-depth analysis of tissue-specific regulation would be necessary to uncover such elements.

Mediation analysis

Our mediation analysis approach is adapted from previous studies in DO mice [19, 20, 56] but adds the use of QTL effect sizes to establish consistency with the directionality of the relationships and the formal calculation of a mediation p-value through permutation. Related conditional regression approaches were used in the incipient pre-CC lines [73, 74]. Mediation analyses largely reflect the correlations between the variables after adjusting for additional sources of variation, such as covariates and batch effects. We additionally require the mediator QTL to have a larger effect size than the outcome QTL in order to identify trios that are consistent with the proposed mediation models. Measurement of mediator effects may be noisier, resulting in a reduced mediator QTL effect size and false negatives. Despite these limitations, mediation analysis provides specific causal candidates of QTL. For example, we show strong evidence for ZFP985 mediating the genetic regulation of *Akr1e1* (Fig 9). Additional evidence suggests this is done by ZFP985 contributing to reduced *Akr1e1* promoter activity. It is possible that *Zfp985* expression is simply strongly correlated with the true mediator, and others have alternatively proposed *Rex2* expression as a candidate [64], which we did not consider due to low expression in all three tissues. Regardless of the identity of the true mediator, our analysis has shown strong evidence that it acts causally by inducing a heterochromatic state near the *Akr1e1* promoter.

QTL mapping power and their effect size estimates

Our reduced sample size of 47 CC strains motivated our use of multiple scopes of statistical significance, from testing for QTL locally to genome-wide. As expected, the additional local-QTL detected by less stringent methods have smaller effect sizes (orange and purple dots for C and L, respectively; S19 Fig). Based on a recent evaluation of QTL mapping power in the CC using simulation [42], this study has approximately 80% power to detect genome-wide QTL with a 55% effect size or greater. Effect size estimates for genome-wide QTL (green dots; S19 Fig) are consistent with this expectation albeit potentially inflated due to the Beavis effect [75]. Effect size estimates provide interpretable point summaries for haplotype-based QTL mapping, analogous to minor allele effect estimates in SNP-based studies. We calculated estimates of effect size in two ways, one based on a fixed effect fitting of the QTL term and the other as a random term [48]. We primarily report fixed effects estimates due to their consistency with reported expectations [42] for a study of this size. These QTL mapping results are largely consistent with primarily Mendelian genetic regulation of molecular traits (large effect sizes: $> 60\%$). Shrunken effect size estimates can be used to identify putative QTL with unstable fixed effect estimates that result from extreme data points, and thus aid in the identification of low confidence QTL. Notably, a small number of the distal-eQTL had low random effect size estimates (S20 Fig) compared to their fixed effects-based estimates. Alternatively, a residual variation estimate, *i.e.* RSS, could be calculated from the shrunken haplotype effects to identify likely false positives, but not be as aggressively reduced as the variance component-based estimates, representing a middle ground approach that was found to be effective [51].

Materials and methods

Animals

Adult male mice (8-12 weeks old) from 47 CC strains were acquired from the University of North Carolina Systems Genetics Core (listed in S1 Appendix) and maintained on an NTP 2000 wafer diet (Zeigler Brothers, Inc., Gardners, PA) and water *ad libitum*. The housing room was maintained on a 12-h light-dark cycle. Our experimental design sought to maximize the number of strains relative to within-strain replications based on the power analysis for QTL mapping in mouse populations [21]; therefore, one mouse was used per strain. Mice were euthanized and lungs, liver and kidney tissues were collected, flash frozen in liquid nitrogen, and stored at -80°C. These studies were approved by the Institutional Animal Care and Use Committees at Texas A&M University and the University of North Carolina. The experimental design and subsequent analyses performed for this study are diagrammed in Fig 1.

mRNA sequencing and processing

Total RNA was isolated from flash-frozen tissue samples using a Qiagen miRNeasy Kit (Valencia, CA) according to the manufacturer's protocol. RNA purity and integrity were evaluated using a Thermo Scientific Nanodrop 2000 (Waltham, MA) and an Agilent 2100 Bioanalyzer (Santa Clara, CA), respectively. A minimum RNA integrity value of 7.0 was required for RNA samples to be used for library preparation and sequencing. Libraries for samples with a sufficient RNA integrity value were prepared using the Illumina TruSeq Total RNA Sample Prep Kit (Illumina, Inc., San Diego, USA) with ribosomal depletion. Single-end (50 bp) sequencing was performed (Illumina HiSeq 2500).

Sequencing reads were filtered (sequence quality score ≥ 20 for $\geq 90\%$ of bases) and adapter contamination was removed (TagDust). Reads were mapped to strain-specific pseudo-genomes (Build37, `http://csbio.unc.edu/CCstatus/index.py?run=Pseudo`) and psuedo-transcriptomes (C57BL/6J RefSeq annotations mapped to pseudo-genomes) using RSEM with STAR (v2.5.3a). Uniquely aligned reads were used to quantify expression as transcripts per million (TPM) values.

ATAC-seq processing

Flash frozen tissue samples were pulverized in liquid nitrogen using the BioPulverizer (Biospec) to break open cells and allow even exposure of intact chromatin to Tn5 transposase [22]. Pulverized material was thawed in glycerol containing nuclear isolation buffer to stabilize nuclear structure and then filtered through Miracloth (Calbiochem) to remove large tissue debris. Nuclei were washed and directly used for treatment with Tn5 transposase. Paired-end (50 bp) sequencing was performed (Illumina HiSeq 2500).

Reads were similarly filtered as with RNA-seq. Reads were aligned to the appropriate pseudo-genome using GSNAP (parameter set: -k 15, -m 1, -i 5, –sampling=1, –trim-mismatch-score=0, –genome-unk-mismatch=1, –query-unk-mismatch=1). Uniquely mapped reads were converted to mm9 (NCBI37) mouse reference genome coordinates using the associated MOD files (UNC) to allow comparison across strains. Reads overlapping regions in the mm9 blacklist (UCSC Genome Browser) were removed. Exact sites of Tn5 transposase insertion were determined as the start position +5 bp for positive strand reads, and the end position -5 bp for negative strand reads [23]. Peaks were called using F-seq with default parameters. A union set of the top 50,000 peaks (ranked by F-seq score) from each sample was

derived. Peaks were divided into overlapping 300 bp windows [24]. Per sample read coverage of each window was calculated using coverageBed from BedTools [25].

Sequence trait filtering for QTL analysis

Trimmed mean of M-values (TMM) normalization (edgeR; [26]) was applied to TPM values from read counts of genes and chromatin windows, respectively. Genes with TMM-normalized TPM values ≤ 1 and chromatin windows with normalized counts ≤ 5 for $\geq 50\%$ of samples were excluded. For each gene and chromatin window, we applied K-means clustering with $K = 2$ to identify outcomes containing outlier observations that could cause spurious, outlier-driven QTL calls. Any gene or chromatin window where the smaller K-means cluster had a cardinality of 1 was removed.

CC strain genotypes and inferred haplotype mosaics

CC genomes are mosaics of the founder strain haplotypes. The founder haplotype contributions for each CC strain was previously reconstructed by the UNC Systems Genetics Core (`http://csbio.unc.edu/CCstatus/index.py?run=FounderProbs`) with a Hidden Markov Model [27] on genotype calls [MegaMUGA array [28]] from multiple animals per strain, representing ancestors to the analyzed mice. Notably, QTL mapping power is reduced at loci with segregating variants in these ancestors, and where these specific animals likely differ [29]. To reduce the number of statistical tests, adjacent genomic regions were merged through averaging if the founder mosaics for all mice were similar, defined as L2 distance $\leq 10\%$ of the maximum L2 distance ($\sqrt{2}$ for a probability vector). This reduced the number of tested loci from 77,592 to 14,191.

Differential expression and accessibility analyses

Read counts for each sample were converted to counts per million (CPM), followed by TMM normalization (edgeR). Windows in which > 70 % of samples had a CPM ≤ 1 were removed. Differentially expressed genes and accessible chromatin windows were determined using *limma* [30], which fit a linear model of the TMM-normalized CPM value as the response and fixed effect covariates of strain, batch, and tissue (lung, liver, or kidney). To account for mean-variance relationships in gene expression and chromatin accessibility data, precision weights were calculated using the *limma* function *voom* and incorporated into the linear modeling procedure. The p-values were adjusted using a false discovery rate (FDR) procedure [31], and differentially expressed genes and accessible chromatin windows were called based on the q-value ≤ 0.01 and $\log_2$ fold-change ≥ 1. Adjacent significantly differential chromatin windows in the same direction were merged with a p-value computed using Simes' method [32], and chromatin regions were re-evaluated for significance using the Simes p-values.

Gene set association analysis

Biological pathways enriched with differentially expressed genes or accessible chromatin were identified with GSAASeqSP [33] with Reactome Pathway Database annotations (July 24, 2015 release). A list of assayed genes were input to GSAASeqSP along with a weight for each gene g, calculated as:

$$\text{weight}_g = \text{sign}(\Delta_g) \times (1 - q_g), \tag{1}$$

where $\text{sign}(\Delta_g)$ is the sign of the fold change in gene g expression, and q_g is the FDR-adjusted differential expression p-value. Pathways with gene sets of cardinality $<$ 15 or > 500 were excluded.

For pathway analysis of differentially accessible chromatin near genes, each chromatin region was mapped to a gene using GREAT v3.0.0 (*basal plus extension* mode, *5 kb upstream*, *1 kb downstream*, and no distal extension). Weights were calculated as with gene expression, but with $\text{sign}(\Delta_g)$ representing the sign of the fold-change in accessibility of the chromatin region with minimum FDR-adjusted p-value that is associated with gene g.

Haplotype-based QTL mapping

QTL analysis was performed for both gene expression and chromatin accessibility using regression on the inferred founder haplotypes [34], a variant of Haley-Knott regression [35, 36] commonly used for mapping in the CC [37–42] and other multiparent populations (MPPs), such as *Drosophila* [43].

For a given trait—the expression of a gene or the accessibility of a chromatin region—a genome scan was performed in which at each of the 14,191 loci spanning the genome, we fit the linear model,

$$y_i = \mu + \text{batch}_{b[i]} + \text{QTL}_i + \varepsilon_i \tag{2}$$

where y_i is the trait level for individual i, μ is the intercept, batch_b is a categorical fixed effect covariate with five levels $b = 1, \ldots, 5$ representing five sequencing batches for both gene expression and chromatin accessibility and $b[i]$ denoting the batch relevant to i, $\varepsilon_i \sim \text{N}(0, \sigma^2)$ models the residual error, and QTL_i models the genetic effect at the locus, namely that of the eQTL for expression or cQTL for chromatin accessibility. Specifically, the QTL term models the (additive) effects of alternate haplotype states and is defined as $\text{QTL}_i = \boldsymbol{\beta}^{\text{T}}\mathbf{x}_i$, where $\mathbf{x}_i = (x_{i,\text{AJ}}, \ldots, x_{i,\text{WSB}})^{\text{T}}$ is a vector of haplotype dosages (*i.e.*, the posterior expected count, from 0 to 2, inferred by the haplotype reconstruction) for the eight founder haplotypes, and $\boldsymbol{\beta} = (\beta_{\text{AJ}}, \ldots, \beta_{\text{WSB}})^{\text{T}}$ is a corresponding vector of fixed effects. Note that in fitting this term as a fixed effects vector, the linear dependency among the dosages in $\mathbf{x}_i$ results in at least one founder haplotype effect being omitted to achieve identifiability; estimation of effects for all eight founders is performed using a modification described below. Prior to model fitting, to avoid sensitivity to non-normality and strong outliers, the response $\{y_i\}_{i=1}^n$ was subject to a rank inverse normal transformation (RINT). The fit of Eq 2 was compared with the fit of the same model omitting the QTL term (the null model) by an F-test, leading to a nominal p-value, reported as the $\text{logP} = -\log_{10}(p\text{-value})$.

QTL detection: local (L), chromosome-wide (C), and genome-wide (G)

QTL detections were declared according to three distinct protocols of varying stringency and emphasis. The first protocol, termed Analysis L, was concerned only with detection of "local QTL", that is, QTL located at or close to the relevant expressed gene or accessible chromatin region. The second protocol, Analysis C, broadened the search for QTL to anywhere on the chromosome on which the trait is located; the greater number of loci considered meant that the criterion used to call detected QTL for this protocol was more stringent than Analysis L. The third protocol, Analysis G, further broadened the search to the entire genome with the most stringent detection criterion. These protocols used two types of multiple test correction: a permutation-based control of the family-wise error rate (FWER) and the Benjamini-Hochberg False Discovery Rate (FDR; [31]). Below we describe in detail the permutation procedure and then the three protocols.

Permutation-based error rate control: chromosome- and genome-wide

The FWER was controlled based on permutations specific to each trait. 1,000 permutations of the sample index were recorded, and then genome scans performed for each trait, using the same permutation orderings. Given a trait, the maximum logP from either the entire genome or the chromosome for each permutation was collected and used to fit a null generalized extreme value distribution (GEV) in order to control the genome- and chromosome-wide error rates, respectively [44]. Error rates were controlled by calculating a p-value for each QTL based on the respective cumulative density functions from the GEVs: $\text{permP} = 1 - F_{\text{GEV}}(\text{logP})$, where F_{GEV} is the cumulative density function of the GEV. We denote genome- and chromosome-wide error rate controlling p-values as permP_{G} and permP_{C} respectively.

The appropriateness of permutation-derived thresholds [45] is dependent on the CC having balanced founder haplotype contributions. This assumption was supported by simulations of the funnel breeding design [37]. More recently, power simulations based on the observed CC strain genomes were performed [42], which found low levels of population structure for highly polygenic genetic architectures. Nevertheless, we use permutations given that molecular traits often have strong effect QTL that are detectable in the presence of subtle population structure.

Local analysis - Analysis L

Our detection criteria for local-QTL leverages our strong prior belief in local genetic regulation. For a given trait, this local analysis involved examining QTL associations at all loci within a 10Mb window of the gene TSS or chromatin region midpoint. A QTL was detected if the $\text{permP}_{\text{C}} < 0.05$. Notably, we can also check whether the corresponding $\text{permP}_{\text{G}} < 0.05$, as an additional characterization of statistical significance of detected local-QTL.

Chromosome-wide analysis - Analysis C

Chromosome-wide analysis involved examining QTL associations at all loci on the chromosome harboring the gene or chromatin region in question. The peak logP is input into a chromosome-wide GEV, producing a permP_{C}, which are further subjected to adjustment by FDR to account for multiple testing across the traits [46]. Compared to Analysis L, this procedure is more stringent with respect to local-QTL because of the additional FDR adjustment. In addition, it can detect distal-QTL outside the local region of the trait. Note, since only the most significant QTL is recorded, this procedure will disregard local-QTL if a stronger distal-QTL on the chromosome is observed.

Genome-wide analysis - Analysis G

The genome-wide analysis is largely equivalent to Analysis C but examines all genomic loci, while an FDR adjustment is made to the genome-wide permP_{G}. Unlike Analysis C, it incorporates additional scans conditioned on detected QTL to potentially identify multiple QTL per trait (*i.e.* both local- and distal-QTL). Briefly, after a QTL is detected, a subsequent scan (and permutations) is performed in which the previously detected QTL is included in both the null and alternative models, allowing for additional independent QTL to be detected with FDR control. See S2 Appendix for greater detail on this conditional scan procedure and a clear example in Fig S21 Fig. Notably, the local/distal status of the QTL does not factor into Analysis G.

Analyses L, C, and G should detect many of the same QTL, specifically strong local-QTL. Collectively, they allow for efficient detection of QTL with varying degrees of statistical support while strongly leveraging local status.

QTL effect size

The effect size of a detected QTL was defined as the R^2 attributable to the QTL term; specifically, as $1 - \mathrm{RSS}_{\mathrm{QTL}}/\mathrm{RSS}_0$, where $\mathrm{RSS} = \sum_{i=1}^{n}(y_i - \widehat{y_i})^2$ is the residual sum of squares, *i.e.* the sum of squares around predicted value $\widehat{y_i}$, and $\mathrm{RSS}_{\mathrm{QTL}}$ and RSS_0 denote the RSS calculated for the QTL and null models respectively. We also calculated a more conservative QTL effect size estimate with a QTL random effects model to compare with the R^2 estimate.

Stable estimates of founder haplotype effects

The fixed effects QTL model used for mapping, though powerful for detecting associations, is sub-optimal for providing stable estimates of the haplotype effects vector, $\boldsymbol{\beta}$ (calculated as $(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$, where $\mathbf{X}$ is the full design matrix). This is because, among other things, 1) the matrix of haplotype dosages that forms the design matrix of $\{\mathrm{QTL}_i\}_{i=1}^{n}$ in Eq 2, is multi-collinear, which leads to instability, and 2) because the number of observations for some haplotypes will often be few, leading to high estimator variance [47]. More stable estimates were therefore obtained using shrinkage. At detected QTL, the model was refit with $\boldsymbol{\beta}$ modeled as a random effect, $\boldsymbol{\beta} \sim \mathrm{N}(\mathbf{0}, \mathbf{I}\tau^2)$ [48], to give an 8-element vector of the best linear unbiased estimates (BLUPs; [49]), $\widetilde{\boldsymbol{\beta}} = (\widetilde{\beta}_{\mathrm{AJ}}, \ldots, \widetilde{\beta}_{\mathrm{WSB}})^{\mathrm{T}}$. These BLUPs, after being centered and scaled, were then used for further comparison of QTL across tissues.

Comparing QTL effects across tissues

To summarize patterns of the genetic regulation of gene expression and chromatin accessibility across tissues, we calculated correlations between the founder haplotype effects of QTL that map to approximately the same genomic region of the genome for the same traits but in different tissues. For cross-tissue pairs of local-QTL, we required that both be detected within the 20Mb window around the gene TSS or the chromatin window midpoint. For cross-tissue pairs of distal-QTL, the QTL positions had to be within 10Mb of each other. All detected QTL were considered, including QTL from Analyses G and C controlled at an FDR of 0.2, allowing for consistent signal across tissues to provide further evidence of putative QTL with only marginal significance in a single tissue.

For a pair of matched QTL j and k from different tissues, we calculated the Pearson correlation coefficient of their haplotype effects, estimated as BLUPs, $r_{jk} = \mathrm{cor}(\widetilde{\boldsymbol{\beta}}_j, \widetilde{\boldsymbol{\beta}}_k)$. Since each $\widetilde{\boldsymbol{\beta}}$ is an 8-element vector, the corresponding r are distributed such that $r\sqrt{6}(1-r^2)^{-1} \sim t_6$ according to the null model of independent variables. Testing alternative models of $r_{jk} > 0$ and $r_{jk} < 0$ produced two p-values per pair of QTL. These were then subject to FDR control [31] to give two q-values, $q_{ij}^{\{r>0\}}$ and $q_{ij}^{\{r<0\}}$. These were then used to classify cross-tissue pairs of QTL as being significantly correlated or anti-correlated, respectively.

Variant association

Variant association has been used previously in MPP (*e.g.*, [50, 51]), and uses the same underlying model as the haplotype-based mapping (Eq 2), with the QTL term now representing imputed dosages of the minor allele. Variant association can more powerfully detect QTL than haplotype mapping if the simpler variant model is closer to the underlying biological mechanism [52], though it will struggle to detect multi-allelic QTL.

Variant genotypes from mm10 (NCBI38) were obtained using the ISVdb [53] for the CC strains, which were converted to mm9 coordinates with the liftOver tool [54]. Variants were filtered out if their minor allele frequencies ≤ 0.1 or they were not genotyped in one of the CC founder strains to avoid false signals. Tests of association at individual SNPs (variant association) were performed within the local windows of genes with local-QTL detected in more than one tissue. Genes with multiple tissue-specific local-eQTL potentially reflect different functional variants, and could potentially have less consistent patterns of variant association compared to variants that are functionally active in multiple tissues.

Mediation analysis

Mediation analysis has recently been used with genomic data, including in humans (*e.g.*, [10]) and rodents (*e.g.*, [51, 55]), to identify and refine potential intermediates of causal paths underlying phenotypes. We use a similar approach to mediation analyses used with the DO [19, 56] to detect mediators of eQTL effects on gene expression. In our study, a mediator is a molecular trait. either chromatin accessibility or the expression of another gene, that explains a portion of the association between a genetic locus and a gene's expression levels Fig 6.

Statistically significant candidate mediators are identified using a genome-wide scan. Similar to the QTL mapping genome scan described in Eq 2, the mediation scan involves a comparison between alternative and null models at loci across the genome. The alternative model is

$$y_i^{\mathrm{G}_j} = \mu + \mathrm{eQTL}_i^{\mathrm{G}_j} + m_{ik} + \mathrm{batch}_{b[i]} + \varepsilon_i, \tag{3}$$

which is compared with the null model

$$y_i^{\mathrm{G}_j} = \mu + m_{ik} + \mathrm{batch}_{b[i]} + \varepsilon_i$$

where $y_i^{\mathrm{G}_j}$ is individual i's expression level for gene j, which possesses an eQTL, μ is the intercept term, $\mathrm{eQTL}_i^{\mathrm{G}_j}$ is the eQTL effect on gene j, m_{ik} is the effect of the kth mediator, either chromatin accessibility or gene expression depending on the mediation model being evaluated, batch_b is the effect of the sequencing batch b, $b[i]$ is the batch associated with individual i, and ε_i is random noise. Conditioned loci can be included for certain eQTL, but not shown here for clarity. In comparison to the QTL genome scan in which the QTL effect is fit at positions across the genome, the mediation scan has the locus fixed at the eQTL, and the mediator is fit for candidates across the genome.

Because the eQTL is always included in the alternative model but not the null model, the logP of the mediation scan should fluctuate around the observed eQTL logP. At mediators that possess some or all of the information present in the eQTL, we will observe a reduced logP. Significant logP reductions represent candidate mediators. Similar to the QTL mapping procedures, FWER was controlled by performing mediation scans on 1,000 permutations of the mediator variable to characterize GEVs from the minimum logP, that is, $\mathrm{permP}^m = F_{\mathrm{GEV}}(\mathrm{logP})$, for both chromosome- and genome-wide levels. See S3 Appendix for discussion on mediation analysis, including the permutation approach and the formal criterion for mediation detection.

Detection of mediation is dependent on a number of assumptions about the underlying variables, their relationships, and the directionality of those relationships [4]. Many of these cannot be satisfied in a system as complex as chromatin accessibility and transcriptional regulation in whole living organisms. Nevertheless, signals in the data that are consistent with the mediation model represent putative causal factors that regulate gene expression.

Software

All statistical analyses were conducted with the R statistical programming language [57]. The R package miQTL was used for all the mapping and mediations analyses, and is available on GitHub at `https://github.com/gkeele/miqtl`.

Supporting information

S1 Fig. Principle components analysis identifies tissue type as key source of variation for gene expression and chromatin accessibility. Molecular traits for liver (pink), lung (green), and kidney (orange) tissue samples were derived from RNA-seq and ATAC-seq data. Principal components (PC) 1 and 2 capture a majority of the variation and show a greater amount of between tissue variability than within tissue variability. (PDF)

S2 Fig. Concordance between differentially expressed genes and differentially accessible regions in between-tissue comparisons. Genes were categorized by the direction of the difference in expression and chromatin accessibility in their promoter regions. (PDF)

S3 Fig. Overlap across tissues of genes and chromatin windows used for QTL analysis. Sequence traits were filtered to remove outcomes more likely to cause in spurious QTL signal. Genes with TPM ≤ 1 and chromatin windows with TMP ≤ 5 for $\geq 50\%$ of samples were removed from analysis. After this filtering process, lung had the greatest number of traits analyzed, for both genes and chromatin windows, followed by kidney and then liver. (PDF)

S4 Fig. Overlap across tissues of genes (A) and chromatin windows (B) with local-QTL detected. The majority of sequence traits with a local-QTL detected were identified in only a single tissue. Kidney had the highest number of local-eQTL, whereas lung had the highest number of local-cQTL. Liver had a relative lack of local-cQTL, which may relate to its having the fewest chromatin windows analyzed (S3 FigB). Results included local-QTL detected with Analysis G (FDR < 0.1), Analysis C (FDR < 0.1), and Analysis L (genome-wide and chromosome-wide). (PDF)

S5 Fig. QTL mapping results using only Analysis G or Analysis C. QTL map plots of eQTL (A) and cQTL (B) with FDR controlled at 0.1 and 0.2 for liver, lung, and kidney. Detected QTL from Analysis G (multi-stage FDR) and Analysis C (chromosome-wide FDR) are included. Analysis C, which uses FDR control for chromosome-wide significant QTL, produces a large number of intra-chromosomal distal QTL. The y-axis represents the genomic position of the gene or chromatin site, and the x-axis represents the genomic position of the QTL. Local-QTL appear as dots along the diagonal. (PDF)

S6 Fig. Highly significant QTL map nearby the gene TSS and chromatin window midpoint. The permutation-based p-value (permP) from Analyses G (A) and C (B) for eQTL and cQTL by the distance (Mb) from the gene TSS and the midpoint of the chromatin site. Inter-chromosomal distal-QTL are not included. The red dashed lines represent 10Mb upstream and downstream of the gene TSS or the midpoint of the chromatin site for classifying QTL as local or distal. Significant signals (yellow or blue), based on FDR < 0.1, are largely local. Analysis C detects many more intra-chromosomal distal-QTL. (PDF)

S7 Fig. QTL effect sizes by local/distal status. Each dot represents a QTL detected through either Analyses G (A) or C (B) with FDR < 0.1. The three horizontal bars represent the 25th, 50th, and 75th quantiles of QTL effect sizes for all local-QTL per tissue. More local-eQTL are detected and have higher effects than distal-QTL. Analysis C detects a large number of intra-chromosomal distal-QTL that Analysis G does not, many of which have low effect sizes. Effect size estimates are based on a fixed effects model. (PDF)

S8 Fig. CAST and PWK haploytpes have more extreme haplotype effects for eQTL and cQTL compared with the other strains. Haplotype effects were estimated as BLUPs, which are constrained and centered around 0. Each QTL is represented by an 8-element effect vector. Founders with more extreme effects are identified by comparing the absolute values of effects. Haploytpe effects for eQTL (A() are are similar to cQTL (B). Trends are unstable in distal-cQTL because so few are identified. (PDF)

S9 Fig. Effect sizes between cross-tissue QTL pairs are lowly but significantly correlated. Comparisons of QTL effects sizes between (liver/lung) are in the left column, (liver/kidney) middle column, and (lung/kidney) right column. eQTL are yellow and cQTL are blue. Local-eQTL are plotted in the top row, distal-eQTL in the second row, local-cQTL in the third row, and distal-cQTL in the bottom row, with only four pairs detected in (lung/kidney). (PDF)

S10 Fig. Consistent genetic regulation of gene expression and chromatin accessibility observed across tissues. Enrichment of significantly correlated founder haplotype effects detected in QTL pairs for gene expression and chromatin accessibility. Pairs of QTL observed in multiple tissues were defined for local-eQTL (left column), distal-eQTL (middle column), and local-cQTL (right column). Only four pairs of distal-cQTL were observed, all shared between lung and kidney. A right-tailed test the correlation between founders effects ($H_A : r > 0$) was performed for each QTL pair, producing p-values that were then FDR adjusted. Null simulations of uncorrelated 8-element vector pairs for each class of QTL and pairwise tissue comparison emphasize the observed enrichment in correlated founder effects between QTL pairs. (PDF)

S11 Fig. Cross-tissue QTL pairs with highly correlated founder haplotype effects map proximally to each other. Founder effects were estimated as constrain BLUPs. Pairwise correlations of the 8-element effect vectors were calculated for QTL pairs, and plotted again the distance between the QTL coordinates inMb for (liver/lung) in the left column, (liver/kidney) in the middle column, and (lung/kidney) in the right column. Single eQTL and cQTL pairs are represented as a yellow and blue dots, respectively. Local-eQTL are shown in the top row, distal-eQTL in the second row, local-cQTL in the third row, and distal c-QTL in the bottom row, for which only four pairs were detected in (lung/kidney). (PDF)

S12 Fig. The gene *Ubc* has consistent strong local-eQTL observed in the three tissues. The local-eQTL consistently drove higher expression when the B6, NOD, NZO, and WSB haploytpes were present. Expression levels in liver and lung were found to be significantly different ($q = 0.022$). The estimated haplotype effects, scaled BLUPs, were highly consistent with the expression data, represented as interquartile bars by most likely diplotype. The haplotype and variant associations in the eQTL regions were similar across tissues, suggesting they may represent the same causal origin. The red tick represents the *Ubc* TSS, the black tick represents the peak variant

association, and the colored ticks represent the peak haplotype association, each each tissue. (PDF)

S13 Fig. The gene *Rnf13* has tissue-specific patterns of genetic regulation. A strong local-eQTL detected in liver. After conditioning on the local-eQTL, a statistically significant (Analysis G) distal-eQTL was detected on chromosome 12, largely driven by the B6 haplotype, distinct from the local-eQTL. The unique haplotype effect patterns for each eQTL can be seen in both the expression data, represented by interquartile bars by most likely diplotype, and the estimated effects (scaled BLUPs). On the association scans, the red tick marks the *Rnf13* TSS and the black tick marks the location of distal-eQTL. (PDF)

S14 Fig. Co-localizing eQTL and cQTL are not sufficient for statistical mediation. The approach used to detect mediation through chromatin accessibility requires that the eQTL and cQTL co-localize (both within 10Mb of the gene TSS), as well as possess similar founder haplotype effect patterns. Co-localizing cQTL are observed for local-eQTL for both *Hdhd3* in liver (A) and *Acbd4* in kidney (B). QTL and mediation scans are shown, with chromosomes 4 and 11 blown up for *Hdhd3* and *Acbd4*, respectively. The red ticks denote the TSS for both genes. The founders effects were estimated as centered and scaled BLUPs. The founder effects for the eQTL and cQTL are highly correlated ($r = 0.96$) for *Hdhd3*, but not for *Acbd4* ($r = 0.52$). Strong mediation of the *Hdhd3* eQTL through chromatin is detected, but not for *Acbd4*. The effect size of the proximal cQTL to *Acbd4* is smaller than its eQTL, also inconsistent with the relationship depicted in Fig 6A. (PDF)

S15 Fig. Mediation of *Ccnyl1* distal-eQTL through *Zfp979* expression. Expression of *Ccnyl1* and *Zfp979* are correlated ($r = 0.72$) in lung, which is also observed in the expression data categorized by diplotype and the founder effects estimated as scaled BLUPs. The distal-eQTL on chromosome 4 for *Ccnyl1* corresponds closely to local-eQTL of *Zfp979*. *Ccnyl1* is located on chromosome 1, indicated by the red tick. *Zfp979* and *Zfp985*, both zinc finger proteins likely with DNA binding properties, are identified as strong candidate mediators of the distal-eQTL at genome-wide significance. The correlations, magnitude of effects, and mediation are consistent with the simple relationship depicted in the graph. The distal-eQTL and candidate mediators are located in a region of interest that regulates *Akr1e1* expression. (PDF)

S16 Fig. Mediation of *Akr1e1* distal-eQTL through *Zfp985* expression. Expression of *Akr1e1* and *Zfp985* are anti-correlated ($r = -0.69$) in lung. This relationship is also observed in the expression data with bars representing the interquartile range, categorized by most likely diplotype, and the founder haplotype effects, estimated as scaled BLUPs. The QTL and mediation scans reveal that *Akr1e1*, TSS marked with a red tick on chromosome 13, possesses a distal-eQTL on chromosome 4 that is proximal to the strong local-eQTL of *Zfp985*. The mediation scan identifies *Zfp985* as a strong candidate mediator consistent with the mediation model. A more complete picture of the genetic regulation of *Akr1e1* expression is pieced together by looking across all three tissues and includes a potential chromatin mediator (Fig 9). (PDF)

S17 Fig. Confirmation of *Akr1e1* distal-eQTL and mediation by *Zfp985* in kidney tissue of Diversity Outbred mice. A genome-wide significant distal-eQTL was detected for *Akr1e1* in liver, lung (shown here), and kidney tissues

from 47 CC strains. In a larger sample of kidney tissue from outbred DO mice, the same distal-eQTL and mediation relationship were observed. As expected, the larger sample results in greater statistical significance, and confirms that the NOD effect is more strongly negative than NZO, PWK, and WSB, which the effects plots the *Zfp985* local-eQTL suggested. Notably, *Zfp985* was not tested in the CC kidney because of low levels, though the distal-eQTL for *Akr1e1* is consistent with its activity, which is here confirmed in the DO. (PDF)

S18 Fig. Observed relationships across the three tissues related to the genetic regulation of *Akr1e1* expression. The model for the distal genetic regulation of *Akr1e1* expression, described in Figure 9, was reconstructed from these observed relationships. Solid arrows were observed, whereas dashed arrows are assumed. Effect sizes represent the proportion of variance explained by the QTL and mediation p-values (permP) were defined using a permutation procedure. The assumed relationships are supported by the presence of the distal-eQTL in all three tissues. The *Zfp985* mediator relationship in kidney, though not observed in the CC, was observed in the related DO population. (PDF)

S19 Fig. Local-QTL effect sizes by mapping analysis. Based on ranking mapping analyses with respect to the extent of scope, local (L; magenta) to chromosome (C; plum) to genome-wide (G; cyan), the greater the scope corresponded to reduced power to detect QTL, shown in liver, lung, and kidney tissues for gene expression (yellow line) and chromatin accessibility (blue line). Each dot represents a detected local-QTL, colored according to the highest scope mapping procedure that detected it. The three horizontal bars represent the 25th, 50th, and 75th quantiles of QTL effect sizes for all local-QTL per tissue. Analysis G generally detects QTL with effect size $>$ 60%, whereas Analyses C and L detect QTL effect sizes $>$ 45%. Effect size estimates correspond to a fixed effects model of the QTL. (PDF)

S20 Fig. Comparison of QTL effect sizes estimates from fixed effects and random effects models. The effect size corresponding to the random effect fit is harshly penalized compared to the fixed effect estimate, likely due to a small sample size of 47 individuals. Notably, there are a number of distal-eQTL that are more harshly reduced by the random effects model compared to the other QTL, likely representing signals resulting from extreme observations or imbalances in founder contributions at the locus. QTL detected by Analysis G (FDR ≤ 0.1), 2 (FDR ≤ 0.1), and 3 are shown. (PDF)

S21 Fig. Detection of local-eQTL after conditioning on distal-eQTL for *Gpn3*. The multi-stage conditional regression approach of Analysis G allows for the detection of multiple genome-wide significant QTL, which can then be appropriately incorporated into a FDR procedure across many outcomes. In this example in lung tissue, the gene *Gpn3* initially has a strong distal-eQTL on chromosome 8 (A). Though a peak is detected near the TSS of *Gpn3*, it does not meet genome-wide significance. However, after conditioning on the distal-eQTL, the local-eQTL is detected (B). Horizontal dashed lines represent empirical 95% significance thresholds based on 1,000 permutations. (PDF)

S1 Appendix. CC strains used in study. (PDF)

S2 Appendix. Detailed description of conditional genome-wide scans (Analysis G). (PDF)

S3 Appendix. Detailed description of mediation analysis. (PDF)

S1 Table. Number of differentially expressed genes and accessible chromatin regions detected between liver, lung, and kidney tissues. (PDF)

S2 Table. Number of genes with eQTL detected in liver, lung, and kidney tissues at FDR < 0.1. (PDF)

S3 Table. Number of genes with eQTL detected in liver, lung, and kidney tissues at FDR < 0.2. (PDF)

S4 Table. Number of chromatin accessibility sites with cQTL detected in liver, lung, and kidney tissues at FDR < 0.1. (PDF)

S5 Table. Number of chromatin accessibility sites with cQTL detected in liver, lung, and kidney tissues at FDR < 0.2. (PDF)

S6 Table. Number of genes with chromatin mediation of local-eQTL in liver, lung, and kidney tissues. (PDF)

S7 Table. Genes with distal-eQTL with gene mediators detected in lung and kidney tissues. (PDF)

Acknowledgments

This work was funded by grants from the National Institute of Environmental Health Sciences (NIEHS) (R01-ES023195 to IR and GEC and P30-ES025128), with GRK and WV funded by grants from the National Institute of General Medical Sciences (NIGMS) (R01-GM104125 and R35-GM127000 to WV). From UNC, we thank Lucas T. Laudermilk for his insights on the functional mechanisms of zinc finger protein-induced gene silencing, and Samir N. P. Kelada and Lauren J. Donoghue for reading the manuscript and providing thoughtful suggestions. We thank Fred A. Wright of North Carolina State University for helpful discussions on eQTL and mediation analyses. We thank Gary A. Churchill of The Jackson Laboratory for discussions on mediation analysis that influenced this work. Additionally, GRK was partially supported through funding from Dr. Churchill and The Jackson Laboratory.

Author Contributions

Conceptualization: Ivan Rusyn, Terrence S. Furey, Gregory E. Crawford, Gregory R. Keele, William Valdar
Data Curation: Jennifer W. Israel, Jeremy M. Simon, Paul Cotney
Formal Analysis: Gregory R. Keele, Bryan C. Quach
Funding Acquisition: Terrence S. Furey, Ivan Rusyn, Gregory E. Crawford
Investigation: Grace A Chappell, Lauren Lewis, Alexias Safi,
Methodology: Gregory R. Keele, William Valdar
Project Admin: Terrence S. Furey, Ivan Rusyn, Gregory E. Crawford, William Valdar
Software: Gregory R. Keele, Bryan C. Quach
Supervision: Terrence S. Furey, Ivan Rusyn, William Valdar, Gregory E. Crawford
Visualization: Gregory R. Keele, Bryan C. Quach

Writing – Original: Gregory R. Keele, Bryan C. Quach
Writing – Review: Terrence S. Furey, William Valdar, Ivan Rusyn

References

1. Schadt EE. Molecular networks as sensors and drivers of common human diseases. Nature. 2009;461(7261):218–223. doi:10.1038/nature08454.

2. Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. Nature Reviews Genetics. 2018;19(8):491–504. doi:10.1038/s41576-018-0016-z.

3. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology. 1986;51(6):1173–1182. doi:10.1037/0022-3514.51.6.1173.

4. MacKinnon DP, Fairchild AJ, Fritz MS. Mediation Analysis. Annual Review of Psychology. 2007;58(1):593–614. doi:10.1146/annurev.psych.58.110405.085542.

5. Richmond RC, Hemani G, Tilling K, Davey Smith G, Relton CL. Challenges and novel approaches for investigating molecular mediation. Human Molecular Genetics. 2016;25(R2):R149–R156. doi:10.1093/hmg/ddw197.

6. Degner JF, Pai AA, Pique-Regi R, Veyrieras JB, Gaffney DJ, Pickrell JK, et al. DNase I sensitivity QTLs are a major determinant of human expression variation. Nature. 2012;482(7385):390–4. doi:10.1038/nature10808.

7. Pai AA, Pritchard JK, Gilad Y. The genetic and mechanistic basis for variation in gene regulation. PLoS Genetics. 2015;11(1):e1004857. doi:10.1371/journal.pgen.1004857.

8. Battle A, Khan Z, Wang SH, Mitrano A, Ford MJ, Pritchard JK, et al. Genomic variation. Impact of regulatory variation from RNA to protein. Science (New York, NY). 2015;347(6222):664–7. doi:10.1126/science.1260793.

9. Yang F, Wang J, GTEx Consortium, Pierce BL, Chen LS. Identifying cis-mediators for trans-eQTLs across many human tissues using genomic mediation analysis. Genome Research. 2017;27(11):1859–1871. doi:10.1101/gr.216754.116.

10. Battle A, Mostafavi S, Zhu X, Potash JB, Weissman MM, McCormick C, et al. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. Genome Research. 2014;24(1):14–24. doi:10.1101/gr.155192.113.

11. Alasoo K, Rodrigues J, Mukhopadhyay S, Knights AJ, Mann AL, Kundu K, et al. Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. Nature Genetics. 2018; p. 102392. doi:10.1038/s41588-018-0046-7.

12. Roytman M, Kichaev G, Gusev A, Pasaniuc B. Methods for fine-mapping with chromatin and expression data. PLoS Genetics. 2018;14(2):e1007240. doi:10.1371/journal.pgen.1007240.

13. Wu Y, Zeng J, Zhang F, Zhu Z, Qi T, Zheng Z, et al. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. Nature Communications. 2018;9(1):918. doi:10.1038/s41467-018-03371-0.

14. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. American Journal of Human Genetics. 2017;101(1):5–22. doi:10.1016/j.ajhg.2017.06.005.

15. Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, et al. The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nature Genetics. 2004;36(11):1133–1137. doi:10.1038/ng1104-1133.

16. Collaborative Cross Consortium. The genome architecture of the Collaborative Cross mouse genetic reference population. Genetics. 2012;190(2):389–401. doi:10.1534/genetics.111.132639.

17. Srivastava A, Morgan AP, Najarian ML, Sarsani VK, Sigmon JS, Shorter JR, et al. Genomes of the mouse collaborative cross. Genetics. 2017;206(2):537–556. doi:10.1534/genetics.116.198838.

18. Churchill G, Gatti D, Munger S, Svenson K. The Diversity outbred mouse population. Mammalian Genome. 2012;23:713–8. doi:10.1007/s00335-012-9414-2.The.

19. Chick JM, Munger SC, Simecek P, Huttlin EL, Choi K, Gatti DM, et al. Defining the consequences of genetic variation on a proteome-wide scale. Nature. 2016;534(7608):500–5. doi:10.1038/nature18270.

20. Skelly DA, Czechanski A, Byers C, Aydin S, Spruce C, Olivier C, et al. Genetic variation influences pluripotent ground state stability in mouse embryonic stem cells through a hierarchy of molecular phenotypes. bioRxiv. 2019;doi:10.1101/552059.

21. Kaeppler SM. Quantitative trait locus mapping using sets of near-isogenic lines: Relative power comparisons and technical considerations. Theoretical and Applied Genetics. 1997;95(3):384–392. doi:10.1007/s001220050574.

22. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. In: Current Protocols in Molecular Biology. vol. 2015. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2015. p. 21.29.1–21.29.9. Available from: `http://doi.wiley.com/10.1002/0471142727.mb2129s109`.

23. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature Methods. 2013;10(12):1213–8. doi:10.1038/nmeth.2688.

24. Shibata Y, Sheffield NC, Fedrigo O, Babbitt CC, Wortham M, Tewari AK, et al. Extensive evolutionary changes in regulatory element activity during human origins are associated with altered gene expression and positive selection. PLoS Genetics. 2012;8(6):e1002789. doi:10.1371/journal.pgen.1002789.

25. Quinlan AR, Hall IM. BEDTools : a flexible suite of utilities for comparing genomic features. 2010;26(6):841–842. doi:10.1093/bioinformatics/btq033.

26. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–140.

27. Fu CP, Welsh CE, de Villena FPM, McMillan L. Inferring ancestry in admixed populations using microarray probe intensities. In: Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine - BCB '12. New York, New York, USA: ACM Press; 2012. p. 105–112. Available from: `http://dl.acm.org/citation.cfm?id=2382936.2382950http://dl.acm.org/citation.cfm?doid=2382936.2382950`.

28. Morgan AP, Fu CP, Kao CY, Welsh CE, Didion JP, Yadgary L, et al. The Mouse Universal Genotyping Array: From Substrains to Subspecies. G3 (Bethesda, Md). 2016;6(2):263–279. doi:10.1534/g3.115.022087.

29. Shorter JR, Najarian ML, Bell TA, Blanchard M, Ferris MT, Hock P, et al. Whole Genome Sequencing and Progress Towards Full Inbreeding of the Mouse Collaborative Cross Population. G3: Genes, Genomes, Genetics. 2019;doi:10.1534/g3.119.400039.

30. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research. 2015;43(7):e47.

31. Benjamini Y, Hochberg Y, Society RS, Benajmini Y, Hochberg Y, Society RS, et al. Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society, Series B. 1995;57(1):289–300.

32. Sarkar SK, Chang CK. The Simes Method for Multiple Hypothesis Testing With Positively Dependent Test Statistics. Journal of the American Statistical Association. 1997;92(440):1601. doi:10.2307/2965431.

33. Xiong Q, Mukherjee S, Furey TS. GSAASeqSP: a toolset for gene set association analysis of RNA-Seq data. Scientific Reports. 2014;4:6347. doi:10.1038/srep06347.

34. Mott R, Talbot CJ, Turri MG, Collins AC, Flint J. A method for fine mapping quantitative trait loci in outbred animal stocks. Proceedings of the National Academy of Sciences. 2000;97(23):12649–54. doi:10.1073/pnas.230304397.

35. Haley CS, Knott SA. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity. 1992;69:315–24.

36. Martínez O, Curnow RN. Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. Theoretical and Applied Genetics. 1992;85(4):480–488. doi:10.1007/BF00222330.

37. Valdar W, Flint J, Mott R. Simulating the collaborative cross: power of quantitative trait loci detection and mapping resolution in large sets of recombinant inbred strains of mice. Genetics. 2006;172(3):1783–97. doi:10.1534/genetics.104.039313.

38. Aylor DL, Valdar W, Foulds-mathes W, Buus RJ, Verdugo RA, Baric RS, et al. Genetic analysis of complex traits in the emerging Collaborative Cross. Genome Research. 2011;21:1213–22. doi:10.1101/gr.111310.110.mentally.

39. Gralinski LE, Ferris MT, Aylor DL, Whitmore AC, Green R, Frieman MB, et al. Genome Wide Identification of SARS-CoV Susceptibility Loci Using the Collaborative Cross. PLoS Genetics. 2015;11(10):e1005504. doi:10.1371/journal.pgen.1005504.

40. Kelada SNP. Plethysmography Phenotype QTL in Mice Before and After Allergen Sensitization and Challenge. G3 (Bethesda, Md). 2016;6(9):2857–2865. doi:10.1534/g3.116.032912.

41. Donoghue LJ, Livraghi-Butrico A, McFadden KM, Thomas JM, Chen G, Grubb BR, et al. Identification of trans Protein QTL for Secreted Airway Mucins in Mice and a Causal Role for Bpifb1. Genetics. 2017;207(2):801–812. doi:10.1534/genetics.117.300211.

42. Keele GR, Crouse WL, Kelada SNP, Valdar W. Determinants of QTL Mapping Power in the Realized Collaborative Cross. G3 (Bethesda, Md). 2019;9(May):459966. doi:10.1534/g3.119.400194.

43. King EG, Merkes CM, McNeil CL, Hoofer SR, Sen S, Broman KW, et al. Genetic dissection of a model complex trait using the Drosophila Synthetic Population Resource. Genome Research. 2012;22(8):1558–66. doi:10.1101/gr.134031.111.

44. Dudbridge F, Koeleman BPC. Efficient Computation of Significance Levels for Multiple Associations in Large Studies of Correlated Data, Including Genomewide Association Studies. The American Journal of Human Genetics. 2004;75(3):424–435. doi:10.1086/423738.

45. Doerge R, Churchill G. Permutation tests for multiple loci affecting a quantitative character. Genetics. 1996;142(1977):285–94.

46. Chesler EJ, Lu L, Shou S, Qu Y, Gu J, Wang J, et al. Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. Nature Genetics. 2005;37(3):233–242. doi:10.1038/ng1518.

47. Zhang Z, Wang W, Valdar W. Bayesian modeling of haplotype effects in multiparent populations. Genetics. 2014;198(1):139–56. doi:10.1534/genetics.114.166249.

48. Wei J, Xu S. A random-model approach to QTL mapping in multiparent advanced generation intercross (MAGIC) populations. Genetics. 2016;202(2):471–486. doi:10.1534/genetics.115.179945.

49. Robinson GK. That BLUP is a Good Thing: The Estimation of Random Effects. Statistical Science. 1991;6(1):15–51. doi:10.2307/2246134.

50. Baud A, Guryev V, Hummel O, Johannesson M, Rat Genome Sequencing and Mapping Consortium, Flint J. Genomes and phenomes of a population of outbred rats and its progenitors. Scientific Data. 2014;1:140011. doi:10.1038/sdata.2014.11.

51. Keele GR, Prokop JW, He H, Holl K, Littrell J, Deal A, et al. Genetic Fine-Mapping and Identification of Candidate Genes and Variants for Adiposity Traits in Outbred Rats. Obesity. 2018;26(1):213–222. doi:10.1002/oby.22075.

52. Yalcin B, Flint J, Mott R. Using progenitor strain information to identify quantitative trait nucleotides in outbred mice. Genetics. 2005;171(2):673–81. doi:10.1534/genetics.104.028902.

53. Oreper D, Cai Y, Tarantino LM, de Villena FPM, Valdar W. Inbred Strain Variant Database (ISVdb): A Repository for Probabilistically Informed Sequence Differences Among the Collaborative Cross Strains and Their Founders. G3 (Bethesda, Md). 2017;7(6):1623–1630. doi:10.1534/g3.117.041491.

54. Lawrence M, Gentleman R, Carey V. rtracklayer: an R package for interfacing with genome browsers. Bioinformatics. 2009;25:1841–1842. doi:10.1093/bioinformatics/btp328.

55. Oreper D, Schoenrock SA, McMullan R, Ervin R, Farrington J, Miller DR, et al. Reciprocal F1 Hybrids of Two Inbred Mouse Strains Reveal Parent-of-Origin and Perinatal Diet Effects on Behavior and Expression. G3 (Bethesda, Md). 2018;8(11):3447–3468. doi:10.1534/g3.118.200135.

56. Keller MP, Gatti DM, Schueler KL, Rabaglia ME, Stapleton DS, Simecek P, et al. Genetic Drivers of Pancreatic Islet Function. Genetics. 2018;209(1):335–356. doi:10.1534/genetics.118.300864.

57. R Core Team. R: A Language and Environment for Statistical Computing; 2019. Available from: `https://www.R-project.org/`.

58. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics. 2017;33(18):2938–2940. doi:10.1093/bioinformatics/btx364.

59. Rieseberg LH, Archer MA, Wayne RK. Transgressive segregation, adaptation and speciation. Heredity. 1999;83(4):363–372. doi:10.1038/sj.hdy.6886170.

60. Yang H, Wang JR, Didion JP, Buus RJ, Bell TA, Welsh CE, et al. Subspecific origin and haplotype diversity in the laboratory mouse. Nature Genetics. 2011;43(7):648–55. doi:10.1038/ng.847.

61. Crowley JJ, Zhabotynsky V, Sun W, Huang S, Pakatci IK, Kim Y, et al. Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance. Nature Genetics. 2015;47(4):353–60. doi:10.1038/ng.3222.

62. Braccini L, Ciraolo E, Campa CC, Perino A, Longo DL, Tibolla G, et al. PI3K-C2γ is a Rab5 effector selectively controlling endosomal Akt2 activation downstream of insulin signalling. Nature Communications. 2015;6(1):7400. doi:10.1038/ncomms8400.

63. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize implements and enhances circular visualization in R. Bioinformatics. 2014;30(19):2811–2812. doi:10.1093/bioinformatics/btu393.

64. Hamilton-Williams EE, Wong SBJ, Martinez X, Rainbow DB, Hunter KM, Wicker LS, et al. Idd9.2 and Idd9.3 Protective Alleles Function in CD4+ T-Cells and Nonlymphoid Cells to Prevent Expansion of Pathogenic Islet-Specific CD8+ T-Cells. Diabetes. 2010;59(6):1478–1486. doi:10.2337/db09-1801.

65. Ecco G, Imbeault M, Trono D. KRAB zinc finger proteins. Development. 2017;144(15):2719–2729. doi:10.1242/dev.132605.

66. Hamilton-Williams EE, Rainbow DB, Cheung J, Christensen M, Lyons PA, Peterson LB, et al. Fine mapping of type 1 diabetes regions Idd9.1 and Idd9.2 reveals genetic complexity. Mammalian Genome. 2013;24(9-10):358–75. doi:10.1007/s00335-013-9466-y.

67. Aguet F, Brown AA, Castel SE, Davis JR, He Y, Jo B, et al. Genetic effects on gene expression across human tissues. Nature. 2017;550(7675):204–213. doi:10.1038/nature24277.

68. Huang GJ, Shifman S, Valdar W, Johannesson M, Yalcin B, Taylor MS, et al. High resolution mapping of expression QTLs in heterogeneous stock mice in multiple tissues. Genome Research. 2009;19(6):1133–1140. doi:10.1101/gr.088120.108.

69. Fu J, Wolfs MGM, Deelen P, Westra HJ, Fehrmann RSN, Te Meerman GJ, et al. Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression. PLoS Genetics. 2012;8(1):e1002431. doi:10.1371/journal.pgen.1002431.

70. Sul JH, Han B, Ye C, Choi T, Eskin E. Effectively Identifying eQTLs from Multiple Tissues by Combining Mixed Model and Meta-analytic Approaches. PLoS Genetics. 2013;9(6):e1003491. doi:10.1371/journal.pgen.1003491.

71. Flutre T, Wen X, Pritchard J, Stephens M. A statistical framework for joint eQTL analysis in multiple tissues. PLoS Genetics. 2013;9(5):e1003486. doi:10.1371/journal.pgen.1003486.

72. Acharya CR, McCarthy JM, Owzar K, Allen AS. Exploiting expression patterns across multiple tissues to map expression quantitative trait loci. BMC Bioinformatics. 2016;17(1):257. doi:10.1186/s12859-016-1123-5.

73. Rutledge H, Aylor DL, Carpenter DE, Peck BC, Chines P, Ostrowski LE, et al. Genetic regulation of Zfp30, CXCL1, and neutrophilic inflammation in murine lung. Genetics. 2014;198(2):735–45. doi:10.1534/genetics.114.168138.

74. Kelada SNP, Carpenter DE, Aylor DL, Chines P, Rutledge H, Chesler EJ, et al. Integrative genetic analysis of allergic inflammation in the murine lung. American Journal of Respiratory Cell and Molecular Giology. 2014;51(3):436–45. doi:10.1165/rcmb.2013-0501OC.

75. Xu S. Theoretical basis of the Beavis effect. Genetics. 2003;165(4):2259–68. doi:10.1534/genetics.114.161968.

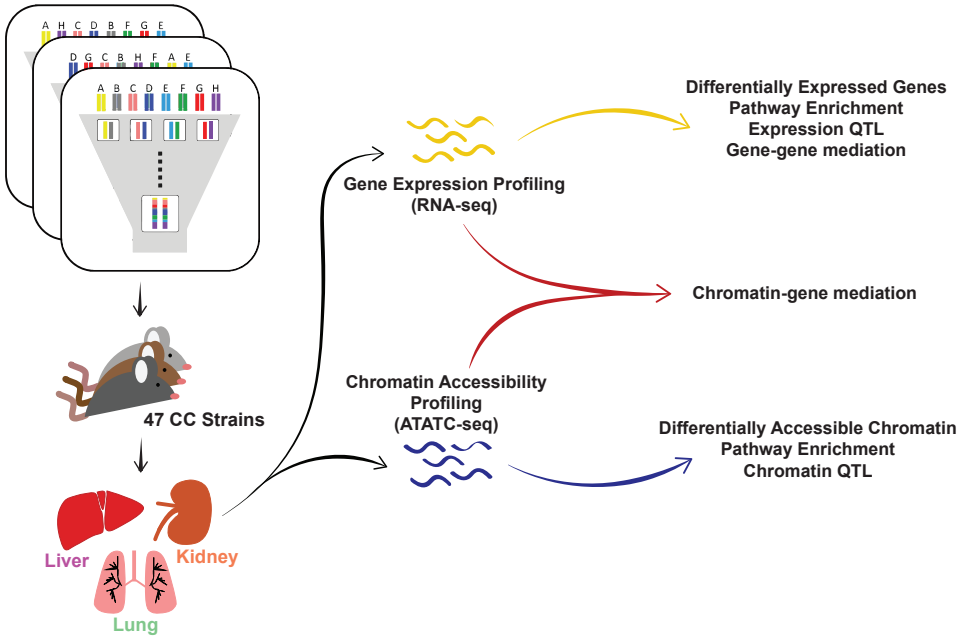
A H C D B F G E
D G C B H F A E
A B C D E F G H
47 CC Strains
Liver
Kidney
Lung
Gene Expression Profiling
(RNA-seq)
Differentially Expressed Genes
Pathway Enrichment
Expression QTL
Gene-gene mediation
Chromatin-gene mediation
Chromatin Accessibility
Profiling
(ATATC-seq)
Differentially Accessible Chromatin
Pathway Enrichment
Chromatin QTL

Fig 1

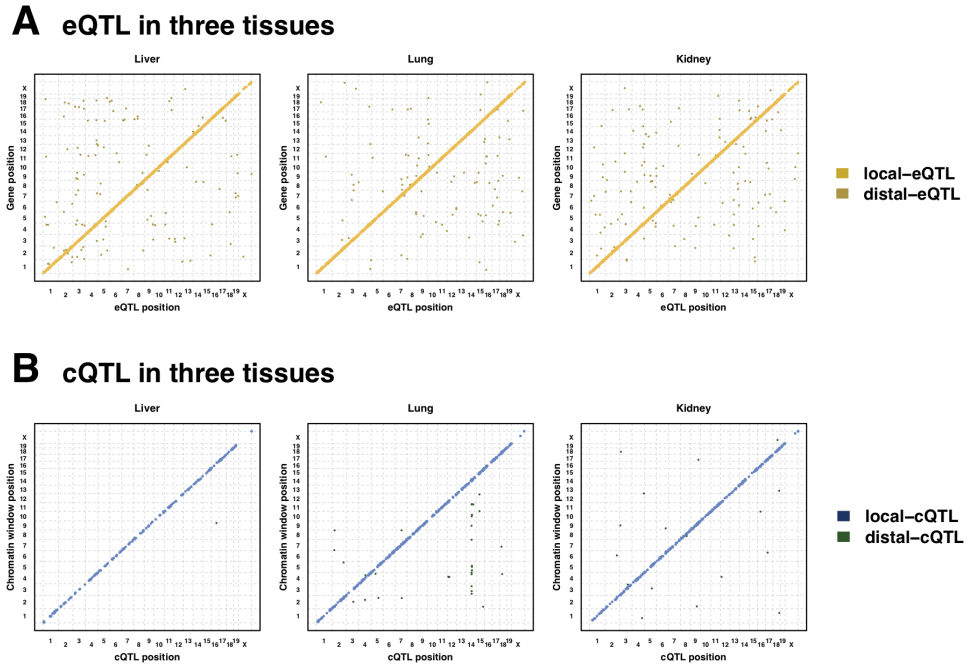
A eQTL in three tissues
Liver
Lung
Kidney
Gene position
eQTL position
local–eQTL
distal–eQTL
B cQTL in three tissues
Liver
Lung
Kidney
Chromatin window position
cQTL position
local–cQTL
distal–cQTL

Fig 2

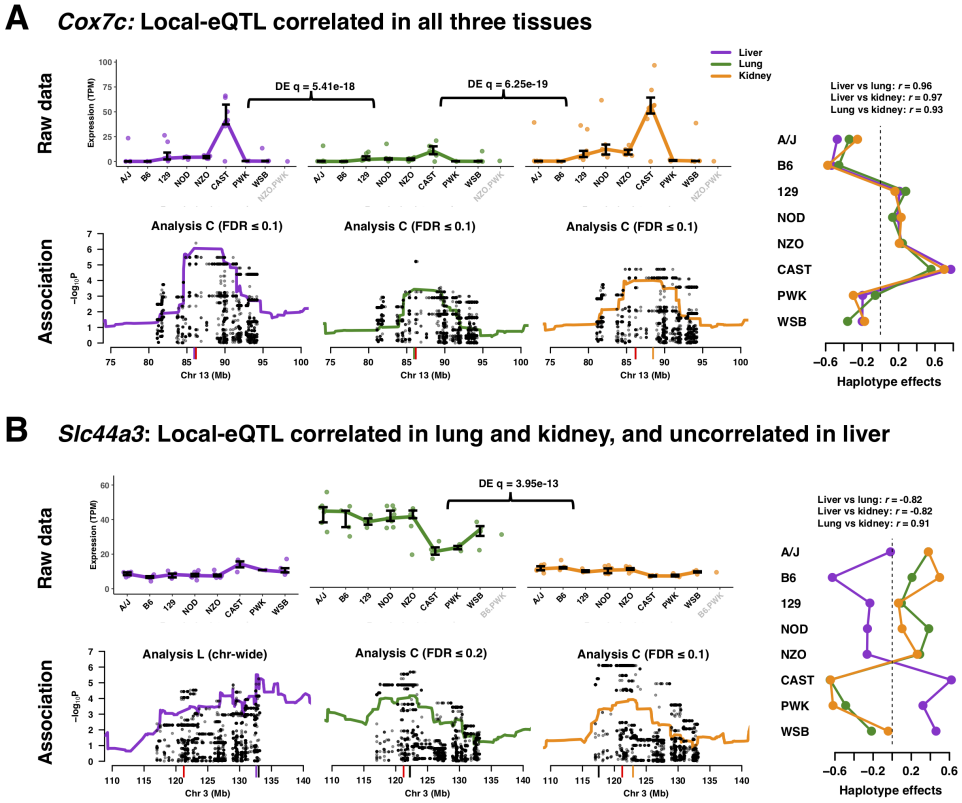
A
Cox7c: Local-eQTL correlated in all three tissues
Raw data
Association
DE q = 5.41e-18
DE q = 6.25e-19
Liver
Lung
Kidney
Liver vs lung: r = 0.96
Liver vs kidney: r = 0.97
Lung vs kidney: r = 0.93
Analysis C (FDR ≤ 0.1)
Chr 13 (Mb)
Haplotype effects
B
Slc44a3: Local-eQTL correlated in lung and kidney, and uncorrelated in liver
DE q = 3.95e-13
Liver vs lung: r = -0.82
Liver vs kidney: r = -0.82
Lung vs kidney: r = 0.91
Analysis L (chr-wide)
Analysis C (FDR ≤ 0.2)
Analysis C (FDR ≤ 0.1)
Chr 3 (Mb)

Fig 3

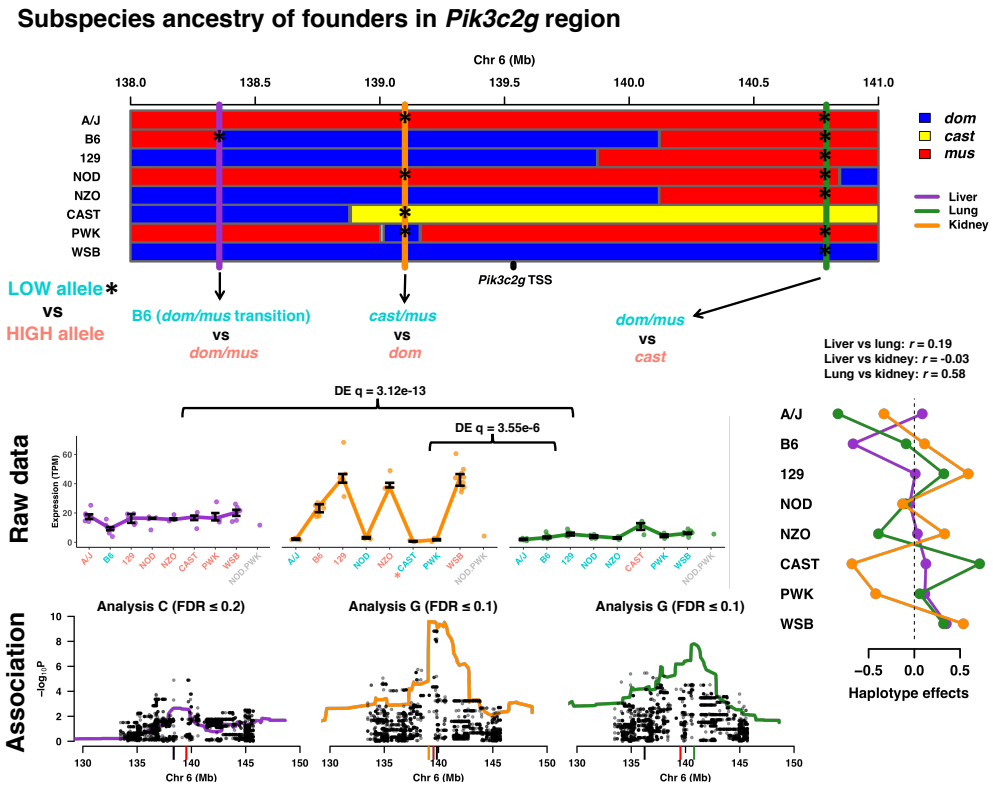
Subspecies ancestry of founders in Pik3c2g region
Chr 6 (Mb)
138.0
138.5
139.0
139.5
140.0
140.5
141.0
A/J
B6
129
NOD
NZO
CAST
PWK
WSB
dom
cast
mus
Liver
Lung
Kidney
Pik3c2g TSS
LOW allele
vs
HIGH allele
B6 (dom/mus transition)
vs
dom/mus
cast/mus
vs
dom
dom/mus
vs
cast
Liver vs lung: r = 0.19
Liver vs kidney: r = -0.03
Lung vs kidney: r = 0.58
DE q = 3.12e-13
DE q = 3.55e-6
Raw data
Expression (TPM)
Association
Analysis C (FDR ≤ 0.2)
Analysis G (FDR ≤ 0.1)
Analysis G (FDR ≤ 0.1)
−log10P
Chr 6 (Mb)
Haplotype effects
−0.5
0.0
0.5

Fig 4

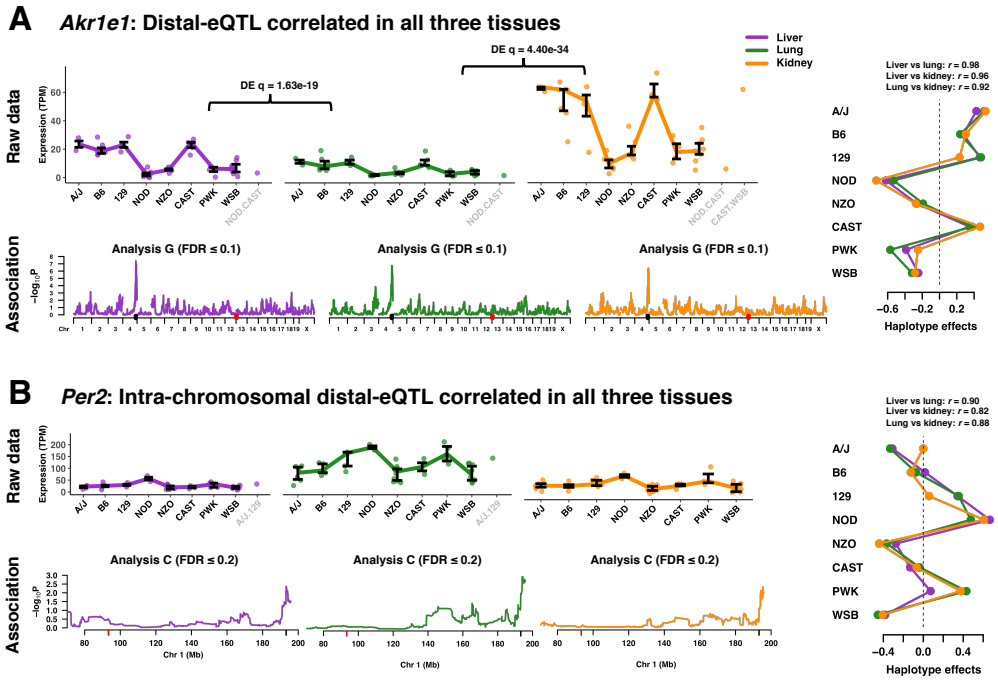
A
Akr1e1: Distal-eQTL correlated in all three tissues
Liver
Lung
Kidney
DE q = 1.63e-19
DE q = 4.40e-34
Raw data
Expression (TPM)
Liver vs lung: r = 0.98
Liver vs kidney: r = 0.96
Lung vs kidney: r = 0.92
A/J
B6
129
NOD
NZO
CAST
PWK
WSB
NOD.CAST
CAST.WSB
Association
Analysis G (FDR ≤ 0.1)
Chr
Haplotype effects
B
Per2: Intra-chromosomal distal-eQTL correlated in all three tissues
Liver vs lung: r = 0.90
Liver vs kidney: r = 0.82
Lung vs kidney: r = 0.88
A/J.129
Analysis C (FDR ≤ 0.2)
Chr 1 (Mb)

Fig 5

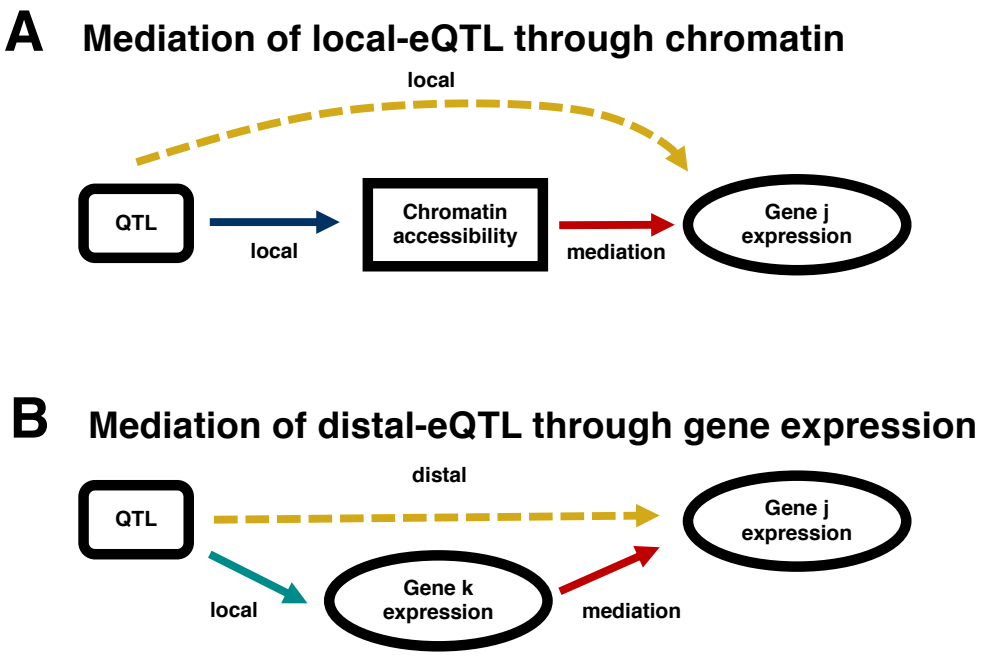
A
Mediation of local-eQTL through chromatin
local
QTL
local
Chromatin accessibility
mediation
Gene j expression
B
Mediation of distal-eQTL through gene expression
distal
QTL
Gene j expression
local
Gene k expression
mediation

Fig 6

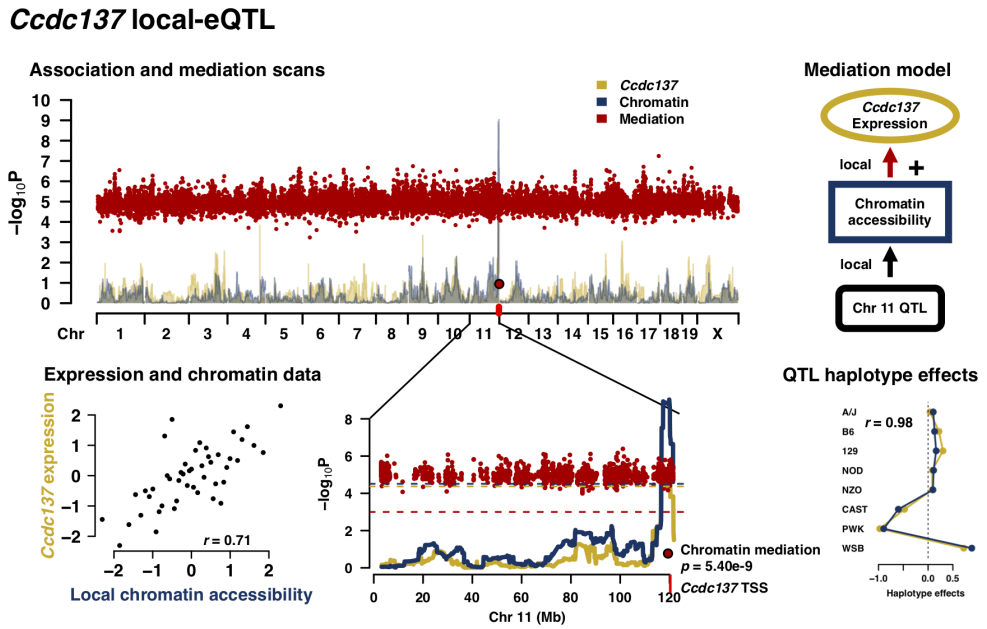
Ccdc137 local-eQTL
Association and mediation scans
Ccdc137
Chromatin
Mediation
−log10P
Chr 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 X
Mediation model
Ccdc137 Expression
local
+
Chromatin accessibility
local
Chr 11 QTL
Expression and chromatin data
Ccdc137 expression
r = 0.71
Local chromatin accessibility
Chromatin mediation
p = 5.40e-9
Ccdc137 TSS
Chr 11 (Mb)
QTL haplotype effects
A/J
B6
129
NOD
NZO
CAST
PWK
WSB
r = 0.98
Haplotype effects

Fig 7

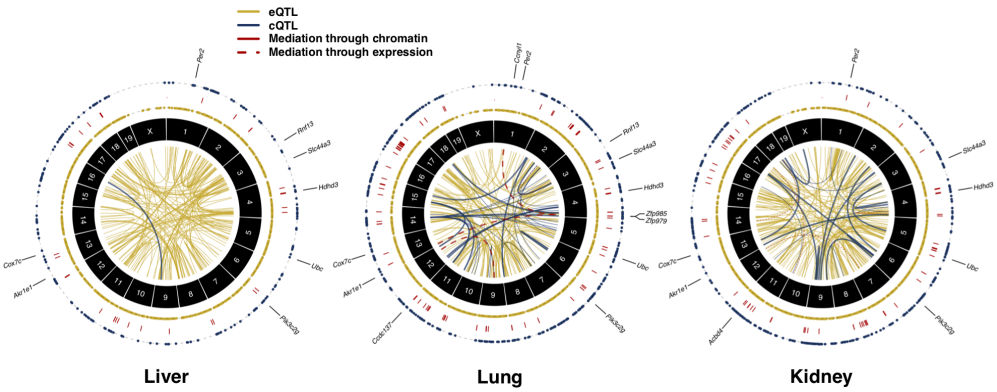
eQTL
cQTL
Mediation through chromatin
Mediation through expression
Liver
Lung
Kidney

Fig 8

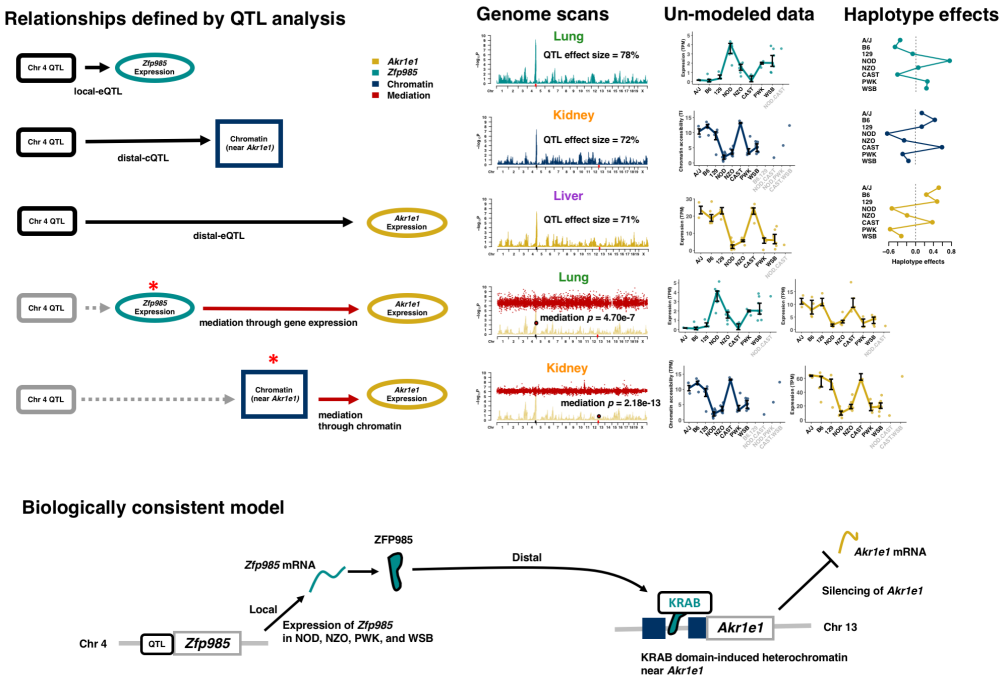
Relationships defined by QTL analysis
Genome scans
Un-modeled data
Haplotype effects
Chr 4 QTL
Zfp985 Expression
local-eQTL
Akr1e1
Zfp985
Chromatin
Mediation
Chromatin (near Akr1e1)
distal-cQTL
Akr1e1 Expression
distal-eQTL
mediation through gene expression
mediation through chromatin
Lung
QTL effect size = 78%
Kidney
QTL effect size = 72%
Liver
QTL effect size = 71%
mediation p = 4.70e-7
mediation p = 2.18e-13
A/J
B6
129
NOD
NZO
CAST
PWK
WSB
Haplotype effects
Biologically consistent model
ZFP985
Zfp985 mRNA
Distal
Akr1e1 mRNA
Silencing of Akr1e1
Local
Expression of Zfp985 in NOD, NZO, PWK, and WSB
Chr 4
QTL
Zfp985
KRAB
Akr1e1
Chr 13
KRAB domain-induced heterochromatin near Akr1e1

Fig 9