Candidate Regulators of Dyslipidemia in Chromosome 1 Substitution Lines Using Liver Co-expression Profiling Analysis

CURRENT STATUS: POSTED

maochun wang
College of Chemistry, Chemical Engineering and Biotechnology
ORCiD: 0000-0002-6056-1606

Fuyi Xu
Genetics, Genomics and Informatics

Shixian Hu
Gastroenterology and Hepatology

Yuxun Zhou
Chemistry, Chemical Engineering and Biotechnology

Kai Li
Chemistry, Chemical Engineering and Biotechnology

Hongyan Xu
Biostatistics and Epidemiology

Junhua Xiao  xiaojunhua@dhu.edu.cn
Corresponding Author

DOI: 10.21203/rs.2.10142/v1

SUBJECT AREAS
Epigenetics & Genomics

KEYWORDS
plasma lipids, glucose, C1SLs, gene network, candidate gene
Abstract

Background

Dyslipidemia is the major risk factor of cardiovascular disease. Although many genetic factors have been unveiled, large fraction of the phenotypic variance still needs further investigation. Chromosome 1 (Chr 1) harbors multiple gene loci related to blood lipid regulation, nevertheless, it's challenging to identify the functional genes.

Results

We constructed a mouse population, chromosome 1 substitution lines (C1SLs), of which Chr 1 differ from recipient C57BL/6J (B6) mouse while others remain the same. Therefore, any phenotype variance between C1SLs and B6 can be easily inferred to Chr 1. In the current study, we assayed plasma lipids and glucose in 13 C1SLs and their recipient strain B6. Through weighted gene co-expression network analysis of the liver transcriptome and “guilty-by-association” study, eight associated modules of plasma lipids and glucose were identified. Further joint analysis of human genome wide association studies revealed 48 candidate blood lipids or glucose regulating genes. In addition, 38 genes on Chr 1 were also uncovered, in which 13 have been functional validated in mouse models.

Conclusions

These results suggest that C1SLs are ideal mouse models to study the complex traits and identify the candidate genes on Chr 1 with gene co-expression network analysis.

Background

Plasma lipids levels of total cholesterol (CHOL), high-density lipoprotein cholesterol (HDL-C), Low-density lipoprotein cholesterol (LDL-C), and triglycerides (TG), are major contributors to cardiovascular disease [1]. Current evidence demonstrated both environmental and genetic factors can influence these multi-genic traits. Therefore,
discovery of the genetic regulators of plasma lipids metabolism would benefit medical 
diagnosis and gene therapies. Recent genome wide association studies (GWAS) in human 
have linked hundreds of genetic loci to plasma lipids metabolism, including those 
pathogenic genes such as APOE, PCSK9, CETP, LIPC, LPL APOA5[2, 3]. Further, several 
rare variants are also uncovered with next generation sequencing [4]. Although significant 
achievements have been made, the identified genetic loci only explained a small portion 
of the phenotypic variance, which suggests most of the genetic regulators remain 
unknown.

Mouse models have been widely used for deciphering regulatory genes of quantitative 
traits. Thousands of genetic loci have been identified through quantitative trait loci (QTL) 
mapping in F2 or backcross mouse populations (http://www.informatics.jax.org/). However, 
it’s challenging to identify the causative genes within the QTL. During the past decades, 
several newly developed mouse genetic reference populations, such as BXD recombinant 
inbred strains [5], Collaborative Cross [6], Hybrid Mouse Diversity Panel [7], and 
chromosome substitution strains (CSSs) [8], significantly accelerated the precise QTL 
mapping and functional gene identification due to the improved mapping power and 
resolution [9]. CSSs, which typically involve two inbred strains with significant phenotype 
difference, are a panel of inbred strains by backcrossing the donor and recipient parents 
over 10 generations. The final panel contains the entire genome information of the donor 
strain, and each CSS carries one intact donor chromosome in the genetic background of 
the recipient strain. Therefore, any phenotype difference between CSSs and recipient 
strain can be easily inferred to one particular chromosome. This enables to detect the 
multi-genic trait genes easily, and identify the QTLs quickly through linkage analysis in F2 
population and fine mapping with congenic strains. Previously, we proposed a novel 
strategy, the population of specific CSSs, to dissect the complex traits and built a
Chromosome 1 (Chr 1) substitution lines (C1SLs), in which, Chr 1 of recipient strain C57BL/6J was replaced by 30 different wild mice individually [10, 11]. Different from CSSs, C1SLs introduce more genetic alleles and can perform association studies to identify QTLs quickly without construct F2 or backcross segregation population and can be applied to systems genetics analysis easily.

It is well known that genes do not act individually, but interact with each other to act on complex traits. Therefore, building such gene interaction networks will give us deeper insights into the underlying mechanisms[12]. In addition, co-expressed genes have similar biological functions or are involved in same biological processes. Therefore, analysis based on gene networks can provide us alternate way to identify the potential regulators. Several algorithms have been developed to construct gene networks [13], among which weighted gene co-expression network analysis (WGCNA) is the most widely used[14]. This method not only construct gene networks but also summarize the hub genes and module eigengenes, which can be used to perform “guilt-by-association” between phenotypes and eigengenes to identify the traits-associated networks.

Several studies have demonstrated that Chr 1 harboring multiple genetic loci that control the plasma lipids and glucose levels [15, 16]. In order to identify the causal genes, we measured the plasma lipids and fasting glucose levels in C1SLs and quantified the transcriptomes of liver with RNA-seq technique. Through gene co-expression network analysis, combined with human GWAS and gene functional annotation, several plasma lipids and glucose regulating candidate genes especially those located on Chr 1 were identified (Figure 1).

Figure 1 Schematic of the methodology 13 C1SLs and recipient train B6 were involved in this study, blood lipids and fasting glucose levels were measured at 20 weeks. Liver gene co-expression network were constructed with WGCNA, then traits-associated modules were
identified through testing the association between traits and module eigengenes. For the trait-associated modules, the candidate gene were further dominated combing several other evidence such human GWAS or mouse phenotypes etc.

Results

C1SLs exhibits broad phenotype variability

In this study, plasma lipids (CHOL, HDL-C, LDL-C, and TG) and fasting glucose levels of 13 C1SLs and corresponding recipient strain B6 were examined using enzymatic assays (Figure 2A). The results showed broad phenotype variability with fold change 1.62 in GLU, 1.55 in CHOL, 1.51 in HDL-C, 2.11 in LDL-C and 1.58 in TG (Figure 2B-F, Supplementary Data S1). Due to the genetic variance on chromosome 1, different C1SLs showed great phenotypic variability. Compared with C1SLs, recipient strain B6 exhibited relatively low levels for GLU, CHOL, HDL-C, and LDL-C and a relatively high TG level.

Figure 2 Phenotype distributions across C1SLs and B6. (A) Schematic of traits collection (B-F) distribution of fasting glucose and plasma lipids levels across C1SLs and B6. Each bar represent one strain, the black corresponds to the recipient strain B6.

WGCNA identified several modules significantly associated with plasma lipids and fasting glucose levels

We carried out high throughput RNA-seq using Ilumina X-ten platform to comprehensively quantify the transcripts abundance of liver tissue for 29 samples (two samples per C1SLs and three B6). A total of ~2.3 billion reads were obtained, ranging from 26 million to 0.42 billion per sample. The raw reads were mapped onto the mouse genome with an average of 80% concordant pair alignment rate. Gene expression levels were generated and normalized with Cuffnorm program, further filtration was applied (See Materials and Methods) which resulted in 10525 genes for the subsequent analysis (Supplementary Data
To identify gene co-expression networks and regulatory genes for plasma lipids and glucose levels.

We constructed gene co-expression networks using WGCNA. With the soft-thresholding power parameter ($\beta = 6$) determined by the scale-free topology (Figure 3A-B), a total of 24 modules (exclude module gray) were identified (Figure 3C-D, Supplementary Data S2). The module size (i.e., the total number of genes in a module) varies significantly, ranging from 39 genes in module M5 to 2141 genes in module M24. Among those modules, module M19 (83 genes) was significantly associated with all five traits (Figure 3D). While modules M7 (491 genes), M8 (311 genes), M12 (247 genes) were significantly associated with glucose levels and modules M1 (930 genes), M14 (99 genes), M20 (389 genes), M21 (117 genes) are significantly linked to TG (Figure 3D).

**Figure 3** Weighted gene co-expression network analysis of liver transcriptomes (A) The soft thresholding index R2 (y-axis) as a function of different thresholding power $\beta$ (x-axis). (B) Mean connectivity (y-axis) as a function of the power $\beta$ (x-axis). (C) 24 co-expression modules identified from the liver RNA-seq dataset. WGCNA cluster dendrogram groups genes ($n = 10520$) measured across C1SLs and its recipient strain B6 liver into distinct gene modules (M1–24) defined by dendrogram branch cutting (D) Heat maps of Pearson correlation and p-value between modules and traits. Each cell represents the correlation coefficient (and p-value) computing from correlating module eigengenes (MEs) (rows) to traits (columns). The hub genes for each module were indicated aside each module.

**Gene prioritizing of the associated traits modules**

Among the 83 genes in module M19, 74 are significantly correlated ($p<0.05$) with phenotypes and module eigengenes simultaneously (Supplementary Data S3). Gene ontology (biological process) enrichment analysis revealed these genes are significantly
enriched in lipid metabolism and gluconeogenesis regulation (Supplementary Data S4). In addition, 14 genes are found in human GWAS with p < 10^{-4} (Figure 4A, Supplementary Data S5), 11 genes have been identified as functional genes for blood lipids or glucose metabolism (Figure 4C). Four of them, Creg1, Abcc3, Cyp2b9, and Cyp26a1 are highly expressed in liver tissue (Supplementary Figure 1). The module hub gene Tmem176a are significantly correlated with blood lipid levels (Figure 4B, Supplementary Data S3), and also have been mapped with human GWAS CHOL signal with a p = 2x10^{-8} (Supplementary Data S5).

In other modules associated with TG, 505 genes are significantly correlated with TG levels and eigengenes simultaneously (Supplementary Data S6), in which, 26 genes are overlapped with human GWAS signal (Figure 5A, Supplementary Data S5) and 40 genes have been demonstrated to play significant roles in TG metabolism (Figure 5B-E). In addition, six genes, Egfr, Hsd17b13, Cyp3a11, Arg1, Fads2, and Ahcy, are specifically highly expressed in liver tissue (Supplementary Figure 1).

There are 377 genes in module M7, M8, and M12 that are significantly associated with fasting glucose and their module eigengenes (Supplementary Data S7). Among them, eight genes were found to overlap with GWAS signals (Figure 6A, Supplementary Data S5) and 27 are glucose metabolism regulatory genes (Figure 6B-D). Furthermore, three of them, Pck1, Fads1, and Gckr are highly expressed in liver tissue (Supplementary Figure 1).

Figure 4 Gene prioritize for module M19. (A) Human GWAS overlapped genes for module M19, Gene with GWAS p < 10^{-4} for fasting glucose and blood lipids were retrieved from GRASP and GWAS Catalog (B) Correlation between module M19 hub gene Tmem176a and module eigengenes and traits. (C) Gene subnetwork for module M19, circles filled with green color represent genes overlapped with human GWAS, and circles filled with blue
represent genes functionally validated in mouse models, while genes with both functionally validated and overlapped with GWAS are marked with blue rectangles.

Figure 5 Gene prioritize for TG associated modules (A) Human GWAS overlapped genes for TG associated modules. Gene with GWAS p < 10^{-4} of TG levels were retrieved from GRASP and GWAS Catalog (B-E) Gene subnetwork for module M1, M14, M20, and M21. Circles filled with green color represent genes overlapped with human GWAS, and circles filled with blue represent genes functionally validated in mouse models, while genes with both functionally validated and overlapped with GWAS are marked with blue rectangles.

Figure 6 Gene prioritize for fasting glucose associated modules (A) Human GWAS overlapped genes for fasting glucose associated modules. Genes with GWAS p < 10^{-4} of fasting glucose levels were retrieved from GRASP and GWAS Catalog (B-D) Gene subnetworks for module M7, M8, and M12. Circles filled with green color represent genes overlapped with human GWAS, and circles filled with blue represent genes functionally validated in mouse models, while genes with both functionally validated and overlapped with GWAS are marked with blue rectangles.

Prioritizing causative genes on Chr1

Due to the specific genomic structure of C1SLs that only one chromosome differs from B6 (Figure 1), we believe the phenotypic difference are driven by the genetic variants on Chr 1. A total of 38 genes in the trait-associated modules were found to be located on Chr 1, of which, 35 harbor missense SNPs and all have 3’ or 5’ UTR variants (Table 1). In addition, several genes have been associated with the traits in mouse modules, such as Creg1 and Aox1 in module M19; Phlpp1, Nr5a2, Rnf149, Ncoa2, and Abl2 in module M1; Mogat1, Igfbp2, and Col3a1 in module M20; G0s2, Crp, and Ppox in module M21, M7, and M12, respectively.

Table 1 lists of module-trait associated genes on Chr 1
Recent work has demonstrated that gene co-expression network analysis is a powerful way to link genes to phenotypes. Here, WGCNA was applied to investigate liver transcriptomes of C1SLs mouse. A total of 24 modules were identified, in which, module M19 was significantly associated with blood lipids and glucose levels (Figure 3D). Through searching the MGI database, 13% (11 out of 84) of the M19 genes have been reported and involved in blood lipids or glucose metabolism (Figure 4C), including several functional genes, such as acyl-CoA thioesterase 11 (Acot11) [17], cellular repressor of E1A-stimulated genes 1 (Creg1) [18], carboxylesterase 1E (Ces1e) [19], carboxylesterase 1G (Ces1g) [19], and lamin A (Lmna) [20], which suggests both glucose and lipid metabolism are linked to each other and may share some common genetic architecture [21]. We also identified several TG (M1, M14, M20, and M21) and fasting glucose (M7, M8, and M12) associated modules, respectively. These modules include several known functional genes (Figure 5-6), such as peroxisome proliferator activated receptor gamma (Pparg) [22], cell death-inducing DFFA-like effector c (Cidec) [23], monoacylglycerol O-acyltransferase 1 (Mogat1) [24], glucokinase regulatory protein (Gckr) [25], and phosphoenolpyruvate carboxykinase 1 (Pck1) [26] etc. Since co-expressed genes are believed to be involved in functionally related biological processes or pathways [27], together with the known functional genes, we believe other genes in trait-associated modules could also have regulatory roles on those investigated traits.

In the past decades, human GWAS of blood lipids and glucose metabolism have identified hundreds of associated genes [1, 3, 28-32]. However, like other complex traits or disease, most variants identified so far only explained a small portion of phenotype variance,
leading to much ‘missing’ heritability unexplained [33]. This could be contributed to
multiple factors, including sample size, genetic structure, rare variants, and gene-gene
interactions et al [33, 34]. In addition, strict p-value threshold with high multiple testing
corrections was also believed to lose many positive loci [35, 36]. Joint analysis of human
GWAS and mouse genetics would help to “rescue” some of the ‘missing’ heritability [5, 34,
37-39]. In the present study, we identified 48 genes in the trait-associated modules which
have been reported in human GWAS with p < 10^{-4}. Among them, several genes not only
achieved the GWAS significance threshold (< 1 x 10^{-8}), but also have been functionally
validated in mouse models, including acyl-CoA thioesterase 11 (Acot11) [17, 40], estrogen
receptor 1(Esr1) [40, 41], Cd36 molecule (Cd36) [40, 42], fatty acid desaturase 2 (Fads2)
[28, 43], phospholipase A2, group VI (Pla2g6) [44, 45], glucokinase regulatory protein
(Gckr) [25, 46], and fatty acid desaturase 1(Fads1) [46]. Furthermore, we also found some
genes with modest GWAS p values which have also been validated to be functional in
mouse models, such as cellular repressor of E1A-stimulated genes 1 (Creg1) [18, 47],
cytochrome P450 family 7 subfamily b polypeptide 1 (Cyp7b1)[48, 49], NAD(P)H
dehydrogenase quinone 1 (Nqo1) [40, 50], peroxisome proliferator activated receptor
gamma (Pparg) [22, 28], and phosphoenolpyruvate carboxykinase 1(Pck1) [26, 31].
Although these genes have GWAS p value > 10^{-4}, they are possible candidate genes based
on the genetic evidence from our results (Figure 4-6). Therefore, we believe that by
reducing the threshold of strict GWAS p values and screening mouse genetic databases, it
is possible to locate more pathogenic genes more efficiently and accurately.
C1SLs were aimed to identify complex traits associated genes on Chr 1 by performing
association studies or systems genetics analysis. However, in the current study, only 13
C1SLs and recipient strain B6 were used which might lack adequate statistical power and
introduce many false positives [51]. But gene co-expression network analysis was able to prioritize the candidate genes especially for those located on Chr 1. The results showed only 38 genes located on Chr 1 with an average of 4.75 genes in eight trait-associated modules. The number of candidate genes is far less than QTL candidate genes identified by linkage or association studies with F2 mouse segregation population or other reference populations [9]. This will help to locate the disease-causing genes more efficiently. In addition, the identified modules contain at least one candidate gene located on Chr 1 that has been verified to be functional, which demonstrate that this strategy is very effective in C1SLs. Therefore, with the upcoming of other C1SLs and sequence data, we believe co-expression network analysis on C1SLs would provide us a new perspective for identifying other complex traits causative genes on Chr 1.

**Conclusion**

A total of eight gene networks associated with blood lipids and glucose levels were identified by gene co-expression network analysis. Further joint analysis of human GWAS resulted in 48 candidate genes. In addition, 38 genes on Chr 1 which include 13 well characterized were also prioritized as the causative genes (Table 1). However, these candidate genes still need further studies to illustrate their potential functional roles.

**Methods**

**Mice and diet**

Thirteen C1SLs (B6-Chr1JD, B6-Chr1HZ, B6-Chr1CM, B6-Chr1LY, B6-Chr1KM, B6-Chr1ZZ2, B6-Chr1SMX, B6-Chr1BLD, B6-Chr1ZZ1, B6-Chr1SJ, B6-Chr1TW, B6-Chr1DX and B6-Chr1YP) and its recipient strain B6 were housed in a room maintained at 18-22°C with a 12-hour
light and 12-hour dark cycle (6:00 A.M. to 6:00 P.M.). All mice were given a chow diet (M01-F25; Shanghai SLAC Laboratory Animal Co., Ltd.) for eight weeks, then fed with D12450B diet containing 4.3% fat, 19.2% protein, and 67.3% carbohydrate (Research Diets, New Brunswick, NJ) until sacrificed by cervical dislocation at 20 weeks.

**Experiment measurement**

Blood was collected into 1.5-ml tubes with EDTA by retro-orbital bleeding from male mice fasted for 4 hours in the morning. Blood serum was separated by centrifugation at 2500 g for 15 mins and frozen at -20°C until assay. Enzymatic assays for total cholesterol (CHOL), high-density lipoprotein cholesterol (HDL-C), Low-density lipoprotein cholesterol (LDL-C), triglycerides (TG), and Glucose (GLU) were performed with biochemical blood analyzer (Hitachi 7180; Hitachi, Tokyo, Japan) by Sino-British SIPPR/B&K Lab Animal Ltd. (Shanghai, China).

**RNA isolation and quality control**

RNA was extracted from liver tissue using RNAiso Plus reagent [Takara Biotechnology (Dalian) Co., Ltd, Liaoning China] according to the manufacturer’s protocol. RNA quality was analyzed using NanoDrop 2000c and Bioanalyzer. Samples with A260/A280 of 1.8–2.0 and RNA integrity number (RIN) greater than 8 were subsequently used for sequencing library preparation.

**RNA sequencing and read mapping**

Ploy-A enriched mRNA was sequenced on Illumina X-Ten platform (Illumina Inc., San Diego, CA) with 2 x 150 bp read length by WuXi AppTec (Shanghai, China) according to the manufacturer’s protocol. Reads were aligned to the mouse reference genome (GRCm38) using tophat2[52] with default parameters. Cuffnorm program implemented in cufflinks[53] was used to generate tables of expression values (Fragments Per Kilobase of
transcript per Million mapped reads, FPKM) which properly normalized for library size based on GRCm38 gene annotation downloaded from iGenome (https://support.illumina.com/sequencing/sequencing_software/igenome.html). Expression data were further filtered to remove genes that had less than 1 count in 20% or more samples and then log-transformed with log2(FPKM+1).

**Weighted gene co-expression network analysis (WGCNA)**

Log2 transformed expression values were analyzed with WGCNA package[14] in R to construct gene coexpression networks as previously described. Briefly, correlation matrix was obtained by calculating pair-wise pearson correlation coefficients between all genes across all samples, then soft thresholding power $\beta = 6$ was chosen based on scale-free topology ($R^2 > 0.9$) to generate weighted adjacency matrix. The adjacency was further transformed into Topological Overlap Matrix (TOM) which assess transcript interconnectedness. Following this, a dissimilarity measure was calculated. Genes were aggregated into modules by hierarchical clustering based on TOM and further refined using the dynamic tree cut algorithm. The module eigengene (ME) is the first principal component of a given module, and it was used to evaluate the module membership, which assessed the importance of genes in the network.

**Candidate gene analysis using publicly available resources**

We prioritized the candidates using the following public resources:

1. Human - Mouse: Disease Connection (HMDC). This resource included mouse and human gene-trait relationships from several databases, including Mouse Genome Informatics database (MGI), National Center for Biotechnology Information (NCBI), Online Mendelian Inheritance in Man (OMIM), and the Human Phenotype Ontology (HPO).

2. Human GWAS. Human GWAS for the plasma lipids and fasting glucose were obtained
from GRASP (https://grasp.nhlbi.nih.gov) [54] and GWAS Catalog (https://www.ebi.ac.uk/gwas/) [55]. GRASP includes available genetic association studies with p value less than 0.05 while GWAS Catalog collects SNP-trait associations with p value less than 1 \times 10^{-6}. In the present study, the mapped or nearest genes of the marker with p value less than 1 \times 10^{-4} were used to looking for overlap with the module gene lists.

3. Gene expression atlas across mouse tissue. Gene expression profiles across 22 mouse tissues were queried from NCBI (https://www.ncbi.nlm.nih.gov/) which generated by the Mouse ENCODE project using RNA-seq [56]. We define liver highly expressed genes by the expression level in liver is greater than threefold of the mean expression value across the 22 tissues.

**Genetic variants of the candidates**

Genetic variants between C1SLs and B6 were identified with whole genome sequencing as previously described [11]. Variant annotation were performed using Variant Effect Predictor [57].

**Abbreviations**

Chr 1: chromosome 1; C1SLs: chromosome 1 substitution lines; CHOL: total cholesterol ; HDL-C: high-density lipoprotein cholesterol ; LDL-C: low-density lipoprotein cholesterol ; TG: triglycerides; GWAS: genome wide association studies; QTL: quantitative trait loci; CSSs: chromosome substitution strains; WGCNA: weighted gene co-expression network analysis; GLU: Glucose; ME: eigengene
Declarations

Ethics approval:

All animal procedures were performed in accordance with guidelines of the Laboratory Animal Committee of Donghua University.

Consent for publication:

Not applicable.

Availability of data and material:

Raw data is being uploaded. The datasets supporting the conclusions of this article are included within the article and its Additional files.

Competing interests:

The authors declare that no competing interests exist.

Funding:

This work was supported by National Natural Science Foundation of China (Grant no. 31772550), the Key Project of Science and Technology Commission of Shanghai Municipality (No. 16140901300, 16140901302).

Authors’ contributions

JHX conceived and supervised the study. MCW and FYX performed the experiment and data analysis. SXH helped to collect RNA. MCW and FYX wrote the manuscript. YXZ, KL and HYX edited the manuscript. All authors read and approved the final version of the manuscript.

Acknowledgements:

The authors appreciate Jiatao Lu and Weiwang Duan for help doing some experiments and data analysis.
Authors' information:

1. College of Chemistry, Chemical Engineering, and Biotechnology, Donghua University, Shanghai 201620, China
2. Department of Genetics, Genomics, and Informatics, University of Tennessee Health Science Center, Memphis 38163, TN, United States
3. Department of Gastroenterology and Hepatology, University of Groningen and University Medical Center Groningen, Groningen, Netherlands
4. Department of Biostatistics and Epidemiology, Medical College of Georgia, Augusta University, Augusta 30912, GA, United States

References

1. Kathiresan S, Manning AK, Demissie S, D'agostino RB, Surti A, Guiducci C, Gianniny L, Burtt NP, Melander O, Orho-Melander M: A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study. *BMC Med Genet* 2007, 8(1):S17.

2. Helgadottir A, Gretarsdottir S, Thorleifsson G, Hjartarson E, Sigurdsson A, Magnusdottir A, Jonasdottir A, Kristjansson H, Sulem P, Oddsson A: Variants with large effects on blood lipids and the role of cholesterol and triglycerides in coronary disease. *Nat Genet* 2016, 48(6):634.

3. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, Mora S: Discovery and refinement of loci associated with lipid levels. *Nat Genet* 2013, 45(11):1274.

4. Natarajan P, Peloso GM, Zekavat SM, Montasser M, Ganna A, Chaffin M, Khera AV, Zhou W, Bloom JM, Engreitz JM: Deep-coverage whole genome sequences and blood lipids among 16,324 individuals. *Nature communications* 2018, 9(1):3391.

5. Wang X, Pandey AK, Mulligan MK, Williams EG, Mozhui K, Li Z, Jovaisaite V, Quarles LD,
Xiao Z, Huang J: Joint mouse–human phenome-wide association to test gene function and disease risk. *Nature communications* 2016, 7:10464.

6. Consortium CC: The genome architecture of the Collaborative Cross mouse genetic reference population. *Genetics* 2012, 190(2):389-401.

7. Ghazalpour A, Rau CD, Farber CR, Bennett BJ, Orozco LD, van Nas A, Pan C, Alayee H, Beaven SW, Civelek M: Hybrid mouse diversity panel: a panel of inbred mouse strains suitable for analysis of complex genetic traits. *Mamm Genome* 2012, 23(9-10):680-692.

8. Nadeau JH, Singer JB, Matin A, Lander ES: Analysing complex genetic traits with chromosome substitution strains. *Nat Genet* 2000, 24(3):221.

9. Buchner DA, Nadeau JH: Contrasting genetic architectures in different mouse reference populations used for studying complex traits. *Genome Res* 2015, 25(6):775-791.

10. Xiao J, Liang Y, Li K, Zhou Y, Cai W, Zhou Y, Zhao Y, Xing Z, Chen G, Jin L: A novel strategy for genetic dissection of complex traits: the population of specific chromosome substitution strains from laboratory and wild mice. *Mamm Genome* 2010, 21(7-8):370-376.

11. Xu F, Chao T, Liang Y, Li K, Hu S, Wang M, Zhou Y, Xu H, Xiao J: Genome sequencing of chromosome 1 substitution lines derived from Chinese wild mice revealed a unique resource for genetic studies of complex traits. *G3: Genes, Genomes, Genetics* 2016, 6(11):3571-3580.

12. Stuart JM, Segal E, Koller D, Kim SK: A gene-coexpression network for global discovery of conserved genetic modules. *Science* 2003, 302(5643):249-255.

13. Henry VJ, Bandrowski AE, Pepin A-S, Gonzalez BJ, Desfeux A: OMICtools: an informative directory for multi-omic data analysis. *Database* 2014, 2014.

14. Langfelder P, Horvath S: WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008, 9(1):559.

15. Leduc MS, Lyons M, Darvishi K, Walsh K, Sheehan S, Amend S, Cox A, Orho-Melander
M, Kathiresan S, Paigen B: The mouse QTL map helps interpret human genome-wide association studies for HDL cholesterol. *J Lipid Res* 2011, 52(6):1139-1149.

16. Orozco LD, Cokus SJ, Ghazalpour A, Ingram-Drake L, Wang S, Van Nas A, Che N, Araujo JA, Pellegrini M, Lusis AJ: Copy number variation influences gene expression and metabolic traits in mice. *Hum Mol Genet* 2009, 18(21):4118-4129.

17. Zhang Y, Li Y, Niepel MW, Kawano Y, Han S, Liu S, Marsili A, Larsen PR, Lee C-H, Cohen DE: Targeted deletion of thioesterase superfamily member 1 promotes energy expenditure and protects against obesity and insulin resistance. *Proceedings of the National Academy of Sciences* 2012, 109(14):5417-5422.

18. Tian X, Yan C, Liu M, Zhang Q, Liu D, Liu Y, Li S, Han Y: CREG1 heterozygous mice are susceptible to high fat diet-induced obesity and insulin resistance. *PLoS One* 2017, 12(5):e0176873.

19. Wei E, Ali YB, Lyon J, Wang H, Nelson R, Dolinsky VW, Dyck JR, Mitchell G, Korbutt GS, Lehner R: Loss of TGH/Ces3 in mice decreases blood lipids, improves glucose tolerance, and increases energy expenditure. *Cell Metab* 2010, 11(3):183-193.

20. Arimura T, Helbling-Leclerc A, Massart C, Varnous S, Niel F, Lacene E, Fromes Y, Toussaint M, Mura A-M, Keller DI: Mouse model carrying H222P-Lmna mutation develops muscular dystrophy and dilated cardiomyopathy similar to human striated muscle laminopathies. *Hum Mol Genet* 2004, 14(1):155-169.

21. Parhofer KG: Interaction between glucose and lipid metabolism: more than diabetic dyslipidemia. *Diabetes Metab J* 2015, 39(5):353-362.

22. Heikkinen S, Argmann C, Feige JN, Koutnikova H, Champy M-F, Dali-Youcef N, Schadt EE, Laakso M, Auwerx J: The Pro12Ala PPARγ2 variant determines metabolism at the gene-environment interface. *Cell Metab* 2009, 9(1):88-98.

23. Toh SY, Gong J, Du G, Li JZ, Yang S, Ye J, Yao H, Zhang Y, Xue B, Li Q: Up-regulation of
mitochondrial activity and acquirement of brown adipose tissue-like property in the white adipose tissue of fsp27 deficient mice. *PLoS One* 2008, 3(8):e2890.

24. Agarwal AK, Tunison K, Dalal JS, Yen C-LE, Farese RV, Horton JD, Garg A: Mogat1 deletion does not ameliorate hepatic steatosis in lipodystrophic (Agpat2−/−) or obese (ob/ob) mice. *J Lipid Res* 2016, 57(4):616-630.

25. Farrelly D, Brown KS, Tieman A, Ren J, Lira SA, Hagan D, Gregg R, Mookhtiar KA, Hariharan N: Mice mutant for glucokinase regulatory protein exhibit decreased liver glucokinase: a sequestration mechanism in metabolic regulation. *Proceedings of the National Academy of Sciences* 1999, 96(25):14511-14516.

26. Hakimi P, Johnson MT, Yang J, Lepage DF, Conlon RA, Kalhan SC, Reshef L, Tilghman SM, Hanson RW: Phosphoenolpyruvate carboxykinase and the critical role of cataplerosis in the control of hepatic metabolism. *Nutr Metab (Lond)* 2005, 2(1):33.

27. Weirauch MT: Gene coexpression networks for the analysis of DNA microarray data. *Applied statistics for network biology: methods in systems biology* 2011, 1:215-250.

28. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ: Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010, 466(7307):707.

29. Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, Cooper GM, Roos C, Voight BF, Havulinna AS: Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet* 2008, 40(2):189.

30. Siewert KM, Voight BF: Bivariate GWAS scan identifies six novel loci associated with lipid levels and coronary artery disease. *bioRxiv* 2018:319848.

31. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, Wheeler E, Glazer NL, Bouatia-Naji N, Gloyn AL: New genetic loci implicated in fasting glucose
homeostasis and their impact on type 2 diabetes risk. *Nat Genet* 2010, 42(2):105.

32. Hwang J-Y, Sim X, Wu Y, Liang J, Tabara Y, Hu C, Hara K, Tam CH, Cai Q, Zhao Q: Genome-wide association meta-analysis identifies novel variants associated with fasting plasma glucose in East Asians. *Diabetes* 2015, 64(1):291-298.

33. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A: Finding the missing heritability of complex diseases. *Nature* 2009, 461(7265):747.

34. Parker CC, Palmer AA: Dark matter: are mice the solution to missing heritability? *Frontiers in genetics* 2011, 2:32.

35. Lee SH, Wray NR, Goddard ME, Visscher PM: Estimating missing heritability for disease from genome-wide association studies. *The American Journal of Human Genetics* 2011, 88(3):294-305.

36. Lee T, Lee I: araGWAB: Network-based boosting of genome-wide association studies in Arabidopsis thaliana. *Sci Rep* 2018, 8(1):2925.

37. Ashbrook DG, Williams RW, Lu L, Stein JL, Hibar DP, Nichols TE, Medland SE, Thompson PM, Hager R: Joint genetic analysis of hippocampal size in mouse and human identifies a novel gene linked to neurodegenerative disease. *BMC Genomics* 2014, 15(1):850.

38. Cox RD, Church CD: Mouse models and the interpretation of human GWAS in type 2 diabetes and obesity. *Dis Model Mech* 2011, 4(2):155-164.

39. Ashbrook DG, Williams RW, Lu L, Hager R: A cross-species genetic analysis identifies candidate genes for mouse anxiety and human bipolar disorder. *Front Behav Neurosci* 2015, 9:171.

40. Asselbergs F, Guo Y, Sivapalaratnam S, Tragante V, Lanktree M, Lange L, Almoguera B, Appelman Y, Barnard J, Baumert J: Large-scale gene-centric meta-analysis across 32 studies identifies multiple lipid loci. *The American Journal of Human Genetics* 2012,
91(5):823-838.

41. Ohlsson C, Hellberg N, Parini P, Vidal O, Bohlooly M, Rudling M, Lindberg MK, Warner M, Angelin B, Gustafsson J-Å: Obesity and disturbed lipoprotein profile in estrogen receptor-α-deficient male mice. *Biochem Biophys Res Commun* 2000, 278(3):640-645.

42. Goudriaan JR, Dahlmans VE, Teusink B, Ouwens DM,Febbraio M, Maassen JA, Romijn JA, Havekes LM, Voshol PJ: CD36 deficiency increases insulin sensitivity in muscle, but induces insulin resistance in the liver in mice. *J Lipid Res* 2003, 44(12):2270-2277.

43. Stroud CK, Nara TY, Roqueta-Rivera M, Radlowski EC, Lawrence P, Zhang Y, Cho BH, Segre M, Hess RA, Brenna JT: Disruption of FADS2 gene in mice impairs male reproduction and causes dermal and intestinal ulceration. *J Lipid Res* 2009, 50(9):1870-1880.

44. Spracklen CN, Chen P, Kim YJ, Wang X, Cai H, Li S, Long J, Wu Y, Wang YX, Takeuchi F: Association analyses of East Asian individuals and trans-ancestry analyses with European individuals reveal new loci associated with cholesterol and triglyceride levels. *Hum Mol Genet* 2017, 26(9):1770-1784.

45. Zhang L, Zhong S, Li Y, Ji G, Sundaram M, Yao Z: Global inactivation of the Pla2g6 gene in mice does not cause dyslipidemia under chow or high-fat diet conditions. *Journal of cancer prevention* 2013, 18(3):235.

46. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan Ja, Mägi R, Strawbridge RJ, Rehnberg E, Gustafsson S: Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet* 2012, 44(9):991.

47. Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ: Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007, 316(5829):1331-1336.

48. Del-Aguila J, Beitselshees A, Cooper-Dehoff R, Chapman A, Gums J, Bailey K, Gong Y,
Turner ST, Johnson J, Boerwinkle E: Genome-wide association analyses suggest NELL1 influences adverse metabolic response to HCTZ in African Americans. *The pharmacogenomics journal* 2014, 14(1):35.

49. Li-Hawkins J, Lund EG, Turley SD, Russell DW: Disruption of the oxysterol 7α-hydroxylase gene in mice. *J Biol Chem* 2000, 275(22):16536-16542.

50. Gaikwad A, Long DJ, Stringer JL, Jaiswal AK: In vivo role of NAD (P) H: quinone oxidoreductase 1 (NQO1) in the regulation of intracellular redox state and accumulation of abdominal adipose tissue. *J Biol Chem* 2001, 276(25):22559-22564.

51. Flint J, Eskin E: Genome-wide association studies in mice. *Nature Reviews Genetics* 2012, 13(11):807.

52. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL: TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 2013, 14(4):R36.

53. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Van Baren MJ, Salzberg SL, Wold BJ, Pachter L: Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 2010, 28(5):511.

54. Leslie R, O’donnell CJ, Johnson AD: GRASP: analysis of genotype–phenotype results from 1390 genome-wide association studies and corresponding open access database. *Bioinformatics* 2014, 30(12):i185-i194.

55. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H, McMahon A, Milano A, Morales J: The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res* 2016, 45(D1):D896-D901.

56. Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, Sandstrom R, Ma Z, Davis C, Pope BD: A comparative encyclopedia of DNA elements in the mouse genome. *Nature* 2014,
515(7527):355.

57. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F: The ensembl variant effect predictor. Genome Biol 2016, 17(1):122.

Table

Table 1 lists of module-trait associated genes on Chr 1

| Gene Symbol | Entrez ID | C h r | Start | End   | Mod ule | Associated Phenotype | # Missense SNP | # UTR SNP |
|-------------|-----------|-------|-------|-------|---------|---------------------|----------------|-----------|
| Creg1       | 43337      | 5     | 16576 | 16577 | M19     | CHOL, HDL, LDL, TG, GLU | 1              | 118       |
| Atic        | 10814      | 7     | 71579 | 71579 | M19     | CHOL, HDL, LDL, TG, GLU | 4              | 31        |
| Aox1        | 11761      | 1     | 58029 | 58106 | M19     | CHOL, HDL, LDL, TG, GLU | 18             | 5         |
| Smyd3       | 69726      | 1     | 17895 | 17951 | M19     | CHOL, HDL, LDL, TG, GLU | 10             | 152       |
| Igsf8       | 14055      | 9     | 17226 | 17231 | M19     | CHOL, HDL, LDL, TG, GLU | 7              | 29        |
| Inpp4a      | 26918      | 0     | 37299 | 37410 | M1      | TG                  | 39             | 98        |
| Phlpp1      | 98432      | 1     | 10617 | 10639 | M1      | TG                  | 9              | 20        |
| Eprs        | 10750      | 8     | 18536 | 18542 | M1      | TG                  | 14             | 7         |
| Ino80d      | 22719      | 5     | 62958 | 63114 | M1      | TG                  | 36             | 231       |
| Nr5a2       | 26424      | 1     | 13684 | 13696 | M1      | TG                  | 5              | 93        |
| Rnf149      | 67702      | 1     | 39551 | 39577 | M1      | TG                  | 2              | 15        |
| Ncoa2       | 17978      | 1     | 13139 | 13374 | M1      | TG                  | 6              | 59        |
| Abl2        | 11352      | 1     | 15655 | 15664 | M1      | TG                  | 5              | 113       |
| Kmo         | 98256      | 1     | 17562 | 17566 | M1      | TG                  | 20             | 105       |
| Myo1b       | 17912      | 1     | 51749 | 51916 | M1      | TG                  | 4              | 78        |
| Wdr26       | 22675      | 7     | 18117 | 18121 | M1      | TG                  | 3              | 100       |
| Rabgap1l    | 29809      | 1     | 16021 | 16079 | M1      | TG                  | 16             | 62        |
| Sde2        | 20876      | 8     | 18085 | 18086 | M1      | TG                  | 21             | 28        |
| Gm38394     | NA         | 1     | 13361 | 13366 | M1      | TG                  | 6              | 112       |
| 1700034P13Rik | 73331   | 1     | 97476 | 97919 | M1      | TG                  | 0              | 0         |
| Etnk2       | 21425      | 3     | 13336 | 13338 | M14     | TG                  | 12             | 62        |
| Protein   | GenBank Accession | Chromosome | Start (bp) | End (bp) | Mapped to Plate | Plate | TG | Figures |
|-----------|-------------------|------------|------------|----------|-----------------|-------|----|---------|
| Mogat1    | 68393             | 1          | 78510      | 78538    | M20             | TG    | 0  | 3       |
| Igfbp2    | 16008             | 1          | 72824      | 72852    | M20             | TG    | 0  | 5       |
| Cps1      | 22723             | 1          | 67123      | 67231    | M20             | TG    | 1  | 11      |
| Col3a1    | 12825             | 1          | 45311      | 45349    | M20             | TG    | 2  | 10      |
| Rpl28-ps1 | 10004             | 1          | 12803      | 12803    | M20             | TG    | 0  | 0       |
| Aox3      | 71724             | 1          | 58113      | 58200    | M20             | TG    | 14 | 35      |
| 2810459M  | 72792             | 1          | 86045      | 86055    | M20             | TG    | 6  | 56      |
| 11Rik     | 14373             | 1          | 19327      | 19327    | M21             | TG    | 0  | 7       |
| G0s2      | 69527             | 1          | 42851      | 42905    | M21             | TG    | 5  | 22      |
| Crp       | 12944             | 1          | 17269      | 17283    | M7              | GLU   | 2  | 10      |
| Ppi3      | NA                | 1          | 82233      | 82235    | M7              | GLU   | 6  | 40      |
| Tmem131   | 56030             | 1          | 36792      | 36943    | M7              | GLU   | 16 | 19      |
| Tmem185   | 22635             | 1          | 11952      | 11952    | M7              | GLU   | 1  | 26      |
| Tmem185b  | 75623             | 1          | 44086      | 44102    | M7              | GLU   | 0  | 24      |
| Tex30     | 70225             | 1          | 58430      | 58445    | M7              | GLU   | 0  | 0       |
| 9430016H  | 68115             | 1          | 57406      | 57417    | M7              | GLU   | 1  | 10      |
| 0Rik      | 19044             | 1          | 17127      | 17128    | M12             | GLU   | 9  | 10      |

Figures
Schematic of the methodology 13 C1SLs and recipient train B6 were involved in this study, blood lipids and fasting glucose levels were measured at 20 weeks. Liver gene co-expression network were constructed with WGCNA, then traits-associated modules were identified through testing the association between traits and module eigengenes. For the trait-associated modules, the candidate gene were further dominated combing several other evidence such human GWAS or mouse phenotypes etc.
Figure 2

Phenotype distributions across C1SLs and B6. (A) Schematic of traits collection (B-F) distribution of fasting glucose and plasma lipids levels across C1SLs and B6. Each bar represent one strain, the black corresponds to the recipient strain B6.
Weighted gene co-expression network analysis of liver transcriptomes (A) The soft thresholding index R2 (y-axis) as a function of different thresholding power β (x-axis). (B) Mean connectivity (y-axis) as a function of the power β (x-axis). (C) 24 co-expression modules identified from the liver RNA-seq dataset. WGCNA cluster dendrogram groups genes (n = 10520) measured across C1SLs and its recipient strain B6 liver into distinct gene modules (M1–24) defined by dendrogram branch cutting (D) Heat maps of Pearson correlation and p-value between modules and traits. Each cell represents the correlation coefficient (and p-value) computing from correlating module eigengenes (MEs) (rows) to traits (columns). The hub genes for each module were indicated aside each module.
Gene prioritize for module M19. (A) Human GWAS overlapped genes for module M19, Gene with GWAS p < 10^{-4} for fasting glucose and blood lipids were retrieved from GRASP and GWAS Catalog (B) Correlation between module M19 hub gene Tmem176a and module eigengenes and traits. (C) Gene subnetwork for module M19, circles filled with green color represent genes overlapped with human GWAS, and circles filled with blue represent genes functionally validated in mouse models, while genes with both functionally validated and overlapped with GWAS are marked with blue rectangles.
Gene prioritize for TG associated modules (A) Human GWAS overlapped genes for TG associated modules. Gene with GWAS p < 10^-4 of TG levels were retrieved from GRASP and GWAS Catalog (B-E) Gene subnetwork for module M1, M14, M20, and M21. Circles filled with green color represent genes overlapped with human GWAS, and circles filled with blue represent genes functionally validated in mouse models, while genes with both functionally validated and overlapped with GWAS are marked with blue rectangles.
Gene prioritize for fasting glucose associated modules (A) Human GWAS overlapped genes for fasting glucose associated modules. Genes with GWAS p < 10^{-4} of fasting glucose levels were retrieved from GRASP and GWAS Catalog (B-D) Gene subnetworks for module M7, M8, and M12. Circles filled with green color represent genes overlapped with human GWAS, and circles filled with blue represent genes functionally validated in mouse models, while genes with both functionally validated and overlapped with GWAS are marked with blue rectangles.

Supplementary Files
This is a list of supplementary files associated with the primary manuscript. Click to
