PLOS | GENETICS

Meta-Analysis Identifies Gene-by-Environment Interactions as Demonstrated in a Study of 4,965 Mice

Eun Yong Kang[1][9], Buhm Han[2,3,4][9], Nicholas Furlotte[1][9], Jong Wha J Joo[5], Diana Shih[6], Richard C. Davis[6], Aldons J. Lusis[6,7][¶], Eleazar Eskin[1,7][¶]*

1 Department of Computer Science, University of California, Los Angeles, Los Angeles, California, United States of America, 2 Division of Genetics, Brigham & Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America, 3 Partners HealthCare Center for Personalized Genetic Medicine, Boston, Massachusetts, United States of America, 4 Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, United States of America, 5 Interdepartmental Program in Bioinformatics, University of California, Los Angeles, Los Angeles, California, United States of America, 6 Department of Medicine, University of California, Los Angeles, Los Angeles, California, United States of America, 7 Department of Human Genetics, University of California, Los Angeles, Los Angeles, California, United States of America

Abstract

Identifying environmentally-specific genetic effects is a key challenge in understanding the structure of complex traits. Model organisms play a crucial role in the identification of such gene-by-environment interactions, as a result of the unique ability to observe genetically similar individuals across multiple distinct environments. Many model organism studies examine the same traits but under varying environmental conditions. For example, knock-out or diet-controlled studies are often used to examine cholesterol in mice. These studies, when examined in aggregate, provide an opportunity to identify genomic loci exhibiting environmentally-dependent effects. However, the straightforward application of traditional methodologies to aggregate separate studies suffers from several problems. First, environmental conditions are often variable and do not fit the standard univariate model for interactions. Additionally, applying a multivariate model results in increased degrees of freedom and low statistical power. In this paper, we jointly analyze multiple studies with varying environmental conditions using a meta-analytic approach based on a random effects model to identify loci involved in gene-by-environment interactions. Our approach is motivated by the observation that methods for discovering gene-by-environment interactions are closely related to random effects models for meta-analysis. We show that interactions can be interpreted as heterogeneity and can be detected without utilizing the traditional uni- or multi-variate approaches for discovery of gene-by-environment interactions. We apply our new method to combine 17 mouse studies containing in aggregate 4,965 distinct animals. We identify 26 significant loci involved in High-density lipoprotein (HDL) cholesterol, many of which are consistent with previous findings. Several of these loci show significant evidence of involvement in gene-by-environment interactions. An additional advantage of our meta-analysis approach is that our combined study has significantly higher power and improved resolution compared to any single study thus explaining the large number of loci discovered in the combined study.

Introduction

Identifying environmentally specific genetic effects is a key challenge in understanding the structure of complex traits. In humans, gene-by-environment (GxE) interactions have been widely discussed [1–12] yet only a few have been replicated. One reason for this discrepancy is the inability to accurately control for environmental conditions in humans as well as the inability to observe the same individuals in multiple distinct environments. Model organisms do not share such difficulties and for this reason can play a crucial role in the identification of gene-by-environment interactions. For example, in many mouse genetic studies the same traits are examined under different environmental conditions. Specifical-ly, knock-out or diet-controlled mice are often utilized in the study of cholesterol levels. The availability of these studies presents a unique opportunity to identify genomic loci involved in gene-by-environment interactions as well as those loci involved in the trait independent of the environment.

In order to utilize genetic studies in model organisms to identify gene-by-environment interactions, one needs to directly compare the effects of genetic variations in studies conducted under different conditions. This practice is complicated for a number of reasons, when combining more than two studies. First, environmental conditions are often variable across studies and do not fit to the standard univariate model for interactions. For example, in one study, cholesterol may be examined under different diet conditions

Author Summary

Identifying gene-by-environment interactions is important for understand the architecture of a complex trait. Discovering gene-by-environment interaction requires the observation of the same phenotype in individuals under different environments. Model organism studies are often conducted under different environments. These studies provide an unprecedented opportunity for researchers to identify the gene-by-environment interactions. A difference in the effect size of a genetic variant between two studies conducted in different environments may suggest the presence of a gene-by-environment interaction. In this paper, we propose to employ a random-effect-based meta-analysis approach to identify gene-by-environment interaction, which assumes different or heterogeneous effect sizes between studies. Our approach is motivated by the observation that methods for discovering gene-by-environment interactions are closely related to random effects models for meta-analysis. We show that interactions can be interpreted as heterogeneity and can be detected without utilizing the traditional approaches for discovery of gene-by-environment interactions, which treats the gene-by-environment interactions as covariates in the analysis. We provide a intuitive way to visualize the results of the meta-analysis at a locus which allows us to obtain the biological insights of gene-by-environment interactions. We demonstrate our method by searching for gene-by-environment interactions by combining 17 mouse genetic studies totaling 4,965 distinct animals.

(eg. low fat and high fat) and then in another study cholesterol is examined using gene knockouts. In this case, it is not straightforward to analyze these studies in aggregate using a single variable to represent the environmental condition. Applying a multivariate model, one in which the environment is represented using multiple environmental variables, results in increased degrees of freedom and low statistical power. Second, model organisms such as the mouse exhibit a large degree of population structure. Population structure is well-known for causing false positives and spurious associations [13,14] in association analysis and can be expected to complicate the ability to combine separate studies.

In this paper, we propose a random-effects based meta-analytic approach to combine multiple studies conducted under varying environmental conditions and show that this approach can be used to identify both genomic loci involved in gene-by-environment interactions as well as those loci involved in the trait independent of the environment. By making the connection between gene-by-environment interactions and random effects model meta-analysis, we show that interactions can be interpreted as heterogeneity and detected without requiring uni- or multi-variate models. We also define an approach for correcting population structure in the random effects model meta-analysis, extending the methods developed for fixed effects model meta-analysis [15]. We show that this method enables the analyses of large scale meta-analyses with dozens of heterogeneous studies and leads to dramatic increases in power. We demonstrate that insights regarding gene-by-environment interactions are obtained by examining the differences in effect sizes among studies facilitated by the recently developed m-value statistic [16], which allows us to distinguish between studies having an effect and studies not having an effect at a given locus.

We applied our approach, which we refer to as Meta-GxE, to combine 17 mouse High-density lipoprotein (HDL) studies

containing 4,965 distinct animals. To our knowledge, this is the largest mouse genome-wide association study conducted to date. The environmental factors of the 17 studies vary greatly and include various diet conditions, knock-outs, different ages and mutant animals. By applying our method, we have identified 26 significant loci. Consistent with the experience of meta-analysis in human studies, our combined study finds many loci which were not discovered in any of the individual studies. Among the 26, 24 loci have been previously implicated in having an effect on HDL cholesterol or closely related lipid levels in the blood, while 2 loci are novel findings. In addition, our study provides insights into genetic effects on several disease loci and their relationship between environment and sex. For example, we identified 3 loci (Chr10: 21399819, Chr19:3319089, ChrX:151384614), where female mice show a more significant effect on HDL phenotypes than male mice. We also identified 7 loci (Chr1:171199523, Chr8:46903188, Chr8: 64150094, Chr8:84073148, Chr10:90146088, Chr11:69906552, Chr15:21194226) where male mice show a more significant effect on HDL than female mice. In addition, many of the loci show strong gene-by-environment interactions. Using additional information describing the studies and our predictions of which studies do and do not contain an effect, we gain insights into the interaction. For example, locus on chromosome 8 (Chr8:84073148) shows a strong sex by mutation-driven LDL level interaction, which affects HDL cholesterol levels.

Part of the reason for our success in identifying a large number of loci is that our study combined multiple mouse genetic studies many of which use very different mapping strategies. Over the past few years, many new strategies have been proposed beyond the traditional F2 cross [17] which include the hybrid mouse diversity panel (HMDP) [18,19], heterogeneous outbred stocks [20], commercially available outbred mice [21], and the collaborative cross [22]. In our current study, we are combining several HMDP studies with several F2 cross studies and benefit from the statistical power and resolution advantages of this combination [15]. The methodology presented here can serve as a roadmap for both performing and planning large scale meta-analysis combining the advantages of many different mapping strategies. Meta-GxE is publicly available at http://genetics.cs.ucla.edu/metagxe/.

Results

Discovering environmentally-specific loci using meta-analysis

The Meta-GxE strategy uses a meta-analytic approach to identify gene-by-environment interactions by combining studies that collect the same phenotype under different conditions. Our method consists of four steps. First, we apply a random effects model meta-analysis (RE) to identify loci associated with a trait considering all of the studies together. The RE method explicitly models the fact that loci may have different effects in different studies due to gene-by-environment interactions. Second, we apply a heterogeneity test to identify loci with significant gene-by-environment interactions. Third, we compute the *m-value* of each study to identify in which studies a given variant has an effect and in which it does not. Forth, we visualize the result through a forest plot and PM-plot to understand the underlying nature of gene-by-environment interactions.

We illustrate our methodology by examining a well-known region on mouse chromosome 1 harboring the *Apoa2* gene, which is known to be strongly associated with HDL cholesterol [23]. Figure 1 shows the results of applying our method to this locus. We first compute the effect size and its standard deviation for each of the 17 studies. These results are shown as a forest plot in Figure 1
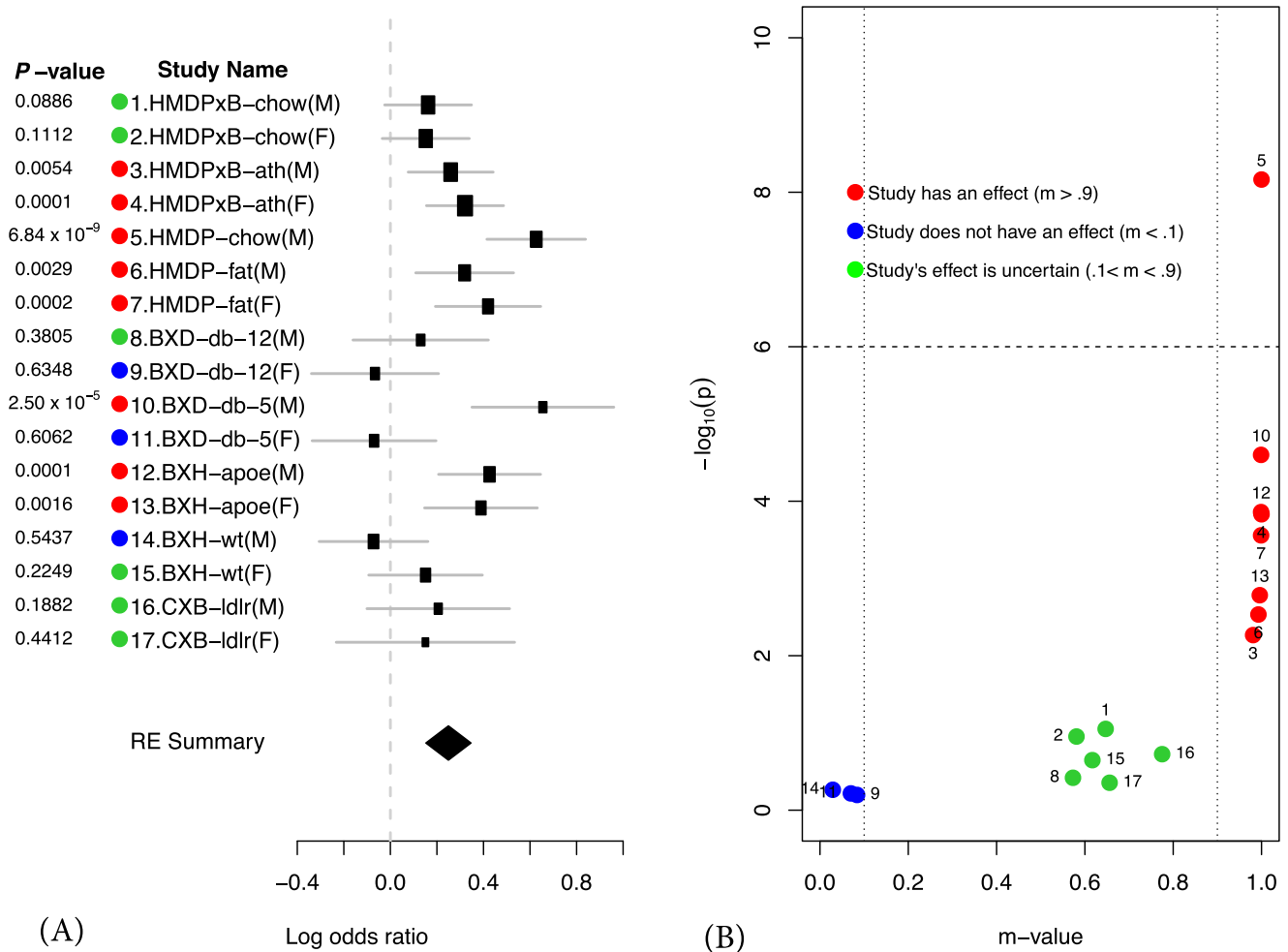
Figure 1. Application of Meta-GxE to *Apoa2* locus. The forest plot (A) shows heterogeneity in the effect sizes across different studies. The PM-plot (B) predicts that 7 studies have an effect at this locus, even though only 1 study (HMDP-chow(M)) is genome-wide significant with *P*-value.
doi:10.1371/journal.pgen.1004022.g001

(a). Second we compute the *P*-value for each individual study also shown in Figure 1 (a). If we were to follow traditional methodology and evaluate each study separately, we would declare an effect present in a study if the *P*-value exceeds a predefined genome-wide significance threshold ($P < 1.0 \times 10^{-6}$). In this case, we would only identify the locus as associated in a single study, HMDP-chow(M) ($P = 6.84 \times 10^{-9}$). On the other hand, in our approach, we combine all studies to compute a single *P*-value for each locus taking into account heterogeneity between studies. This approach leads to increased power over the simple approach considering each study separately. The combined meta *P*-value for the *Apoa2* locus is very significant (4.41×10^{-22}), which is consistent with the fact that the largest individual study only has 749 animals compared to 4,965 in our combined study.

In order to evaluate how significantly different the effect sizes of the locus are between studies, we apply a heterogeneity test. The statistical test is based on Cochran's Q test [24,25], which is a non-parametric test for testing if studies have the same effect or not. In this locus, the effect sizes are clearly different and not surprisingly the *P*-value of the heterogeneity test is significant (5.80×10^{-5}). This provides strong statistical evidence of a gene-by-environment

interaction at the locus. Below we more formally describe how heterogeneity in effect size at a given locus can be interpreted as gene-by-environment interaction.

If a variant is significant in the meta-analytic testing procedure, then this implies that the variant has an effect on the phenotype in one or more studies. Examining in which subset of the studies an effect is present and comparing to the environmental conditions of the studies can provide clues to the nature of gene-by-environment interactions at the locus. However, the presence of the effect may not be reflected in the study-specific *P*-value due to a lack of statistical power. Therefore, it is difficult to distinguish only by a *P*-value if an effect is absent in a particular study due to a gene-by-environment interaction at the locus or a lack of power. In order to identify which studies have effects, we utilize a statistic called the m-value [16], which estimates the posterior probability of an effect being present in a study given the observations from all other studies. We visualize the results through a PM-plot, in which *P*-values are simultaneously visualized with the m-values at each tested locus. These plots allow us to identify in which studies a given variant has an effect and in which it does not. M-values for a given variant have the following interpretation: a study with a

small m-value(≤ 0.1) is predicted not to be affected by the variant, while a study with a large m-value(≥ 0.9) is predicted to be affected by the variant.

The PM-plot for the *Apoa2* locus is shown in Figure 1 (b). If we only look at the separate study *P*-values (y-axis), we can conclude that this locus only has an effect in HMDP-chow(M). However, if we look at m-value (x-axis), then we find 8 studies (HMDPxB-ath(M), HMDPxB-ath(F), HMDP-chow(M), HMDP-fat(M), HMDP-fat(F), BxD-db-5(M), BxH-apoe(M), BxH-apoe(F)), where we predict that the variation has an effect, while in 3 studies (BxD-db-12(F), BxD-db-5(F), BxH-wt(M)) we predict there is no effect. The predictions for the remaining 6 studies are ambiguous.

From Figure 1, we observe that differences in effect sizes among the studies are remarkably consistent when considering the environmental factors of each study as described in Table 1. For example, when comparing study 1–4, the effect size of the locus decreases in both the male and female HMDPxB studies in the chow diet (chow study) relative to the fat diet (ath study). Thus we can see that when the mice have Leiden/CETP transgene, which cause high total cholesterol level and high LDL cholesterol level, effect size of this locus on HDL cholesterol level in blood is affected by the fat level of diet. Similarly, when comparing study 12–15, the knockout of the *Apoe* gene affects the effect sizes for both male and female BxH crosses. However, in the BxD cross (study 8–11), where each animal is homozygous for a mutation causing a deficiency of the leptin receptor, the effect of the locus is very strong in the young male animals, while as animals get older and become fatter, the effect becomes weaker. However in the case of female mice, the effect of the locus is nearly absent at both 5 and 12 weeks of age. Thus we can see that sex plays an important role in affecting HDL when the leptin receptor activity is deficient. We note that there are many genes in this locus and the genetic mechanism of interactions may involve genes other than *Apoa2*. Despite this caveat, the results of Meta-GxE at this locus provides insights into the nature of GxE and can provide a starting point for further investigation.

We note that an alternate explanation for differences in effect sizes between studies is the presence of gene-by-gene interactions and differences in the genetic backgrounds of the studies. While this is a possible explanation for differences in effect sizes between the different crosses and the HMDP studies, in Figure 1, we see many differences in effect sizes among studies with the same genetic background. Thus gene-by-gene interactions can only partially explain the differences in observed effect sizes.

The connection between random effects meta-analysis and gene-by-environment interactions

Gene-by-environment interactions, random effects meta-analysis and heterogeneity testing are closely related. Suppose we have k studies each conducted under different environmental conditions. We define the following linear model, where y_i is the observed phenotype for study i, α_i is the phenotype mean for study i, δ_i is the genetic effect on the phenotype for study i, X is the genotype, and e is the residual error.

$$y_i = \alpha_i + \delta_i X + e \qquad (1)$$

Since each environment is different, the effect size δ_i is partially determined by environmentally-specific factors and partially determined by factors common to all studies. Given that we can

Table 1. 17 HDL studies for meta analysis.

Study ID	Strains	Conditions	Age	Sex	# Strains	# Samples	# Sig Loci	Ref
HMDPxB-chow(M)	(HMDP×BL/6) F1	Leiden/CETP TG,chow diet	8 weeks	M	97	516	1	U
HMDPxB-chow(F)	(HMDP×BL/6) F1	Leiden/CETP TG, chow diet	8 weeks	F	95	468	0	U
HMDPxB-ath(M)	(HMDP×BL/6) F1	Leiden/CETP TG, highfat diet	24 weeks	M	97	408	0	U
HMDPxB-ath(F)	(HMDP×BL/6) F1	Leiden/CETP TG, highfat diet	24 weeks	F	93	457	3	U
HMDP-chow(M)	HMDP	chow diet	12 weeks	M	111	749	6	[18]
HMDP-fat(M)	HMDP	highfat diet	16 weeks	M	106	586	0	[14]
HMDP-fat(F)	HMDP	highfat diet	16 weeks	F	92	475	0	[44]
BxD-db-12(M)	(DBA×BL/6) F2	BXD db/db, chow diet	12 weeks	M	125	125	0	[45]
BxD-db-12(F)	(DBA×BL/6) F2	BXD db/db, chow diet	12 weeks	F	122	122	0	[45]
BxD-db-5(M)	(DBA×BL/6) F2	BXD db/db, chow diet	5 weeks	M	109	109	1	[45]
BxD-db-5(F)	(DBA×BL/6) F2	BXD db/db, chow diet	5 weeks	F	139	139	0	[45]
BxH-apoe(M)	(C3H×BL/6) F2	BXH Apoe -/-	24 weeks	M	161	161	0	[46]
BxH-apoe(F)	(C3H×BL/6) F2	BXH Apoe -/-	24 weeks	F	174	174	0	[46]
BxH-wt(M)	(C3H×BL/6) F2	BXH wildtype, highfat diet	20 weeks	M	164	164	0	[47]
BxH-wt(F)	(C3H×BL/6) F2	BXH wildtype, highfat diet	20 weeks	F	144	144	0	[47]
CxB-ldlr(M)	(BALB/cJ×BL/6) F2	CXB LDLR -/-, highfat diet	12 weeks	M	124	124	0	U
CxB-ldlr(F)	(BALB/cJ×BL/6) F2	CXB LDLR -/-, highfat diet	12 weeks	F	64	64	0	U

Seventeen HDL studies are combined in the meta analysis. U in the Ref column represents a data set that is not yet published. Mice for the HMDPxB panel were created by breeding females of the various HMDP inbred strains to males carrying transgenes for both Apoe Leiden and for human Cholesterol Ester Transfer Protein (CETP) on a C57BL/6 genetic background. The Leiden/CETP transgenes [48,49] cause high total cholesterol level and high LDL cholesterol level in the circulation, along with reduced HDL cholesterol. BxD db/db denotes a population of F2 mice from a cross between C57BL/6 DBA/2 with homozygous deficiency in leptin receptor (db/db), which results in obese mice. BxH Apoe -/- denotes denotes a population of F2 mice from a cross between C57BL/6 and C3H also carrying a deficiency in apolipoprotein E. CxB LDLR -/- denotes a population of F2 mice from a cross between C57BL/6 and BALB/cBy also carrying a deficiency in LDL receptor, which results in high LDL cholesterol level in the circulation BXH wildtype denotes a population of F2 mice from a cross between C57BL/6 and C3H.
doi:10.1371/journal.pgen.1004022.t001

decompose the effect δ_i into environment-independent and environment-dependent factors. Then we define the following linear model, where β is the environment-independent genetic effect and γ_i is the environment-dependent genetic effect for study i.

$$y_i = \alpha_i + \beta X + \gamma_i X + e \qquad (2)$$

In order to test for the presence of an effect shared across environments, we test the null hypothesis $\beta = 0$ and to test for the presence of a gene-by-environment interaction, we test the hypothesis that $\gamma_i = 0$.

In the random effects meta-analysis, we assume that the effect size δ_i is sampled from a normal distribution with mean μ and variance τ^2, denoted $\delta_i \sim N(\mu, \tau^2)$. Under this assumption, we test the null hypothesis $\mu = 0$ and $\tau^2 = 0$, in order to obtain a study-wide P-value. Additionally, we perform a heterogeneity test to test the null hypothesis $\delta_1 = \ldots = \delta_k$ versus the alternative hypothesis $NOT(\delta_1 = \ldots = \delta_k)$. We posit that by conducting hypotheses tests in the meta-analysis framework, we are simultaneously testing for the presence of environmentally-independent and environmentally-specific effects and that by applying heterogeneity testing we are testing for only environmentally-specific effects.

Consider that in the meta-analysis framework μ is analogous to β and the variation (τ^2) around μ is analogous to variation among γ_is. In the random effects meta analysis testing framework we are testing if $\mu = 0$ and $\tau^2 = 0$. This is equivalent to testing both environmentally-independent ($\beta = 0$) and environmentally-dependent ($\gamma_i = 0$) effects simultaneously. In heterogeneity testing, we test the null hypothesis $\delta_1 = \ldots = \delta_k$ versus the alternative hypothesis $NOT(\delta_1 = \ldots = \delta_k)$. When the environmentally-dependent effect (γ_i) is 0 it means that $\tau^2 = 0$ and thus $\delta_1 = \ldots = \delta_k$. When $\tau^2 \neq 0$, we expect that δ_i will vary around μ, so that we do not expect that $\delta_i = \delta_j$. Since the variation (τ^2) of δ_i around μ is analogous to the variable γ_i, heterogeneity testing in the meta-analysis framework is approximately equivalent to testing for environmentally-specific effects.

Gene-by-environment interactions are prevalent in mouse association studies

The presence of heterogeneity in the effect size at causal genetic loci due to gene-by-environment interactions is naturally expected in mouse genetic studies when combining studies with varying environmental conditions. One extreme example comes from a knock-out experiment. If the knocked-out gene is causal for a particular trait, then we can expect that the gene would have no effect on a knock-out mouse, while the gene would have an effect on the wild type mouse. This is a binary form of heterogeneity. In a less extreme form of heterogeneity, the effect of a given gene may be affected by an environmental factor which varies in different mice – ranging from small effects to large effects.

To see the relationship between significance of the association and gene-by-environment interactions, we compute and compare this P-value for each SNP from the 17 studies using the random effects meta-analysis to a measure of heterogeneity. Heterogeneity can be assessed by I^2 statistic, which describes the percentage of variation across studies that is due to heterogeneity rather than chance [26].

Figure 2 compares I^2 statistic with the meta-analysis P-value for each SNP. In this figure, we see that I^2 is uniformly distributed for the non-significant SNPs (blue dots), while it is right skewed for significant SNPs (red dots), indicating that more significant SNPs

have a greater potential for exhibiting heterogeneity in effect. Since heterogeneity in this case can be interpreted as representing gene-by-environment interactions, as heterogeneity is induced by differences in the environment, we see that the presence of a GxE interaction confers higher power to detect an association.

Power of meta-analysis for detecting gene-by-environment interactions

The power to identify both gene-by-environment and main effects in a meta-analysis of mouse studies depends on both the main effect size and the amount of heterogeneity. We performed simulations using the genotypes of the 17 mouse studies analyzed in this paper. We simulated a range of main effect (mean effect) sizes and a range of gene-by-environment effects. We are simulating the realistic scenario in which we do not know exactly the set of covariates which are responsible for the gene-by-environment effects. We simulated gene-by-environment effects by drawing the effect in each study from a distribution with a mean given by the main effect size and a variance controlling the magnitude of gene-by-environment interactions. If this variance is small, then all of the studies have close to the same effect size and there are few gene-by-environment effects. If the variance is high, then there are strong gene-by-environment effects. Figure 3 shows the results of our simulations. 1000 simulated phenotypes were generated for each mean and variance pair. Statistical power is estimated by computing the proportion of the datasets in which a simulated effect is detected. We observe that the power is high for a wide range of main effect sizes and gene-by-environment effect sizes which is explained by the large sample size of the study. We also observe that even for small main effects, if there are strong gene-by-environment effects, we can still identify the locus. This is because in this case a subset of the studies will have strong effect sizes due to gene-by-environment effects.

Our approach is not the only way to analyze a meta-analysis study. We compare the power to two other meta-analytic approaches. The first is the traditional meta-analysis strategy which uses a fixed effects model (FE) in which all of the effect sizes across studies are assumed to be the same. We utilize an extension of the fixed effects model which corrects for population structure [15]. A second alternate strategy is to simply apply the heterogeneity test (HE), which in our framework is only applied to loci first identified using random effects meta-analysis. The HE test follows the intuition that loci with high heterogeneity will harbor gene-by-environment interactions. For the purposes of the comparison we refer to Meta-GxE as the random effects (RE) model.

The level of gene-by-environment interactions can be simulated by changing both the environment-dependent and environment-independent effect simultaneously, when simulating the phenotype. Figure 4 (a)–(c) shows the power of the three approaches (RE, FE, HE) respectively when we vary the mean and variance of the effect size distribution we sampled from. In this simulation study, mean effect represents shared effect and variance of the effect size represents interaction effect. As expected, RE has high power in cases where the shared effect or the interaction effect is large. FE has high power when the shared effect is large and the HE test has high power when the interaction effect is large. Figure 4 (d) shows the heatmap which is colored with the color of highest powered approach. FE is most powerful at the top-left region, HE is most powerful at the bottom-right region, while RE is most powerful for a majority of the simulations. In the Text S1, we show through simulations that our methodology outperforms the alternative fixed effects and heterogeneity testing approaches when the effect is present in a subset of the studies, which is another possible
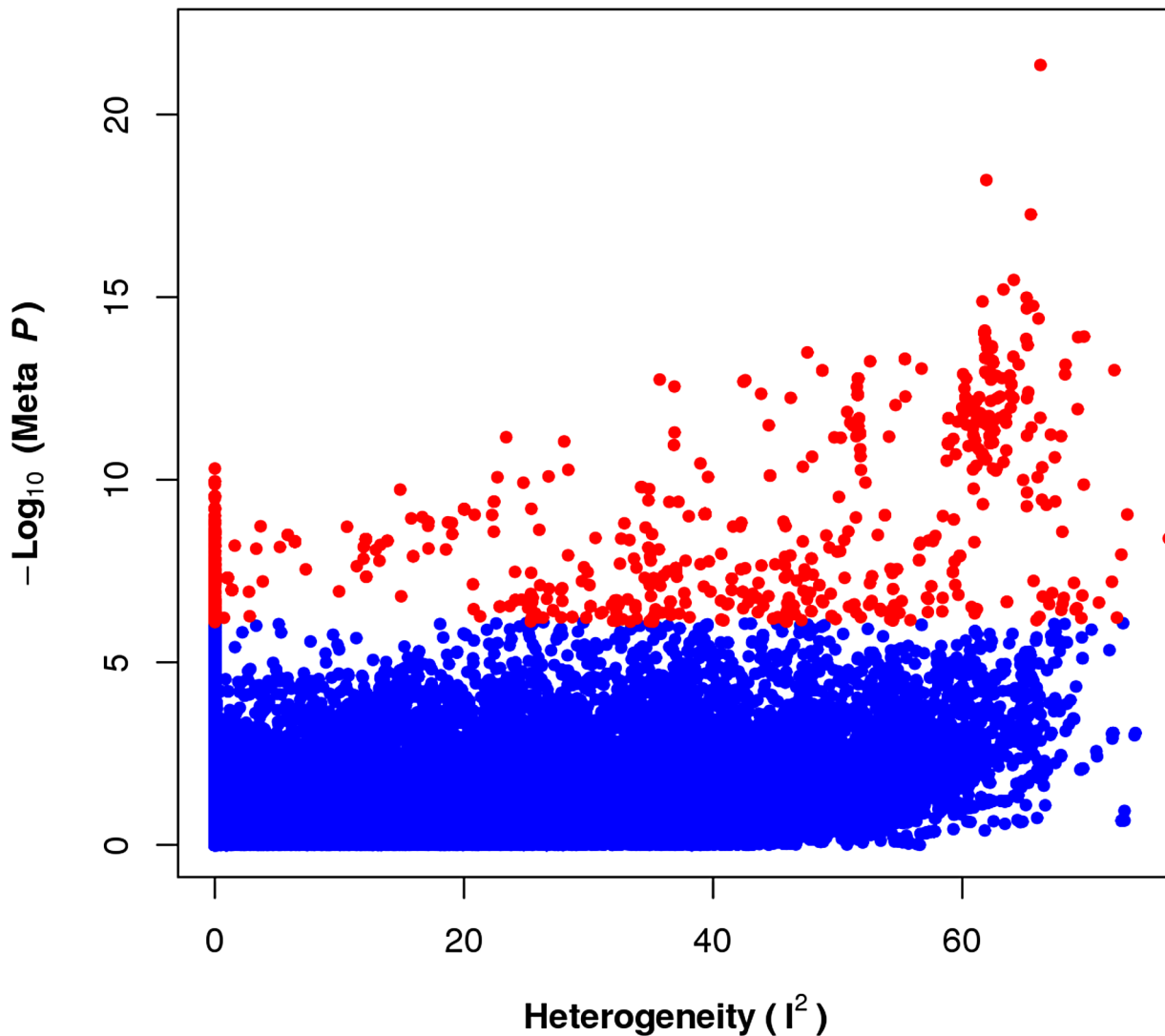
Figure 2. The prevalence of heterogeneity in effect size of significant loci. Each dot represents association between SNPs and HDL phenotype from applying random effects based meta-analysis approach. Dots with larger I^2 value represents the existence of more heterogeneity at the locus between studies. The distribution of the heterogeneity statistic for significant SNPs (red dots) in the meta analysis is skewed toward higher heterogeneity while the non-significant SNPs are much less skewed.
doi:10.1371/journal.pgen.1004022.g002

interaction model we can assume. We also show in the Text S1 that our approach is more powerful than the traditional uni- or multi-variate gene-by-environment association approach which assumes knowledge of the covariates involved in gene-by-environment interactions. For the traditional uni- or multi-variate approach, required knowledge includes kinds of variable (e.g. sex, age, gene knockouts) and encoding of the variables (e.g. binary values, continuous values). In the Text S1, we also show the our proposed approach controls the false positive rate.?

Application to 17 mouse HDL studies

We applied Meta-GxE to 17 mouse genetic studies conducted under various environmental conditions where each study measured HDL cholesterol. Table 1 summarizes each study. More details are provided in the Materials and Methods section and in Text S1. We analyzed all 17 studies together and we also analyzed the 9 male and 8 female studies separately. Some significant

associations are shared and some associations are specific to males and females.

The Manhattan plots in Figure 5 show the meta-GxE result when applied to the 17 studies, 9 male only studies and 8 female only studies. Table 2 summarizes 26 significant peaks ($P < 1.0 \times 10^{-6}$) showing the P-values obtained by applying meta-GxE to the male only studies (9 studies), the female only studies (8 studies) and the male+female studies (17 studies). For each significant locus, we computed m-values, interpreted as the posterior probability of having an effect on the phenotype and report the number of studies with an effect (E), the number of studies with ambiguous effect size (A) and the number of studies without an effect (N). We also report the number of individual studies where the locus was significant ($P < 1.0 \times 10^{-6}$). As seen in the table, many of the loci were not significant in any of the individual studies and would not have been discovered without combining the studies. We note that we use a more stringent genome wide threshold of $P < 1.0 \times 10^{-6}$ than was used in the
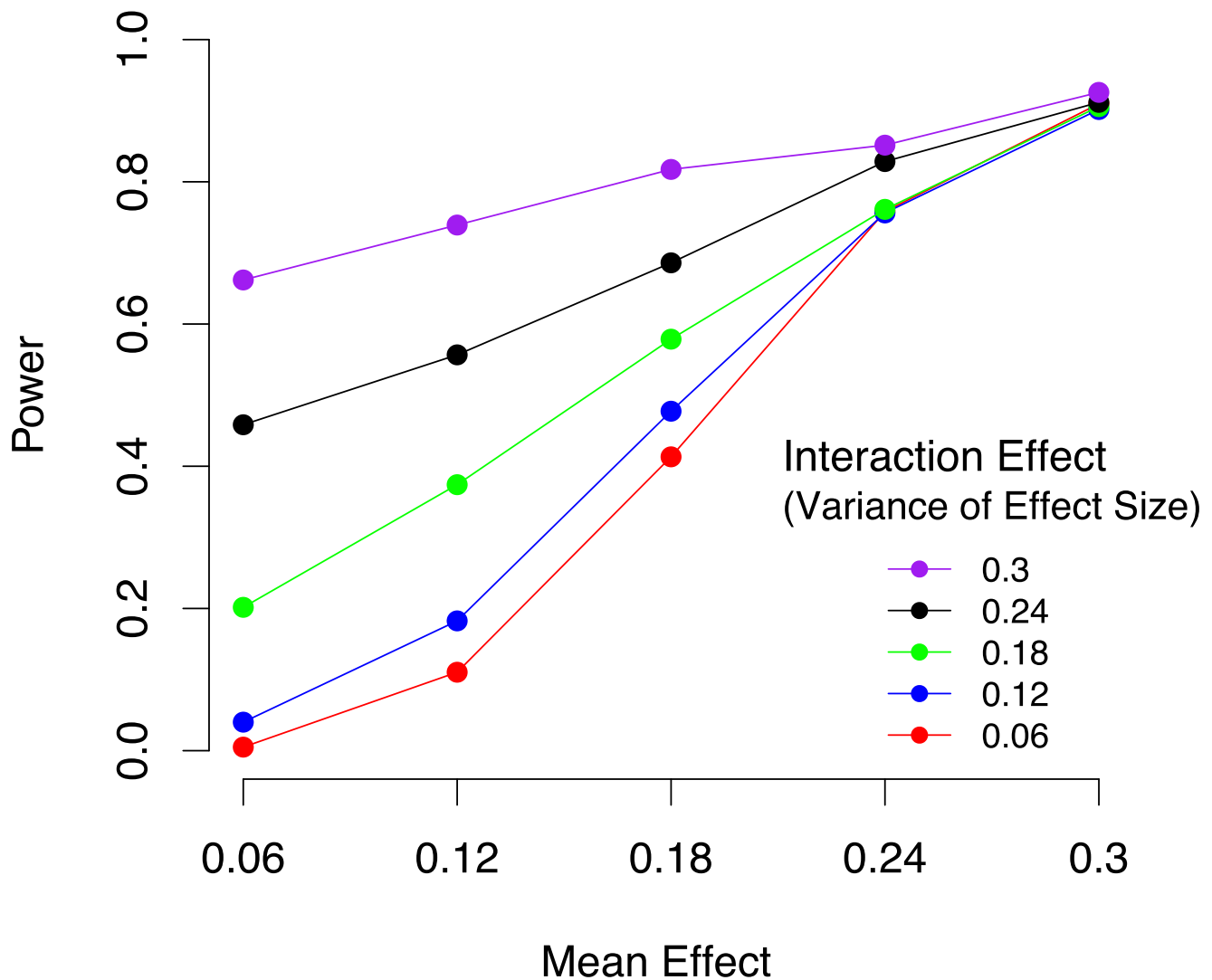
Figure 3. Power of mouse meta-analysis to identify gene-by-environment interactions in 4,965 animals from 17 studies under varying mean effect sizes and the per study variance of the effect size which corresponds to gene-by-environment effects.
doi:10.1371/journal.pgen.1004022.g003

original studies. The Genes in Region and Gene Refs columns contain the gene names near the locus previously known to affect HDL cholesterol level or closely related lipid level in the blood and associated literature citations.

Among the 26 loci that we identified by applying Meta-GxE, 24 loci are near the genes (mostly genes are located within 1MB of the peak) known to affect HDL or closely related lipid level in the blood, while 3 loci are novel.

For example, we identified 3 loci (Chr10:21399819, Chr19: 3319089, ChrX:151384614) female mice show a more significant effect on HDL phenotypes than male mice. We also identified 7 loci (Chr1:171199523, Chr8:46903188, Chr8:64150094, Chr8: 84073148, Chr10:90146088, Chr11:69906552, Chr15:21194226) where male mice show a more significant effect on HDL than female mice.

Interestingly, we observed that in 3 loci (Chr10:21399819, Chr19:3319089, ChrX:151384614), female mice are more highly affected, while in 7 loci (Chr1:171199523, Chr8:46903188, Chr8: 64150094, Chr8:84073148, Chr10:90146088, Chr11:69906552, Chr15:21194226) male mice are more highly affected. Among 26 loci, many show a significant heterogeneity in effect sizes be-

tween the 17 studies, which we interpret as gene-by-environment interactions.

One interesting example showing strong gene-by-environment interaction is a locus in Chr8:84073148. This locus is located near the gene *Prkaca*, which is known to affect the abnormal lipid levels in blood [27]. Figure 6 shows the forest plot and PM-plot for this locus. If we look at the forest plot of the locus in Figure 6, we can easily see that there are two groups: 12 studies with an effect (red dots) and 5 studies with an ambiguous prediction of the existence of an effect (green dots). Interestingly, the log odds ratios of effect size for the 12 studies with an effect is about the same (around 0.2). The common characteristic in 4 of the 5 studies (HMDPxB-chow(F), HMDPxB-ath(F), BXH-apoe(F), CXB-ldlr(F)) is that they are female mice with high LDL levels in the blood. In addition, in all 4 cases, these high LDL levels are caused by mutant genes. Mice in HMDPxB-chow and HMDPxB-ath studies have transgenes for both Apoe Leiden and for human Cholesterol Ester Transfer Protein (CETP), while mice in the BXH-apoe and CXB-ldlr studies carried knockouts of the genes for Apoe and LDL receptor, respectively. This is a strong evidence that there is an interaction between sex×muta-
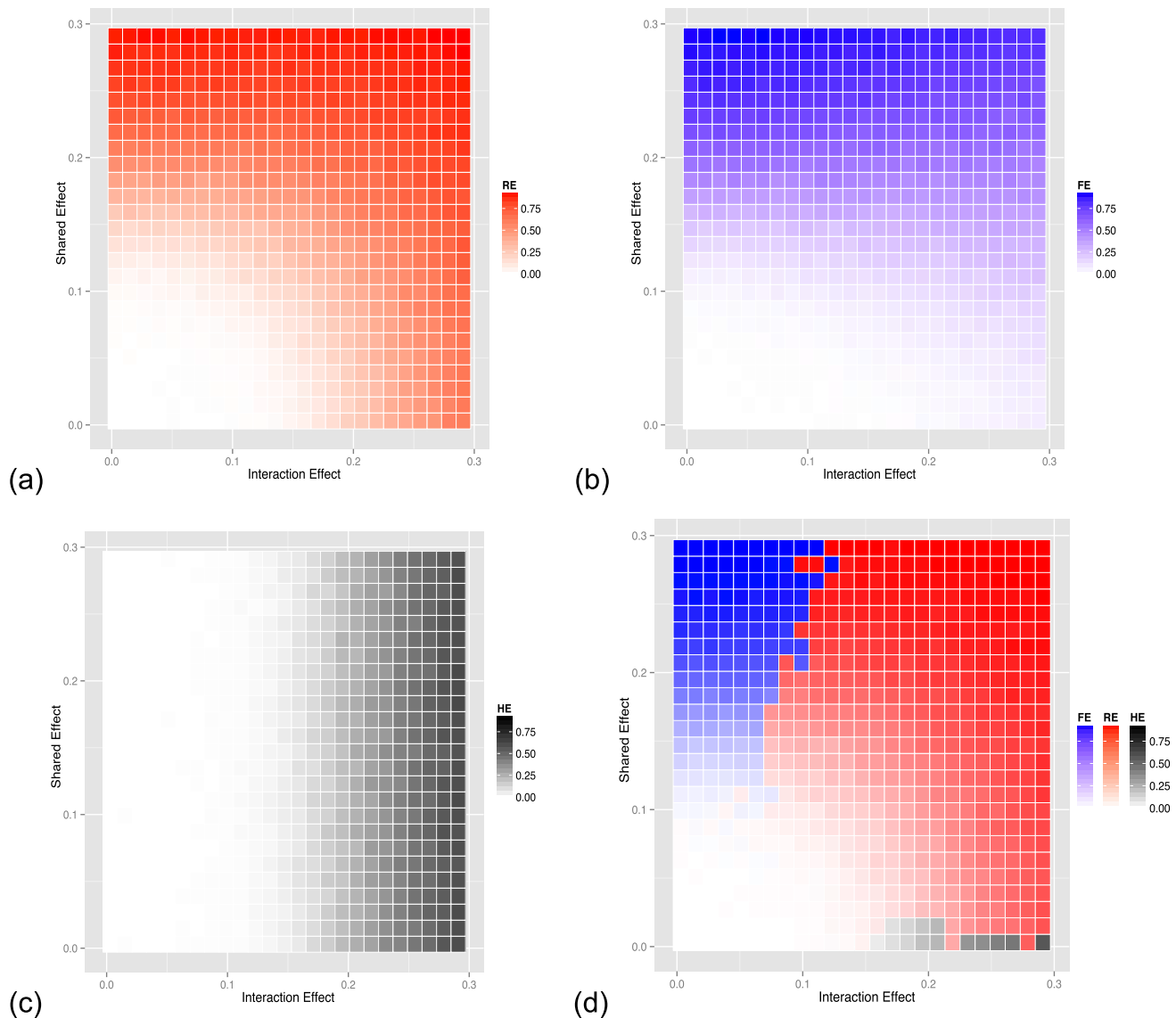
Figure 4. Power of (a) random-effect, (b) fixed-effect meta-analysis and (c) heterogeneity meta-analysis methods as a function of the effect size and the strength of the interaction effect (heterogeneity). (d) shows a comparison of the three methods with the color corresponding to the method with the highest power.
doi:10.1371/journal.pgen.1004022.g004

tion-driven LDL levels through this locus (Chr8:84073148) when affecting HDL levels in mice.

Figures S5, S6, S7, S8, S9, S10, S11, S12, S13, S14, S15, S16, S17, S18, S19, S20, S21, S22, S23, S24, S25, S26, S27, S28, S29, S30 show the forest plots and PM-plots for each locus, which show information such as effect sizes, the direction of the effect, which study has an effect and which study does not have an effect for each of 17 studies at the given locus.

Discussion

In this paper, we present a new meta-analysis approach for discovering gene-by-environment interactions that can be applied to a large number of heterogeneous studies each conducted in different environments and with animals from different genetic backgrounds. We show the practical utility of the proposed method by applying it to 17 mouse HDL studies containing 4,965

mice, and we successfully identify many known loci involved in HDL. Consistent with the results of meta-analysis in human studies, our combined study finds many loci which were not discovered in any of the individual studies.

A point of emphasis is that in our study design, in each of the combined studies, all of the individuals in the study are subject to only a single environment. This is distinct from other approaches for discovery of gene-by-environment interactions using meta-analysis such as those described in [28]. In these approaches, in each of the combined studies, the individuals in the study are subject to multiple environments and information on each individual's environment is collected. Gene-by-environment statistics are then computed in each study and then combined in the meta-analysis. In our study design, we compute main effect sizes for each SNP and then look for variants where the effect sizes are different suggesting the presence of a gene-by-environment interaction.
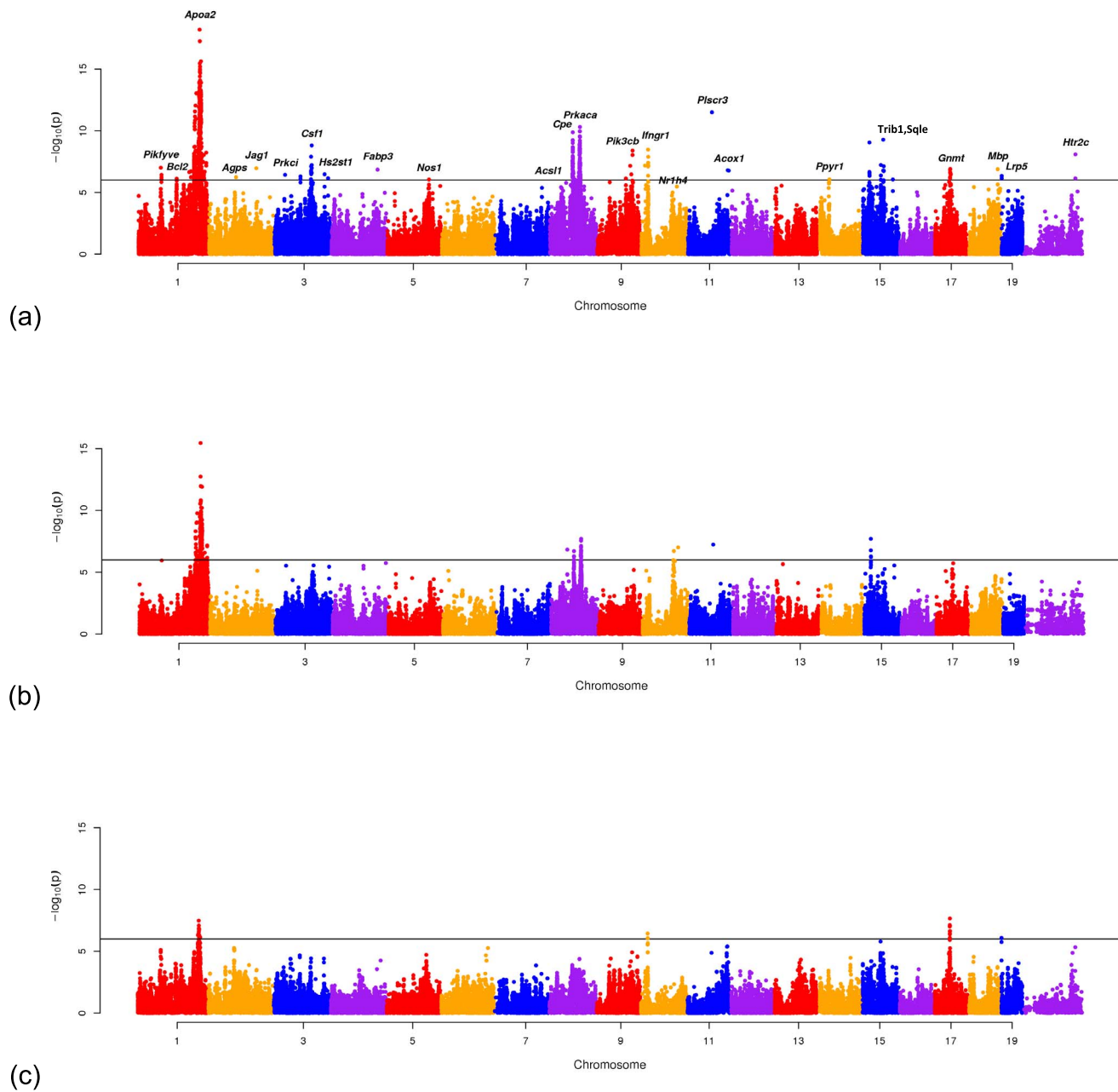
Figure 5. Manhattan plots showing the results of Meta-GxE applied to (a) 17 HDL studies, (b) 9 HDL studies consisting only of male animals and (c) 8 studies consisting only of female animals.
doi:10.1371/journal.pgen.1004022.g005

In our meta-analysis approach, we assume that we do not have any prior knowledge of the effect size in any particular study. However one might incorporate prior knowledge of the specific environmental effects. In some cases, one might know that some of the studies have similar effect sizes as compared to others. Or the prior knowledge might suggest that one specific study needs to be eliminated in the meta analysis. If we utilize such prior knowledge, we may be able to achieve even higher statistical power.

In this paper we have addressed how to perform meta-analysis when the studies have different genetic structures, building off the results of our previous study [15]. While in this paper we combine 7 HMDP studies with 10 genetic crosses, the approach in principle can be used to combine any variety of study types. Recently,

several strategies for mouse genome-wide association mapping have been proposed [29] [17]. These include HMDP [18], collaborative cross [30] and outbredstock [21] [17]. The approach presented here can be utilized to combine these different kinds of studies and is a roadmap for integrating the results of different strategies for mouse GWAS.

Although we have focused on explaining heterogeneity by gene-by-environment interaction, it is possible that the differences in effect sizes can be caused by gene-by-gene interactions on different genetic backgrounds, where the interacting variants differ in frequency in the different studies. While gene-by-gene interactions certainly contribute to locus heterogeneity, we predict that, in combining studies with similar genetic structures, locus heteroge-

Table 2. 26 significant loci identified by applying Meta-GxE analysis.

SNP Location	Meta GxE P (Male)	Meta GxE P (Female)	Meta GxE P (Male+Female)	# of Studies w/ Significant P	HE Meta P (Male+Female)	# Studies E/A/N	Genes in Region	Gene Refs
Chr1:64752822 (rs31078051)	1.12×10^{-6}	2.37×10^{-3}	$\mathbf{9.82\times10^{-8}}$	0	3.82×10^{-3}	2/14/1	**Pikfyve**	[50]
Chr1:107271282 (rs32203839)	6.69×10^{-4}	2.66×10^{-4}	$\mathbf{7.67\times10^{-7}}$	0	7.55×10^{-2}	6/0/11	**Bcl2**	[51]
Chr1:171199523 (rs32075748)	$\mathbf{3.45\times10^{-16}}$	$\mathbf{1.39\times10^{-7}}$	$\mathbf{4.41\times10^{-22}}$	1	5.80×10^{-5}	8/6/3	**Apoa2**	[52]
Chr2:77837584 (rs6273567)	1.55×10^{-4}	7.17×10^{-4}	$\mathbf{5.74\times10^{-7}}$	0	2.25×10^{-2}	3/14/0	**Agps**	[53]
Chr2:134421733 (rs27238693)	7.69×10^{-6}	1.66×10^{-3}	$\mathbf{1.09\times10^{-7}}$	0	1.63×10^{-2}	5/12/0	**Jag1**	[54]
Chr3:32944259 (rs29869794)	2.97×10^{-6}	2.49×10^{-2}	$\mathbf{3.68\times10^{-7}}$	0	2.36×10^{-5}	3/8/6	**Prkci**	[55]
Chr3:76066632 (rs31487078)	5.89×10^{-3}	2.29×10^{-5}	$\mathbf{5.03\times10^{-7}}$	0	6.56×10^{-2}	4/13/0	**Novel**	-
Chr3:107430396 (rs30013147)	9.59×10^{-6}	3.84×10^{-5}	$\mathbf{1.56\times10^{-9}}$	0	7.74×10^{-2}	7/10/0	**Csf1**	[56]
Chr3:143466942 (rs30206761)	1.82×10^{-3}	3.97×10^{-5}	$\mathbf{3.34\times10^{-7}}$	0	8.60×10^{-2}	7/10/0	**Hs2st1**	[57]
Chr4:131925523 (rs32595861)	1.72×10^{-4}	2.84×10^{-4}	$\mathbf{1.42\times10^{-7}}$	0	8.65×10^{-4}	6/8/3	**Fabp3**	[58]
Chr5:119034507 (rs33131194)	1.23×10^{-4}	2.59×10^{-3}	$\mathbf{9.00\times10^{-7}}$	0	3.94×10^{-1}	9/8/0	**Nos1**	[59,60]
Chr8:46903188 (rs33272858)	$\mathbf{1.47\times10^{-7}}$	6.52×10^{-1}	1.66×10^{-6}	1	1.62×10^{-4}	2/11/4	**Acsl1**	[61]
Chr8:64150094 (rs31750594)	$\mathbf{1.96\times10^{-7}}$	1.89×10^{-4}	$\mathbf{1.33\times10^{-10}}$	0	8.34×10^{-1}	11/6/0	**Cpe**	[62]
Chr8:84073148 (rs33435859)	$\mathbf{1.95\times10^{-8}}$	4.53×10^{-4}	$\mathbf{4.94\times10^{-11}}$	0	8.33×10^{-1}	12/5/0	**Prkaca**	[27]
Chr9:101972687 (rs6333310)	1.22×10^{-4}	1.22×10^{-5}	$\mathbf{4.05\times10^{-9}}$	0	$\mathbf{1.98\times10^{-8}}$	2/1/14	**Pik3cb**	[63]
Chr10:21399819 (rs29363941)	9.07×10^{-4}	$\mathbf{3.64\times10^{-7}}$	$\mathbf{3.36\times10^{-9}}$	0	1.18×10^{-2}	3/12/2	**Ifngr1**	[64]
Chr10:90146088 (rs29370592)	$\mathbf{1.93\times10^{-7}}$	0.756	1.02×10^{-5}	1	8.94×10^{-4}	2/14/1	**Nr1h4**	[65]
Chr11:69906552 (rs29477071)	$\mathbf{5.77\times10^{-8}}$	1.35×10^{-5}	$\mathbf{3.17\times10^{-12}}$	0	$\mathbf{2.37\times10^{-9}}$	6/9/2	**Plscr3**	[66]
Chr11:114083173 (rs29416888)	1.10×10^{-4}	7.83×10^{-5}	$\mathbf{1.71\times10^{-7}}$	0	5.28×10^{-5}	3/13/1	**Acox1**	[67]
Chr14:33632464 (rs31061259)	1.96×10^{-4}	1.65×10^{-3}	$\mathbf{8.90\times10^{-7}}$	0	2.02×10^{-5}	3/10/4	**Ppyr1**	[68]
Chr15:21194226 (rs31670969)	$\mathbf{1.96\times10^{-8}}$	1.29×10^{-2}	$\mathbf{8.97\times10^{-10}}$	1	$\mathbf{5.65\times10^{-7}}$	3/2/12	**Novel**	-
Chr15:59860191 (rs3718217)	5.64×10^{-6}	1.45×10^{-5}	$\mathbf{5.31\times10^{-10}}$	0	9.92×10^{-5}	5/10/2	**Trib1, Sqle**	[69,70]
Chr17:46530712 (rs33259313)	1.09×10^{-5}	4.90×10^{-3}	$\mathbf{3.26\times10^{-7}}$	0	3.53×10^{-5}	5/10/2	**Gnmt**	[71]
Chr18:82240606 (rs13483466)	2.05×10^{-4}	2.23×10^{-4}	$\mathbf{1.32\times10^{-7}}$	0	9.05×10^{-1}	5/12/0	**Mbp**	[72]
Chr19:3319089 (rs31004232)	5.58×10^{-2}	$\mathbf{8.53\times10^{-7}}$	$\mathbf{4.56\times10^{-7}}$	0	1.08×10^{-1}	3/14/0	**Lrp5**	[73]
ChrX:151384614 (rs31202008)	2.59×10^{-4}	4.72×10^{-6}	$\mathbf{8.09\times10^{-9}}$	0	1.12×10^{-1}	5/5/0	**Htr2c**	[74]

Twentysix significant loci identified by applying Meta-GxE analysis of both random effects meta-analysis and heterogeneity testing to 17 mouse HDL studies under different environments containing 4,965 total animals. # studies E denotes the number of studies with an effect on HDL phenotype. # studies N denotes the number of studies with no effect on HDL phenotype. # studies A denotes the number of studies with an ambiguous effect size. Genes in region denotes candidate genes for each locus based on close proximity to the peak SNP and previously suggested role in lipid or apolipoprotein metabolism: **Pikfyve** (phosphoinositide kinase), **Bcl2** (B cell leukemia/lymphoma 2), **Apoa2** (apolipoprotein A-II), **Agps** (alkylglycerone phosphate synthase), **Jag1** (jagged 1), **Prkci** (protein kinase C), **Prkci** (colony stimulating factor 1 (macrophage)), **Hs2st1** (heparan sulfate 2-O-sulfotransferase 1), **Fabp3** (fatty acid binding protein 3), **Nos1** (nitric oxide synthase 1), **Acsl1** (acyl-CoA synthetase long-chain family member 1), **Cpe** (carboxypeptidase E), **Prkaca** (protein kinase, cAMP dependent, catalytic, alpha), **Acox1** (peroxisomal acyl-coenzyme A oxidase 1), **Ppyr1** (pancreatic polypeptide receptor 1), **Trib1** (tribbles homolog 1), **Sqle** (squalene epoxidase), **Gnmt** (glycine N-methyltransferase), **Mbp** (myelin basic protein), **Lrp5** (low density lipoprotein receptor-related protein 5), **Htr2c** (5-hydroxytryptamine (serotonin) receptor 2C).
doi:10.1371/journal.pgen.1004022.t002

neity more likely arises from gene-by-environment interactions. In any case, determining whether or not these heterogeneous loci are environment-driven or interaction-driven is an important and interesting direction for future study.

Materials and Methods

Standard study design for testing gene-by-environment interactions

In the model organism studies for which we can control the environment, the standard study design for testing gene-by-environment interactions is to combine multiple cohorts whose environments are known. The environmental value that we vary is typically a quantitative measure that we can model with a single random variable. Thus, the standard univariate linear model can be applied

$$y = \mu + \alpha D + \beta X + \gamma X \cdot D + e$$

where y is $n \times 1$ vector of phenotype measurements from n

individuals, μ is the phenotype mean, α is the main environmental effect mean, D is $n \times 1$ environmental status vector, β is the genetic effect, X is $n \times 1$ genotype vector, γ is GxE interactions effect, $\cdot$ denotes the dot-product between two vectors, and e is the residual error, which follows normal distribution. In this model, vector D is a vector of indicators which describes the environmental status of each individual. study. For example, Suppose the environmental condition of one study is wildtype and that of another is gene knockout. In this case, the environmental condition of wildtype is described as 0 and that of knockout is described as 1. In order to test if there are interactions, we test the null hypothesis $\gamma = 0$ versus the alternative hypothesis $\gamma \neq 0$. Another possible testing strategy is to test the interactions effect together with the genetic effect, that is, the null hypothesis $\beta = 0$ and $\gamma = 0$ versus the alternative hypothesis $\beta \neq 0$ or $\gamma \neq 0$. This strategy is powerful in detecting loci exhibiting both the genetic effects and the interactions effects.

Multivariate interactions model

For more complicated scenarios where the different environments can not be modeled with a single variable, a straightforward
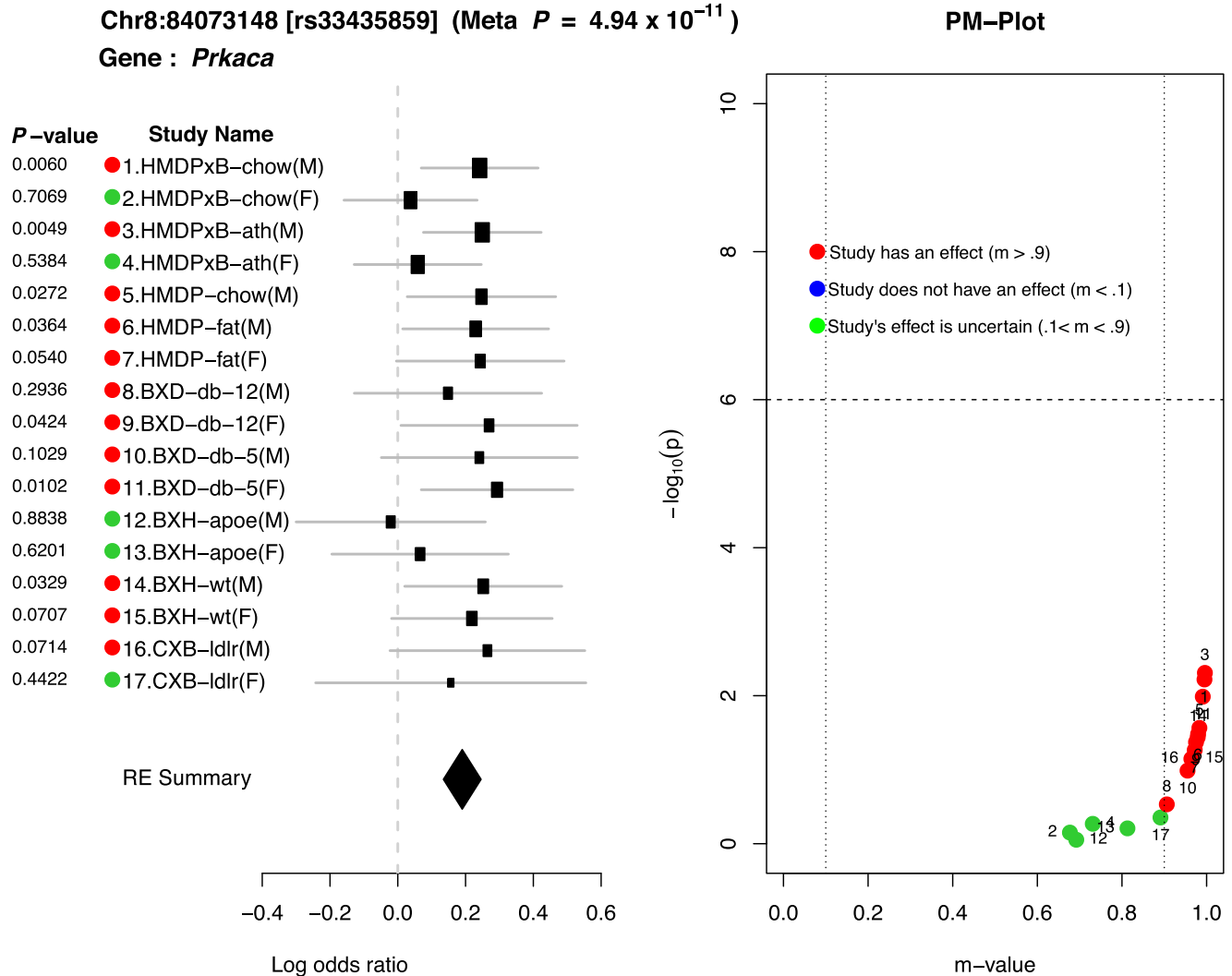

Figure 6. Peak SNP in chromosome 8 shows interesting gene-by-environment interactions between sex×mutation-driven LDL levels.
doi:10.1371/journal.pgen.1004022.g006

extension of the standard univariate interactions model is the multivariate model. Suppose that there are k different possible environments and the information on the environments of each individual are captured by a matrix D which has k columns where each column corresponds to one environment. Then, the standard multivariate interactions model will be

$$y = \sum_{i=1}^{k} \alpha_i D_i + \beta X + \sum_{i=1}^{k} \gamma_i X \cdot D_i + e \qquad (3)$$

D_i is the ith column of the D matrix, α_i is the environment specific mean, y denotes the phenotype measurements, X denotes the genotypes, β denotes the fixed genetic effect, γ_i denotes GxE interactions effect of ith environmental variable and, and e denotes the residual error. Then the testing will be between the null hypothesis $\gamma_1 = 0$ and ... and $\gamma_k = 0$ versus the alternative hypothesis $\gamma_1 \neq 0$ or ... or $\gamma_k \neq 0$. The test statistic will be

$$S_{Mult} = \sum_{i=1}^{k} Z_i^2$$

where Z_i is the z-score corresponding to γ_i. S_{Mult} follow $\chi^2_{(k)}$ under the null. Similarly to the univariate model, if we want to test the interactions effect together with genetic effect, we add the z-score corresponding to β into the statistic, in which case the statistic will follow $\chi^2_{(k+1)}$.

Standard meta-analysis approach

Before we describe the relationship between gene-by-environment interactions and meta-analysis, we first describe the standard fixed effects and random effects meta-analysis in details.

Fixed effects model meta analysis. In standard meta-analysis, we have N studies. In each of the N studies, we estimate the effect size of interest. Suppose that we estimate the genetic effect in study i,

$$y_i = \alpha_i + \delta_i X_i + e_i \qquad (4)$$

We can obtain the estimates of δ_i and its variance V_i. In the fixed effects model meta-analysis, we assume that the underlying effect sizes are the same as δ ($\delta = \delta_1 = \ldots = \delta_N$). The best estimate of δ is the inverse-variance weighted effect size,

$$\bar{\delta} = \frac{\sum W_i \delta_i}{\sum W_i}, \qquad (5)$$

where $W_i = 1/V_i$ is the so-called inverse variance. Then we test the null hypothesis $\delta = 0$ versus the alternative hypothesis $\delta \neq 0$.

Testing heterogeneity. The phenomenon that the underlying effect sizes differ between studies is called *heterogeneity*. The presence of heterogeneity is tested using the Cochran's Q test [24,25]. Cochran's Q test is a non-parametric test for testing if N studies have the same effect or not. Particularly it tests the null hypothesis $\delta_1 = \ldots = \delta_N$ versus the alternative hypothesis $NOT(\delta_1 = \ldots = \delta_N)$. Cochran's Q statistic can be calculated as the weighted sum of squared differences between individual study effects and the pooled effect across studies.

$$Q = \sum_{i=1}^{N} W_i (\delta_i - \bar{\delta})^2 \qquad (6)$$

Cochran's Q statistic has a chi-square statistic with $N-1$ degrees of freedom.

Random effects model meta analysis. Under the random effects model meta-analysis, we explicitly model heterogeneity by assuming a hierarchical model. We assume that the effect size of each study δ_i is a random variable sampled from a distribution with a mean δ and variance τ^2,

$$\delta_i \sim N(\delta, \tau^2)$$

Traditional formulations of a random effects meta-analysis method are known to be overly conservative [24,31,32]. However, we recently developed a random effects model that addresses this issue [33]. The method assumes that there is no heterogeneity under the null, a modification that is natural in the context of association studies because the effect size should be fixed to be zero under the null hypothesis. This random effects model tests the null hypothesis $\delta = 0$ and $\tau^2 = 0$ versus the alternative hypothesis $\delta \neq 0$ or $\tau^2 \neq 0$.

Similarly to the traditional random effects model [24], we use the likelihood ratio framework considering each statistic as a single observation. Since we assume no heterogeneity under the null, $\mu = 0$ and $\tau^2 = 0$ under the null hypothesis. The likelihoods are then

$$L_0 = \prod_i \frac{1}{\sqrt{2\pi V_i}} \exp\left(-\frac{\delta_i^2}{2V_i} \right)$$

$$L_1 = \prod_i \frac{1}{\sqrt{2\pi(V_i + \tau^2)}} \exp\left(-\frac{(\delta_i - \mu)^2}{2(V_i + \tau^2)} \right).$$

The maximum likelihood estimates $\hat{\mu}$ and $\hat{\tau}^2$ can be found by an iterative procedure suggested by Hardy and Thompson [34]. Then the likelihood ratio test statistic can be built

$$S_{meta} = -2\log(\lambda) = \sum \log\left(\frac{V_i}{V_i + \hat{\tau}^2} \right) + \sum \frac{\delta_i^2}{V_i} - \sum \frac{(\delta_i - \hat{\mu})^2}{V_i + \hat{\tau}^2}, \quad (7)$$

whose *P*-value is calculated using tabulated values [33].

Relation between gene-by-environment interactions and meta-analysis

Here we explain more about the relationship between gene-by-environment interactions and meta-analysis based on the explanation in Results section. If we do not consider the interactions, it has been already known that the fixed effects model meta-analysis is approximately equivalent to the linear model of combined cohorts [35]. That is, the fixed effects model equation (5) gives approximately equivalent results to the combined linear model

$$y = \sum_{i=1}^{k} \alpha_i A_i + \beta X + e \qquad (8)$$

where X is the combined genotype vector from all cohorts, A is a

matrix that includes indicator columns which identify which individual is in each cohort, A_i is the ith column of matrix A, and α_i is the cohort specific mean. The two methods are approximately equivalent because they both test the fixed mean effect (β in equation (8) and δ in equation (5)). The subtle difference between the two models is that in equation (8), we assume the error e follows a single normal distribution (e.g. $N(0,\sigma^2)$), whereas in equation (5), the variance of the distributions may differ between studies (e.g. $e_j \sim N(0,\sigma_j^2)$ for each j). In other words, under the constant error variance assumption ($\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_N^2$), the two models become equivalent and β in equation (8) equals δ in equation (5),

$$\beta = \delta \ .$$

Similarly, by considering interactions, we extend this argument to show the relationship between gene-by-environment interactions and meta-analysis. We consider the relationship between equation (3) and equation (4). For simplicity of the notation, we consider the case where the matrix D is defined in such a way that each individual is only in one environment such that the D matrix is equivalent to the matrix A described above. If we assume the constant error variance assumption, we establish the following relationship,

$$\delta_i = \beta + \gamma_i$$

where the left hand side is the coefficient of the genotype X_i of study i from the meta-analysis equation (4) and the right hand side is the same coefficient of X_i (the study i's part within the combined genotype matrix X) from the equation (3).

Suppose that there are no interactions (null hypothesis of interaction testing). Then, $\gamma_i = 0$ for each study i. Thus, the effect size of meta-analysis δ_i is equivalent to β, the genetic effects that are invariant across studies. Therefore, $\tau^2 = 0$ (null hypothesis of heterogeneity testing). On the other hand, suppose that $\tau^2 = 0$ (null hypothesis of heterogeneity testing). Naturally, $\gamma_i = 0$ for all studies (null hypothesis of interaction testing). This shows that the null hypothesis of the interactions test in the model (3) and the null hypothesis of the heterogeneity test in meta-analysis are equivalent. As a result, we can utilize meta-analytic heterogeneity testing to detect interactions.

Using reasoning, it is straightforward to show that we can utilize the random effects model meta-analysis method to detect the mean effect and the interaction effect at the same time, which can be powerful for identifying loci bearing both kinds of effects.

Controlling for population structure within studies

Model organism such as the mouse are well-known to exhibit population structure or cryptic relatedness [36,37], where genetic similarities between individuals both inhibit the ability to find true associations and cause the appearance of a large number of false or spurious associations. Mixed effects models are often used in order to correct this problem [38–42]. Methods employing a mixed effects correction account for the genetic similarity between individuals with the introduction of a random variable into the traditional linear model.

$$y_i = \mu + \delta_i X + u_i + \epsilon \tag{9}$$

In the model in equation (9), the random variable u_i represents the vector of genetic contributions to the phenotype for individuals in population i. This random variable is assumed to follow a normal distribution with $u_i \sim N(0,\sigma_g^2 K_i)$, where K_i is the $n_i \times n_i$ kinship coefficient matrix for population i. With this assumption, the total variance of y_i is given by $\Sigma_i = \sigma_g^2 K_i + \sigma_e^2 I$. A z-score statistic is derived for the test $\delta_i = 0$ by noting the distribution of the estimate of $\hat{\delta}_i$. In order to avoid complicated notation, we introduce a more basic matrix form of the model in equation (9), shown in equation (10).

$$y_i = S_i \Gamma + u_i + \epsilon \tag{10}$$

In equation (10), S_i is a $n_i \times 2$ matrix with the first column being a vector of 1 s representing the global mean and the second vector is the vector and Γ is a 2×1 coefficient vector containing the mean α_i and genotype effect (δ_i). We note that this form also easily extends to models with multiple covariates. The maximum likelihood estimate for Γ in population i is given by $\hat{\Gamma}_i = (S_i' \Sigma_i^{-1} S_i)^{-1} S_i' \Sigma_i^{-1} y_i$ which follows a normal distribution with a mean equal to the true Γ and variance $(S_i' \Sigma_i^{-1} S_i)^{-1}$. The estimates of the effect size δ_i and standard error of the δ_i ($SE(\delta_i)$) are then given in equation (11) and equation (12), where $R = [0 \quad 1]$ is a vector used to select the appropriate entry in the vector $\hat{\Gamma}_i$.

$$\delta_i = R(S_i' \Sigma_i^{-1} S_i)^{-1} S_i' \Sigma_i^{-1} y_i \tag{11}$$

$$SE(\delta_i) = [R(S_i' \Sigma_i^{-1} S_i)^{-1} R']^{1/2} \tag{12}$$

Meta-analysis of studies with population structure

When we test gene-by-environment interactions with meta analysis approaches, one important step is correcting for population structure. This can be achieved by correcting for population structure within each study first as described above. For example, consider the random effects model meta-analysis method that we primarily focus on. We employ population structure control, using (11) and (12). Then the likelihood ratio test statistic will be

$$S_{\text{Pop}} = -2\log(\lambda) = \sum \log\left(\frac{V_i}{V_i + \hat{\tau}^2}\right) + \sum \frac{\delta_i^2}{V_i} - \sum \frac{(\delta_i - \hat{\mu})^2}{V_i + \hat{\tau}^2}, \tag{13}$$

where $\delta_i = R(S_i' \Sigma_i^{-1} S_i)^{-1} S_i' \Sigma_i^{-1} y_i$ and $V_i = [R(S_i' \Sigma_i^{-1} S_i)^{-1} R']$.

Identifying studies with an effect

After identifying loci exhibiting interaction effects, we employ the meta-analysis interpretation framework that we recently developed. The *m-value* [16] is the posterior probability that the effect exists in each study. Suppose we have n number of studies we want to combine. Let $E = [\delta_1, \delta_2, \ldots, \delta_n]$ be the vector of estimated effect sizes and $V = [V_1, V_2 \ldots, V_n]$ be the vector of estimated variance of n effect sizes. We assume that the effect size δ_i follows the normal distribution.

$$P(\delta_i | \text{no effect}) = N(\delta_i; 0, V_i) \qquad (14)$$

$$P(\delta_i | effect) = N(\delta_i; \mu, V_i) \qquad (15)$$

We assume that the prior for the effect size is

$$\mu \sim N(0, \sigma^2) \qquad (16)$$

A possible choice for σ in GWASs is 0.2 for small effect and 0.4 for large effect [43]. We also denote C_i be a random variable whose value is 1 if a study i has an effect and 0 otherwise. We also denote C as a vector of C_i for n studies. Since C has n binary values, C can be 2^n possible configurations. Let $U = [c_1, \ldots, c_{2^n}]$ be a vector containing all the possible these configurations. We define $m\text{-}value$ m_i as the probability $P(C_i = 1 | E)$, which is the probability of study i having an effect given the estimated effect sizes. We can compute this probability using the Bayes' theorem in the following way.

$$m_i = P(C_i = 1 | E) = \frac{\sum_{c \in U_i} P(E | C = c) P(C = c)}{\sum_{c \in U} P(E | C = c) P(C = c)} \qquad (17)$$

where U_i is a subset of U whose elements' i^{th} value is 1. Now we need to compute $P(E | C = c)$ and $P(C = c)$. $P(C = c)$ can be computed as

$$P(C = c) = \frac{B(|c| + \alpha, n - |c| + \beta)}{B(\alpha, \beta)} \qquad (18)$$

where $|c|$ denotes the number of 1's in c and B denotes the beta function and we set α and β as 1 [16]. The probability E given configuration c, $P(E | C = c)$, can be computed as

$$P(E | C = c) = \int_{-\infty}^{\infty} \prod_{i \in c_0} N(\delta_i; 0, V_i) \prod_{i \in c_1} N(\delta_i; \mu, V_i) p(\mu) d\mu \qquad (19)$$

$$= \bar{C} N(\bar{\delta}; 0, \bar{V} + \sigma^2) \prod_{i \in c_0} N(\delta_i; 0, V_i) \qquad (20)$$

$$\bar{\delta} = \frac{\sum_i W_i \delta_i}{\sum_i W_i} \qquad \text{and} \qquad \bar{V} = \frac{1}{\sum_i W_i} \qquad (21)$$

where where c_0 is the indices of 0 in c and c_1 is the indices of 1 in c, $N(\delta; a, b)$ denotes the probability density function of the normal distribution with mean a and variance b. $W_i = V_i^{-1}$ is the inverse variance or precision and $\bar{C}$ is a scaling factor.

$$\bar{C} = \frac{1}{(\sqrt{2\pi})^{N-1}} \sqrt{\frac{\prod_i W_i}{\sum_i W_i}} exp \\ \left\{ -\frac{1}{2} \left(\sum_i W_i \delta_i^2 - \frac{(\sum_i W_i \delta_i)^2}{\sum_i W_i} \right) \right\} \qquad (22)$$

All summations appeared for computing $\bar{\delta}$, $\bar{V}$ and $\bar{C}$ are with respect to $j \in t_1$.

The $m\text{-}value$s have the following interpretations: small $m\text{-}value$s(0.1) represent a study that is predicted to not have an effect, large $m\text{-}$ $value$s(0.9) represent a study that is predicted to have an effect, otherwise it is ambiguous to make a prediction. It was previously reported that $m\text{-}value$s can accurately distinguish studies having an effect from the studies not having an effect [16]. For interpreting and understanding the result of the meta-analysis, it is informative to look at the P-value and m-value at the same time. We propose to apply the PM-plot framework [16], which plots the P-values and m-values of each study together in two dimensions. Figure 1 (b) shows one example of a PM-plot. In this example, studies with an $m\text{-}value$ less than 0.1 are interpreted as studies not having an effect while studies with an $m\text{-}value$ greater than 0.9 are interpreted as studies having an effect. For studies with an $m\text{-}value$ between 0.1 and 0.9, we cannot make a decision. One reason that studies are ambiguous ($0.9 \leq m - value \leq 0.1$) is that they are underpowered due to small sample size. If the sample size increases, the study can be drawn to either the left or the right side.

Supporting Information

Figure S1 Power comparison between random-effect, fixed-effect meta-analysis and heterogeneity testing.
(EPS)

Figure S2 The association result of the fixed effects meta analysis.
(EPS)

Figure S3 Power comparison between random-effect meta-analysis and traditional wald test based approach.
(EPS)

Figure S4 Power comparison between heterogeneity testing approach and traditional wald test based approach.
(EPS)

Figure S5 Forest plot and PM-plot for Chr1:64752822 locus.
(EPS)

Figure S6 Forest plot and PM-plot for Chr1:107271282 locus.
(EPS)

Figure S7 Forest plot and PM-plot for Chr1:171199523 locus.
(EPS)

Figure S8 Forest plot and PM-plot for Chr2:77837584 locus.
(EPS)

Figure S9 Forest plot and PM-plot for Chr2:134421733 locus.
(EPS)

Figure S10 Forest plot and PM-plot for Chr3:32944259 locus.
(EPS)

Figure S11 Forest plot and PM-plot for Chr3:76066632 locus.
(EPS)

Figure S12 Forest plot and PM-plot for Chr3:107430396 locus.
(EPS)

Figure S13 Forest plot and PM-plot for Chr3:143466942 locus.
(EPS)

Figure S14 Forest plot and PM-plot for Chr4:131925523 locus.
(EPS)

Figure S15 Forest plot and PM-plot for Chr5:119034507 locus.
(EPS)

Figure S16 Forest plot and PM-plot for Chr8:46903188 locus.
(EPS)

Figure S17 Forest plot and PM-plot for Chr8:64150094 locus.
(EPS)

Figure S18 Forest plot and PM-plot for Chr8:84073148 locus.
(EPS)

Figure S19 Forest plot and PM-plot for Chr9:101972687 locus. (EPS)

Figure S20 Forest plot and PM-plot for Chr10:21399819 locus. (EPS)

Figure S21 Forest plot and PM-plot for Chr10:90146088 locus. (EPS)

Figure S22 Forest plot and PM-plot for Chr11:69906552 locus. (EPS)

Figure S23 Forest plot and PM-plot for Chr11:114083173 locus. (EPS)

Figure S24 Forest plot and PM-plot for Chr14:33632464 locus. (EPS)

Figure S25 Forest plot and PM-plot for Chr15:21194226 locus. (EPS)

Figure S26 Forest plot and PM-plot for Chr15:59860191 locus. (EPS)

Figure S27 Forest plot and PM-plot for Chr17:46530712 locus. (EPS)

Figure S28 Forest plot and PM-plot for Chr18:82240606 locus. (EPS)

Figure S29 Forest plot and PM-plot for Chr19:3319089 locus. (EPS)

Figure S30 Forest plot and PM-plot for ChrX:151384614 locus. (EPS)

Table S1 False positive rate of RE versus traditional Wald Test based approaches at thresholds of increasing significance. (EPS)

Table S2 False positive rate of HE versus traditional Wald Test based approaches at thresholds of increasing significance. (EPS)

Text S1 Details on power and type I error simulations for gene by environment and 17 HDL mouse studies. (PDF)

Author Contributions

Conceived and designed the experiments: EYK BH NF JWJ DS RCD AJL EE. Performed the experiments: EYK BH NF JWJ DS RCD. Analyzed the data: EYK BH NF JWJ DS RCD AJL EE. Contributed reagents/materials/analysis tools: EYK BH NF JWJ DS RCD AJL EE. Wrote the paper: EYK BH NF DS RCD AJL EE.

References

1. Gerke J, Lorenz K, Ramnarine S, Cohen B (2010) Gene-environment interactions at nu- cleotide resolution. PLoS Genet 6: e1001144.
2. Murea M, Ma L, Freedman BI (2012) Genetic and environmental factors associated with type 2 diabetes and diabetic vascular complications. Rev Diabet Stud 9: 6–22.
3. Smith EN, Kruglyak L (2008) Gene-environment interaction in yeast gene expression. PLoS Biol 6: e83.
4. Talmud PJ (2007) Gene-environment interaction and its impact on coronary heart disease risk. Nutr Metab Cardiovasc Dis 17: 148–52.
5. Forsyth JK, Ellman LM, Tanskanen A, Mustonen U, Huttunen MO, et al. (2012) Genetic risk for schizophrenia, obstetric complications, and adolescent school outcome: Evidence for gene-environment interaction. Schizophr Bull 39: 1067–76.
6. Orozco LD, Bennett BJ, Farber CR, Ghazalpour A, Pan C, et al. (2012) Unraveling inammatory responses using systems genetics and gene-environment interactions in macrophages. Cell 151: 658–70.
7. Dai X, Wu C, He Y, Gui L, Zhou L, et al. (2013) A genome-wide association study for serum bilirubin levels and gene-environment interaction in a chinese population. Genet Epidemiol 37: 293–300.
8. Patel CJ, Chen R, Kodama K, Ioannidis JPA, Butte AJ (2013) Systematic identification of interaction effects between genome- and environment-wide associations in type 2 diabetes mellitus. Hum Genet 132: 495–508.
9. Wu C, Kraft P, Zhai K, Chang J, Wang Z, et al. (2012) Genome-wide association analyses of esophageal squamous cell carcinoma in chinese identify multiple susceptibility loci and gene-environment interactions. Nat Genet 44: 1090–7.
10. Ma J, Xiao F, Xiong M, Andrew AS, Brenner H, et al. (2012) Natural and orthogonal interaction framework for modeling gene-environment interactions with application to lung cancer. Hum Hered 73: 185–94.
11. Gao J, Nalls MA, Shi M, Joubert BR, Hernandez DG, et al. (2012) An exploratory analysis on gene-environment interactions for parkinson disease. Neurobiol Aging 33: 2528.e1–6.
12. Wei S, Wang LEE, McHugh MK, Han Y, Xiong M, et al. (2012) Genome-wide gene- environment interaction analysis for asbestos exposure in lung cancer susceptibility. Carcinogenesis 33: 1531–7.
13. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics 38: 904.
14. Devlin B, Roeder K, Wasserman L (2001) Genomic control, a new approach to genetic- based association studies. Theor Popul Biol 60: 155–66.
15. Furlotte NA, Kang EY, Van Nas A, Farber CR, Lusis AJ, et al. (2012) Increasing asso- ciation mapping power and resolution in mouse genetic studies through the use of meta- analysis for structured populations. Genetics 191: 959–67.
16. Han B, Eskin E (2012) Interpreting meta-analyses of genome-wide association studies. PLoS Genet 8: e1002555.
17. Flint J, Eskin E (2012) Genome-wide association studies in mice. Nature Reviews Genetics 13: 807.
18. Bennett BJ, Farber CR, Orozco L, Kang HM, Ghazalpour A, et al. (2010) A high- resolution association mapping panel for the dissection of complex traits in mice. Genome Res 20: 281–90.
19. Ghazalpour A, Rau CD, Farber CR, Bennett BJ, Orozco LD, et al. (2012) Hybrid mouse diversity panel: a panel of inbred mouse strains suitable for analysis of complex genetic traits. Mamm Genome 23: 680–92.
20. Valdar W, Solberg LC, Gauguier D, Burnett S, Klenerman P, et al. (2006) Genome-wide genetic association of complex traits in heterogeneous stock mice. Nat Genet 38: 879–87.
21. Yalcin B, Nicod J, Bhomra A, Davidson S, Cleak J, et al. (2010) Commercially available outbred mice for genome-wide association studies. PLoS Genet 6: e1001085.
22. Aylor DL, Valdar W, Foulds-Mathes W, Buus RJ, Verdugo RA, et al. (2011) Genetic analysis of complex traits in the emerging collaborative cross. Genome Res 21: 1213–22.
23. Warden CH, Hedrick CC, Qiao JH, Castellani LW, Lusis AJ (1993) Atherosclerosis in transgenic mice overexpressing apolipoprotein a-ii. Science 261: 469–72.
24. DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. Control Clin Trials 7: 177–88.
25. Cochran WG (2009) The Combination of Estimates from Different Experi- ments. doi: 10.2307/3001666.
26. Higgins JPT, Thompson SG (2002) Quantifying heterogeneity in a meta-analysis. Stat Med 21: 1539–58.
27. Niswender CM, Willis BS, Wallen A, Sweet IR, Jetton TL, et al. (2005) Cre recombinase- dependent expression of a constitutively active mutant allele of the catalytic subunit of protein kinase a. Genesis 43: 109–19.
28. Manning AK, LaValley M, Liu CTT, Rice K, An P, et al. (2011) Meta-analysis of gene- environment interaction: joint estimation of snp and snp×environment regression coeffi- cients. Genetic Epidemiology 35: 11.
29. Kirby A, Kang HM, Wade CM, Cotsapas C, Kostem E, et al. (2010) Fine mapping in 94 inbred mouse strains using a high-density haplotype resource. Genetics 185: 1081–95.
30. Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, et al. (2004) The collaborative cross, a community resource for the genetic analysis of complex traits. Nat Genet 36: 1133–7.
31. Ioannidis JPA, Patsopoulos NA, Evangelou E (2007) Heterogeneity in meta-analyses of genome-wide association investigations. PLoS One 2: e841.
32. Ioannidis JPA, Patsopoulos NA, Evangelou E (2007) Uncertainty in heteroge- neity esti- mates in meta-analyses. BMJ 335: 914–6.
33. Han B, Eskin E (2011) Random-effects model aimed at discovering associations in meta- analysis of genome-wide association studies. Am J Hum Genet 88: 586–98.
34. Hardy RJ, Thompson SG (1996) A likelihood approach to meta-analysis with random effects. Stat Med 15: 619–29.
35. Lin DYY, Sullivan PF (2009) Meta-analysis of genome-wide association studies with over- lapping subjects. Am J Hum Genet 85: 862–72.
36. Devlin B, Roeder K, Bacanu SA (2001) Unbiased methods for population-based association studies. Genet Epidemiol 21: 273–84.

37. Voight BF, Pritchard JK (2005) Confounding from cryptic relatedness in case-control association studies. PLoS Genet 1: e32.

38. Lange K (2002) Mathematical and statistical methods for genetic analysis. Springer Verlag.

39. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, et al. (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38: 203–8.

40. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, et al. (2008) Efficient control of population structure in model organism association mapping. Genetics 178: 1709.

41. Lippert C, Quon G, Kang EY, Kadie CM, Listgarten J, et al. (2013) The benefits of selecting phenotype-specific variants for applications of mixed models in genomics. Scientific Reports 3: 1815.

42. Listgarten J, Lippert C, Kang EY, Xiang J, Kadie CM, et al. (2013) A powerful and efficient set test for genetic markers that handles confounders. Bioinformatics 29: 1526–1533.

43. Stephens M, Balding DJ (2009) Bayesian statistical methods for genetic association stud- ies. Nat Rev Genet 10: 681–90.

44. Parks BW, Nam E, Org E, Kostem E, Norheim F, et al. (2013) Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. Cell Metab 17: 141–52.

45. Davis RC, van Nas A, Castellani LW, Zhao Y, Zhou Z, et al. (2012) Systems genetics of susceptibility to obesity-induced diabetes in mice. Physiol Genomics 44: 1–13.

46. Wang S, Yehya N, Schadt E, Wang H, Drake T, et al. (2006) Genetic and genomic analysis of a fat mass trait with complex inheritance reveals marked sex specificity. PLoS genetics 2: e15.

47. van Nas A, Ingram-Drake L, Sinsheimer JS, Wang SS, Schadt EE, et al. (2010) Expression quantitative trait loci: replication, tissue- and sex-specificity in mice. Genetics 185: 1059–68.

48. van den Maagdenberg AM, Hofker MH, Krimpenfort PJ, de Bruijn I, van Vlijmen B, et al. (1993) Transgenic mice carrying the apolipoprotein e3-leiden gene exhibit hyperlipopro- teinemia. J Biol Chem 268: 10540–5.

49. Jiang XC, Agellon LB, Walsh A, Breslow JL, Tall A (1992) Dietary cholesterol increases transcription of the human cholesteryl ester transfer protein gene in transgenic mice. dependence on natural anking sequences. J Clin Invest 90: 1290–5.

50. Takasuga S, Horie Y, Sasaki J, Sun-Wada GHH, Kawamura N, et al. (2013) Critical roles of type iii phosphatidylinositol phosphate kinase in murine embryonic visceral endoderm and adult intestine. Proc Natl Acad Sci U S A 110: 1726–31.

51. He C, Bassik MC, Moresi V, Sun K, Wei Y, et al. (2012) Exercise-induced bcl2-regulated autophagy is required for muscle glucose homeostasis. Nature 481: 511–5.

52. Plump AS, Azrolan N, Odaka H, Wu L, Jiang X, et al. (1997) Apoa-i knockout mice: characterization of hdl metabolism in homozygotes and identification of a post-rna mechanism of apoa-i up-regulation in heterozygotes. J Lipid Res 38: 1033–47.

53. Liegel R, Chang B, Dubielzig R, Sidjanin DJ (2011) Blind sterile 2 (bs2), a hypomorphic mutation in agps, results in cataracts and male sterility in mice. Mol Genet Metab 103: 51–9.

54. Hofmann JJ, Zovein AC, Koh H, Radtke F, Weinmaster G, et al. (2010) Jagged1 in the portal vein mesenchyme regulates intrahepatic bile duct development: insights into alagille syndrome. Development 137: 4061–72.

55. Farese RV, Sajan MP, Yang H, Li P, Mastorides S, et al. (2007) Muscle-specific knockout of pkc-lambda impairs glucose transport and induces metabolic and diabetic syndromes. J Clin Invest 117: 2289–301.

56. Qiao JH, Tripathi J, Mishra NK, Cai Y, Tripathi S, et al. (1997) Role of macrophage colony-stimulating factor in atherosclerosis: studies of osteopetrotic mice. Am J Pathol 150: 1687–99.

57. Stanford KI, Wang L, Castagnola J, Song D, Bishop JR, et al. (2010) Heparan sulfate 2-o-sulfotransferase is required for triglyceride-rich lipoprotein clearance. J Biol Chem 285: 286–94.

58. Morgan H, Beck T, Blake A, Gates H, Adams N, et al. (2010) Europhenome: a repository for high-throughput mouse phenotyping data. Nucleic Acids Res 38: D577–85.

59. Leshan RL, Greenwald-Yarnell M, Patterson CM, Gonzalez IE, Myers MG (2012) Leptin action through hypothalamic nitric oxide synthase-1-expressing neurons controls energy balance. Nat Med 18: 820–3.

60. Nakata S, Tsutsui M, Shimokawa H, Suda O, Morishita T, et al. (2008) Spontaneous myocardial infarction in mice lacking all nitric oxide synthase isoforms. Circulation 117: 2211–23.

61. Li LO, Ellis JM, Paich HA, Wang S, Gong N, et al. (2009) Liver-specific loss of long chain acyl-coa synthetase-1 decreases triacylglycerol synthesis and beta-oxidation and alters phospholipid fatty acid composition. J Biol Chem 284: 27816–26.

62. Nishina PM, Naggert JK, Verstuyft J, Paigen B (1994) Atherosclerosis in genetically obese mice: the mutants obese, diabetes, fat, tubby, and lethal yellow. Metabolism 43: 554–8.

63. Ciraolo E, Iezzi M, Marone R, Marengo S, Curcio C, et al. (2008) Phosphoinositide 3- kinase p110beta activity: key role in metabolism and mammary gland cancer but not development. Sci Signal 1: ra3.

64. Gupta S, Pablo AM, Jiang Xc, Wang N, Tall AR, et al. (1997) Ifn-gamma potentiates atherosclerosis in apoe knock-out mice. J Clin Invest 99: 2752–61.

65. Kim I, Ahn SHH, Inagaki T, Choi M, Ito S, et al. (2007) Differential regulation of bile acid homeostasis by the farnesoid x receptor in liver and intestine. J Lipid Res 48: 2664–72.

66. Wiedmer T, Zhao J, Li L, Zhou Q, Hevener A, et al. (2004) Adiposity, dyslipidemia, and insulin resistance in mice with targeted deletion of phospholipid scramblase 3 (plscr3). Proc Natl Acad Sci U S A 101: 13296–301.

67. Fan CY, Pan J, Chu R, Lee D, Kluckman KD, et al. (1996) Hepatocellular and hepatic peroxisomal alterations in mice with a disrupted peroxisomal fatty acyl-coenzyme a oxidase gene. J Biol Chem 271: 24698–710.

68. Sainsbury A, Baldock PA, Schwarzer C, Ueno N, Enriquez RF, et al. (2003) Synergistic effects of y2 and y4 receptors on adiposity and bone mass revealed in double knockout mice. Mol Cell Biol 23: 5225–33.

69. Edmondson AC, Braund PS, Stylianou IM, Khera AV, Nelson CP, et al. (2011) Dense genotyping of candidate gene loci identifies variants associated with high-density lipoprotein cholesterol. Circ Cardiovasc Genet 4: 145–55.

70. Foresti O, Ruggiano A, Hannibal-Bach HK, Ejsing CS, Carvalho P (2013) Sterol home- ostasis requires regulated degradation of squalene monooxygenase by the ubiquitin ligase doa10/teb4. eLife 2: e00953.

71. Liu SPP, Li YSS, Chen YJJ, Chiang EPP, Li AFY, et al. (2007) Glycine n-methyltransferase-/- mice develop chronic hepatitis and glycogen storage disease in the liver. Hepatology 46: 1413–25.

72. Bourre JM, Cl_ement M, G_erard D, Chaudi_ere J (1989) Alterations of cholesterol synthe- sis precursors (7-dehydrocholesterol, 7-dehydrodesmosterol, desmosterol) in dysmyelinat- ing neurological mutant mouse (quaking, shiverer and trembler) in the pns and the cns. Biochim Biophys Acta 1004: 387–90.

73. Fujino T, Asaba H, Kang MJJ, Ikeda Y, Sone H, et al. (2003) Low-density lipoprotein receptor-related protein 5 (lrp5) is essential for normal cholesterol metabolism and glucose- induced insulin secretion. Proc Natl Acad Sci U S A 100: 229–34.

74. Kawahara Y, Grimberg A, Teegarden S, Mombereau C, Liu S, et al. (2008) Dysregulated editing of serotonin 2c receptor mrnas results in energy dissipation and loss of fat mass. J Neurosci 28: 12834–44.