Large-scale in silico mapping of complex quantitative traits in inbred mice

Pengyuan Liu
Washington University School of Medicine in St. Louis

Haris Vikis
Washington University School of Medicine in St. Louis

Yan Lu
Washington University School of Medicine in St. Louis

Daolong Wang
Washington University School of Medicine in St. Louis

Ming You
Washington University School of Medicine in St. Louis

Follow this and additional works at: https://digitalcommons.wustl.edu/open_access_pubs

Please let us know how this document benefits you.

Recommended Citation
Liu, Pengyuan; Vikis, Haris; Lu, Yan; Wang, Daolong; and You, Ming, "Large-scale in silico mapping of complex quantitative traits in inbred mice." PLoS One. 2, 7. e651. (2007).
https://digitalcommons.wustl.edu/open_access_pubs/874

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.
Large-Scale In Silico Mapping of Complex Quantitative Traits in Inbred Mice

Pengyuan Liu, Haris Vikis, Yan Lu, Daolong Wang, Ming You*

Department of Surgery and the Alvin J. Siteman Cancer Center, Washington University School of Medicine, St. Louis, Missouri, United States of America

Understanding the genetic basis of complex disease and disease-related quantitative traits will aid in the development of diagnostics and therapeutics. The process of gene discovery can be sped up by rapid and effective integration of well-defined mouse genome and phenome data resources. We describe here an in silico gene-discovery strategy through genome-wide association (GWA) scans in inbred mice with a wide range of genetic variation. We identified 937 quantitative trait loci (QTLs) from a survey of 173 mouse phenotypes, which include models of human disease (atherosclerosis, cardiovascular disease, cancer and obesity) as well as behavioral, hematological, immunological, metabolic, and neurological traits. 67% of QTLs were from a survey of 173 mouse phenotypes, which include models of human disease (atherosclerosis, cardiovascular disease, cancer and obesity) as well as behavioral, hematological, immunological, metabolic, and neurological traits. 67% of QTLs were refined into genomic regions <0.5 Mb with ~40-fold increase in mapping precision as compared with classical linkage analysis. This makes for more efficient identification of the genes that underlie disease. We have identified two QTL genes, Adam12 and Cdh2, as causal genetic variants for atherogenic diet-induced obesity. Our findings demonstrate that GWA analysis in mice has the potential to resolve multiple tightly linked QTLs and achieve single-gene resolution. These high-resolution QTL data can serve as a primary resource for positional cloning and gene identification in the research community.


INTRODUCTION

Humans and mice have been a common ancestor for 75 million years ago. Of mouse genes, 99% have homologues in humans and 96% are in the same syntenic location. Regions of the genome that are highly conserved between the two species align well with functionality [1]. The mouse has been a powerful force in elucidating the genetic basis of human physiology and pathophysiology [2]. Many genetic alterations identified in mouse models have a common genetic equivalent (or ortholog) in humans. In the past decades, quantitative trait locus (QTL) mapping has identified hundreds of chromosomal regions containing genes affecting cancer, diabetes, hypertension, obesity and other disease-related phenotypes in mice. Most of these complex traits show clear genetic components, yet the underlying genes remain largely elusive.

QTL analysis using mouse models can potentially identify genes that are important for human diseases, which are often difficult to identify in human populations, due to the complexity of human genetic data. However, a major obstacle of identifying QTL genes is the difficulty of resolving these chromosomal regions (10–20 cM) into sufficiently small intervals to make positional cloning possible. Recent advances in genomic sequence analysis and SNP discovery have provided researchers with the necessary resources to explore a wide range of genetic variation in laboratory inbred mice [3]. The use of dense SNP maps in laboratory inbred mice has proven successful in the refinement of previous QTL regions and the identification of new genetic determinants of complex traits [4–8]. Recently, positional cloning of a novel gene underlying a complex trait was done immediately after a genome-wide association (GWA) analysis [4].

A large and diverse set of traits has been systematically organized in the publicly accessible Mouse Phenome Database (MPD) (www.jax.org/phenome), housed at The Jackson Laboratory (Bar Harbor, Maine, United States). Rapid and effective integration of mouse genome and phenome data resources can dramatically speed up the process of gene discovery [Figure 1]. Here, we describe an in silico gene-discovery strategy through GWA scans to identify high-resolution QTLs in inbred mice. We further demonstrate that GWA analysis in mice has the potential to resolve multiple tightly linked QTLs and achieve single-gene resolution.

RESULTS

The mouse haplotype map

We first evaluated the laboratory inbred mouse haplotype map that was used in the analysis. This haplotype map was generated by a joint effort of The Broad Institute of Harvard and MIT, and the Wellcome Trust Center for Human Genetics (Text S1). In total, 148,062 evenly distributed single-nucleotide polymorphisms (SNPs), which span the mouse genome at an average density of ~18 kb per SNP, were genotyped in classical inbred mouse strains (Figure S1). Of these SNPs, 96.2% have minor allele frequencies greater than 0.05 and 76.7% of the inter-SNP distances are smaller than 20 kb. LD decay over the physical distance was observed where the mean $r^2$ fell to <0.5 within 2 kb and to less than 0.23 within 200 kb. To capture large-scale patterns of LD, we also measured the number of proxies, i.e. SNPs, showing a strong correlation with one or more others [9], with a window size of 500 kb across the genome (Figure S2). Using a stringent threshold of $r^2 ≥ 1.0$ (that is complete LD), the average number of proxies per SNP is 2.5, and one in two SNPs

Academic Editor: Katrina Gwinn, DER Neurogenetics, National Institute of Neurological Disorders and Stroke, United States of America

Received March 28, 2007; Accepted June 21, 2007; Published July 25, 2007

Copyright: © 2007 Liu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Our research was supported by grants from the US National Institutes of Health (CA099187, CA099147, ES012063 and ES013340).

Competing Interests: The authors have declared that no competing interests exist.

* To whom correspondence should be addressed. E-mail: youm@wudosis.wustl.edu
Figure 1. Integration of genomics and phenomics approaches in mapping complex traits in mice. (A) Pulmonary adenoma susceptibility locus 1 (Pas1) was previously mapped to the distal region of chromosome 6 by linkage analysis in A×B and B×A recombinant inbred mice. (B) *In silico* LD mapping of the Pas1 locus using existing mouse SNP and tumor incidence data. (C) Enhanced view of LD mapping between 144.0–146.5 Mb. (D) Physical map of the complex multigenic Pas1 locus. The Casc1 gene carries a functional polymorphism at codon 60 (N60S) between lung tumor susceptible strains (e.g. A/J, SWR/J, MA/My) and resistant strains (e.g. C57BL/6J, DBA/2J, C3H/HeJ and AKR/J). (E) Lung tumorigenesis of Casc1 knockout mice. Treatment of Casc1*ko*/*ko* mice with urethane produced more than six times as many tumors per mouse than did treatment of Casc1*wt*/*B6*/*B6* mice. The horizontal coordinates were plotted using physical distance (Mb). In panel A, the horizontal line indicates lod = 3.3, and tumor multiplicity data were used in the linkage analysis. In panel B and C, the horizontal lines indicate –log(P) = 6.2, and urethane-induced lung adenoma incidence data were used in the analysis [4].

doi:10.1371/journal.pone.0000651.g001
have five or more proxies. Of all the SNPs, 89.5% have an $r^2 = 1.0$ with at least one other SNP. Using a looser threshold of $r^2 > 0.5$, the average number of proxies is increased to 6.5, and two in three SNPs have more than five proxies. Of all the SNPs, 91.3% have an $r^2 > 0.5$ with at least one other SNP. These data indicate that the mouse haplotype map provides substantial coverage of genetic variation in inbred mice and is useful for high-resolution genetic mapping studies.

### Analysis of 173 mouse phenotypes

We performed a comprehensive survey of 173 complex quantitative traits from the mouse PHENOME projects instituted by the Jackson Laboratory (www.jax.org/phenome). In most PHENOME projects, each phenotype was measured for about 40 strains from a total of 59 inbred strains of mice with a wide range of genetic variation (Figure S3 and Text S1). The strains used for phenotyping varied slightly among different PHENOME projects. Generally, at least 10 males and 10 females from each strain were measured for each trait. These complex traits include models of human disease (such as atherosclerosis, cardiovascular disease, cancer, diabetes, obesity, and osteoporosis) as well as behavioral, biochemical, hematological, immunological, metabolic, neurological, and reproductive traits (Table 1). Most of these 173 mouse quantitative traits are highlyheritable, with a mean heritability of 0.51, and 74% of these traits have a heritability greater than 0.40 (Figure S4). In general, cancer, taste and drinking preferences show less genetic control ($h^2 < 0.3$), while body composition and plasma lipid concentrations show larger genetic control ($h^2 > 0.7$). Of these mouse traits, 47% display significant phenotypic differences between males and females ($P < 0.05$), contributing to 1–31% of the phenotypic variation in the inbred mouse populations. Notably, obesity-related traits such as body weight and fat mass have larger sex effects and explain more than 20% of the phenotypic variation in inbred mice. Nevertheless, sex-specific genetic variations are generally small (less than 6%) in the inbred mouse populations.

### Identification of 937 QTLs

To identify the genetic basis of this large and diverse set of mouse phenotypes, we designed an in silico approach for gene identification through GWA scans in inbred mice. In our approach, GWA scans were implemented in an automatic processing pipeline which constitutes data retrieving, outlier detection, data preprocessing, hypothesis testing, permutation testing and QTL identification. The potential problem of population structure in inbred mice was also inspected (Figure S5 and Text S1). Totally, we identified 937 QTLs from the survey of 173 complex quantitative traits in inbred mice. These QTLs affect almost all types of mouse phenotypes and are distributed unevenly on all of the mouse chromosomes except chromosome Y (Figure 2A, and Table S1 and S2). The number of QTLs per trait ranges from 0 to 38, with a mean of 5.4 (Figure 2B). The average size of a QTL region is 0.54 Mb and 67% of the QTLs were mapped to genomic regions less than 0.5 Mb (Figure 2C). This translates to an extraordinary ~40-fold increase in mapping precision as compared with classical linkage analysis. In addition, about 20% of the QTLs were located within genomic regions less than 100 kb away from at least one other QTL, implying the prevalence of pleiotropic effects of genetic manipulation in mice.

Figure S6 shows a general QTL landscape from our GWA scans of several obesity-related phenotypes, including body weight and lean mass on normal chow and high fat, high cholesterol atherogenic diets. The analyses identified a large number of QTLs, most of which contained multiple significant SNPs clustering in the same genomic region. Several QTLs affecting multiple obesity phenotypes were repeatedly mapped to the identical locations on distal chromosome 1 and proximal chromosome 13 in different datasets. Most of these QTLs are broadly concordant with those detected previously by linkage analyses which, however, normally encompassed 10–20 cM (about 20 to 40 Mb on the mouse genome) [10]. In some scenarios, several separated QTLs from the GWA scans were mapped to a single linkage-defined QTL region. This lack of precision is mainly due to limited recombinants in these linkage mapping populations. Our GWA scans also found several novel QTLs that do not overlap with previous linkage analysis regions (Figure S6). For example, we fine mapped a novel obesity QTL to a region of approximately 700 Kb (33.98–34.71 Mb) on proximal chromosome 9, which covers two candidates: KirreI3 and A130066N16Rik; and mapped a novel QTL to a region of

### Table 1. Summary of the analyzed mouse traits and QTLs

<table>
<thead>
<tr>
<th>Categories</th>
<th>Description</th>
<th>Traits</th>
<th>QTLs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood biochemistry</td>
<td>Blood calcium and pH, coagulation factors, glucose, leptin, insulin, peptides and proteins</td>
<td>13</td>
<td>59</td>
</tr>
<tr>
<td>Body composition</td>
<td>Body weight, fat mass and lean mass on normal chow and atherogenic diet</td>
<td>24</td>
<td>118</td>
</tr>
<tr>
<td>Cancer</td>
<td>Metastatic progression, tumor latency, and tumor burden</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>Total cholesterol, HDL cholesterol, non-HDL cholesterol, and hepatic cholesterol</td>
<td>20</td>
<td>206</td>
</tr>
<tr>
<td>Drinking preference</td>
<td>Drinking preference for ammonium, calcium, potassium, sodium chloride and/or lactate</td>
<td>29</td>
<td>137</td>
</tr>
<tr>
<td>Gallbladder and gallstones</td>
<td>Gallbladder volume, mucin, liquid crystals, monohydrate crystals, sandy stones, and gallstones</td>
<td>10</td>
<td>23</td>
</tr>
<tr>
<td>Hematology</td>
<td>Complete blood count, red blood cell parameters, and platelets</td>
<td>19</td>
<td>128</td>
</tr>
<tr>
<td>Liver patholgy</td>
<td>Diet-induced liver pathology including liver weight, alanine aminotransferase, bile salts, microvacuoles, macrovacuoles and inflammation</td>
<td>6</td>
<td>22</td>
</tr>
<tr>
<td>Musculoskeletal</td>
<td>Bone mineral content and density</td>
<td>4</td>
<td>59</td>
</tr>
<tr>
<td>Neurological</td>
<td>Acoustic startle response, hearing, and prepulse inhibition</td>
<td>18</td>
<td>123</td>
</tr>
<tr>
<td>Reproduction</td>
<td>Reproductive performance such as number of litters per dam and number of mice per litter</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>Triglycerides</td>
<td>Triglycerides on normal chew and atherogenic diet</td>
<td>4</td>
<td>56</td>
</tr>
<tr>
<td>Others</td>
<td>Daily average food and water intake</td>
<td>7</td>
<td>0</td>
</tr>
</tbody>
</table>

DOI: 10.1371/journal.pone.0000651.t001
approximately 100 Kb (56.28–56.37 Mb) on distal chromosome 19, which covers Halp2 and Nrap. These novel QTLs identified in the GWA are not unexpected since association mapping detects allelic associations at the level of the population, which is normally comprised of 30–40 genetically diverse inbred mouse strains, rather than allelic difference between two specific parental strains in linkage mapping. Furthermore, all the obesity QTL regions detected by the GWA analyses using natural inbred mouse population samples are much smaller (less than 0.6 Mb on average with the present SNP density of 18 kb per SNP) as compared with linkage analyses. This makes positional cloning feasible immediately after GWA scans without the need for further time-consuming fine mapping with congenic strains.

Dissection of multiple tightly linked QTLs

GWA scans help resolve multiple, tightly linked QTLs. We again used obesity QTLs to demonstrate this potential. So far, at least 12 obesity QTLs have been mapped to distal mouse chromosome 1, and most of them overlap with each other and may be linked to the same gene (Figure 3A). Among these, Bwtq1 was originally identified in a cross between DBA/2J and C57BL/6J [11] and later refined to an 8-cM region (~33 Mb) between D1Mit30 and D1Mit57 [12]. This QTL has a general effect on body size, affecting the length of the tail and most bones studied, as well as having a weaker effect on mass. Using congenic strains, Bwtq1 was narrowed down to a 1.4-cM region (~4 Mb), approximately the region from D1Mit451 to D1Mit219 [13]. The location of Bwtq1 was further refined to 0.91 Mb between D1Icp7 and D1Icp10.
using interval-specific subcongenic strains, which led to the identification of a strong candidate, Pappa2, encoding an enzyme that cleaves an insulin-like growth factor binding protein [14]. However, our GWA scans identified two closely linked QTLs within 5 Mb of each other on distal chromosome 1. One is associated with several different obesity phenotypes in three independent datasets and is located within 153.5–155.5 Mb (-log(P) = 8.2–11.4) and the other reproduced the refined Bwtq1 QTL, showing significant effects on body weight and bone mineral content (-log(P) = 7.2). These results illustrate the robustness and reproducibility of GWA scans for complex quantitative traits in inbred mice.

**Murine knockout and/or transgenic models**

The refined region of previous QTLs was reduced, on average, to 0.54 Mb, which covers on average 6 genes on the mouse genome. Characterization of such a small number of genes in each of these QTLs can greatly accelerate the discovery of genes underlying complex traits. We further performed a detailed survey of QTLs for all obesity-related phenotypes such as body weight, fat mass, lean mass and leptin levels. Of the QTLs detected in the present study, we identified 10 candidate genes within these QTLs that, when mutated or overexpressed as transgenes in the mouse, result in phenotypes affecting body weight and adiposity. These genes include Gpr39 (G protein-coupled receptor 39), Crhr2 (corticotropin releasing hormone receptor 2), Cdkn1b (cyclin-dependent kinase inhibitor 1B), Adam12 (ADAM metallopeptidase domain 12), Rasgrf1 (RAS protein-specific guanine nucleotide-releasing factor 1), Chrm3 (cholinergic receptor, muscarinic 3), Mapk8 (mitogen activated protein kinase 8), Wnt10b (wingless related MMTV integration site 10b), Dusp1 (dual specificity phosphatase 1) and Cdh2 (cadherin 2) (Table 2). For example, an obesity QTL on the proximal chromosome 13 has been replicated in at least four linkage studies [15–18], and has not yet been narrowed down. Our GWA scans identified a refined QTL (8.0–10.0 Mb) covering the gene Chrm3 to be associated with four obesity phenotypes in three independent datasets (-log(P) = 6.7–8.6) (Figure 4A). Chrm3 is a member of the muscarinic acetylcholine receptors that play critical roles in regulating the activity of many important functions of the central and peripheral nervous system [19]. Chrm3 knockout mice had a significant reduction in food intake, body weight (25%), and peripheral fat deposits. The lack of Chrm3 also led to pronounced reductions (~5–10-fold) in serum leptin and insulin levels [20].

Similarly, we performed a survey of QTLs for plasma lipid levels which are correlated with the risk of atherosclerotic coronary heart disease [21]. We identified 11 candidate genes in the refined regions of QTLs affecting levels of high density lipoprotein (HDL)

<table>
<thead>
<tr>
<th>Genea</th>
<th>Chr</th>
<th>QTL peakb</th>
<th>Analyzed traitc</th>
<th>-log(P)</th>
<th>Phenotypes in KO or TG miced</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gpr39</td>
<td>1</td>
<td>127,875,052</td>
<td>Percent of tissue mass that is fat, after 8 wks on atherogenic diet (Naggert1)</td>
<td>6.1</td>
<td>Increased body weight and body fat composition (KO)</td>
<td>[27]</td>
</tr>
<tr>
<td>Crhr2</td>
<td>6</td>
<td>55,858,855</td>
<td>Leptin, after 18 wks on atherogenic diet (Naggert1)</td>
<td>6.3</td>
<td>Lower body fat and plasma lipids, increased insulin sensitivity, and lower feed efficiency on high-fat diet (KO)</td>
<td>[47]</td>
</tr>
<tr>
<td>Cdkn1b</td>
<td>6</td>
<td>134,947,686</td>
<td>Percent fat (Tordoff3)</td>
<td>7.4</td>
<td>Increased body fat percentage, and increased adiposity (KO)</td>
<td>[48]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>134,947,686</td>
<td>Percent lean (Tordoff3)</td>
<td>7.3</td>
<td>Moderate resistance to diet-induced obesity due to an impairment in the increase of the number of adipocytes in high-fat-fed mice (TG)</td>
<td>[24]</td>
</tr>
<tr>
<td>Rasgrf1</td>
<td>9</td>
<td>89,769,476</td>
<td>Percent fat (Tordoff3)</td>
<td>7.5</td>
<td>Reduced body weight, adiposity and ~20% decreased gonadal fat percentage (KO)</td>
<td>[49]</td>
</tr>
<tr>
<td>Chrm3</td>
<td>13</td>
<td>8,305,114</td>
<td>Final body weight after 8 wks on atherogenic diet (Paigen1)</td>
<td>7.9</td>
<td>Reduced adiposity and decreased ~25% body weight (KO)</td>
<td>[20]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10,048,424</td>
<td>Calculated weight of lean tissue (Tordoff3)</td>
<td>8.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>10,065,064</td>
<td>Initial body weight (Paigen1)</td>
<td>8.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>10,065,799</td>
<td>Body weight after 8 wks on atherogenic diet (Naggert1)</td>
<td>6.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mapk8</td>
<td>14</td>
<td>32,376,554</td>
<td>Leptin, after 18 wks on atherogenic diet (Naggert1)</td>
<td>6.3</td>
<td>Reduced body weight and increased insulin sensitivity (KO)</td>
<td>[50]</td>
</tr>
<tr>
<td>Wnt10b</td>
<td>15</td>
<td>98,434,034</td>
<td>Percent of tissue mass that is fat, after 8 wks on atherogenic diet (Naggert1)</td>
<td>6.1</td>
<td>Resistant to diet-induced obesity and loss of brown adipose tissue (TG)</td>
<td>[51]</td>
</tr>
<tr>
<td>Dusp1</td>
<td>17</td>
<td>25,910,596</td>
<td>Percent of tissue mass that is fat, after 8 wks on atherogenic diet (Naggert1)</td>
<td>7.4</td>
<td>Resistant to diet-induced obesity but succumb to glucose intolerance on a high fat diet (KO)</td>
<td>[52]</td>
</tr>
<tr>
<td>Cdh2</td>
<td>18</td>
<td>16,917,241</td>
<td>Liver weight at sacrifice, after 8 wks on atherogenic diet (Paigen1)</td>
<td>8.2</td>
<td>Increased adiposity (TG)</td>
<td>[25]</td>
</tr>
</tbody>
</table>

*aGenes were located within the identified QTLs.

*bPhysical position (bp) was based on the NCBI mouse genome build 36.1.

*cTraits were analyzed in the GWA scans; name of the PHENOME projects was indicated in the parentheses.

*dRelevant phenotypes were observed in murine knockout (KO) and/or transgenic (TG) models.

d doi:10.1371/journal.pone.0000651.t002
cholesterol, low density lipoprotein (LDL) cholesterol, and triglycerides. These genes are Gpr39 (G protein-coupled receptor 39), Apoa2 (apolipoprotein A-II), Cd36 (CD36 antigen), Cckar (cholecystokinin A receptor), Npy (neuropeptide Y), Pparg (peroxisome proliferator activated receptor gamma), Ccr2 (chemokine (C-C motif) receptor 2), Acaca (acetyl-Coenzyme A carboxylase alpha), Apob (apolipoprotein B), Soat2 (sterol O-aeryltransferase 2) and Lipg (endothelial lipase) (Table 3). All of these genes were observed to have strong involvement in plasma lipid phenotypes in murine knockout and/or transgenic models. Notably, we identified a QTL containing Apoa2 on chromosome 1 located at 173.0–173.9 Mb, which affects the percent of total plasma cholesterol in the HDL fraction after 8 weeks on an atherogenic diet (-log(P) = 7.9) (Figure 4B). Apoa2 is located within a strong HDL QTL, hdlq5, which has been repeatedly detected by linkage analyses in 12 different genetic crosses. Relevant phenotypes for these two genes were already observed in murine knockout models. The horizontal gray lines indicate genome-wide empirical thresholds.

doi:10.1371/journal.pone.0000651.g004

Adam12 and Cdh2 mapped with single-gene resolution by GWA scans

GWA scans have the potential to achieve single-gene resolution, depending on local LD patterns and gene density on the mouse genome. A closer view of the above 10 obesity candidate-gene regions (Table 2) further identified two QTLs that were mapped with single-gene resolution. One QTL significantly affects the percent of fat mass after 8 weeks on an atherogenic diet (-log(P) = 6.5) and is located on chromosome 7 spanning less than 100 kb where Adam12 is the only candidate gene (Figure 5A). Meltrin alpha (Adam12) is a metalloprotease-disintegrin that belongs to the ADAM (for “a disintegrin and metalloprotease”) family. Compared with wild-type mice, meltrin alpha(-/-) mice displayed moderate resistance to weight gain induced by a high-fat diet, mainly because of an increase in the number of adipocytes [24]. Another QTL is significantly associated with mouse liver weight after 8 weeks on an atherogenic diet (-log(P) = 8.2) and is located on chromosome 18, spanning less than 100 kb, where Cdh2 is the only candidate gene (Figure 5B). N-cadherin (Cdh2)
GWA scans identified a QTL on chromosome 1 affecting percent knockout phenotypes on both obesity and plasma lipid levels. Among these 21 candidate genes (Gpr39 mapped in a pleiotropic QTL rated with murine knockout or transgenic models can the casual variants associated with the disease, are incorporated with murine knockout or transgenic models can the causal relationship be firmly established.

Table 3. Candidate genes for plasma lipid levels validated in murine knockout and/or transgenic models

<table>
<thead>
<tr>
<th>Gene*</th>
<th>Chr</th>
<th>QTL peak</th>
<th>Analyzed trait</th>
<th>-log(P)</th>
<th>Phenotypes in KO or TG mice</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gpr39</td>
<td>1</td>
<td>127,218,849</td>
<td>Total cholesterol, after 17 wks on atherogenic diet (Paigen2)</td>
<td>8.8</td>
<td>Increased 11% and 42% cholesterol levels in males and females, respectively (KO)</td>
<td>[27]</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>127,775,854</td>
<td>Fold change in total cholesterol after 17 wks on atherogenic diet (Paigen2)</td>
<td>6.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>127,775,854</td>
<td>Non-HDL cholesterol, after 17 wks on atherogenic diet (Paigen2)</td>
<td>9.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apoa2</td>
<td>1</td>
<td>173,030,894</td>
<td>Percent of total plasma cholesterol in HDL fraction, after 8 wks on atherogenic diet (Paigen1)</td>
<td>7.9</td>
<td>Reduced HDL cholesterol levels and HDL particle size (KO)</td>
<td>[23]</td>
</tr>
<tr>
<td>Cd36</td>
<td>5</td>
<td>16,119,658</td>
<td>Non-HDL cholesterol, after 17 wks on atherogenic diet (Paigen2)</td>
<td>7.1</td>
<td>Increased plasma free fatty acid and triglyceride levels and decreased glucose levels (KO)</td>
<td>[53]</td>
</tr>
<tr>
<td>Cckar</td>
<td>5</td>
<td>53,817,933</td>
<td>Triglycerides (Paigen4)</td>
<td>7.5</td>
<td>Increased cholesterol gallstone formation (KO)</td>
<td>[54]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>54,607,854</td>
<td>Triglycerides (Paigen2)</td>
<td>7.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Npy</td>
<td>6</td>
<td>50,629,612</td>
<td>Triglycerides (Paigen4)</td>
<td>7.7</td>
<td>Decreased 18% and 27% Triglycerides in fasted and refed mice (KO)</td>
<td>[55]</td>
</tr>
<tr>
<td>Pparg</td>
<td>6</td>
<td>116,114,107</td>
<td>Hepatic cholesterol concentration, after 8 wks on atherogenic diet (Paigen1)</td>
<td>6.9</td>
<td>Reduced cholesterol efflux and HDL levels (KO)</td>
<td>[56]</td>
</tr>
<tr>
<td>Ccr2</td>
<td>9</td>
<td>123,799,254</td>
<td>Non-HDL cholesterol, after 17 wks on atherogenic diet (Paigen2)</td>
<td>6.8</td>
<td>Decreased 20% and 17% total cholesterol levels in low- and high-fat diet, respectively (KO)</td>
<td>[57]</td>
</tr>
<tr>
<td>Acaca</td>
<td>11</td>
<td>85,561,821</td>
<td>Non-HDL cholesterol, after 17 wks on atherogenic diet (Paigen2)</td>
<td>6.8</td>
<td>Decrease in the de novo fatty acid synthesis and triglyceride accumulation in the liver (KO)</td>
<td>[58]</td>
</tr>
<tr>
<td>Apob</td>
<td>12</td>
<td>7,813,147</td>
<td>Hepatic cholesterol concentration, after 8 wks on atherogenic diet (Paigen1)</td>
<td>7.1</td>
<td>Decreased ~37–39% total cholesterol, LDL, and HDL levels (KO)</td>
<td>[59]</td>
</tr>
<tr>
<td>Soat2</td>
<td>15</td>
<td>101,357,816</td>
<td>Hepatic cholesterol concentration, after 8 wks on atherogenic diet (Paigen1)</td>
<td>6.5</td>
<td>Resistance to diet-induced hypercholesterolemia and gallstone formation (KO)</td>
<td>[60]</td>
</tr>
<tr>
<td>Lipg</td>
<td>18</td>
<td>74,969,883</td>
<td>Triglycerides (Paigen4)</td>
<td>7.2</td>
<td>Fasting plasma HDL cholesterol was increased by 57% in Lipg+/– and 25% in Lipg−/− mice and was decreased by 19% in Lipg transgenic mice (KO/TG)</td>
<td>[61]</td>
</tr>
</tbody>
</table>

is a classical cadherin from the cadherin superfamily. Targeted expression of a dominant-negative N-cadherin in vivo delays peak bone mass and increases adipogenesis when compared with wild-type littermates [25]. Since Adam12 and Cdh2 are the only candidate gene in each of their QTL regions and further supported by the evidence from knockout and/or transgenic mice, we can safely establish these two QTL genes as casual genetic variants for atherogenic diet-induced obesity in natural inbred mouse populations. It should be noted that evidence solely from knockout or transgenic mice is not sufficient to infer causal relationship between a particular gene and disease in natural populations since the gene that was knocked out or overexpressed in mice may be only one of the several genes involved in the same biological pathway leading to the disease, rather than the true gene that is mutated in natural populations. Only when genetic mapping data, especially from high-resolution QTL analyses that are overlapped QTLs are indistinguishable which are very likely controlled by the same gene with pleiotropic effects on obesity and plasma lipid levels. The critical region of this pleiotropic QTL spans about 0.7 Mb (127.2–127.9 Mb) and covers six annotated genes based on the latest NCBI mouse genome build 36.1, including Actr3, Sla35f5, Gpr39, Loc667043, Lypd1 and E030049G20Rik. Gpr39 is a member of a family that includes the receptors for ghrelin and motilin [26]. Mature body weight and percent fat mass were increased by 38% and 5%, respectively, in Gpr39 loss-of-function (Gpr39−/−) mice; cholesterol levels were increased by 11% and 42% in male and female Gpr39−/− mice, respectively [27]. This establishes Gpr39 as a candidate for this pleiotropic QTL affecting both obesity and plasma lipid levels, even in light of the fact of recent observations clouding its role [28–31]. Nevertheless, these remarkable phenotypic differences between Gpr39−/− and wild-type mice mentioned above were only observed at old age (about 50 weeks for obesity and 25 weeks for cholesterol) [27], suggesting that deleterious mutations in Gpr39 may increase the risk for late-onset obesity and atherosclerotic coronary heart disease.

Gpr39 mapped in a pleiotropic QTL

Among these 21 candidate genes (Table 2 and 3), Gpr39 has knockout phenotypes on both obesity and plasma lipid levels. Our GWA scans identified a QTL on chromosome 1 affecting percent fat mass on atherogenic diet (-log(P) = 6.1). This QTL colocalizes with another QTL identified by our GWA scans, which affects total cholesterol and non-HDL cholesterol levels, and fold change in total cholesterol levels after 17 weeks on atherogenic diet (-log(P) = 6.6–9.9) (Figure 6). The genomic locations of these two overlapped QTLs are indistinguishable which are very likely controlled by the same gene with pleiotropic effects on obesity and plasma lipid levels. The critical region of this pleiotropic QTL spans about 0.7 Mb (127.2–127.9 Mb) and covers six annotated genes based on the latest NCBI mouse genome build 36.1, including Actr3, Slc35f5, Gpr39, Loc667043, Lypd1 and E030049G20Rik. Gpr39 is a member of a family that includes the receptors for ghrelin and motilin [26]. Mature body weight and percent fat mass were increased by 38% and 5%, respectively, in Gpr39 loss-of-function (Gpr39−/−) mice; cholesterol levels were increased by 11% and 42% in male and female Gpr39−/− mice, respectively [27]. This establishes Gpr39 as a candidate for this pleiotropic QTL affecting both obesity and plasma lipid levels, even in light of the fact of recent observations clouding its role [28–31]. Nevertheless, these remarkable phenotypic differences between Gpr39−/− and wild-type mice mentioned above were only observed at old age (about 50 weeks for obesity and 25 weeks for cholesterol) [27], suggesting that deleterious mutations in Gpr39 may increase the risk for late-onset obesity and atherosclerotic coronary heart disease.
Figure 5. Adam12 and Cdh2 were mapped to single-gene resolution. (A) Adam12 is associated with the percent of fat mass after 8 weeks on an atherogenic diet (Naggert1_pctfat_fat8). (B) Cdh2 is associated with liver weight after 8 weeks on an atherogenic diet (Paigen1_liverweight). Relevant phenotypes for these two genes were already observed in murine knockout or transgenic models. In each subfigure, the upper panel is the enhanced view of LD mapping in a short region around the LD peak, and the lower panel is the physical map of the associated refined QTL. Genes shown on the upper side of the chromosome (turquoise lines) are transcribed in the–orientation (from right to left), and those on the lower side (pink lines) in the+orientation (from left to right). The horizontal gray lines indicate genome-wide empirical thresholds. Physical position was based on the latest NCBI mouse genome build 36.1.
doi:10.1371/journal.pone.0000651.g005
DISCUSSION

We have demonstrated the robustness and reproducibility of GWA analysis by comparing QTLs derived from GWA with those derived from previous linkage analyses in inbred mice. In a recent GWA analysis of lung tumor incidence, we reproduced the pulmonary adenoma susceptibility 1 (Pas1) locus identified in previous linkage studies and further narrowed this QTL to a region of less than 0.5 Mb in which at least two genes, Kras2 and Casc1, are strong candidates [4]. The refined region is completely coincident with that from traditional fine mapping by using congenic strains of mice [32]. Casc1 knockout mouse tumor bioassays further confirmed Casc1 as a primary candidate for the Pas1 locus [4]. In the present study, we again reproduced and narrowed the Bwtq1 locus to a region of less than 1.0 Mb where Pappa2 is a primary candidate. The refined Bwtq1 region is identical to that achieved by developing a series of congenic mice and interval-specific subcongenic mice [13,14]. These results demonstrate that GWA scans using dense SNP maps in laboratory mice are a powerful tool for the refinement of previously identified QTL regions. These results can be also served as an internal control for the methods that we used in our GWA scans. Given that known refined QTLs were exactly reproduced in the analyses, the newly refined QTLs identified by our GWA scans are expected to be a useful resource for further positional cloning and gene discovery in the mouse. It should be also noted that 937 QTLs identified in our study represent about one half of all QTLs derived from linkage analyses in mouse cross-breeding experiments in the past decades (http://www.informatics.jax.org).

Research tools such as genetrap mutations, transgenes, and standard knockout technology can be used to verify causal variants from highly-refined QTL regions immediately after GWA scans [33,34]. Using resources from murine gene-deficiency and/or transgenic models, we have identified 10 candidate genes affecting obesity-related phenotypes and 11 candidate genes affecting plasma cholesterol levels from a survey of several highly refined QTLs (Table 2 and 3). These QTL genes are most likely the causal genes based on the following evidence. First, the genomic regions where these genes are located are not only highly associated with the analyzed phenotypes in the GWA scans, but are also under the linkage peak from two or more independent linkage studies with crossing-breeding experiments [10,22]. Second, these candidates were identified from a highly refined genomic region, as narrow as 0.54 Mb (with an average of 6 genes per region), which is extremely more precise than those chosen from a much broader linkage region (about 20 Mb). Third, and more importantly, relevant phenotypes for these genes were observed in murine knockout and/or transgenic models. Notably, due to the single-gene resolution achieved by our GWA scans, we firmly establish two QTL genes, Adam12 and Cdh2, as causal variants for obesity in inbred mice. It is also worth noting that in the present study we only focused on 28 obesity and cholesterol QTLs out of 937 QTLs, which have available data from murine knockout and/or transgenic models (Table 2 and 3). Other QTLs...
genes identified in this study are now ready to be evaluated for their functional relevance to the analyzed mouse phenotypes. We have summarized this extremely valuable QTL resource in online Table S1 and S2; this resource can greatly facilitate positional cloning and the identification of new genetic determinants of complex traits.

Finally, several caveats for our findings should be mentioned. Firstly, selection and inbreeding play important roles in the formation of the genetic architecture of mouse inbred strains. Potential population structure may arise in these mouse strains, which inflates type I error rates and may lead to spurious associations in the analysis [3]. Wild-derived inbred mice have divergent evolutionary histories and thus have strong potential to generate spurious associations. There are systematic differences in obesity-related trait values between wild-derived and other inbred strains. Therefore, the population structure was carefully inspected in our association analysis, and we removed wild-derived inbred strains from the analysis when population structure was detected. Spurious associations were largely reduced in the analysis after the removal of wild-derived inbred mouse strains (Figure S7 and Text S1). Several methods have been developed that adjust for the effect of population structure on spurious associations [35–30], but these may not be sufficient for the analysis of a small number of mouse inbred strains (~30–40 strains) [39]. Secondly, we used permutations to establish stringent genome-wide thresholds for declaring significant association for each phenotype. Although most genomic regions show correct type I error rates, some regions may show increased error rates due to unequal relatedness among the strains [40]. The risk of these spurious associations can be alleviated when incorporating prior linkage evidence into the analysis. The identified QTL genes located within previous linkage-defined regions from mouse crosses which may show increased error rates due to unequal relatedness among the strains [40]. The risk of these spurious associations can be alleviated when incorporating prior linkage evidence into the analysis. The identified QTL genes located within previous linkage-defined regions from mouse crosses which may show increased error rates due to unequal relatedness among the strains [40]. The risk of these spurious associations can be alleviated when incorporating prior linkage evidence into the analysis. The identified QTL genes located within previous linkage-defined regions from mouse crosses which may show increased error rates due to unequal relatedness among the strains [40]. The risk of these spurious associations can be alleviated when incorporating prior linkage evidence into the analysis. The identified QTL genes located within previous linkage-defined regions from mouse crosses which may show increased error rates due to unequal relatedness among the strains [40]. The risk of these spurious associations can be alleviated when incorporating prior linkage evidence into the analysis. The identified QTL genes located within previous linkage-defined regions from mouse crosses which may show increased error rates due to unequal relatedness among the strains [40]. The risk of these spurious associations can be alleviated when incorporating prior linkage evidence into the analysis. The identified QTL genes located within previous linkage-defined regions from mouse crosses which may show increased error rates due to unequal relatedness among the strains [40]. The risk of these spurious associations can be alleviated when incorporating prior linkage evidence into the analysis. The identified QTL genes located within previous linkage-defined regions from mouse crosses which may show increased error rates due to unequal relatedness among the strains [40].

Whole-genome LD patterns
The commonly used pairwise LD measure, r², which is the squared correlation coefficient of alleles at two loci [42], was calculated for pairwise SNPs on the mouse genome. To capture large-scale patterns of LD, we measured the number of proxies, i.e. SNPs showing a strong correlation with one or more others [9], with a window size of 500 kb across the genome.

GWA scans
A full description of statistical methods for GWA scans is provided in Text S1. Briefly, the associations of a mouse phenotype with SNP markers on the genome were tested by using linear regression models. To maximize power in GWA scans, we examined several competitive models with different variables based on Bayesian information criteria. The log-likelihood-ratio test statistic for the existence of a QTL is calculated by comparing the likelihood values under full model with null model. In the results, the negative 10-base logarithmic p value from x² tests, i.e. –log(P), was presented. Population structure was investigated by the cumulative distribution of p values in genome-wide association analysis for each phenotype [35]. We removed wild-derived inbred strains in the association analysis when population structure was detected.

In the GWA scans, 1,000 permutations were used to establish a genome-wide threshold (a global p value of 0.05) for declaring significant associations for each phenotype. Specifically, the analyzed phenotype was randomly reshuffled among subjects while fixing the genotypes. For each of the 1,000 permutations, the GWA scan was repeated, and the most significant –log(P) was recorded. Sorting the maximum –log(P) from large to small, the 5% quantile of the empirical distribution was taken as the genome-wide threshold (a global p value of 0.05). Since LD extends a very short genomic region (normally less than a few hundred kb) on the mouse genome, once a significant SNP is identified, a causal variant responsible for the QTL should be close to the identified SNP. To define
approximate QTL regions, we first identified significant SNPs with an established empirical threshold. If two significant SNPs were less than 1,000 kb from each other, they were treated as one association signal. Then these genomic regions were extended to cover closely linked SNPs that showed suggestive associations with the analyzed traits. We also reported p values, physical positions and genomic domains for each significant SNP. The identified QTLs were compared with those detected by previous linkage analyses from mouse crosses which were retrieved from Mouse Genome Informatics (MGI) at http://www.informatics.jax.org.

**Data transformation**
To avoid the violation of assumption of normal distribution in the above statistical analysis, 173 quantitative phenotypes were checked for their normality by Shapiro-Wilk tests [43] before GWA scans. Phenotypes with Shapiro-Wilk tests P<0.05 were converted to approach normality by Box-Cox transformation or normal scores [44,45]. For those phenotypes still showing marked departure from the normal distribution (p<0.01) after data transformation, the Mann-Whitney rank tests were performed in departure from the normal distribution (p

**Heritability**
The heritability of inbred strains can be estimated by linear regression models with sex, strain and sex-strain interaction as independent variables. Each strain was treated as a categorical variable in the linear model analyses. The difference between the adjusted R² of the regression models with and without strain and its associated sex-strain interaction is approximately equal to the percentage of genetic variation (that is, heritability) in inbred mouse populations. Similarly, the phenotypic variation due to sex differences can also be estimated.

**Phylogenetic analysis**
The phylogenetic tree was constructed with 148,062 SNPs from 59 inbred mouse strains, implemented in the *dndist* program in the PHYLIP 3.66 package [http://evolution.genetics.washington.edu/phylip.html]. Branch length information was used to plot evolutionary distance between strains in the drawtree program in the PHYLIP 3.66 package.

**SUPPORTING INFORMATION**

**Text S1**  Supporting methods: A detailed description of methods and materials.
Found at: doi:10.1371/journal.pone.0000651.s001 (0.11 MB DOC)

**Table S1**  173 mouse quantitative traits analyzed in genome-wide association analyses. The table gives empirical genome-wide threshold, the number of QTLs and significant SNPs detected in GWA scans for each trait. The table also lists project names, mouse phenome database (MPD) accession numbers, short and full name of traits, and categories of traits.
Found at: doi:10.1371/journal.pone.0000651.s002 (0.36 MB DOC)

**Table S2**  Summary of 937 QTL identified in genome-wide association analyses. The table gives the chromosome, position of QTL peak (bp, based on the NCBI genome build 36), QTL interval, and -log(P) for each QTL identified in GWA scans. The table also provides the SNP ID of the most significant SNP and its dbSNP's annotation if available, and total number of significant and suggestive SNPs for each QTL.
Found at: doi:10.1371/journal.pone.0000651.s003 (1.99 MB DOC)

**Figure S1**  Distribution of SNP positions and LD structure across the mouse genome. For each chromosome, the top panel shows number of SNPs in a window size of 300 kb. The bottom panel shows number of proxies (red lines for r²>0.5 and blue lines for r²>0.8) per SNP in a window size of 500 kb.
Found at: doi:10.1371/journal.pone.0000651.s004 (0.41 MB PDF)

**Figure S2**  Characteristics of the mouse SNP map. (A) Distribution of SNP allele frequency. (B) Distribution of inter-SNP distances. (C) LD decay as a function of physical distance. (D) Number of proxies per SNP in a window size of 500 kb, as a function of the threshold for correlation (r²).
Found at: doi:10.1371/journal.pone.0000651.s005 (0.90 MB PDF)

**Figure S3**  Phylogenetic tree of 59 inbred mouse strains. The phylogenetic tree was constructed with a total of 148,062 SNPs from the WTTCHG and Broad Institute, implemented in the dndist program in the PHYLIP 3.66 package [http://evolution.genetics.washington.edu/phylip.html]. The branch length information was used to plot evolutionary distance between strains in the drawtree program. 59 inbred mouse strains are organized into seven groups: Bagg albino derivatives, C57-related strains, Castle’s mice, Japanese and New Zealand inbred strains, Little’s DBA and related strains, Swiss mice, and wild-derived strains.
Found at: doi:10.1371/journal.pone.0000651.s006 (1.41 MB PDF)

**Figure S4**  Characteristics of 173 complex quantitative traits in inbred mice. (A) Heritability of quantitative traits. (B) Phenotypic variation due to sex effects.
Found at: doi:10.1371/journal.pone.0000651.s007 (0.70 MB PDF)

**Figure S5**  An in silico strategy for high-throughput gene discovery in inbred mice. GWA scans were implemented in an automatic processing pipeline which constitutes data retrieving, outlier detection, data preprocessing, hypothesis testing, permutations and QTL identification.
Found at: doi:10.1371/journal.pone.0000651.s008 (0.21 MB PDF)

**Figure S6**  Genome-wide association analysis of several obesity-related phenotypes in inbred mice. The obesity-related phenotypes are (A) body weight at the start of testing (8 weeks) (Tordoff3_bw_start); (B) calculated weight of lean tissue (14 weeks) (Tordoff3_lean wt); (C) body weight after 8 weeks on an atherogenic diet (Nagger1_bw_fat8); (D) total tissue mass after 8 weeks on an atherogenic diet (Nagger1_tissuemass_fat8); (E) weight of lean portion of tissue mass after 8 weeks on an atherogenic diet (Nagger1_leanwt_fat8); (F) bone mineral content after 8 weeks on an atherogenic diet (Nagger1_BMC_fat8); (G) initial body weight (7–9 weeks), day 0 of an atherogenic diet (Paigen1_inibw); and (H) final body weight after 8 weeks on an atherogenic diet (Paigen1_finalbw). In Paigen1 and Nagger1 projects, mice at 7–9 weeks of age were weighed and then administered a high fat, high cholesterol atherogenic diet. The scatter plots were drawn for -log(P) against the SNP position in the chromosomes. The horizontal gray lines indicate genome-wide empirical thresholds (global p value = 0.05). The horizontal coordinates were plotted using physical distance (Mb) in each panel.
Found at: doi:10.1371/journal.pone.0000651.s009 (7.96 MB PDF)
REFERENCES


14. Brockmann GA, Haley CS, Schwerin M, et al. (2000) Associated autosomal M-1–M-5 linkage for body weight after 8 weeks on altherogenic diet (Naggert1_bw_fat8) was used to illustrate the population structure in inbred mice. (A) Cumulative distribution of p values from GWA analysis of samples with (red line) and without (blue line) wild-derived inbred strains. cdf, cumulative distribution function. (B) Comparison of GWA analysis of samples with (upper, red) and without (lower, blue) wild-derived inbred strains. The two green horizontal lines are genome-wide thresholds (a global p value of 0.05). Spurious associations were largely reduced in the analysis after the removal of wild-derived inbred mouse strains. Found at: doi:10.1371/journal.pone.0000651.s011 (0.48 MB PDF).


16. Brockmann GA, Haley CS, Renne U, Schwerin M, et al. (1998) Association analysis with (red line) and without (blue line) wild-derived inbred strains. The two green horizontal lines are genome-wide thresholds (a global p value of 0.05). Spurious associations were largely reduced in the analysis after the removal of wild-derived inbred mouse strains. Found at: doi:10.1371/journal.pone.0000651.s010 (1.78 MB PDF).


ACKNOWLEDGMENTS

We thank the Jackson Laboratory and investigators (a detailed list of investigators of the mouse PHENOME projects is provided in the Text S1) for collecting and phenotyping inbred mouse strain characteristics. We thank the Broad Institute of Harvard/Massachusetts General Hospital and MIT, the Wellcome Trust Center for Human Genetics at Oxford University, and Perlegen Sciences for releasing inbred laboratory mouse SNP data. The mouse phenotype and SNP data were crucial for the key findings described in this paper.

Author Contributions

Conceived and designed the experiments: MY PL. Performed the experiments: PL HV. Analyzed the data: PL DW YL. Wrote the paper: PL. Other: Revise the manuscript HV.