Genetic characterization of outbred Sprague Dawley rats and utility for genome-wide association studies

Alexander F. Gileta[1,3], Christopher J. Fitzpatrick[2], Apurva S. Chitre[3], Celine L. St. Pierre[1,4], Elizabeth V. Joyce[4], Rachael J. Maguire[4], Africa M. McLeod[4], Natalia M. Gonzales[1], April E. Williams[3], Jonathan D. Morrow, Terry E. Robinson[6], Shelly B. Flagel[2,5,7], Abraham A. Palmer[3, 8*]

[1] Department of Human Genetics, University of Chicago, Chicago, Illinois, United States of America

[2] Neuroscience Graduate Program, University of Michigan, Ann Arbor, Michigan, United States of America

[3] Department of Psychiatry, University of California, San Diego, La Jolla, California, United States of America

[4] Biological Sciences Collegiate Division, University of Chicago, Chicago, Illinois, United States of America

[5] Department of Psychiatry, University of Michigan, Ann Arbor, Michigan, United States of America

[6] Department of Psychology and Neuroscience Program, University of Michigan, Ann Arbor, Michigan, United States of America

[7] Molecular and Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, Michigan, United States of America

[8] Institute for Genomic Medicine, University of California, San Diego, La Jolla, California, United States of America

* Corresponding author

Email: aapalmer@ucsd.edu

Abstract

Sprague Dawley (**SD**) rats are one of the most commonly used outbred rat strains. Despite this, the genetic characteristics of SD are poorly understood. We collected behavioral data from 4,625 SD rats acquired predominantly from two commercial vendors, Charles River Laboratories and Harlan Sprague Dawley Inc. Using double-digest genotyping-by-sequencing (**ddGBS**), we obtained dense, high-quality genotypes at 234,887 SNPs across 4,061 rats. This genetic data allowed us to characterize the variation present in Charles River vs. Harlan SD rats. We found that the two populations are highly diverged ($F_{ST} > 0.4$). We also used these data to perform a genome-wide association study (**GWAS**) of Pavlovian conditioned approach (**PavCA**), which assesses the propensity for rats to attribute motivational value to discrete, reward-associated cues. Due to the genetic divergence between rats from Charles River and Harlan, we performed two separate GWAS by fitting a linear mixed model that accounted for within vendor population structure and using meta-analysis to jointly analyze the two studies. We identified 18 independent loci that were significantly associated with one or more metrics used to describe PavCA; we also identified 3 loci that were body weight, which was only measured in a subset of the rats. The genetic characterization of SD rats is a valuable resource for the rat community that can be used to inform future study design.

Author Summary

Outbred Sprague Dawley rats are among the most commonly used rats for neuroscience, physiology and pharmacological research. SD rats are sold by several commercial vendors, including Charles River Laboratories and Harlan Sprague Dawley Inc. (now Envigo). Despite their wide spread use, little is known about the genetic diversity of SD. We genotyped more than 4,000

SD rats, which we used to characterize genetic differences between SD rats from Charles River Laboratories and Harlan. Our analysis revealed that the two SD colonies are highly divergent. We also performed a genome-wide association study (**GWAS**) for Pavlovian conditioned approach (**PavCA**), which assesses the propensity for rats to attribute motivational value to discrete, reward-associated cues. Our results demonstrate that, despite sharing an identical name, SD rats are obtained from different vendors are genetically very different. We conclude that results obtained using SD rats should not be presented without also carefully noting the vendor.

Introduction

Rats are among the most commonly used organisms for experimental psychology and biomedical research. Whereas research using mice makes extensive use of inbred strains, in rats, it is more common to use commercially available *outbred* populations. Among the commercially available outbred rat populations, the Sprague Dawley strain (SD) is one of the most widely used. SD rats are distributed by several vendors. Each vendor has multiple breeding locations, and each breeding location has one or more barriers in which the rats are housed. Prior studies have identified numerous physiological differences between SD rats obtained from different vendors [1–5]. Despite these observations, many researchers appear to assume that SD rats obtained from different vendors or barrier facilities are largely interchangeable. There has been little research into the genetic diversity and population structure that exists among commercially available outbred rats [6]. Prior rat genetic studies have used F_2 and more complex, multi-parental crosses of inbred strains for QTL mapping and GWAS [7–9]; however, we are not aware of any such studies that have employed commercially available outbred rats. Recently, we and others have demonstrated the potential benefits and challenges associated with the use of commercially

available outbred mice for GWAS [10,11], suggesting that similar studies in rats might also be of value.

SD rats originated in 1925 at the Sprague-Dawley Animal Company (Madison, WI), where they were created by a cross between a hooded male hybrid of unknown origin and an albino Wistar female [12]. In 1950, Charles River Inc. began to distribute SD rats. In 1980, Harlan Inc. (now Envigo, Inc.) began to distribute SD rats after their acquisition of Sprague-Dawley, Inc. [13]. In 1992, Charles River reestablished a foundation colony of SD rats, using 100 breeder pairs from various existing colonies [14]. The resulting litters were used to populate SD colonies globally and have since been bred using a mating system that minimizes inbreeding. Every 3 years, Charles River replaces 25% of their male breeders in each production colony with rats from a single foundation colony. Charles River also replaces 5% of their foundation colony breeding pairs with rats from the production colonies on a yearly basis. These practices are designed to reduce genetic drift between the production colonies [15]. Harlan also reports using a rotational breeding system to limit inbreeding; however, more detailed information is not publicly available. Since Harlan's acquisition by Envigo, the process has become more transparent [16]. Envigo follows a Poiley rotational breeding scheme [17], whereby animals are cycled through different sections of the colony with each generation, reducing genetic drift and inbreeding.

Here we used SD rats from multiple vendors, breeding locations, and barrier facilities to elucidate the genetic background of SD and to perform a GWAS of components of a complex behavior. DNA samples were obtained from rats used in multiple studies as part of an unrelated Program Project grant (P01DA031656) concerned with individual variation in the propensity to

attribute incentive value to food and drug cues [18,19]. All rats were first screened for Pavlovian conditioned approach behavior (PavCA) [20], which provides one index of the degree to which a reward cue has been attributed with incentive salience. Although the genetic analyses reported here were not part of the original design, we took advantage of the opportunities afforded by that large, behavioral study. We extracted genomic DNA from available tissue samples and then used double digest genotype-by-sequencing (ddGBS) to obtain dense genotypes for 4,625 SD rats. We used these genotypes to first genetically characterize different populations of SD, and then in conjunction with behavioral data from PavCA, to perform the largest rodent GWAS to date. Because most of the rats were obtained from two vendors (Harlan and Charles River) we performed two separate GWAS and combined the results using meta-analysis. Our results provide insights about the population structure and suitability of SD for GWAS, and also explore specific loci associated with Pavlovian conditioned approach.

Results

Phenotype

Our final dataset consisted of 4,061 genotyped individuals that were also phenotyped for Pavlovian conditioned approach; 2,281 from Harlan and 1,780 from Charles River, from 5 and 4 different breeding locations, respectively. As noted previously [20], we found that the metrics used to describe performance in PavCA are highly correlated (S1 Fig). Additionally, several of the base and composite PavCA metrics had tail-heavy distributions due to biased responding in sign- and goal-tracking from the animals during the testing periods (S2 Fig). For this reason, we chose to quantile normalize all measurements prior to mapping. The PavCA index score [20], which has been used previously to categorize rats into sign-trackers (**STs**), goal-trackers (**GTs**), and

intermediate responders (**IRs**), showed the expected divergence and stabilization for STs and GTs over the five days of testing (Fig 1A). We therefore focused initial analyses on days 4 and 5 of testing and the average of days 4 and 5 as previously reported [20].

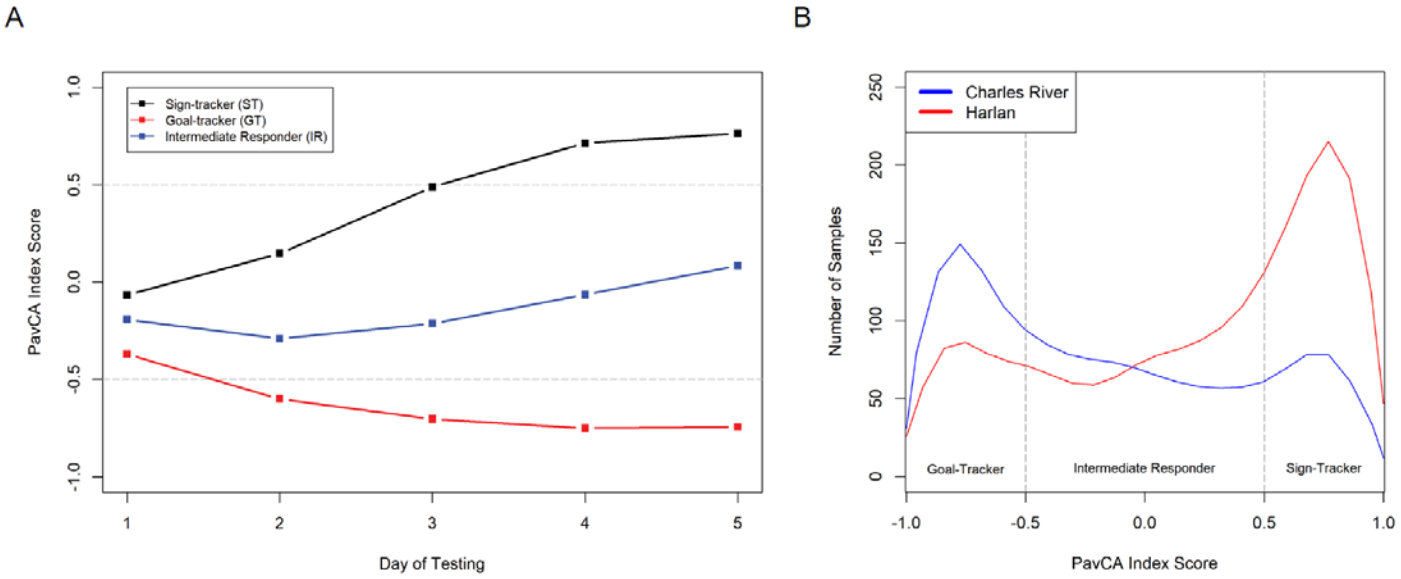
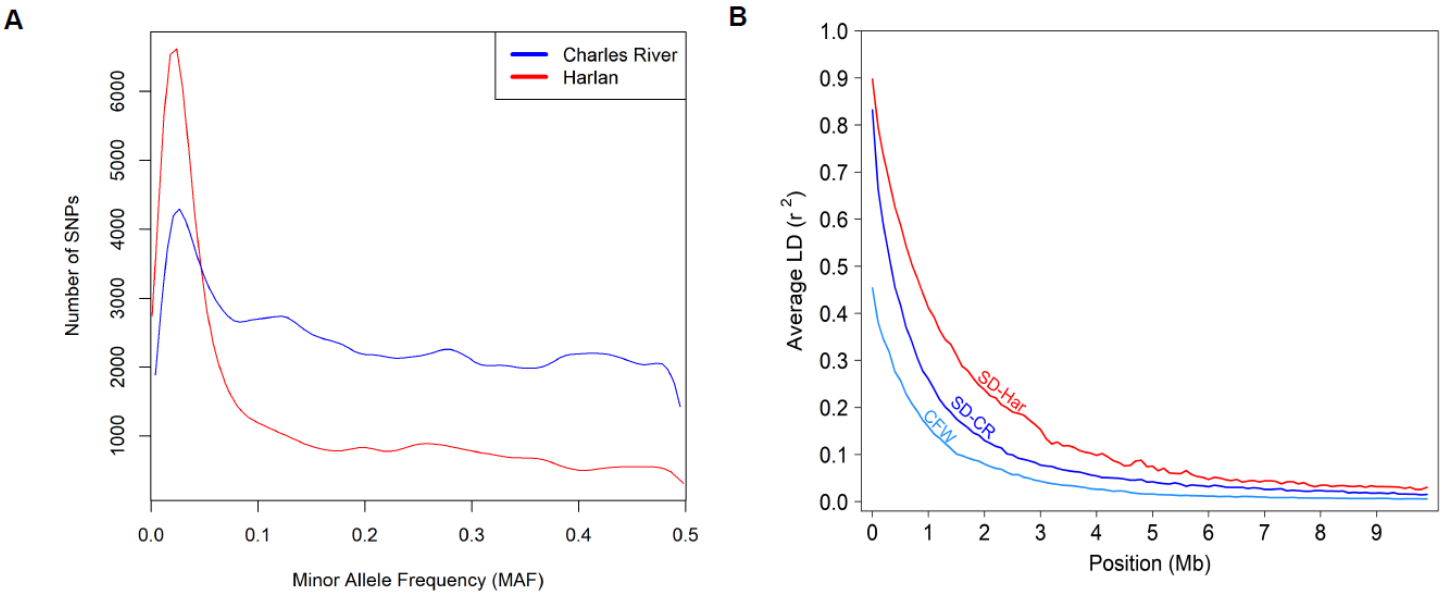


Fig 1. PavCA index score progression across days and distribution between Charles River and Harlan.
(A) Samples were classified as ST, GT, or IR based on their average day 4 and 5 PavCA index score. PavCA index scores for each day of training (1-5) were averaged across all STs, GTs, and IRs. (B) Density curves of average day 4 and 5 PavCA index score in Harlan (n=2,281) vs. Charles River (n=1,780) SD rats.

Charles River and Harlan rats had significantly different average PavCA index score distributions (Fig 1B; Welch's 2 -sample t-test, p-value < 2.2×10^{-16}), with Charles River rats biased towards goal-tracking and Harlan rats biased towards sign-tracking. We divided the samples further into the breeding locations that the rats originated from (S1 Table). The differences in PavCA index score between breeding locations within vendor were smaller, but still significant (S3 Fig) [6].

<u>Genotyping and genetic characterization of SD rats</u>

We identified more single nucleotide polymorphisms (SNPs) for the rats from Charles River compared to Harlan (214,309 vs 114,568; S2 Table). Fig 2C compares the distribution of SNPs across each chromosome for each vendor. There were some regions where Harlan had few SNPs, but Charles River had many, and other regions where both Harlan and Charles River had few SNPs. We also observed a large difference in the minor allele frequency (MAF) distributions for Charles River and Harlan (Fig 2A), with Charles River having a far greater proportion of SNPs with high MAF (>0.05). This observation could reflect the fact that Charles River has adhered to their International Genetics Standardization Protocol for 25+ years, whereas Harlan appears to have focused on maintaining diversity within breeding colonies and may have allowed for a moderate degree of drift between them. After combining the two SNP sets, we identified a total of 234,887 unique, bi-allelic SNPs. Using the 381 duplicate samples to evaluate our genotyping accuracy, we calculated our genotype discordance rate to be 0.85% (S3 Table).



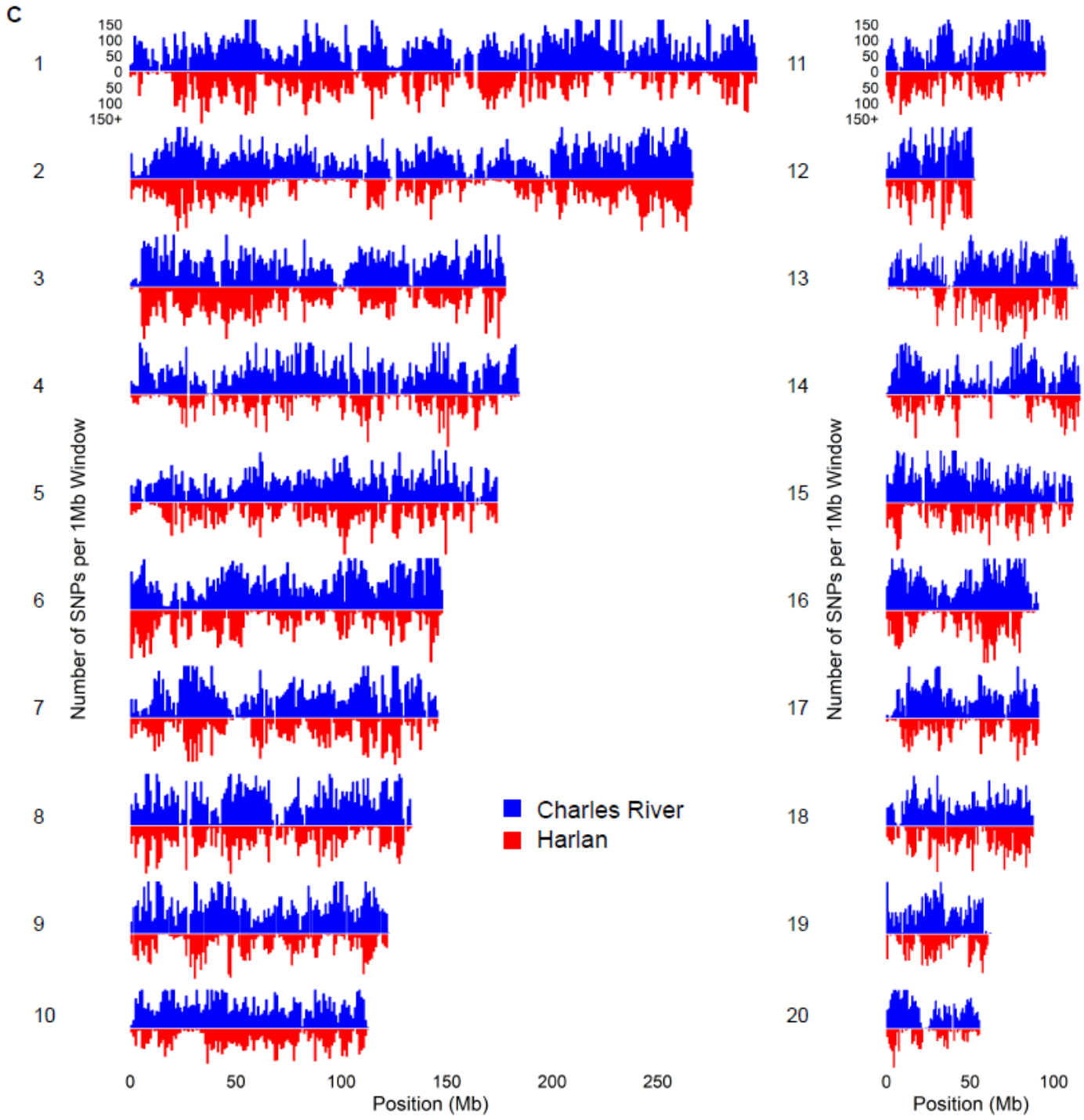


Fig 2. Genetic architecture of SD rats from Charles River vs. Harlan.
(A) Density curves of minor allele frequencies for 214,309 SNPs in Charles River and 114,568 SNPs in Harlan, after removing SNPs with MAF < 0.01. (B) Linkage disequilibrium decay rates in SD rats from both vendors and outbred Swiss Webster (CFW) mice. (C) Filtered SNP density per 1Mb window in Charles River vs. Harlan samples for all 20 autosomes.

In addition to the variants described above, which were obtained using ANGSD/Beagle, we also used STITCH as an alternative genotyping pipeline [21]. This approach identified a larger total number of variants (S2 Table); however, after pruning SNPs with high linkage disequilibrium (LD; $r^2 \geq 0.95$), STITCH produced fewer SNPs compared to ANGSD/Beagle. Preliminary GWAS suggested the two SNP sets produced broadly similar results. Ultimately, we chose to use the ANGSD/Beagle variants for all subsequent analyses.

To examine the levels of linkage disequilibrium (LD) in Charles River and Harlan, we constructed LD decay curves (Fig 2B). These curves show the rate at which LD between two genetic loci dissipates as a function of the distance between the loci. Harlan rats had more extensive LD compared to Charles River. For contrast, we included the decay curve for the Swiss Webster (CFW) line, a commercially available outbred mouse population that has been successfully used for GWAS in the past [10,11].

To further investigate the observed genetic divergence between Charles River and Harlan, we performed principal component analysis (PCA) on a set of 4,502 LD-pruned ($r^2 < 0.5$) SNPs with MAF > 0.05 across both populations (Fig 3B). The first PC corresponded to vendor (Charles River or Harlan) and accounted for ~33.7% of the variance. The second PC accounted for just ~0.9% of the variance and reflected population structure within Harlan SD rats. To investigate within vendor structure further (Fig 3A), we performed PCA on the samples from each vendor separately using the same set of SNPs. Fig 3C and 3D show evidence of substructure at both the level of breeding location (i.e. the city) and barrier facility (i.e., the segregated breeding areas

within the building). Interestingly, rats from some barrier facilities showed greater differentiation from barrier facilities within the same breeding location than barriers in other locations.

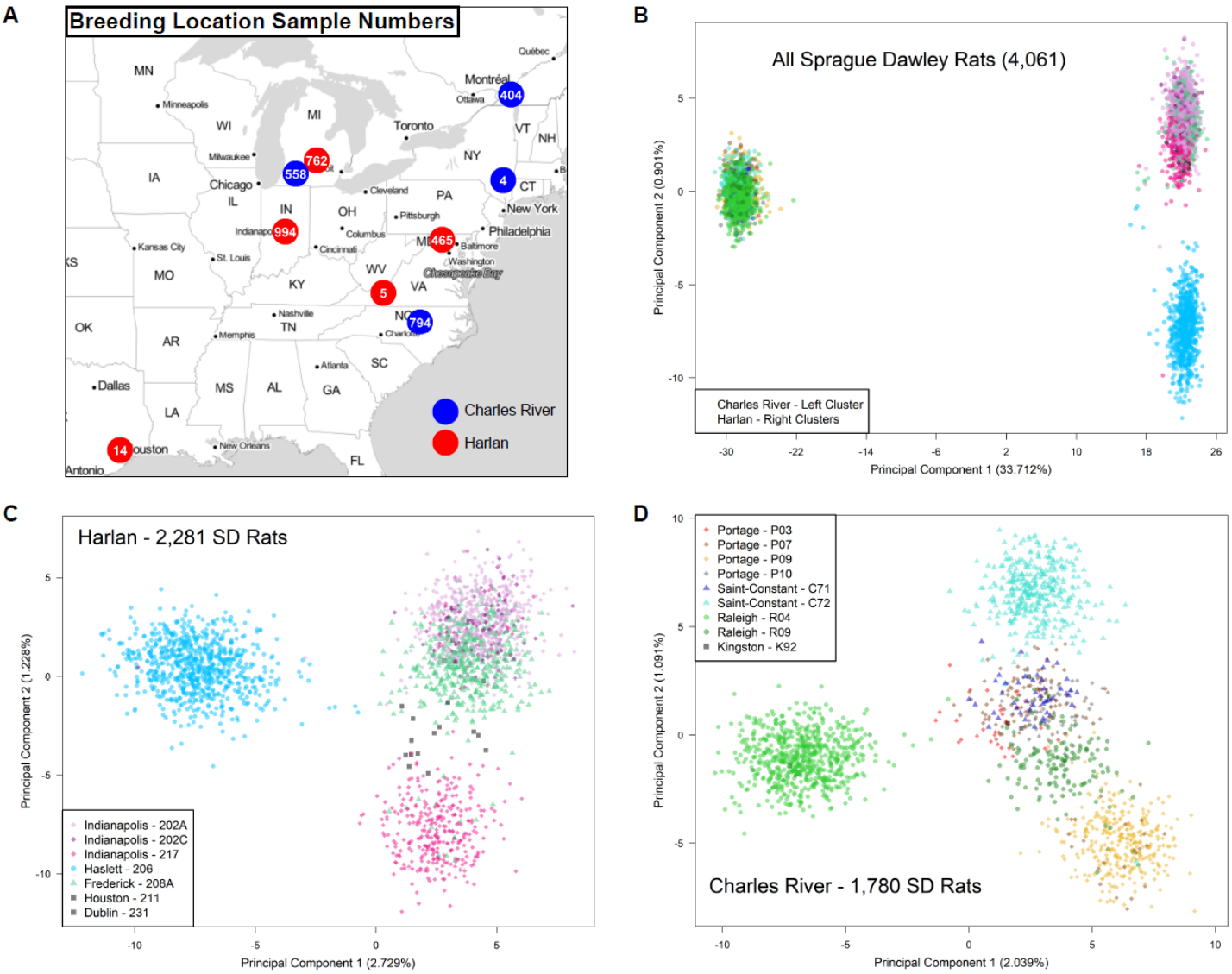


Fig 3. SD population structure and comparison of linkage disequilibrium decay rates
(A) Map of the nine vendor breeding locations and the number of SD rats obtained from each location. (B) A summary of the genetic data from all 4,061 SD rats based on principal components 1 and 2 from PCA. Each point represents a sample. The left cluster is composed of samples from Charles River and the right clusters are composed of samples from Harlan. (C-D) Repeated PCA analyses on subsets of the samples from Harlan and Charles River, colored by barrier facility of origin.

The fixation index (F_{ST}) is a statistic widely employed by population geneticists to measure the level of structure in populations [22]. It is calculated using the variance in allele frequencies

among populations; values closer to 0 indicate genetic homogeneity, and values closer to 1 indicate genetic differentiation. We calculated the F_{ST} between all Harlan and Charles River breeding locations (Table 1) and barrier facilities (S4 Table) with a sufficiently large number of samples (N > 30). F_{ST} values between vendors were ~0.435, which is much higher than corresponding values for major human lineages [23], whereas the values for different breeding locations within a vendor were substantially lower (S4 Table).

Table 1. Pairwise F_{ST} statistics for Harlan and Charles River breeding locations.

	Harlan			Charles River		
Breeding Locations	Frederick, MD	Haslett, MI	Indianapolis, IN	Portage, MI	Raleigh, NC	St. Constant, CAN
Frederick, MD	0	0.028	0.026	0.437	0.434	0.440
Haslett, MI		0	0.009	0.435	0.432	0.438
Indianapolis, IN			0	0.432	0.429	0.434
Portage, MI				0	0.016	0.014
Raleigh, NC					0	0.018
St. Constant, CAN						0

We speculated that some of the rats might share close genetic relationships with one another. We used *plink 1.9* [24,25] to estimate the pairwise proportions of identity-by-descent (IBD; Panels A and C in S4 Fig), which showed that while most rats were only distantly related, a subset shared closer familial relationships. We removed several rats that showed high levels of relatedness with many other samples (presumably due to technical error), as well as any with unreasonably high levels of IBD (S5 Table; Panels B and D in S4 Fig).

SNP heritability and genome-wide association analyses

Although evidence from selective breeding studies has suggested that behavior in the Pavlovian conditioned approach procedure is heritable [26], we are not aware of any specific

heritability estimates using inbred strains or outbred populations. We used GCTA to calculate the proportion of the variance in the base and composite PavCA metrics that could be explained by the union set of variants from Harlan and Charles River (S6 Table). The SNP heritability estimates for all PavCA metrics in Harlan ranged from ~4-11%, whereas they were ~4-21% for Charles River; on average the estimated heritability was about ~1.9-fold greater in Charles River. Importantly, some of the highest heritabilities were for metrics used to designate sign-trackers vs. goal-trackers, such as average of day 4 and day 5 response bias, probability difference, and PavCA index score. However, even the heritability estimates from Charles River were lower than SNP heritability estimates for many other behavioral traits [9–11,27]. We also estimated the heritability of body weight, for which we had data in 957 rats from Charles River and 901 from Harlan. The estimates were much higher than for the behavioral metrics: 42.7% (s.e. = 0.070) for Charles River and ~63.2% (s.e. = 0.056) for Harlan.

Next, we performed GWAS for various metrics derived from Pavlovian conditioned approach separately for rats from Harlan and Charles River using GEMMA to fit a linear mixed model that allowed us to account for population structure. We also performed meta-analyses of the two populations using the subset of 93,990 SNPs that overlapped between our two filtered sets. S7 Table contains information on all loci that obtained permutation-derived genome-wide significance in any of these analyses. S1 File contains Manhattan plots for all analyses. Because the meta-analysis only contained the overlapping SNPs, we present 5 stacked Manhattan plots for each PavCA metric: Harlan and Charles River with all SNPs, Harlan and Charles River with only the ~94k overlapping SNPs, and the meta-analysis, which only uses the overlapping SNPs.

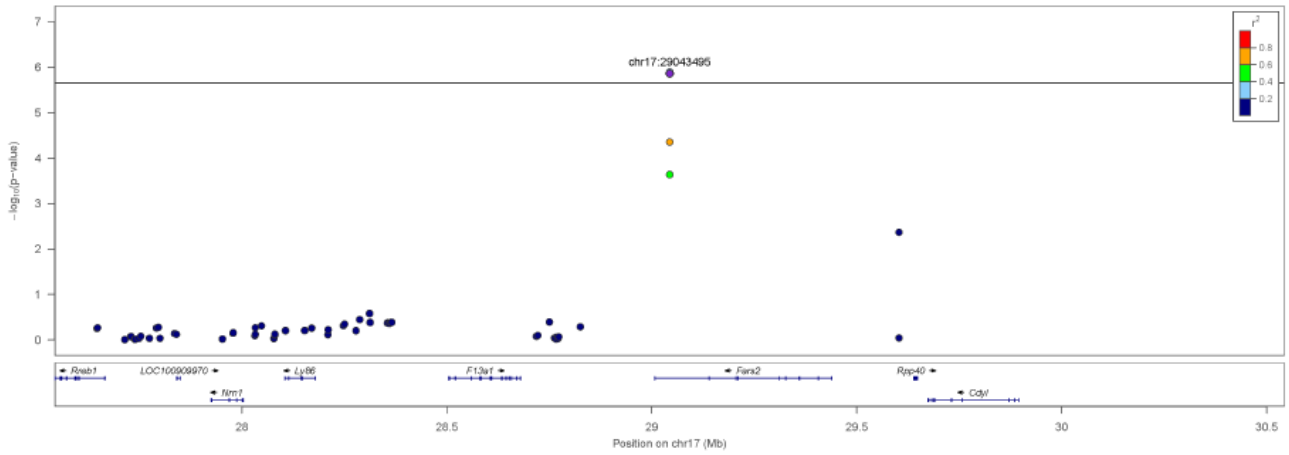
Previous work on Pavlovian conditioned approach in rats has focused on the average PavCA index score of days 4 and 5 to phenotype rats [20]. When analyzed as a quantitative trait,

we did not identify any genome-wide significant results for this metric at a threshold of $-\log_{10}(p) = 5.66$ for Harlan and $-\log_{10}(p) = 6.05$ for Charles River. We also coded PavCA index score as a binary trait using various thresholds to define cases (goal-trackers) and controls (sign-trackers); this approach identified two genome-wide significant associations. The first locus was unique to Harlan and present on chromosome 4 in an intron of *Cntn4* (Page 67 in S1 File; S5 Fig), which encodes a cell-adhesion molecule involved in synaptic signaling, neuronal network formation, and neuropsychiatric disorders such as addiction [28–30]. The second locus only reached genome-wide significance in the meta-analysis and resided on chromosome 17 in the intronic region of *Fars2* (Page 68 in S1 File; Fig 4B), a mitochondrial phenylalanyl-tRNA synthetase involved in oxidative phosphorylation and neuronal functioning [31–33].

A

Harlan 114k SNPs - Average Latency to Magazine Entry - Days 4 & 5 Average - n =2245
Harlan 94k Meta-SNPs - Average Latency to Magazine Entry - Days 4 & 5 Average - n =2245
Charles River 214k SNPs - Average Latency to Magazine Entry - Days 4 & 5 Average - n =1728
Charles River 94k Meta-SNPs - Average Latency to Magazine Entry - Days 4 & 5 Average - n =1728
Meta-analysis 94k SNPs - Average Latency to Magazine Entry - Days 4 & 5 Average - n =3973
-log10(P value)
Chromosomal Position

B


chr17:29043495
-log10(p-value)
Position on chr17 (Mb)
Rreb1
LOC100909970
Nm1
Ly86
F13a1
Fars2
Rpp40
Cdyl

242

Fig 4. Stacked Manhattan plots and a LocusZoom plot for day 4 and 5 average latency to magazine entry.

(A) The five Manhattan plots in descending order are: (1) GWAS in Harlan with the full set of 114k SNPs, (2) GWAS in Harlan with the overlapping set of 94k SNPs, (3) GWAS in Charles River with the full set of 214k SNPs, (4) GWAS in Charles River with the overlapping set of 94k SNPs, and (5) meta-analysis of the 94k overlapping SNPs. (B) LocusZoom plot of the genome-wide significant loci on chromosome 17 identified for day 4 and 5 average latency to magazine entry.

We then expanded our search to the average of days 4 and 5 for each of the 10 remaining metrics individually. We identified a genome-wide significant association for the number of head entries into the food magazine during the intertrial interval (i.e. in the absence of a conditioned stimulus; CS) in Harlan, but none of the other 9 metrics analyzed in Harlan and Charles River produced significant associations. However, the meta-analysis of the average day 4 and 5 metrics produced several additional genome-wide significant loci. There was a strong peak for the probability of a magazine entry during the CS interval on chromosome 7 (Page 42 in S1 File) that spanned ~1.2Mb and included the genes *Atxn10*, *Ppara*, and *Wnt7b*. Additionally, we replicated the *Fars2* association seen for the binary coding of the PavCA index score in the meta-analysis of the day 4 and 5 average latency to magazine entry during CS presentation (Fig 4A).

In an effort to further examine this large dataset, we also performed exploratory GWAS for all 11 metrics on days 1-5 for both Harlan, Charles River and the meta-analysis of the two. This large number of additional analyses (110 GWAS and 55 meta-analyses) produced only 5 additional genome-wide significant hits for Harlan, 1 for Charles River, and 5 from the meta-analyses. Because many of the metrics being tested are correlated, a Bonferroni correction would not be appropriate; however, the modest number of significant associations relative to the number of GWAS was not different from expectations under the null hypothesis (no true associations).

None of the meta-analyses results replicated results from the individual populations. The meta-analyses did show that there were associations that occurred for the same metrics across multiple days of testing. For example, magazine entries with the CS showed an association with the same SNP on chromosome 1 on days 3, 4, and 5 in the meta-analysis (Pages 21-24 in S1 File).

We sought to use PCA to analyze all 11 metrics across all 5 days separately for both Harlan and Charles River. A summary of the population-specific factor loadings and the percent variance explained by each factor for the first 5 PCs is presented in S8 Table. The percent of variation explained by each of these PCs in each population is shown in S6 Fig. Fig 5 shows the correlations between each metric and the first two PCs in each population, helping to visualize which of the 55 metrics loaded most strongly on PC1 and PC2. PC1 accounted for slightly over 50% of the explained variance and loaded metrics from days 4 and 5, supporting our overall approach to Pavlovian conditioned approach. PC2 predominantly loaded metrics from day 1 of training. The factor structure and percent variance explained were very similar between Harlan and Charles River, suggesting that PCA was an effective way to summarize these data. Since the first 3 PCs cumulatively accounted for more than 70% of the variance in both populations, we only considered these for mapping and performed parallel GWAS for Harlan and Charles River. We identified 3 genome-wide significant associations; 2 of them were for PC2 on chromosomes 1 and 17 and were seen exclusively in Harlan. The third came from the meta-analysis of PC1 and was located on chromosome 17 at the same locus as associations for 5 other metrics.

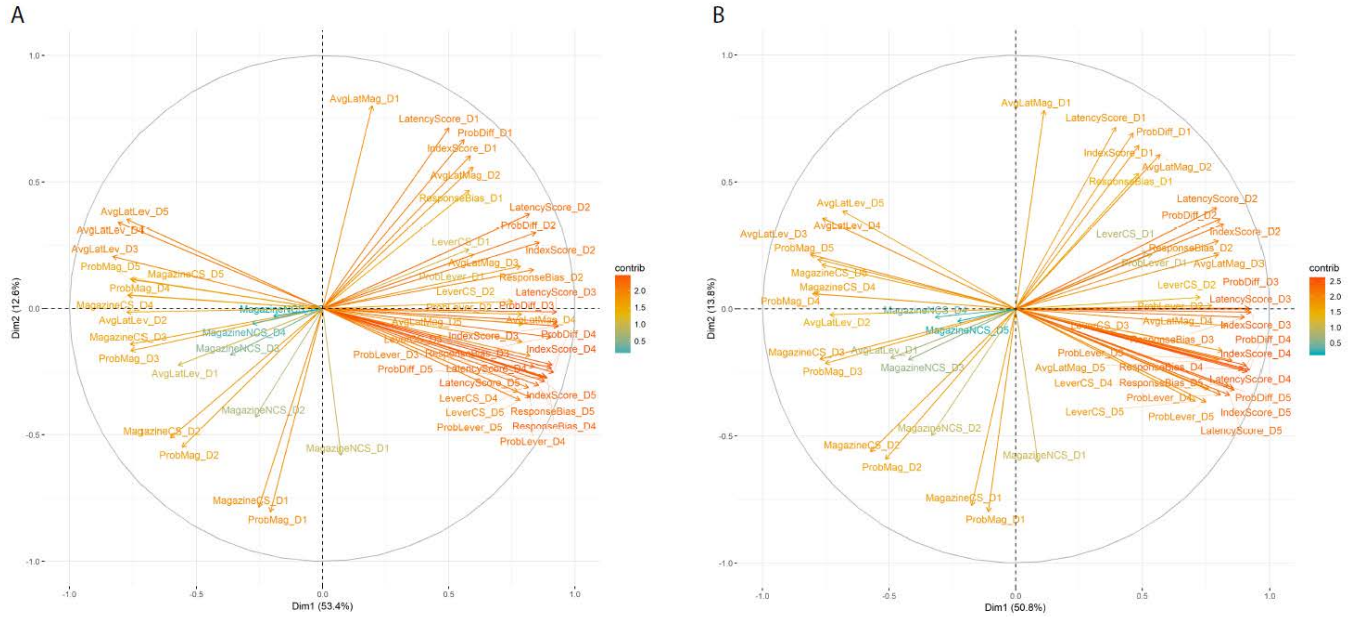


Fig 5. Correlation circle plots for 55 metric PCA in Charles River and Harlan

Correlation circles representing the strength and direction of each metrics' loading onto the first two PCs from PCA for (A) Charles River and (B) Harlan.

Finally, we mapped the first PC from running PCA on each metric across all 5 days and on all 11 metrics for each day (S8 Table), which yielded 3 additional genome-wide significant associations. Two associations with PC1 of magazine entries during CS presentation on days 1-5 were discovered on chromosomes 1 and 2. The last associated locus was on chromosome 3 for all PCA metrics on day 4 of training.

For each metric GWAS, we made quantile-quantile (Q-Q) plots (S2 File) to examine the distribution of the test statistics in comparison to the null expectation. We observed moderate levels of inflation for several of the GWAS, which we interpreted as evidence for polygenic signal rather than population structure, since the LMM should have accounted for population structure.

Given the large sample size for a rodent GWAS, the modest number of genome-wide significant associations was surprising. To determine if this was a phenomenon unique to our

behavioral metrics, which had relatively low SNP heritabilities (mean 7.8% in Harlan and 14.8% in Charles River), we also performed a GWAS for body weight, for which we had observed much higher heritability (63.2% and 42.7%). One limitation of this analysis is that we only had body weight data a subset of 1,858 rats (Charles River n=957; Harlan n=901). We found three genome wide significant associations for body weight, two in Harlan and one in Charles River. The meta-analysis supported the association from Charles River, but did not yield any additional associations. Though only 3 loci were identified, perhaps due to our limited sample size, the Q-Q plots (Page 89 of S2 File) show that there is an increased polygenic signal for body weight compared to the PavCA metrics, supporting our hypothesis that a more heritable trait would show stronger association in these populations.

Discussion

Although both rats and mice are widely used in the biomedical sciences, most studies in mice utilize inbred strains whereas studies in rats more commonly use outbred strains. One of the most extensively used outbred rat strains is Sprague Dawley [34]. While SD and other outbred rats have been used for selective breeding studies [35–37], this study was the first to use SD for GWAS. We densely genotyped more than 4,000 SD rats and used the data to characterize the genetic background of SD rats and to perform GWAS for the behaviors that comprise Pavlovian conditioned approach. This represents the largest rodent GWAS ever undertaken, and the first performed using a commercially available outbred rat population. We found dramatic genetic differences between SD rats obtained from Harlan versus Charles River. F_{ST} estimates show that SD rats from Harlan and Charles River are more differentiated than the major human subpopulations [23] and nearly as diverged as some subspecies of mice [38]. We also found evidence of population structure among the various breeding locations and barrier facilities for

each vendor. We found that SD rats from both Harlan and Charles River showed a rapid decay in LD; however, SD from Charles River have more polymorphisms and more favorable MAF and LD profiles, suggesting that future GWAS in SD are best done with rats obtained from Charles River.

We estimated that our meta-analysis was well powered for identify loci that accounted for only 1%-2% of total trait variance (S7 Fig). When we began this study, we were unaware of the large genetic differences between Sprague Dawley rats from Charles River and Harlan. To cope with the observed differentiation, we emulated the approach often taken in human genetics in which multiple groups are analyzed separately before being combined in a meta-analysis (e.g. [39]); however, it is well known that different human populations (e.g. East Asian and European) often do not share the same causal variants. Similarly, our power may have been reduced because causal variants were not shared between the Charles River and Harlan. Modifiers of causal variants might also be dissimilar between Charles River and Harlan, further hindering our meta-analyses. [40]

A notable success of this study was a locus we detected on chromosome 17, which was identified by the meta-analysis of Harlan and Charles River. This locus was identified for multiple metrics on days 4 and 5, as well as in the univariate analysis of the first PC from PCA on the comprehensive set of all 55 metrics. All genome-wide significant SNPs were located in an intron of *Fars2* (Fig3B), which is highly expressed in the Purkinje cells of the cerebellum [31] but has no previously known ties to behavior. Another promising locus for magazine entries and latency to magazine entry during CS presentation was on chromosome 1. That locus was consistent across multiple days of testing for both of these measures. However, the locus is in an intergenic region

with no known genes, indicating either the presence of an unannotated gene, or a regulatory site that influences a nearby gene.

Since PavCA metrics had never been used for a GWAS before, it was unclear which measurements and transformations would yield the best results. Therefore, we used several approaches, which led to the identification of over 20 genome-wide significant loci but involved 267 total tests with a significance threshold of alpha = 0.05. Despite high correlations between many of the metrics, some of the loci we detected were unique to a single PavCA metric, raising the possibility that they could be false positives. Furthermore, nearly half of the QTL identified occurred on earlier training days (days 1-3), which are harder to interpret since the literature surrounding Pavlovian conditioned approach focuses on behavior after training (4/5).

Using genome-wide genotype data, we have provided the first quantitative estimates of the SNP heritability of the component measurements of PavCA. The heritability estimates were lower than we had anticipated, but perhaps reasonable given the complexity of the behavior. The highest heritabilities (16-20% in Charles River) were seen for measures typically used to assess the propensity to attribute incentive salience to reward cues, including the averages of the day 4 and day 5 PavCA index score, response bias, and latency score. Previous work had shown that SD rats selectively bred for ~15 generations for high or low responses to a novel environment were also highly divergent for behaviors in PavCA [26]. Those selection studies demonstrated that SD rats had alleles that could exert a strong influence on PavCA performance. The present results are consistent with this conclusion but highlight the extent to which alleles can be concentrated over many generations of selective breeding. We obtained lower heritability estimates for SD rats from Harlan compared to Charles River, further emphasizing the genetic differences between SD rats

from the two vendors. With more highly heritable behavioral traits, we suspect the SD rats would have yielded better results.

Although our study is the first to carefully document population structure within SD rats, ours is not the first to highlight phenotypic differences among SD from different vendors. In 1973, Prejean et al. reported that the incidence of endocrine tumors varied among SD rats from different vendors [1]. Then Clark et al. (1991) [41] reported differences in noradrenergic neural projections among SD rats from different vendors. Subsequently, Turnbull & Rivier (1999) [4] reported vendor-specific differences in the response to inflammatory stimuli. Then Fuller et al. (2001) [42] reported vendor-specific differences in hypoxic response among SD rats. Even more recently, there have been additional publications reporting differences for a variety of traits among SD rats obtained from different vendors [2,5,43,44], and even suggesting that these phenotypic differences may extend to differences among a single vendor's breeding facility [3]. Our own studies have previously reported both behavioral and genetic differences among SD rats [6], conclusions that are much more comprehensively addressed with the present dataset. Specifically, in addition to wide spread genetic differences, we have also shown that SD rats obtained from Harlan show a much higher proclivity to become sign-trackers compared to SD rats from Charles River. However, neither these prior publications nor the current one can differentiate between two possibilities: that the observed behavioral differences are the result of the different environment in which these animals are raised versus the genetic difference that we have clearly demonstrated. This question could be addressed by future studies in which SD rats are bred in the same facility and the offspring tested in the same manner.

Regardless of whether differences in SD rats obtained from different vendors are due to genetic or environmental differences, our results demonstrate the need for much greater care in the

use of SD rats. Future studies may only want to use rats from a single vendor and from a single breeding facility to maintain a consistent genetic background. The differences among vendors can be a source of unwanted phenotypic variability, which alone might be a reason to avoid heterogeneous samples. Vendor differences, especially when not reported in the methods, can also lead problems with replication, since observations made in SD rats from Harlan may not be valid in SD rats from Charles River. Problems of replication and inadequate reporting also extend to differences between breeding facilities within a given vendor. However, a more subtle consequence of using SD rats from multiple vendors or breeding facilities is that spurious correlations can occur. For example, if SD rats from Charles River were higher for traits A and B, compared to SD rats from Harlan, a heterogeneous cohort of SD rats from both vendors would show a significant positive correlation between traits A and B. Such a correlation is unlikely to be due to a shared biological mechanism; it may instead be the result of either genetic population structure or environmental differences between SD rats from the two vendors.

Methods

Sprague Dawley samples

Tissue samples from 5,206 male Sprague Dawley rats were obtained, predominantly from Charles River and Harlan, with a few samples from Taconic. A subset of 4,625 of these rats went on to be genotyped by ddGBS and/or WGS. After sample filtering, a final set of 4,061 genotyped SD rats were used for the population genetic and association analyses. S1 Table lists the number of many samples that came from each vendor, breeding location, and barrier facility. Detailed information about these 4,061 rats is available in S3 File. Behavioral testing was performed between February 2012 and August 2015 as part of work for multiple studies [19,45–51,51–57]. All experiments were approved by the University of Michigan IACUC. Housing, feeding, lighting

and other relevant environmental conditions have been previously described. Following sacrifice at the University of Michigan, tissue samples were shipped to the University of Chicago; subsequent processing of those samples is described in the following sections.

Pavlovian conditioned approach

Pavlovian conditioned approach procedures have been thoroughly described previously [58,59] as a means to assess the tendency to attribute incentive motivational value or incentive salience to a cue that has been repeatedly paired with a noncontingent reward. Briefly, rats are placed into a testing chamber in which an illuminated lever (conditioned stimulus; CS) enters the chamber and after 8 seconds the lever-CS retracts and a food pellet (unconditioned stimulus; US) is immediately delivered into an adjacent food cup. Rats are scored for their three possible responses to the lever-CS entering the cage: approach and interact with the lever, approach and interact with the food receptacle (magazine), or make neither approach. Conditioned responses are captured during the 8-second period during which the lever-CS has entered the chamber, but before the food reward enters the magazine. The following measures are obtained in automated fashion: the number of lever contacts as measured by lever depressions, number of magazine entries as measured by infrared sensor in the food receptacle, and the latency to both during the 8-second lever-CS presentation. The rats are tested in this manner with 25 trials per session and one session is conducted per day for 5 consecutive days. For the purposes of this project, the number of lever contacts and magazine entries are summed across all 25 trials within a given session, and the latencies are averaged across 25 trials within a session.

Along with response counts and latencies, three additional measurements are recorded: 1) the proportion of trials in a session during which a rat made a lever contact ("probability" of lever press), 2) the proportion of trials during which they made a magazine entry ("probability" of magazine entry), and 3) the number of non-CS (NCS) magazine entries that occurred outside of the 8 second trials (when the cue was not present during the intertrial interval). We also calculated composite scores to categorize rats as sign-trackers (ST; defined as rats that preferentially interacted with the lever-CS), goal-trackers (GT; defined as rats that preferentially interacted with the food magazine), and intermediate responders (IR; rats that vacillated between sign- and goal-tracking behavior) [20]. These scores include: response bias ([lever presses – magazine entries]/[lever presses + magazine entries]), latency score ([average magazine entry latency – average lever press latency]/8), and probability difference ([lever press probability – magazine entry probability]). The PavCA index score is the average of the response bias, latency score, and probability difference. A value of [-1, -0.5] for the PavCA index score indicates a GT, (-0.5, 0.5) indicates an IR, and [0.5, 1] indicates a ST. We performed a Welch's 2-sample t-test to show that the PavCA index score distributions differed significantly between Charles River and Harlan SD rats ($t = 20.161$, $df = 3908.1$, p-value $< 2.2 \times 10^{-16}$). In summary, 11 PavCA metrics were available for analysis, each of which we measured on days 1, 2, 3, 4, and 5 (S10 Table).

Double digest genotype-by-sequencing (ddGBS)

To obtain genotypes, we used ddGBS, a genotyping method that reduces the complexity of the genome by only sequencing regions proximal to restriction enzyme cut sites [6,60]. We have recently described the technical aspects of this protocol in detail [61]. The ddGBS protocol used in this paper is a synthesis of the GBS approach described in Graboski et al. [62] and used more

recently by Parker et al. [10] and Gonzales et al. [27], and an analogous approach known as double digest restriction associated DNA sequencing (ddRADseq) [63].

DNA was extracted from rat tails using the PureLink® Genomic DNA kit. DNA purity was assayed using a Nanodrop 8000 (260/280 ≥ 1.8) and DNA integrity by gel electrophoresis (minimal smearing). Genomic DNA was then digested using two restriction enzymes: *PstI* (6-bp recognition site) and *NlaIII* (4-bp recognition site). Adapter oligos were annealed to overhangs left by *Pst1* and *NlaIII*. The *PstI* adapters contained 48 unique 4-8 bp in-line indexes [10,27,62]. A Y-adapter was annealed to the *NlaIII* cut sites, which controlled the direction of the first round of PCR amplification and thus ensured that the library was primarily composed of fragments with one of each of the adapters. Post-annealing, sets of 24 individual sample libraries were quantified and pooled. Pooled libraries were PCR amplified for 12 cycles, size-selected for 300-450bp using the Pippin Prep, and quality checked by Agilent Bioanalyzer (peak range ~ 300-500bp and conc. ≥ 20nM). Sequences for the 48 barcoded adapters, Y-adapter, and PCR primers are provided in S1 Text.

Sequencing of pooled libraries was performed by Beckman Coulter Genomics (now GENEWIZ). Sequencing was carried out on the Illumina HiSeq 2500 using v4 chemistry and 125-bp single-end reads. Each lane consisted of a pool of 24 samples, resulting in an average of 8.9 million reads per sample. A total of 4,608 unique ddGBS sample libraries were sequenced. Of these samples, 384 were sequenced twice, resulting in two sets of sequencing data for each sample from the same library prep that were used for to check genotype concordance.

Light whole-genome sequencing (WGS)

To discover new variants and support imputation, we performed low-coverage whole-genome sequencing of 80 SD rats from this same cohort. The rats were selected to represent sign-trackers, goal-trackers, and intermediate responders from each of the major barriers within the 6 major subpopulations of Harlan and Charles River. Sample libraries were prepared using the Illumina TruSeq® PCR-Free Library Prep kit and quality checked using an Agilent Bioanalyzer and qPCR on an Applied Biosystems StepOne Real-Time PCR System to ensure they met Illumina quality standards. Sample pooling (10 samples per pool) was performed by Beckman Coulter Genomics. Each pooled library was sequenced on two lanes of an Illumina 2500 flow cell with 125-bp single-end reads, resulting in an average of 51.6 million reads per sample per two lanes. Assuming that the rat genome (rn6) is ~2.87Gb, this provided about 180x coverage of the rat genome, or about 2.25x coverage per rat.

ddGBS Sequence data processing

We have recently described the bioinformatic steps that we use for ddGBS in detail [61]. We follow an analogous approach in this paper, though we deviate at the imputation step due to our use of SD instead of HS rats. Briefly, raw reads from ddGBS were demultiplexed using FASTX Barcode Splitter [64], allowing for 1 mismatch. After demultiplexing, barcodes were trimmed by *cutadapt* [65]. Any reads not matching a sample's barcode within 1-bp were filtered out. We removed 316 samples for which there were less than 4 million reads, leaving 4,292 samples with ddGBS data. We also used *cutadapt* to trim low-quality base pairs (phred quality score < 20) at the ends of the reads, and to remove 3' adapter sequences. Reads trimmed to less than 25-bp were discarded. Next, all reads were aligned to the rat reference genome assembly (rn6) using *bwa* [66].

All ddGBS reads were realigned at known indel sites by GATK's IndelRealigner [67]. Because of a lack of SD-specific variant data, we used variant data from 42 whole-genome sequenced rat strains and substrains [68] as the reference set for indels. We then used GATK to perform base quality score recalibration (BQSR) on the BAM files, using data (SNP & indel) from the 42 rat genomes as the "known" set of variants. For the ddGBS samples that were sequenced twice (381 remaining after filtering for read count), we performed all quality control and variant calling steps in parallel, since our goal was to compare calls made in these samples as a means of estimating the genotyping error rate.

Light whole-genome sequencing (WGS) data processing

Raw reads from WGS were processed in an analogous manner to the ddGBS data (detailed above) through the alignment step. After alignment, duplicates reads were removed using *picard* [69]. Reads were then realigned and underwent BQSR. The final WGS BAM files from each lane (2 files per sample) were merged. The WGS BAM files for the 63 samples that had undergone both ddGBS and WGS were then merged.

Variant discovery and imputation – ANGSD/Beagle

We found that GATK's HaplotypeCaller tool [67] was ineffectual at making high-confidence SNP calls in our dataset, likely due to the unusual distribution of reads produced by ddGBS. Instead, we used the Samtools variant calling model [70] as implemented by ANGSD [71] to estimate genotype likelihoods from the mapped ddGBS reads. Likelihoods were obtained in 10Mb chunks of the genome, which were subsequently concatenated. Major and minor alleles were inferred from the data based on allele frequency estimates made from the genotype

likelihoods. The likelihoods were only estimated at sites where at least 100 samples had reads. ddGBS data results in low call rates at many loci. However, we retained these loci because we anticipated they would be useful for imputation. Next, we used the ANGSD genotype likelihoods to impute missing genotypes (that is for SNPs where only a portion of the rats had genotyping information) using Beagle [72,73], which produced a VCF file containing hard genotype calls (0,1, or 2), dosages ([0,2]), and posterior probabilities for each genotype ([0,1]) for 2,274,118 biallelic SNPs in 4,309 rats (ddGBS+WGS). This is the unfiltered set of SNPs and samples we moved forward with for all subsequent steps. We elected not to pursue variant quality score recalibration using the GATK VariantRecalibrator algorithm [74], because we did not have the required "truth" SNP set. Due to the poorly understood population history of SD rats, it was unclear whether the variation present in the 42 rat genomes would be representative of the variation present in our sample. Using the 42 genomes as a reference for the VariantRecalibrator may also have negatively impacted the calling of novel SD variants.

STITCH (Sequencing To Imputation Through Constructing Haplotypes)

In addition to variant calling and imputation using ANGSD/Beagle, we also explored the use of STITCH [21], since it is specifically designed for low-coverage WGS data lacking haplotype reference panels. However, ddGBS data is higher coverage and sparser than the input for which STITCH was designed. We used the set of alignment files and known variant sites from the 42 genomes [68], as described above. STITCH queries the user for the number of ancestral haplotypes that exist in the population (K). Due to our lack of knowledge about the founder population, we ran STITCH on a single chromosome using 5 different values of K: 2, 3, 4, 5, and 6. We found that K values of 3, 4, and 5 worked similarly well and chose K=4 to maximize the

number of SNPs called and minimize error rate as ascertained by comparison of genotype calls between duplicate samples. STITCH yielded 8,691,886 biallelic SNPs, vastly more than were called with ANGSD/Beagle. However, after applying filters for dosage r^2/INFO score ≥ 0.9, MAF ≥ 0.01, HWE p-value ≤ $1x10^{-7}$, as well as removing sites in near perfect pairwise LD > 0.95, we found that the genotypes from STITCH contained fewer SNPs compared to the comparably filtered output from ANGSD/BEAGLE (see S2 Table). For this reason, we did not use the SNP calls made by STITCH in any of the analyses presented in this paper.

Genotype concordance check

Whereas some of our past projects that used GBS were able to determine the accuracy of GBS genotypes by comparing them to genotypes obtained from SNP microarrays, we did not have microarray-based genotypes for this cohort. Instead, we relied on the remaining 381 duplicate samples whose genotypes were called in parallel. To estimate genotyping accuracy, we compared the rate of concordance of hard genotype calls among the duplicate samples. We first filtered variants by dosage r^2 (DR^2), a measure of the accuracy of the genotype imputation performed by Beagle. We tested three different DR^2 thresholds (≥0.7, ≥0.8, ≥0.9). We then removed variants with MAFs < 1% or that violated Hardy-Weinberg equilibrium at a threshold of $1x10^{-7}$ in either vendor population. Concordance rates were checked by two methods: 1) by using hard calls in the RAW format from *plink 1.9* and dividing the number of times a genotype call matched between duplicate samples by the total number of SNPs and 2) by taking the mean Pearson correlation of the dosages of the duplicate samples. The results are presented in S3 Table. Similar error estimates were obtained by the hard call and dosage approaches. We chose to move forward with the DR^2 threshold of 0.9 for all subsequent analyses, which yielded an error rate of ~0.85%.

Post-genotyping sample filtering

We removed female rats (n=77) and rats from Taconic Farms (n=4) since they made up a very small fraction of the total sample. We also removed rats that showed poor clustering in the PCA analysis, described below. We filtered out individuals with unusually high or low rates of heterozygosity and high degrees of relatedness as detailed below. Lastly, we excluded a small set of duplicate samples (n=7) and samples missing phenotype data for mapping (n=10). All filters and sample numbers are listed in S5 Table. After these steps, 4,061 unique male SD rats from Charles River (n=1,780) and Harlan (n=2,281) remained.

Principal component analysis, identity-by-descent, and heterozygosity

We performed principal component analysis (PCA) on the cohort of 4,228 samples filtered for low read count, Taconic, and females. PCA was performed on hard genotype calls in R using the *prcomp* function in R [75] on a set of variants pruned for SNPs with MAF ≤ 0.05 in the combined sample set, SNPs in pairwise LD > 0.5, and SNPs violating HWE at a p-value $< 1x10^{-7}$ in either Charles River or Harlan. The first PC clearly separated animals from Harlan and Charles River; however, there was a set of 54 rats that did not visually cluster as expected at the level of vendor or subpopulation (data not shown). These animals were removed from all subsequent analyses (we assumed they reflected inaccurate records, sample mix-ups, or some other technical problems).

With this further reduced set of 4,174 rats, we continued on to assess the genetic relationships among the rats in our sample. SD rats were ordered in multiple batches over several years, and we suspected some of these rats would be closely related (siblings, cousins, etc.). We

reapplied the variant filters used for PCA and utilized the --*genome* function in *plink 1.9* on the resulting SNP set to estimate $\hat{\pi}$ (the proportion of genotypes predicted to be identical by descent), for all pairs of samples. Panels A and C in S4 Fig show that while most of the animals were unrelated, there were a significant subset of closely related pairs, as well as some pairs with exceedingly high IBD1 rates in Harlan.

We used the *plink 1.9* function --*het* to examine possible inflation or deflation of heterozygosity rates in our samples. Panels A and C in S8 Fig show that a handful of samples in both populations with uncharacteristically high (> 0.25) or low (< 0.25) rates of heterozygosity. We filtered out 34 such samples, as we found they drove the majority of the anomalous signal in our pre-filtering plots in S4 Fig. We also removed 12 samples with more than 30 close relations (defined as $\hat{\pi} \geq 0.1875$) and 32 samples that had a $\hat{\pi} \geq 0.6$ with another sample (likely sample contamination/mix-up). After applying these sample filters, we were left with Panels B and D in S4 Fig and S8 Fig.

After reaching our final set of 4,061 rats, we reapplied the SNP filters on the reduced sample set, resulting in 4,502 SNPs that were polymorphic in both populations. We then ran PCA as described above. We also repeated PCA on the samples from Charles River and Harlan separately to examine substructure within each population.

Final variant filtering and minor allele frequency spectrum

After establishing the final cohort of 4,061 rats, we sought to establish a final set of SNPs to be used for the association analyses. First, we removed sites with $DR^2 < 0.9$. We then separated

the Charles River and Harlan samples for all subsequent variant filtering performed in *plink 1.9*, which converts the posterior probabilities we received from Beagle to hard genotype calls, so long as the probability is ≥ 0.9. In each population, we removed SNPs with MAF < 0.01 using the *–maf* option in *plink 1.9*. Lastly, we used the *--hardy* function in *plink 1.9* [76] to perform tests for Hardy-Weinberg equilibrium. We did this for samples from each major subpopulation in each vendor (for Harlan: Frederick, Haslett, and Indianapolis; for Charles River: Portage, Raleigh, and St. Constant) because the population structure (Figure XX) would have inflated the HWE statistics. We removed all SNPs with an HWE p-value < $1x10^{-7}$ in any of the 6 subpopulations. At the end of this process, we retained a total of 214,309 SNPs in Charles River and 114,568 SNPs in Harlan. A summary of the filtering process is shown in S2 Table. We used the *--freq* option in *plink 1.9* to estimate the MAFs for these final filtered sets SNPs. The distributions are show in Fig 1D.

Taking the union of these two sets of SNPs, there were a total of 234,887 polymorphic sites passing our filters between the two populations. Our final set of filtered SNPs contained 105,638 novel SNPs that had not be previously reported by Hermsen et al. [68]. We also compared our final set of SNPs to the most recent dbSNP release for rats (Build 149, November 7, 2016) and found that 195,299 of the 234,887 SNPs we discovered were not present in the current dbSNP build. The full set of SNPs for all samples, as well as the filtered VCF files for Harlan and Charles River and the filtered union set BED files are available (S4 File, S5 File, S6 File, & S7 File).

Fixation Index (F_{ST})

To quantify the population divergence between SD rats from Harlan and Charles River, we computed F_{ST} for each SNP in the union set of 234,887 using *smartpca* within the *EIGENSOFT* package [23,77,78]. Due to the substructure within vendor and vendor subpopulation that we saw in our PCA analyses, we chose to approach the F_{ST} estimation in three different ways. We first

grouped the rats solely on vendor to calculate F_{ST} between Harlan and Charles River rats. Then, we broke these populations down further into the 3 major breeding facilities in each of the two vendors (listed previously) and calculated the pairwise F_{ST} between each sample set. Finally, we split the breeding facilities into the barriers that composed them, treating each barrier as a separate population. In each case, we computed all possible pairwise F_{ST} values. Samples from poorly represented barrier facilities were removed from the latter 2 analyses.

Linkage Disequilibrium

We plotted the decay of linkage disequilibrium (LD) using the *--r²* utility in *plink 1.9* and the procedures described in Parker et. al 2016 [10]. Briefly, each population's curve included all SNPs with MAF > 20% and pairwise LD comparisons were restricted to SNPs with allele frequencies within 5% of each other. An average r^2 estimate was obtained using 10,000 randomly selected SNP pairs from each 100kb interval for the distance between two SNPs, starting with 0-100kb and end at 9.9-10Mb. Harlan's curve used 35,770 autosomal SNPs in 2,281 rats and Charles River's used 112,678 autosomal SNPs in 1,780 rats. The CFW mouse curve was downloaded from a repository for Parker et al. (http://datadryad.org/resource/doi:10.5061/dryad.2rs41)[10] and used for comparison as another commercially available, outbred rodent stock.

LMM covariates and phenotype data pre-processing

To select covariates for the GWAS, we performed univariate linear regression for each potential covariate for each PavCA metric. This was done separately for rats from Charles River and Harlan. Any covariates that accounted for 1% or more of the variance for at least one PavCA metric were passed into multivariate model selection with the R package *leaps* [79]. Model

selection with *leaps* was performed for all metrics for all days of testing, as well as the average of days 4 and 5. Out of the 66 models, all covariates that had surpassed the 1% threshold to reach this step were ultimately selected in at least 40% of the *leaps* models. For consistency, this led us to simply use the full set of covariates for all downstream GWAS. These covariates were included: age at testing, housing (binary – single or multiple), light cycle (binary – standard or reverse), and a set of binary 'indicator' variables to model the effects of different experimenters/technicians (10 variables were used for Charles River and 7 were used for Harlan). All covariates were included in the LMMs for association testing by GEMMA, rather than being regressed out prior to GWAS. We excluded an additional 88 rats from the final association analyses because of missing covariate data.

Many of the PavCA metrics were exceedingly non-normally distributed. In most cases this was expected due to how the behaviors were measured and defined. For example, the rats only had a window of 8 seconds in each trial during which to contact the lever and/or make a nose poke into the food magazine. All values for "average latency to lever press" or "average latency to magazine entry" were therefore necessarily between 0 and 8 seconds. As is typical for latency traits, many of the values were near 0 or exactly 8. Similarly, the "probability" of lever press or magazine entry were very skewed towards the limits 0 and 1, especially after conditioned responding had been established on the later test days of training. Given these unusual distributions, we chose to quantile normalize all metrics prior to association testing, accepting a possible loss in power since samples with identical values are ranked randomly during the quantile normalization procedure.

In prior studies, sign trackers, intermediate responders, and goal trackers are categorized based on their behavior (i.e. PavCA) on days 4 and 5 of training [20]. In an attempt to emulate this approach, we created a binary trait (essentially case/control: ST/GT) and treated the intermediate responders as unknown. The threshold of PavCA index score typically used to define ST and GT are 0.5 and -0.5, respectively. We considered three options: 0, +/-0.25, & +/-0.5; for the 0 threshold all rats were either ST or GT, whereas for the other two thresholds some rats were intermediate, so they were excluded from the analysis.

Lastly, we explored the use of principal component analysis (PCA) to summarize the 55 metrics (S10 Table). We regressed out all covariates mentioned above (age, housing, light cycle, & experimenters) and mean-centered and quantile normalized the residuals. We ran PCA on the normalized residuals in 3 different ways: all 55 metrics simultaneously, all 11 metrics for each of the 5 days, and each of the 11 metrics for all 5 days. We choose the latter two approaches because we reasoned that there may be a specific day, or a metric by itself across days, that was uniquely of interest to the behavior. We only mapped the first PC for each, as it accounted for the majority of the variation for all metrics/days. Raw factor loadings and the percentage of the variance in the PCs explained by each metric for the first 5 PCs in each population are summarized in S8 Table. The percent of behavioral variation explained by each of the first 10 PCs in each population is shown by S5 Fig. Fig 5 graphically shows the correlations between each factor and the first two PCs in each population using the *factoextra* package in R [80]. Since the first 3 PCs cumulatively accounted for more than 70% of the variance in both populations, we only considered these for mapping.

A subset of 1,858 rats (Charles River n=957; Harlan n=901) also had body weight measurements, taken when the shipments were received from the breeding facilities, but prior to the onset of behavioral testing. We used this data for mapping.

SNP-based heritabilities

Heritabilities were estimated separately for Charles River and Harlan with the union set of 234,887 SNPs. We used this SNP set to construct genetic relationship matrices (GRMs) for each population using GCTA [81]. We then used the restricted maximum likelihood (REML) approach within GCTA on the GRMs, covariates, and quantile normalized Pavlovian conditioned approach data to calculate the SNP-based heritabilities for each metric, as well the previously mentioned PCs.

GWAS

We used GEMMA [82], which implements an LMM for GWAS analysis. We included a GRM as a random term to account for relatedness and population structure. Though beneficial for preventing false positive associations, GRMs can also reduce power to detect QTLs in populations with greater levels of LD; this is due to proximal contamination [83,84]. To avoid this reduction in power, we used the leave-one-chromosome-out (LOCO) approach [27,85,86]. As described above, we selected covariates that were included as fixed effects in our model (listed in S9 Table). When mapping PCs 1, 2, and 3 for all 55 metrics, the fixed covariates were excluded from the LMM, as they had already been regressed out prior to calculation of the PCs. For the body weight GWAS, only age was used as a fixed covariate in the model. Any samples missing measurements for a given metric were removed from that analysis. For all GWAS, genotypes were represented

as dosages (continuous [0,2]) in lieu of 'hard' genotype calls [0, 1, 2] to account for uncertainty in the genotype calls. Reported p-values come from the likelihood-ratio test (LRT) performed by GEMMA. Results were plotted using a custom R script and LocusZoom plots were created using the stand-alone software [87], custom SD SNP databases and LD calculations, and a 1.5Mb flanking region.

GWAS meta-analysis

We used the beta estimates and LRT p-values to perform a meta-analysis on the sample sets from Harlan and Charles River using the method outlined by Myers et al 2014 [88]. Analysis was limited to the set of 93,990 SNPs that existed at a frequency of at least 1% in both Harlan and Charles River. The allele frequencies, imputation accuracy (DR^2), and sample sizes were factored into the weighting for the z-statistics, which were then summed across the two GWAs. All analyses were completed in R [75].

P-value significance thresholds

In human GWAS, $5x10^{-8}$ is widely used as a significance threshold [89]. However, model organisms have widely different levels of LD, meaning that the effective number of independent tests differs between studies. Therefore, many prior studies have used permutation testing [90,91], where phenotype data is shuffled with respect to fixed genotypes. The computational load of such methods becomes intractable for studies with large sample sizes and several traits being mapped [83]. Thus, we used the sequence of SLIDE [92,93] and MultiTrans [94] to obtain significance thresholds. We used separate thresholds for Charles River and Harlan because the number of SNPs, the LD structure, and the marker-based heritabilities were different. An advantage of this approach

is that only one threshold was needed. We used a sliding window of 1000 SNPs and sampled from the multivariate normal 10 million times to obtain a 0.05 significance threshold of 8.96×10^{-7} ($-\log_{10}(p) = 6.05$) for Charles River and 2.19×10^{-6} ($-\log_{10}(p) = 5.66$) for Harlan. We calculated the thresholds for body weight separately due to its substantially greater heritability; however, the differences in thresholds were negligible (8.82×10^{-7} for Charles River and 2.13×10^{-6} for Harlan). We had difficulty obtaining a significance threshold for the meta-analysis because MultiTrans cannot accommodate such a situation and permutation would have been computationally expensive. Therefore, we used the threshold for Harlan for all meta-analyses, even though the meta-analyses utilized ~20k fewer SNPs, meaning that this threshold is likely to be overly conservative.

Power Analysis

Power analysis was performed in Quanto v 1.2.4 [95,96] using the model for testing an additive genetic effect for a quantitative trait in independent individuals. We estimated our power with a fixed sample size of 2,000, which was approximately the midpoint for the Harlan and Charles River sample sets. We regressed out the covariates in our LMM from the quantile normalized metrics and calculated the mean and standard deviation from the residuals ($\mu \approx 0$, $\sigma \approx 0.98$). The resulting curve is show in S7 Fig.

Acknowledgements

We would like to thank John Novembre, Mark Abney, and Dan Nicolae at the University of Chicago for their advice concerning the statistical analyses involved in this publication.

Citations

1. Prejean JD, Peckham JC, Casey AE, Griswold DP, Weisburger EK, Weisburger JH. Spontaneous tumors in Sprague-Dawley rats and Swiss mice. Cancer Res. 1973;33: 2768–2773.

2. Bodnar TS, Hill LA, Taves MD, Yu W, Soma KK, Hammond GL, et al. Colony-Specific Differences in Endocrine and Immune Responses to an Inflammatory Challenge in Female Sprague Dawley Rats. Endocrinology. 2015;156: 4604–4617. doi:10.1210/en.2015-1497

3. Brower M, Grace M, Kotz CM, Koya V. Comparative analysis of growth characteristics of Sprague Dawley rats obtained from different sources. Lab Anim Res. 2015;31: 166. doi:10.5625/lar.2015.31.4.166

4. Turnbull AV, Rivier CL. Sprague-Dawley Rats Obtained from Different Vendors Exhibit Distinct Adrenocorticotropin Responses to Inflammatory Stimuli. Neuroendocrinology. 1999;70: 186–195. doi:10.1159/000054475

5. Weber K. Differences in Types and Incidence of Neoplasms in Wistar Han and Sprague-Dawley Rats. Toxicol Pathol. 2017;45: 64–75. doi:10.1177/0192623316672075

6. Fitzpatrick CJ, Gopalakrishnan S, Cogan ES, Yager LM, Meyer PJ, Lovic V, et al. Variation in the Form of Pavlovian Conditioned Approach Behavior among Outbred Male Sprague-Dawley Rats from Different Vendors and Colonies: Sign-Tracking vs. Goal-Tracking. Campolongo P, editor. PLoS ONE. 2013;8: e75042. doi:10.1371/journal.pone.0075042

7. Woods LCS, Mott R. Heterogeneous Stock Populations for Analysis of Complex Traits. In: Schughart K, Williams RW, editors. Systems Genetics. New York, NY: Springer New York; 2017. pp. 31–44. doi:10.1007/978-1-4939-6427-7_2

805 8. Parker CC, Chen H, Flagel SB, Geurts AM, Richards JB, Robinson TE, et al. Rats are the smart choice:
806 Rationale for a renewed focus on rats in behavioral genetics. Neuropharmacology. 2014;76: 250–258.
807 doi:10.1016/j.neuropharm.2013.05.047

808 9. Rat Genome Sequencing and Mapping Consortium, Baud A, Hermsen R, Guryev V, Stridh P, Graham
809 D, et al. Combined sequence-based and genetic mapping analysis of complex traits in outbred rats.
810 Nat Genet. 2013;45: 767–775. doi:10.1038/ng.2644

811 10. Parker CC, Gopalakrishnan S, Carbonetto P, Gonzales NM, Leung E, Park YJ, et al. Genome-wide
812 association study of behavioral, physiological and gene expression traits in outbred CFW mice. Nat
813 Genet. 2016;48: 919–926. doi:10.1038/ng.3609

814 11. Nicod J, Davies RW, Cai N, Hassett C, Goodstadt L, Cosgrove C, et al. Genome-wide association of
815 multiple complex traits in outbred mice by ultra-low-coverage sequencing. Nat Genet. 2016;48: 912–
816 918. doi:10.1038/ng.3595

817 12. Charles River. CD® (Sprague Dawley) IGS Rat Details. In: CD® (Sprague Dawley) IGS Rat [Internet].
818 [cited 2 Jul 2018]. Available: https://www.criver.com/products-services/find-model/cd-sd-igs-
819 rat?region=3611

820 13. Envigo, Inc. Sprague Dawley® outbred rat | Envigo. In: Envigo [Internet]. [cited 2 Jul 2018]. Available:
821 https://www.envigo.com/products-services/research-models-services/models/research-
822 models/rats/outbred/sprague-dawley-outbred-rat/

823 14. Charles River Laboratories International, Inc. International Genetic Standardization (IGS) Program
824 [Internet]. 2016. Available:

825 https://www.criver.com/sites/default/files/Technical%20Resources/Charles%20River%20Internatio
826 nal%20Genetic%20Standardization%20Program.pdf

827 15. White WJ, Lee CS. The Development and Maintenance of the Crl:CD(SD) IGS BR Rat Breeding System
828 [Internet]. 1998. Available:
829 https://www.crj.co.jp/cms/cmsrs/pdf/company/rm_rm_a_igs_rat_breeding_system.pdf

830 16. Envigo, Inc. Genetic integrity assurance program | Envigo. In: Envigo [Internet]. 2 Jul 2018 [cited 2 Jul
831 2018]. Available: https://www.envigo.com/products-services/research-models-
832 services/resources/genetic-integrity-assurance-program/

833 17. Poiley SM. A systematic method of breeder rotation for non-inbred laboratory colonies. Proc Anim
834 Care Panel. 1960;10: 159–166.

835 18. Flagel SB, Akil H, Robinson TE. Individual differences in the attribution of incentive salience to reward-
836 related cues: Implications for addiction. Neuropharmacology. 2009;56: 139–148.
837 doi:10.1016/j.neuropharm.2008.06.027

838 19. Robinson TE, Yager LM, Cogan ES, Saunders BT. On the motivational properties of reward cues:
839 Individual differences. Neuropharmacology. 2014;76: 450–459.
840 doi:10.1016/j.neuropharm.2013.05.040

841 20. Meyer PJ, Lovic V, Saunders BT, Yager LM, Flagel SB, Morrow JD, et al. Quantifying Individual Variation
842 in the Propensity to Attribute Incentive Salience to Reward Cues. Zhuang X, editor. PLoS ONE. 2012;7:
843 e38987. doi:10.1371/journal.pone.0038987

844 21. Davies RW, Flint J, Myers S, Mott R. Rapid genotype imputation from sequence without reference
845 panels. Nat Genet. 2016;48: 965–969. doi:10.1038/ng.3594

846 22. Holsinger KE, Weir BS. Genetics in geographically structured populations: defining, estimating and
847 interpreting FST. Nat Rev Genet. 2009;10: 639–650. doi:10.1038/nrg2611

848 23. Bhatia G, Patterson N, Sankararaman S, Price AL. Estimating and interpreting FST: The impact of rare
849 variants. Genome Res. 2013;23: 1514–1521. doi:10.1101/gr.154831.113

850 24. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for
851 Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet. 2007;81: 559–
852 575. doi:10.1086/519795

853 25. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the
854 challenge of larger and richer datasets. GigaScience. 2015;4. doi:10.1186/s13742-015-0047-8

855 26. Flagel SB, Robinson TE, Clark JJ, Clinton SM, Watson SJ, Seeman P, et al. An Animal Model of Genetic
856 Vulnerability to Behavioral Disinhibition and Responsiveness to Reward-Related Cues: Implications
857 for Addiction. Neuropsychopharmacology. 2010;35: 388–400. doi:10.1038/npp.2009.142

858 27. Gonzales NM, Seo J, Hernandez-Cordero AI, St. Pierre CL, Gregory JS, Distler MG, et al. Genome wide
859 association study of behavioral, physiological and gene expression traits in a multigenerational mouse
860 intercross. 2017; doi:10.1101/230920

861 28. Clark SL, Aberg KA, Nerella S, Kumar G, McClay JL, Chen W, et al. Combined Whole Methylome and
862 Genomewide Association Study Implicates *CNTN4* in Alcohol Use. Alcohol Clin Exp Res. 2015;39:
863 1396–1405. doi:10.1111/acer.12790

864 29. Yoshihara Y, Kawasaki M, Tamada A, Nagata S, Kagamiyama H, Mori K. Overlapping and differential
865 expression of BIG-2, BIG-1, TAG-1, and F3: Four members of an axon-associated cell adhesion

866 molecule subgroup of the immunoglobulin superfamily. J Neurobiol. 1995;28: 51–69.
867 doi:10.1002/neu.480280106

868 30. Mikulska-Ruminska K, Kulik AJ, Benadiba C, Bahar I, Dietler G, Nowak W. Nanomechanics of
869 multidomain neuronal cell adhesion protein contactin revealed by single molecule AFM and SMD. Sci
870 Rep. 2017;7. doi:10.1038/s41598-017-09482-w

871 31. Yang Y, Liu W, Fang Z, Shi J, Che F, He C, et al. A Newly Identified Missense Mutation in *FARS2* Causes
872 Autosomal-Recessive Spastic Paraplegia. Hum Mutat. 2016;37: 165–169. doi:10.1002/humu.22930

873 32. Walker MA, Mohler KP, Hopkins KW, Oakley DH, Sweetser DA, Ibba M, et al. Novel Compound
874 Heterozygous Mutations Expand the Recognized Phenotypes of *FARS2* -Linked Disease. J Child Neurol.
875 2016;31: 1127–1137. doi:10.1177/0883073816643402

876 33. Vernon HJ, McClellan R, Batista DA, Naidu S. Mutations in *FARS2* and non-fatal mitochondrial
877 dysfunction in two siblings. Am J Med Genet A. 2015;167: 1147–1151. doi:10.1002/ajmg.a.36993

878 34. Foster JR, Frost D. The History of the Rat. Boorman's Pathology of the Rat. Elsevier; 2018. pp. 7–12.
879 doi:10.1016/B978-0-12-391448-4.00003-4

880 35. Scott PA, Cierpial MA, Kilts CD, Weiss JM. Susceptibility and resistance of rats to stress-induced
881 decreases in swim-test activity: a selective breeding study. Brain Res. 1996;725: 217–230.

882 36. Stead JDH, Clinton S, Neal C, Schneider J, Jama A, Miller S, et al. Selective Breeding for Divergence in
883 Novelty-seeking Traits: Heritability and Enrichment in Spontaneous Anxiety-related Behaviors. Behav
884 Genet. 2006;36: 697–712. doi:10.1007/s10519-006-9058-7

885 37. Bertholomey ML, West CHK, Jensen ML, Li T-K, Stewart RB, Weiss JM, et al. Genetic propensities to
886 increase ethanol intake in response to stress: studies with selectively bred swim test susceptible

887 (SUS), alcohol-preferring (P), and non-preferring (NP) lines of rats. Psychopharmacology (Berl).
888 2011;218: 157–167. doi:10.1007/s00213-011-2381-6

889 38. Geraldes A, Basset P, Smith KL, Nachman MW. Higher differentiation among subspecies of the house
890 mouse (Mus musculus) in genomic regions with low recombination: RECOMBINATION AND
891 SPECIATION IN MICE. Mol Ecol. 2011;20: 4722–4736. doi:10.1111/j.1365-294X.2011.05285.x

892 39. Sanchez-Roige S, Palmer AA, Fontanillas P, Elson S, Adams MJ, Howard DM, et al. Genome-wide
893 association study meta-analysis of the Alcohol Use Disorder Identification Test (AUDIT) in two
894 population-based cohorts (N=141,958). 2018; doi:10.1101/275917

895 40. Sittig LJ, Carbonetto P, Engel KA, Krauss KS, Barrios-Camacho CM, Palmer AA. Genetic Background
896 Limits Generalizability of Genotype-Phenotype Relationships. Neuron. 2016;91: 1253–1259.
897 doi:10.1016/j.neuron.2016.08.013

898 41. Clark FM, Yeomans DC, Proudfit HK. The noradrenergic innervation of the spinal cord: differences
899 between two substrains of Sprague-Dawley rats determined using retrograde tracers combined with
900 immunocytochemistry. Neurosci Lett. 1991;125: 155–158.

901 42. Fuller DD, Baker TL, Behan M, Mitchell GS. Expression of hypoglossal long-term facilitation differs
902 between substrains of Sprague-Dawley rat. Physiol Genomics. 2001;4: 175–181.
903 doi:10.1152/physiolgenomics.2001.4.3.175

904 43. Langer M, Brandt C, Löscher W. Marked strain and substrain differences in induction of status
905 epilepticus and subsequent development of neurodegeneration, epilepsy, and behavioral alterations
906 in rats. [corrected]. Epilepsy Res. 2011;96: 207–224. doi:10.1016/j.eplepsyres.2011.06.005

907 44. Segerström L, Roman E. Response: Commentary: Supplier-dependent differences in intermittent
908 voluntary alcohol intake and response to naltrexone in Wistar rats. Front Neurosci. 2016;10.
909 doi:10.3389/fnins.2016.00442

910 45. Pitchers KK, Flagel SB, O'Donnell EG, Solberg Woods LC, Sarter M, Robinson TE. Individual variation in
911 the propensity to attribute incentive salience to a food cue: Influence of sex. Behav Brain Res.
912 2015;278: 462–469. doi:10.1016/j.bbr.2014.10.036

913 46. Pitchers KK, Wood TR, Skrzynski CJ, Robinson TE, Sarter M. The ability for cocaine and cocaine-
914 associated cues to compete for attention. Behav Brain Res. 2017;320: 302–315.
915 doi:10.1016/j.bbr.2016.11.024

916 47. Kawa AB, Bentzley BS, Robinson TE. Less is more: prolonged intermittent access cocaine self-
917 administration produces incentive-sensitization and addiction-like behavior. Psychopharmacology
918 (Berl). 2016;233: 3587–3602. doi:10.1007/s00213-016-4393-8

919 48. Ahrens AM, Meyer PJ, Ferguson LM, Robinson TE, Aldridge JW. Neural Activity in the Ventral Pallidum
920 Encodes Variation in the Incentive Value of a Reward Cue. J Neurosci. 2016;36: 7957–7970.
921 doi:10.1523/JNEUROSCI.0736-16.2016

922 49. Ahrens AM, Singer BF, Fitzpatrick CJ, Morrow JD, Robinson TE. Rats that sign-track are resistant to
923 Pavlovian but not instrumental extinction. Behav Brain Res. 2016;296: 418–430.
924 doi:10.1016/j.bbr.2015.07.055

925 50. Singer BF, Guptaroy B, Austin CJ, Wohl I, Lovic V, Seiler JL, et al. Individual variation in incentive
926 salience attribution and accumbens dopamine transporter expression and function. Dalley J, editor.
927 Eur J Neurosci. 2016;43: 662–670. doi:10.1111/ejn.13134

928 51. Yager LM, Pitchers KK, Flagel SB, Robinson TE. Individual Variation in the Motivational and
929 Neurobiological Effects of an Opioid Cue. Neuropsychopharmacology. 2015;40: 1269–1277.
930 doi:10.1038/npp.2014.314

931 52. Meyer PJ, Cogan ES, Robinson TE. The Form of a Conditioned Stimulus Can Influence the Degree to
932 Which It Acquires Incentive Motivational Properties. Le Foll B, editor. PLoS ONE. 2014;9: e98163.
933 doi:10.1371/journal.pone.0098163

934 53. Yager LM, Robinson TE. A classically conditioned cocaine cue acquires greater control over motivated
935 behavior in rats prone to attribute incentive salience to a food cue. Psychopharmacology (Berl).
936 2013;226: 217–228. doi:10.1007/s00213-012-2890-y

937 54. Saunders BT, Yager LM, Robinson TE. Cue-Evoked Cocaine "Craving": Role of Dopamine in the
938 Accumbens Core. J Neurosci. 2013;33: 13989–14000. doi:10.1523/JNEUROSCI.0450-13.2013

939 55. Paolone G, Angelakos CC, Meyer PJ, Robinson TE, Sarter M. Cholinergic Control over Attention in Rats
940 Prone to Attribute Incentive Salience to Reward Cues. J Neurosci. 2013;33: 8321–8335.
941 doi:10.1523/JNEUROSCI.0709-13.2013

942 56. Morrow JD, Saunders BT, Maren S, Robinson TE. Sign-tracking to an appetitive cue predicts incubation
943 of conditioned fear in rats. Behav Brain Res. 2015;276: 59–66. doi:10.1016/j.bbr.2014.04.002

944 57. Singer BF, Bryan MA, Popov P, Scarff R, Carter C, Wright E, et al. The sensory features of a food cue
945 influence its ability to act as an incentive stimulus and evoke dopamine release in the nucleus
946 accumbens core. Learn Mem. 2016;23: 595–606. doi:10.1101/lm.043026.116

947 58. Fitzpatrick CJ, Morrow JD. Pavlovian Conditioned Approach Training in Rats. J Vis Exp JoVE. 2016;
948 e53580. doi:10.3791/53580

949 59. Robinson TE, Flagel SB. Dissociating the Predictive and Incentive Motivational Properties of Reward-
950 Related Cues Through the Study of Individual Differences. Biol Psychiatry. 2009;65: 869–873.
951 doi:10.1016/j.biopsych.2008.09.006

952 60. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A Robust, Simple Genotyping-
953 by-Sequencing (GBS) Approach for High Diversity Species. Orban L, editor. PLoS ONE. 2011;6: e19379.
954 doi:10.1371/journal.pone.0019379

955 61. Gao J, Gileta A, Palmer A. An efficient variant calling pipeline for double digest genotype-by-
956 sequencing data in heterogeneous stock rats. 2018.

957 62. Grabowski PP, Morris GP, Casler MD, Borevitz JO. Population genomic variation reveals roles of
958 history, adaptation and ploidy in switchgrass. Mol Ecol. 2014;23: 4059–4073. doi:10.1111/mec.12845

959 63. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double Digest RADseq: An Inexpensive
960 Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. Orlando L,
961 editor. PLoS ONE. 2012;7: e37135. doi:10.1371/journal.pone.0037135

962 64. Hannon Lab. FASTX-Toolkit [Internet]. 2010. Available:
963 http://hannonlab.cshl.edu/fastx_toolkit/index.html

964 65. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
965 EMBnet.journal. 2011;17: 10. doi:10.14806/ej.17.1.200

966 66. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform.
967 Bioinformatics. 2009;25: 1754–1760. doi:10.1093/bioinformatics/btp324

968 67. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis
969 Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res.
970 2010;20: 1297–1303. doi:10.1101/gr.107524.110

971 68. Hermsen R, de Ligt J, Spee W, Blokzijl F, Schäfer S, Adami E, et al. Genomic landscape of rat strain and
972 substrain variation. BMC Genomics. 2015;16. doi:10.1186/s12864-015-1594-1

973 69. Broad Institute. Picard Tools [Internet]. 2018. Available: http://broadinstitute.github.io/picard/

974 70. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population
975 genetical parameter estimation from sequencing data. Bioinformatics. 2011;27: 2987–2993.
976 doi:10.1093/bioinformatics/btr509

977 71. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: Analysis of Next Generation Sequencing Data. BMC
978 Bioinformatics. 2014;15. doi:10.1186/s12859-014-0356-4

979 72. Browning BL, Browning SR. A Unified Approach to Genotype Imputation and Haplotype-Phase
980 Inference for Large Data Sets of Trios and Unrelated Individuals. Am J Hum Genet. 2009;84: 210–223.
981 doi:10.1016/j.ajhg.2009.01.005

982 73. Browning BL, Browning SR. Genotype Imputation with Millions of Reference Samples. Am J Hum
983 Genet. 2016;98: 116–126. doi:10.1016/j.ajhg.2015.11.020

984 74. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation
985 discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43: 491–498.
986 doi:10.1038/ng.806

987 75. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R
988 Foundation for Statistical Computing; 2017. Available: http://www.R-project.org/

989 76. Wigginton JE, Cutler DJ, Abecasis GR. A Note on Exact Tests of Hardy-Weinberg Equilibrium. Am J Hum
990 Genet. 2005;76: 887–893. doi:10.1086/429864

991 77. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis
992 corrects for stratification in genome-wide association studies. Nat Genet. 2006;38: 904–909.
993 doi:10.1038/ng1847

994 78. Patterson N, Price AL, Reich D. Population Structure and Eigenanalysis. PLoS Genet. 2006;2: e190.
995 doi:10.1371/journal.pgen.0020190

996 79. Thomas Lumley based on Fortran code by Alan Miller. leaps: Regression Subset Selection. R package
997 version 3.0. [Internet]. 2017. Available: https://CRAN.R-project.org/package=leaps

998 80. Alboukadel Kassambara and Fabian Mundt. factoextra: Extract and Visualize the Results of
999 Multivariate Data Analyses [Internet]. Available: https://CRAN.R-project.org/package=factoextra

1000 81. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J
1001 Hum Genet. 2011;88: 76–82. doi:10.1016/j.ajhg.2010.11.011

1002 82. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet.
1003 2012;44: 821–824. doi:10.1038/ng.2310

1004 83. Cheng R, Palmer AA. A Simulation Study of Permutation, Bootstrap, and Gene Dropping for Assessing
1005 Statistical Significance in the Case of Unequal Relatedness. Genetics. 2013;193: 1015–1018.
1006 doi:10.1534/genetics.112.146332

1007 84. Listgarten J, Lippert C, Kadie CM, Davidson RI, Eskin E, Heckerman D. Improved linear mixed models
1008 for genome-wide association studies. Nat Methods. 2012;9: 525–526. doi:10.1038/nmeth.2037

1009 85. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the application of
1010 mixed-model association methods. Nat Genet. 2014;46: 100–106. doi:10.1038/ng.2876

1011 86. Loh P-R, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, et al. Efficient Bayesian
1012 mixed-model analysis increases association power in large cohorts. Nat Genet. 2015;47: 284–290.
1013 doi:10.1038/ng.3190

1014 87. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional
1015 visualization of genome-wide association scan results. Bioinformatics. 2010;26: 2336–2337.
1016 doi:10.1093/bioinformatics/btq419

1017 88. Myers RA, Scott NM, Gauderman WJ, Qiu W, Mathias RA, Romieu I, et al. Genome-wide interaction
1018 studies reveal sex-specific asthma risk alleles. Hum Mol Genet. 2014;23: 5251–5259.
1019 doi:10.1093/hmg/ddu222

1020 89. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science. 1996;273:
1021 1516–1517.

1022 90. Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. Genetics.
1023 1994;138: 963–971.

1024 91. Cheng R, Lim JE, Samocha KE, Sokoloff G, Abney M, Skol AD, et al. Genome-Wide Association Studies
1025 and the Problem of Relatedness Among Advanced Intercross Lines and Other Highly Recombinant
1026 Populations. Genetics. 2010;185: 1033–1044. doi:10.1534/genetics.110.116863

1027 92. Han B, Kang HM, Eskin E. Rapid and Accurate Multiple Testing Correction and Power Estimation for
1028 Millions of Correlated Markers. Storey JD, editor. PLoS Genet. 2009;5: e1000456.
1029 doi:10.1371/journal.pgen.1000456

1030 93. Han B, Eskin E. Interpreting Meta-Analyses of Genome-Wide Association Studies. Kerr K, editor. PLoS
1031 Genet. 2012;8: e1002555. doi:10.1371/journal.pgen.1002555

1032 94. Joo JWJ, Hormozdiari F, Han B, Eskin E. Multiple testing correction in linear mixed models. Genome
1033 Biol. 2016;17. doi:10.1186/s13059-016-0903-6

1034 95. Gauderman W, Morrison J. QUANTO documentation. (Technical report no. 157). Los Angeles, CA:
1035 Department of Preventive Medicine, University of Southern California; 2001.

1036 96. Gauderman WJ. Sample size requirements for association studies of gene-gene interaction. Am J
1037 Epidemiol. 2002;155: 478–484.

1038

1039

1040

1041

1042

1043

1044

1045

1046

1047

1048

Supporting information

S1 Fig. Correlation heatmap of PavCA metrics across days 1-5. The heatmap displays the absolute value of the Pearson correlation coefficient between each pair of metrics across all 5 days of testing.

S2 Fig. Distributions of the average of day 4 and day 5 measurements for 10 PavCA metrics. Histograms were constructed using measurements from the combined Harlan and Charles River sample set.

S1 Table. Sample origins for all 4,061 SD rats in final filtered set.

S3 Fig. Distributions of the average of day 4 and day 5 PavCA index scores for 6 major breeding locations. Sample numbers for each breeding location can be found in S1 Table. The three locations in the left column are from Harlan, and the three locations in the right column belong to Charles River.

S2 Table. List of variant filtering steps and the numbers of SNPs remaining after each step for both ANGSD/Beagle and STITCH. The values highlighted in green are the final SNP totals used for GWAS in Charles River and Harlan. The values highlighted in orange are the counts we used as criteria for choosing ANGSD/Beagle over STITCH.

S3 Table. Concordance and error rates for ANGSD/Beagle genotypes at different dosage r2 thresholds. Rates of concordance and genotyping error were calculated by comparing genotypes for 381 duplicate samples called in parallel. Each of the two replicates of the sample was assumed to contribute half the discordant genotypes. Therefore, the per sample genotyping error rate was calculated as half of the observed rate of discordance.

S4 Table. Pairwise F_{ST} estimates between vendor barrier facilities. Charles River – Kingston, NY and Harlan – Dublin, VA and Houston, TX were excluded due to low sample numbers.

S4 Fig. Heatmaps of pairwise identity-by-descent pre- and post-filtering. Panels A and C show pre-filtering values of P(IBD) = 0 plotted against P(IBD) = 1 for Harlan and Charles River. Unrelated samples cluster in the lower right corner. Samples along the diagonal have 2nd and 3rd degree levels of relatedness, while those clustering around (0.5,0.25) are full-siblings. Panels B and D demonstrate that the sample filtering steps removed several spurious relations from the sample.

S5 Table. List of sample filtering criteria and number of samples removed.

S6 Table. Heritability estimates in Charles River and Harlan for 85 mapped PavCA metrics/traits and body weight using the union set of 234,887 SNPs.

S7 Table. All associated loci from GWAS in Charles River and Harlan and the meta-analysis. Cells that are bold indicate the whether the association was made in Charles River, Harlan, and/or the meta-analysis.

S1 File. Stacked Manhattan plots for all analyzed metrics/traits. Each page contains stacked plots in the following order: Harlan 114k SNPs, Harlan 94k SNPs, Charles River 214k SNPs, Charles River 94k SNPs, and meta-analysis 94k SNPs.

S5 Fig. LocusZoom plot of genome-wide associated region containing *Cntn4*.

S8 Table. Factor loadings and percent variance explained by each factor for the first 5 PCs in each PCA analysis. Included are the loadings and PVEs of the PCA analyses of all 55 metrics, all metrics on each day, and all days for each metric for both Charles River and Harlan.

S2 File. Q-Q plots for all analyzed metrics/traits. Each page contains Q-Q plots for Harlan 114k SNPs, Charles River 214k SNPs, and meta-analysis 94k SNPs.

S6 Fig. Scree plots of the PVE for each of the top 10 PCs in the 55 metric PCA analysis. Panel A shows the PVE for Charles River and Panel B shows the PVE for Harlan.

S7 Fig. Power analysis curve for n=2,000 using Quanto.

S9 Table. List of covariates used for the GWAS LMMs for Harlan and Charles River. The first three covariates were used in both Charles River and Harlan analyses. The remaining covariates were unique to each population.

S10 Table. List of all PavCA metrics collected on SD rats. There are 11 total metrics across 5 days of training. The first 7 metrics are direct measurements made during the training periods. The following 3 metrics are calculated from the base measurements. The final metric, PavCA index score, is a composite score from the previous 3 metrics.

S8 Fig. Pre- and post-filtering distributions of heterozygosity for Harlan and Charles River. Panels A and C show pre-filtering distributions of heterozygosity in Harlan and Charles River, as measured by the method-of-moments F coefficient. Panels B and D show the same distributions post-filtering. A value above 0 indicates a deflation of heterozygosity, whereas a value below 0 would be an inflation.

S3 File. Spreadsheet with detailed sample sequencing and phenotype information.

S1 Text. ddGBS protocol used to sequence SD rats.

S4 File – VCF containing the filtered set of 214,309 Charles River SNPs.

S5 File – VCF containing the filtered set of 114,568 Harlan SNPs.