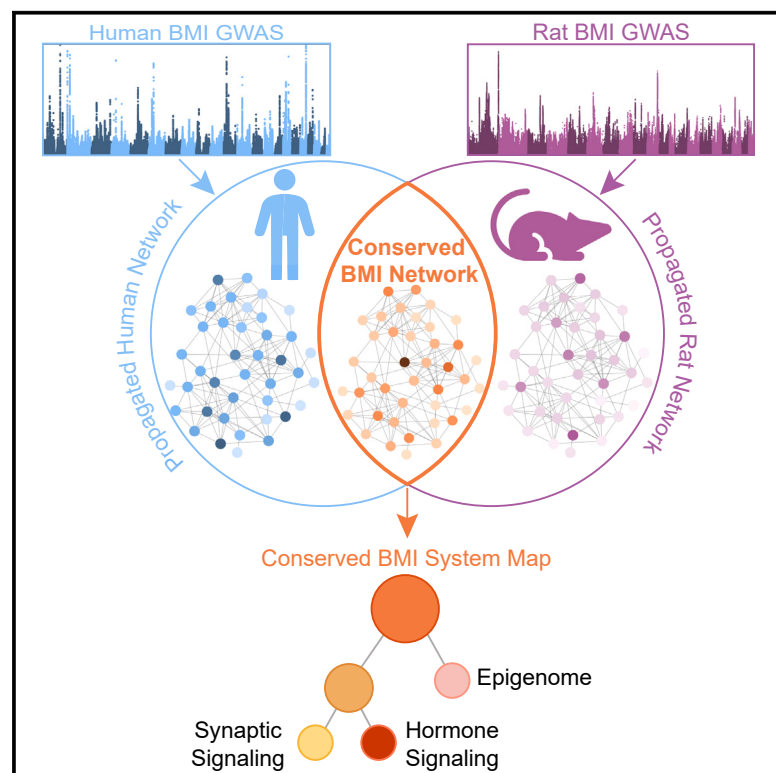


Genome-wide association studies of human and rat BMI converge on synapse, epigenome, and hormone signaling networks

Graphical abstract



Authors

Sarah N. Wright, Brittany S. Leger, Sara Brin Rosenthal, ..., Jason F. Kreisberg, Trey Ideker, Abraham A. Palmer

Correspondence

tideker@ucsd.edu (T.I.),
aap@ucsd.edu (A.A.P.)

In brief

Wright et al. identify a conserved molecular network underlying body mass index in humans and rats. Genes in this network, which are also associated with body-size phenotypes in mice, describe key processes including neuronal, epigenetic, and hormonal regulation. This work advances a general paradigm for cross-species GWAS translation.

Highlights

- Disparate genetics in humans and rats converge on a conserved molecular network
- A systems map of 27 interacting gene communities defines shared BMI biology
- Conserved BMI genes are linked to mechanisms of BMI via the Mouse Genome Database
- Neuronal pathways are a critical shared mechanism of BMI in humans and rodents



Article

Genome-wide association studies of human and rat BMI converge on synapse, epigenome, and hormone signaling networks

Sarah N. Wright,^{1,2,18} Brittany S. Leger,^{3,4,18} Sara Brin Rosenthal,⁵ Sophie N. Liu,¹ Tongqiu Jia,¹ Apurva S. Chitre,³ Oksana Poleskaya,³ Katie Holl,⁶ Jianjun Gao,³ Riyan Cheng,³ Angel Garcia Martinez,⁷ Anthony George,⁸ Alexander F. Gileta,^{3,9} Wenyan Han,⁷ Alesa H. Netzley,¹⁰ Christopher P. King,^{8,11} Alexander Lamparelli,¹¹ Connor Martin,^{8,11} Celine L. St. Pierre,¹² Tengfei Wang,⁷ Hannah Bimschleger,³ Jerry Richards,⁸ Keita Ishiwari,^{8,13} Hao Chen,⁷ Shelly B. Flagel,^{10,14} Paul Meyer,¹¹ Terry E. Robinson,¹⁵ Leah C. Solberg Woods,¹⁶ Jason F. Kreisberg,¹ Trey Ideker,^{1,17,19,*} and Abraham A. Palmer^{3,17,*}

¹Department of Medicine, University of California San Diego, La Jolla, CA 92093, USA

²Program in Bioinformatics and Systems Biology, University of California San Diego, La Jolla, CA 92093, USA

³Department of Psychiatry, University of California San Diego, La Jolla, CA 93093, USA

⁴Program in Biomedical Sciences, University of California San Diego, La Jolla, CA 93093, USA

⁵Center for Computational Biology & Bioinformatics, Department of Medicine, University of California, San Diego, La Jolla, CA 92093, USA

⁶Department of Physiology, Medical College of Wisconsin, Milwaukee, WI 53226, USA

⁷Department of Pharmacology, University of Tennessee Health Science Center, Memphis, TN 38163, USA

⁸Clinical and Research Institute on Addictions, University at Buffalo, Buffalo, NY 14203, USA

⁹Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA

¹⁰Department of Psychiatry, University of Michigan, Ann Arbor, MI 48109, USA

¹¹Department of Psychology, University at Buffalo, Buffalo, NY 14260, USA

¹²Department of Genetics, Washington University, St. Louis, MO 63110, USA

¹³Department of Pharmacology and Toxicology, University at Buffalo, Buffalo, NY 14203, USA

¹⁴Michigan Neuroscience Institute, University of Michigan, Ann Arbor, MI 48109, USA

¹⁵Department of Psychology, University of Michigan, Ann Arbor, MI 48109, USA

¹⁶Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA

¹⁷Institute for Genomic Medicine, University of California San Diego, La Jolla, CA 92093, USA

¹⁸These authors contributed equally

¹⁹Lead contact

*Correspondence: tideker@ucsd.edu (T.I.), aap@ucsd.edu (A.A.P.)

<https://doi.org/10.1016/j.celrep.2023.112873>

SUMMARY

A vexing observation in genome-wide association studies (GWASs) is that parallel analyses in different species may not identify orthologous genes. Here, we demonstrate that cross-species translation of GWASs can be greatly improved by an analysis of co-localization within molecular networks. Using body mass index (BMI) as an example, we show that the genes associated with BMI in humans lack significant agreement with those identified in rats. However, the networks interconnecting these genes show substantial overlap, highlighting common mechanisms including synaptic signaling, epigenetic modification, and hormonal regulation. Genetic perturbations within these networks cause abnormal BMI phenotypes in mice, too, supporting their broad conservation across mammals. Other mechanisms appear species specific, including carbohydrate biosynthesis (humans) and glycerolipid metabolism (rodents). Finally, network co-localization also identifies cross-species convergence for height/body length. This study advances a general paradigm for determining whether and how phenotypes measured in model species recapitulate human biology.

INTRODUCTION

Rodents and other model organisms provide unique advantages for investigating the molecular basis of human traits and diseases. However, the widespread assumption that disease biology is conserved across mammals is primarily based on conjecture and anecdotal evidence and remains controversial.^{1–3} Methods to rigorously assess the validity of animal

models are urgently needed, particularly for complex polygenic traits. One such trait is body mass index (BMI), which is used to diagnose obesity, a condition with a large and growing disease burden worldwide.

Complex traits are commonly examined by genome-wide association studies (GWASs), which have identified thousands of genetic loci underlying numerous diseases. Despite this success, synthesizing the numerous loci identified by GWASs into



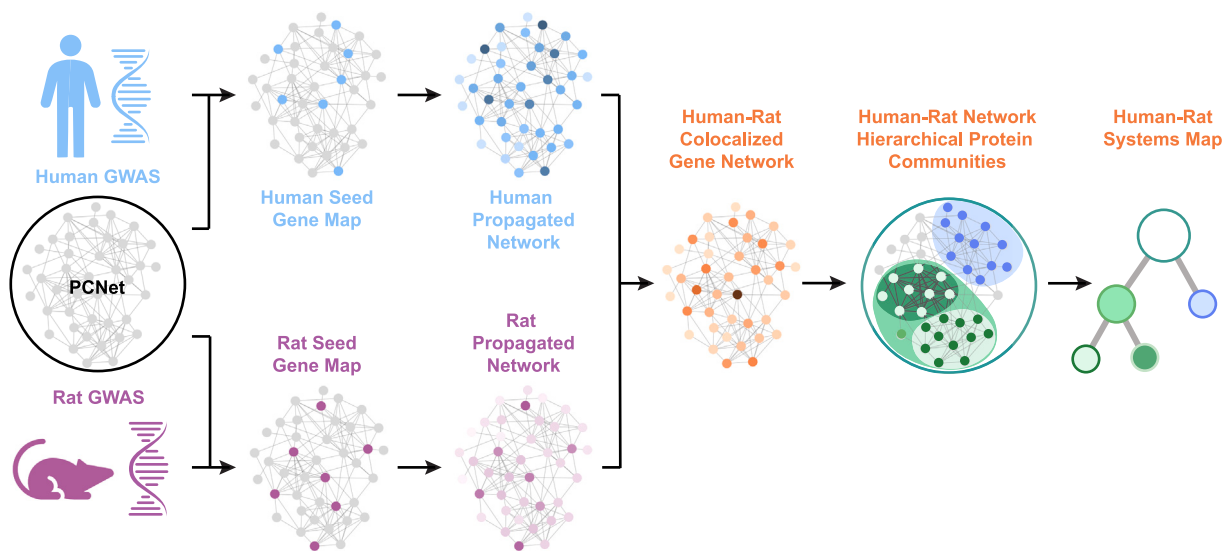


Figure 1. Generation of a cross-species conserved BMI systems map

GWAS results for human BMI (top, blue) and rat BMI (bottom, purple) are independently propagated through a molecular interaction network (PCNet) to measure the proximity of all genes to the input sets. Colocalized genes are identified as genes that are proximal in the network to BMI seed genes in both species (orange). Within this subset, nested communities of densely connected genes are identified to generate a hierarchy of systems representing the shared biology of BMI across humans and rats. A toy network is shown for illustrative purposes.

biologically interpretable results remains challenging.^{4,5} By design, the hypothesis-free nature of GWASs lends impartiality; on the other hand, knowledge about the functional relationships between genes has enormous potential to improve sensitivity and interpretability.

Model organism data are commonly used to provide a more holistic interpretation of GWAS results.⁶ For example, single-gene knockouts can address ambiguity about which variants are causal at a given locus (i.e., the fine-mapping problem), and integration of model organism omics datasets can clarify causal biological mechanisms.^{7,8} However, such approaches do not capture the polygenic architecture of complex traits. An attractive alternative is to perform GWASs in heterogeneous model organism populations, which is a complementary approach that recapitulates the polygenic nature of complex human traits. With this motivation, many GWASs have been performed in rats,^{9–12} mice,^{13–16} fruit flies,^{17–19} and other species.^{20–23}

In cases where similar traits have been studied via GWASs in both humans and model organisms, it is possible to ask whether the results implicate orthologous genes. Thus far, the concordance of genes implicated by separate GWASs in two species has typically been limited by incomplete power, differences in the variants that are common in each population, and many other factors. Furthermore, the most functionally similar gene in a model species can be a family or pathway member rather than the most sequence-similar ortholog.²⁴

One means of coping with these challenges is to leverage biological knowledge networks. Biological networks provide prior information on the interactions among genes and, therefore, aid the translation of implicated genes' mechanistic insights into relevant cellular processes, cell types, tissues, and dynamics.^{25–27} We and others have shown that integrating genetic

data with biological knowledge networks can improve the interpretability of findings,^{28–31} including via studies in model organisms.^{32,33} We recently expanded on these approaches by developing a network “co-localization” framework (NetColoc), which evaluates the convergence of two gene sets within a biological network. This approach was then used to identify convergent genetic circuits underlying autism spectrum disorder and congenital heart disease, which are comorbid.^{34,35}

Here, we explore the network co-localization approach for cross-species translation of GWAS results (Figure 1). As proof of concept, we use this approach to study BMI, a highly polygenic and medically important trait that has been extensively studied in both humans and rodent models.^{36,37} We show that, while the specific genes identified by human and rat GWASs do not show significant overlap, they converge on a conserved molecular interaction network. This conserved network provides insight into the shared etiology of BMI across mammals, and our approach establishes a general paradigm for assessing the extent to which an animal model does, or does not, recapitulate human biology.

RESULTS

GWASs associate different genes with BMI in humans and rats

We obtained genome-wide significant loci from a recent large human GWAS meta-analysis for BMI³⁸ ($n = 681,275$; Figure 2A). Positional mapping of these loci to the human genome identified 1,958 significant BMI-associated genes at a Bonferroni-corrected threshold of $p < 2.5 \times 10^{-6}$. In addition, we performed a separate GWAS for BMI in outbred heterogeneous stock (HS) rats, using data from a previous study,³⁷

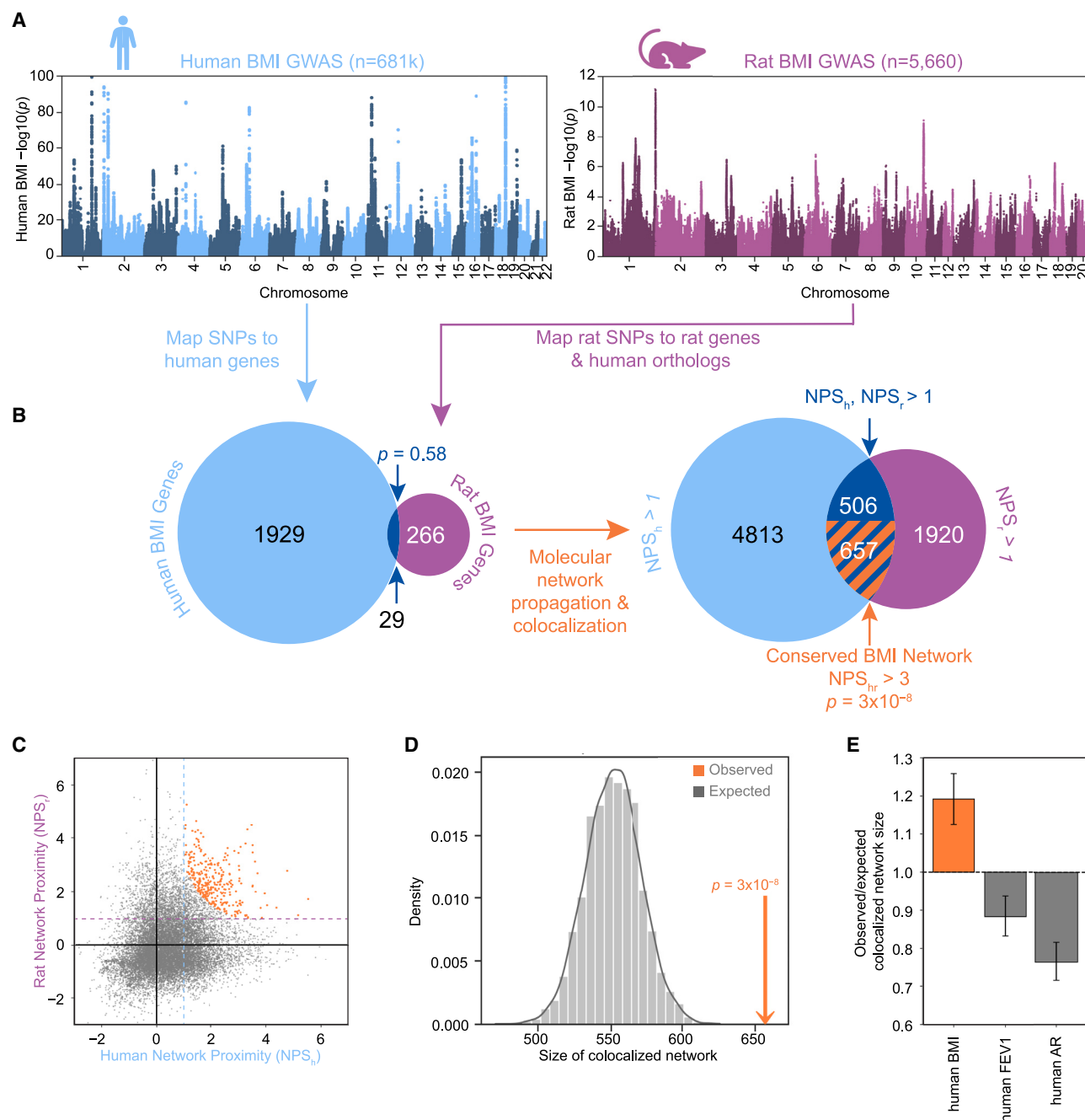


Figure 2. Convergence of rat and human BMI GWAS results at the gene and network level

(A) GWAS Manhattan plots for human BMI (left, light blue) with a maximum displayed $-\log(p) = 100$ and rat BMI (right, purple) with a maximum displayed $-\log(p) = 12$. (B) Left, Venn diagram representing the gene-level overlap of GWAS results after mapping to human genes; dark blue represents the 29 orthologous genes implicated in both species. Right, Venn diagram of genes passing network proximity score (NPS) thresholds after network co-localization. NPS_h > 1 (light blue); NPS_r > 1 (purple); NPS_h > 1 and NPS_r > 1 (dark blue); NPS_h > 1, NPS_r > 1, and NPS_{hr} > 3 (conserved network, orange stripes). Overlap p values were calculated via hypergeometric test. (C) Human and rat NPS for all non-seed genes, with genes passing all thresholds for the conserved BMI network (NPS_{hr} > 3, NPS_h > 1, and NPS_r > 1) shown in orange. Dotted horizontal and vertical lines indicate the single-species thresholds. (D) Observed (orange arrow) versus expected (gray distribution) size of the conserved BMI network following 10,000 permutations of NPS_r labels, with p value calculated via Z test. (E) The observed-to-expected ratio of colocated network size for rat BMI with human BMI (left orange bar) and two negative control comparisons (gray). FEV1, forced expiratory volume per second; AR, allergic rhinitis symptoms. Vertical bars indicate 95% confidence intervals. See also [Figure S2](#) and [Table S1](#).

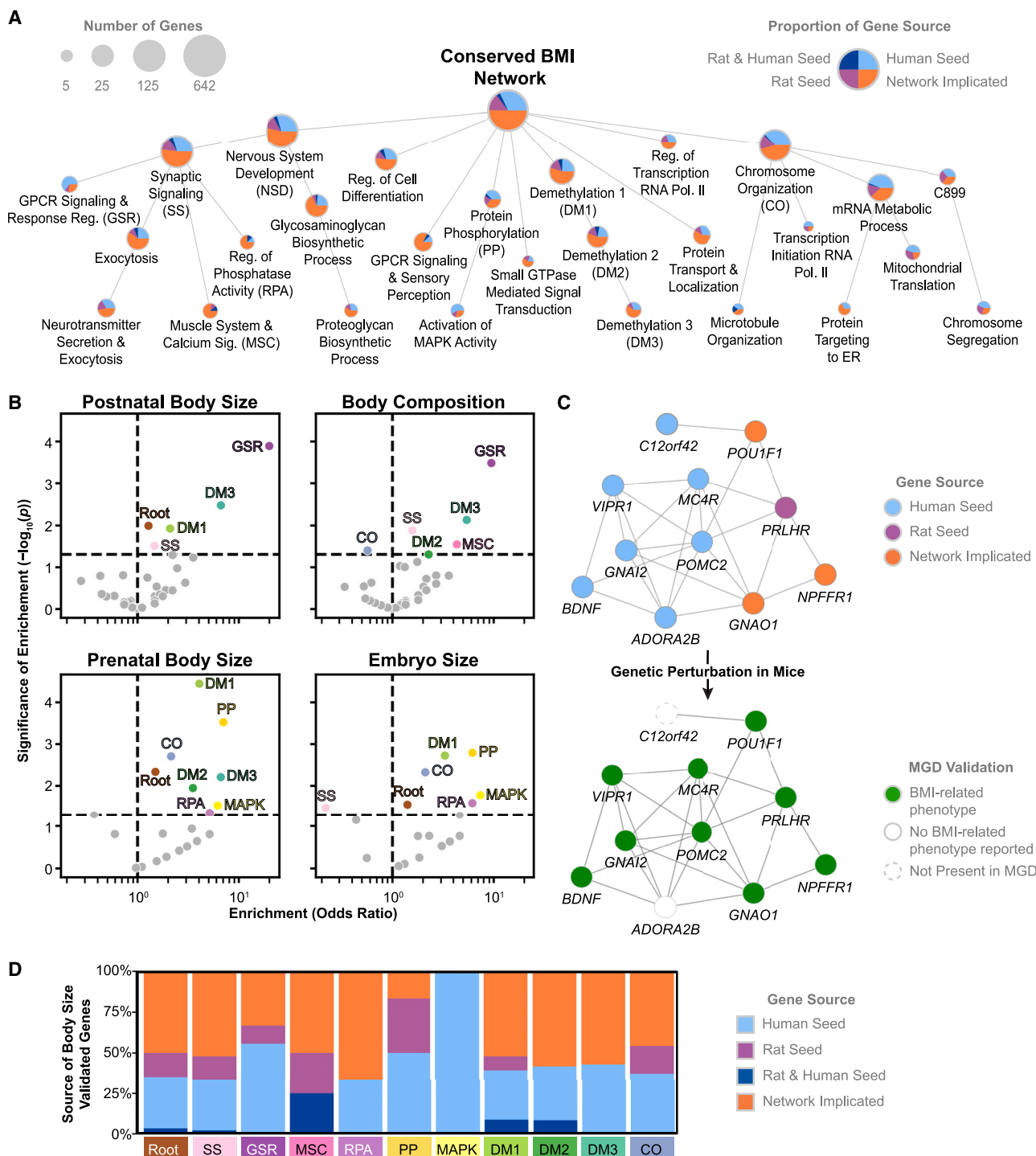


Figure 3. The conserved BMI systems map

(A) The conserved BMI network as a hierarchy of interacting gene communities. Nodes represent individual gene communities, with size indicating the number of genes (minimum five, maximum 642). Edges indicate smaller communities that are nested within larger communities. Pie charts indicate the fraction of community genes that are human seed genes, rat seed genes, seed genes in both species, or network-implied genes. Systems are annotated based on significantly associated Gene Ontology (GO) Biological Process terms. Only annotated communities and those necessary to connect annotated communities are shown.

(legend continued on next page)

together with an additional 2,487 previously unreported HS rats to obtain a single, outbred population ($n = 5,660$; [Figure 2A](#)). Outbred HS rats have larger linkage disequilibrium (LD) blocks compared to humans,³⁹ meaning that a GWAS represents a smaller number of independent tests. Therefore, we used a relaxed significance threshold ($p < 1.0 \times 10^{-4}$) for the rat GWAS, identifying 476 BMI-associated rat genes by positional mapping. Of these, 295 had unambiguous human orthologs ([Figure S1](#)). We define the significant BMI gene sets for humans and rats as “seed” genes and assume that both sets represent a mixture of true and false positives. Overall, we found 29 genes in common between the seed genes for humans and rats, a number that was not significantly different than expected by chance for gene sets of this size ([Figure 2B](#); $p = 0.58$, hypergeometric test).

A conserved molecular network underlies BMI across species

We next examined the molecular pathways in which the BMI genes function. For this purpose, we used the Parsimonious Composite Network (PCNet), a resource of ~ 2.7 million pairwise associations among human genes.⁴⁰ PCNet is formulated from a consensus of 21 physical and functional interaction databases and integrates multiple lines of evidence, including protein-protein interactions, mRNA and protein co-expression across tissues, literature curation, and other measures. We assigned a human BMI network proximity score (NPS_h) to each gene using a random walk algorithm⁴¹ to compute the normalized number of steps through the network to reach that gene from the seed set of 1,958 human BMI genes. We repeated this process using the 295 human orthologs of the rat BMI seed genes, resulting in a rat BMI network proximity score (NPS_r) for each gene. Finally, we calculated the product of the two proximity scores to compute $NPS_{hr} = NPS_h \times NPS_r$ ([Table S1](#)). In this way, genes with the highest NPS_{hr} tended to be close in the molecular network to BMI seed genes from both species, even if they were not seed genes in either species.

We found that NPS_{hr} values were significantly higher than expected for permuted sets of human and rat genes ($p = 1 \times 10^{-3}$; [Figure S2A](#)). We chose to threshold this score ($NPS_{hr} > 3$ and NPS_h and NPS_r , each > 1 ; [Figures 2B and 2C](#)) to create a highly co-localized subnetwork of PCNet that we called the “conserved BMI network.” This network included 657 genes, which was significantly more than expected by chance ([Figure 2D](#), $p = 3 \times 10^{-8}$). Thus, although the two BMI GWASs identified different genes in humans and rats, the molecular networks encoded by those genes show significant convergence. The 657-gene conserved BMI network covered 21 genes that were both human and rat seeds, with an additional 207 human seed genes and 92 rat seed genes. These genes represented the most interconnected

genes from the input sets. The remaining 337 conserved network genes were not directly identified by either the human or the rat GWAS but were instead implicated via network proximity.

To explore whether network convergence extends to other traits, we performed a similar network analysis of human height and its rodent analog, rat body length. These traits have been widely phenotyped in both species and are used to calculate BMI, so they are partially but not completely correlated with BMI.⁴² Unlike BMI, the GWAS for human height and the GWAS for rat body length identified a significant number of orthologous genes ($p = 6 \times 10^{-11}$; [Figure S2B](#)). However, consistent with our BMI results, we found that the molecular networks defined by human height genes and rat body length genes show substantially increased agreement, even when controlling for the number of shared genes ($p = 3 \times 10^{-20}$; [Figures S2C and S2D](#); [STAR Methods](#)).

We also examined the network co-localization of rat BMI genes with human GWAS results for four negative control traits that had SNP heritability similar to human BMI ($0.15 < h^2 < 0.35$) but were not genetically correlated with BMI or height (genetic correlation $|r_g| < 0.1$): allergic rhinitis symptoms, forced expiratory volume per second, smoking status (never smoked cigarettes), and presence of non-cancerous neoplasms. The rat BMI genes did not show significant network convergence with the genes identified by the negative control traits ([Figures 2E and S2E](#)). Because PCNet is a human interaction network, we also tested the network co-localization of the human and rat BMI seed genes within a molecular network consisting of 16,787 rat genes and 277,852 high-confidence rat interactions sourced from the STRING database.⁴³ We observed comparable network conservation with this rat network ([Figure S2E](#); $p = 2 \times 10^{-4}$; [STAR Methods](#)).

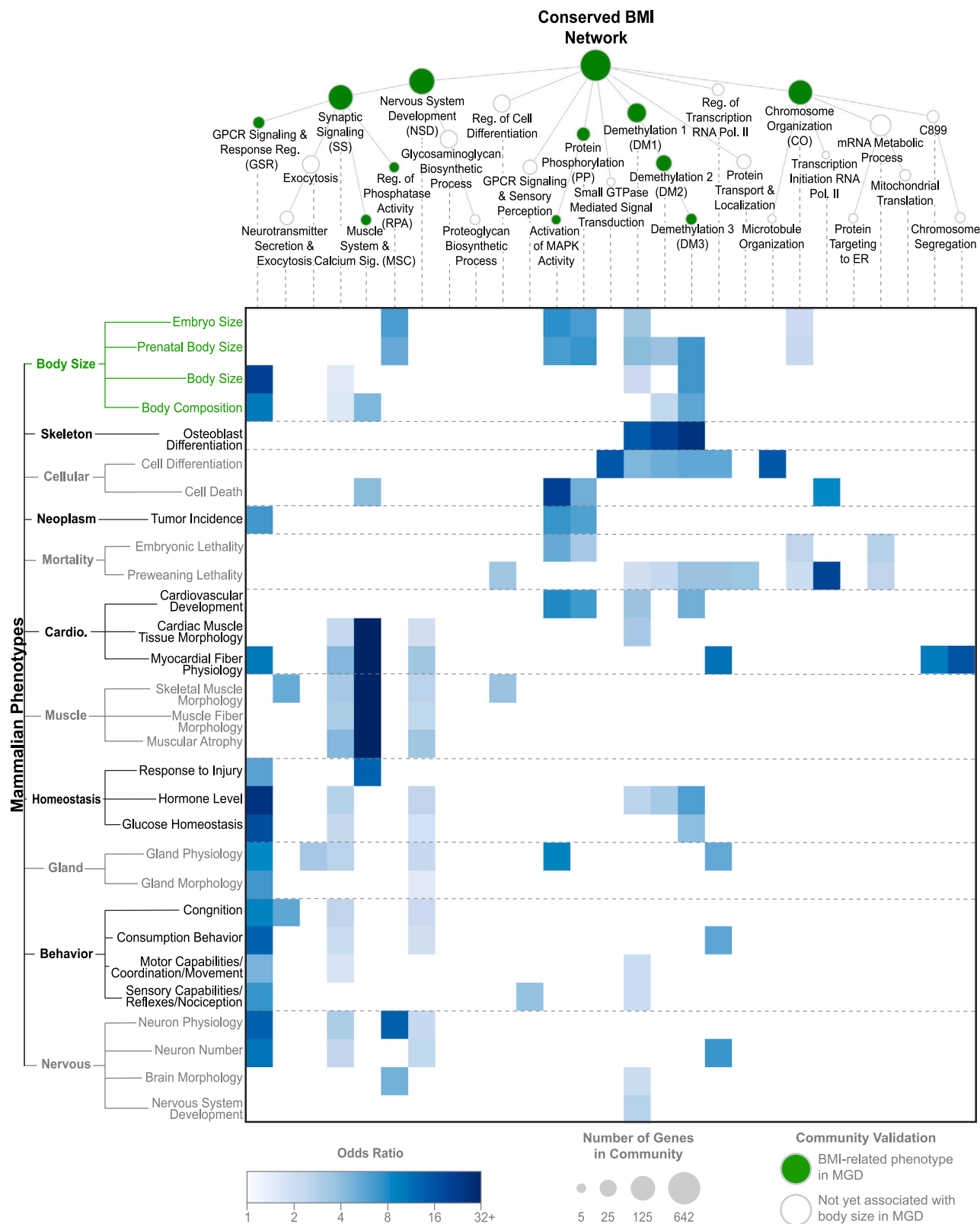
The conserved BMI network implicates distinct gene communities and functions

We observed that genes in the conserved BMI network were not uniformly connected but were organized within densely interacting communities ([Figure S3A](#)). To identify these, we used multi-scale community detection, which yielded a hierarchical map of 61 gene communities, which we term the “BMI systems map” ([Figure S3B](#)). This map contained the 642 genes (98%) that were part of the conserved BMI network’s largest connected component. We found that 27 communities in this map could be confidently annotated with enriched gene functions ([Figure 3A](#)) and that these functions were more numerous and specific than those identified for the entire conserved network or for sets of seed genes alone. We observed 331 significant annotations after filtering, compared to 45 for the conserved network, 59 for the human seed genes, and zero for the rat seed genes. Many of the annotations identified for the seed genes were

(B) Enrichment of each annotated community in (A) for genes associated with four BMI-related phenotypes (defined as postnatal body size, prenatal body size, embryo size, and body composition) in the Mouse Genome Database (MGD). Colored points show the nominally significantly enriched communities (Fisher’s exact test), with all other communities in light gray.

(C) The GPCR signaling and response regulation community (GSR) subnetwork. Node color indicates the source of genes in the subnetwork (top) and associations with any of the four BMI-related phenotypes in mice (bottom).

(D) Sources of all MGD-validated genes for communities significantly enriched for body-size-associated genes in (B). MGD-validated genes were defined as genes linked to at least one BMI-related phenotype in MGD. See also [Figures S3, S6A](#), and [Table S1](#).



(legend on next page)

captured via the annotated BMI systems map, particularly those relating to synaptic signaling and cellular morphogenesis. However, the hierarchical systems map identified additional and more specific functions, such as catecholamine synthesis and transport, demethylation, and monoamine transport (Table S2). The largest branch of the BMI systems map (by the number of genes covered) centered on nervous system development and physiology, with distinct sub-communities including G-protein-coupled receptor (GPCR) signaling, neurotransmitter secretion, and muscle development and function. A second major branch concerned nucleic acid and chromosomal processes, including mRNA transcription and protein translation. The remaining branches included functions such as chromatin regulation via demethylation, cell differentiation, and communities of mitogen-activated protein (MAP) kinases.

To validate this BMI systems map, we consulted the extensive catalog of mouse mutation experiments maintained by the Mouse Genome Database⁴⁴ (MGD). MGD records nearly 364,000 various phenotypic changes after whole-body or tissue-specific genetic perturbations in mice, which we used to assess the functional role of genes within the BMI systems map. Focusing on body size and composition phenotypes, we found that mouse orthologs of 39% of genes within the conserved BMI network had been associated with at least one body-size phenotype in mice. We observed that these validated genes were concentrated within particular branches of the BMI systems map, spanning 14 distinct communities in which disruption to community genes leads to MGD phenotypes related to body size or composition in mice (Figure 3B). Within these branches, we observed stronger enrichments for smaller, more specific systems, and, across the hierarchy, we found differentiation between communities associated with postnatal and prenatal body-size phenotypes.

The strongest enrichment for postnatal body size and composition in mice was observed for genes in the GPCR signaling and response regulation community (GSR, $n = 11$ genes), where disruptions to nine genes led to BMI-related phenotypes in mice (Figure 3C). The MGD-validated genes included some that have been extensively studied in the context of BMI, such as genes for brain-derived neurotrophic factor *BDNF*, melanocortin receptor *MC4R*, and the hormone precursor proopiomelanocortin *POMC*. While these genes were identified in the human BMI GWAS, they lacked equivalent associations in the rat GWAS. In contrast, the MGD-validated prolactin-releasing hormone receptor (*PRLHR*) was not associated with human BMI, but its rat ortholog (*Prhr*) was associated with BMI in the rat GWAS ($p = 1 \times 10^{-12}$). Beyond seed genes, we observed three network-implicated genes that were validated in MGD. For example, neither *NPFFR1* nor its rat ortholog *Npffr1* were directly identified by their respective GWAS, but *NPFFR1* has been associated with insulin-like growth factor 1 (IGF-1) levels in humans,⁴⁵ and constitutive inactivation of *Npffr1* has been associated with

increased susceptibility to weight loss in mice.⁴⁶ As with the GSR community, we observed that network-implicated genes substantially contributed to body-size enrichments throughout the BMI systems map, illustrating the power of our network approach to connect and expand the results of human and rat BMI GWASs (Figures 3D and S3C).

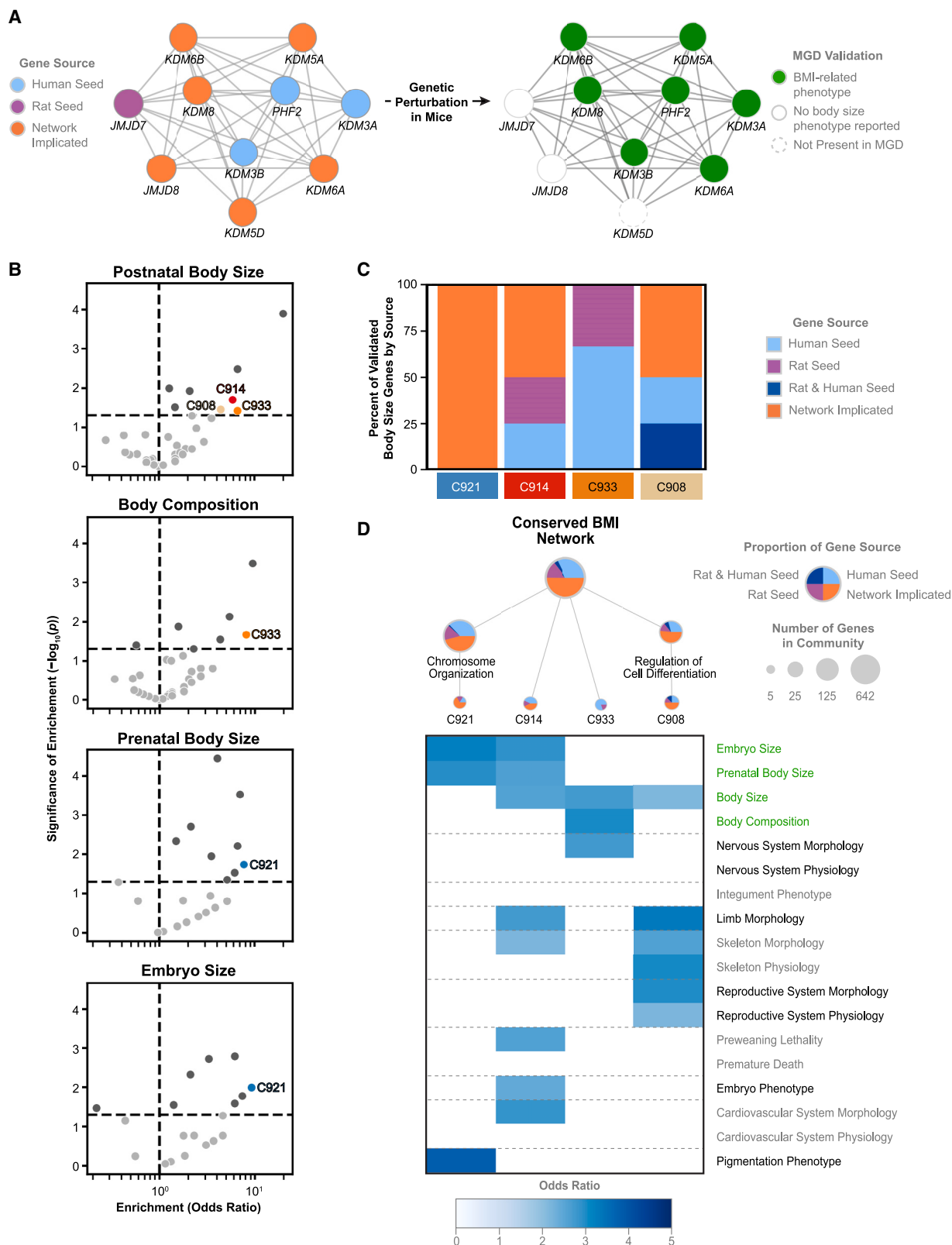
In addition to body-size phenotypes, genetic disruptions in specific communities were enriched for related tissue and behavioral phenotypes (Figure 4). For example, disruptions to genes in the GSR community led to changes in mouse hormone levels (odds ratio [OR] = 25, $p = 3 \times 10^{-7}$), glucose homeostasis (OR = 16, $p = 8 \times 10^{-6}$), consumption behavior (OR = 13, $p = 5 \times 10^{-5}$), neuron physiology (OR = 13, $p = 6 \times 10^{-5}$), and other relevant mouse traits. Given that these same disruptions also affect postnatal body size and composition, the GSR community provides a conserved pathway by which diverse genetic alterations cause key hormone and behavioral changes that, in turn, give rise to gross morphological changes affecting overall body size and mass.

Other notable BMI systems included muscle system and calcium signaling (MSC; Figures 4 and S4A), in which genetic perturbations affect muscle morphology leading to abnormal body size and composition, as well as chromatin regulation via demethylation (DM) communities (Figure S4B), in which mutations cause defects in skeletal and cellular differentiation and changes to hormone levels. Notably, at least 50% of the MGD-validated genes in these communities were implicated by the network analysis rather than by the initial human and rat GWASs. For example, the MSC community successfully recovered the X chromosome gene *DMD*, named for its role in Duchenne muscular dystrophy, a condition with a high prevalence of comorbid obesity.⁴⁷ *DMD* and *Dmd* were not identified by either the human or rat GWAS due to the exclusion of sex chromosomes, but *DMD* has been previously linked to BMI in humans and weight changes in mice.^{48–50} The core DM community (DM3) included distinct demethylase genes identified by the human and rat GWASs, as well as four network-implicated demethylases (KDM5A, KDM6A, KDM6B, and KDM8), which have been linked to body-size phenotypes in mice^{51–56} (Figure 5A).

Finally, we observed four communities that did not have clear associations with known biological functions but showed enrichment for genes associated with BMI-relevant phenotypes in mice (Figure 5B); these communities had contributions from both seed and network-implicated genes (Figures 5C and S3C). For example, the C914 community ($n = 7$; Figure S4C) contained seed genes axin-1 (*AXIN1*) and hedgehog acyltransferase (*HHAT*), as well as network-implicated genes plectin (*PLEC*) and the SKI-like proto-oncogene (*SKIL*). Disruptions to their mouse orthologs *Axin1*, *Hhat*, *Plec*, and *Skil* led to changes in prenatal and postnatal body size, and the gene community was linked to developmental effects on skeletal, nervous, and cardiovascular system morphology in mice (Figure 5D). We observed further

Figure 4. Functional effects of perturbations to conserved BMI systems in mice

Reproduction of the conserved BMI systems map (top) with gene communities significantly enriched for mouse BMI-related traits in green ($p < 0.05$, Fisher's exact test). Heatmap (bottom) shows the enrichment of communities within the system map for body-size and non-body-size traits in mice, selected to capture the spectrum of phenotypes affected by the systems map (STAR Methods). The OR is shown for any community-phenotype associations that were nominally significant ($p < 0.05$) via a Fisher's exact test. Phenotypes are grouped based on the Mammalian Phenotype Ontology (left). See also Table S1.



(legend on next page)

impacts on skeletal and limb morphology in mice via disruptions to genes in community C908 ($n = 8$; [Figure S4D](#)), which was nested within a larger community involved in regulating cellular differentiation. C908 includes four SOX transcription factors: two seed genes (SOX5, SOX6) and two network-implicated genes (SOX3, SOX8). SOX transcription factors have been associated with hyperinsulinemia in obese mice.^{57,58} Increased expression of SOX6 may predispose individuals to obesity through the promotion of adipogenesis,⁵⁹ and Sox8-deficient mice have significant weight reduction in adulthood due to adipose tissue degeneration.

Divergent BMI mechanisms across species

Apart from the conserved BMI network, we also identified non-overlapping BMI networks distinct to each species. We used the previously calculated NPS_h and NPS_r values to find genes highly proximal to seed genes of one species but not the other ([Figures 6A and S5](#)), yielding human-specific ($n = 925$ genes) and rat-specific ($n = 688$ genes) networks ([Figure 6B](#)). While the conserved BMI network was significantly enriched for genes expressed in human and mouse brain tissue ([Figures 6C and 6D](#)), no brain specificity was observed for the species-specific networks, indicating that these networks are functionally distinct from the conserved BMI network. The rat-specific network was most strongly enriched for orthologs expressed in human adipose tissue ($q = 1 \times 10^{-3}$), while the human-specific network showed little tissue specificity across human and mouse tissues. In contrast, the human-specific network contained a significant number of genes associated with body-size phenotypes in mice, while the rat-specific network did not ($p_{\text{human}} = 0.02$, $p_{\text{rat}} = 0.82$, [Figure S6A](#)). The MGD-validation rate for the human-specific network was comparable to the conserved network ($p = 0.02$), which had the highest validation rates for human and rat seed genes, then rat seed genes, then network-implicated genes ([Figure S6A](#)).

To investigate potential species-specific BMI-mediating pathways, we created an expanded cross-species BMI network, including all genes with high proximity to human or rat GWAS seed genes ([Figure 6A](#)). We then formed an expanded systems map from this network ([Figure S6B](#)) and identified communities with an over-representation of conserved, rat-specific, or human-specific genes ([Figure 6E](#)). In this way, we highlighted functional systems influencing BMI that are affected by common genetic variation in a species-specific manner. For rats, we found 12 communities containing a significant number of rat-specific genes, representing processes such as water homeostasis, cytokine-mediated signaling, and glycerolipid metabolism. Simi-

larly, for humans, we observed five communities enriched for human-specific genes, including processes such as biosynthesis of carbohydrate derivatives and DNA damage response. Within the human- and rat-specific communities, only the human-specific cytoplasmic translation branch contained a significant number of genes associated with body-size phenotypes in mice ([Figure S6C](#)). The fact that this branch had overlapping functions with the conserved network (chromosome organization) indicated that, while cytoplasmic translation may affect body size across species, it may be more strongly associated with human BMI.

DISCUSSION

Understanding the shared etiology of a trait across species is crucial for leveraging the many experimental advantages of animal models. Rodent models have long been used to clarify monogenic and polygenic regulation of BMI and obesity.^{60–65} Likewise, outbred rodent populations are also being used for genome-wide studies of complex traits such as BMI,^{11,16,37,66,67} building on successes in human populations. However, similar to the challenge of translating GWAS results between different human populations,⁶⁸ cross-species translation is confounded by numerous factors, both logistical and genetic. Logistical factors include the confounding impact of the environment on human genetic studies and the opaque alignment between model organism paradigms and human traits. Genetic limitations include the incomplete power of GWAS, the potential lack of common variation in genes important for the trait under study, differences in genomic architecture, difficulty in determining the causal gene from loci implicated by GWASs, and imperfect knowledge of gene orthology between species. All these factors limit the translation of genetic findings across species.

Here we have demonstrated that network biology provides a foundation for addressing these limitations, clarifying the extent to which genetic findings in rodent species recapitulate those of human genetic studies. Using a network co-localization framework (NetColoc), we showed that variants affecting different genes in the human and rat populations converge on a conserved BMI network ([Figures 2B, 3A, and S3](#)). In doing so, we successfully extended this approach beyond the original application (to compare two human traits) to analyze a putatively similar trait in two different species.³⁴ In addition to connecting the disparate gene sets from humans and rats, our network approach expands the set of implicated genes in each species; these network-implicated genes validate well in mouse genetic

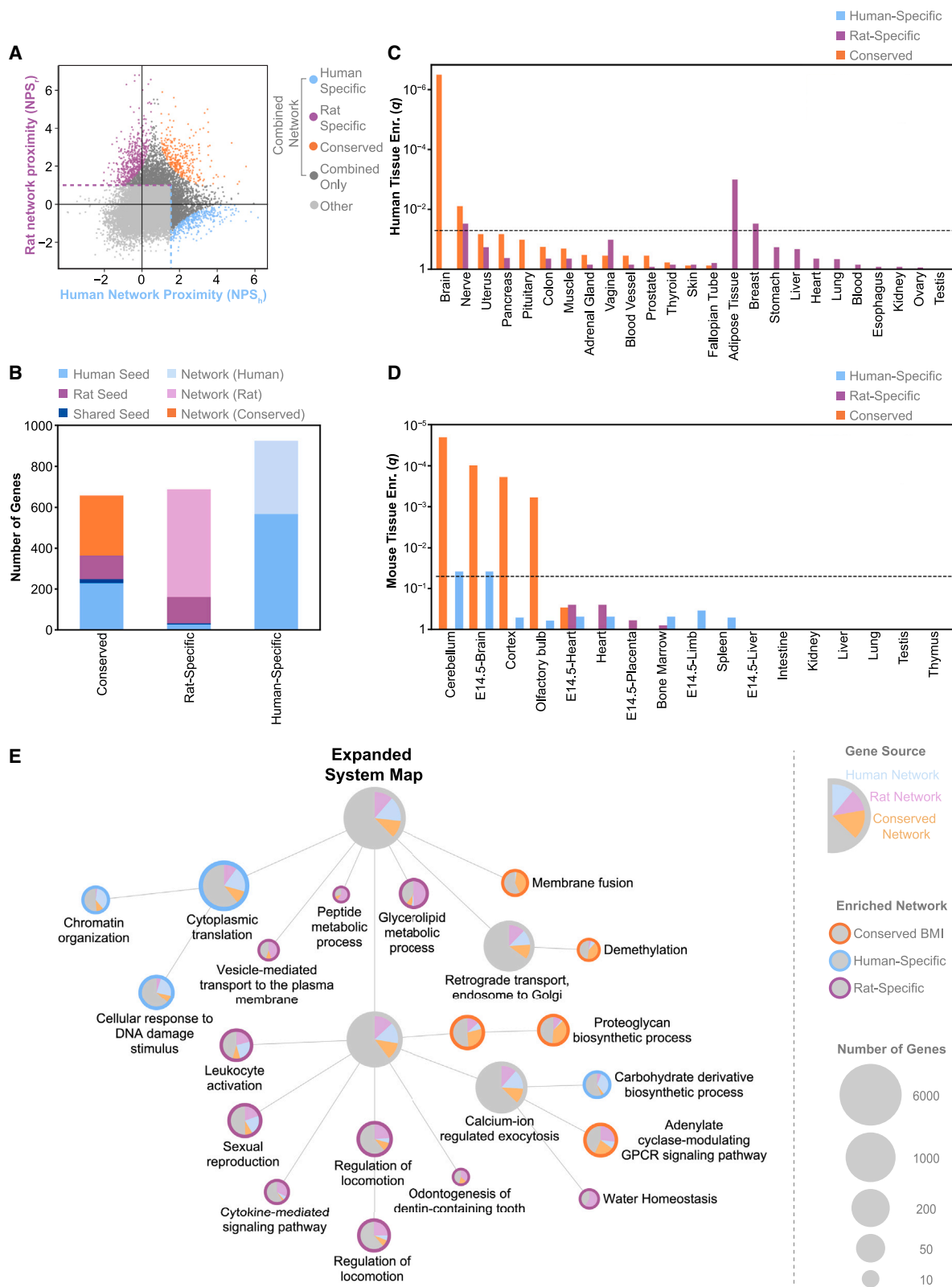
Figure 5. Investigation of gene communities within the conserved BMI systems map

(A) Subnetwork of genes in demethylation community 3 (DM3). Node color indicates the source of the genes (left) and associations with any reported BMI-related abnormal phenotypes (postnatal body size, prenatal body size, embryo size, or body composition) in MGD (right).

(B) Enrichment of systems map communities for genes associated with the four BMI-related traits in mice via MGD. Colored points highlight significant enrichments (Fisher's exact test) for unannotated communities. Light gray indicates unannotated communities that are not significantly enriched, and dark gray indicates annotated communities with significant enrichment, as shown in [Figure 3B](#).

(C) Sources of all MGD-validated genes for communities significantly enriched for body-size-associated genes in (B). MGD-validated genes were defined as genes linked to at least one BMI-related phenotype in MGD.

(D) Unannotated communities within the conserved BMI systems hierarchy (top) with enrichments for selected phenotypes in mice (heatmap, bottom). Odds ratios for all nominally significant ($p < 0.05$, Fisher's exact test) enrichments are shown. See also [Figures S3, S6A](#), and [Table S1](#).



(legend on next page)

studies and are critical for understanding functional roles within the conserved BMI network.

The conserved BMI network shows that BMI shares common etiology between rats and humans and provides an opportunity to explore the similarities and differences in BMI regulation in the two species. The structure of the conserved BMI network reveals distinct communities of genes that correspond to distinct underlying mechanisms of BMI common to humans and rats, with particular emphasis on nervous system processes (Figures 3A, 3B, and 4). Many of the systems are corroborated by known mechanisms of BMI in humans or rodents, but the conservation of such systems between species was not obvious from the independent GWAS results. For example, the regulation of hormones such as leptin and insulin is a well-studied mechanism underlying obesity due to its role in metabolism and hunger signaling.^{69,70} While human and rat BMI GWAS identified genes encoding different hormone receptors and signaling proteins, the proximity of these genes in the molecular interaction network identifies the conserved role of neuronal hormone signaling. The well-studied monogenic obesity genes *BNDF*, *MC4R*, and *POMC* were only identified by human GWAS, but such genes are likely important BMI factors in both species; for example, knockout mutations in *Mc4r* produce obesity phenotypes in rats.⁷¹

The BMI systems map also highlights key hierarchical relationships among functions; for example, gene communities involved in hormone regulation (GSR), regulation of phosphatase activity (RPA), and muscle development and signaling (MSC) all fall within the broader community of synaptic signaling (SS) (Figure 3A). Genes in this community strongly mediate nervous system physiology in mice, particularly via the mediation of protein-protein interactions at the synapse,^{72,73} and synaptic signaling has also been highlighted in rat and human adiposity regulation.⁶⁶ Together with the brain-specific expression of many genes within the conserved BMI network (Figure 6C), these results indicate a central role for nervous system processes in connecting the convergent biology of BMI across species.

It is also crucial to appreciate the differences between species when considering the implications of rodent models in the study of BMI and obesity. While, in some contexts, the conservation of processes such as hormone regulation and muscle development will be sufficient to translate findings, the differences in genes affected by common variants may be consequential for specific applications. It remains to be seen whether species differences in BMI-associated genes are due to differential functional impacts of genetic variants, higher portions of false-positive results, population structure, or a combination thereof.

Limitations of the study

While the identification of a conserved BMI network from non-overlapping GWAS results from two species demonstrates the utility of a network approach for cross-species translation, there are several limitations of note. First, BMI is an imperfect measure of body composition since it does not differentiate between tissue types or assess metabolic factors. Nevertheless, it is widely used as a normalized body mass metric in humans and rats and therefore provides a convenient, high-powered, polygenic, and heritable trait for cross-species study.

Second, the human and rat populations do not perfectly represent BMI in their respective species. The human population only represents individuals of European ancestry, and the HS rats, while genetically diverse for a model species, are laboratory animals that do not capture all genetic diversity in domestic and wild rats. Third, some of the disparity between the human and rat GWAS results may be attributed to the difference in population size between the human and rat GWASs and the process of mapping loci to genes. A relaxed significance threshold for the rat GWAS helps address the sample size and LD disparity but may introduce a greater proportion of false positives. Further, mapping loci to genes based on genomic proximity may introduce false positives and can identify multiple genes from a single significant locus. In humans, the nearest gene is thought to be causal about 70% of the time^{74–76}; however, the more extensive linkage present in the currently available rodent mapping populations may make this nearest-gene approach less accurate. In this respect, we have previously demonstrated that a network approach is robust to up to 80% false positives and boosts signal from potentially noisy inputs.³⁵

Finally, our conversion of rat genes to human orthologs and the use of a human interaction network may introduce bias. Although many genes and cellular processes are similar between humans and other mammals, the mapping of orthologous genes is imperfect, and gene and protein interaction patterns can vary between species due to the rewiring of regulatory and protein-protein interaction networks.^{33,77,78} For this reason, work is ongoing to generate independent molecular interaction networks for mammalian species, including mice,⁷⁹ rats,⁸⁰ and cattle.⁸¹ Network rewiring of interactions also occurs in different tissues and cell types.^{77,82,83} In this study, we used a single global network, meaning we may have missed processes specific to certain contexts.

While future improvements to our methodology and the underlying data will undoubtedly improve our ability to understand cross-species convergence, this work has demonstrated that

Figure 6. Identification of divergent processes underlying BMI in humans and rats

(A) Definition of species-specific subnetworks from NPS_h and NPS_r, showing genes that were not seed genes in either species. Cutoffs for the expanded network are indicated with dotted lines (NPS_h, blue; NPS_r, purple). The expanded network includes all nodes in the human-specific, rat-specific, and conserved network, as well as all dark gray nodes; all other nodes are indicated in light gray. Networks are defined as human specific (NPS_h > 1.5, NPS_r < 0, NPS_h(NPS_r - 1) < -4), rat specific (NPS_h < 1, NPS_r > 1, NPS_h(NPS_r - 1) < -2), or conserved (NPS_h > 1, NPS_r > 1, NPS_h > 3).

(B) The number of genes in the conserved and species-specific subnetworks (x axis) colored by the source of the genes.

(C) Tissue enrichment for species-specific and conserved subnetwork genes in human GTEx data, with p values false discovery rate (FDR) corrected (STAR Methods).

(D) Tissue enrichment for species-specific and conserved subnetwork genes in mouse ENCODE data, with p values FDR corrected (STAR Methods).

(E) Subset of the expanded systems map containing communities enriched for species-specific or conserved genes (q < 0.05, hypergeometric test). Pie charts show the fraction of genes from the conserved and species-specific networks. Systems are annotated based on associated GO Biological Process terms. See also Figures S5, S6, and Table S1.

existing techniques and molecular network knowledge bases already have substantial power to translate GWAS findings from one species to another. We demonstrated that our approach can extend to additional phenotypes beyond BMI and that cross-species network convergence of genetic results is specific to the phenotypes of interest rather than a general property of the method or mammalian genetics as a whole (Figures 2E and S2E). This study can thus act as a road map for future investigations of phenotypes shared across different species. As the quantity, quality, and variety of data generated from model species increase, approaches such as this one have a key role to play in clarifying the translational potential of these various model species across a spectrum of traits related to human health and disease.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead contact
 - Materials availability
 - Data and code availability
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
- **METHOD DETAILS**
 - HS rat phenotyping
 - HS rat DNA sequencing
 - Data acquisition
 - Molecular interaction networks
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - HS rat genotyping, variant calling & imputation
 - HS rat genome-wide association
 - Rat SNP-to-gene mapping
 - Human SNP-to-gene mapping
 - Network propagation and co-localization
 - Definition of the conserved BMI network
 - Definition of the species-specific BMI networks
 - Gene set enrichment analysis
 - Construction of multiscale systems maps
 - Mouse variant validation
 - Tissue enrichment
 - Network visualization

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2023.112873>.

ACKNOWLEDGMENTS

This work was supported by the following grants from the National Institutes of Health: NIMH U01 MH115747 to T.I., NIGMS P41 GM103504 to T.I., and NIDA P50 DA037844 to A.A.P., T.I., T.E.R., H.C., P.M., L.C.S.W., and O.P. This work was partially supported by the Altman Clinical & Translational Research Institute (ACTRI) at the University of California, San Diego. The ACTRI is funded from awards issued by the National Center for Advancing Translational Sci-

ences, NIH UL1TR001442. B.S.L. was supported in part by NIGMS T32 GM008666.

AUTHOR CONTRIBUTIONS

S.N.W., B.S.L., and S.B.R. conceptualized the study, developed methods, performed analysis, and co-wrote the manuscript. S.L. and T.J. developed methods and performed analysis. A.S.C., J.G., R.C., and A.F.G. performed rat genotyping and GWAS analysis. K.H., A.G.M., A.G., W.H., A.H.N., C.P.K., A.L., C.M., C.L.S.P., T.W., and H.B. collected the rat data. O.P., J.R., K.I., H.C., S.B.F., P.M., T.E.R., and L.C.S.W. supervised data acquisition. J.F.K. conceptualized the study. T.I. and A.A.P. conceptualized the study and co-wrote the manuscript.

DECLARATION OF INTERESTS

T.I. is a co-founder, member of the advisory board, and has an equity interest in Data4Cure and Serinus Biosciences. T.I. is a consultant for and has an equity interest in Ideaya BioSciences and Light Horse Therapeutics. A.A.P. is on the Scientific Advisory Board of Vivid Genomics and has stock options. The terms of these arrangements have been reviewed and approved by the University of California San Diego in accordance with its conflict of interest policies.

Received: November 8, 2022

Revised: July 5, 2023

Accepted: July 11, 2023

Published: July 31, 2023

REFERENCES

1. Mestas, J., and Hughes, C.C.W. (2004). Of mice and not men: differences between mouse and human immunology. *J. Immunol.* 172, 2731–2738.
2. Perlman, R.L. (2016). Mouse models of human disease: An evolutionary perspective. *Evol. Med. Public Health* 2016, 170–176.
3. Even, P.C., Virtue, S., Morton, N.M., Fromentin, G., and Semple, R.K. (2017). Editorial: Are Rodent Models Fit for Investigation of Human Obesity and Related Diseases? *Front. Nutr.* 4, 58.
4. Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., and Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* 20, 467–484.
5. French, J.D., and Edwards, S.L. (2020). The Role of Noncoding Variants in Heritable Disease. *Trends Genet.* 36, 880–891.
6. Reynolds, T., Johnson, E.C., Huggett, S.B., Bubier, J.A., Palmer, R.H.C., Agrawal, A., Baker, E.J., and Chesler, E.J. (2021). Interpretation of psychiatric genome-wide association studies with multispecies heterogeneous functional genomic data integration. *Neuropsychopharmacology* 46, 86–97.
7. Mignogna, K.M., Bacanu, S.A., Riley, B.P., Wolen, A.R., and Miles, M.F. (2019). Cross-species alcohol dependence-associated gene networks: Co-analysis of mouse brain gene expression and human genome-wide association data. *PLoS One* 14, e0202063.
8. Li, Z., Votava, J.A., Zajac, G.J.M., Nguyen, J.N., Leyva Jaimes, F.B., Ly, S.M., Brinkman, J.A., De Giorgi, M., Kaul, S., Green, C.L., et al. (2020). Integrating Mouse and Human Genetic Data to Move beyond GWAS and Identify Causal Genes in Cholesterol Metabolism. *Cell Metabol.* 31, 741–754.e5.
9. Gileta, A.F., Fitzpatrick, C.J., Chitre, A.S., St Pierre, C.L., Joyce, E.V., Maguire, R.J., McLeod, A.M., Gonzales, N.M., Williams, A.E., Morrow, J.D., et al. (2022). Genetic characterization of outbred Sprague Dawley rats and utility for genome-wide association studies. *PLoS Genet.* 18, e1010234.
10. Gunturkun, M.H., Wang, T., Chitre, A.S., Garcia Martinez, A., Holl, K., St Pierre, C., Bimschleger, H., Gao, J., Cheng, R., Poleskaya, O., et al. (2022). Genome-Wide Association Study on Three Behaviors Tested in

an Open Field in Heterogeneous Stock Rats Identifies Multiple Loci Implicated in Psychiatric Disorders. *Front. Psychiatr.* **13**, 790566.

11. Keele, G.R., Prokop, J.W., He, H., Holl, K., Littrell, J., Deal, A., Francic, S., Cui, L., Gatti, D.M., Broman, K.W., et al. (2018). Genetic Fine-Mapping and Identification of Candidate Genes and Variants for Adiposity Traits in Outbred Rats. *Obesity* **26**, 213–222.
12. Rat Genome Sequencing and Mapping Consortium; Baud, A., Hermesen, R., Guryev, V., Stridh, P., Graham, D., McBride, M.W., Foroud, T., Calderari, S., Diez, M., et al. (2013). Combined sequence-based and genetic mapping analysis of complex traits in outbred rats. *Nat. Genet.* **45**, 767–775.
13. Yang, C., Wang, Y., Xu, W., Liu, Z., Zhou, S., Zhang, M., and Cui, D. (2019). Genome-wide association study using diversity outcross mice identified candidate genes of pancreatic cancer. *Genomics* **111**, 1882–1888.
14. Long, P.N., Cook, V.J., Majumder, A., Barbour, A.G., and Long, A.D. (2022). The utility of a closed breeding colony of *Peromyscus leucopus* for dissecting complex traits. *Genetics* **221**, iyac026. <https://doi.org/10.1093/genetics/iyac026>.
15. Zou, J., Gopalakrishnan, S., Parker, C.C., Nicod, J., Mott, R., Cai, N., Lionikas, A., Davies, R.W., Palmer, A.A., and Flint, J. (2022). Analysis of independent cohorts of outbred CFW mice reveals novel loci for behavioral and physiological traits and identifies factors determining reproducibility. *G3* **12**, jkab394.
16. Keller, M.P., Rabaglia, M.E., Schueler, K.L., Stapleton, D.S., Gatti, D.M., Vincent, M., Mitok, K.A., Wang, Z., Ishimura, T., Simonett, S.P., et al. (2019). Gene loci associated with insulin secretion in islets from non-diabetic mice. *J. Clin. Invest.* **129**, 4419–4432.
17. Huang, W., Campbell, T., Carbone, M.A., Jones, W.E., Unsel, D., Anholt, R.R.H., and Mackay, T.F.C. (2020). Context-dependent genetic architecture of *Drosophila* life span. *PLoS Biol.* **18**, e3000645.
18. Erickson, P.A., Weller, C.A., Song, D.Y., Bangerter, A.S., Schmidt, P., and Bergland, A.O. (2020). Unique genetic signatures of local adaptation over space and time for diapause, an ecologically relevant complex trait, in *Drosophila melanogaster*. *PLoS Genet.* **16**, e1009110.
19. Wu, K.J., Kumar, S., Serrano Negron, Y.L., and Harbison, S.T. (2018). Genotype Influences Day-to-Day Variability in Sleep in *Drosophila melanogaster*. *Sleep* **41**, zsx205.
20. Kos, M.Z., Carless, M.A., Blondell, L., Leland, M.M., Knape, K.D., Göring, H.H.H., and Szabó, C.Á. (2021). Whole Genome Sequence Data From Captive Baboons Implicate RBFOX1 in Epileptic Seizure Risk. *Front. Genet.* **12**, 714282.
21. Zhang, Q., Cai, Z., Lhomme, M., Sahana, G., Lesnik, P., Guerin, M., Fredholm, M., and Karlsson-Mortensen, P. (2020). Inclusion of endophenotypes in a standard GWAS facilitate a detailed mechanistic understanding of genetic elements that control blood lipid levels. *Sci. Rep.* **10**, 18434.
22. Letko, A., Minor, K.M., Norton, E.M., Marinescu, V.D., Drögemüller, M., Ivansson, E., Megquier, K., Noh, H.J., Starkey, M., Friedenberg, S.G., et al. (2021). Genome-Wide Analyses for Osteosarcoma in Leonberger Dogs Reveal the CDKN2A/B Gene Locus as a Major Risk Locus. *Genes* **12**, 1964.
23. Hédan, B., Cadieu, É., Rimbault, M., Vaysse, A., Dufaure de Citres, C., Devauchelle, P., Botharel, N., Abadie, J., Quignon, P., Derrien, T., and André, C. (2021). Identification of common predisposing loci to hematopoietic cancers in four dog breeds. *PLoS Genet.* **17**, e1009395.
24. Washington, N.L., Haendel, M.A., Mungall, C.J., Ashburner, M., Westfield, M., and Lewis, S.E. (2009). Linking human diseases to animal models using ontology-based phenotype annotation. *PLoS Biol.* **7**, e1000247.
25. Jia, P., and Zhao, Z. (2014). assisted analysis to prioritize GWAS results: principles, methods and perspectives. *Hum. Genet.* **133**, 125–138.
26. Fong, S.H., Carlin, D.E., Ozturk, K., Ideker, T., Arang, N., Bao, B., Bennett, H., Cai, X., Chau, K., Fixsen, B., et al. (2018). UCSD Network Biology Class & Ideker, T. Strategies for Network GWAS Evaluated Using Classroom Crowd Science. *Cell Syst.* **8**, 275–280.
27. Keller, M.P., Gatti, D.M., Schueler, K.L., Rabaglia, M.E., Stapleton, D.S., Simecek, P., Vincent, M., Allen, S., Broman, A.T., Bacher, R., et al. (2018). Genetic Drivers of Pancreatic Islet Function. *Genetics* **209**, 335–356.
28. Schwartzentruber, J., Cooper, S., Liu, J.Z., Barrio-Hernandez, I., Bello, E., Kumasaka, N., Young, A.M.H., Franklin, R.J.M., Johnson, T., Estrada, K., et al. (2021). Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. *Nat. Genet.* **53**, 392–402.
29. Carlin, D.E., Fong, S.H., Qin, Y., Jia, T., Huang, J.K., Bao, B., Zhang, C., and Ideker, T. (2019). A Fast and Flexible Framework for Network-Assisted Genomic Association. *iScience* **16**, 155–161.
30. Wang, Q., Chen, R., Cheng, F., Wei, Q., Ji, Y., Yang, H., Zhong, X., Tao, R., Wen, Z., Sutcliffe, J.S., et al. (2019). A Bayesian framework that integrates multi-omics data and gene networks predicts risk genes from schizophrenia GWAS data. *Nat. Neurosci.* **22**, 691–699.
31. Biran, H., Almozilino, T., Kupiec, M., and Sharan, R. (2018). A Web Server for Network Propagation. *J. Mol. Biol.* **430**, 2231–2236.
32. Bogenpohl, J.W., Mignogna, K.M., Smith, M.L., and Miles, M.F. (2017). Integrative Analysis of Genetic, Genomic, and Phenotypic Data for Ethanol Behaviors: A Network-Based Pipeline for Identifying Mechanisms and Potential Drug Targets. *Methods Mol. Biol.* **1488**, 531–549.
33. Podder, A., Raju, A., and Schork, N.J. (2021). Cross-species and human inter-tissue network analysis of genes implicated in longevity and aging reveal strong support for nutrient sensing. *Front. Genet.* **12**, 719713.
34. Rosenthal, S.B., Willsey, H.R., Xu, Y., Mei, Y., Dea, J., Wang, S., Curtis, C., Sempou, E., Khokha, M.K., Chi, N.C., et al. (2021). A convergent molecular network underlying autism and congenital heart disease. *Cell Syst.* **12**, 1094–1107.e6. <https://doi.org/10.1016/j.cels.2021.07.009>.
35. Rosenthal, S.B., Wright, S.N., Liu, S., Churas, C., Chilin-Fuentes, D., Chen, C.H., Fisch, K.M., Pratt, D., Kreisberg, J.F., and Ideker, T. (2023). Mapping the common gene networks that underlie related diseases. *Nat. Protoc.* **18**, 1745–1759. <https://doi.org/10.1038/s41596-022-00797-1>.
36. Kleinert, M., Clemmensen, C., Hofmann, S.M., Moore, M.C., Renner, S., Woods, S.C., Huypens, P., Beckers, J., de Angelis, M.H., Schürmann, A., et al. (2018). Animal models of obesity and diabetes mellitus. *Nat. Rev. Endocrinol.* **14**, 140–162.
37. Chitre, A.S., Polesskaya, O., Holl, K., Gao, J., Cheng, R., Bimschleger, H., Garcia Martinez, A., George, T., Gileta, A.F., Han, W., et al. (2020). Genome-Wide Association Study in 3,173 Outbred Rats Identifies Multiple Loci for Body Weight, Adiposity, and Fasting Glucose. *Obesity* **28**, 1964–1973.
38. Yengo, L., Sidorenko, J., Kemper, K.E., Zheng, Z., Wood, A.R., Weedon, M.N., Frayling, T.M., Hirschhorn, J., Yang, J., and Visscher, P.M.; GIANT Consortium (2018). Meta-analysis of genome-wide association studies for height and body mass index in ~700,000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649.
39. Munro, D., Wang, T., Chitre, A.S., Polesskaya, O., Ehsan, N., Gao, J., Gu-sev, A., Woods, L.C.S., Saba, L.M., Chen, H., et al. (2022). The regulatory landscape of multiple brain regions in outbred heterogeneous stock rats. *Nucleic Acids Res.* **50**, 10882–10895.
40. Huang, J.K., Carlin, D.E., Yu, M.K., Zhang, W., Kreisberg, J.F., Tamayo, P., and Ideker, T. (2018). Systematic Evaluation of Molecular Networks for Discovery of Disease Genes. *Cell Syst.* **6**, 484–495.e5.
41. Vanunu, O., Magger, O., Rupp, E., Shlomi, T., and Sharan, R. (2010). Associating genes and protein complexes with disease via network propagation. *PLoS Comput. Biol.* **6**, e1000641.
42. Abbott, L., Bloom, J., Bryant, S., Carey, C., Churchhouse, C., Ganna, A., Goldstein, J., Howrigan, D., King, D., Palmer, D., et al. (2022).

Nealelab/UKBB_idsc: v2.0.0 (Round 2 GWAS Update). <https://doi.org/10.5281/zenodo.7186871>.

43. Szklarczyk, D., Gable, A.L., Nastou, K.C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N.T., Legeay, M., Fang, T., Bork, P., et al. (2021). The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 49, D605–D612.
44. Blake, J.A., Baldarelli, R., Kadin, J.A., Richardson, J.E., Smith, C.L., and Bult, C.J.; Mouse Genome Database Group (2021). Mouse Genome Database (MGD): Knowledgebase for mouse-human comparative biology. *Nucleic Acids Res.* 49, D981–D987.
45. Sinnott-Armstrong, N., Tanigawa, Y., Amar, D., Mars, N., Benner, C., Aguirre, M., Venkataraman, G.R., Wainberg, M., Ollila, H.M., Kiiskinen, T., et al. (2021). Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nat. Genet.* 53, 185–194.
46. León, S., García-Galiano, D., Ruiz-Pino, F., Barroso, A., Manfredi-Lozano, M., Romero-Ruiz, A., Roa, J., Vázquez, M.J., Gaytan, F., Blumenrohr, M., et al. (2014). Physiological roles of gonadotropin-inhibitory hormone signaling in the control of mammalian reproductive axis: studies in the NPFF1 receptor null mouse. *Endocrinology* 155, 2953–2965.
47. Weber, D.R., Hadjiyannakis, S., McMillan, H.J., Noritz, G., and Ward, L.M. (2018). Obesity and Endocrine Management of the Patient With Duchenne Muscular Dystrophy. *Pediatrics* 142, S43–S52.
48. Sakaue, S., Kanai, M., Tanigawa, Y., Karjalainen, J., Kurki, M., Koshiba, S., Narita, A., Konuma, T., Yamamoto, K., Akiyama, M., et al. (2021). A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* 53, 1415–1424.
49. Akiyama, M., Okada, Y., Kanai, M., Takahashi, A., Momozawa, Y., Ikeda, M., Iwata, N., Ikegawa, S., Hirata, M., Matsuda, K., et al. (2017). Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nat. Genet.* 49, 1458–1467.
50. Beasom, N., Lu, H., Macke, A., Canan, B.D., Johnson, E.K., Penton, C.M., Kaspar, B.K., Rodino-Klapac, L.R., Zhou, L., Janssen, P.M.L., and Montanaro, F. (2011). mdx^{6cv} mice manifest more severe muscle dysfunction and diaphragm force deficits than do mdx Mice. *Am. J. Pathol.* 179, 2464–2474.
51. Shpargel, K.B., Sengoku, T., Yokoyama, S., and Magnuson, T. (2012). UTX and UTY demonstrate histone demethylase-independent function in mouse embryonic development. *PLoS Genet.* 8, e1002964.
52. Thieme, S., Gyárfás, T., Richter, C., Özhan, G., Fu, J., Alexopoulou, D., Muders, M.H., Michalk, I., Jakob, C., Dahl, A., et al. (2013). The histone demethylase UTX regulates stem cell migration and hematopoiesis. *Blood* 121, 2462–2473.
53. Zhang, F., Xu, L., Xu, L., Xu, Q., Li, D., Yang, Y., Karsenty, G., and Chen, C.D. (2015). JMJD3 promotes chondrocyte proliferation and hypertrophy during endochondral bone formation in mice. *J. Mol. Cell Biol.* 7, 23–34.
54. Oh, S., and Janknecht, R. (2012). Histone demethylase JMJD5 is essential for embryonic development. *Biochem. Biophys. Res. Commun.* 420, 61–65.
55. Ishimura, A., Minehata, K., Terashima, M., Kondoh, G., Hara, T., and Suzuki, T. (2012). Jmjd5, an H3K36me2 histone demethylase, modulates embryonic cell proliferation through the regulation of Cdkn1a expression. *Development* 139, 749–759.
56. Kaelin, W.G. (2008). Generation of a Null Allele of Rbp2, MGI Direct Data Submission to Mouse Genome Database (MGD). J:137333. <https://www.informatics.jax.org/>.
57. Gracz, A.D., and Magness, S.T. (2011). Sry-box (Sox) transcription factors in gastrointestinal physiology and disease. *Am. J. Physiol. Gastrointest. Liver Physiol.* 300, G503–G515.
58. Iguchi, H., Ikeda, Y., Okamura, M., Tanaka, T., Urashima, Y., Ohguchi, H., Takayasu, S., Kojima, N., Iwasaki, S., Ohashi, R., et al. (2005). SOX6 Attenuates Glucose-stimulated Insulin Secretion by Repressing PDX1 Transcriptional Activity and Is Down-regulated in Hyperinsulinemic Obese Mice. *J. Biol. Chem.* 280, 37669–37680.
59. VanHook, A.M. (2016). SOX to be fat. *Sci. Signal.* 9, ec66–ec66.
60. Pellemounter, M.A., Cullen, M.J., Baker, M.B., Hecht, R., Winters, D., Boone, T., and Collins, F. (1995). Effects of the obese gene product on body weight regulation in ob/ob mice. *Science* 269, 540–543.
61. Campfield, L.A., Smith, F.J., Guisez, Y., Devos, R., and Burn, P. (1995). Recombinant mouse OB protein: evidence for a peripheral signal linking adiposity and central neural networks. *Science* 269, 546–549.
62. Attie, A.D., Churchill, G.A., and Nadeau, J.H. (2017). How mice are indispensable for understanding obesity and diabetes genetics. *Curr. Opin. Endocrinol. Diabetes Obes.* 24, 83–91.
63. Doulberis, M., Papaefthymiou, A., Polyzos, S.A., Katsinelos, P., Grigoriadis, N., Srivastava, D.S., and Kountouras, J. (2020). Rodent models of obesity. *Minerva Endocrinol.* 45, 243–263.
64. Tschöp, M., and Heiman, M.L. (2002). Overview of rodent models for obesity research. *Curr. Protoc. Neurosci. Chapter 9*, Unit 9.10.
65. Kanasaki, K., and Koya, D. (2011). Biology of obesity: lessons from animal models of obesity. *J. Biomed. Biotechnol.* 2011, 197636.
66. Crouse, W.L., Das, S.K., Le, T., Keele, G., Holl, K., Seshie, O., Craddock, A.L., Sharma, N.K., Comeau, M.E., Langefeld, C.D., et al. (2022). Transcriptome-wide analyses of adipose tissue in outbred rats reveal genetic regulatory mechanisms relevant for human obesity. *Physiol. Genom.* 54, 206–219.
67. Linke, V., Overmyer, K.A., Miller, I.J., Brademan, D.R., Hutchins, P.D., Trujillo, E.A., Reddy, T.R., Russell, J.D., Cushing, E.M., Schueler, K.L., et al. (2020). A large-scale genome-lipid association map guides lipid identification. *Nat. Metab.* 2, 1149–1162.
68. Martin, A.R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B.M., and Daly, M.J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* 51, 584–591.
69. Miller, G.D. (2019). Appetite Regulation: Hormones, Peptides, and Neurotransmitters and Their Role in Obesity. *Am. J. Lifestyle Med.* 13, 586–601.
70. Qaid, M.M., and Abdelrahman, M.M. (2016). Role of insulin and other related hormones in energy metabolism—A review. *Cogent Food Agric.* 2, 1267691.
71. Mul, J.D., van Bostel, R., Bergen, D.J.M., Brans, M.A.D., Brakkee, J.H., Toonen, P.W., Garner, K.M., Adan, R.A.H., and Cuppen, E. (2012). Melanocortin receptor 4 deficiency affects body weight regulation, grooming behavior, and substrate preference in the rat. *Obesity* 20, 612–621.
72. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. *The Gene Ontology Consortium. Nat. Genet.* 25, 25–29.
73. Gene Ontology Consortium (2021). The Gene Ontology resource: enriching a GOLD mine. *Nucleic Acids Res.* 49, D325–D334.
74. Backman, J.D., Li, A.H., Marcketta, A., Sun, D., Mbatchou, J., Kessler, M.D., Benner, C., Liu, D., Locke, A.E., Balasubramanian, S., et al. (2021). Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* 599, 628–634.
75. Nasser, J., Bergman, D.T., Fulco, C.P., Guckelberger, P., Doughty, B.R., Patwardhan, T.A., Jones, T.R., Nguyen, T.H., Ulirsch, J.C., Lekschas, F., et al. (2021). Genome-wide enhancer maps link risk variants to disease genes. *Nature* 593, 238–243.
76. Pietzner, M., Wheeler, E., Carrasco-Zanini, J., Cortes, A., Koprulu, M., Wörheide, M.A., Oerton, E., Cook, J., Stewart, I.D., Kerrison, N.D., et al. (2021). Mapping the proteo-genomic convergence of human diseases. *Science* 374, eabj1541.
77. Skinner, M.A., Scott, N.E., Prudova, A., Kerr, C.H., Stoykov, N., Stacey, R.G., Chan, Q.W.T., Rattray, D., Gsponer, J., and Foster, L.J. (2021). An atlas of protein-protein interactions across mouse tissues. *Cell* 184, 4073–4089.e17.

78. Ha, D., Kim, D., Kim, I., Oh, Y., Kong, J., Han, S.K., and Kim, S. (2022). Evolutionary rewiring of regulatory networks contributes to phenotypic differences between human and mouse orthologous genes. *Nucleic Acids Res.* 50, 1849–1863.
79. Alanis-Lobato, G., Möllmann, J.S., Schaefer, M.H., and Andrade-Navarro, M.A. (2020). the mouse integrated protein-protein interaction reference. *Database* 2020, baaa035.
80. Tao, Y.-T., Ding, X.-B., Jin, J., Zhang, H.-B., Guo, W.-P., Ruan, L., Yang, Q.-L., Chen, P.-C., Yao, H., and Chen, X. (2020). Predicted rat interactome database and gene set linkage analysis. *Database* 2020. <https://doi.org/10.1093/database/baaa086>.
81. Chen, Y., Liu, Y., Du, M., Zhang, W., Xu, L., Gao, X., Zhang, L., Gao, H., Xu, L., Li, J., and Zhao, M. (2017). Constructing a comprehensive gene co-expression based interactome in *Bos taurus*. *PeerJ* 5, e4107.
82. Bossi, A., and Lehner, B. (2009). Tissue specificity and the human protein interaction network. *Mol. Syst. Biol.* 5, 260.
83. Huttlin, E.L., Bruckner, R.J., Navarrete-Perea, J., Cannon, J.R., Baltier, K., Gebreab, F., Gygi, M.P., Thornock, A., Zarraga, G., Tam, S., et al. (2021). Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. *Cell* 184, 3022–3040.e28.
84. Smith, C.L., and Eppig, J.T. (2009). The mammalian phenotype ontology: enabling robust annotation and comparative analysis. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 1, 390–399.
85. Babraham Bioinformatics. FastQC (2017). A Quality Control Tool for High Throughput Sequence Data (Cambridge, UK: Babraham Institute).
86. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.
87. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303.
88. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). Analysis of Next Generation Sequencing Data. *BMC Bioinf.* 15, 356.
89. Durvasula, A., Hoffman, P.J., Kent, T.V., Liu, C., Kono, T.J.Y., Morrell, P.L., and Ross-Ibarra, J. (2016). angsd-wrapper: utilities for analysing next-generation sequencing data. *Mol. Ecol. Resour.* 16, 1449–1454.
90. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575.
91. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4, 7.
92. Browning, B.L., and Browning, S.R. (2016). Genotype Imputation with Millions of Reference Samples. *Am. J. Hum. Genet.* 98, 116–126.
93. Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 5, e1000529.
94. Yang, J., Hong Lee, S., Goddard, M.E., and Visscher, P.M. (2011). GCTA: A Tool for Genome-wide Complex Trait Analysis. *Am. J. Hum. Genet.* 88, 76–82. <https://doi.org/10.1016/j.ajhg.2010.11.011>.
95. Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z., and Bergmann, S. (2016). Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. *PLoS Comput. Biol.* 12, e1004714.
96. Zheng, F., Zhang, S., Churas, C., Pratt, D., Bahar, I., and Ideker, T. (2021). HiDeF: identifying persistent structures in multiscale 'omics data. *Genome Biol.* 22, 21. Preprint at. <https://doi.org/10.1186/s13059-020-02228-4>.
97. Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., and Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 47, W191–W198.
98. Dougherty, J.D., Schmidt, E.F., Nakajima, M., and Heintz, N. (2010). Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. *Nucleic Acids Res.* 38, 4218–4230.
99. Jain, A., and Tuteja, G. (2019). TissueEnrich: Tissue-specific gene enrichment analysis. *Bioinformatics* 35, 1966–1967.
100. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504.
101. Loos, R.J. (2018). The genetics of adiposity. *Curr. Opin. Genet. Dev.* 50, 86–95.
102. Eyre, T.A., Wright, M.W., Lush, M.J., and Bruford, E.A. (2007). HCOP: a searchable database of human orthology predictions. *Briefings Bioinf.* 8, 2–5.
103. Yates, B., Gray, K.A., Jones, T.E.M., and Bruford, E.A. (2021). Updates to HCOP: the HGNC comparison of orthology predictions tool. *Briefings Bioinf.* 22, bbab155.
104. Hermesen, R., de Ligt, J., Spee, W., Blokzijl, F., Schäfer, S., Adami, E., Boymans, S., Flink, S., van Bostel, R., van der Weide, R.H., et al. (2015). Genomic landscape of rat strain and substrain variation. *BMC Genom.* 16, 357.
105. Ramdas, S., Ozel, A.B., Treutelaar, M.K., Holl, K., Mandel, M., Woods, L.C.S., and Li, J.Z. (2019). Extended regions of suspected mis-assembly in the rat reference genome. *Sci. Data* 6, 39.
106. Gileta, A.F., Gao, J., Chitre, A.S., Bimschleger, H.V., St Pierre, C.L., Gopalakrishnan, S., and Palmer, A.A. (2020). Adapting Genotyping-by-Sequencing and Variant Calling for Heterogeneous Stock Rats. *G3* 10, 2195–2205.
107. Cheng, R., Parker, C.C., Abney, M., and Palmer, A.A. (2013). Practical considerations regarding the use of genotype and pedigree data to model relatedness in the context of genome-wide association studies. *G3* 3, 1861–1867.
108. Gonzales, N.M., Seo, J., Hernandez Cordero, A.I., St Pierre, C.L., Gregory, J.S., Distler, M.G., Abney, M., Canzar, S., Lionikas, A., and Palmer, A.A. (2018). Genome wide association analysis in a mouse advanced intercross line. *Nat. Commun.* 9, 5162.
109. 1000 Genomes Project Consortium; Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. *Nature* 526, 68–74.
110. Davies, R.B. (1980). Algorithm AS 155: The Distribution of a Linear Combination of χ^2 Random Variables. *Appl. Stat.* 29, 323–333.
111. Farebrother, R.W., and Algorithm, A.S. (1984). The Distribution of a Positive Linear Combination of χ^2 Random Variables. *Appl. Stat.* 33, 332–339. 204.
112. Guney, E., Menche, J., Vidal, M., and Barabási, A.L. (2016). Network-based in silico drug efficacy screening. *Nat. Commun.* 7, 10331.
113. Singhal, A., Cao, S., Churas, C., Pratt, D., Fortunato, S., Zheng, F., and Ideker, T. (2020). Multiscale community detection in Cytoscape. *PLoS Comput. Biol.* 16, e1008239.
114. Bult, C.J., Blake, J.A., Smith, C.L., Kadin, J.A., and Richardson, J.E.; Mouse Genome Database Group (2019). Mouse Genome Database (MGD) 2019. *Nucleic Acids Res.* 47, D801–D806.
115. GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* 45, 580–585.
116. Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., and Ren, B. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature* 488, 116–120.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
GWAS summary statistics (Human BMI, height)	Yengo et al. ³⁸	https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files
Rat genotypes and phenotypes (Rat BMI, body length)	This paper	EVA: PRJEB63638 (Project); ENA: ERZ19474633 (Analysis)
Rat gene-level summary statistics (BMI, body length)	This paper	Table S4
Human gene-level summary statistics (BMI, Height)	This paper	Table S5
Control GWAS summary statistics	Neale Lab ⁴²	http://www.nealelab.is/uk-biobank ; Zenodo: https://doi.org/10.5281/zenodo.7186871
PCNet Interactome (v1.3)	Huang et al. ⁴⁰	Ndexbio.org, NDEx: 4de852d9-9908-11e9-bcaf-0ac135e8bacf
STRING High Confidence Rat Interactome	Szklarczyk et al., ⁴³ this paper	Ndexbio.org, NDEx: 880c7d8c-f5ad-11ec-ac45-0ac135e8bac
Conserved Human-Rat BMI Network	This paper	Ndexbio.org, NDEx: https://doi.org/10.18119/N9102V
Conserved Human-Rat BMI Systems Map	This paper	Ndexbio.org, NDEx: https://doi.org/10.18119/N9RG73
Human-Specific BMI Network	This paper	Ndexbio.org, NDEx: https://doi.org/10.18119/N9MS56
Rat-Specific BMI Network	This paper	Ndexbio.org, NDEx: https://doi.org/10.18119/N9H03K
Expanded BMI Systems Map	This paper	Ndexbio.org, NDEx: https://doi.org/10.18119/N9C89C
Expanded BMI Network	This paper	Ndexbio.org, NDEx: https://doi.org/10.18119/N9W90Z
Mouse genotype-phenotype associations	Blake et al. ⁴⁴	https://www.informatics.jax.org/downloads/
Mammalian Phenotype Ontology	Smith and Eppig ⁴⁴	https://www.informatics.jax.org/downloads/
Experimental models: Organisms/strains		
Heterogeneous Stock (HS) Rats	Medical College of Wisconsin, WI	MCW: NMcw:HS #2314009; RRID
Software and algorithms		
CrossSpeciesBMI v.1.0.0	This paper	Zenodo: https://doi.org/10.5281/zenodo.7868889
FASTX-Toolkit v0.0.13	Hannon Lab, 2010	RRID: SCR_005534; http://hannonlab.cshl.edu/fastx_toolkit/
FastQC v0.11.6	Babraham Bioinformatics ⁸⁵	RRID: SCR_014583; http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Burrows-Wheeler Aligner (BWA) v0.7.5a	Li and Durbin ⁸⁶	RRID: SCR_010910; http://bio-bwa.sourceforge.net/
GATK v3.5	McKenna et al. ⁸⁷	RRID: SCR001876; https://software.broadinstitute.org/gatk/

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ANGSD-SAMtools	Korneliussen et al. ⁸⁸ & Durvasula et al. ⁸⁹	RRID: SCR_021865; https://github.com/ANGSD/angsd ;
PLINK v1.9	Purcell et al. ⁹⁰ & Chang et al. ⁹¹	RRID: SCR_001757; https://www.cog-genomics.org/plink/
BEAGLE v4.1	Browning and Browning ⁹²	RRID: SCR_001789; http://faculty.washington.edu/browning/beagle/beagle.html
IMPUTE2	Howie et al. ⁹³	RRID: SCR_013055; https://mathgen.stats.ox.ac.uk/impute/impute_v2.html
GCTA	Yang et al. ⁹⁴	https://yanglab.westlake.edu.cn/software/gcta/#Overview
PASCAL	Lamparter et al. ⁹⁵	https://www2.unil.ch/cbg/index.php?title=Pascal
Python 3.9.13	N/A	https://www.python.org/downloads/release/python-3913/
NetworkX 2.8.4	N/A	RRID: SCR_016864; https://networkx.org/
NetColoc v0.1.6.post1	Rosenthal et al. ³⁴	https://pypi.org/project/netcoloc/0.1.6.post1/
HiDeF v1.0.0	Zheng et al. ⁹⁶	https://apps.cytoscape.org/apps/cycommunitydetection
Gprofiler	Raudvere et al. ⁹⁷	RRID: SCR_006809; https://biit.cs.ut.ee/gprofiler/
TSEA	Dougherty et al. ⁹⁸	http://genetics.wustl.edu/jdlab/tsea/
TissueEnrich	Jain and Tuteja ⁹⁹	http://tissueenrich.gdcb.iastate.edu/
Cytoscape 3.9.1	Shannon et al. ¹⁰⁰	RRID: SCR_003032; https://cytoscape.org

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to the lead contact, Trey Ideker tideker@ucsd.edu.

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All data analyzed in this paper are publicly available. Accession numbers are listed in the [key resources table](#). Rat genotype and phenotype data have been deposited at the European Variation Archive (Project: EVA: [PRJEB63638](https://www.eva-epi.org/PRJEB63638), Analyses: ENA: [ERZ19474633](https://www.eva-epi.org/ERZ19474633)). All networks and systems maps generated in this study are available as a network set on the Network Data Exchange (NDEx: <https://doi.org/10.18119/N97G8T>). Gene-level GWAS summary statistics are included in [Tables S4](#) and [S5](#).
- All code used for analysis and data visualization is freely available in public repositories. All original code is publicly available at GitHub: <https://github.com/sarah-n-wright/CrossSpeciesBMI> as of the date of publication and has been deposited at Zenodo (Zenodo: <https://doi.org/10.5281/zenodo.7868889>).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The N/NIH heterogeneous stock (HS) rats colony was initiated by the NIH in 1984 using the following eight inbred founder strains: ACI/N, BN/SsN, BUF/N, F344/N, M520/N, MR/N, WKY/N, and WN/N,¹⁰¹ which have previously been shown to differ significantly for adiposity traits.¹¹ Rats were housed in micro-isolation cages in a conventional facility using autoclaved bedding (sani-chips from PJ Murphy). They were provided reverse osmosis water chlorinated to 2–3 ppm, and had *ad libitum* access to diets as listed below.

The rats used for this study are part of a large multi-site project focused on genetic analysis of behavioral phenotypes related to drug abuse (www.ratgenes.org), and are a larger updated cohort from previously published.³⁷ HS rats bred at NMcw:HS colony (RRID:RGD_2314009) (Medical College of Wisconsin (MCW), WI) were sent to three institutions throughout the United States: University of Tennessee Health Science Center (TN), University at Buffalo (NY), and University of Michigan (MI). The NMcw:HS breeding colony consisted of 64–96 breeding pairs, bred from 8 founder inbred strains (the number of breeding pairs varied over the course of the experiment). No more than 2 siblings were sent to the same Project, and these siblings had different sex. Rats were shipped at 3–6 weeks of age, and each site received multiple shipments over five years (10/27/2014–03/07/2019). In these datasets, the information was aggregated from multiple projects across different centers, leading to four cohorts, with siblings present in different phenotyping centers. Rats in TN were fed Teklad Irradiated LM-485 Mouse/Rat Diet; rats in NY were fed Envigo Teklad 18% Protein Rodent Diet, and rats in MI were fed Labdiet Picolab Laboratory Rodent Diet Irradiated.

Rats were exposed to a different battery of behavioral testing at each site (Table S3), followed by euthanasia, which occurred at different ages at each site. All phenotypes presented in this paper were collected at the time of euthanasia, in roughly equal numbers of male and female rats. Briefly, in MI, rats were housed in trios, exposed to a single modest dose of cocaine (15 mg/kg) each day for five days, and then euthanized, weighed, and body length was measured 4–7 days after the final cocaine exposure (89 s.d. 6 days of age). In NY, rats were housed in pairs, tested for multiple behaviors over 16 weeks, exposed to a modest dose of cocaine (10 mg/kg) once daily for three days, and then euthanized, weighed, and body length was measured 7–10 days after the last dose of cocaine (198 s.d. 13 days of age). In TN, there were two separate cohorts: breeders (sent from MCW) and experimental rats (bred in TN). Female breeders had mostly one, sometimes two litters and underwent no behavioral testing. After breeding was completed, males and females were euthanized, weighed, and body length was measured (169 s.d. 34 days of age). The experimental rats were tested for multiple behaviors, exposed to nicotine (self-administration, resulting in a range of doses) for 12 days, and euthanized, weighed, and body length was measured 10 days after the final dose of nicotine (73 s.d. 12 days of age). All protocols were approved by the Institutional Animal Care and Use Committees (IACUC) for each of the relevant institutions.

METHOD DETAILS

HS rat phenotyping

Several days after the completion of behavioral experiments (males and females), rats were fasted overnight (17 ± 2 h), and body weight was measured. Under anesthesia (pentobarbital at MI and NY, isoflurane at TN), the rats were measured for body length (from nose to base of the tail (tail length not included in the measurement of the body length)), which we then used to calculate body mass index. BMI was calculated as: $(\text{body weight/body length}^2) \times 10$. Several tissues were dissected and weighed, including spleens, which were used for DNA extraction. All protocols were approved by the Institutional Animal Care and Use Committees (IACUC) for each of the relevant institutions.

HS rat DNA sequencing

HS rat DNA was extracted from spleen tissue using the Agencourt DNAdvance Kit (Beckman Coulter Life Sciences, Indianapolis, IN). All genomic DNA quality and purity was assessed by NanoDrop 8000 (Thermo Fisher Scientific, Waltham, MA). Sample DNA, PstI barcoded adapters, and NlaIII Y-adapters were combined in a 96-well plate and allowed to evaporate at 37°C overnight. The PstI adapter barcode is “in-line,” such that each sequencing read from a given sample contains both the PstI overhang sequence (4bps) and a unique adapter sequence (4–8bps) prior to the beginning of the insert sequence. Sample DNA and adapters were re-eluted on day two with a PstI/NlaIII digestion mix and incubated at 37°C for 2 h to allow for complete digestion. Ligation reagents were then added and incubated at 16°C for 1 h to anneal the adapters to the DNA fragments, followed by a 30-min incubation at 80°C to inactivate the enzymes. Sample libraries were purified using a plate from a MinElute 96 UF PCR Purification Kit (QIAGEN Inc., Hilden, Germany), vacuum manifold, and ddH₂O. Once re-eluted, libraries were quantified in duplicate with Quant-IT PicoGreen (Thermo Fisher Scientific, Waltham, MA) and pooled to the desired level of multiplexing (i.e., 12, 24, or 48 samples per library). Each pooled library was then concentrated by splitting the pooled volume across 2–3 wells of the MinElute vacuum plate and resuspending the library at desired volume for use in the Pippin Prep. The concentrated pool was quantified to ensure the gel cassette was not overloaded with DNA (0.5mg). The pool was then loaded into the Pippin Prep for size selection (300–450 bps) using a 2% agarose gel cassette on a Pippin Prep (Sage Science, Beverly, MA). Size-selected libraries were then PCR amplified for 12 cycles to add the Illumina sequencing primers and increase the quantity of DNA, concentrated again, and checked for quality on an Agilent 2100 Bioanalyzer with a DNA 1000 Series II chip (Agilent Technologies, Santa Clara, CA). Bioanalyzer results were used to ensure sufficient library concentration and to identify excessive primer dimer peaks. Each library was run on a single flow cell lane on an Illumina HiSeq 4000 with 100bp single-end reads at the Institute for Genomic Medicine (IGM) Genomics Center (University of California San Diego, La Jolla, CA).

Data acquisition

The summary statistics for human body mass index and height were obtained from the GIANT Consortium ([GIANT consortium data files](#)), computed via a meta-analysis of GWAS results representing an average of 681,275 (BMI) and 693,529 (height) individuals of European ancestry.³⁸ Summary statistics for all human negative control traits were obtained from the Neale Lab Round 2 UK Biobank

(UKB) GWAS⁴² (<http://www.nealelab.is/uk-biobank>). Phenotype codes are as follows: FEV1: forced expiratory volume per second (20153_irmt), AR: allergic rhinitis symptoms (6152_9), SMOK: smoking status - never smoked cigarettes (20116_0), Neo: non-cancer neoplasms (C3_OTHER_SKIN). Negative control traits were selected to have a substantial number of implicated human genes ($n > 150$), a similar SNP heritability to BMI ($0.15 < h^2 < 0.35$), and minimal genetic correlation with BMI ($|r_g| < 0.1$). These estimates were obtained from the UKB Heritability browser (https://nealelab.github.io/UKBB_Idsc/h2_browser.html) and UKB Genetic Correlation browser (<https://ukbb-rs.hail.is/>), both generated by the Neale Lab.⁴² Bulk human-rat ortholog mapping from HGNC Comparison of Orthology Predictions (HCOP)^{102,103} was downloaded from the European Bioinformatics Institute (<http://ftp.ebi.ac.uk/pub/databases/genenames/hcop/>) on June 26, 2022. Mouse genotype-phenotype associations and the Mammalian Phenotype Ontology were downloaded from the Mouse Genome Database⁴⁴ (<http://www.informatics.jax.org/>) on April 24, 2022. Datasets acquired: Phenotype-Genotype Associations (MGI_PhenoGenoMP.rpt); Marker Information (MRK_List2.rpt); Human Ortholog Mapping (MRK_List2.rpt); Mammalian Phenotype Ontology⁸⁴ (MPheno_OBO.ontology). Specific mappings between conserved BMI network genes and BMI-relevant phenotypes in MGD can be found in Table S1. Stable versions of Mouse Genome Database reference data used in this study are contained in the code repository (Zenodo: <https://doi.org/10.5281/zenodo.7868889>).

Molecular interaction networks

The Parsimonious Composite Network⁴⁰ (PCNet v1.3) was obtained from the network data exchange (ndexbio.org), NDEx: 4de852d9-9908-11e9-bcaf-0ac135e8bacf. PCNet is a molecular interaction resource formed from integrating 21 interaction databases that contain various evidence types, including physical protein-protein, genetic, co-expression, and co-citation evidence. Each interaction in PCNet is supported by at least two of the component databases, a threshold chosen to maximize the ability of PCNet to perform gene set recovery tasks via network propagation. All human genes and human orthologs of rat genes were mapped to the nodes of PCNet via gene symbols. Rat interactome data was downloaded from STRING⁴³ (v11.5, <https://string-db.org/cgi/download.pl>) and filtered to only interactions in *R. norvegicus*. We defined high-confidence interactions as those with a combined score > 700 and mapped all network node identifiers to rat gene symbols using STRING's protein information (10116.protein.info.v11.5.txt.gz). All rat genes and rat orthologs of human genes were mapped to the nodes of the high-confidence rat interactome via rat symbols.

QUANTIFICATION AND STATISTICAL ANALYSIS

HS rat genotyping, variant calling & imputation

The PstI adapter barcodes were used to demultiplex FASTQ files into individual sample files. Demultiplexing was completed using FASTX Barcode Splitter v0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/, RRID: SCR_005534). Reads were discarded if they could not be matched with any barcode (maximum of 1 mismatch allowed) or they lacked the appropriate enzyme cut site. Samples with less than two million reads after demultiplexing were discarded. Read quality was assessed using FastQC v0.11.6.⁸⁵ Reads were trimmed of the barcode, adapter sequences, and low-quality base pairs at the ends of reads using FASTX Clipper/Trimmer/Quality Trimmer tools v0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/, RRID: SCR_005534). A base quality threshold of 20 was used and reads shorter than 25bp were discarded.

The *R. norvegicus* genome assembly rn6 was used as the reference genome for read alignment with the Burrows-Wheeler Aligner v0.7.5a (BWA)⁸⁶ using the mem algorithm. We then used GATK IndelRealigner v3.5⁸⁷ to improve alignment quality by locally realigning reads around a reference set of known indels in 42 whole-genome sequenced inbred rat strains, including the eight HS progenitor strains.¹⁰⁴

Variants were called, and genotype likelihoods were computed at variant sites using ANGSD v0.911, under the SAMtools model for genotype likelihoods (ANGSD-SAMtools).^{88,89} Further, using ANGSD-SAMtools, we inferred the major and minor alleles (-domajor-minor 1) from the genotype likelihoods, retaining only high confidence polymorphic sites (-snp_pval 1e-6), and estimated the allele frequencies based on the inferred alleles (-domaf 1). We discarded sites missing read data in more than 4% of samples (-minInd). Additionally, we tested multiple thresholds for minimum base (-minQ) and mapping (-minMapQ) qualities. Variants for X- and Y chromosomes were not called. Prior to GWAS, SNPs in high linkage disequilibrium were removed using PLINK v1.9^{90,91} with an r^2 cutoff of 0.95.

Beagle v4.1⁹² was used to improve the genotyping within the samples without the use of an external reference panel. Missing and low-quality genotypes were imputed by borrowing information from other individuals in the dataset with high-quality information at these same variant sites. To verify the quality of the "internally" imputed genotypes prior to imputing SNPs from the 42 inbred strain reference panel¹⁰⁴ we checked concordance rates for the 96 HS animals with array genotypes, examined the TSTV ratio, and assessed whether the sex as recorded in the pedigree records agreed with the sex empirically determined by the proportion of reads on the X chromosome out of the total number of reads. We also identified Mendelian errors using the -mendel option in PLINK and known pedigree information for 1,136 trios from 214 families within the HS sample. Using the fraction of the trios that were informative for a given SNP and the equation $1 - (1 - 2p(1 - p))^3$, where p represents the minor allele frequency of the allele, we formed curves for the distributions of the expected number of Mendelian errors for both SNPs and samples and chose the inflection points as thresholds for the number of Mendelian errors allowed.

Only variants previously identified in the eight HS founder strains were retained as we expected the polymorphisms in our samples to be limited to the variation present in the founders.^{104,105} Estimated haplotypes from our dataset for imputation with multiple different reference panels were calculated previously,¹⁰⁶ and were applied to this dataset.

For imputation, we used a reference panel calculated previously¹⁰⁶ from a combination of existing sequencing and array genotyping data from the HS rat founder and other inbred laboratory rat strains.¹⁰⁴ Genotype data underwent QC and were phased by Beagle into single chromosome haplotype files. Imputation by IMPUTE2 was performed in 5Mb windows using the aforementioned reference panels and genetic maps. Final genotypes for all samples are reported in [EVA: PRJEB63638](#).

HS rat genome-wide association

Each trait within each research site was quantile-normalized separately for males and females; an approach similar to using sex as a covariate. Other relevant covariates (including age, batch number, and dissector) were identified for each trait, and covariate effects were regressed out if they were significant and if they explained more than 2% of the variance (BMI covariates: age and test center for all rats; additionally for TN, experimental or breeders, and TN experimental: behavioral experiment, Technician JS, and for TN breeder: Technician XH. Body length covariates: age, test center for all rats; additionally for TN experimental: Technician JS, TN experimental, and for TN breeders: age, Technician XH). Residuals were then quantile-normalized again, after which the data for each sex and site were pooled prior to further analysis. This approach removed mean differences due to sex; however, it did not attempt to model gene-by-sex interactions. By quantile-normalizing the three centers separately, we also addressed the numerous and confounded differences among the three cohorts, such that quantitative trait loci (QTLs) identified in the current study are resistant to environmental influences that differed among the sites. Raw phenotypes and post-regression quantile-normalized values for body length without tail and body mass index without tail are reported in [EVA: PRJEB63638](#).

The software GCTA⁹⁴ was used to perform the GWAS analysis. First, we pre-calculated a genetic relatedness matrix (GRM) from autosomal chromosomes (–autosome –make-grm-bin). We then performed the GWAS with a linear mixed model (–mlma), using the GRM to account for complex family relationships within the HS rat population (–mlma-subtract-grm). This method employs a Leave One Chromosome Out (LOCO) method to avoid proximal contamination.^{107,108}

Rat SNP-to-gene mapping

We assigned gene-level significance values by taking the lowest p value of any GWAS variant within a ± 10 kb region encompassing each gene region. For each gene's set of assigned SNPs, the SNP with the lowest p value was assigned as the lead SNP for that gene and used to assess the significance of the gene ([Table S4](#)). Using a threshold of $p < 1 \times 10^{-4}$, we defined sets of rat seed genes significantly associated with rat body mass index (BMI) and rat body length (without tail; BL). For BMI, 20,079 top SNPs were assigned to 23,884 rat genes, with 3,108 being assigned to more than one gene. After significance filtering, 413 lead SNPs were assigned to 476 genes, with 54 SNPs being assigned to more than one of the 476 significant genes identified. Rat genes were then mapped to human orthologs using the HCOPI^{102,103} search tool (<https://www.genenames.org/tools/hcop/>), which incorporates orthology information from 16 databases. We required that the ortholog be present in at least three databases to map a gene. To plot the position of rat seed genes in the rat genome ([Figure S1](#)), each gene's genomic range ± 10 kb was plotted in a single column per chromosome, and manually aligned to each chromosome. Due to variable genomic ranges listed for some genes, the smallest start site and greatest end site from those listed in [Table S4](#) were used to define the genomic range.

Human SNP-to-gene mapping

We generated gene-level significance values ([Table S5](#)) from the SNP-level summary statistics for each human trait using the PASCAL algorithm⁹⁵ using default parameters. This approach assigns SNPs to a gene if they fall within a window of ± 50 kb of the gene region and then calculates an aggregate p value after correcting for linkage disequilibrium. To model the expected distribution of SNP p values, an SNP-by-SNP correlation matrix is formed from the 1,000 Genomes Project¹⁰⁹ reference data for European individuals (1KG-EUR) for all n SNPs in the gene region and used as the basis of a multivariate normal null distribution, allowing calculation of z-scores for each SNP. A weighted sum of χ^2_1 -distributed random variables is used to model the sum of all z-scores within a gene region, with the resulting gene p value of the sum of chi-squared statistics calculated using the Davies algorithm (or Farebrother algorithm) for weighted sums of independent χ^2_1 -distributed random variables.^{110,111} Following a Bonferroni correction of the PASCAL p values, we defined sets of trait-associated genes using a threshold of $q < 0.05$.

Network propagation and co-localization

We used the Python package NetColoc^{34,35} (<https://pypi.org/project/netcoloc/>) for network propagation and co-localization. The sets of significant trait-associated genes from GWAS were used as “seed” genes for network propagation using a Random Walk with Restart⁴¹ algorithm. This algorithm models the diffusion of heat along the edges of a network from an initial set of “hot” seed genes. In each iteration, a unit of heat is added to each seed gene, and the heat diffuses from each gene to its neighbors. To conserve the total heat within the system, a constant fraction of heat dissipates from each gene with each iteration. The heat diffusion process converges to a closed-form solution³⁵ described by [Equation 1](#):

$$\mathbf{F} = (\mathbf{I} - \alpha \mathbf{W})^{-1} (1 - \alpha) \mathbf{Y}_0 \quad (\text{Equation 1})$$

Where F is the stable heat vector for all nodes, Y_0 is the vector of seed genes, W is the column-normalized adjacency matrix of the network (PCNet), and $\alpha \in (0, 1)$ is the dissipation constant. Following network propagation with $\alpha = 0.5$, we calculated a network proximity score (NPS) for each gene in the network by comparing the observed results to a null distribution. The null distribution was formed by propagating 1,000 randomly selected seed gene sets. Each set was sampled to preserve the size and degree distribution of the original input set. As previously implemented,¹¹² we binned all genes in the network by degree with a minimum of 10 nodes per bin. For each gene g , the NPS was calculated as a Z score comparing the observed heat at that gene $F_{g,S}$ after network propagation of gene set S , to the mean of the null distribution heats at that gene $\bar{F}_{g,rand}$ as given by Equation 2:

$$NPS_{g,S} = \frac{\log(F_{g,S}) - \langle \log(\bar{F}_{g,rand}) \rangle}{\sigma(\log(\bar{F}_{g,rand}))} \quad (\text{Equation 2})$$

Where $\langle \rangle$ denotes the mean of a vector, and σ denotes the standard deviation of a vector. All heat values are log-transformed to ensure the distributions are approximately normal.

NetColoc recommends fewer than 500 input seed genes given the sample space of PCNet (~18,000 genes). Therefore, we employed a weighted sampling procedure for any trait having more than 500 significantly associated genes. We sampled 500 genes from the set of all significant genes (weighted by $-\log_{10}p$ from GWAS) and ran the propagation analysis from this subset. After 100 repetitions, the 75% percentile NPS score was selected to approximate a consensus score for each gene.

From input seed genes for humans and rats, we independently calculated NPS_h (human) and NPS_r (rat) for each trait. We then defined a gene as colocalized between two species if it had high proximity to inputs in both species. Therefore, we defined the combined network proximity NPS_{hr} as the product of the independent species scores in Equation 3:

$$NPS_{hr} = NPS_h \circ NPS_r \quad (\text{Equation 3})$$

We repeated this approach using our high-confidence rat interaction network from STRING⁴³ (v11.5). All human input seed genes were mapped to rat gene symbols using the bulk download human-rat ortholog mapping from HCOP.^{102,103} Where multiple complete mappings existed for a gene, the one supported by the most databases was selected.

Definition of the conserved BMI network

From NPS_{hr} , we selected genes with high proximity scores in both species to define the conserved cross-species network using thresholds: $NPS_{hr} > 3$, $NPS_h > 1$, and $NPS_r > 1$ (Figure S5A). To calculate the significance of the network co-localization, we compared the conserved network size and the mean NPS_{hr} to a permuted null distribution. We permuted the labels of NPS_r 10,000 times, and each time calculated the mean NPS_{hr} across all genes and the number of genes passing the above thresholds. For genes present in both the human and rat input sets, labels were permuted separately to maintain the higher expected distribution for these genes. The significance of the conserved network size and mean NPS_{hr} was calculated by Z-test.

Definition of the species-specific BMI networks

We also defined human-specific ($n = 925$ genes) and rat-specific ($n = 688$ genes) BMI networks to investigate the biology underlying BMI in only one species (Figures S5B and S5C). We used modified individual and combined NPS thresholds to capture genes with high NPS in only one species. For the individual thresholds on NPS_h and NPS_r , we set cutoffs based on the NPS distributions, calculated from non-seed genes. For the same-species threshold (e.g., the threshold on NPS_r for defining the rat-specific network), we required that the NPS be a minimum of one standard deviation above the mean in that species. Therefore, for the human-specific network, we required $NPS_h > 1.5$, and for the rat-specific network, we required $NPS_r > 1$. For the other-species threshold (e.g., the threshold on NPS_h for defining the rat-specific network), we required that the NPS be below the mean of all scores in that species, resulting in a threshold of $NPS_r < 0$ for the human-specific network and $NPS_h < 0.5$ for the rat-specific network. We further defined combined thresholds as $NPS_r(NPS_h - 1) < -2$ for the rat-specific network and $NPS_h(NPS_r - 1) < -4$ for the human-specific network. Compared to the $NPS_{hr} > 3$ threshold for the conserved network, the shift in directionality captured genes with high NPS in one species but not the other, and the $NPS - 1$ transformation allowed the inclusion of genes with slightly positive, but below the mean, NPS in the other species. We set the combined thresholds (-2 and -4) to generate subnetworks of comparable size (500–1000 genes) to the conserved network, resulting in a stricter threshold for the human-specific threshold (Figure S5D). Finally, we defined an expanded network comprising all genes with an NPS score one standard deviation above the mean in at least one species ($NPS_h > 1.5$ or $NPS_r > 1$) to enable the analysis of the conserved and species-specific networks under a common framework ($n = 6025$). By design, this expanded network contained the conserved, rat-specific, and human-specific networks as subnetworks.

Gene set enrichment analysis

Sets of genes were assessed for biological function by quantifying significant enrichment in Gene Ontology (GO) Biological Process terms using gProfiler.⁹⁷ The gProfiler g:GOST method maps genes to functional terms and assesses the statistical significance of functional enrichment. We maintained communities with 1) at least five genes, 2) at least three genes mapped to a GO term of

size 50–1,000 genes, and 3) an FDR-corrected gene-set enrichment p value $< 1 \times 10^{-3}$. Enrichment results for seed genes and network sets are included in Table S2.

Construction of multiscale systems maps

We generated a multiscale systems map of the conserved BMI network using the Hierarchical community Decoding Framework (HiDeF)⁹⁶ algorithm as implemented in Cytoscape’s Python package CDAPS.¹¹³ HiDeF is an efficient hierarchical community detection method that makes use of persistent homology to identify structures at all scales simultaneously. First, the interaction network is reformulated as a fully connected similarity network and scanned across a range of modularity resolutions to identify densely connected network communities using a Louvain algorithm for community detection. At high resolutions, community detection tends to identify a larger number of small communities, whereas at low resolutions it tends to identify a smaller number of large communities. From this pool of all candidate communities, persistent communities are then defined as those that are redundantly identified across resolutions based on pairwise Jaccard similarities. Finally, the communities are organized into a hierarchical structure by finding those that are partially or fully nested within others based on the *containment index* (CI, Equation 4) which measures the overlap between communities v and w .

$$CI(v, w) = \frac{|s(v) \cap s(w)|}{|s(w)|} \quad (\text{Equation 4})$$

If $CI(v, w)$ is greater than a threshold σ , then an edge is added in the hierarchy from v to w . Any redundant relations are then reduced to produce the final multiscale hierarchy. To identify the conserved BMI systems map, we implemented HiDeF with maximum resolution set to 10, and all other parameters set to defaults. The resulting communities were annotated based on significantly enriched GO Biological Processes using gProfiler,⁹⁷ followed by manual curation (See [gene set enrichment analysis](#)). For communities containing multiple seed genes colocated on the same chromosome arm within 250kb, we verified that these seed genes were drawn from multiple independent loci (Table S6) to ensure that systems did not over-represent a single locus.

We performed the same hierarchical community detection and annotation procedure for the expanded BMI network, retaining all communities with more than ten genes to form the expanded BMI Systems Map. For each community, we calculated the enrichment of genes from each of the conserved, rat-specific, and human-specific networks via a hypergeometric test followed by Benjamini-Hochberg correction of p values.

Mouse variant validation

We accessed mouse knockout, knock-in, and mutation data from the Mouse Genome Database,¹¹⁴ including genotype-phenotype associations mapped to the Mammalian Phenotype Ontology⁸⁴ (MPO), similar to the analysis previously published.^{34,35} We extracted mouse genes from the listed alleles, including the target genes from conditional genotypes, and mapped them to human orthologs using MGD’s marker and ortholog mappings. We used a set of BMI-relevant phenotypes within the MPO (MP:0001697 - abnormal embryo size; MP:0010866 - abnormal prenatal body size; MP:0003956 - abnormal postnatal body size; MP:0005451 - abnormal body composition) to assess relevance to BMI. For the full conserved network and each community, we reported the odds ratio of the number of genes linked to a phenotype in the MGD and tested this value for enrichment using a Fisher’s Exact Test. The validation rate of genes within a network was defined as the number of genes associated with at least one BMI-relevant phenotype divided by the number of genes with at least one normal or abnormal phenotype reported in MGD. The MPO is hierarchical, so we considered all genes associated with a phenotype to be associated with all parent phenotypes in the MPO hierarchy. When performing multiple simultaneous tests, we used the Benjamini-Hochberg (BH) correction to calculate adjusted q values with FDR = 0.05.

We tested all communities in the conserved and expanded systems maps for the enrichment of genes with mouse orthologs associated with BMI-relevant phenotypes in MGD. For the conserved systems map, we chose communities with the highest significant odds ratios ($p < 0.05$) for at least one BMI-relevant phenotype for further investigation. To assess the functional effects of the gene communities, for each phenotype term, we calculated the positive predictive value (PPV) from the set of genes in the community (*community*), and the set of genes associated with the phenotype in MGD (*phenotype*) as:

$$PPV_{phenotype} = \frac{|community \cap phenotype|}{|phenotype|} \quad (\text{Equation 5})$$

We took a top-down approach from the root “mammalian phenotype” term (MP:0000001). At each level of the hierarchy, we kept all child terms with $PPV_{child} > PPV_{parent}$ and performed a Benjamini-Hochberg correction on the enrichment p values for the set of child terms for each parent phenotype. A representative subset of higher-level terms was selected to enable readability in the phenotype enrichment heatmaps (Figures 4 and 5D).

Tissue enrichment

To assess the tissue-specific expression of genes in the conserved and species-specific networks, we used the Tissue-Specific Enrichment Tool (TSEA,⁹⁸ <http://genetics.wustl.edu/jdlab/tsea/>) and the TissueEnrich⁹⁹ tool (<http://tissueenrich.gdcb.iastate.edu/>). We used TSEA to calculate the enrichment of gene sets for selective expression in 25 tissue types based on human GTEx¹¹⁵

data, and defined an FDR-corrected p value threshold of $q < 0.05$ as a significant enrichment. We used TissueEnrich to calculate the enrichment of mouse orthologs of gene sets for tissue-specific genes in 17 mouse tissue types based on mouse ENCODE data.¹¹⁶ We defined significant enrichments as those with an FDR-corrected p value threshold of $q < 0.05$.

Network visualization

We generated all hierarchy and network figures using Cytoscape.¹⁰⁰ Tree and radial layouts were used as the basis for all hierarchy figures (Figures 3A, 5B, 6E, S3B, and S6B), and a spring-embedded layout algorithm was used to determine the layouts of network and subnetwork figures (Figures 3C, 4A, 5A, S3A, and S4). For communities in systems maps, node graphics were used to display the source of the genes, and node size was used to indicate the number of genes. For gene networks, node color was used to indicate the source of genes.