RatXcan: Framework for translating genetic results between species via transcriptome-wide association analyses

Natasha Santhanam[1†], Sandra Sanchez-Roige[2,3†], Yanyu Liang[1], Apurva S. Chitre[3], Daniel Munro[3, 4], Denghui Chen[3], Riyan Cheng[3], Festus Nyasimi[1], Margaret Perry[1], Jianjun Gao[3], Anthony M. George[5], Alex Gileta[3], Katie Holl[6], Alesa Hughson[7], Christopher P. King[8], Alexander C. Lamparelli[9], Connor D. Martin[10], Angel Garcia Martinez[11], Sabrina Mi[1], Celine L. St. Pierre[3], Jordan Tripi[11], Tengfei Wang[9], Hao Chen[9], Shelly Flagel[7], Keita Ishiwari[8,12], Paul Meyer[6], Laura Saba[13], Leah C. Solberg Woods[14], Oksana Polesskaya[3], Abraham A. Palmer[3,4*], Hae Kyung Im[1*]

***For correspondence:**
aap@ucsd.edu (AAP);
haky@uchicago.edu (HKI)

[†]These authors contributed equally to this work

Present address:
[§]Department of Psychiatry, University of California San Diego, La Jolla, CA, USA;
[¶]Genetic Medicine, University of Chicago, US

[1]Department of Medicine, Section of Genetic Medicine, The University of Chicago, Chicago, IL, 60637, USA; [2]Department of Medicine, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN, USA; [3]Department of Psychiatry, University of California San Diego, La Jolla, CA, 92093, USA; [4]Department of Integrative Structural and Computational Biology, Scripps Research, La Jolla, CA; [5]University at Buffalo, Clinical and Research Institute on Addictions University at Buffalo, Buffalo, NY, 14203, USA; [6]Medical College of Wisconsin, Department of Pediatrics, Milwaukee, WI, 53226, USA; [7]University of Michigan, Department of Psychiatry, Ann Arbor, MI, 48109, USA; [8]University at Buffalo, Department of Psychology, Buffalo, NY, 14260, USA; [9]Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, 92093, USA; [10]Department of psychology, University of California Los Angeles, Los Angeles, CA, 90095, USA; [11]University of Tennessee Health Science Center, Department of Pharmacology, Addiction Science and Toxicology, Memphis, TN, 38163, USA; [12]University at Buffalo, Pharmacology and Toxicology University at Buffalo, Buffalo, NY, 14203, USA; [13]University of Colorado Anschutz Medical Campus, Department of Pharmaceutical Sciences, Aurora, CO 80045, USA; [14]Wake Forest University School of Medicine, Department of Internal Medicine, Winston-Salem, NC, 27157, USA

Abstract We developed a framework for identifying trait-associated genes in rats and facilitating the transfer of polygenic evidence across species by expanding the transcriptome-wide association (TWAS) approach to rats. Our analysis successfully trained transcript predictors for over 8000 genes in each of the five brain regions of rats, revealing several shared properties of gene regulation with humans. Moreover, mirroring trends observed in humans, our findings showed that sparse predictors using variants in cis are more effective than polygenic predictors and that gene expression prediction in rats is highly correlated across brain regions. Importantly, our analysis also identified a significant overlap between genes associated with rat and human body length and BMI, indicating rat models may be useful for studying the genetic basis of complex traits in humans. RatXcan represents a valuable tool for uncovering shared biological mechanisms of complex traits across species, with potential applications in a wide range of research fields.

Introduction

Over the last decade, genome-wide association studies (GWAS) have identified numerous genetic loci that contribute to biomedically important traits [***Visscher et al., 2017***]. GWAS have demonstrated that most traits have a highly polygenic architecture, meaning that numerous genetic variants with individually small effects confer risk [***Loos, 2020***]. However, translating these results into meaningful biological discoveries remains extremely challenging [***Lewis and Vassos, 2020***, ***Martin et al., 2019***, ***Alliance et al., 2021***].

Model organisms provide a system in which the effect of genotype, genetic manipulations, and environmental exposures can be experimentally tested. Whereas the tools for using model organisms to study *individual* genes are well established, there are no satisfactory methods for studying the *polygenic* signals obtained from GWAS in model organisms.

To start addressing this problem, we extend the TWAS framework [***Gamazon et al., 2015***] to rats so that the unit of analysis are genes rather than rats. We call this approach RatXcan. Following our human pipeline, we investigate the genetic architecture of gene expression traits in rats and compare them to humans. Then, we train genetic predictors of gene expression traits in rats and perform association between the latter and rat body size traits.

Results

Experimental setup

To build a framework for translating genetic results between species, we followed the experimental setup illustrated in Fig. 1. In the *training stage* (Fig. 1a), we investigated the genetic architecture of gene expression and built prediction models of gene expression in rats. We used genotype and transcriptome data from five brain regions sampled from 88 heterogeneous stock (HS) rats, generated by the NIDA Center for GWAS for Outbred rats (Fig. 1a). We selected HS rats because they are a well characterized, outbred mammalian population for which dense genotype, phenotype, and gene expression data are available in thousands of subjects [***Solberg Woods and Palmer, 2019***, ***Chitre et al., 2020***, ***Keele et al., 2018***, ***Crouse et al., 2022***]. In the *association stage* (Fig. 1b), we used genotype data to predict the transcriptome in a non-overlapping *target set* of 3,407 rats. We tested for associations between the genetically predicted gene expression and body length by adapting the PrediXcan software, which was originally developed for use in humans [***Gamazon et al., 2015***], to rats ('RatXcan').

Genetic Architecture of Gene Expression across Brain Tissues

To inform the optimal prediction model training, we examined the genetic architecture of gene expression in HS rats by quantifying heritability and polygenicity

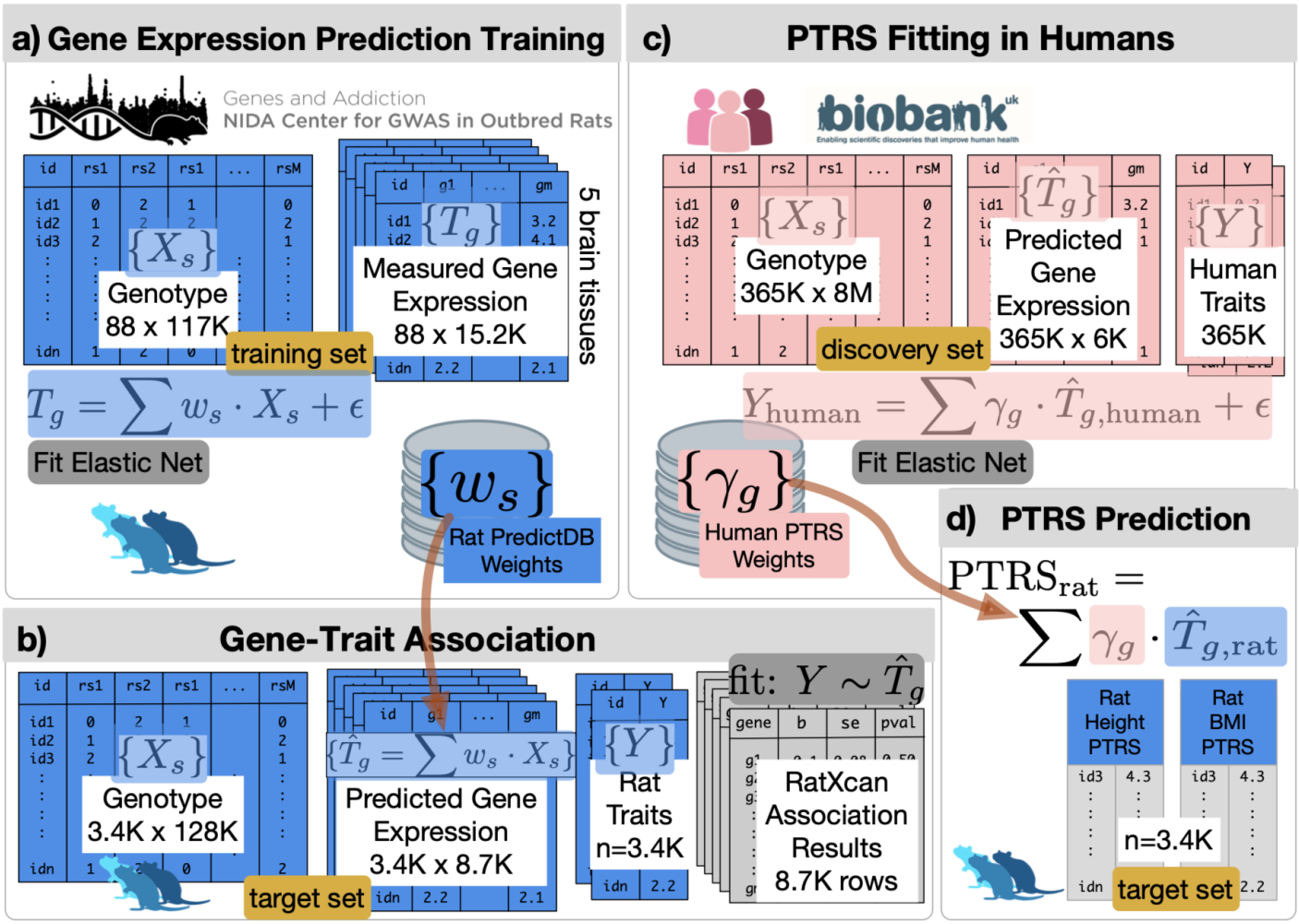

Figure 1. Schematic representation of cross-species polygenic translation framework.
The workflow was divided into 4 stages: a) gene expression prediction training, b) gene-trait association, c) PTRS fitting in humans, d) PTRS prediction. a) In the gene expression prediction training stage, we used genotype (117,155 SNPs) and gene expression data (15,216 genes) from samples derived from 5 brain regions in 88 rats. The prediction weights (rat PredictDB weights) are stored in predictdb.org. Rats used in this stage constitute the training set. b) In the gene-trait association stage, we used genotype and phenotype data from the target set of 3,407 rats (no overlap with training set rats). Predicted gene expression (8,567 genes for which prediction was possible) was calculated for all the 3,407 target set rats, and gene-trait associations were tested using RatXcan (N=1,463-3,110). We queried human gene-level associations from PhenomeXcan to estimate enrichment levels with our rat findings. c) Human PTRS weights were fitted using elastic net regression of height on predicted whole blood gene expression levels (7,002 genes) in the UK Biobank (N=356,476). d) The human PTRS weights will be used for complex trait prediction in rats. Prediction performance of PTRS will be used to calculate as the correlation (and partial correlation) between the predicted scores in rats and the observed traits. Analyses in rats are shown in blue and analyses in humans are shown in pink.

for five areas of brain tissue. Because the results for each tissue are similar, in the main text we summarize results for all tissues, highlighting the results for nucleus accumbens core; we present the remaining tissues in more detail in the supplement.

Brain Region	# Rats	# Genes Predicted	Average R^2	Average cis h^2
Nucleus Accumbens Core (NAcc)	78	8,567	8.51%	9.82%
Infralimbic Cortex (IL)	83	8,856	8.87%	9.77%
Lateral Habenula (LHb)	83	8,244	7.78%	8.86%
Prelimibic Cortex (PL)	81	8,315	9.33%	10.12%
Orbitofrontal Cortex (OFC)	82	8,821	9.13%	9.82%

Table 1. Summary of heritability and prediction performance in rats. The table shows the number of rats used in the prediction, number of genes predicted per model, the average predicion performance R^2, and average cis-heritability cis h^2, for all gene transcripts.

We calculated the heritability of expression for each gene by estimating the proportion of variance explained (PVE) using a Bayesian Sparse Linear Mixed Model (BSLMM) [***Zhou et al., 2013***]. We restricted the feature set to variants within 1 Mb of the transcription start site of each gene since this is expected to capture most cis-eQTLs. Among the 15,216 genes considered, 3,438 genes were heritable (defined as having a 95% credible set lower boundary greater than 1%) in the nucleus accumbens core. The mean heritability ranged from 8.86% to 10.12% for all brain tissues tested (Table 1). Fig. 2a shows the heritability estimates for gene expression in the nucleus accumbens core, while Fig. S1 shows heritability estimates for other tissues. We identified a similar heritability distribution in humans (Fig. 2b, Fig. S2) based on whole blood samples from GTEx.

Next, to evaluate the polygenicity of gene expression levels, we examined whether predictors with more polygenic or sparse architecture correlate better with observed expression. We fitted elastic net regression models using a range of mixing parameters from 0 to 1 (Fig. 2c). The leftmost parameter value of 0 corresponds to ridge regression, which is fully polygenic and uses all cis-variants. Larger values of the mixing parameters yield more sparse predictors, with the number of variants decreasing as the mixing parameter increases. The rightmost value of 1 corresponds to lasso regression, which yields the most sparse predictor within the elastic net family.

We used the 10-fold cross-validated Pearson correlation (R) between predicted and observed values as a measure of performance (Spearman correlation yielded similar results). We observed a substantial drop in performance towards the

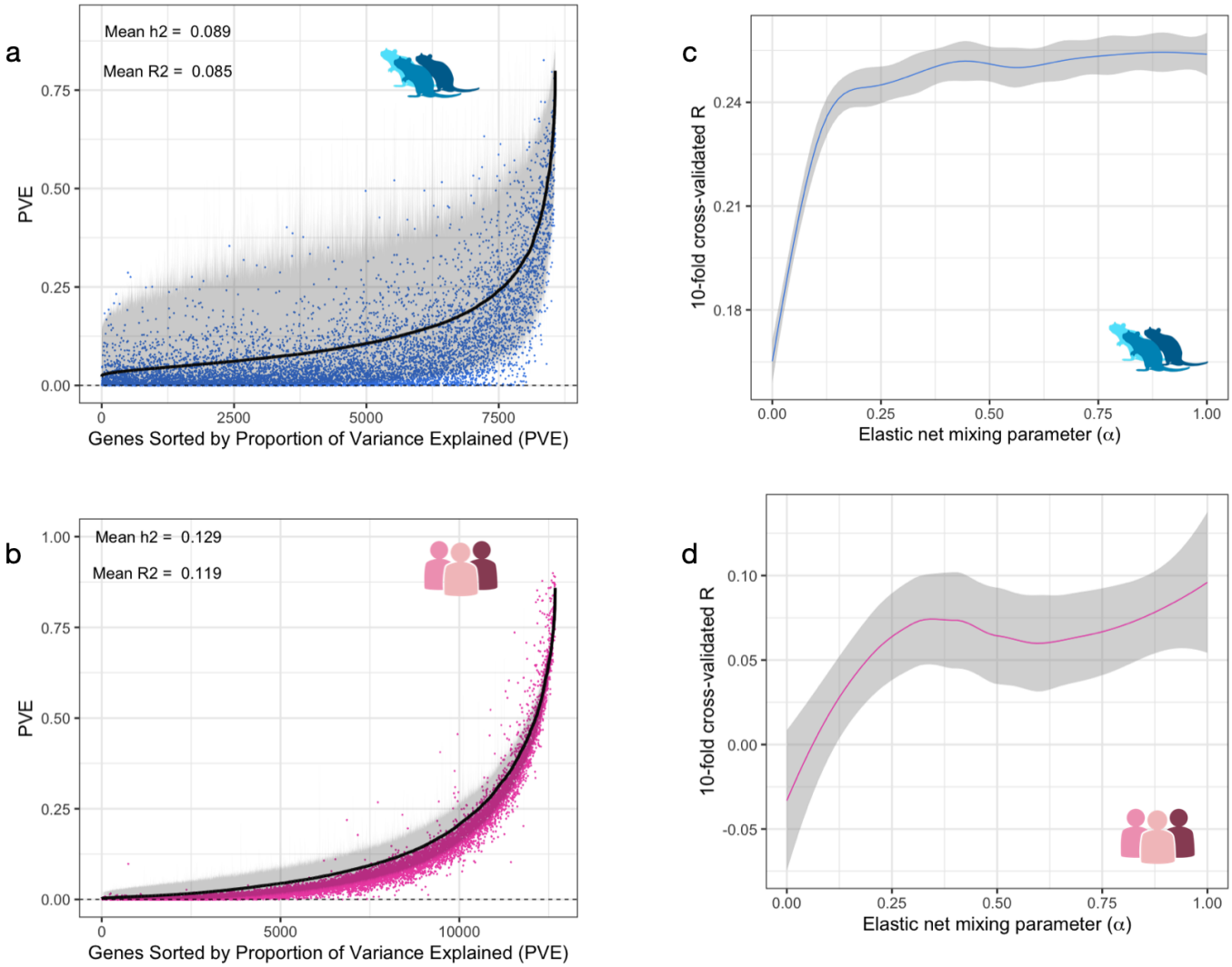

Figure 2. Heritability and sparsity of gene expression in both rats and humans. a) cis-heritability of gene expression levels in the nucleus accumbens core of rats calculated using BSLMM (black). We show only genes (N = 10,268) that have an equivalent ortholog in the GTEx population. On the x-axis, genes are ordered by their heritability estimates. 95% credible sets are shown in gray for each gene. Blue dots indicate the prediction performance (cross validated R^2 between predicted and observed expression). b) cis heritability of gene expression levels in whole blood tissue in humans from GTEx. We show only the same 10,268 orthologous genes. On the x-axis, genes are ordered by their heritability estimates. 95% credible sets are shown in gray for each gene. Pink dots indicate the prediction performance (cross validated R^2 between predicted and observed expression). c) Cross validated prediction performance in rats (Pearson correlation R) as a function of the elastic net parameter ranging from 0 to 1. d) Cross validated prediction performance in humans (Pearson correlation R) as a function of the elastic net parameter ranging from 0 to 1.

more polygenic end of the mixing parameter spectrum (Fig. 2c). We observed similar results using human gene expression data from whole blood samples in GTEx individuals (Fig. 2d). Overall, these results indicate that the genetic architecture of gene expression in HS rats (detectable with the currently available sample size) is sparse, similar to that of humans [***Wheeler et al., 2016***].

Generation of Prediction Models of Gene Expression in Rats

We trained elastic net predictors for all genes in all five brain regions. Based on the relative performance across different elastic net mixing parameters, we chose a parameter value of 0.5, which yielded slightly less sparse predictors than lasso but provided robustness to missing or low quality variants; this is the same value that we have chosen in the past for humans datasets [***Gamazon et al., 2015***]. The procedure yielded 8,244-8,856 genes across five brain tissues from the available 15,216 genes (Table 1). The 10-fold cross-validated prediction performance (R^2) ranged from 0 to 80% with a mean of 8.51% in the nucleus accumbens core. As shown in Table 1, mean prediction R^2 was consistently lower than mean heritability for all tissues, as is expected since genetic prediction performance is restricted by its heritability. Prediction performance values followed the heritability curve, confirming that genes with highly heritable expression tend to be better predicted than genes with low heritability in both HS rats and humans (Fig. 2a-b). Interestingly, we identified better prediction performance in HS rats than in humans (Fig. S3), despite heritability of gene expression being similar across species (Fig. 2a-b).

In Fig. 3a-b, we show the prediction performance of the best predicted genes in HS rats (*Mgmt*, $R^2 = 0.72$) and humans (*RPS26*, $R^2 = 0.74$). Across all genes, we found that the prediction performance in HS rats was correlated with that of humans ($R = 0.061$, $P = 8.03 * 10^{-6}$; Fig. 3c). Furthermore, performance per gene in different tissues was similar in both HS rats (Fig. 3d) and humans (Fig. 3e), namely, genes that were well-predicted in one tissue were also well-predicted in another tissue. Correlation of prediction performance across tissues ranged from 58 to 84% in HS rats and 42 to 69% in humans.

Having established the similarity of the genetic architecture of gene expression between rats and humans, we transitioned to the *association stage*.

PrediXcan/TWAS Implementation in Rats (RatXcan)

To extend the PrediXcan/TWAS framework to rats, we developed RatXcan. We used the predicted weights from the *training stage* to estimate the genetically regulated expression in the *target set* of 3,407 densely genotyped HS rats. We then tested the association between predicted expression and body length in the target set.

We identified 90 Bonferroni significant genes ($P(0.05/5388) = 9.28 \times 10^{-6}$) in 57

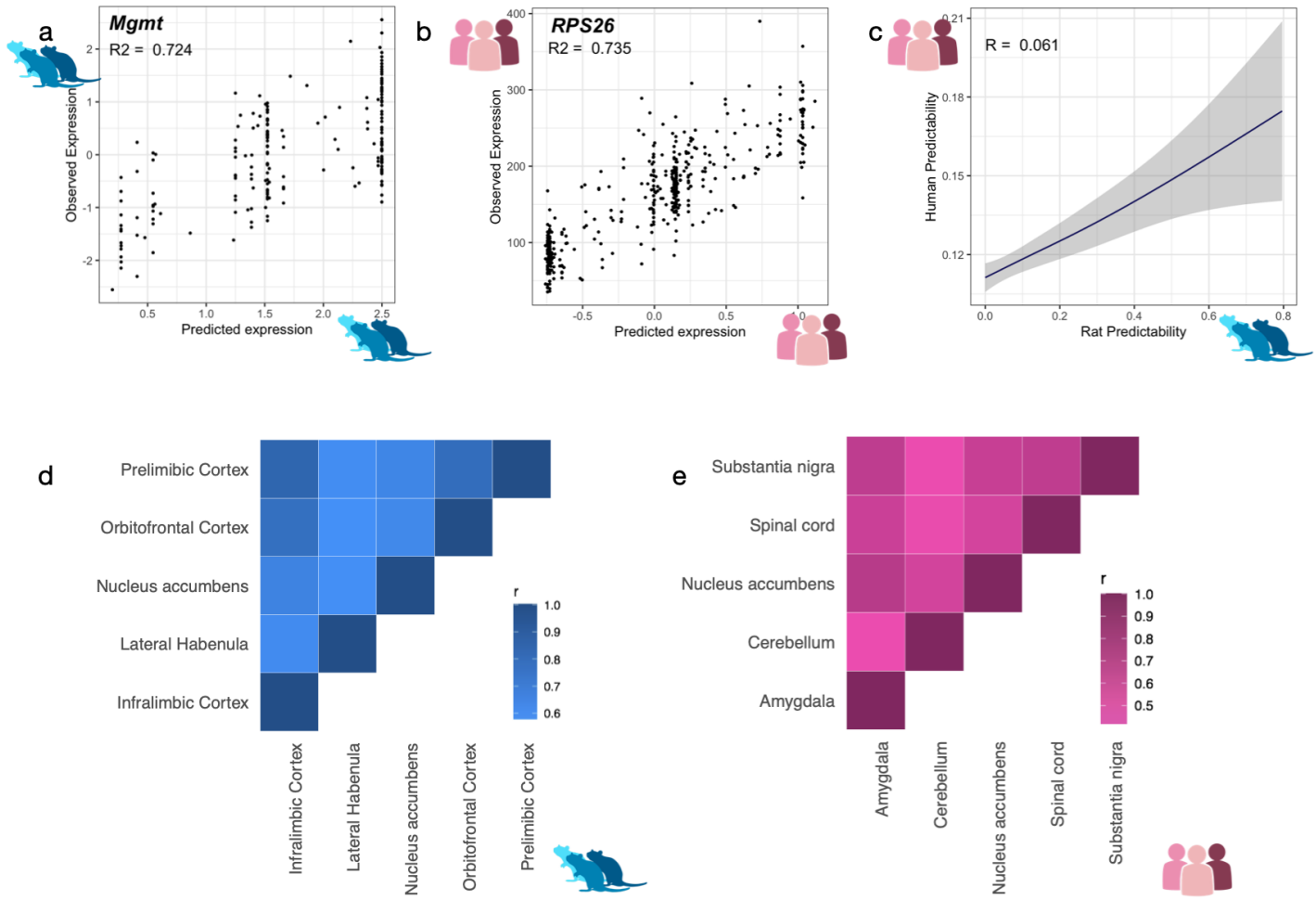

Figure 3. Shared genetic architecture of gene expression in rats and humans a) Comparison of predicted vs. observed expression for a well predicted gene in rats (*Mgmt*, $R^2 = 0.72$, $R = 0.65$, $P < 2.20 \times 10^{-16}$). b) In humans, predicted and observed expression for *RPS26* were significantly correlated ($R^2 = 0.74$, $R = 0.86$, $P < 2.20 \times 10^{-16}$). c) Prediction performance was significantly correlated across species ($R = 0.06$, $P = 8.03 \times 10^{-06}$) d-e) and across all five brain tissues tested in rats and humans. In rats, within tissue prediction performance ranged from ($R = [0.58 - 0.84]$, $P < 2.20 \times 10^{-16}$). In humans, the range was [$R = 0.42 - 0.69$, $P < 2.20 \times 10^{-16}$] .

distinct loci separated by ± 1 Mb for rat body length (Fig. 4a; Supplementary Table 1). Among the 90 significant genes, 30.46% had human orthologs previously associated with height in GWAS. For example, *Tgfa*, which is related to growth pathways, including epidermal growth factor, was associated with body length in rats ($P = 1.18 \times 10^{-9}$) and nominally associated with height in humans [***Comuzzie et al., 2012***] ($P = 8.00 \times 10^{-6}$). To evaluate whether trait-associated genes identified in HS rats were more significantly associated with the corresponding traits in humans, we performed enrichment analysis. Specifically, we selected genes that were nominally associated with HS rat body length ($P < 0.05$) and compared the p-value from the analogous human trait (height) against the background distribution of height-associated genes identified in GWAS. Given the large sample size of human height GWAS, we expected the p-values for of height-associated genes (shown in pink, Fig. 4b)to depart substantially from the identity line (in gray). The subset of genes that were associated with rat body length (in blue, Fig. 4b) showed a major departure from the background distribution, indicating that body-length genes in rats were more significantly associated with human height than expected. To quantify the enrichment, we compared the p-value distribution of all the genes with the distribution of the subset of genes that were nominally significantly associated with rat body length ($P = 6.55 \times 10^{-10}$).

Discussion

Overwhelming evidence demonstrates that most complex diseases are extremely polygenic; however, there is an unmet need for methods that translate polygenic results to other species.

A critical first step to achieve the transfer of polygenic scores is the development of RatXcan, which is the rat version of PrediXcan [***Gamazon et al., 2015***], a well-established statistical tool that is used in human genetics. We showed that the genetic architecture of gene expression in rats is broadly similar to humans: they are heritable, sparse, and the degree of heritability is preserved across tissues; some of these observations are consistent with another recent publication that mapped eQTLs in HS rats [***Munro et al., 2022***]. Interestingly, despite the smaller sample sizes used to train our prediction models, rats showed better prediction than humans. This might reflect the fact that HS rats have a preponderance of common alleles [***Chitre et al., 2020***] whereas humans have numerous rare alleles that influence gene expression but are difficult to capture in prediction models. The superior prediction may also reflect the longer haplotype blocks that are present in HS rats relative to humans [***Chitre et al., 2020***], which reduces the multiple testing burden when mapping cis-eQTLs and likely facilitates predictor training.

Using RatXcan, we tested gene-level associations of body length, which had

been previously measured in rats. We chose height because of the availability of large human GWAS , relatively large genotyped HS rat cohort in which body length was known, and relatively unambiguous similarity between humans height and rat body length. We found substantial enrichment of trait-associated genes among orthologous human trait-associated genes.

There are several limitations in the current study. The sample size of the reference transcriptome data in rats was limited. We would expect better predictability estimates in our elastic-net trained models with larger sample sizes. Furthermore, we used gene expression data from human blood and rat nucleus accumbens core because they were convenient datasets, but these tissues are not likely to be major mediators of height or body length. Second, we suspect that in both humans and rats, some gene-level associations may be confounded by linkage disequilibrium contamination and co-regulation. This problem is likely to be more serious in model organisms where even longer range LD exists. Finally, integration of other omic data types (e.g., protein, methylation, metabolomics) and the use of cell-specific data may improve prediction accuracy and cross-species portability. It is worth noting that while we have shown success with humans and HS rats, it is still not clear whether more distantly related species, such as non-mammalian vertebrates or even insects, might also lend themselves to ortholog analysis and ultimately a cross-species transciptome-based polygenic risk score.

Despite these limitations, we have developed a methodology for effectively and efficiently identifying orthologs between rats and humans, which should support new and transformatice experimental designs involving model organisms and enable the future development of a transcriptome-based polygenic risk score that is portable across species. Moreover, the RatXcan methodology provides a method to empirically validate traits that are intended to model or recapitulate aspects of human diseases in model systems. While the validity of these animal models has been a source of passionate debate, empirical evidence has been limited. Our polygenic approach provides a empirical approach to this debate that has been urgently needed.

Methods

Resource availability

Lead contact

Requests for further information, resources, and reagents should be directed to and will be fulfilled by one of the lead contacts, Hae Kyung Im (haky@uchicago.edu) or Abraham Palmer (aapalmer@ucsd.edu)

Material availability

This study did not generate new unique reagents.

Experimental model and subject details

The rats used for this study are part of a large multi-site project focused on genetic analysis of complex traits (www.ratgenes.org). N/NIH heterogeneous stock (HS) outbred rats are the most highly recombinant rat intercross available and are a powerful tool for genetic studies ([***Solberg Woods and Palmer, 2019***]; [***Chitre et al., 2020***]). HS rats were created in 1984 by interbreeding eight inbred rat strains (ACI/N, BN/SsN, BUF/N, F344/N, M520/N, MR/N, WKY/N and WN/N) and been maintained as an outbred population for almost 100 generations.

Method details

Genotype and expression data in the training rat set

For training the gene expression predictors, we used RNAseq and genotype data pre-processed for ***Munro et al.*** [***2022***]. We used 88 HS male and female adult rats, for which whole genome and RNA-sequencing information was available across five brain tissues [nucleus accumbens core (NAcc), infralimbic cortex (Il), prelimbic cortex (PL), orbitofrontal cortex (OFC), and lateral habenula (Lhb); Table 1]. Mean age was 85.7 ± 2.2 for males and 87.0 ± 3.8 for females. All rats were group housed under standard laboratory conditions and had not been through any previous experimental protocols. Genotypes were determined using genotyping-by-sequencing, as described previously in [***Parker et al., 2016***], [***Chitre et al., 2020***] and [***Gileta et al., 2020***]. Bulk RNA-sequencing was performed using Illumina HiSeq 4000 with polyA libraries, 100 bp single-end reads, and mean library size of 27M. Read alignment and gene expression quantification were performed using RSEM and counts were upper-quartile normalized, followed by additional quality-control filtering steps as described in ***Munro et al.*** [***2022***]. Gene-expression levels refer to transcript abundance for reads aligned to the gene's exons using the Ensembl Rat Transcriptome.

For each gene, we inverse normalized the TPM values to account for outliers and fit a normal distribution. We then performed PEER factor analysis [***Stegle et al., 2010***]. We regressed out sex, batch number, batch center and 7 PEER factors from the gene expression and saved the residuals for all downstream analyses.

Genotype and phenotype data in the target rat set

We used genotype and phenotype data from 3,407 HS rats (i.e., target set) reported in ***Chitre et al.*** [***2020***]. We used phenotypic information on body length (including tail), and fasting glucose. For each trait, sex, age, batch number and site were regressed out if they were significant and if they explained more than 2% of the variance, as described in [***Chitre et al., 2020***].

Querying human gene-trait association results

To retrieve analogous human gene–trait association results, we queried PhenomeXcan, a web-based tool that serves gene-level association results for 4,091 traits based on predicted expression in 49 GTEx tissues [***Pividori et al., 2020***]. Orthologous genes (N = 22,777) were mapped with Ensembl annotation, using the *biomart* R package and were one to one matched.

Estimating gene expression heritability

We calculated the cis-heritability of gene expression from the training set using a Bayesian sparse linear mixed model, BSLMM [***Zhou et al., 2013***], as implemented in GEMMA. We used variants within the ± 1Mb window up- and down-stream of the transcription start and end of each gene annotated by Gencode v26 [***Frankish et al., 2021***]. We used the proportion of variance explained (PVE) generated by GEMMA as the measure of cis-heritability of gene expression. We then display only the PVE estimates of 10,268 genes that were also present in the human gene expression data.

Heritability of human gene expression, which was also calculated with BSLMM, was downloaded from the database generated by ***Wheeler et al.*** [***2016***]. Genes were also limited to the same 10,268 as above.

Examining polygenicity versus sparsity of gene expression

To examine the polygenicity versus sparsity of gene expression in rats, we identified the optimal elastic net mixing parameter α, as described in ***Wheeler et al.*** [***2016***]. Briefly, we compared the prediction performance of a range of elastic net mixing parameters spanning from 0 to 1 (11 values from 0 to 1, with steps of 0.1). If the optimal mixing parameter was closer to 0, corresponding to ridge regression, we deemed gene expression trait to be polygenic. In contrast, if the optimal mixing parameter was closer to 1, corresponding to lasso, then the gene expression trait was considered to be more sparse. We also restricted the number of genes in the pipeline to the 10,268 orthologous genes.

Training gene expression prediction in rats

To train prediction models for gene expression in rats, we used the training set of 88 rats described above and followed the elastic net pipeline from predictdb.org. Briefly, for each gene, we fitted an elastic net regression using the *glmnet* package in R. We only included variants in the cis region (i.e., 1Mb up and downstream of the transcription start and end). The regression coefficient from the best penalty parameter (chosen via *glmnet*'s internal 10-fold cross validation [***Zou and Hastie, 2005***]) served as the weight for each gene. The calculated weights (w_s) are available in predictdb.org. For the comparison of number of predictable genes across species, we ran the same cross-validated elastic net pipeline in four GTEx tissues with sample sizes similar to that of the rats: Substantia Nigra, Kidney Cortex,

Uterus and Ovary. To ensure fair comparison, we used the same number of genes that were orthologous across all four human tissues and rat tissues.

Estimating overlap and enrichment of genes between rats and humans

For human transcriptome prediction used in the comparison with rats, we simply downloaded elastic net predictors trained in GTEx whole blood samples from the PredictDB portal, as previously done in humans [***Barbeira et al., 2021***]. This model was different from the ones used in the UK Biobank for calculating the PTRS weights (See Calculating PTRS in a rat target set).

We quantified the accuracy of the prediction models using a 10-fold cross validated correlation (R) and correlation squared (R^2) between predicted and observed gene expression [***Zou and Hastie, 2005***]. For the rat prediction models, we only included genes whose prediction performance was greater than 0.01 and had a non-negative correlation coefficient, as these genes were considered well predicted.

We tested the prediction performance of our elastic net model trained in nucleus accumbens core in an independent rat reference transcriptome set. We predicted expression in the reference set of 188 individuals and compared to observed genetic expression in the nucleus accumbens core.

Quantification and Statistical Analysis

Implementing RatXcan

We developed RatXcan, based on PrediXcan [***Gamazon et al., 2015***] [***Barbeira et al., 2018***] in humans. RatXcan uses the elastic net prediction models generated in the training set. In the prediction stage, we generated a predicted expression matrix for all genes in the rat target set, by fitting an additive genetic model:

$$Y_g = \sum_k w_{k,g} X_k + \epsilon$$

Y_g is the predicted expression of gene g, $w_{k,g}$ is the effect size of marker k for gene g, X_k is the number of reference alleles of marker k, and ϵ is the contribution of other factors that determine the predicted gene expression, assumed to be independent of the genetic component.

We then tested the association between the predicted expression matrix and body length. We fitted a linear regression of the phenotype on the predicted expression of each gene, which generated gene-level association results for all gene trait pairs.

Estimating overlap and enrichment of genes between rats and humans

We queried PhenomeXcan to identify genes associated with human height. PhenomeXcan provides gene-level associations aggregated across all available GTEx tissues, as calculated by MultiXcan (an extension of PrediXcan) [***Barbeira et al., 2019***]. To this aim, we adapted MultiXcan to similarly aggregate our results across

the 5 tested brain tissues in rats. We used a Q-Q plot to inspect the level of enrichment across rat and human findings. To quantify enrichment, we used a Mann-Whitney test as implemented in R to discern whether the distribution of the p-values for genes in humans was the same for the genes that were and were not nominally significant in rats.

Calculating PTRS weights in the UK Biobank

We calculated human-derived height PTRS weights using elastic net with a mixing parameter of 0.5, as described in ***Liang et al.*** [***2022***]. We predicted expression levels in 356,476 UK Biobank unrelated participants of European descent using whole blood prediction models trained in GTEx. We used the prediction models trained with UTMOST based on grouped lasso, which borrows information across tissues to improve prediction performance [***Barbeira et al., 2020***, ***Hu et al., 2019***]. The predicted expression was generated using high quality SNPs from Hapmap2 [***McCarthy et al., 2016***]. We performed elastic net regression with height as the predicted variable and the predicted expression matrix from 356,476 UK Biobank unrelated individuals of European descent. More specifically, for each regularization parameter λ, we selected weight parameters γ_g that minimized the mean squared difference between the predicted variable Y and prediction model $X\gamma+\gamma_0$ where $\hat{T}_g \in \mathbb{R}^{N\times 1}$ is the standardized predicted expression level of gene g across N individuals and $\hat{C}_l \in \mathbb{R}^{N\times 1}$ is the the observed value of the lth standardized covariate:

$$\gamma^{EN} = argmin_\gamma \overbrace{\frac{1}{N} \parallel Y - X\gamma - \gamma_0 \parallel_2^2}^{loss:l\gamma} + \lambda\alpha \parallel \gamma \parallel_1 + \lambda_a(1-\alpha)(\parallel \gamma \parallel)_2^2$$

$$X := [\hat{T}_1, ..., \hat{T}_m, C_1, ..., C_L]$$

where γ_0 is the intercept, m the number of genes, L is the number of covariates, $||B||_2^2$ is the l_2 norm and the $||B||_1$ is the l_1 norm of the effect size vector. α denotes the elastic net mixing parameter and λ is the regularization parameter. 37 different λ's were used, generating 37 different sets of predictors. Covariates included age at recruitment (Data-Field 21022), sex (Data-Field 31), and the first 20 genetic PCs. For more details, see ***Liang et al.*** [***2022***]. The values of the regularization parameters were chosen in a region likely to cover a wide range of sparsity in the resulting models, from very sparse, containing a couple of genes, to dense, containing all genes ***Liang et al.*** [***2022***].

Code and Data Availability

The code used for this work is available at https://github.com/hakyimlab/Rat_Genomics_Paper_Pipeline. Genotype and expression data are available through [***Munro et al., 2022***]. Prediction models for gene expression in all five brain tissues in rats are available at predictdb.org

Acknowledgments

This research has been conducted using the UK Biobank Resource under Application Number 19526. We thank Natalia Gonzales and Christian Jones for help editing the paper. The abstract's style was improved by using chatGPT iteratively. This work was partially supported by DP1DA054394 (SSR), P30DK020595 and R01CA242929 (HKI, NS, MP), P30DA044223 and R24 AA013162 (LS), P50DA037844 (AAP)

Author contributions

A.A.P. and H.K.I. conceived the cross species PTRS and supervised the work. N.S. and Y.L. performed a large portion of the analyses. N.S. and S.S-R. analyzed and interpreted the results and wrote the initial draft of the manuscript. MP and FN performed analysis of some of the PTRS results. S.M., D.M., A.C., D.C., L.S-W, and O.P. pre-processed and analyzed the RNAseq, genotype, and phenotype data. R.C., J.G., A.M.G., A.G., K.H., A.H., C.P.K., C.L.S-P., J.T., T.W., H.C., S.F., K.I., P.M., L.S. were involved in various aspects of the collection of the rat physiological traits. All authors read, edited and approved the final version of the manuscript.

Competing interests

The authors declare no conflict of interest.

Ethics declaration

Not applicable.

References

Alliance ICD, Adeyemo A, Balaconis MK, Darnes DR, Ripatti S, Widen E, Zhou A. Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps. Nature Medicine. 2021; 27(11):1876–1884.

Barbeira AN, Bonazzola R, Gamazon ER, Liang Y, Park Y, Kim-Hellmuth S, Wang G, Jiang Z, Zhou D, Hormozdiari F, et al. Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. Genome biology. 2021; 22(1):1–24.

Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, Torstenson ES, Shah KP, Garcia T, Edwards TL, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nature communications. 2018; 9(1):1–20.

Barbeira AN, Melia OJ, Liang Y, Bonazzola R, Wang G, Wheeler HE, Aguet F, Ardlie KG, Wen X, Im HK. Fine-mapping and QTL tissue-sharing information improves the reliability of causal gene identification. Genet Epidemiol. 2020 Sep; n/a(n/a).

400 **Barbeira AN**, Pividori M, Zheng J, Wheeler HE, Nicolae DL, Im HK. Integrating predicted
401 transcriptome from multiple tissues improves association detection. PLoS genetics.
402 2019; 15(1):e1007889.

403 **Chitre AS**, Polesskaya O, Holl K, Gao J, Cheng R, Bimschleger H, Garcia Martinez A, George
404 T, Gileta AF, Han W, et al. Genome-Wide Association Study in 3,173 Outbred Rats Iden-
405 tifies Multiple Loci for Body Weight, Adiposity, and Fasting Glucose. Obesity. 2020;
406 28(10):1964–1973.

407 **Comuzzie AG**, Cole SA, Laston SL, Voruganti VS, Haack K, Gibbs RA, Butte NF. Novel
408 genetic loci identified for the pathophysiology of childhood obesity in the Hispanic
409 population. PloS one. 2012; 7(12):e51954.

410 **Crouse WL**, Das SK, Le T, Keele G, Holl K, Seshie O, Craddock AL, Sharma NK, Comeau
411 ME, Langefeld CD, Hawkins GA, Mott R, Valdar W, Solberg Woods LC. Transcriptome-
412 wide analyses of adipose tissue in outbred rats reveal genetic regulatory mechanisms
413 relevant for human obesity. Physiological Genomics. 2022 Jun; 54(6):206–219. doi:
414 10.1152/physiolgenomics.00172.2021.

415 **Frankish A**, Diekhans M, Jungreis I, Lagarde J, Loveland JE, Mudge JM, Sisu C, Wright JC,
416 Armstrong J, Barnes I, et al. GENCODE 2021. Nucleic acids research. 2021; 49(D1):D916–
417 D923.

418 **Gamazon ER**, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, Eyler AE,
419 Denny JC, Nicolae DL, Cox NJ, et al. A gene-based association method for mapping
420 traits using reference transcriptome data. Nature genetics. 2015; 47(9):1091–1098.

421 **Gileta AF**, Gao J, Chitre AS, Bimschleger HV, St Pierre CL, Gopalakrishnan S, Palmer AA.
422 Adapting genotyping-by-sequencing and variant calling for heterogeneous stock rats.
423 G3: Genes, Genomes, Genetics. 2020; 10(7):2195–2205.

424 **Hu Y**, Li M, Lu Q, Weng H, Wang J, Zekavat SM, Yu Z, Li B, Gu J, Muchnik S, et al. A statistical
425 framework for cross-tissue transcriptome-wide association analysis. Nature genetics.
426 2019; 51(3):568–576.

427 **Keele GR**, Prokop JW, He H, Holl K, Littrell J, Deal A, Francic S, Cui L, Gatti DM, Broman KW,
428 Tschannen M, Tsaih SW, Zagloul M, Kim Y, Baur B, Fox J, Robinson M, Levy S, Flister MJ,
429 Mott R, et al. Genetic Fine-Mapping and Identification of Candidate Genes and Variants
430 for Adiposity Traits in Outbred Rats. Obesity (Silver Spring, Md). 2018 Jan; 26(1):213–
431 222. doi: 10.1002/oby.22075.

432 **Lewis CM**, Vassos E. Polygenic risk scores: from research tools to clinical instruments.
433 Genome medicine. 2020; 12(1):1–11.

434 **Liang Y**, Pividori M, Manichaikul A, Palmer AA, Cox NJ, Wheeler HE, Im HK. Polygenic
435 transcriptome risk scores (PTRS) can improve portability of polygenic risk scores across
436 ancestries. Genome Biol. 2022 Jan; 23(1):23.

437 **Loos RJ**. 15 years of genome-wide association studies and no signs of slowing down.
438 Nature Communications. 2020; 11(1):1–3.

Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nature genetics. 2019; 51(4):584.

McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, Kang HM, Fuchsberger C, Danecek P, Sharp K, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nature genetics. 2016; 48(10):1279.

Munro D, , Palmer A, Mohammadi P. The regulatory landscape of multiple brain regions in outbred heterogeneous stock rats. . 2022; .

Parker CC, Gopalakrishnan S, Carbonetto P, Gonzales NM, Leung E, Park YJ, Aryee E, Davis J, Blizard DA, Ackert-Bicknell CL, et al. Genome-wide association study of behavioral, physiological and gene expression traits in outbred CFW mice. Nature genetics. 2016; 48(8):919–926.

Pividori M, Rajagopal PS, Barbeira A, Liang Y, Melia O, Bastarache L, Park Y, Consortium G, Wen X, Im HK. PhenomeXcan: Mapping the genome to the phenome through the transcriptome. Science Advances. 2020; 6(37):eaba2083.

Solberg Woods LC, Palmer AA. Using heterogeneous stocks for fine-mapping genetically complex traits. Rat genomics. 2019; p. 233–247.

Stegle O, Parts L, Durbin R, Winn J. A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. PLoS computational biology. 2010; 6(5):e1000770.

Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. 10 years of GWAS discovery: biology, function, and translation. The American Journal of Human Genetics. 2017; 101(1):5–22.

Wheeler HE, Shah KP, Brenner J, Garcia T, Aquino-Michaels K, Consortium G, Cox NJ, Nicolae DL, Im HK. Survey of the heritability and sparse architecture of gene expression traits across human tissues. PLoS genetics. 2016; 12(11):e1006423.

Zhou X, Carbonetto P, Stephens M. Polygenic modeling with bayesian sparse linear mixed models. PLoS Genet. 2013 Feb; 9(2):e1003264–e1003264.

Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the royal statistical society: series B (statistical methodology). 2005; 67(2):301–320.

Supplementary information

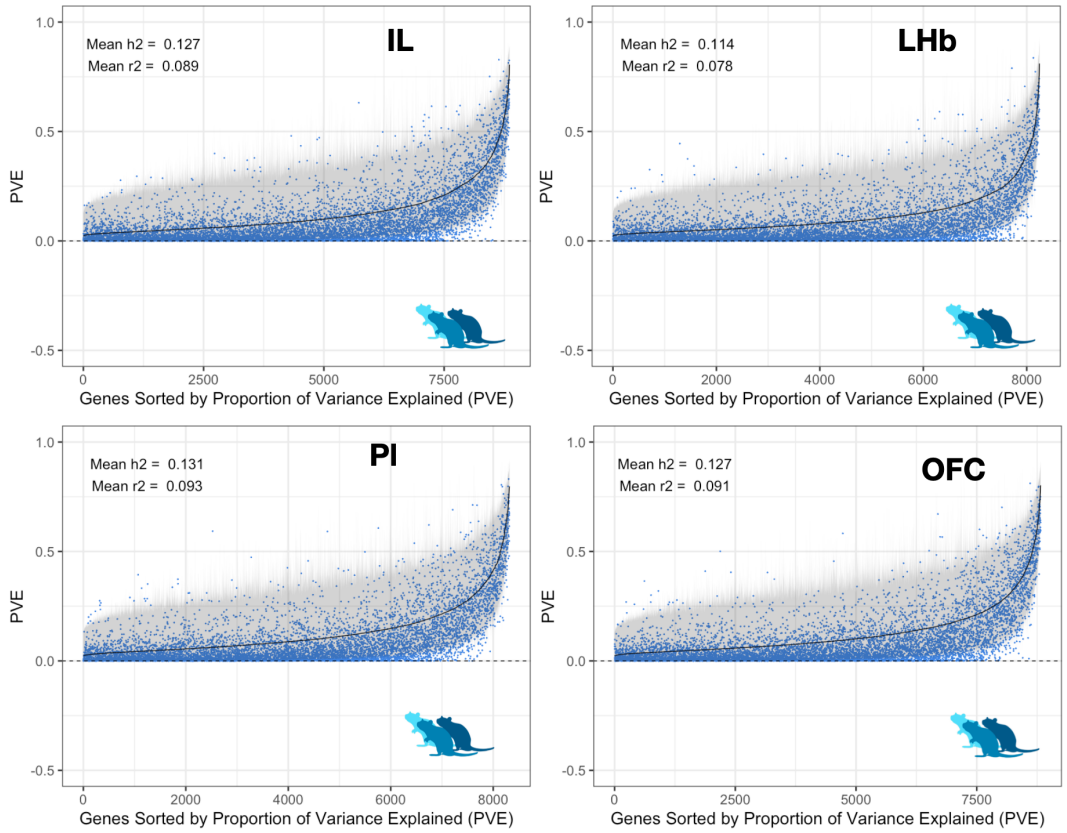

Figure S1. Gene expression was heritable [8.86-10.12%] and comparable across several brain tissues tested (Infralimbic Cortex, IL; Lateral Habenula, LHb; Prelimibic Cortex, PL; Orbitofrontal Cortex, OFC) in rats. We refer to heritability (h^2, cis-heritability within 1Mb) as the proportion of variance explained (PVE). Across all brain tissues tested, heritability estimates were significantly correlated ($R = [0.58 - 0.83]$, $P < 2.20 \times 10^{-16}$).

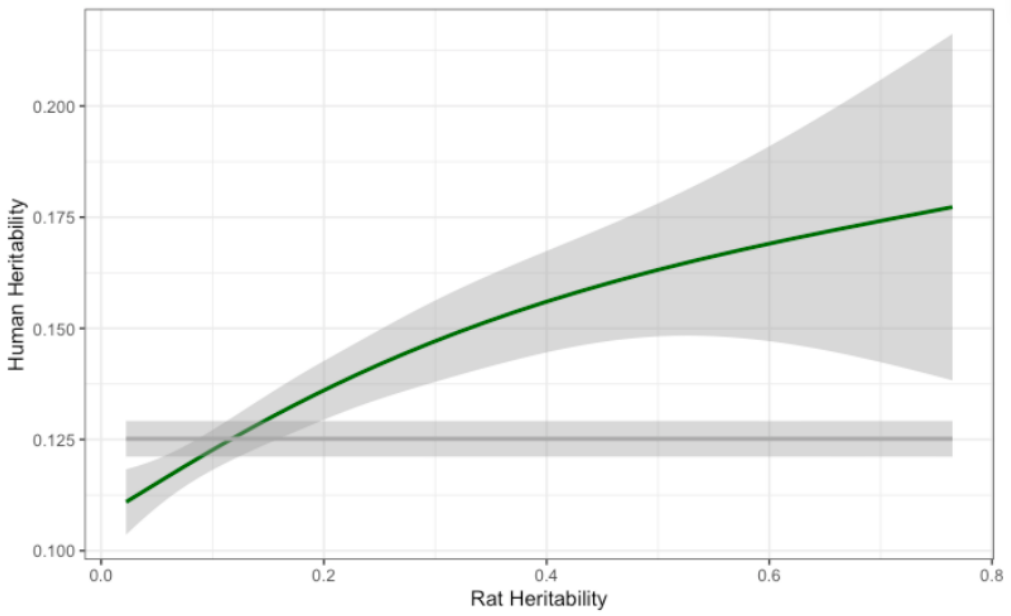

Figure S2. Heritability of gene expression was correlated between rats and humans. We found a significant correlation ($R = 0.07$, $P = 4.34 \times 10^{-12}$) between heritability estimates in rats and humans. Confidence intervals are represented as gray bars. The gray line represents the null distribution.

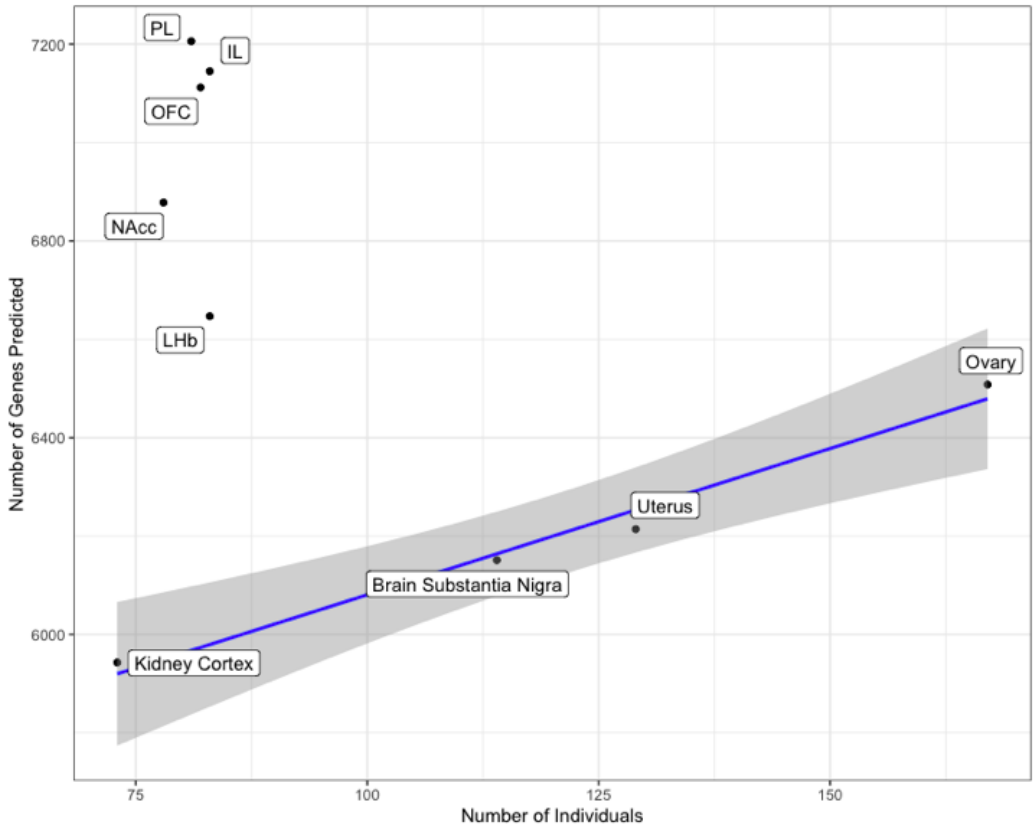

Figure S3. Prediction was greater in rat tissues than that in human GTEx tissues. The number of predicted genes across all five rat tissues was greater than those in GTEx human tissues with similar sample size. To ensure fair comparison, we included the same subset of genes that were orthologous across all tested tissues. Nucleus Accumbens Core (NAcc) Infralimbic Cortex (IL) Lateral Habenula (LHb) Prelimibic Cortex (PL) Orbitofrontal Cortex (OFC)

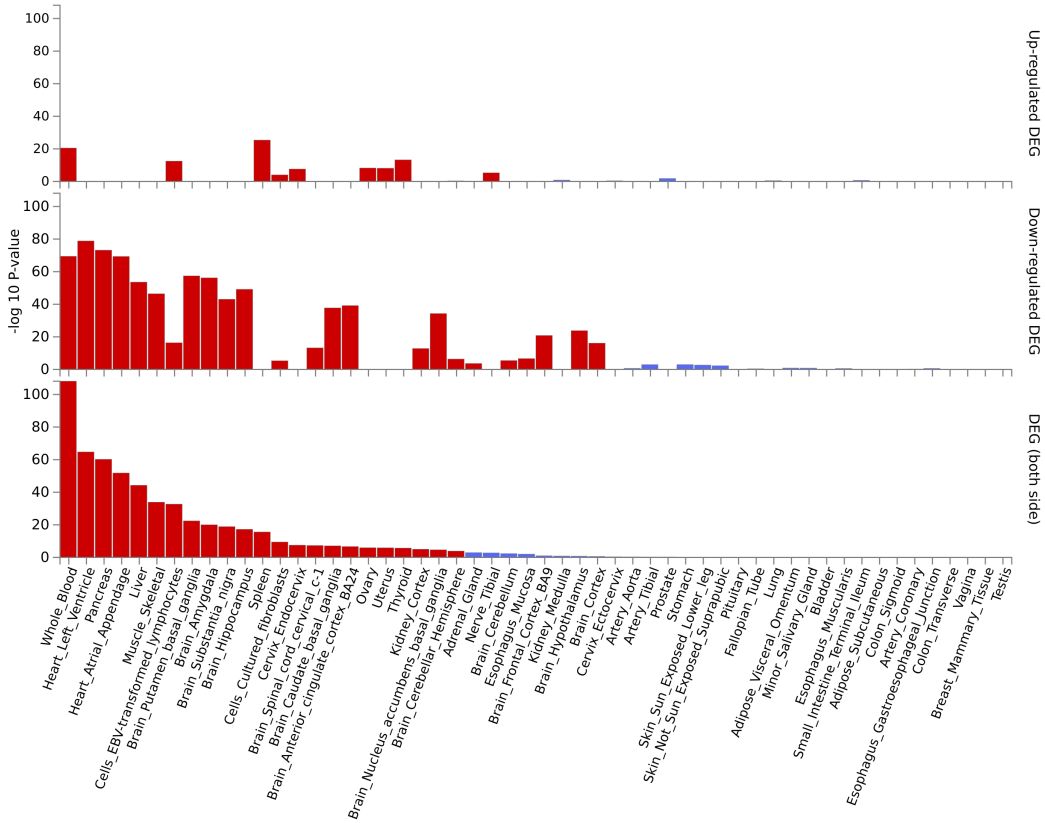

Figure S4. Tissue analysis revealed substantial enrichment in multiple relevant tissues, including heart, pancreas, muscle, liver, and central nervous system. Significantly enriched sets ($P < 0.05$) are highlighted in red.