Genome-wide association study in two cohorts from a multi-generational mouse advanced intercross line highlights the difficulty of replication

Xinzhu Zhou[1], Celine L. St. Pierre[2], Natalia M. Gonzales[3], Riyan Cheng[4], Apurva Chitre[4], Greta Sokoloff[5], Abraham A. Palmer[4,6*]

[1] Biomedical Sciences Graduate Program, University of California San Diego, La Jolla, CA, USA

[2] Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA

[3] Department of Human Genetics, University of Chicago, Chicago, IL, USA

[4] Department of Psychiatry, University of California San Diego, La Jolla, CA, USA

[5] Department of Psychological & Brain Sciences, University of Iowa, Iowa City, IO, USA

[6] Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA

[*]Corresponding author

Email: aap@ucsd.edu (AAP)

Author contributions

AAP and XZ designed the study, oversaw data collection and analysis, and co-wrote the manuscript. XZ imputed genotypes, performed SNP- and subject-level QC, and conducted GWAS in F_{34} and $F_{39\text{-}43}$ AILs under supervision of AAP. CS prepared GBS libraries for sequencing, as well as organizing portions of the $F_{39\text{-}43}$ phenotypes. NMG de-multiplexed GBS sequencing results and performed alignment and variant calling. RC helped with kinship relatedness matrix calculated from AIL pedigree and with power analysis. AC provided technical support for running programs and scripts. GS collected $F_{39\text{-}43}$ phenotypes.

Data Availability

All relevant data are within the paper and its Supporting Information files. Genotypes and phenotypes of F_{34} ("AIL LGSM F34 (Array)": GN655; "AIL LGSM F34 (GBS)": GN656), $F_{39\text{-}43}$ ("AIL LGSM F39-43 (GBS)": GN657), and mega-analysis cohort ("AIL LGSM F34 and F39-43 (GBS)": GN654) of AIL are uploaded to GeneNetwork2 (http://gn2.genenetwork.org/).

Abstract

Replication is considered to be critical for genome-wide association studies (**GWAS**) in humans, but is not routinely performed in model organisms. We explored replication using an advanced intercross line (**AIL**) which is the simplest possible multigenerational intercross. We re-genotyped a previously published cohort of LG/J x SM/J AIL mice (F_{34}; n=428) using a denser marker set and also genotyped a novel cohort of AIL mice ($F_{39\text{-}43}$; n=600) for the first time. We identified 110 significant loci in the F_{34} cohort, 36 of which were new discoveries attributable to the denser marker set; we also identified 27 novel significant loci in the $F_{39\text{-}43}$ cohort. For traits measured in both cohorts (locomotor activity, body weight, and coat color), the genetic correlations were high, although, the $F_{39\text{-}43}$ cohort showed systematically lower SNP-heritability estimates. We then attempted to replicate loci identified in either F_{34} or $F_{39\text{-}43}$ in the other cohort. Albino coat color was robustly replicated; we observed only partial replication of associations for locomotor activity and body weight. Finally, we performed a mega-analysis of locomotor activity and body weight by combining F_{34} and $F_{39\text{-}43}$ cohorts (n=1,028), which identified four novel loci. The incomplete replication was inconsistent with simulations we performed to estimate our power to replicate. This may reflect: 1) false positives errors in the discovery cohort, 2) environmental or genetic heterogeneity between the two samples, or 3) the systematic over estimation of the effect sizes at significant loci (“Winner’s Curse”). Our results demonstrate that it is difficult to replicate GWAS results even when using similarly sized discovery and replication cohorts drawn from the same population.

Introduction

Despite its importance in human GWAS, replication has been infrequently used for GWAS in model organisms. Challenges in replicating human GWAS findings include genetic, demographic or environmental differences between cohorts. In contrast, model organism GWAS can use genetically identical cohorts phenotyped under extremely similar conditions, which would be expected to enhance the success of replication. We set out to investigate replication in model organism GWAS using an intercross mouse population. The use of multi-parental crosses and commercially available outbred populations for GWAS in model organisms such as mice (1–17), rats (18), chickens (19,20), zebrafish (21,22), fruit flies (23–27), *C. elegans* (28) and various plant species (29–31) has become increasingly common over the last decade. These mapping populations can further be categorized as multi-parental crosses, which are created by interbreeding two or more inbred strains, and commercially available outbred populations, in which the founders are of unknown provenance. An F_2 cross between two inbred strains is the prototypical mapping population; however, F_2s provide poor mapping resolution (32). To improve mapping resolution, Darvasi and Soller (33) proposed the creation of advanced intercross lines (**AILs**), which are produced by intercrossing F_2 mice for additional generations. AILs accumulate additional crossovers with every successive generation, leading to a population with shorter LD blocks, which improves mapping precision, albeit at the expense of power (32,34).

The longest running mouse AIL was generated by crossing LG/J and SM/J inbred strains, which were selectively bred for large and small body size. We obtained this AIL in 2006 at generation 33 from Dr. James Cheverud (Jmc: LG,SM-G_{33}). Since then, we have collected genotype and phenotype information from multiple generations, including F_{34} (16,35–38), F_{39}-F_{43} and $F_{50\text{-}56}$ (39). Our previous publications using the F_{34} generation employed a custom Illumina Infinium genotyping microarray to obtain genotypes for 4,593 SNPs (35,36), we refer to this set

of SNPs as the 'sparse markers'. Those genotypes were used to identify significant associations for numerous traits, including methamphetamine sensitivity (35), pre-pulse inhibition (16), musculoskeletal measurements (17), muscle weight (37), body weight (40), open field (36), conditioned fear (36), red blood cell parameters (41), and murine soleus muscle (38). Although not previously published, we also collected phenotype information from the F_{39}-F_{43} generations, including body weight, fear conditioning, locomotor activity in response to methamphetamine, and the light dark test for anxiety.

While the prior GWAS using the F_{34} generation detected many significant loci, the sparsity of the markers likely precluded the discovery of some true loci, and also made it difficult to clearly define the boundaries of the loci that we did identify. For example, Parker et al conducted an integrated analysis of F_2 and F_{34} AILs (40). One of their body weight loci spanned from 87.93–102.70 Mb on chromosome 14. Denser markers might have more clearly defined implicated region. In the present study, we used genotyping-by-sequencing (**GBS**), which is a reduced-representation sequencing method (42–44), to obtain a much denser set of SNPs in the F_{34} and to genotype mice from the F_{39}-F_{43} generations for the first time. With this denser set of SNPs, we attempted to identify novel loci in the F_{34}s that were not detected using the sparse SNPs. We also performed a GWAS using the mice from the F_{39}-F_{43} AILs. We explored whether imputation from the array SNPs could have provided the additional coverage we obtained using the denser GBS genotypes. Because F_{39}-F_{43} AILs are direct descendants of the F_{34}, they are uniquely suited to serve as a replication population for GWAS in the F_{34} generation. Using multiple cohorts of the same mouse intercross strain, we attempted to examine which loci are replicable between the F_{34} and $F_{39\text{-}43}$ LG/J x SM/J AILs and also performed simulations to estimate the power for these replica on studies. In addition to their use as a replication sample, the F_{39}-F_{43} can also provide improved resolution and allow for the discovery of novel loci not detected in the F_{34} generation. Therefore, we performed a mega-analysis of F_{34} mice and F_{39}-

F_{43} mice to identify loci that were not identified in either individual dataset. Apart from identifying novel loci with a range of physiological and behavioral traits in mice, the present study explores replication in a mouse intercross line by comparing association findings of 1) the same cohort with two genotype panels, 2) two cohorts in the same intercross line, 3) mega-analysis of combined cohorts, and 4) two cohorts with imputed genotypes.

Results

We used 214 males and 214 females from generation F_{34} (Aap:LG,SM-G34) and 305 males and 295 females from generations $F_{39\text{-}43}$. For the F_{34} AIL 79 traits were available from previous published and unpublished, for the $F_{39\text{-}43}$ AIL 49 unpublished traits were available (S1 Table). F_{34} mice had been previously genotyped on a custom SNP array (35,36). The average minor allele frequency (**MAF**) of those 4,593 array SNPs was 0.388 (Fig 1). To obtain a denser set of SNP markers, we used genotyping-by-sequencing in F_{34} and $F_{39\text{-}43}$ AIL mice. Since data about the $F_{39\text{-}43}$ AIL mice had been collected over the span of approximately two years, we carefully considered the possibility of sample contamination and sample mislabeling (45). We removed samples based on four major features: heterozygosity distribution, number of reads aligned to sex chromosomes, discrepancies between pedigree and genetic kinship relatedness, and coat color genotype to phenotype mismatch (see Methods; S1 and S2 Figs). The final SNP sets included 60,392 GBS-derived SNPs in 428 F_{34} AIL mice, 59,790 GBS-derived SNPs in 600 $F_{39\text{-}43}$ AIL mice, and 58,461 GBS-derived SNPs that existed in both F_{34} and $F_{39\text{-}43}$ AIL mice (S2 Table). The MAF for the GBS SNPs was 0.382 in F_{34}, 0.358 in $F_{39\text{-}43}$, and 0.370 in F_{34} and $F_{39\text{-}43}$ (Fig 1). There were 66 SNPs called from our GBS data that were also present on the genotyping array. The genotype concordance rate for those 66 SNPs, which reflects the sum of errors from both sets of genotypes, was 95.4% (S3 Fig). We found that LD decay rates using

F_{34} array, F_{34} GBS, $F_{39\text{-}43}$ GBS, and F_{34} and $F_{39\text{-}43}$ GBS genotypes were generally similar to one another, though levels of LD using the GBS genotypes appear to be slightly reduced in the later generations of AILs (S4 Fig).

GBS genotypes produced more significant associations than array genotypes in F_{34}

We used a linear mixed model (**LMM**) as implemented in GEMMA (46) to perform GWAS. We used the leave-one-chromosome-out (**LOCO**) approach to address the problem of proximal contamination, as previously described (39,47–49). We performed GWAS using both the sparse array SNPs and the dense GBS SNPs to determine whether additional SNPs would produce additional genome-wide significant associations. Autosomal and X chromosome SNPs were included in all GWAS. We obtained a significance threshold for each SNP set using MultiTrans and SLIDES (50,51).To select independently associated loci ("lead loci"), we used a LD-based clumping method implemented in PLINK to group SNPs that passed the adjusted genome-wide significance thresholds over a large genomic region flanking the index SNP (52). Applying the most stringent clumping parameters (r^2 = 0.1 and sliding window size = 12,150kb, S3 Table), we identified 110 significant lead loci in 49 out of 79 F_{34} phenotypes using the GBS SNPs. Table 1 contains the significant associations that contained 5 or less genes; additional significant associations that contained more than 5 genes are listed in S4 Table. In contrast, we identified 83 significant lead loci in 45 out of 79 F_{34} phenotypes using the sparse array SNPs (Table 1, S4 Table). Among the loci identified in the F_{34}, 36 were uniquely identified using the GBS genotypes, whereas 11 were uniquely identified using the array genotypes. GBS SNPs consistently yielded more significant lead loci compared to array SNPs regardless of the clumping parameter values (S3 Table), indicating that a dense marker panel was able to detect more association signals compared to a sparse marker panel.

Table 1. Select lead SNPs with association regions containing less than 5 coding genes. Credible set analysis was performed to define the boundaries of the locus (r^2 threshold = 0.8, posterior probability threshold = 0.99). Genes contained in and/or immediately downstream of the credible set interval were included as associated genes.

Generation	Genotypes	Phenotype	chr	lead snp	pos	P-value	af	beta	se	Associated genes
F34	array	Quadriceps femoris muscle weight (mg)	1	chr1:22249888	22249888	5.69×10^{-06}	0.596	1.93×10^{-01}	4.21×10^{-02}	*Kcnq5, Rims1*
F34	array	Quadriceps femoris muscle weight (mg)	4	chr4:54609724	54609724	1.29×10^{-05}	0.707	1.99×10^{-01}	4.54×10^{-02}	*Zfp462, Rad23b*
F34	array	Extensor digitorum longus muscle weight (mg)	4	chr4:54609724	54609724	1.54×10^{-06}	0.553	2.23×10^{-01}	4.59×10^{-02}	*Zfp462, Rad23b*
F34	array	Gastrocnemius muscle weight (mg)	6	chr6:80659015	80659015	1.69×10^{-05}	0.484	1.78×10^{-01}	4.11×10^{-02}	*Lrrtm4, Gcfc2*
F34	array	Quadriceps femoris muscle weight (mg)	6	chr6:81465256	81465256	1.41×10^{-05}	0.553	1.90×10^{-01}	4.33×10^{-02}	*Lrrtm4, Gcfc2*
F34	array	Coat color, albino	7	chr7:87689885	87689885	3.76×10^{-109}	0.568	5.84×10^{-01}	1.90×10^{-02}	*Tyr*
F34	array	Locomotor test day 2, total distance travelled in 30min (cm)	8	chr8:17159295	17159295	7.15×10^{-06}	0.827	-7.72×10^{-02}	1.70×10^{-02}	*Csmd1*
F34	array	Quadriceps femoris muscle weight (mg)	8	chr8:128443469	128443469	1.05×10^{-07}	0.496	2.23×10^{-01}	4.14×10^{-02}	*Nrp1, Itgb1, Ccdc7b*
F34	array	Retic parameters concentration of hemoglobin in reticulocytes (pg)	8	chr8:128607127	128607127	1.83×10^{-05}	0.414	2.70×10^{-01}	6.26×10^{-02}	*Nrp1, Itgb1, Ccdc7b*
F34	array	Retic parameters concentration of hemoglobin in reticulocytes, repeat (pg)	8	chr8:128607127	128607127	3.50×10^{-06}	0.349	-3.45×10^{-01}	7.38×10^{-02}	*Nrp1, Itgb1, Ccdc7b*
F34	array	Routine CBC cell hemoglobin concentration mean (g/dL)	15	chr15:29893263	29893263	3.86×10^{-06}	0.483	2.93×10^{-01}	6.30×10^{-02}	*Ctnnd2*
F34	array	Routine CBC hemoglobin distribution width (g/dL)	16	chr16:70166221	70166221	2.01×10^{-05}	0.511	2.85×10^{-01}	6.63×10^{-02}	*Gbe1, Speer2, Robo1*
F34	GBS	Routine CBC hematocrit (%)	1	chr1:41580321	41580321	1.37×10^{-05}	0.724	-3.08×10^{-01}	6.99×10^{-02}	*Mfsd9, Pou3f3, 930448I06Rik*
F34	GBS	Tibia length (cm)	1	chr1:77255381	77255381	1.39×10^{-05}	0.433	-3.70×10^{-01}	8.40×10^{-02}	*Epha4, Sgpp2, Pax3*
F34	GBS	Routine CBC red blood cell count (x10^6 cells/uL blood)	1	chr1:148977286	148977286	1.30×10^{-06}	0.526	-3.05×10^{-01}	6.25×10^{-02}	*Bring3, Pla2g4a*
F34	GBS	Routine CBC hemoglobin (g/dL)	1	chr1:148977286	148977286	1.88×10^{-06}	0.522	-3.46×10^{-01}	7.16×10^{-02}	*Brinp3, Pla2g4a*
F34	GBS	Iron content in spleen (ug/kg)	2	chr2:95390277	95390277	4.96×10^{-06}	0.528	2.49×10^{-01}	5.38×10^{-02}	*Lrrc4c, Api5, Alkbh3, Hsd17b12*
F34	GBS	Routine CBC cell hemoglobin concentration mean (g/dL)	2	chr2:128412068	128412068	7.07×10^{-06}	0.449	3.03×10^{-01}	6.63×10^{-02}	*Bcl2l11, Spdye4c, Anapc1, Mertk*
F34	GBS	Iron content in spleen (ug/kg)	3	chr3:52749950	52749950	9.06×10^{-06}	0.508	2.55×10^{-01}	5.70×10^{-02}	*Cog6, Lhfp*
F34	GBS	Iron content in spleen z score	3	chr3:52749950	52749950	1.31×10^{-05}	0.508	2.68×10^{-01}	6.10×10^{-02}	*Cog6, Lhfp*
F34	GBS	Gastrocnemius muscle weight (mg)	6	chr6:80767933	80767933	2.68×10^{-06}	0.483	1.92×10^{-01}	4.06×10^{-02}	*Lrrtm4*
F34	GBS	Quadriceps femoris muscle weight (mg)	6	chr6:81267890	81267890	6.35×10^{-07}	0.597	2.14×10^{-01}	4.24×10^{-02}	*Lrrtm4, Gcfc2*
F34	GBS	Body weight, day 1	6	chr6:81405109	81405109	5.89×10^{-06}	0.497	2.06×10^{-01}	4.51×10^{-02}	*Lrrtm4, Gcfc2*
F34	GBS	Quadriceps femoris muscle weight (mg)	6	chr6:93488270	93488270	3.28×10^{-08}	0.497	2.30×10^{-01}	4.11×10^{-02}	*Prickle2, Adamts9, Magi1*
F34	GBS	Coat color, albino	7	chr7:87642045	8764204	5.00×10^{-106}	0.432	-5.81×10^{-01}	1.91×10^{-02}	*Tyr*
F34	GBS	Routine CBC red blood cell distribution width (%)	7	chr7:115540157	115540157	1.04×10^{-06}	0.379	-3.35×10^{-01}	6.79×10^{-02}	*Sox6, Plekha7*
F34	GBS	Locomotor test day2, total distance travelled in 30min (cm)	8	chr8:17410225	17410225	4.96×10^{-06}	0.171	7.91×10^{-02}	1.71×10^{-02}	*Csmd1*
F34	GBS	Retic parameters concentration of hemoglobin in reticulocytes (pg)	8	chr8:128524048	128524048	3.01×10^{-06}	0.255	-3.54×10^{-01}	7.50×10^{-02}	*Nrp1, Itgb1, Ccdc7b*
F34	GBS	Retic parameters concentration of hemoglobin in reticulocytes, repeat (pg)	8	chr8:128524048	128524048	1.04×10^{-05}	0.471	-3.00×10^{-01}	6.75×10^{-02}	*Nrp1, Itgb1, Ccdc7b*
F34	GBS	Quadriceps femoris muscle weight (mg)	8	chr8:128571869	128571869	1.32×10^{-06}	0.531	-1.97×10^{-01}	4.03×10^{-02}	*Nrp1, Itgb1, Ccdc7b*
F34	GBS	Routine CBC red blood cell distribution width (%)	10	chr10:66602511	66602511	3.42×10^{-08}	0.671	-4.10×10^{-01}	7.32×10^{-02}	*Reep3, Jmjd1c, Nrbf2*
F34	GBS	Retic parameters cell hemoglobin distribution width (pg)	10	chr10:99307553	99307553	3.30×10^{-06}	0.499	-2.84×10^{-01}	6.00×10^{-02}	*Dusp6*
F34	GBS	Retic parameters hemoglobin distribution width (g/dL)	13	chr13:45281420	45281420	1.86×10^{-06}	0.45	2.61×10^{-01}	5.43×10^{-02}	*Dtnbp1, Mylip, Gmpr*
F34	GBS	Routine CBC hemoglobin distribution width (g/dL)	16	chr16:70311614	70311614	5.81×10^{-06}	0.674	3.15×10^{-01}	6.90×10^{-02}	*Speer2, Gbe1*
F39-43	GBS	Coat color, albino	7	chr7:87255156	87255156	5.91×10^{-167}	0.389	-6.24×10^{-01}	1.57×10^{-02}	*Tyr*
F34, F39-43	GBS	Locomotor test day 2, total distance travelled in 30min (cm)	1	chr1:40907532	40907532	3.53×10^{-08}	0.289	-2.53×10^{-01}	4.56×10^{-02}	*Mfsd9, Tmem182*
F34, F39-43	GBS	Locomotor test day 1, total distance travelled in 30min (cm)	1	chr1:122479820	122479820	1.94×10^{-07}	0.373	2.16×10^{-01}	4.13×10^{-02}	*Htr5b, Ddx18, Dpp10*
F34, F39-43	GBS	Body weight	4	chr4:66866758	66866758	1.64×10^{-09}	0.403	-1.84×10^{-01}	3.01×10^{-02}	*Pappa, Trim32, Astn2, Tlr4*
F34, F39-43	GBS	Body weight	6	chr6:81267890	81267890	8.61×10^{-07}	0.507	1.40×10^{-01}	2.83×10^{-02}	*Lrrtm4, Gcfc2*
F34, F39-43	GBS	Coat color, albino	7	chr7:87665727	87665727	6.71×10^{-275}	0.592	5.97×10^{-01}	1.18×10^{-02}	*Tyr*

To determine the boundaries of each locus, we performed a Bayesian-framework credible set analysis, which estimated a posterior probability for association at each SNP (r^2 threshold = 0.8, posterior probability threshold = 0.99; (53)). The physical positions of the SNPs in the credible set were used to determine the boundaries of each locus. As expected, the greater density of the GBS genotypes allowed us to better define each interval. For instance, the lead locus at chr17:27130383 was associated with distance travelled in periphery in the open filed test in F_{34} AILs (Fig 2). However, no SNPs were genotyped between 26.7 and 28.7 Mb in the array SNPs, which makes the size of this LD block ambiguous. In contrast, the LocusZoom plot portraying GBS SNPs in the same region shows that SNPs in high LD with chr17:27130383 are between 27 Mb and 28.3 Mb. The more accurate definition of the implicated intervals allowed us to better refine the list of the coding genes and non-coding variants associated with the phenotype (Table 1).

In our prior studies using the sparse marker set, we did not attempt to increase the number of available markers by using imputation. Therefore, we examined whether the disparity between the numbers of loci identified by the two SNP sets could be resolved by imputation, which should increase the number of markers available for GWAS. We used LG/J and SM/J whole genome sequencing data as reference panels (54) and performed imputation on array and GBS SNPs using Beagle v4.1 (55). After QC filtering, we obtained 4.3M SNPs imputed from the array SNPs and 4.1M SNPs imputed from the GBS SNPs. More imputed GBS SNPs were filtered out because GBS SNPs were called from genotype probabilities, thus introducing uncertainty in imputed SNPs. We found that imputed array genotypes and imputed GBS genotypes did not meaningfully increase the number of loci discovered (Fig S5).

Under a polygenic model where a large number of additive common variants contribute to a complex trait, heritability estimates could be higher when more SNPs are considered (56). Given that there were more GBS SNPs than array SNPs, we used autosomal SNPs to examine

whether GBS SNPs would generate higher SNP heritability estimates compared to the sparse array SNPs. Heritability estimates were similar for the two SNP sets, with the exception of agouti coat color, which showed marginally greater heritability for the GBS SNPs (S6 Fig; S5 Table). Our results show that while the denser GBS SNP set was able to identify more genome-wide significant loci, greater SNP density did not improve the polygenic signal.

Partial replication of loci indented in F_{34} or $F_{39\text{-}43}$ and mega-analysis

We identified 27 genome-wide significant loci for 21 phenotypes in the $F_{39\text{-}43}$ cohort. A subset of those traits, including coat color, body weight, and locomotor activity, were also phenotyped in the F_{34} AILs (Table 2; S8 Table). To assess replication, we determined whether loci that were significant in one cohort (either F_{34} or $F_{39\text{-}43}$) would also be significant in the other. We termed the cohort in which a locus was initially discovered as its “discovery set” and the cohort we attempted replication in as the “replication set” (Table 2). Coat color phenotypes (both albino and agouti) are Mendelian traits and thus served as positive control. As expected, all coat color and body weight loci were replicated. The three body weight loci identified in the F_{34} were replicated at nominal levels of significance ($p<0.05$) in $F_{39\text{-}43}$; similarly, the one body weight locus identified in $F_{39\text{-}43}$ was replicated in F_{34} ($p<0.05$). However, none of the five locomotor activity loci were replicated in the reciprocal (replication) cohorts. We then considered the more liberal “sign test” to determine whether the directions of the effect (beta) of the coat color, body weight and activity loci were in the same direction between the discovery and replication cohorts. We found that 11 of 13 loci passed this much less stringent test of replication. The two loci that did not pass the sign test were the two locomotor loci “discovered” in $F_{39\text{-}43}$ (Table 2).

In light of the failure to replicate the locomotor activity findings, we conducted a series of 2,500 simulations per trait to estimate the expected power of our replication cohorts. For each

phenotype we used the kinship relatedness matrix and variance components estimated from the replication set. For the coat color traits, we found that we had 100% power to replicated the association at either genome-wide significant levels or the more liberal $p<0.05$ threshold (S8 Fig). For body weight and locomotor activity, power to replicate at a genome-wide significance threshold ranged from 50% to 80%, whereas power to replicate at the $p<0.05$ threshold was nearly 100% (S8 Fig). These power estimates were clearly inconsistent with our empirical observations for the locomotor activity traits, none of which replicated at even the $p<0.05$ threshold, where we should have had almost 100% power (Table 2). However, our power simulations did not account for “Winner’s Curse”, which would be expected to systematically overestimate of effect size estimates used in our simulations (57).

Table 2. Replication of significant SNPs between F_{34} and $F_{39\text{-}43}$ AIL association analyses. "Discovery set" indicates the AIL generation that significant SNPs were identified. "Replication set" shows the association p-value, β estimates, etc. of the "discovery set" significant SNPs in the replication AIL generation. SNPs that replicated (p<0.05, same sign for the beta) between F_{34} and $F_{39\text{-}43}$ are highlighted in bold italics. Genetic correlations for phenotypes measured in both F_{34} and $F_{39\text{-}43}$ are listed (see also Supplementary Table 6).

			Discovery set				Replication set			
Phenotype	**rG(*s.e.*)(*p<0.05)**	**SNP**	**P**	**af**	**beta**	**se**	**P**	**af**	**beta**	**se**
			F34 GBS				**F3943 GBS replicate**			
Body weight		***chr4.66414508***	***8.06×10^{-08}***	0.419	-2.50×10^{-01}	4.56×10^{-02}	***1.56×10^{-02}***	0.406	-1.05×10^{-01}	4.33×10^{-02}
	0.711(0.25)*	***chr6.81405109***	***5.89×10^{-06}***	0.497	2.06×10^{-01}	4.51×10^{-02}	***2.36×10^{-02}***	0.518	9.48×10^{-02}	4.18×10^{-02}
		chr14.79312393	***7.53×10^{-06}***	0.514	-2.01×10^{-01}	4.45×10^{-02}	***1.44×10^{-02}***	0.566	-1.04×10^{-01}	4.26×10^{-02}
Coat color, albino	0.967(0.04)*	***chr7.87642045***	***5.00×10^{-106}***	0.432	-5.81×10^{-01}	1.91×10^{-02}	***2.85×10^{-163}***	0.387	-6.07×10^{-01}	1.55×10^{-02}
Coat color, agouti	0.971(0.04)*	***chr2.154464466***	***9.43×10^{-191}***	0.129	9.39×10^{-01}	1.25×10^{-02}	***5.70×10^{-93}***	0.207	7.20×10^{-01}	2.57×10^{-02}
Locomotor test day 1, total distance travelled in 30min	0.968(0.24)*	chr19.21812298	9.28×10^{-07}	0.461	$-6.90\times10^{+02}$	$1.39\times10^{+02}$	7.72×10^{-01}	0.52	-1.94×10^{-02}	6.74×10^{-02}
Locomotor test day2, total distance travelled in 30min	0.988(0.19)*	chr7.45084416	1.12×10^{-05}	0.246	$6.77\times10^{+02}$	$1.53\times10^{+02}$	3.40×10^{-01}	0.217	7.38×10^{-02}	7.77×10^{-02}
		chr8.17410225	4.96×10^{-06}	0.171	$7.91\times10^{+02}$	$1.71\times10^{+02}$	4.03×10^{-01}	0.192	7.04×10^{-02}	8.43×10^{-02}
			F3943 GBS				**F34 GBS replicate**			
Body weight	0.711(0.25)*	***chr14.82586326***	***2.63×10^{-06}***	0.658	-2.09×10^{-01}	4.43×10^{-02}	***2.87×10^{-05}***	0.575	-1.89×10^{-01}	4.50×10^{-02}
Coat color, albino	0.967(0.04)*	***chr7.87255156***	***5.91×10^{-167}***	0.389	-6.24×10^{-01}	1.57×10^{-02}	***7.80×10^{-97}***	0.444	-5.75×10^{-01}	2.07×10^{-02}
Coat color, agouti	0.971(0.04)*	***chr2.155091628***	***1.78×10^{-115}***	0.218	7.42×10^{-01}	2.17×10^{-02}	***1.51×10^{-185}***	0.135	8.98×10^{-01}	1.33×10^{-02}
Locomotor test day 2, total distance travelled in 30min	0.988(0.19)*	chr15.67235072	3.21×10^{-06}	0.47	3.14×10^{-01}	6.63×10^{-02}	1.69×10^{-01}	0.522	$-1.72\times10^{+02}$	$1.25\times10^{+02}$
Locomotor test day 3, total distance travelled in 30min	0.577(0.22)*	chr7.113250866	5.88×10^{-06}	0.389	3.29×10^{-01}	7.20×10^{-02}	8.27×10^{-01}	0.483	$-1.21\times10^{+02}$	$5.45\times10^{+02}$

We also evaluated whether or not the traits showed genetic correlations across the two cohorts; low genetic correlations between the two cohorts could indicate that environmental sources of heterogeneity had limited the potential for replication (F_{34} and $F_{39\text{-}43}$). We used autosomal SNPs to calculate genetic correlations between the F_{34} and $F_{39\text{-}43}$ generations for body weight, coat color, and locomotor activity phenotypes (S6 Table), using GCTA-GREML (58). Albino and agouti coat color, body weight and locomotor activity on days 1 and 2 were highly genetically correlated (r_Gs >0.7; S6 Table). In contrast, locomotor activity on day 3 showed a significant but weaker genetic correlation (r_G=0.577), perhaps reflecting variability in the quality of the methamphetamine injection, which were only given on day 3. Overall, these results suggest that genetic influences on these traits were broadly similar in the two cohorts; however, the genetic correlations were less than 1, suggesting an additional barrier to replication that was not accounted for in our power simulations.

We also calculated the SNP heritabilities for all traits using GCTA. SNP heritability was consistently lower in the $F_{39\text{-}43}$ compared to the F_{34}, possibly a result of increased experimental variance introduced by our extended phenotype collection period (Fig 3; S7 Table).

Due to the relatively high genetic correlations (S6 Table), we suspected that a mega-analysis using the combined sample set would allow for the identification of additional loci; indeed, mega-analysis identified four novel genome-wide significant associations (Fig 4; S9 Table). The significance of the associations identified by the mega-analysis was often greater than in either individual cohort. For instance, the p-values obtained by mega-analysis for chr4:66866758 ($p = 1.64\times10^{-9}$) and chr14:82672838 ($p = 2.06\times10^{-10}$) for body weight were lower than the corresponding p-values for the same loci for F_{34} (chr4:65246120, $p = 9.06\times10^{-8}$; chr14:78926547, $p = 6.24\times10^{-6}$) and $F_{39\text{-}43}$ (chr4:66414508, $p = 8.06\times10^{-8}$; chr14:79312393, $p = 7.53\times10^{-6}$; S7 Fig).

Discussion

We used F_{34} and $F_{39\text{-}43}$ generations of a LG/J x SM/J AIL to perform GWAS, SNP heritability estimates, genetic correlations, replication and mega-analysis. We had previously performed several GWAS using a sparse marker set in the F_{34} cohort. In this study we used a denser set of SNPs, obtained using GBS, to reanalyze the F_{34} cohort. We found 110 significant loci, 36 of which had not been identified in our prior studies using the sparse marker set. We used a new, previously unpublished $F_{39\text{-}43}$ cohort and showed that genetic correlations were high for the subset of traits that were measured in both cohorts. Despite this, we found that some but not all loci replicated between cohorts, even when we used a relatively liberal definition of replication ($p<0.05$). The failure to replicate some of our findings was not predicted by our power simulations. Imputation to reference panels increased the number of SNPs available for analysis but did not meaningfully enhance the number of loci we discovered presumably because it did not improve our ability to capture recombination events. Finally, mega-analysis of the two cohorts allowed us to discover 4 additional loci. Taken together, we have identified refined regions of associations for numerous physiological and behavioral traits in multiple generations of AILs.

Previous publications from our lab used a sparse set of array genotypes for GWAS of various behavioral and physiological traits in 688 F_{34} AILs (16,17,35–38,40,41). In this study we obtained a much denser marker set for 428 of the initial 688 AIL mice using GBS. The denser genotypes allowed us to identify most of the loci obtained using the sparse set, as well as many additional loci. For instance, using the sparse markers we identified significant locus on chromosome 8 for locomotor day 2 activity that contained only one gene: *Csmd1* (CUB and sushi multiple domains 1). Gonzales et al. (39) replicated this finding in $F_{50\text{-}56}$ AIL and identified a *cis*-eQTL mapped to the same region. *Csmd1* mutant mice showed increased locomotor activity compared to wild-type and heterozygous mice, indicating that *Csmd1* is likely a causal

gene for locomotor and related traits (39). We replicated this locus in the analysis of the F_{34} cohort that used the denser marker set (S7 Fig). We also replicated a locus on chromosome 17 for distance traveled in the periphery in the open field test (Fig 4; (36,39)), three loci on chromosomes 4, 6, and 14 for body weight (Supplementary Fig 7; (40)), one locus on chromosome 7 for mean corpuscular hemoglobin concentrations (MCHC, complete blood count; S7 Fig; (41)), and numerous loci on chromosome 4, 6, 7, 8, and 11 for muscle weights (Supplementary Fig 7; (37)). We noticed that even using original spares markers, some previously published loci were not replicated in the current GWAS. The most likely explanation is that we had only 428 of the 688 mice used in the previous publications. Methodological differences between prior studies and the current study, such as the use of QTLRel rather than GEMMA and the choice of pedigree rather than genotypes for estimating relatedness, may also lead to lack of complete replication (56).

$F_{39\text{-}43}$ AILs replicated some, but not all, significant loci identified in F_{34}, despite generally high genetic correlations between the two cohorts. Significant loci for coat color, which are monogenic and served as positive controls, were consistent between the two cohorts. Loci for body weight were fully replicated ($p<0.05$) between F_{34} and $F_{39\text{-}43}$, while loci for locomotor activity were not. Nevertheless, the beta estimates for all but two loci shared the same sign, which provides modest evidence for replication. Several possibilities may cause the lack of replication for locomotor activity. Firstly, the loci for locomotor activity identified in F_{34} could be false positives. We controlled the genome-wide false positive error rate at 5% using permutation, however, our replication study considered six phenotypes, which was not accounted for by our permutations, thus somewhat increasing the chances that at least one of our significant associations could have been a false positive. Second, unlike the F_{34} dataset, the $F_{39\text{-}43}$ used multiple technicians to conduct the behavioral tests and occurred over a prolonged period of time in which numerous environmental factors may have changed. The circumstances under

which the F_{39-43} data were collected may have introduced greater environmental heterogeneity, possibly contributing to their lower heritability and the limited replication. The genetic correlations (S6 Table) were high but still less than 1, suggesting a role for heterogeneity. Finally, our simulation did not account for the systematic overestimation of the effect sizes ("the Winner's Curse" (57)). The failure to replicate the locomotor activity loci is likely due to a combination of these factors: false positives, heterogeneity, and Winner's Curse. Thus, while it may seem intuitive that a genome-wide significant result should be reliable at a nominal threshold of $p<0.05$, when a similarly sized cohort drawn from the same population is used, our results show that this is not the case.

Finally, we performed a mega-analysis using F_{34} and F_{39-43} AIL mice. The combined dataset increased our power and allowed us to identify four novel genome-wide significant associations that were not detected in either the F_{34} or the F_{39-43} cohorts. For example, the mega-analysis identified a locus for body weight on chromosome 2 (Fig S7). Parker et al. (40) identified the same locus using an integrated analysis of LG/J x SM/J F_2 and F_{34} AILs.

QTL mapping studies have traditionally used a 1.0~2.0 LOD support interval to approximate the size of the association region (see (72,73)). The LOD support interval, proposed by Conneally et al. (74) and Lander & Botstein (75), is a simple confidence interval method involving converting the p-value of the peak locus into a LOD score, subtracting "drop size" from the peak locus LOD score, and finding the two physical positions to the left and to the right of the peak locus location that correspond to the subtracted LOD score. Although Mangin et al. (76) showed via simulation that the boundaries of LOD support intervals depend on effect size, others observed that a 1.0 ~ 2.0 LOD support interval accurately captures ~95% coverage of the true location of the loci when using a dense set of markers (75,77,78). In the present study, we considered using LOD support intervals but found that the sparse array SNPs produced misleadingly large support intervals. Various methods

have been proposed for calculating confidence intervals in analogous situations (e.g. (12,79)). We performed credible set analysis and compared LocusZoom plots of the same locus region between array SNPs and the GBS SNPs (S7 Fig; (80)). For example, the benefit of the denser SNP coverage is easily observed in the locus on chromosome 7 (array lead SNP chr7:44560350; GBS lead SNP chr7:44630890) for the complete blood count trait "retic parameters cell hemoglobin concentration mean, repeat" (Fig S7). Thus, there are advantages of dense SNP sets that go beyond the ability to discover additional loci.

LD in the LG/J x SM/J AIL mice is more extensive than in the Diversity Outbred mice and Carworth Farms White mice (1). Some of the loci that we identified are relatively large, making it difficult to infer which genes are responsible for the association. We focused on loci that contained 5 or fewer genes (Table 1). We highlight a few genes that are supported by the existing literature for their role in the corresponding traits. The lead SNP at chr1:77255381 is associated with tibia length in F_{34} AILs (Table 1; S7 Fig). One gene at this locus, *EphA4*, codes for a receptor for membrane-bound ephrins. EphA4 plays an important role in the activation of the tyrosine kinase Jak2 and the signal transducer and transcriptional activator Stat5B in muscle, promoting the synthesis of insulin-like growth factor 1 (IGF-1) (64–66). Mice with mutated EphA4 shows significant defect in body growth (66). Curiously, another gene at this locus, *Pax3*, has been shown as a transcription factor expressed in resident muscle progenitor cells and is essential for the formation of skeletal muscle in mice (67). It is possible that both *EphA4* and *Pax3* are associated with the trait tibia length because they are both involved in organismal growth. Another region of interest is the locus at chr4:66866758, which is associated with body weight in all AIL cohorts (Table 1; S4 Table; S8 Table; S9 Table). The lead SNP is immediately upstream of *Tlr4*, Toll-like receptor 4, which recognizes Gram-negative bacteria by its cell wall component, lipopolysaccharide (68,69). TLR4 responds to the high circulating level of fatty acids and induces inflammatory signaling, which leads to insulin resistance (70). Kim et al showed TLR4-deficient mice were protected

from the increase in proinflammatory cytokine level and gained less weight than wild-type mice when fed on high fat diet (71). The association between *Tlr4* and body weight in the AILs corroborates these findings.

Our study has notable limitations. First, not all F_{34} and $F_{39\text{-}43}$ animals that were phenotyped were later genotyped by GBS due to missing DNA samples, which in turn lowered our sample size and reduced the power of association analyses. Second, $F_{39\text{-}43}$ traits have been collected by different technicians over the span of several years, which introduced environmental heterogeneity and diminished trait heritability (Fig 2). Finally, our power simulations did not account for common factors that can limit replication, including false positives errors, heterogeneity and the Winner's Curse.

The present study explored replication of GWAS results, the role of marker density, and imputation in GWAS. The combination of denser marker coverage and the addition of 600 $F_{39\text{-}43}$ AIL mice allowed us to identify novel loci for locomotor activity, open field test, fear conditioning, light dark test for anxiety, complete blood count, iron content in liver and spleen, and muscle weight. An important conclusion is that power for replication is modest even when a similarly sized replication cohort is used in a genetically identical population tested under conditions that are designed to be as similar as possible.

Materials and Methods

Animals

All mice used in this study were members of the advanced intercross line (**AIL**) between LG/J and SM/J that was originally created by Dr. James Cheverud (Loyola University Chicago, Chicago, IL). The AIL line has been maintained in the Palmer laboratory since generation F_{33}. Age and exact number of animals tested in each phenotype are described in S1 Table. Several

previous publications (16,35–38,40,41) have reported on association analyses of the F_{34} mice (n=428). No prior publications have described the $F_{39\text{-}43}$ generations (n=600). The sample size of F_{34} mice reported in this study (n=428) is smaller than that in previous publications of F_{34} (n=688) because we only sequenced a subset of F_{34} animals using GBS. With the exception of coat color, we quantile normalized all phenotypes. Coat color traits were coded in binary numbers (albino: 1 = white, 0 = non-white; agouti: 1 = tan, 0 = black, NA=white). Locomotor activity traits in F_{34} were not quantile transformed in order to follow the guideline described in Cheng et al. (35) for direct comparison.

F_{34}, $F_{39\text{-}43}$ Phenotypes

We have previously described the phenotyping of F_{34} animals for locomotor activity (35), fear conditioning (36), open field (36), coat color, body weight (40), complete blood counts (41), heart and tibia measurements (37), muscle weight (37). Iron content in liver and spleen, which have not been previously reported in these mice, was measured by atomic absorption spectrophotometry, as described in Gardenghi et al. (59) and Graziano, Grady and Cerami (60). Although the phenotyping of $F_{39\text{-}43}$ animals has not been previously reported, we used method that were identical to those previously reported for locomotor activity (35), open field (36), coat color, body weight (40), and light/dark test for anxiety (15).

F_{34} AIL Array Genotypes

F_{34} animals had been genotyped on a custom SNP array on the Illumina Infinium platform (35,36), which yielded a set of 4,593 SNPs on autosomes and X chromosome that we refer to as ‘array SNPs’.

F_{34} and $F_{39\text{-}43}$ GBS Genotypes

F_{34} and $F_{39\text{-}43}$ animals were genotyped using genotyping-by-sequencing (**GBS**), which is a reduced-representation genome sequencing method (1,39). We used the same protocol for GBS library preparation that was described in Gonzales et al (39). We called GBS genotype probabilities using ANGSD (61). GBS identified 1,667,920 autosomal and 43,015 X-chromosome SNPs. To fill in missing genotypes at SNPs where some but not all mice had calls, we ran within-sample imputation using Beagle v4.1, which generated hard call genotypes as well as genotype probabilities (55). After imputation, only SNPs that had dosage $r^2 > 0.9$ were retained. We removed SNPs with minor allele frequency < 0.1 and SNPs with $p < 1.0\times10^{-6}$ in the Chi-square test of Hardy–Weinberg Equilibrium (**HWE**) (S2 Table). All phenotype and GBS genotype data are deposited in GeneNetwork (http://www.genenetwork.org).

QC of individuals

We have found that large genetic studies are often hampered by cross-contamination between samples and sample mix-ups. We used four features of the data to identify problematic samples: heterozygosity distribution, proportion of reads aligned to sex chromosomes, pedigree/kinship, and coat color. We first examined heterozygosity across autosomes and removed animals where the proportion of heterozygosity that was more than 3 standard deviations from the mean (S1 Fig). Next, we sought to identify animals in which the recorded sex did not agree with the sequencing data. We compared the ratio of reads mapped to the X and Y chromosomes. The 95% CI for this ratio was 196.84 to 214.3 in females and 2.13 to 2.18 in males. Twenty-two F_{34} and $F_{39\text{-}43}$ animals were removed because their sex (as determined by

reads ratio) did not agree with their recorded sex; we assumed this discrepancy was due to sample mix-ups. To further identify mislabeled samples, we calculated kinship coefficients based on the full AIL pedigree using QTLRel. We then calculated a genetic relatedness matrix (**GRM**) using IBDLD, which estimates identity by descent using genotype data. The comparison between pedigree kinship relatedness and genetic kinship relatedness identified 7 pairs of animals that showed obvious disagreement between kinship coefficients and the GRM. Lastly, we excluded 14 $F_{39\text{-}43}$ animals that showed discordance between their recorded coat color and their genotypes at markers flanking *Tyr*, which causes albinism in mice. The numbers of animals filtered at each step are listed in S2 Table. Some animals were detected by more than one QC step, substantiating our belief that these samples were erroneous.

At the end of SNP and sample filtering, we had 59,561 autosomal and 831 X chromosome SNPs in F_{34}, 58,966 autosomal and 824 X chromosome SNPs in $F_{39\text{-}43}$, and 57,635 autosomal and 826 X chromosome SNPs in the combined F_{34} and $F_{39\text{-}43}$ set (S2 Table). GBS genotype quality was estimated by examining concordance between the 66 SNPs that were present in both the array and GBS genotyping results.

LD decay

Average LD (r^2) was calculated using allele frequency matched SNPs (MAF difference < 0.05) within 100,000bp distance, as described in Parker et al. (1).

Imputation to LG/J and SM/J reference panels

F_{34} array genotypes (n=428) and F_{34} GBS genotypes (n=428) were imputed to LG/J and SM/J whole genome sequence data (54) using BEAGLE. For F_{34} array imputation, we used a

large window size (100,000 SNPs and 45,000 SNPs overlap). Imputation to reference panels yielded 4.3 million SNPs for F_{34} array and F_{34} GBS imputed sets. Imputed SNPs with $DR^2 > 0.9$, MAF > 0.1, HWE p value > 1.0×10^{-6} were retained, resulting in 4.1M imputed F_{34} GBS SNPs and 4.3M imputed F_{34} array SNPs.

Genome-wide association analysis (GWAS)

We used the linear mixed model, as implemented in GEMMA (46), to perform a GWAS that accounted for the complex familial relationships among the AIL mice (35,39). We used the leave-one-chromosome-out (**LOCO**) approach to calculate genetic relatedness matrix, which effectively circumvented the problem of proximal contamination (48). Separate GWAS were performed using F_{34} array genotypes, F_{34} GBS genotypes, and $F_{39\text{-}43}$ GBS genotypes. Apart from coat color (binary trait) and locomotor activity, raw phenotypes were quantile normalized prior to analysis. Locomotor activity was not quantile normalized because the trait was reasonably normally distributed already and because we wanted our analysis to match those performed in Cheng et al (35). Because F_{34} AIL had already been studied using array genotypes (35) and mapped using QTLRel (62), we used the same covariates as described in Cheng et al. (35) in order to examine whether our array and GBS GWAS would replicate their findings. We included sex and body weight as covariates for locomotor activity traits (see covariates used in (35))and sex, age, and coat color as covariates for fear conditioning and open field test in F_{34} AILs (see covariates used in (36)). We used sex and age as covariates for all other phenotypes. Covariates for each analysis are shown in S1 Table. Finally, we performed mega-analysis of F_{34} and $F_{39\text{-}43}$ animals (n=1,028) for body weight, coat color, and locomotor activity, since these traits were measured in the same way in both cohorts. For the mega-analyses, locomotor

activity was quantile normalized after the combination of the two datasets to ensure that data were normally distributed across generations.

Identifying suspicious SNPs

Some significant SNPs in F_{34} GWAS and in the mega-analysis of F_{34} and $F_{39\text{-}43}$ were suspicious because nearby SNPs, which would have been expected to be in high LD (a very strong assumption in an AIL), did not have high -log10 values. We only examined SNPs that obtained significant p-values; these examinations reveled that these SNPs had suspicious ratios of heterozygotes to homozygotes calls and had corresponding HWE p-values that were close to our 1.0×10^{-6} threshold (S10 and S11 Tables). To avoid counting these as novel loci, we removed those SNPs prior to summarizing our results as they likely reflected genotyping errors.

Selecting independent significant SNPs

To identify independent “lead loci” among significant GWAS SNPs that surpass the significance threshold, we used the LD-based clumping method in PLINK v1.9. We empirically chose clumping parameters (r^2 = 0.1 and sliding window size = 12,150kb) that gave us a conservative set of independent SNPs (S3 Table). For the coat color phenotypes, we found that multiple SNPs remained significant even after LD-based clumping, presumably due to the extremely significant associations at these Mendelian loci. In these cases, we used a stepwise model selection procedure in GCTA (58) and performed association analyses conditioning on the most significant SNPs.

Significance thresholds

We used MultiTrans and SLIDE to set significance thresholds for the GWAS (50,51). MultiTrans and SLIDE are methods that assume multivariate normal distribution of the phenotypes, which in LMM models, contain a covariance structure due to various degrees of relatedness among individuals. We were curious to see whether MultiTrans/SLIDE produces significance thresholds drastically different from the threshold we obtained from a standard permutation test ('naïve permutation' as per Cheng et al. (48)). We performed 1,000 permutations using the F_{34} GBS genotypes and the phenotypic data from locomotor activity (days 1, 2, and 3). We found that the 95th percentile values for these permutations were 4.65, 4.79, and 4.85, respectively, which were very similar to 4.85, the threshold obtained from MultiTrans using the same data. Thus, the thresholds presented here were obtained from MultiTrans but are similar (if anything slightly more conservative) than thresholds we would have obtained had we used permutation. Because the effective number of tests depends on the number of SNPs and the specific animals used in GWAS, we obtained a unique adjusted significance threshold for each SNP set in each animal cohort (S12 Table).

Power analysis

To estimate the power of replication of a SNP from the discovery set in the replication set, we simulated GWAS with 50 varying effect sizes for the discovery SNP using the LMM model. We first fit the trait in a null model (i.e., no genotype effect), and obtained estimates of model parameters including the intercept and the genetic variance component. Using these model parameters, we added the genotype effect to the random numbers generated from the null model to recreate a trait. For each

simulated effect size, we scanned every simulated trait 2,500 times and examined the ratio of association tests whose test statistics surpassed the significance thresholds (both the genome-wide significance threshold for the cohort and the nominal P value of 0.05). In order to compare effect sizes measured in phenotypes with different scales and units, we converted beta estimates (of discovery SNPs and simulated effect sizes) to z scores based on the standard errors of the beta estimates of discovery SNPs in the simulation cohort.

Credible set analysis

We followed the method described in (53). The R script could be found on GitHub: https://github.com/hailianghuang/FM-summary/blob/master/getCredible.r

Genetic correlation and heritability estimates between F_{34} and $F_{39\text{-}43}$ phenotypes

Locomotor activity, body weight, and coat color had been measured in both F_{34} and $F_{39\text{-}43}$ populations. We calculated both SNP heritability and genetic correlations between F_{34} and $F_{39\text{-}43}$ animals using GCTA bivariate GREML analysis (58). Because $F_{39\text{-}43}$ day 1 locomotor activity data were not normally distributed, we quantile normalized locomotor activity data when estimating SNP heritabilities and genetic correlations.

LocusZoom Plots

LocusZoom plots were generated using the standalone implementation of LocusZoom (63), using LD scores calculated from PLINK v.1.9 --ld option and mm10 gene annotation file downloaded from USCS genome browser.

Ethics Statement

All procedures were approved by the Institutional Animal Care and Use Committee (IACUC protocol: S15226) Euthanasia was accomplished using CO_2 asphyxiation followed by cervical dislocation.

Acknowledgements

We would like to recognize Jackie Lim and Kaitlin Samocha for collecting F_{34} AIL phenotype data and Ryan Walters for collecting $F_{39\text{-}43}$ AIL phenotype data. We wish to acknowledge Alex Gileta for input on a draft of this manuscript.

References

1. Parker CC, Gopalakrishnan S, Carbonetto P, Gonzales NM, Leung E, Park YJ, et al. Genome-wide association study of behavioral, physiological and gene expression traits in outbred CFW mice. Nat Genet. 2016;48(8):919.

2. Gatti DM, Svenson KL, Shabalin A, Wu L-Y, Valdar W, Simecek P, et al. Quantitative trait locus mapping methods for diversity outbred mice. G3 Genes Genomes Genet. 2014;4(9):1623–1633.

3. Chesler EJ. Out of the bottleneck: the Diversity Outcross and Collaborative Cross mouse populations in behavioral genetics research. Mamm Genome. 2014;25(1–2):3–11.

4. Churchill GA, Gatti DM, Munger SC, Svenson KL. The diversity outbred mouse population. Mamm Genome. 2012;23(9–10):713–718.

5. Consortium CC. The genome architecture of the Collaborative Cross mouse genetic reference population. Genetics. 2012;190(2):389–401.

6. Talbot CJ, Nicod A, Cherny SS, Fulker DW, Collins AC, Flint J. High-resolution mapping of quantitative trait loci in outbred mice. Nat Genet. 1999;21(3):305.

7. Demarest K, Koyner J, McCaughran J, Cipp L, Hitzemann R. Further characterization and high-resolution mapping of quantitative trait loci for ethanol-induced locomotor activity. Behav Genet. 2001;31(1):79–91.

8. Valdar W, Solberg LC, Gauguier D, Burnett S, Klenerman P, Cookson WO, et al. Genome-wide genetic association of complex traits in heterogeneous stock mice. Nat Genet. 2006;38(8):879.

9. Ghazalpour A, Doss S, Kang H, Farber C, Wen P-Z, Brozell A, et al. High-resolution mapping of gene expression using association in an outbred mouse stock. PLoS Genet. 2008;4(8):e1000149.

10. Svenson KL, Gatti DM, Valdar W, Welsh CE, Cheng R, Chesler EJ, et al. High-resolution genetic mapping using the Mouse Diversity outbred population. Genetics. 2012;190(2):437–447.

11. Yalcin B, Willis-Owen SA, Fullerton J, Meesaq A, Deacon RM, Rawlins JNP, et al. Genetic dissection of a behavioral quantitative trait locus shows that Rgs2 modulates anxiety in mice. Nat Genet. 2004;36(11):1197.

12. Nicod J, Davies RW, Cai N, Hassett C, Goodstadt L, Cosgrove C, et al. Genome-wide association of multiple complex traits in outbred mice by ultra-low-coverage sequencing. Nat Genet. 2016;48(8):912.

13. Carbonetto P, Cheng R, Gyekis JP, Parker CC, Blizard DA, Palmer AA, et al. Discovery and refinement of muscle weight QTLs in B6$\times$ D2 advanced intercross mice. Physiol Genomics. 2014;46(16):571–582.

14. Coyner J, McGuire JL, Parker CC, Ursano RJ, Palmer AA, Johnson LR. Mice selectively bred for High and Low fear behavior show differences in the number of pMAPK (p44/42 ERK) expressing neurons

563 in lateral amygdala following Pavlovian fear conditioning. Neurobiol Learn Mem. 2014;112:195–
564 203.

565 15. Parker CC, Cheng R, Sokoloff G, Palmer AA. Genome-wide association for methamphetamine
566 sensitivity in an advanced intercross mouse line. Genes Brain Behav. 2012;11(1):52–61.

567 16. Samocha KE, Lim JE, Cheng R, Sokoloff G, Palmer AA. Fine mapping of QTL for prepulse inhibition in
568 LG/J and SM/J mice using F2 and advanced intercross lines. Genes Brain Behav. 2010;9(7):759–767.

569 17. Hernandez Cordero AI, Carbonetto P, Riboni Verri G, Gregory JS, Vandenbergh DJ, P Gyekis J, et al.
570 Replication and discovery of musculoskeletal QTLs in LG/J and SM/J advanced intercross lines.
571 Physiol Rep. 2018;6(4).

572 18. Baud A, Guryev V, Hummel O, Johannesson M, Hermsen R, Stridh P, et al. Genomes and phenomes
573 of a population of outbred rats and its progenitors. Sci Data. 2014;1:140011.

574 19. Besnier F, Wahlberg P, Rönneg\a ard L, Ek W, Andersson L, Siegel PB, et al. Fine mapping and
575 replication of QTL in outbred chicken advanced intercross lines. Genet Sel Evol. 2011;43(1):3.

576 20. Johnsson M, Henriksen R, Höglund A, Fogelholm J, Jensen P, Wright D. Genetical genomics of
577 growth in a chicken model. BMC Genomics. 2018;19(1):72.

578 21. Guryev V, Koudijs MJ, Berezikov E, Johnson SL, Plasterk RH, Van Eeden FJ, et al. Genetic variation in
579 the zebrafish. Genome Res. 2006;16(4):491–497.

580 22. Patowary A, Purkanti R, Singh M, Chauhan R, Singh AR, Swarnkar M, et al. A sequence-based
581 variation map of zebrafish. Zebrafish. 2013;10(1):15–20.

582 23. Mackay TF, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, et al. The Drosophila
583 melanogaster genetic reference panel. Nature. 2012;482(7384):173.

584 24. Vonesch SC, Lamparter D, Mackay TF, Bergmann S, Hafen E. Genome-wide analysis reveals novel
585 regulators of growth in Drosophila melanogaster. PLoS Genet. 2016;12(1):e1005616.

586 25. King EG, Macdonald SJ, Long AD. Properties and power of the Drosophila Synthetic Population
587 Resource for the routine dissection of complex traits. Genetics. 2012;genetics–112.

588 26. Kislukhin G, King EG, Walters KN, Macdonald SJ, Long AD. The genetic architecture of
589 methotrexate toxicity is similar in Drosophila melanogaster and humans. G3 Genes Genomes
590 Genet. 2013;g3–113.

591 27. Marriage TN, King EG, Long AD, Macdonald SJ. Fine-mapping nicotine resistance loci in Drosophila
592 using a multiparent advanced generation inter-cross population. Genetics. 2014;198(1):45–57.

593 28. Doitsidou M, Jarriault S, Poole RJ. Next-generation sequencing-based approaches for mutation
594 mapping and identification in Caenorhabditis elegans. Genetics. 2016;204(2):451–474.

595 29. Diouf IA, Derivot L, Bitton F, Pascual L, Causse M. Water Deficit and Salinity Stress Reveal Many
596 Specific QTL for Plant Growth and Fruit Quality Traits in Tomato. Front Plant Sci. 2018;9:279.

30. Cockram J, Mackay I. Genetic Mapping Populations for Conducting High-Resolution Trait Mapping in Plants. 2018;

31. Rishmawi L, Bühler J, Jaegle B, Hülskamp M, Koornneef M. Quantitative trait loci controlling leaf venation in Arabidopsis. Plant Cell Environ. 2017;40(8):1429–1441.

32. Parker CC, Palmer AA. Dark matter: are mice the solution to missing heritability? Front Genet. 2011;2:32.

33. Darvasi A, Soller M. Advanced intercross lines, an experimental population for fine genetic mapping. Genetics. 1995;141(3):1199–1207.

34. Gonzales NM, Palmer AA. Fine-mapping QTLs in advanced intercross lines and other outbred populations. Mamm Genome. 2014;25(7–8):271–292.

35. Cheng R, Lim JE, Samocha KE, Sokoloff G, Abney M, Skol AD, et al. Genome-wide association studies and the problem of relatedness among advanced intercross lines and other highly recombinant populations. Genetics. 2010;

36. Parker CC, Carbonetto P, Sokoloff G, Park YJ, Abney M, Palmer AA. High-resolution genetic mapping of complex traits from a combined analysis of F2 and advanced intercross mice. Genetics. 2014;198(1):103–116.

37. Lionikas A, Cheng R, Lim JE, Palmer AA, Blizard DA. Fine-mapping of muscle weight QTL in LG/J and SM/J intercrosses. Physiol Genomics. 2010;42(1):33–38.

38. Carroll AM, Cheng R, Collie-Duguid ESR, Meharg C, Scholz ME, Fiering S, et al. Fine-mapping of genes determining extrafusal fiber properties in murine soleus muscle. Physiol Genomics. 2017;49(3):141–150.

39. Gonzales NM, Seo J, Hernandez-Cordero AI, Pierre CLS, Gregory JS, Distler MG, et al. Genome wide association study of behavioral, physiological and gene expression traits in a multigenerational mouse intercross. bioRxiv. 2017 Dec 8;230920.

40. Parker CC, Cheng R, Sokoloff G, Lim JE, Skol AD, Abney M, et al. Fine-mapping alleles for body weight in LG/J × SM/J F<Subscript>2</Subscript> and F<Subscript>34</Subscript> advanced intercross lines. Mamm Genome. 2011 Oct 1;22(9–10):563.

41. Bartnikas TB, Parker CC, Cheng R, Campagna DR, Lim JE, Palmer AA, et al. QTLs for murine red blood cell parameters in LG/J and SM/J F2 and advanced intercross lines. Mamm Genome. 2012 Jun;23(5–6):356–66.

42. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS One. 2011;6(5):e19379.

43. Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet. 2011;12(7):499.

632 44. Fitzpatrick CJ, Gopalakrishnan S, Cogan ES, Yager LM, Meyer PJ, Lovic V, et al. Variation in the form
633 of Pavlovian conditioned approach behavior among outbred male Sprague-Dawley rats from
634 different vendors and colonies: sign-tracking vs. goal-tracking. PloS One. 2013;8(10):e75042.

635 45. Toker L, Feng M, Pavlidis P. Whose sample is it anyway? Widespread misannotation of samples in
636 transcriptomics studies. F1000Research [Internet]. 2016 Sep 30 [cited 2018 Jul 11];5. Available
637 from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5034794/

638 46. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat
639 Genet. 2012;44(7):821.

640 47. Listgarten J, Lippert C, Kadie CM, Davidson RI, Eskin E, Heckerman D. Improved linear mixed
641 models for genome-wide association studies. Nat Methods. 2012;9(6):525.

642 48. Cheng R, Parker CC, Abney M, Palmer AA. Practical considerations regarding the use of genotype
643 and pedigree data to model relatedness in the context of genome-wide association studies. G3
644 Genes Genomes Genet. 2013;g3–113.

645 49. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the application of
646 mixed-model association methods. Nat Genet. 2014;46(2):100.

647 50. Joo JWJ, Hormozdiari F, Han B, Eskin E. Multiple testing correction in linear mixed models. Genome
648 Biol. 2016;17(1):62.

649 51. Han B, Kang HM, Eskin E. Rapid and accurate multiple testing correction and power estimation for
650 millions of correlated markers. PLoS Genet. 2009;5(4):e1000456.

651 52. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for
652 whole-genome association and population-based linkage analyses. Am J Hum Genet.
653 2007;81(3):559–575.

654 53. The Wellcome Trust Case Control Consortium, Maller JB, McVean G, Byrnes J, Vukcevic D, Palin K,
655 et al. Bayesian refinement of association signals for 14 loci in 3 common diseases. Nat Genet. 2012
656 Dec;44(12):1294–301.

657 54. Nikolskiy I, Conrad DF, Chun S, Fay JC, Cheverud JM, Lawson HA. Using whole-genome sequences
658 of the LG/J and SM/J inbred mouse strains to prioritize quantitative trait genes and nucleotides.
659 BMC Genomics. 2015;16(1):415.

660 55. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for
661 whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet.
662 2007;81(5):1084–1097.

663 56. Yang J, Zeng J, Goddard ME, Wray NR, Visscher PM. Concepts, estimation and interpretation of
664 SNP-based heritability. Nat Genet. 2017 Sep;49(9):1304–10.

665 57. Zöllner S, Pritchard JK. Overcoming the Winner's Curse: Estimating Penetrance Parameters from
666 Case-Control Data. Am J Hum Genet. 2007 Apr 1;80(4):605–15.

667 58. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis.
668 Am J Hum Genet. 2011 Jan 7;88(1):76–82.

669 59. Gardenghi S, Marongiu MF, Ramos P, Guy E, Breda L, Chadburn A, et al. Ineffective erythropoiesis
670 in β-thalassemia is characterized by increased iron absorption mediated by down-regulation of
671 hepcidin and up-regulation of ferroportin. Blood. 2007;109(11):5027–5035.

672 60. Graziano JH, Grady RW, Cerami A. The identification of 2, 3-dihydroxybenzoic acid as a potentially
673 useful iron-chelating drug. J Pharmacol Exp Ther. 1974;190(3):570–575.

674 61. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: analysis of next generation sequencing data.
675 BMC Bioinformatics. 2014;15(1):356.

676 62. Cheng R, Abney M, Palmer AA, Skol AD. QTLRel: an R package for genome-wide association studies
677 in which relatedness is a concern. BMC Genet. 2011 Jul 27;12:66.

678 63. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional
679 visualization of genome-wide association scan results. Bioinformatics. 2010;26(18):2336–2337.

680

Main figure legends

Fig 1. Minor allele frequency distributions for F_{34} array, F_{34} GBS, F_{39}-F_{43} GBS, and F_{34} and F_{39}-F_{43} GBS SNP sets. MAF distributions are highly comparable between AIL generations.

Fig 2. Significant loci on chromosome 17 for open field, distance traveled in periphery in F_{34} AIL. As exemplified in this pair of LocusZoom plots, GBS SNPs defined the boundaries of the loci much more precisely than array SNPs. GBS SNPs that are in high LD (r_2 > 0.8, red dots) with lead SNP chr17:27130383 resides between 27 ~ 28.3 Mb. In contrast, too few SNPs are present in the array plot to draw any definitive conclusion about the boundaries or LD pattern in this region. Purple track shows the credible set interval. LocusZoom plots for all loci identified in this paper are in Fig S7.

Fig 3. Chip-heritability estimates in F_{34} and $F_{39\text{-}43}$ AILs. All heritability estimates are highly significant (p < 1.0×10-05; see S7 Table).

Fig 4. Manhattan plots comparing F_{34} GBS, $F_{39\text{-}43}$ GBS, and mega-analysis on locomotor day 1 test using 57,170 shared SNPs in all AIL generations. Mega-analysis identified a locus on chromosome 10 (chr10.104988207) that was not detected in the F_{34} or F_{39}-F_{43} alone, suggesting that mega-analysis enhanced power to detect some loci.

Supporting information

S1 Fig. Autosomal heterozygosity distribution in F_{34}, $F_{39\text{-}43}$ AILs. Animals with excessive or insufficient heterozygosity (3 *s.d.* away from mean) were removed from further analysis. As controls, we have sequenced two F_2s of LG and SM, four LG mice and four SM mice (see annotated data points with 1 and 0 heterozygosity).

S2 Fig. Kinship coefficients in F_{34} and $F_{39\text{-}43}$ AILs calculated from pedigree against genetic relatedness matrix calculated using IBDLD [49]. Each circle represents a pair of animals, which their genetic kinship relatedness on the x-axis and pedigree kinship relatedness on the y-axis. Color signifies relatedness based on AIIL pedigree. Blue circles represent identical twins, red full siblings, yellow parent-offspring pairs, grey other relationships. Seven animal pairs that deviate from the pedigree relationship clusters were excluded (see black arrows).

S3 Fig. Heatmap showing F_{34} array and F_{34} GBS genotype concordance in percentages, using 66 shared SNPs. "A" codes for the LG/J allele, and "B" codes for the SM/J allele. "AA" genotype concordance between array and GBS is 24.54%, "AB" 43.23%, "BB" 27.60%.

S4 Fig. LD decay in F_{34} array, F_{34} GBS, $F_{39\text{-}43}$ GBS, and F_{34} and $F_{39\text{-}43}$ GBS SNP sets.

S5 Fig. Manhattan plots comparing 4,593 F_{34} array, 60.3K F_{34} GBS, 4.3M imputed F_{34} array, and 4.1M imputed F_{34} GBS (N=428) SNPs on day 2 locomotor activity. Adjusted significance

thresholds for imputed array and GBS SNPs were estimated using LD pruned SNPs (r2=0.1, window size=20kb; PLINK v1.9). Notice that even though the imputed sets have more SNPs (the two right panels), they are frequently blocks of many SNPs with almost identical position and LD=1, therefore making it hard to visualize the additional SNPs.

S6 Fig. SNP heritability using F_{34} GBS and F_{34} array SNPs (slope=1).

S7 Fig. LocusZoom for F_{34} array, F_{34} GBS, $F_{39\text{-}43}$ GBS, and mega-analysis QTLs. Purple track shows the credible set interval (r^2 threshold = 0.8, posterior probability threshold = 0.99).

S8 Fig. Power simulations for discovery SNPs in the replication set. Power was simulated at both the genome-wide significance level for the cohort and the nominal P value of 0.05. Each data point represents the estimated power at the simulated beta (plotted as Z score). The vertical dashed line in orange indicates the effect size of the discovery SNP.

S1 Table. List of phenotypes used in GWAS.

S2 Table. SNP and individual QC filter table. Numbers of animals and SNPs remained after each step of filtering are shown per GBS SNP set.

S3 Table. Effect of PLINK v1.9 clump-based pruning parameters on number of independent SNPs remained. At all r_2 values examined, a sliding window size of 12150kb was the first smallest window that yield the most stringent number of clumped SNPs in both array and GBS GWAS.

S4 Table. Lead QTL in F_{34} GBS and F_{34} array GWAS studies across phenotypes. Significant SNPs are clumped using parameters r_2=0.1, 12150kb.

S5 Table. F_{34} GBS and array SNP heritability estimates.

S6 Table. F_{34} and $F_{39\text{-}43}$ genetic correlations in locomotor activity, coat color, and body weight.

S7 Table. SNP-heritability comparison between F_{34} and $F_{39\text{-}43}$ GBS.

S8 Table. Lead QTL in $F_{39\text{-}43}$ N=600 GBS GWAS studies across phenotypes. Significant SNPs are clumped using parameters r_2=0.1, 12150kb.

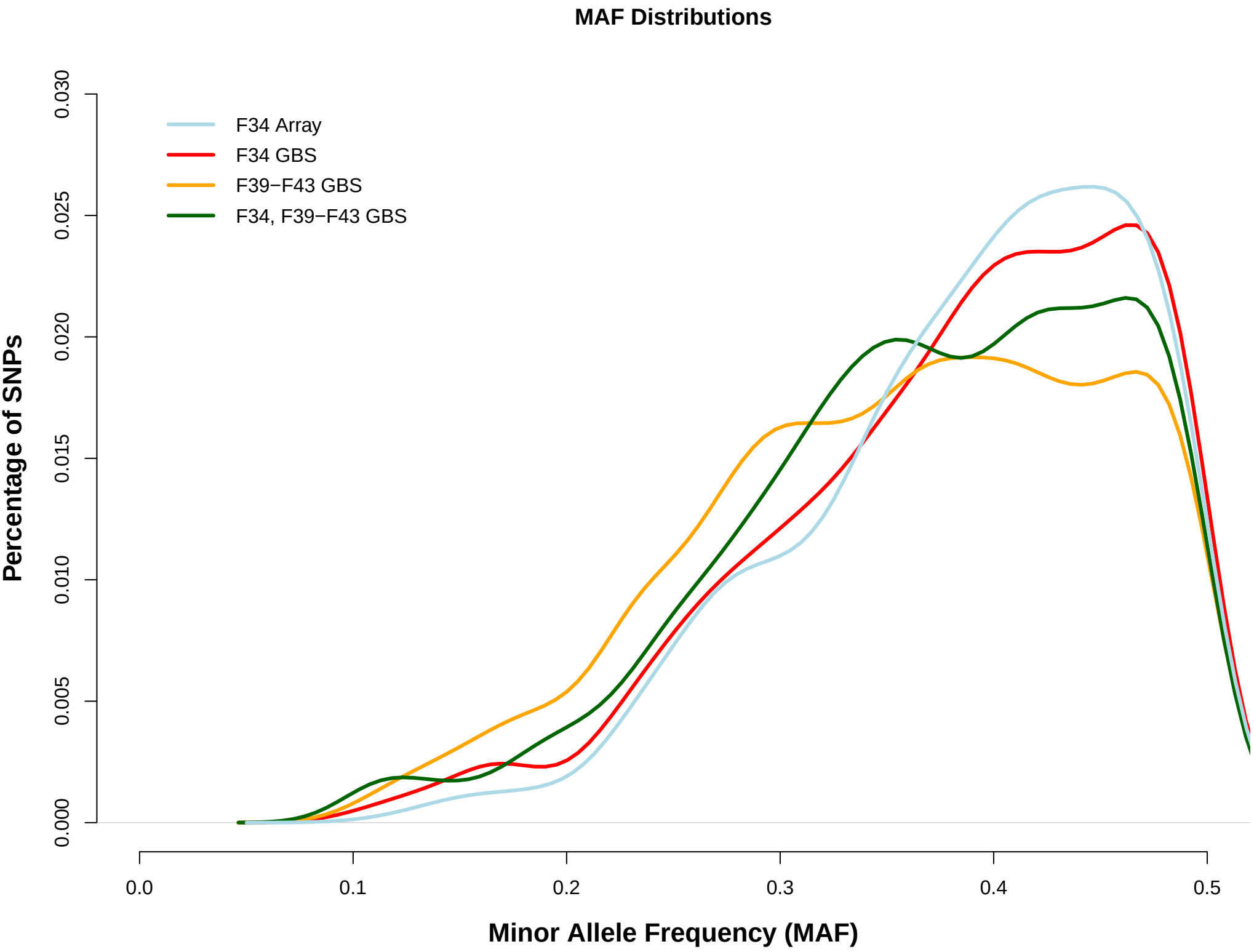
S9 Table. Lead QTL in F_{34} and $F_{39\text{-}43}$ (N=1028) mega-analysis across phenotypes. Significant SNPs are clumped using parameters r_2=0.1, 12150kb.

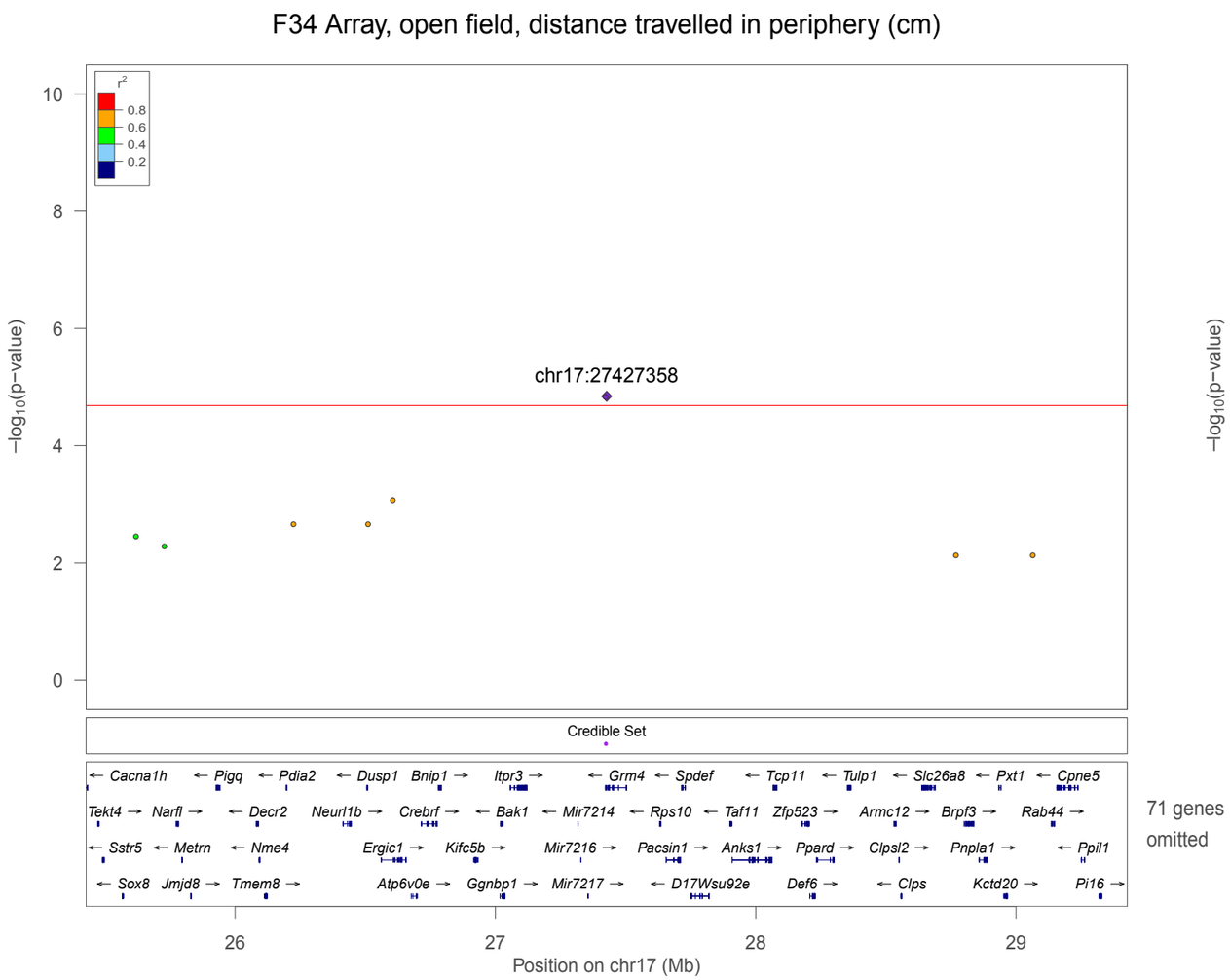
S10 Table. SNPs in F_{34} GBS set with HWE p-values close to 1.0×10-6 cutoff threshold. These SNPs are removed from QTL summary tables.

S11 Table. SNPs in F_{34} and $F_{39\text{-}43}$ mega-analysis GBS set with HWE p values close to 1.0×10-6 cutoff threshold. These SNPs are removed from QTL summary tables.

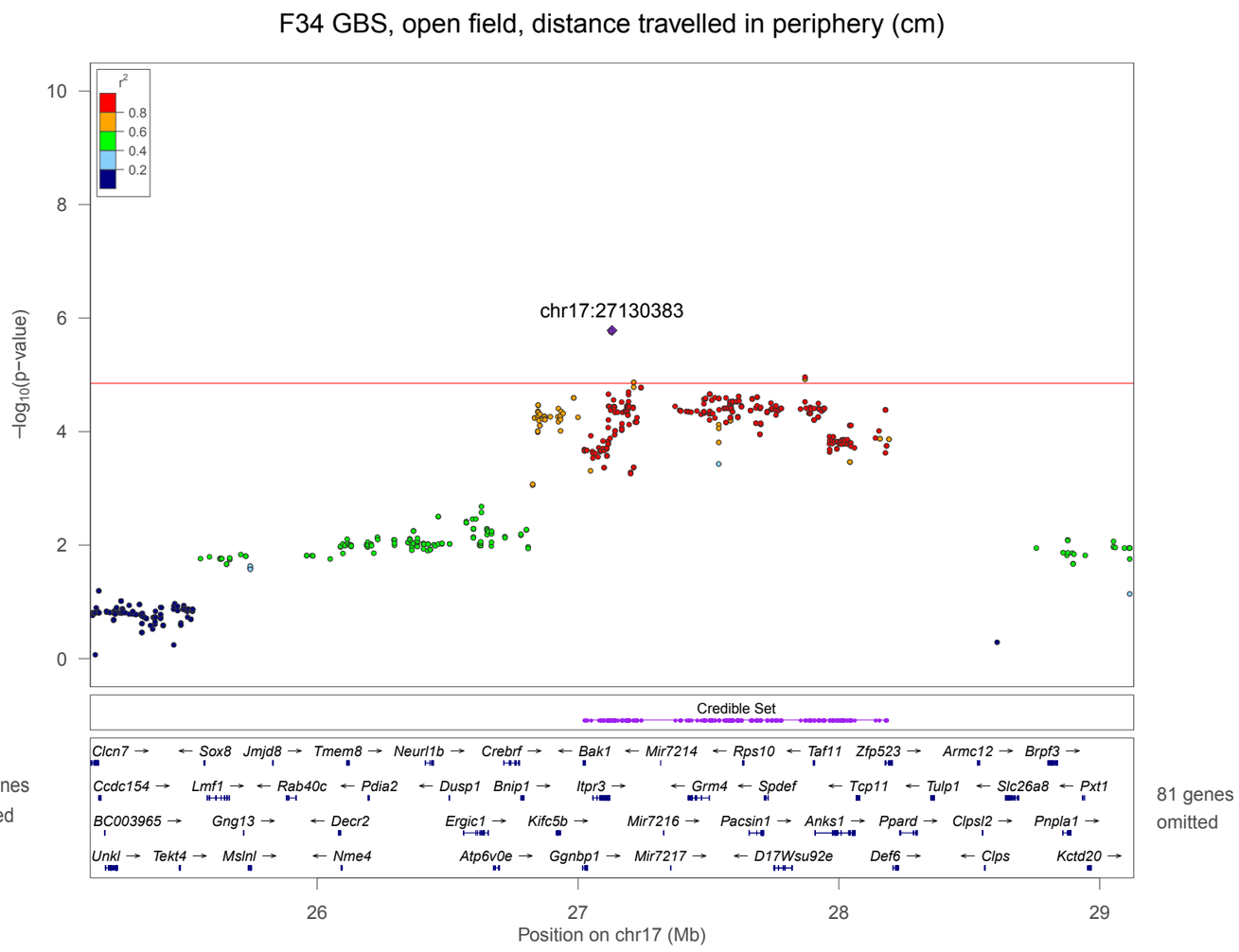
S12 Table. Adjusted significance threshold for each SNP set and GWAS cohort.

770

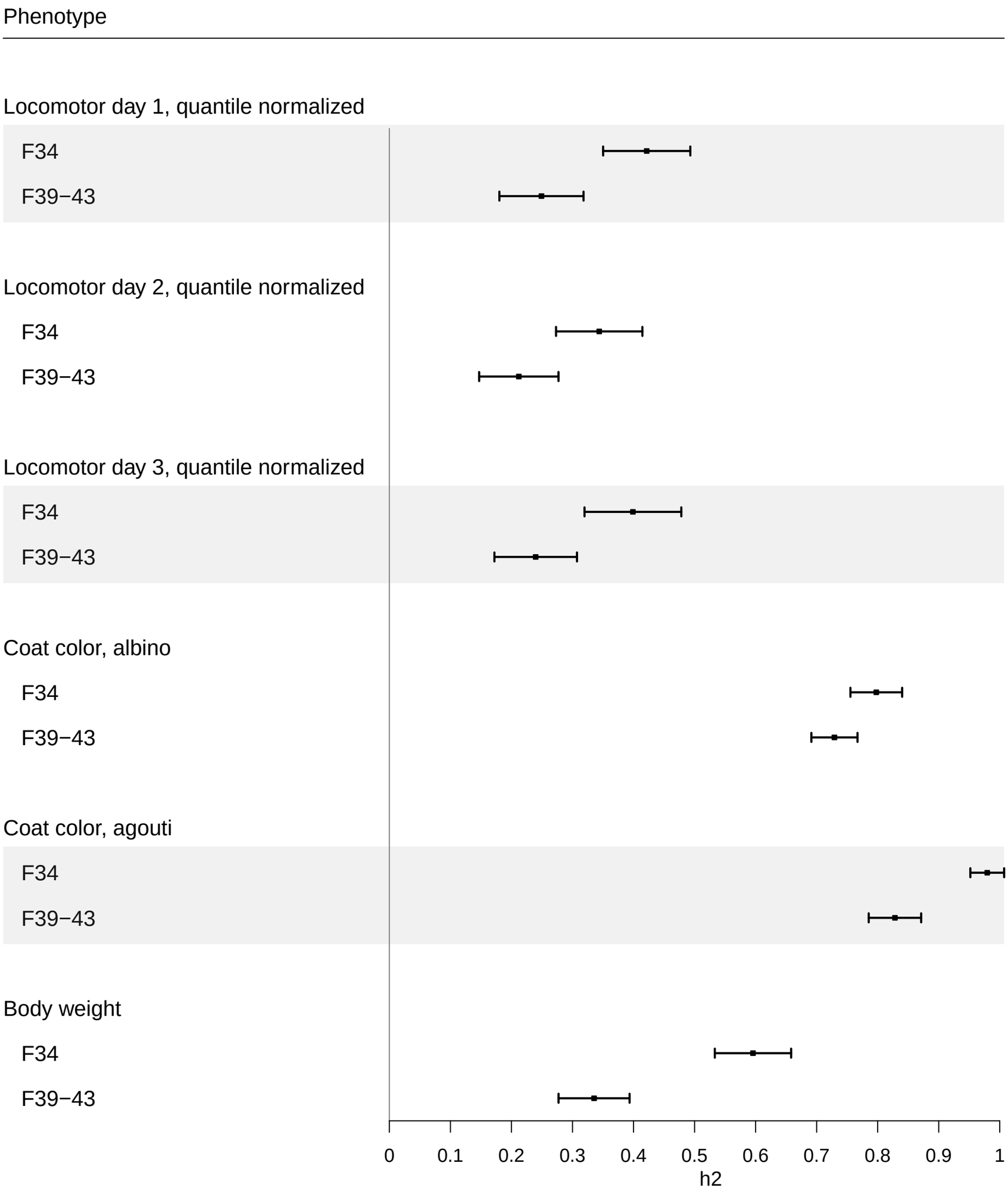

MAF Distributions
F34 Array
F34 GBS
F39−F43 GBS
F34, F39−F43 GBS
Percentage of SNPs
0.000
0.005
0.010
0.015
0.020
0.025
0.030
0.0
0.1
0.2
0.3
0.4
0.5
Minor Allele Frequency (MAF)

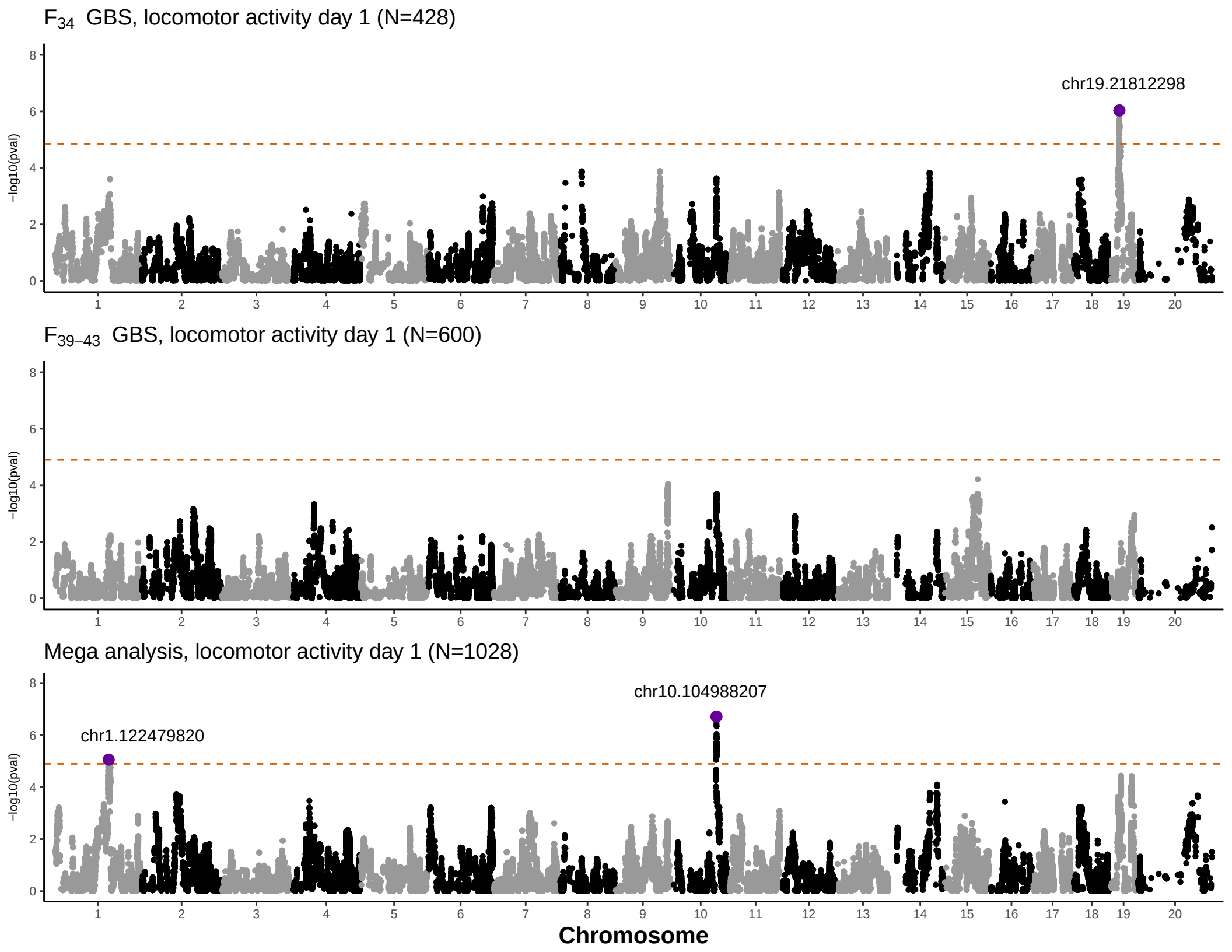


71 genes omitted



81 genes omitted


Phenotype
Locomotor day 1, quantile normalized
F34
F39−43
Locomotor day 2, quantile normalized
F34
F39−43
Locomotor day 3, quantile normalized
F34
F39−43
Coat color, albino
F34
F39−43
Coat color, agouti
F34
F39−43
Body weight
F34
F39−43
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
h2


F34 GBS, locomotor activity day 1 (N=428)
chr19.21812298
−log10(pval)
F39–43 GBS, locomotor activity day 1 (N=600)
−log10(pval)
Mega analysis, locomotor activity day 1 (N=1028)
chr10.104988207
chr1.122479820
−log10(pval)
Chromosome