



Exploratory locomotion, a predictor of addiction vulnerability, is oligogenic in rats selected for this phenotype

Zhifeng Zhou^{a,1}, Peter Blandino^{b,c,1}, Qiaoping Yuan^a, Pei-Hong Shen^a, Colin A. Hodgkinson^a, Matti Virkkunen^d, Stanley J. Watson^{b,c}, Huda Akil^{b,c,2}, and David Goldman^{a,2}

^aLaboratory of Neurogenetics, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda, MD 20852; ^bMolecular and Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, MI 48109; ^cDepartment of Psychiatry, University of Michigan, Ann Arbor, MI 48109; and ^dDepartment of Psychiatry, University of Helsinki, 00014 Helsinki, Finland

Contributed by Huda Akil, May 10, 2019 (sent for review November 30, 2018; reviewed by Elissa J. Chesler and Robert Hitzemann)

Artificially selected model organisms can reveal hidden features of the genetic architecture of the complex disorders that they model. Addictions are disease phenotypes caused by different intermediate phenotypes and pathways and thereby are potentially highly polygenic. High responder (bHR) and low responder (bLR) rat lines have been selectively bred (b) for exploratory locomotion (EL), a behavioral phenotype correlated with novelty-seeking, impulsive response to reward, and vulnerability to addiction, and is inversely correlated with spontaneous anxiety and depression-like behaviors. The rapid response to selection indicates loci of large effect for EL. Using exome sequencing of HR and LR rats, we identified alleles in gene-coding regions that segregate between the two lines. Quantitative trait locus (QTL) analysis in F2 rats derived from a bHR × bLR intercross confirmed that these regions harbored genes affecting EL. The combined effects of the seven genome-wide significant QTLs accounted for approximately one-third of the total variance in EL, and two-thirds of the variance attributable to genetic factors, consistent with an oligogenic architecture of EL estimated both from the phenotypic distribution of F2 animals and rapid response to selection. Genetic association in humans linked *APBA2*, the ortholog of the gene at the center of the strongest QTL, with substance use disorders and related behavioral phenotypes. Our finding is also convergent with molecular and animal behavioral studies implicating *Apba2* in locomotion. These results provide multilevel evidence for genes/loci influencing EL. They shed light on the genetic architecture of oligogenicity in animals artificially selected for a phenotype modeling a more complex disorder in humans.

addiction | genetics | locomotion | novelty seeking | oligogenic

Addictions are common, etiologically complex disorders that lead to diverse morbid outcomes and early mortality. Environmental and genetic factors play important roles in vulnerability and resilience to addiction, and the unfolding of the addiction process (1, 2). Although addictive disorders, defined in the *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition* (DSM5) as use disorders, are contingent on availability and use of an addictive agent, people are differentially vulnerable at multiple stages of the cycle of addiction, including the initial propensity to use, transition from use to abuse, long-term addiction, and relapse (3). Genetic variation influences risks, responses, and trajectories of addiction. Response to novelty and novelty seeking play a particularly important role in gene × environment correlation and interaction, because exposure and use set the stage for addiction.

Measured as extraversion, novelty seeking and response to novel environments is moderately heritable ($h^2 = 0.4$) in people. Addictions themselves are also moderately to highly heritable, as observed in large samples of twins representative of populations (4). However, despite the discovery of causal variants at several genes involved in drug metabolism, receptor/signaling and stress

resilience, the search for genetic components that influence response to novelty and vulnerability to addiction has proven challenging, with only a fraction of the genetic variance in addiction liability accounted for by genome-wide association studies (GWAS). This may be due to multiple confounding factors including genetic heterogeneity, small effect sizes of each gene/locus, and phenotypic heterogeneity resulting from variable causes for use emergent at different stages of addiction and normal development. The failure to identify loci via GWAS has been interpreted to mean that behavioral traits are polygenic and has led to speculation that behavioral diseases are omnigenic and caused by alleles of infinitesimal effect (5). As well as implicating novel genes, addiction GWAS have succeeded in recovering several genes previously implicated by study of genes and pathways known to be important in addiction, for example *ADH1B* and *ALDH2* for alcoholism and *CHRNA3* for nicotine addiction (6–8). Cellular and molecular studies of addictions, including studies of the transcriptome, have identified diverse changes

Significance

While addiction disorders have a genetic basis, the expected number of genes contributing to them is large, hence their classification as polygenic disorders. Animal models can enable discovery of genes of larger effect by targeting specific components of genetic vulnerability. High responder and low responder rats were selectively bred based on exploratory locomotion (EL) in a novel environment, a predictor of drug-seeking behavior. Using exome sequencing and quantitative trait locus analysis, we identified seven genome-wide significant loci accounting for approximately one-third of total variance and two-thirds of genetic variance in EL. We found convergent evidence for a role of *APBA2* in humans. Thus, EL is oligogenic, being strongly influenced by a limited number of loci of large effect size.

Author contributions: S.J.W., H.A., and D.G. designed research; Z.Z. and P.B. performed research; P.B., C.A.H., S.J.W., H.A., and D.G. contributed new reagents/analytic tools; Z.Z., P.B., Q.Y., P.-H.S., M.V., and D.G. analyzed data; and Z.Z., C.A.H., H.A., and D.G. wrote the paper.

Reviewers: E.J.C., The Jackson Laboratory; and R.H., Oregon Health & Science University and Portland VA Medical Center.

The authors declare no conflict of interest.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Data deposition: Sequences of the bHR/bLR exomes have been deposited in NCBI as a Bioproject (accession no. [PRJNA521139](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA521139)).

¹Z.Z. and P.B. contributed equally to this work.

²To whom correspondence may be addressed. Email: akil@umich.edu or David.Goldman@nih.hhs.gov.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1820410116/-DCSupplemental.

Published online June 10, 2019.

associated with neural adaptive processes (9–11). Perturbations of molecular components critical for neurotransmission such as glutamatergic, GABAergic, and dopaminergic systems have been recognized (12–15). However, the main successes in the discovery of processes involved in addiction have been in understanding intermediate mechanisms and tertiary consequences, rather than primary causes.

Animal models enable well-controlled designs to minimize confounding factors. Animals selectively bred for targeted phenotypes are of particular value for the identification of alleles naturally occurring in those species (16, 17) because selection enriches and even genetically fixes alleles that influence the phenotype. In such models, genetic, phenotypic, and environmental complexity can be reduced, enabling detection of loci with relatively large effect sizes. For example, two stop codons influencing alcohol preference together account for 6% of the total variance for this phenotype and were genetically fixed by selection in alcohol-preferring (P) and nonpreferring (NP) rats (18). Fixation of the *Gm2* stop codon, which is either common or genetically fixed in outbred Wistar rats from different sources, leads to uncompensated loss of metabotropic glutamate receptor 2 (mGluR2) function and glutamate-specific transcriptome changes in P rats. This finding, and other molecular neuroscience studies showing the importance of mGluR2 in responses to alcohol and cocaine, point to the importance of genes influencing glutamatergic neurotransmission in addictions and the response to artificial selection of an allele strongly influencing mGluR2.

Selectively bred high responder (bHR) and low responder (bLR) rats are a well-established model for novelty-seeking traits and vulnerability to behavioral disinhibition and responsiveness to reward-related cues based on exploratory locomotion (EL) in a novel environment (19, 20). bHR and bLR rats represent extremes of two contrasting modes of environmental interaction. bHR rats are highly exploratory, novelty-seeking, impulsive, and prone to seeking drugs under basal unstressed conditions. They show greater psychomotor sensitization to cocaine (21), greater dopamine “release event” in the core of the nucleus accumbens (19), and altered histone methylation at the dopamine receptor D2 gene promoter and D2 mRNA expression (22). By contrast, bLR rats are inhibited, exhibit greater anxiety- and depression-like behaviors, and tend to seek drugs only following repeated psychosocial stress. Thereby, the bHR and bLR rats model two distinct paths to substance abuse that have been observed in humans, where patients have been classified as externalizing or internalizing based on cross-inheritance of addictions with other disorders (20). Both outbred rat lines can learn to self-administer psychoactive drugs for extended periods, but even with equivalent exposure to the drugs, bHR are more likely to transition to addiction and are more likely to relapse (20).

EL is highly heritable in the bHR/bLR model as shown by the remarkable divergence of bHR and bLR after only a few generations of artificial selection (20). Although other behavioral phenotypes are closely cotransmitted across generations, the focus of this work is on the primary selection phenotype in this rat model: EL. To identify loci modulating EL, we performed exome-based sequencing, uncovering alleles in gene-coding regions that segregate between bHR and bLR. We then performed quantitative trait loci (QTL) analysis with enhanced coverage in the segregating genomic regions, in F2 rats derived from a bHR/bLR intercross. The discovery of genome-wide significant loci allowed us to test whether EL in our selectively bred lines is strongly influenced by a small set of loci (oligogenic) and whether the loci act epistatically or additively. To investigate whether the strongest rat QTL is relevant to human behavior, we targeted orthologous genes within the syntenic region and performed genetic association analyses in subjects phenotyped for the personality trait of novelty seeking, with and without addictions.

Results

Exome Sequencing Identifies Genetic Variants Segregating between bHR and bLR Rats. Broad sense heritability (H^2) of EL was 0.489, calculated based on variance of this phenotype in bHR and bLR F0, F1, and F2 offspring (*SI Appendix, Fig. S1*). Therefore, EL is substantially heritable and the bHR and bLR rats are potentially a model for identification of loci influencing novelty response. By applying Wright’s estimator (23, 24) to the means and variances of locomotor scores in bHR and bLR and variance in F2 offspring, we calculated that the minimum number of loci that may contribute to total variance in EL is 15.7 (*SI Appendix, Fig. S1*). EL in bHR/bLR might therefore be regarded as oligogenic, but not omnigenic, a hypothesis potentially resolvable by discovery of loci and measurement of their individual and joint contributions.

We performed exome sequencing in 12 bHR and 12 bLR F0 animals that were the product of 18 generations of selection and 37 generations of breeding, identifying genetic variants in the gene-coding regions that segregate between bHR and bLR lines (25). The exome sequencing generated adequate coverage (average 86 \times , across the 50-Mb exome target) and identified a total of 110,172 single nucleotide variants (SNVs) across the 24 individual rats sequenced. Of these, 1,584 SNVs (1.44%) showed full allelic segregation between bHR and bLR rats. We compared sequence variation in bHR and bLR rats to the selectively bred P and NP rats, produced by 70 generations of breeding. P and NP rats were previously analyzed using the same methods, including informatics pipeline (18). bHR/bLR rats had higher nucleotide heterozygosity, far fewer fully segregating variants, and a lower inbreeding coefficient (*SI Appendix, Fig. S2*), indicating that the bHR and bLR lines have relatively lower degrees of inbreeding and random fixation and thereby may be more favorable for locus identification.

To better parse the genetic differences between HR and LR rats and to map patterns and dynamics of selection and random fixation on the rat chromosomes, we used Fisher’s exact test to characterize the SNVs that either partially or fully segregated between the two lines. A total of 8,123 partially segregating SNVs were identified. Their genomic locations were graphed (*Fig. 1 and SI Appendix, Table S1*). Most genetic fixation between bHR and bLR was randomly dispersed across the genome. Although some of the regions with segregating variants (genomic islands) may result from random fixation by selective breeding, others may harbor genetic loci that influence the trait under selection. Causal significance of these genomic islands with high densities of segregating variants was evaluated by F2 linkage.

QTL Analysis Targeting Segregating Genomic Regions Identifies Seven Loci Strongly Influencing EL.

The genetic variants segregating between bHR and bLR rats identified by exome sequencing laid a foundation for whole-genome QTL analysis in F2 animals. By selecting SNVs divergent between bHR and bLR, we were able to ensure that the SNVs interrogated were informative for the bHR \times bLR cross. Of a total of 416 SNVs genotyped in 314 F2 rats derived from the intercross, 383 SNVs segregated ($P < 10^{-4}$, Fisher’s exact test). The QTL analysis was also enhanced by overtagging regions where exome sequencing indicated that causal genetic loci were likely to reside (*Fig. 1*) with significant linkage disequilibrium (LD) (average $r^2 = 0.73$) between adjacent genotyped SNVs in the regions dense with segregating variants (*SI Appendix, Fig. S3*). This approach ensured adequate coverage for QTL analysis in those regions, in addition to genome-wide coverage.

Seven genome-wide significant QTL peaks (*Fig. 24*) were identified using R/qtl (26) with the Haley–Knott regression method (1-cM resolution), and a genome-wide threshold LOD (logarithm of the odds) score of 3.99 ($\alpha = 0.05$), determined by permutation (*SI Appendix, Fig. S4*). The strongest QTL locus (LOD = 7.77) was SNV S126771200 [chromosome 1 (Chr1): 126,771,200 bp]. In

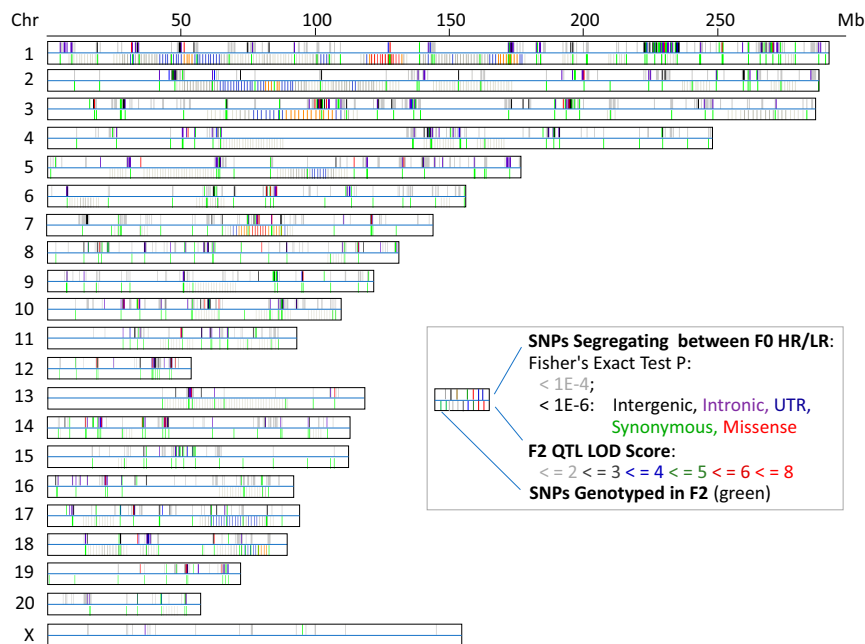


Fig. 1. Genomic locations of SNVs segregating between the F0 bHR and bLR rats and QTL signals in bHR \times bLR F2 rats. SNVs that segregate between the two F0 lines are shown in the top track with genomic positions for each chromosome. The degrees of segregation of the SNVs are noted with either gray (Fisher's exact test $P < 1E-4$) or black ($P < 1E-6$) with their locations and functions coded in color. The QTL signals are shown in the bottom track with QTL LOD scores denoted in color and the SNVs that were genotyped in the F2 rats are shown in green.

addition to the peak near S126771200, there were two other genome-wide significant QTL peaks on chromosome 1, and these were near S54809837 and S171600300. These SNVs are in weak LD with S126771200 and thereby are likely to represent at least partially independent QTLs on chromosome 1 (Fig. 2 *B* and *C*). Four other genome-wide significant QTLs were identified in regions of chromosomes 2, 3, 7, and 18. Another peak with an LOD score of 3.81 just below the genome-wide threshold was located on chromosome 17 (Fig. 2*D*). To verify robustness of the linkage results against the method used, we also used maximum likelihood (EM algorithm) and multiple imputation methods implemented in R/qtl. Consistent QTL results were obtained with those methods (*SI Appendix*, Fig. S5). Details of the genomic locations and genotype effects of the QTLs and SNVs on F2 locomotor scores are provided in Fig. 3.

We performed pairwise genome scans via a two-QTL model to detect additive and epistatic interactions. Significantly higher LOD scores were obtained with the full (F) two-QTL model but primarily at the same QTLs identified by single-locus (S) genome scan and even if all locus pairs were evaluated (*SI Appendix*, Fig. S6). The LOD scores of the full two-QTL model [LOD (F)] for the seven genome-wide significant loci ranged from 8.57 to 15.35 and the difference between the full two-QTL model and single-locus models [LOD (F – S)] ranged from 4.10 to 7.61. Except for the chromosome 1 QTLs, there were only weak epistatic interactions [LOD (I)], between pairs of QTLs (*SI Appendix*, Fig. S6). Failure to detect gene \times gene effects could be in part due to more limited power to detect interactions. Details of additive and interactive effects of the QTL-associated SNV pairs on the F2 locomotor scores are shown in Fig. 4.

To examine the individual and combined effects of the QTLs on EL, we performed a multiple-QTL fit for locomotor score (Fig. 5). To avoid overestimating or overfitting combined effects, we stringently selected the QTLs with genome-wide significance (LOD score > 3.99) determined by single locus scan. Cumulative effects were calculated by adding one QTL at a time, beginning with the QTL of strongest effect and thus testing only one model

(Fig. 5). The strongest QTL at S126771200 on chromosome 1 itself accounted for over 10% of the total variance in locomotor score ($P = 4.44 \times 10^{-8}$). The combined effect of this QTL and the second most significant QTL on chromosome 7 near S74657149 explained $\sim 17\%$ of the variance ($P = 6.13 \times 10^{-12}$), and the top three QTLs accounted for nearly 25% of the variance ($P = 3.33 \times 10^{-16}$). The combined effect of the seven loci above genome-wide significance accounted for around 32.5% of the total variance in exploratory locomotion and approximately two-thirds of the genetic variance ($H^2 = 0.489$) in this phenotype. In addition to these QTLs, there were several loci with suggestive LOD scores (LOD > 3) in other genomic regions (*SI Appendix*, Table S2), including a chromosome 17 QTL that approached genome-wide significance. Together, these results demonstrate that QTL analysis with enhanced coverage of segregating genomic regions uncovered by exome sequencing was an effective strategy for identifying genetic loci influencing EL in artificially selected bHR and bLR rats.

We tested effects of sex and kinship, the former to define sex-influenced loci and the latter to detect residual, and potentially confounding, effects of genetic background. There was a noticeable difference in locomotor scores between male and female rats (*SI Appendix*, Fig. S7*A*). When both sex and kinship were used as covariates in a linear mixed model there were some minor changes of the signals at some loci, but the overall QTL results were unchanged (*SI Appendix*, Fig. S7*B*). Using sex as a covariate did not identify additional QTLs, and there was not a sex-specific origin of any of the seven genome-wide QTLs.

Genetic Association of *APBA2* with Novelty Seeking and Drug Addiction in Human Samples. The largest F2 QTL, at S126771200, implicates a genomic region centered on 126–127 Mb of chromosome 1. To narrow this region, we examined the exome sequences of bHR and bLR F0 rats, observing that no segregating SNVs are found beyond a 500-kb region around S126771200 (Fig. 6*A* and *SI Appendix*, Fig. S8). Therefore, it is likely that the functional locus revealed by QTL analysis resides within this 500-

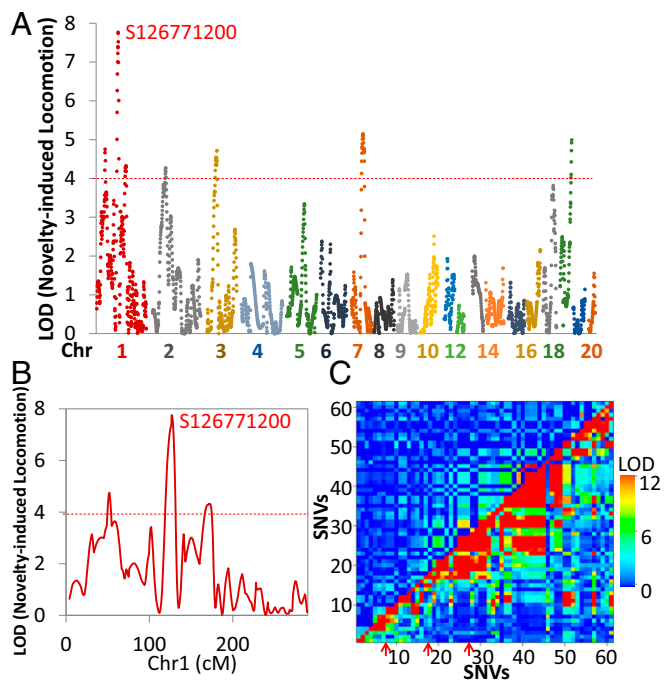


Fig. 2. Distribution of the F2 QTL LOD scores across the genome showing seven genome-wide significant QTLs for EL. (A) Manhattan plot of the QTL LOD scores [Haley–Knott (H-K) regression method implemented in R/qtl] across the genome at 1-Mb resolution. The most significant QTL SNV S126771200 on chromosome 1 is also labeled. (B) QTL peaks and LOD scores on chromosome 1. (C) Pairwise SNV recombination fractions (*Upper Left*) and LD (*Lower Right*) for all SNVs on chromosome 1. Arrows indicate the locations of the three genome-wide significant QTLs on chromosome 1.

kb region. S126771200 itself is a potentially functional missense variant (D219N) of an intronless gene, *Nsmce3*. The human ortholog *NSMCE3* is part of the SMC5–6 chromatin reorganizing

complex involved in neural development (27, 28). However, several other genes in this interval contain single nucleotide variants segregating between HR and LR F0 rats (*SI Appendix, Fig. S8*). *Apba2* [amyloid beta (A4) precursor protein-binding, family A, member 2], is located between S126348556 (LOD score 7.40) and S126771200, the variant that generated the highest LOD score. *Apba2* is of great interest for EL because it encodes a neuronal adaptor protein known to be a component of a multimeric complex mediating synaptic vesicle docking/fusion (29, 30). More importantly, *Abpa2* was shown to influence locomotor behavior in mice and knockout of *Abpa2* results in impaired response to novel objects (31). Although we only identified one intronic segregating variant (located 24 bp upstream of the fourth exon) in *Apba2*, it is possible that a genetic variant(s) in introns of *Apba2* or outside the gene coding region is responsible for the strong QTL signal.

To explore the effect of genetic variation in the human genomic region syntenic to rat Chr1 126- to 127-Mb QTL, we conducted a targeted genetic association analysis in a Finnish Caucasian sample. The sample included individuals with alcohol and drug addictions and related personality disorders, and controls free of psychiatric diagnosis. The target phenotypes were novelty seeking, measured by the Tridimensional Personality Questionnaire (TPQ), and addiction diagnoses. The human Chr15 genomic region syntenic to the rat Chr1 region has relatively narrow boundaries and includes *APBA2*, *NSMCE3* (*Ndn12*), and *FAM189A1* (Fig. 6A). A total of 149 single nucleotide polymorphisms (SNPs) were genotyped to determine LD blocks and identify tag SNPs across the region (*SI Appendix, Fig. S9*). No significant association with novelty seeking and addiction diagnosis were observed for the tag SNPs in the *NSMCE3* and *FAM189A1* gene region (Fig. 6A). In the *APBA2* gene, significant allele and genotype frequency differences were observed between controls and cases with alcohol and drug dependence and antisocial personality disorder among several tag SNPs (Fig. 6 and *SI Appendix, Fig. S9* and Table S3). The *APBA2* SNP rs8030727 was also associated with novelty seeking and the effect was more clearly observed in individuals with substance dependence and personality disorders (Fig. 6). To identify functional implications of associated SNPs, we searched the Genotype-Tissue

Loc #	Loc ID	Chr	Position (Mb)	LOD (hk)	LOD (em)	LOD (imp)	Nearest Genotyped SNV	Func/Loc	Gene Sym	AA	Sub	SIFT	Score	Polyphen	Prob
Q1	c1.loc48	1	52.1	4.75	4.91	4.61	S54809837	missense	Ttll2	R214H	Del*	0.02	Prob		1
Q2	S126771200	1	126.87	7.76	7.63	7.47		missense	Ndn12	D219N	Del	0	Damaging Poss		0.89
Q3	c1.loc168	1	172.14	3.32	4.29	4.37	S171600300	missense	C2cd3	S212F	Del	0.04	Prob		0.97
Q4	S83382082	2	83.4	4.26	4.21	4.01		syn	Adamt5						
Q5	S66926470	3	66.9	4.71	4.65	4.61		Intron	Chrna1						
Q6	c7.loc66	7	79.2	5.14	5.06	4.99	S74657149	missense	Vps13b	Y1865C	Del		Benign		0
Q7	S80534580	18	80.5	4.99	4.89	4.86		missense	Cndp1	E11D	Tol	0.48	Benign		0

* Del: deleterious; Tol: tolerated

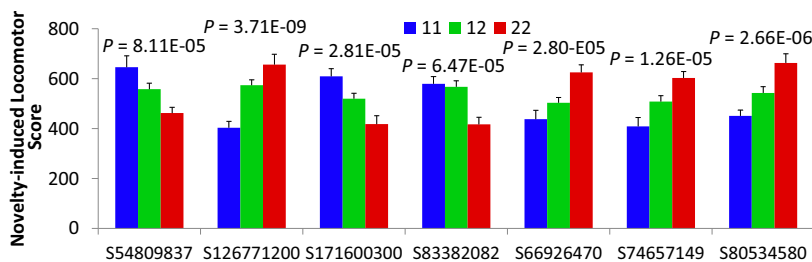


Fig. 3. Top SNVs at seven genome-wide significant QTLs on EL and their genotype effects on locomotor scores in bHR × bLR F2 rats. (Top) Genomic locations of the QTL peak SNVs, gene symbols, and the potential functional implications of the variant alleles predicted by SIFT and Polyphen. (Bottom) Genotype effects of the QTL peak SNVs on locomotor scores in the F2 rats. The numbers 11, 12, and 22 represent homozygous, heterozygous, and homozygous genotypes with “1” the reference allele. *P* values are from one-way ANOVA test.

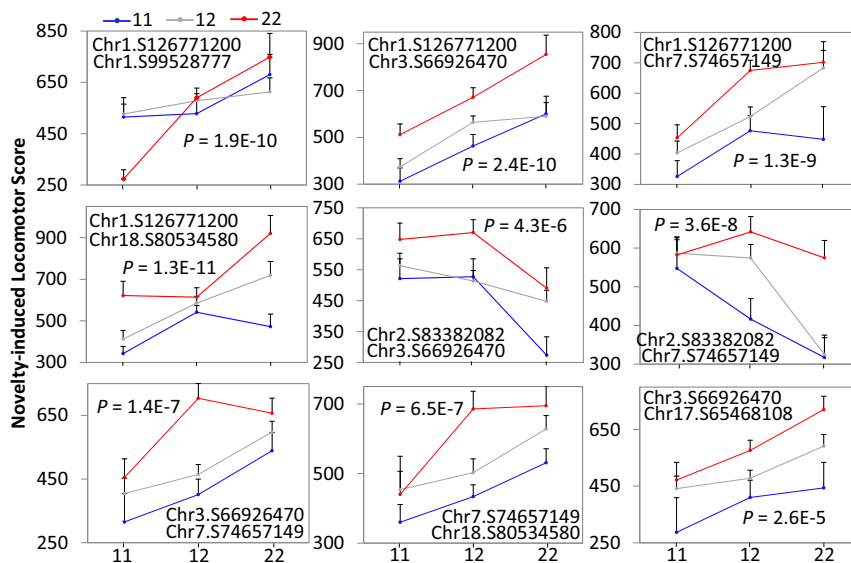


Fig. 4. Genotype effects of QTL-associated SNV pairs on locomotor scores in bHR × BLR F2 rats. Locomotor scores (least square mean and SE) of the F2 rats are plotted according to their combined genotypes at each locus pair. *P* values represent least square fits for the whole model.

Expression (GTEx) Portal <https://gtexportal.org/home> for *APBA2* eQTLs. rs8030727 and rs12439432 were identified by GTEx as significant eQTLs for *APBA2*, although the signals were identified in tissues other than the brain. rs12439432 is in an *APBA2* intronic antisense noncoding RNA gene expressed only in brain (*SI Appendix, Fig. S10*). These findings point to potential roles of the associated SNPs in cis-regulation of *APBA2* expression. *APBA2* plays critical roles in synaptic functions and neural transmission and interacts with a network of proteins involved in these functions (*SI Appendix, Fig. S11*) (32, 33). In rats, *Apba2* or a locus near it contributes a large portion of the variance in EL. The relatively more modest association signals of *APBA2* in the human sample may be attributed to the overlapping but distinctive phenotypes between the rat lines and the human sample which includes people at different stages of addiction and is subject to greater genetic and environmental variation. To evaluate the possibility that the *APBA2* association was confounded by population stratification, we used ethnic factor scores generated from a panel of ancestry informative markers (AIMs) genotyped on Illumina arrays (34). No significant difference in ethnic factor scores was observed in the Finnish sample to differentiate subjects with and without addictions (*SI Appendix, Fig. S12*).

Discussion

Genes that influence propensity to addiction are likely to do so via a series of processes that predict liability to experimentation with drugs, abuse, and conversion to addiction. In humans, twin and family studies have revealed that negative emotionality, including anxiety and depression, and externalization, including novelty seeking, impulsivity, and response to novel stimuli, are predictors of addictive liability. Via influence of genes on dimensions of behavior, addictions can be cross-inherited together, and with other disorders as well (1, 35). In rats, novelty-induced locomotion, herein referred to as exploratory locomotion, strongly predicts impulsive response to reward and vulnerability to heavy use of addictive drugs such as cocaine (19, 20, 36, 37). EL varies between strains, and is subject to genetic selection, showing moderate to high heritability.

GWAS is increasingly able to detect novel loci and pathways for complex traits in humans, but a major impediment to understanding the genetic architecture of addictions is that the loci identified only account for a small percent of the heritable var-

iance. An intriguing explanation for why most of the heritability of addictions is unexplained by GWAS is that the phenotype, as defined, is strongly polygenic. Moreover, as noted above, addiction involves multiple stages and sources of vulnerability, each of which may be influenced by different genes. Human and animal phenotypes representing specific vulnerability mechanisms such as externalization, and focused on specific addiction stages, could be causally less complex and therefore less polygenic or genetically heterogeneous. Animal models can offer unique advantages for locus discovery, by taking advantage of the ability to homogenize environmental exposures, by exploiting the repertoire of natural variation found in these species, and by taking

Number of Loci	Multiple-QTL Model	DF	SS	LOD	% Variance	P value (Chi2)	P value (F)
1	$y \sim Q2$	2	2612853	7.42	10.32	3.8E-08	4.4E-08
2	$y \sim Q2+Q6$	4	4352916	12.86	17.19	4.3E-12	6.1E-12
3	$y \sim Q2+Q6+Q5$	6	6090264	18.75	24.04	2.2E-16	3.3E-16
4	$y \sim Q2+Q6+Q5+Q7$	8	7206690	22.83	28.45	0	0
5	$y \sim Q2+Q6+Q5+Q7+Q4$	10	7596969	24.31	29.99	0	0
6	$y \sim Q2+Q6+Q5+Q7+Q4+Q1$	12	7904757	25.51	31.21	0	0
7	$y \sim Q2+Q6+Q5+Q7+Q4+Q1+Q3$	14	8251181	26.88	32.58	0	0
Total		313	25329252				

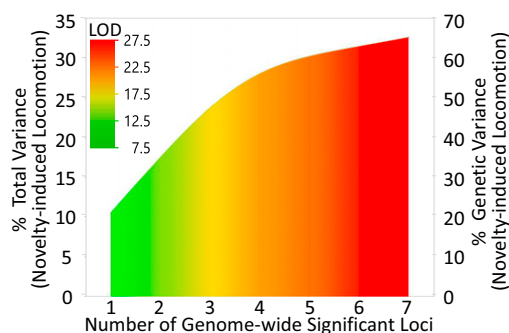


Fig. 5. Combined effects of the QTLs on EL. The multiple-QTL models including one to seven loci (see Fig. 3 for locus details) and their *P* values are shown in the table. The number of loci and the percent of the total variance each model accounts for are plotted in the *Bottom* figure. The composite LOD score for each model is represented by the color scale.

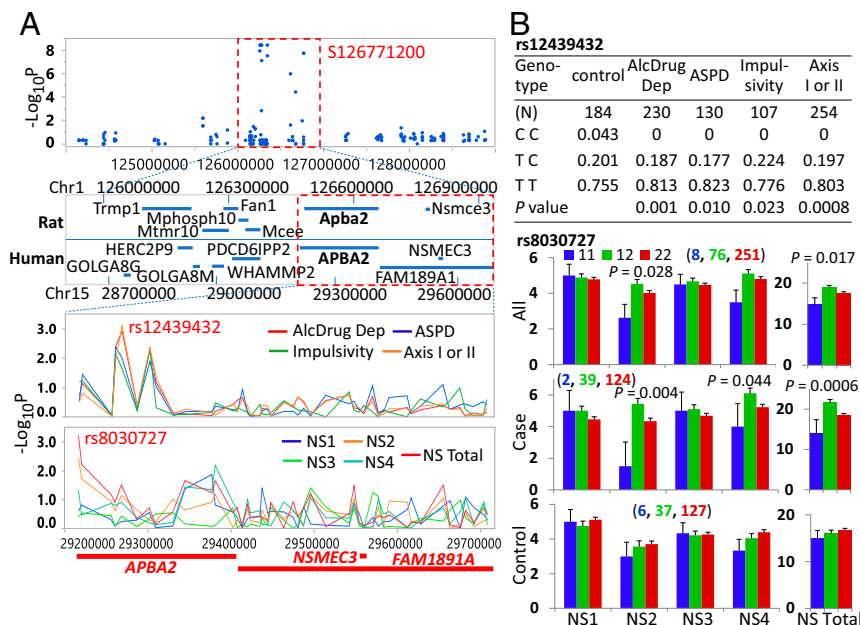


Fig. 6. Genetic association of *APBA2* SNPs with TPQ novelty seeking, alcohol/drug dependence, and related diagnoses in the Finnish sample. (A) (Top) Variants segregating in F0 bHR and bLR rats in the 124- to 129-Mb region of chromosome 1. y axis shows Fisher's exact test *P* values ($-\log_{10}$) of the variants. (Middle) Syntenic regions between rat and human within the 1-Mb location containing segregating variants are highlighted with red dashed lines. (Bottom) Association ($-\log_{10}P$) of the tag SNPs with psychiatric diagnoses (Upper) and TPQ novelty seeking subscales (Lower) across the human genomic region syntenic to the rat. (B) (Top) Genotype frequency comparisons of rs12439432 between controls and cases with the related diagnosis. (Bottom) Genotype-based association of rs8030727 with novelty seeking subscales for controls, cases, and both.

advantage of artificially selected strains in which multiple alleles that directionally contribute to a trait have been collected to high frequency, or even to genetic fixation.

Throughout the selection of the bHR and bLR lines, population size was relatively small. This directly leads to random genetic fixation and to the requirement for confirmation of loci detected by independent genetic analysis, such as the eQTL analysis performed here. However, compared with other selectively bred rat and mouse lines that were bred for many generations, bHR and bLR lines have higher degrees of heterozygosity and lower numbers of randomly fixed variants and inbreeding coefficient (*SI Appendix, Fig. S2*). This is in part due to a breeding strategy that attempted to minimize inbreeding (18). Selected strains with lower degrees of inbreeding are more suitable for locus identification, with the regions showing genetic fixation being likely to harbor variants under selection. A second, and more profound implication, is that rapid response to selection through several generations under these conditions directly indicates the presence of genes of large effect in the progenitor stock. Rapid response to artificial selection has been observed for multiple addiction-related phenotypes, including alcohol preference, alcohol withdrawal, and sedation (38). Some 50 y ago, Kimura (39) recognized that in populations with relatively few breeders (N_e) the stochastic effects of genetic drift tend to overwhelm the effects of selection, unless the locus-specific selection coefficient (*s*) is large.

Artificial selection of the HR/LR lines was initiated with 60 male and 60 female Sprague–Dawley rats and 12 litters with an approximate total of about 144 animals (12 rats \times 12 litters) being maintained each generation, with the high 20% and the low 20% being bred. In the HR/LR model, and in the many other rodent models where there may be somewhat more or less numbers of breeding pairs, N_e is less than 200. Thus, it was expected that loci of large effect would account for rapid response to selection. However, it was not obvious that genetic analysis of this externalizing behavior in a model organism would succeed in mapping loci accounting for any substantial portion of

the variance, much less loci of large effect size. Success of the QTL analysis was contingent on the stability and measurement properties of the EL phenotype, as well as the validity of the oligogenic model.

Although seven genome-wide significant loci for EL account for two-thirds of the variance attributable to genetic factors, and about one-third of the overall variance in EL, other genes contributing to the remaining portion of the variance in EL are likely to be of smaller effect, and furthermore more numerous. These findings likely indicate that a major portion of the variance in a phenotype related to novelty seeking, impulsive response to reward, and vulnerability to drug addiction in the bHR/LR rats is oligogenic, with one locus alone explaining $>10\%$ of the total variance in EL. However, in people, vulnerability to addiction can be multifactorial and attributable to a combination of phenotypes that are each influenced by multiple genes. Before molecular analyses, we hypothesized that novelty-induced locomotion in our model was neither monogenic nor polygenic because the expected minimum number of loci responsible for the trait to explain the total variance of the phenotype was 15.7 (*SI Appendix, Fig. S1*). Genes of major effect have been mapped in many artificially selected organisms, including the P/NP rat (18) and all of the way back to an early QTL paper, in which 15 QTLs of large (“Mendelian”) effect were mapped for mass, soluble solids, and pH in the tomato, the tomato having been domesticated over the generations to enhance these parameters (40).

To avoid overestimating or overfitting the combined effect size of loci on the phenotype, the QTL loci included for the multiple QTL fit were all individually genome-wide significant (LOD score > 3.99) and entered in one sequence. Epistatically acting combinations of genes might be better identified using other methods or in other contexts, for example isolated populations rather than rats artificially selected over several generations. Nevertheless, the observed locus additivity is consistent with the rapid divergence of novelty-induced locomotion through a few generations of selection (20). Epistatically acting alleles that are not

colocalized randomly assort from generation to generation, disrupting multilocus combinations. Furthermore, additivity is seen in other animal behavior models, and in people. Some 52% of the variance in “animal personality” was attributed to additive genetic factors, via a metaanalysis of 71 estimates with both heritability and repeatability (41). Lack of epistasis of gene effects in addictions is also consistent with human twin concordance data for personality traits and addictions. For personality traits, including extraversion, monozygotic/dizygotic (MZ/DZ) concordance ratios are $\sim 2:1$. For eight addictive disorders for which extensive twin data were available, the MZ/DZ twin concordance ratios could reach as high as 3.7/1 for cocaine addiction, with the MZ/DZ ratio for most addictive disorders hovering at $\sim 2:1$ (4). These twin data are informative because epistasis, as distinct from polygenicity, requires gene combinations. Multilocus combinations are conserved between pairs of MZ twins, but are unlikely to be shared between pairs of dizygotic twins.

A significant challenge for behavioral QTL analysis in animal models or genetic association studies in human populations is identification of functional loci/genes in the linked or associated regions. Some animal models display extended LD or greater genetic diversity, impeding locus identification. However, the exome sequencing of bHR/bLR F0 lines provided valuable information for the most significant QTL locus centered at S126771200 on chromosome 1, where segregating variants between bHR and bLR were only identified in a relatively small region (*SI Appendix, Fig. S8*). It is unlikely that the functional locus/gene is located outside the narrow interval implicated by both QTL analysis and genetic segregation of variants observed via exome sequencing. Two genes, *Nsmce3* (*Ndn12*) and *Apba2*, located near the center of this region, are of interest due to their functions. *Nsmce3* is part of the SMC5–6 chromatin reorganizing complex involved in neural development (27, 28). However, we did not detect association of *NSMCE3* in humans. *Apba2* was observed in previous studies to play roles in neural development (30) and locomotor behavior in mice (31). Genetic association of *APBA2* in our human population sample, the rat QTL results, and previous findings on *Apba2* and locomotion convergently implicate *APBA2*.

Apba2 encodes a neuron-specific multimodular adaptor protein that functions in vesicular transport and signal transduction (31, 42). In addition to amyloid precursor protein (APP), *APBA2* interacts physically with several other proteins critically involved in synaptic functions (*SI Appendix, Fig. S10A*), including neuroligin 1 (NRXN1) and neuroligin 3 (NLGN3)—two members of the Ca²⁺-dependent receptor/ligand complex involved in the formation of the synaptic contacts and essential for neurotransmission, calyntenin 1 (CLSTN1), and calyntenin 3 (CLSTN3)—members of a subset of the cadherin superfamily known to mediate axonal vesicle transport, and syntaxin binding protein 1 (STXB1) involved in regulation of neurotransmitter release. Gene ontology analysis of the *APBA2* gene network also revealed significant overrepresentation of the genes in the gene ontology groups involved in synapse assembly, regulation of synaptic transmission, cell adhesion, and neuron parts (*SI Appendix, Fig. S11B*). Beyond Alzheimer’s disease, studies have also implicated *APBA2* in other neurological and psychiatric disorders (43–45). These lines of evidence indicate the critical roles of *APBA2* and related genes in synaptic functions and neurotransmission, further tying addictions and the externalizing domain to these functions.

Except for rs8040932, we did not use *APBA2* coding variants for genetic association analysis, due to the fact that few *APBA2* coding variants are common (minor allele frequency > 0.01) (46) (*SI Appendix, Table S4*), but we cannot rule out that rare *APBA2* coding sequence variants are also associated with behavior. The locus driving the association probably alters RNA expression, and as hinted by GTEx data we hypothesize that the linked *APBA2* SNPs are cis-eQTLs for *APBA2*. Effects of genes on

addictive disorders themselves are likely to be weaker than effects on intermediate phenotypes. The relatively modest association signals of *APBA2* in the human sample may be attributed to the phenotypic heterogeneity which includes subjects differing in novelty seeking with or without addiction and individuals at different stages of addiction. Furthermore, the effect of *APBA2* may be diminished by more diverse genetic background, environmental variance, and smaller effect size of *APBA2* alleles humans may be carrying. This also indicates the need for further replication of genetic association of this gene in independent samples and functional knockout or knockdown in animal studies for the validation of causal effects.

In sum, selectively bred animal lines provide valuable models for genome-based identification of loci influencing addictions and other behavioral traits. In human studies, addictions have proven to be genetically complex and resistant to genetic analysis by GWAS, with only a small portion of the variance accounted for and no loci of large effect having been discovered, or likely to be discovered in the future by comparison of DSM-defined cases to controls. EL in rats, and externalization in people, are traits critical for at least one key component of addiction liability. However, externalization in people, while it may be modeled by EL, is not itself the direct equivalent of EL, an experimentally measured parameter. The bHR/bLR model is an experimental model for externalization and is also unlikely to have interrogated all genes potentially influencing externalization in the rat. Furthermore, the effects of genes selected for novelty response were not isolated to one behavior but were pleiotropic. Our approach combining exome-based genomic sequencing and enhanced QTL analysis of rats artificially selected for novelty-induced locomotion identified seven loci of major effect, accounting for a large portion of the variance of novelty-induced locomotion, a behavioral trait closely related to impulsive response to reward and vulnerability to drug use. Furthermore, we were able to show that the effects of these loci are primarily additive, consistent with human twin data, and as such may be useful in guiding searches for addiction genes in humans. The converging evidence of our human genetic association analysis and the molecular and behavioral analyses of other studies link *Apba2*, a gene at the center of the strongest QTL, to response to novelty and at least to the early stage of drug-seeking behavior seen in externalizing disorders. The possible role of these genes in subsequent stages of drug abuse, including conversion to addiction and propensity to relapse, requires further study.

Materials and Methods

Subjects.

Animals. Selectively bred HR and LR rats were established at the University of Michigan and have been described in detail (19–21). A total of 12 bHR and 12 bLR rats (six male and six female each) at breeding generation 37 were sequenced. The F2 rats were created by reciprocally crossing bHRs and bLRs and by selecting the most phenotypically extreme male and female from each of the 12 bHR and 12 bLR families in generation 37 to maximize phenotypic variance in the F2 animals. Thus, we set up 24 bHR–bLR mating pairs in the F0 generation, and this created 24 F1 litters, comprised of 261 F1 male and female rats, all of which were phenotyped. The next round of breeding, involving 48 F1 mating pairs, yielded 48 F2 litters for a total of 540 male and female F2 rats. A total of 323 F2 rats were phenotyped for exploratory locomotion (20). All research protocols were approved by the University of Michigan animal care and use committee.

Human samples. The Finnish Caucasian sample consisted of a total of 443 unrelated males. There were 188 controls free of psychiatric disorders and 255 individuals diagnosed as having various psychiatric disorders, including alcohol and drug dependence and abuse, antisocial personality disorder, impulsivity, and other personality disorders. They were studied after informed consent was obtained under a human research protocol approved by the National Institute of Mental Health and the University of Helsinki institutional review boards. The Structured Clinical Interview for DSM-III-R was administered to all of the participants, and DSM-III-R lifetime psychiatric

diagnoses were blind rated. Findings from this population sample have been reported previously with additional details (47).

Exome library construction, sequencing base calling, mapping, and SNV calling. Exome sequencing libraries were constructed using Agilent SureSelect protocols. Briefly, HR and LR rat genomic DNA was fragmented to around a 150-base pair range by sonication. The genomic fragments were ligated with universal adaptors and amplified by PCR for eight cycles. Exonic regions were captured by hybridization with the Mouse All Exon bait library at 60°C for 24 h. The captured exome genomic libraries were amplified again by PCR for eight cycles with individually barcoded primers before sequencing.

Sequencing of exomes was carried out on a SOLiD 5500 Wildfire sequencer. Sequences of 50-base single reads were called from image files and aligned to the rat reference genome (RGSC 5.0/rn5) using the ABI pipeline. An average of 74× coverage with uniquely mapped exome reads for each sample was obtained. SNVs from exome-seq reads were called with LifeScope-2.5.1 with default settings. The sequenced base counts at individual SNV locations for each sample were parsed from the mapped reads.

Ampliseq Genotyping in dHR × bLR F2 Rats. Sequencing multiplexing amplicons (final $n = 416$) for SNV genotyping in the F2 rats were custom designed by Life Technologies. Ampliseq libraries were constructed using the standard protocol. Briefly, F2 rat DNA was amplified by PCR in a 96-well (sample) format with the Ampliseq primer pool for 16 cycles. The amplicon pools were ligated with a universal adaptor and sample-specific barcoded adaptor. Amplicon libraries were then pooled (96 samples per pool), amplified on the Ion Sphere Particles, and purified before being sequenced on an Ion Proton sequencer. The sequenced SNV amplicon reads were mapped to the rat reference genome (RGSC 5.0/rn5) and genotypes were called for each variant. An average of 162× coverage was obtained for the genotyped SNVs. Both SNVs and samples having less than 20% missed calls were included in

the final analysis. Genotypes for selected SNVs were also validated by TaqMan genotyping with a concordance of 0.9995.

Human Sample Genotyping. Genotyping of SNPs was carried out by selected TaqMan assays and Affymetrix Axiom custom array-based analysis. The TaqMan assays were either predesigned or custom designed (Life Technologies) and genotyping results were analyzed on a 7900 instrument (Life Technologies).

QTL Analysis and Statistics. F2 locomotor score QTL analyses were performed using R/qtl (26). Haley–Knott regression was used for single-locus and pairwise genome scans. Maximum likelihood (EM algorithm) and multiple imputation methods were also used for comparisons. Genome-wide significance was determined by permutation test ($n = 1,000$, $\alpha = 0.05$) by R/qtl. The multiple-QTL fit was carried out using the multiple imputation method. The additive and interactive effects of the individual SNV pairs were tested using the standard least squares fit (JMP, 11.0.0, SAS Institute, Inc). Fisher's exact test was performed to test allelic segregations of the SNVs between HR and LR lines. For human sample genetic association analysis, haplotypes were constructed from genotyping data using PHASE (48). Distributions of alleles, genotypes, haplotypes, and diplotypes between case and control were analyzed using χ^2 test (likelihood ratio test). Genotyped-based analysis of quantitative measures were analyzed by one-way ANOVA.

ACKNOWLEDGMENTS. We thank Drs. Jun Z. Li and Megan Hagenauer (University of Michigan) for critical reading and feedback on this manuscript and Dr. Elaine Hebda Bauer for input on technical issues relating to the bHR/bLR colony. This work was supported by the Intramural Research Program of the National Institute on Alcohol Abuse and Alcoholism; National Institute on Drug Abuse U01DA043098 (to H.A. and Jun Li, co-PIs); NIH Grant R01MH104261 (to H.A. and S.J.W., co-PIs); Office of Naval Research Grant N00014-12-1-0366 (to H.A., PI); and the Hope for Depression Research Foundation (H.A., PI).

- K. S. Kendler, C. A. Prescott, J. Myers, M. C. Neale, The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Arch. Gen. Psychiatry* **60**, 929–937 (2003).
- M. D. Li, R. Cheng, J. Z. Ma, G. E. Swan, A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins. *Addiction* **98**, 23–31 (2003).
- L. E. Kwako, R. Momenan, R. Z. Litten, G. F. Koob, D. Goldman, Addictions neuroclinical assessment: A neuroscience-based framework for addictive disorders. *Biol. Psychiatry* **80**, 179–189 (2016).
- D. Goldman, G. Oroszi, F. Ducci, The genetics of addictions: Uncovering the genes. *Nat. Rev. Genet.* **6**, 521–532 (2005).
- E. A. Boyle, Y. I. Li, J. K. Pritchard, An expanded view of complex traits: From polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).
- W. Berrettini et al., Alpha-5/alpha-3 nicotinic receptor subunit alleles increase risk for heavy smoking. *Mol. Psychiatry* **13**, 368–373 (2008).
- J. Frank et al., Genome-wide significant association between alcohol dependence and a variant in the ADH gene cluster. *Addict. Biol.* **17**, 171–180 (2012).
- E. E. Quillen et al., ALDH2 is associated to alcohol dependence and is the major genetic determinant of “daily maximum drinks” in a GWAS study of an isolated rural Chinese sample. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **165B**, 103–110 (2014).
- G. F. Koob, N. D. Volkow, Neurocircuitry of addiction. *Neuropsychopharmacology* **35**, 217–238 (2010).
- E. J. Nestler, Transcriptional mechanisms of drug addiction. *Clin. Psychopharmacol. Neurosci.* **10**, 136–143 (2012).
- Z. Zhou, Q. Yuan, D. C. Mash, D. Goldman, Substance-specific and shared transcription and epigenetic changes in the human hippocampus chronically exposed to cocaine and alcohol. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 6626–6631 (2011).
- L. Chen, D. M. Segal, C. T. Moraes, D. C. Mash, Dopamine transporter mRNA in autopsy studies of chronic cocaine users. *Brain Res. Mol. Brain Res.* **73**, 181–185 (1999).
- M. A. Enoch et al., GABAergic gene expression in postmortem hippocampus from alcoholics and cocaine addicts; corresponding findings in alcohol-naïve P and NP rats. *PLoS One* **7**, e29369 (2012).
- W. X. Tang, W. H. Fasulo, D. C. Mash, S. E. Hemby, Molecular profiling of midbrain dopamine regions in cocaine overdose victims. *J. Neurochem.* **85**, 911–924 (2003).
- N. D. Volkow et al., Cocaine cues and dopamine in dorsal striatum: Mechanism of craving in cocaine addiction. *J. Neurosci.* **26**, 6583–6588 (2006).
- J. C. Crabbe, Genetic animal models in the study of alcoholism. *Alcohol Clin. Exp. Res.* **13**, 120–127 (1989).
- K. R. Kaun, R. Azanchi, Z. Maung, J. Hirsh, U. Heberlein, A Drosophila model for alcohol reward. *Nat. Neurosci.* **14**, 612–619 (2011).
- Z. Zhou et al., Loss of metabotropic glutamate receptor 2 escalates alcohol consumption. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 16963–16968 (2013).
- S. B. Flagel et al., An animal model of genetic vulnerability to behavioral disinhibition and responsiveness to reward-related cues: Implications for addiction. *Neuropsychopharmacology* **35**, 388–400 (2010).
- J. D. Stead et al., Selective breeding for divergence in novelty-seeking traits: Heritability and enrichment in spontaneous anxiety-related behaviors. *Behav. Genet.* **36**, 697–712 (2006).
- M. J. García-Fuster, J. A. Perez, S. M. Clinton, S. J. Watson, H. Akil, Impact of cocaine on adult hippocampal neurogenesis in an animal model of differential propensity to drug abuse. *Eur. J. Neurosci.* **31**, 79–89 (2010).
- S. B. Flagel et al., Genetic background and epigenetic modifications in the core of the nucleus accumbens predict addiction-like behavior in a rat model. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E2861–E2870 (2016).
- R. Lande, The minimum number of genes contributing to quantitative variation between and within populations. *Genetics* **99**, 541–553 (1981).
- Z. B. Zeng, D. Houle, C. C. Cockerham, How informative is Wright's estimator of the number of genes affecting a quantitative character? *Genetics* **126**, 235–247 (1990).
- Zhou Z, Yuan Q, Akil H, Goldman D, HR and LR rats exome sequencing. BioProject. <https://www.ncbi.nlm.nih.gov/bioproject/521139>. Deposited 6 February 2019.
- K. W. Broman, H. Wu, S. Sen, G. A. Churchill, R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890 (2003).
- K. Kuwako, H. Taniura, K. Yoshikawa, Necdin-related MAGE proteins differentially interact with the E2F1 transcription factor and the p75 neurotrophin receptor. *J. Biol. Chem.* **279**, 1703–1712 (2004).
- E. M. Taylor, A. C. Copesey, J. J. Hudson, S. Vidot, A. R. Lehmann, Identification of the proteins, including MAGEG1, that make up the human SMCS-6 protein complex. *Mol. Cell. Biol.* **28**, 1197–1206 (2008).
- T. Biederer, T. C. Südhof, Mints as adaptors. Direct binding to neuroligins and recruitment of munc18. *J. Biol. Chem.* **275**, 39803–39806 (2000).
- Y. Zhang et al., Interaction of Mint2 with TrkA is involved in regulation of nerve growth factor-induced neurite outgrowth. *J. Biol. Chem.* **284**, 12469–12479 (2009).
- Y. Sano et al., X11-like protein deficiency is associated with impaired conflict resolution in mice. *J. Neurosci.* **29**, 5884–5896 (2009).
- D. Warde-Farley et al., The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* **38**, W214–W220 (2010).
- H. Mi, A. Muruganujan, J. T. Casagrande, P. D. Thomas, Large-scale gene function analysis with the PANTHER classification system. *Nat. Protoc.* **8**, 1551–1566 (2013).
- C. A. Hodgkinson et al., Addictions biology: Haplotype-based analysis for 130 candidate genes on a single array. *Alcohol Alcohol* **43**, 505–515 (2008).
- D. Goldman, A. Bergen, General and specific inheritance of substance abuse and alcoholism. *Arch. Gen. Psychiatry* **55**, 964–965 (1998).
- D. Belin, A. C. Mar, J. W. Dalley, T. W. Robbins, B. J. Everitt, High impulsivity predicts the switch to compulsive cocaine-taking. *Science* **320**, 1352–1355 (2008).
- N. Vanhille, A. Belin-Rauscent, A. C. Mar, E. Ducret, D. Belin, High locomotor reactivity to novelty is associated with an increased propensity to choose saccharin over cocaine: New insights into the vulnerability to addiction. *Neuropsychopharmacology* **40**, 577–589 (2015).
- J. C. Crabbe, Review. Neurogenetic studies of alcohol addiction. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **363**, 3201–3211 (2008).
- M. Kimura, Evolutionary rate at the molecular level. *Nature* **217**, 624–626 (1968).
- A. H. Paterson et al., Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* **335**, 721–726 (1988).

41. N. A. Dochtermann, T. Schwab, A. Sih, The contribution of additive genetic variation to personality variation: Heritability of personality. *Proc. Biol. Sci.* **282**, 20142201 (2015).
42. Y. Nakajima *et al.*, Neuronal expression of mint1 and mint2, novel multimodular proteins, in adult murine brain. *Brain Res. Mol. Brain Res.* **92**, 27–42 (2001).
43. T. D. Babatz, R. A. Kumar, J. Sudi, W. B. Dobyns, S. L. Christian, Copy number and sequence variants implicate APBA2 as an autism candidate gene. *Autism Res.* **2**, 359–364 (2009).
44. G. Kirov *et al.*, Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. *Hum. Mol. Genet.* **17**, 458–465 (2008).
45. S. E. Sullivan, G. M. Dillon, J. M. Sullivan, A. Ho, Mint proteins are required for synaptic activity-dependent amyloid precursor protein (APP) trafficking and amyloid β generation. *J. Biol. Chem.* **289**, 15374–15383 (2014).
46. M. Lek *et al.*; Exome Aggregation Consortium, Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
47. Z. Zhou *et al.*, Haplotype-based linkage of tryptophan hydroxylase 2 to suicide attempt, major depression, and cerebrospinal fluid 5-hydroxyindoleacetic acid in 4 populations. *Arch. Gen. Psychiatry* **62**, 1109–1118 (2005).
48. M. Stephens, N. J. Smith, P. Donnelly, A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68**, 978–989 (2001).