Rapid Evolution of Genes Involved in Learning and Energy Metabolism for Domestication of the Laboratory Rat

Lin Zeng,[1,2] Chen Ming,[2,3] Yan Li,[1,4] Ling-Yan Su,[2,5] Yan-Hua Su,[6] Newton O. Otecko,[1,2] He-Qun Liu,[1] Ming-Shan Wang,[1] Yong-Gang Yao,[5] Hai-Peng Li,[3] Dong-Dong Wu,*[,1] and Ya-Ping Zhang*[,1]

[1]State Key Laboratory of Genetic Resources and Evolution, Yunnan Laboratory of Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China

[2]University of Chinese Academy of Sciences, Beijing, China

[3]CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

[4]State Key Laboratory for Conservation and Utilization of Bio-resources in Yunnan, Yunnan University, Kunming, China

[5]Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences & Yunnan Province, Kunming Institute of Zoology, Kunming, China

[6]College of Animal Science and Technology, Yunnan Agricultural University, Kunming, China

***Corresponding authors:** E-mails: wudongdong@mail.kiz.ac.cn; zhangyp@mail.kiz.ac.cn.

Associate editor: Gregory Wray

Abstract

The laboratory rat, widely used in biomedical research, is domesticated from wild brown rat. The origin and genetic mechanism underlying domestication of the laboratory rat remain largely elusive. In the present study, large scale genomes supported a single origin for the laboratory rat, possibly from a sister group to wild rats from Europe/Africa/Middle East. Genomic and transcriptomic analyses uncovered many artificially selected genes (e.g., *FOXP2*, *B3GAT1*, and *CLOCK*) involved in the nervous system. These genes associate with learning ability and regulation of circadian rhythm, which likely enabled the successful domestication of the laboratory rat. Particularly, many genes, including mitochondrial genes responsible for energy metabolism, displayed a substantially increased expression in the brain of laboratory rats compared with wild rats. Our findings demystify the origin and evolution of this model animal, and provide insight into the process of its domestication.

Key words: **laboratory rat, origin, domestication, artificial selection.**

Letter

The laboratory rat, which has been widely used in biomedical research as an animal model for ~160 years, is commonly believed to be domesticated from wild brown rat in Europe in the 19th century (Suckow and Weisbroth 2006). Compared with their wild ancestors, laboratory rats exhibit different morphological, physiological, and behavioral attributes such as coat color, organ size, energy metabolism, reproductive performance, and tameness (Whishaw and Kolb 2004; Baker et al. 2013). However, the genetic mechanisms underlying these phenotypic differences remain elusive.

To explore the origin and evolution of the laboratory rat, we generated whole genomes for six wild brown rats from Europe, Africa, and Asia, as well as nine laboratory rats (SD and SHR strains) with ~10–20× depth in the present study (supplementary table S1, Supplementary Material online). We also added 110 whole genome sequences of wild brown rats from China ($n = 38$), Russia ($n = 5$), Southeast Asia ($n = 8$), Middle East ($n = 12$), Europe ($n = 26$), and Africa ($n = 21$) generated by one of our studies (Zeng L et al., unpublished data), and another 24 whole genome sequences of laboratory strains generated from a previously reported study (Atanur et al. 2013).

In order to understand the relationship between laboratory rats and wild rats, we performed a host of analyses including phylogenetic tree construction (fig. 1), Bayesian clustering analysis by ADMIXTURE (Alexander et al. 2009) (fig. 1) and principal components analysis (PCA) (supplementary fig. S1, Supplementary Material online). All these analyses supported a single origin of the laboratory rat (Kuramoto et al. 2012). It is commonly believed that the laboratory rat was domesticated in Europe (Suckow and Weisbroth 2006). However, our analyses illustrated that laboratory rats did not cluster specifically with European wild rats, but unexpectedly formed a sister group to the Europe/Africa/Middle East rat group. Outgroup f3-statistic also corroborated this pattern (supplementary fig. S2, Supplementary Material online). It is possible that the laboratory rat was domesticated from a subpopulation of wild brown rats in Europe, a sister group to the Europe/Africa/Middle East brown rats that we collected.

To assess the genetic mechanisms underlying initial domestication of the laboratory rat, we evaluated differentiation of each SNP between the laboratory rats population (including 26 different stains), and wild rats population using F_{ST}

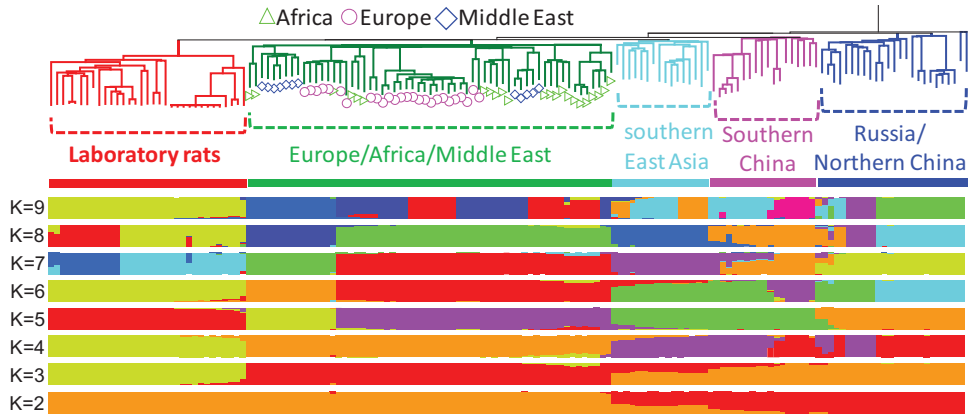


Fig. 1. Phylogenetic tree and population structure of laboratory rats and wild brown rats. (Top) Phylogenetic tree constructed by neighbor joining method based on pairwise distances among individual rats (Bruno et al. 2000). (Bottom) Population structure according to ADMIXTURE. In the ADMIXTURE analysis, East Asian rats are separated with other rats when the number of presumed ancestral population (K) is 2. Laboratory rat population emerges when $K = 3$. Rats from Europe, Africa, and Middle East mix together, and cannot be distinguished geographically.

(Akey et al. 2002). A sliding window analysis (window size: 100 kb, step size: 50 kb) was performed to identify regions/genes harboring high levels of differentiation. As a result, a total of 292 genes displaying high F_{ST} values in the top 1% outlier windows were identified as candidate positively selected genes (supplementary table S2, Supplementary Material online). These genes might contribute to the phenotypic difference between the laboratory and wild rat populations, although demographic history could also generate high level of population differentiation (Sabeti et al. 2006). Gene enrichment analysis found that many of the genes have roles in "*neurological system process*" (54 genes, GO: 0050877, $P = 2.11 \times 10^{-5}$ corrected by Benjamini–Hochberg FDR, supplementary table S3, Supplementary Material online). In particular, seven genes (*AFF2*, *MECP2*, *NAA10*, *NSDHL*, *SLC6A8*, *SLITRK1*, and *ENSRNOG00000049488*) were enriched for the HPO category "*abnormally aggressive, impulsive, or violent behavior*" ($P = 0.05$, corrected by Benjamini–Hochberg FDR, supplementary table S3, Supplementary Material online). Positive selection of gene variants in this category might be related with the behavioral modifications observed in laboratory rats compared with wild rats. We propose that positive selection on genes involved in the nervous system might have played key roles in the successful domestication of the laboratory rat from wild brown rat ancestor, concordant with the findings in other domesticated animals like rabbits, dogs, horses, and goats (Axelsson et al. 2013; Li et al. 2013; Wang et al. 2013; Carneiro et al. 2014; Dong et al. 2015; Librado et al. 2017). Notably, *FOXP2*, a central gene in vocal behavior and learning ability (Enard et al. 2002; Schreiweis et al. 2014), showed signals of high level of population differentiation (fig. 2A). This gene region also harbored lower nucleotide diversity and high XP-EHH value (supplementary fig. S3, Supplementary Material online). It indicated potential positive selection on *FOXP2* in the laboratory rat, although we cannot exclude absolutely the confounding factor of demographic history. The expression level of *FOXP2* was significantly up-regulated in the hypothalamus of laboratory rat examined by real-time quantitative PCR (qPCR) (fig. 2A, $P = 0.026$). Compared with wild rats, learning ability is more advanced in laboratory rats (Boice 1981; Domjan 2014), which probably promote their quick adaptability to human and laboratory environments. We proposed that positive selection driving up-regulation of *FOXP2* might be coupled to the changed learning ability in the laboratory rats compared with wild rats (Boice 1981; Domjan 2014), although functional consequences of up-regulation of *FOXP2* need further experimental validation. Another interesting positively selected gene is *CLOCK*, a central regulator in the control of circadian rhythms (Vitaterna et al. 1994). Additional signals of positive selection with low nucleotide diversity and high XP-EHH value were located upstream of this gene (supplementary fig. S3, Supplementary Material online). *CLOCK* also exhibited higher expression level in the hypothalamus of laboratory rats than wild rats (fig. 2A, $P = 0.004$), and might contribute to the changes in circadian rhythm after domestication.

Genomic loci under positive selection possess other distinctive patterns such as low genetic diversity and long haplotype homozygosity (Sabeti et al. 2006). We therefore assessed signals of artificial selection in the laboratory rats using another statistic, XP-EHH (cross-population extended haplotype homozygosity) (Sabeti et al. 2007). In total, 447 potential candidate positively selected protein coding genes with high XP-EHH values at the top 1% outlier were identified (supplementary table S4, Supplementary Material online). Gene enrichment analysis found that 13 of these genes were enriched for the category "*regulation of developmental growth*" (GO: 0048638, $P = 0.0076$, supplementary table S5, Supplementary Material online). Laboratory rats substantially differ morphologically from wild rats. For example, laboratory rats are larger and have weaker bone structure and smaller internal organs (including brain, heart, liver, and spleen) (Stryjek et al. 2012). Positive selection on developmental genes in the laboratory rat might account for this morphological differentiation.

It is noteworthy that the five topmost windows exhibiting the highest XP-EHH values clustered together and overlapped only with the protein-coding gene *B3GAT1* (fig. 2B). The region also exhibited other signals of positive selection, marked by low nucleotide diversity and high level of population

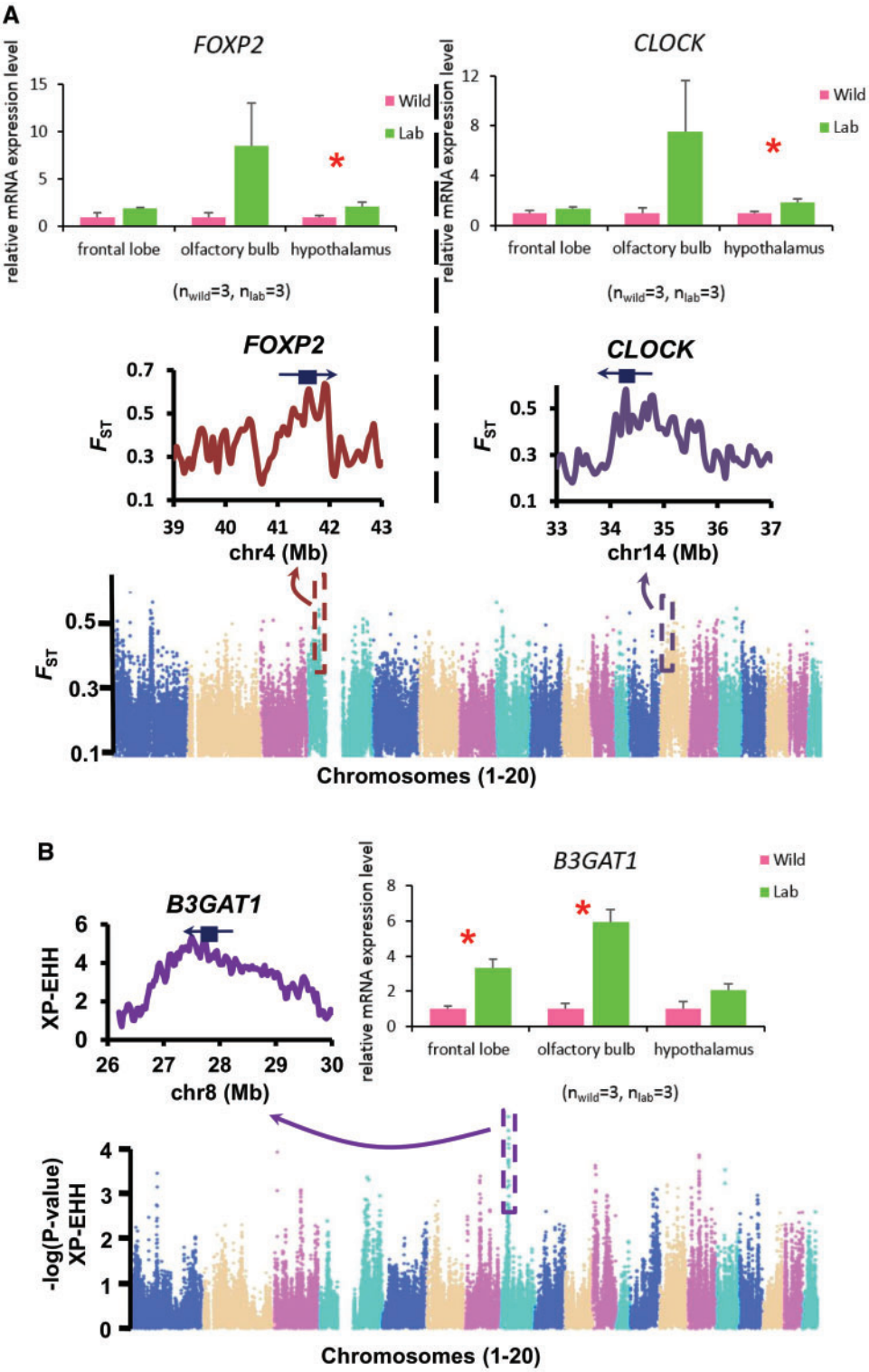


FIG. 2. Artificial selection in the laboratory rat. (*A*) Population differentiation (F_{ST}) between the laboratory rats and wild rats uncover positive selection on *FOXP2* and *CLOCK*. F_{ST} values of the region cross *FOXP2* and CLOCK are presented. (*B*) Genomic landscape of the $-\log_{10}(P \text{ value})$ of XP-EHH values of laboratory rats compared with wild brown rats, and selection at the *B3GAT1* gene are presented. The mRNA expression levels of *FOXP2*, *CLOCK*, and *B3GAT1* genes were detected by qPCR in the frontal lobe, olfactory bulb and hypothalamus from three laboratory rats and three wild brown rats (with two technical replications in each sample). Expression level of each gene was firstly normalized to the expression value of the housekeeping gene β-actin. To clearly show the expression difference between the wild and laboratory rats, we have further normalized the expression value of each gene in each individual as the relative mRNA expression level of this gene in this individual divided by the mean expression level of this gene in three wild rats. The error bars represent standard error of mean (SEM). Asterisk (*) identifies statistically significant differences ($P<0.05$).

differentiation (supplementary fig. S3, Supplementary Material online). The qPCR experiments indicated a significantly higher up-regulation of mRNA expression for *B3GAT1* in the brain of laboratory rats compared with wild brown rats (fig. 2*B*). *B3GAT1* is involved in the biosynthesis of HNK1 (Mitsumoto et al. 2000), which is widely expressed in the brain. *B3GAT1* knockout mice exhibit reduced long-term potentiation at Schaffer collateral-CA1 synapses and defects in spatial learning and memory (Yamamoto et al. 2002). Although the functional consequence of up-regulation of *B3GAT1* remains largely unclear, we propose that the up-regulation of *B3GAT1* likely enhanced spatial learning and memory in laboratory rats, enabling them to adapt to captive situations in the process of domestication. Generally, wild animals are more active and reactive and show extreme levels of stress in captive environments, properties that contribute to their higher mortality in captivity (Price 2002). The increased spatial learning and memory abilities, due to the changes of *B3GAT1*, might have facilitated better stress management in the domesticated rat.

Phenotypic evolution is tightly coupled to changes in gene expression. Therefore, to further explore potential genetic events for domestication of the laboratory rat, we profiled transcriptomes of the cerebral cortex, hypothalamus, olfactory bulb, liver, spleen, and heart from the laboratory rat and wild rat by RNA-sequencing (supplementary fig. S4, Supplementary Material online). Positively selected genes showed significant difference in expression patterns in the nervous system between the wild brown and laboratory rat (supplementary fig. S5, Supplementary Material online, $P < 0.05$, Mann–Whitney *U* test). No significant difference in expression pattern of positively selected genes was found for the other tissues (supplementary fig. S5, Supplementary Material online). Differential expression pattern of positively selected genes in these tissues provides a plausible explanation for the changes in behavior of laboratory rats compared with wild brown rats.

Overall, 777 genes were found to be differentially expressed in the three brain regions between the wild brown and laboratory rat (supplementary fig. S6, Supplementary Material online). In a gene enrichment analysis, 39 genes fell in the category "*behavior*" (GO: 0007610, $P = 1.2 \times 10^{-4}$, supplementary table S6, Supplementary Material online). Differentially expressed genes were also significantly over-represented in categories related with brain development such as "*neurogenesis*," "*gliogenesis*," "*neuron differentiation*," "*neuron development*," and "*neuron projection development*" (supplementary table S6, Supplementary Material online). Changes in the expression levels of these genes might have facilitated the domestication of laboratory rats by influencing the evolution of the nervous system. A similar trend has been reported in other domestic animals, such as dogs, rabbits, horses, and goats (Axelsson et al. 2013; Li et al. 2013; Wang et al. 2013; Carneiro et al. 2014; Dong et al. 2015; Librado et al. 2017). Gene enrichment analysis also found many genes involved in "*immune system development*," which is consistent with the finding by Albert et al (2012), that many immune system genes show differential expression in the brain of domestic animals (like dog, pig, guinea pigs) compared with their wild ancestors (Albert et al. 2012). Domesticated animals are likely particularly exposed to strong selection pressures on their immune systems as a result of living in more crowded captive conditions and/or increased exposure to risks suffered by humans and other domesticated species (Albert et al. 2012).

Another key feature of the transcriptome data was that the differentially expressed genes were over-represented in categories associated with energy metabolism like "*oxidative phosphorylation*," "*mitochondrion*," and "*mitochondrial inner membrane*" (supplementary table S7, Supplementary Material online). Assessing changes in energy metabolism and their consequent nervous system disorders is a key pillar in evolutionary studies of the nervous system (e.g., human brain). Genes related to energy metabolism have been implicated in the evolution and maintenance of human-specific cognitive abilities (Khaitovich et al. 2008). Since the mitochondrion is the energy producing machinery of a cell, we examined the expression of mitochondrial protein-coding genes and found ten genes harbored significant differential expression levels in frontal lobe tissue between the laboratory and wild brown rats (fig. 3*A*). Except for *ND3*, all other identified mtDNA genes were up-regulated in the laboratory rats. However, no significant difference in the copy number of mitochondrial DNA was found between the laboratory and wild brown rats (fig. 3*B*). In contrast, only *ND3* demonstrated significant differential expression in heart and liver tissues. In addition, we validated the up-regulation of nuclear genes (i.e., *ATP5D* and *COX8A*) involved in energy metabolism in the laboratory rats by qPCR (supplementary fig. S7, Supplementary Material online). These findings suggest that substantial evolutionary changes in energy metabolism have occurred in the brain of the laboratory rat during the process of domestication.

In conclusion, compared with wild brown rat, many genes involved in the nervous system, particularly *FOXP2*, *B3GAT1*, and *CLOCK*, have evolved under artificial selection in the laboratory rat. These changes likely enhanced the learning ability and regulation of circadian rhythm to promote successful domestication of laboratory rat, including ability to readily adapt to the human environment. Genes under artificial selection were detected by comparing 26 different laboratory strains with wild rats from different locations. Our gene list did not show overlaps with the positively selected genes in each laboratory breed reported by Atanur et al (2013). In addition, the differentially expressed genes did not show differences among different breeds (Walker et al. 2004) (supplementary fig. S8, Supplementary Material online). We propose that these genes are possibly associated with initial domestication of rats. Brain functions are quite energy intensive relative to the rest of the body (Gómez-Pinilla 2008), enhanced learning ability in the laboratory rats likely placed additional energy demands. As expected, many genes responsible for energy metabolism including mitochondrial genes exhibited a substantially increased expression in the brain of laboratory rats compared with wild rats. Our findings will be helpful for understanding the origin and evolution of the laboratory rats as well as the process of domestication.

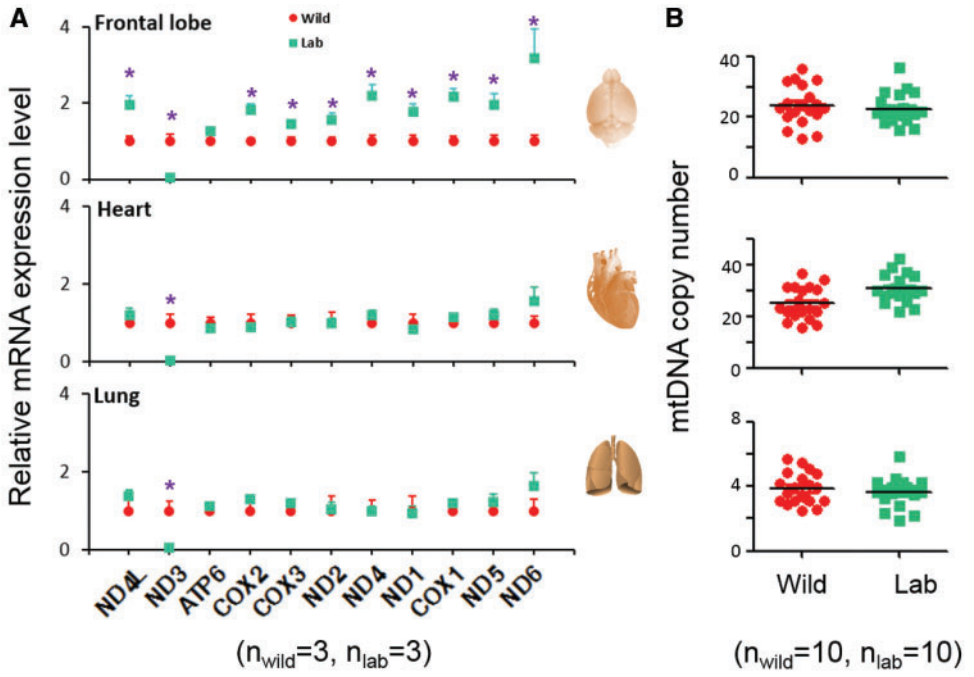


FIG. 3. Relative mRNA expression levels of mitochondrial coding genes (*A*) and mtDNA copy number (*B*) in wild brown and laboratory rats. The mRNA expression levels of mitochondrial genes were detected by qPCR from three laboratory rats and three wild brown rats (with two technical replications in each sample). Expression level of each gene was firstly normalized to the expression values of the housekeeping gene β-actin. To clearly show the expression difference between the wild and laboratory rats, we have further normalized the expression value of each gene in each individual as the relative mRNA expression level of this gene in this individual divided by the mean expression level of this gene in three wild rats. The error bars represent standard error of mean (SEM). Asterisk (*) indicate statistically significant differential expression ($P < 0.05$).

Materials and Methods

Detailed methods are described in the supplementary materials online. Briefly, the genomes of six wild brown rats from Mali ($n = 1$), Morocco ($n = 1$), Russia ($n = 2$), and China ($n = 2$), and nine laboratory rats were sequenced at high coverage (~10–20×) in this study. Whole genome sequences of 24 laboratory strains except Brown Norway breed were obtained from a previously published study (Atanur et al. 2013). The genomes of another 110 *Rattus norvegicus* from Russia, Northern China, Southern China, Southeast Asia, Europe, Africa, and Middle East were obtained from one of our other studies (Zeng L et al., unpublished data). Quality filtered reads were mapped to the reference *Rattus norvegicus* genome (rn5, ENSEMBL version 72) (Gibbs et al. 2004) using the program BWA-MEM (Li 2013). Single nucleotide polymorphisms were detected and filtered using the Genome Analysis Toolkit (GATK) (McKenna et al. 2010). Phylogenetic trees were constructed by the weighted neighbor joining method (Bruno et al. 2000) using pairwise distances among individual rats. To further reveal the relationships among the different *Rattus norvegicus* populations, we performed a principle components analysis (PCA) using GCTA (Yang et al. 2011). Subsequently, population structure was deduced by the program ADMIXTURE (*K* values 2–10), a tool for maximum likelihood estimation of individual ancestries from multi locus SNP genotype data sets (Alexander et al. 2009). The outgroup f3-statistic (Patterson et al. 2012; Raghavan et al. 2014) was used to estimate the genetic proximity of laboratory rat populations to each wild brown rat individual.

To reveal genetic mechanisms underlying initial domestication of the laboratory rats and phenotypic difference between laboratory rats and wild rats, we treated all the laboratory strains together as laboratory population, and all the wild rats as wild population. We used F_{ST} (Akey et al. 2002), XP-EHH (Sabeti et al. 2007) to evaluate differentiation at each SNP between the laboratory population and wild population. The F_{ST} of each SNP was calculated as described elsewhere (Akey et al. 2002). F_{ST} with negative values having no biological explanation were arbitrarily set to 0. XPEHH program (http://hgdp.uchicago.edu/Software/) was used to calculate the XP-EHH value for each SNP. The nucleotide diversity was calculated using an in-house perl script. Sliding window analysis was performed with a window size of 100 kb and a step size of 50 kb. F_{ST} and XP-EHH values for each sliding window were calculated by averaging the values of all SNPs in the window. We employed an outlier approach based on genome-wide empirical data to retrieve the top 1% of windows showing high level F_{ST} values or XP-EHH values.

RNA-sequencing data were analyzed using Tophat–Cufflinks–Cuffdiff pipeline. Filtered reads were aligned to the rat reference genome using TopHat (v2.0.4) (Trapnell et al. 2009) and then assembled using Cufflinks (v2.0.2 with –G parameter) (Trapnell et al. 2012). The differential expression of genes in different tissues was calculated using Cuffdiff (Trapnell et al. 2012). Gene Ontology analysis of the protein-coding genes was conducted using an online annotation tool g: Profiler (Reimand et al. 2011) and DAVID (Dennis et al. 2003), and *P* values were corrected by Benjamini–Hochberg FDR.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (grant no. XDB13020600), and the National Natural Science Foundation of China (91731304).

References

Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. 2002. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res.* 12(12):1805–1814.

Albert FW, Somel M, Carneiro M, Aximu-Petri A, Halbwax M, Thalmann O, Blanco-Aguiar JA, Plyusnina IZ, Trut L, Villafuerte R, et al. 2012. A comparison of brain gene expression levels in domesticated and wild animals. *PLoS Genet.* 8(9):e1002962.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19(9):1655–1664.

Atanur SS, Diaz AG, Maratou K, Sarkis A, Rotival M, Game L, Tschannen MR, Kaisaki PJ, Otto GW, Ma MCJ, et al. 2013. Genome sequencing reveals loci under artificial selection that underlie disease phenotypes in the laboratory rat. *Cell* 154(3):691–703.

Axelsson E, Ratnakumar A, Arendt M-L, Maqbool K, Webster MT, Perloski M, Liberg O, Arnemo JM, Hedhammar A, Lindblad-Toh K, et al. 2013. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495(7441):360–364.

Baker HJ, Lindsey JR, Wesibroth SH. 2013. The laboratory rat: biology and diseases. New York: Academic Press.

Boice R. 1981. Behavioral Comparability of Wild and Domesticated Rats. *Behav Genet.* 11(5):545–553.

Bruno WJ, Socci ND, Halpern AL. 2000. Weighted neighbor joining: a likelihood-based approach to distance-based phylogeny reconstruction. *Mol Biol Evol.* 17(1):189–197.

Carneiro M, Rubin C-J, Di Palma F, Albert FW, Alföldi J, Martinez BA, Pielberg G, Rafati N, Sayyab S, Turner-Maier J, et al. 2014. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science* 345(6200):1074–1079.

Dennis G Jr, Sherman BT, Hosack DA, et al. 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* 4:P3.

Domjan M. 2014. The principles of learning and behavior. Belmont, California, US: Wadsworth Publishing.

Dong Y, Zhang X, Xie M, Arefnezhad B, Wang Z, Wang W, Feng S, Huang G, Guan R, Shen W, et al. 2015. Reference genome of wild goat (capra aegagrus) and sequencing of goat breeds provide insight into genic basis of goat domestication. *BMC Genomics* 16:431.

Enard W, Przeworski M, Fisher SE, Lai CSL, Wiebe V, Kitano T, Monaco AP, Pääbo S. 2002. Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* 418(6900):869–872.

Gibbs RA, Weinstock GM, Metzker ML, et al. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428:493–521.

Gómez-Pinilla F. 2008. Brain foods: the effects of nutrients on brain function. *Nat Rev Neurosci.* 9(7):568–578.

Khaitovich P, Lockstone HE, Wayland MT, Tsang TM, Jayatilaka SD, Guo AJ, Zhou J, Somel M, Harris LW, Holmes E, et al. 2008. Metabolic changes in schizophrenia and human brain evolution. *Genome Biol.* 9(8):R124.

Kuramoto T, Nakanishi S, Ochiai M, Nakagama H, Voigt B, Serikawa T. 2012. Origins of albino and hooded rats: implications from molecular genetic analysis across modern laboratory rat strains. *PLoS One* 7(8):e43059.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *Bioinformatics* arXiv: 1303.3997v2.

Li Y, Vonholdt BM, Reynolds A, Boyko AR, Wayne RK, Wu D-D, Zhang Y-P. 2013. Artificial selection on brain-expressed genes during the domestication of dog. *Mol Biol Evol.* 30(8):1867–1876.

Librado P, Gamba C, Gaunitz C, Der Sarkissian C, Pruvost M, Albrechtsen A, Fages A, Khan N, Schubert M, Jagannathan V, et al. 2017. Ancient genomic changes associated with domestication of the horse. *Science* 356(6336):442–445.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20(9):1297–1303.

Mitsumoto Y, Oka S, Sakuma H, Inazawa J, Kawasaki T. 2000. Cloning and chromosomal mapping of human glucuronyltransferase involved in biosynthesis of the HNK-1 carbohydrate epitope. *Genomics* 65(2):166–173.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient admixture in human history. *Genetics* 192(3):1065–1093.

Price EO. 2002. Animal domestication and behavior. Oxon, UK: Cabi Publishing.

Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, Rasmussen S, Stafford TW, Orlando L, Metspalu E, et al. 2014. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* 505(7481):87–91.

Reimand J, Arak T, Vilo J. 2011. g: Profiler–a web server for functional interpretation of gene lists (2011 update). *Nucleic Acids Res.* 39:W307–W315.

Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES, et al. 2006. Positive natural selection in the human lineage. *Science* 312(5780):1614–1620.

Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne E, McCarroll SA, Gaudet R, et al. 2007. Genome-wide detection and characterization of positive selection in human populations. *Nature* 449(7164):913–918.

Schreiweis C, Bornschein U, Burguière E, Kerimoglu C, Schreiter S, Dannemann M, Goyal S, Rea E, French CA, Puliyadi R, et al. 2014. Humanized Foxp2 accelerates learning by enhancing transitions from declarative to procedural performance. *Proc Natl Acad Sci.* 111(39):14253–14258.

Stryjek R, Modlińska K, Pisula W, McGregor AP. 2012. Species specific behavioural patterns (digging and swimming) and reaction to novel objects in wild type, Wistar, Sprague-Dawley and Brown Norway rats. *PLoS One* 7(7):e40642.

Suckow AM, Weisbroth SH. 2006. The laboratory rat. 2nd ed. San Diego, CA, USA: Academic Press.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25(9):1105–1111.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L, et al. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc.* 7(3):562–578.

Vitaterna MH, King DP, Chang AM, Kornhauser JM, Lowrey PL, McDonald JD, Dove WF, Pinto LH, Turek FW, Takahashi JS, et al. 1994. Mutagenesis and mapping of a mouse gene, Clock, essential for circadian behavior. *Science* 264(5159):719–725.

Walker JR, Su AI, Self DW, Hogenesch JB, Lapp M, Maier R, Hoyer D, Bilbe G. 2004. Applications of a rat multiple tissue gene expression data set. *Genome Res.* 14(4):742–749.

Wang G-D, Zhai W, Yang H-C, Fan R-X, Cao X, Zhong L, Wang L, Liu F, Wu H, Cheng L-G, et al. 2013. The genomics of selection in dogs and the parallel evolution between dogs and humans. *Nat Commun.* 4:1860.

Whishaw IQ, Kolb B. 2004. The behavior of the laboratory rat: a handbook with tests. New York: Oxford University Press.

Yamamoto S, Oka S, Inoue M, Shimuta M, Manabe T, Takahashi H, Miyamoto M, Asano M, Sakagami J, Sudo K, et al. 2002. Mice deficient in nervous system-specific carbohydrate epitope HNK-1 exhibit impaired synaptic plasticity and spatial learning. *J Biol Chem.* 277(30):27227–27231.

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

Supplementary Materials and Methods

The handling of animals used in this study followed the guidelines and regulations of the Kunming Institute of Zoology on animal experimentation and was approved by the Institutional Animal Care and Use Committee of the Kunming Institute of Zoology.

DNA sample collection and genome sequencing

Samples used for sequencing in the present study included six wild brown rats which were collected from Mali (n=1), Morocco (n=1), Russia (n=2) and China (n=2), as well as nine laboratory rats (SD and SHR strains). Genomes of another 110 *Rattus norvegicus* samples of Asia, Europe, Africa and Middle East were obtained from one of our other studies (unpublished data). Whole genome sequence data sets of 24 laboratory rats used in this study were obtained from a previously published work (Atanur et al. 2013).

Tissues for DNA extraction were stored in alcohol at -80°C. Ten micrograms of genomic DNA, prepared by the standard phenol chloroform extraction protocol, was used to construct libraries with 350 base pair insert size. Sequence libraries were constructed according to the Illumina library preparation pipeline, and were sequenced by the Illumina Hiseq 2000 platform.

Reads mapping, single nucleotide polymorphisms (SNPs) calling and filtering

After mapping the reads to the reference *Rattus norvegicus* genome rn5 (ENSEMBL version 72) (Gibbs et al. 2004) with BWA-MEM (Li 2013), we called SNPs using Genome Analysis Toolkit (GATK) (McKenna et al. 2010). After the realignment and

recalibration steps of GATK, we applied hard filters criteria "QD < 2.0, FS > 60.0, MQ < 40.0, HaplotypeScore > 13.0, MappingQualityRankSum < -12.5, ReadPosRankSum < -8.0, -cluster 3 -window 10" to filter the raw SNPs. Eventually, we got 20,568,882 high quality SNPs for the subsequent analyses.

Phylogeny and population structure analysis

The phylogenetic relationships were constructed by the weighted neighbor joining method (Bruno et al. 2000) using pair-wise distances among individual rats. In order to uncover the relationships among the different *Rattus norvegicus* populations, we performed a principle components analysis (PCA) using GCTA (Yang et al. 2011) after pruning the SNPs by plink with --indep-pairwise 50 10 0.1 parameter to get the relatively independent sites (Purcell et al. 2007). Subsequently, population structure was deduced by the program ADMIXTURE with different K values (from 2 to 10) (Alexander et al. 2009). This tool allows for maximum likelihood estimation of individual ancestries from multi locus SNP genotype datasets.

Outgroup f3 analysis

The outgroup f3-statistic (Patterson et al. 2012; Raghavan et al. 2014) was used to estimate the genetic proximity of laboratory rat populations to each wild brown rat individual. We used black rat as the outgroup in the the statistic f3(outgroup; A, B) with laboratory rat population as A and each one of wild rat individual as B.

Analysis of the signatures of positive selection

We used the population pairwise estimate of differentiation (F_{ST}) values (Akey et al. 2002) to identify differentiated loci between the laboratory rat and wild brown rat by

performing sliding window analysis with a window size of 100kb and a step size of 50kb. For the analysis of extended haplotype homozygosity (XP-EHH), haplotypes of each chromosome were deduced using the software SHAPEIT (Delaneau et al. 2013) and XP-EHH values were calculated with the software XPEHH (Sabeti et al. 2007) (http://hgdp.uchicago.edu/Software/) with a window size of 100kb and a step size of 50kb. The nucleotide diversity was calculated using an in-house perl script.

RNA sequencing and analysis

Tissues for RNA-seq, including cortex (n_{wild}=4, n_{lab}=3), olfactory bulb (n_{wild}=2, n_{lab}=3), hypothalamus (n_{wild}=2, n_{lab}=3), heart (n_{wild}=1, n_{lab}=1), liver (n_{wild}=1, n_{lab}=1), spleen (n_{wild}=1, n_{lab}=1). were stored in RNAlater (Ambion, Austin, TX, USA). Total RNA was extracted using the standard Trizol (Qiagen, Chatsworth, CA, USA) protocol and RNeasy mini kit according to manufacturers' instructions (Qiagen, Chatsworth, CA, USA). Before library construction, we assessed the quality of the RNA by spectrophotometry using NanoDrop 2000, gel electrophoresis and Agilent 2100. The library was prepared following the Illumina Genomic RNA sample prep kit protocol and then sequenced on Illumina HiSeq 2000 and HiSeqXTen platform following the manufacturer's instructions.

Adapter sequences were first removed from our own RNA-seq raw data using Cutadapt (v1.2.1) (Martin 2011). Before alignment, reads were trimmed based on their quality scores using the quality trimming program Btrim (Kong 2011). Reads were aligned to the rat reference genome (rn5) (Gibbs et al. 2004) using TopHat (v2.0.4) (Trapnell et al. 2009) and then assembled using Cufflinks (v2.0.2 with –g parameter) (Trapnell et al. 2012). The differential expression of genes in the different

tissues was calculated using Cuffdiff (Trapnell et al. 2012).

To assess if difference in the levels of gene expression in the heart, liver, spleen, cerebral cortex, hypothalamus and olfactory bulb observed between wild brown and laboratory rat might have been driven by positive selection at local regulatory sites during domestication, a series of statistical tests were performed. Expression levels (FPKM) for each gene in each tissue were retrieved and transformed as $\log_2$(FPKM+1). Differences in expression levels for each gene between the wild brown rat and the laboratory rat were calculated using $\log_2((FPKM_{wild} +1) /(FPKM_{lab} +1))$. We then compared the differences in the expression pattern of positively selected genes (PSGs), identified based on their significant F_{ST} and XP-EHH values, to all other genes (all genes in the whole genome excluding PSGs) by Mann-Whitney U test.

Analysis of microarray expression data

To identify the differential expression genes were not altered in the domesticated rat strains, we sought for data based on different rat strains. Fortunately, we found cortex expression data (n_{Wistar}=10, n_{WKY}=6, n_{SD}=6) of three laboratory strains (Wistar, WKY and SD) generated by microarray experiments (Walker JR et al. 2004). We extracted the differentially expressed genes between wild rats and laboratory rats which overlap with the microarray data. The analysis based on this data showed a good consistency among the three domesticated rat strains, indicating a relatively stable expression level among these strains.

Real-time quantitative PCR (qPCR) of selected genes and energy metabolism related genes

We synthesized single-stranded cDNA from 1 μg of total RNA using the PrimeScript RT-PCR Kit (TaKaRa, Japan) in a 25 μL final reaction volume according to the manufacturer's instructions. To validate the differential expression of genes detected in the above RNA-seq analysis, the relative abundance of the mRNAs for eight genes involved in the nervous system and energy metabolism, i.e. *B3GAT1*, *CLOCK*, *FOXP2*, *COX8A* and *ATP5D* genes were measured by the qPCR using RNA samples from the cerebral cortex, olfactory bulb and hypothalamus from three wild brown and three laboratory rats. Expression level of each gene was firstly normalized to the expression values of the housekeeping gene β-actin. To clearly show the expression difference between the wild and laboratory rats, we have further normalized the expression value of each gene in each individual as the relative mRNA expression level of this gene in this individual divided by the mean expression level of this gene in three wild rats.

Detection of mRNA expression of mitochondrial protein coding genes and mitochondrial DNA copy number

The relative abundance of the mRNAs encoding 13 mitochondrial protein-coding genes was estimated in cerebral cortex, heart and lung tissues from three wild brown and three laboratory rats using qPCR as described above. Two mitochondrial genes could not be detected possibly due to technical difficulties. To detect mitochondrial DNA copy number, genomic DNA was extracted by the Genomic DNA Miniprep Kit (Axygen, AP-MN-BL-GDNA-250), and mitochondrial DNA copy number measured using real-time quantitative PCR from samples of cerebral cortex, heart and lung

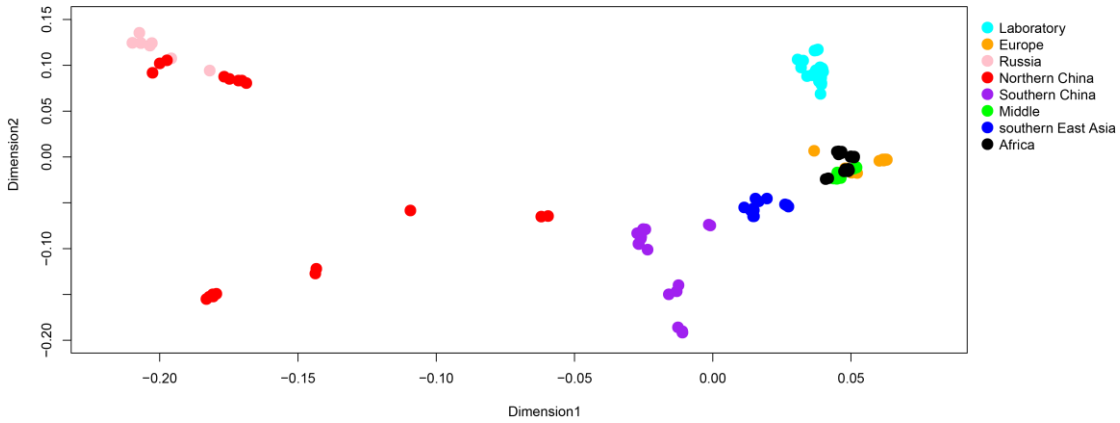
drawn from ten wild brown and ten laboratory rats. Normalization was done to the *Hbb* (β-globin) gene. The qPCR was performed on the iQ2 system platform (BioRad Laboratories) with SYBR Premix Ex Taq II kit (TaKaRa, DRR081A).

Analysis of functional term enrichment

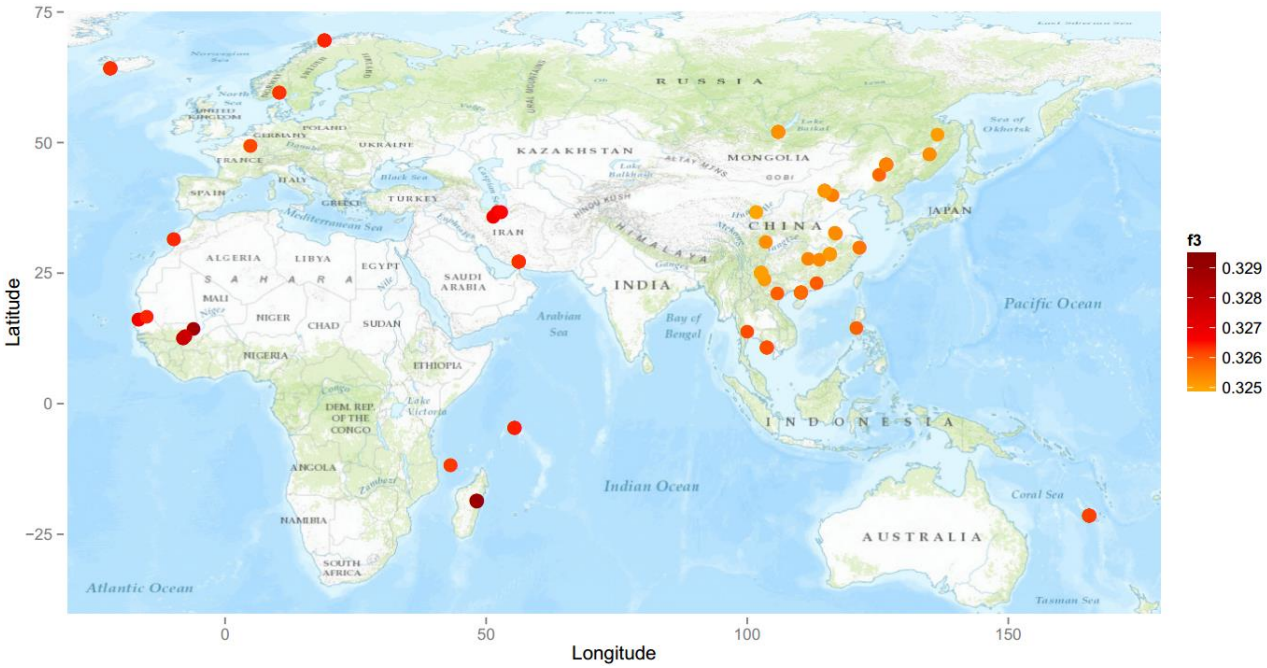
Gene Ontology analysis of protein-coding genes identified as potential candidate positively selected genes was conducted using an online annotation tool g:Profiler and DAVID (Dennis et al. 2003), and P-values were corrected by Benjamini-Hochberg FDR (Reimand et al. 2011).

Accession number

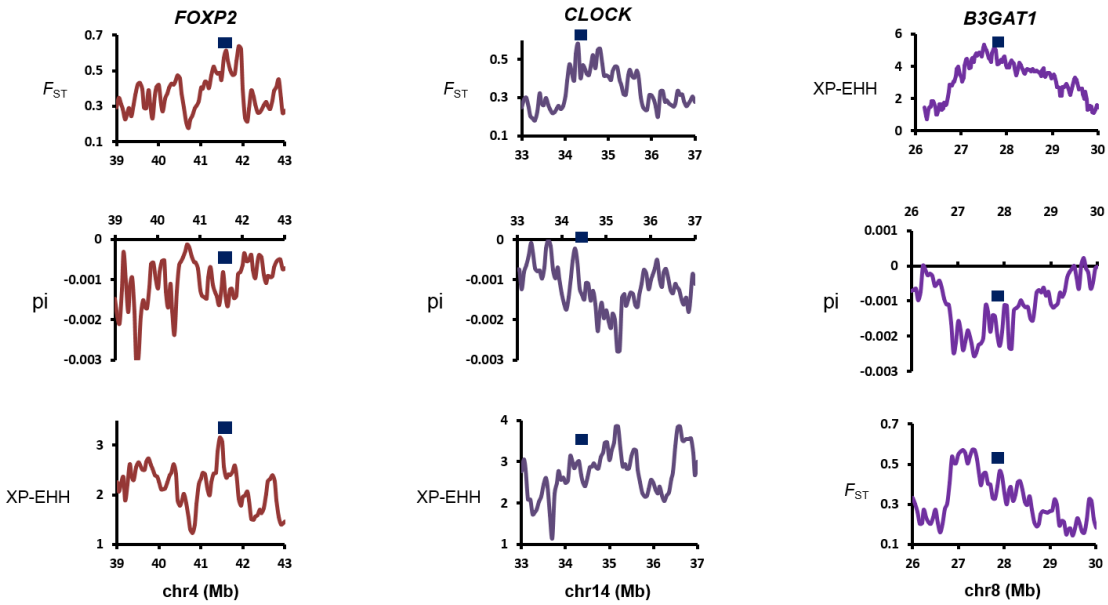
All the sequences reported in this study are deposited in the Genome Sequence Archive database, http://gsa.big.ac.cn/) under Accession ID (PRJCA000251).



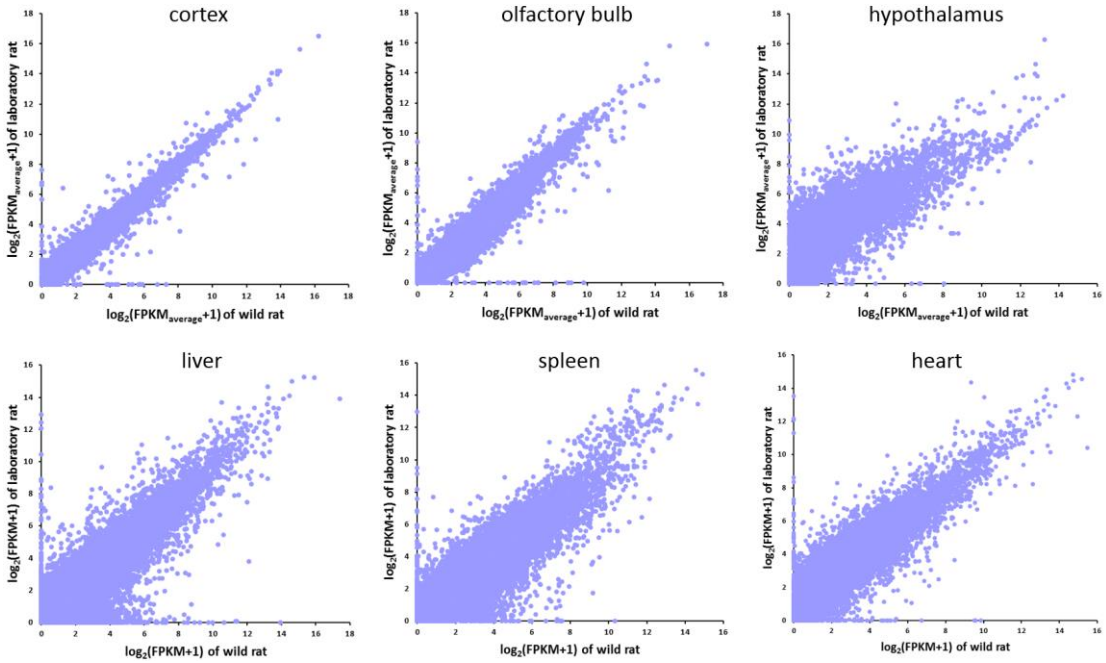
Supplementary Figure S1. PCA analysis of wild brown and laboratory rats.



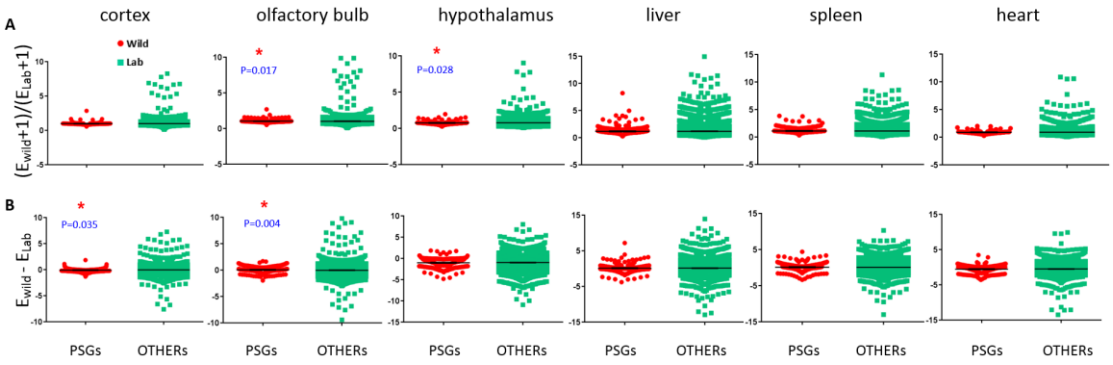
Supplementary Figure S2. Genetic proximity of laboratory rats to wild brown rat individuals evaluated by outgroup f3 statistics.



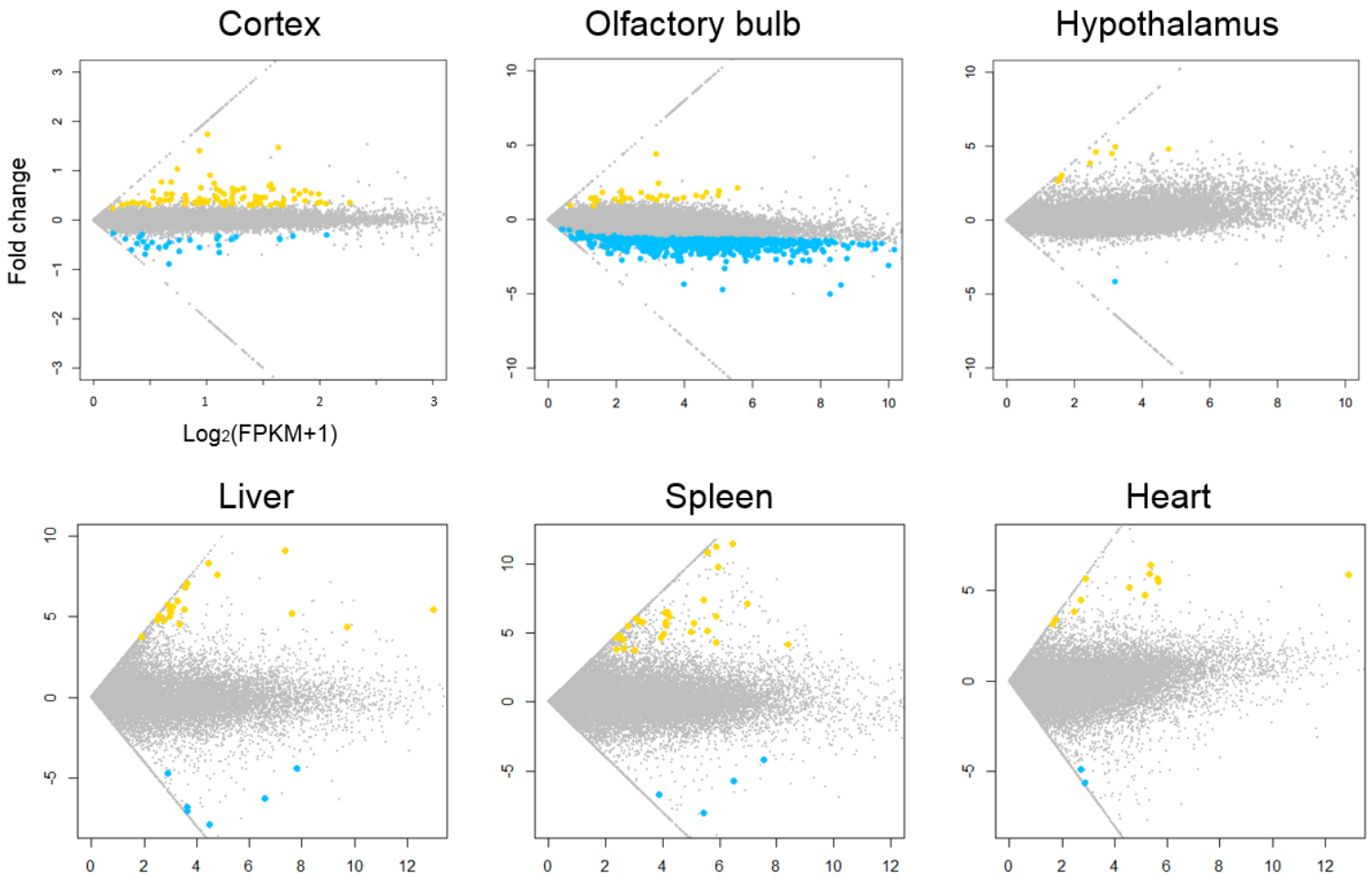
Supplementary Figure S3. Population differentiation value (F_{ST}), nucleotide diversity (pi), and XP-EHH values of *FOXP2*, *CLOCK* and *B3GAT1*.



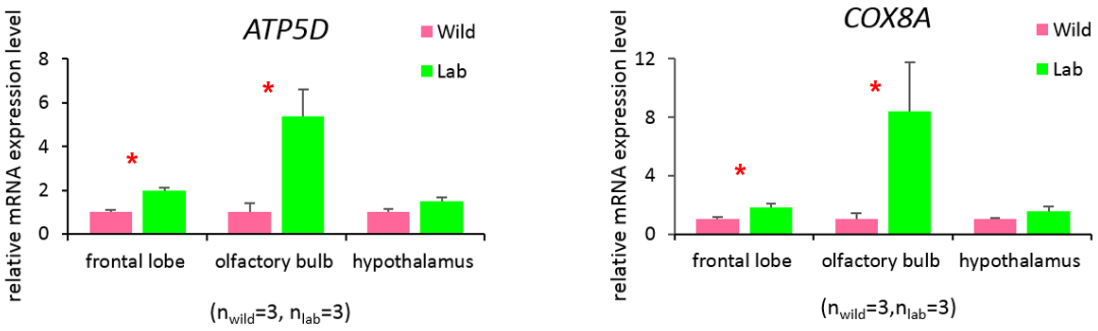
Supplementary Figure S4. The dynamic expression range of the RNA-seq data.



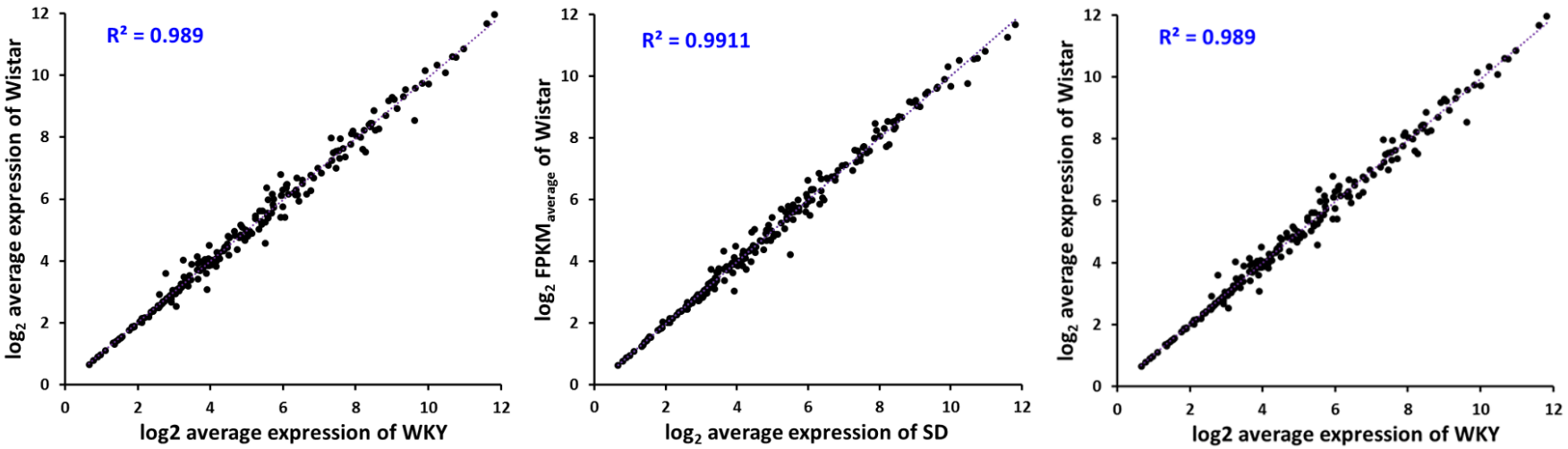
Supplementary Figure S5. Comparisons of the difference in the pattern of mRNA expression between wild brown and laboratory rats. The expression value (E) for each gene was calculated by log2(FPKM+1) transformation, where FPKM (Fragments Per Kilobase of transcript per Million fragments mapped) is the expression value of gene calculated by Cufflinks program based on the RNA-sequencing data. (A): Difference of each gene was calculated by (Ewild +1)/(ELaboratory +1); (B): Expression difference of each gene was calculated by Ewild - ELaboratory. PSGs are positively selected genes, and OTHERs are other genes. The statistical significances were evaluated by Mann–Whitney U test.



Supplementary Figure S6. Comparison of gene expression in six tissues between the wild brown and laboratory rat. A summary dot plot is shown, with blue dots represent down-regulated and yellow dots represent up-regulated genes (P-corrected < 0.05).



Supplementary Figure S7. Comparison of the mRNA expression levels of 2 nuclear genes involved in energy metabolism between wild brown (n=3) and laboratory (n=3) rats evaluated by real-time quantitative PCR. Expression level of each gene was firstly normalized to the expression values of the housekeeping gene β-actin. To clearly show the expression difference between the wild and laboratory rats, we have further normalized the expression value of each gene in each individual as the relative mRNA expression level of this gene in this individual divided by the mean expression level of this gene in three wild rats. The error bars represent standard error of mean (SEM). Asterisk (*) indicate statistically significant differential expression ($P<0.05$).



Supplementary Figure S8. The expression correlation among three domesticated rat strains. We retrieved the expression values of our differential expressed genes between wild rats and laboratory rats in three laboratory rat stains in the microarray data generated by Walker JR et al. (2004). These differential expressed genes between wild rats and laboratory rats exhibited highly consistency among the three different domesticated rat strains, and didn't show expression difference among different stains.

Supplementary Table S1. Sample information and the genomic data information.

Individual	Sex	Longitude	Latitude	Runs	Data size(Gb)	Depth	Coverage
China1		125.3248	43.88686	2	71	21.80	0.87
China2		102.7146	25.04915	2	84	19.38	0.73
Russia1		106	52	2	56	21.63	0.88
Russia2		106	52	1	33	7.51	0.83
Mali		-5.989	14.241	2	83	20.86	0.86
Morocco		-9.75833	31.52222	2	81	19.85	0.87
SD	Female			2	84	19.46	0.85
SHR1	Female			2	32	10.37	0.86
SHR2	Female			2	31	10.06	0.86
SHR3	Male			2	30	9.52	0.87
SHR4	Male			2	30	9.62	0.87
SHR5	Male			2	29	9.41	0.87
SHR6	Male			2	30	9.53	0.87
SHR7	Male			2	30	9.57	0.87
SHR8	Male			1	29	8	0.86

Notes:
Another 110 whole genome sequences of wild brown rats were obtained from one of our studies (Zeng et al. unpublished data). Additional 24 whole genome sequences of laboratory rat strains were retrieved from a previously reported study (Atanur et al. 2013).

Supplementary Table S2. IDs of genes differentiated between wild brown rats and laboratory rats detected by F_{ST}

ENSRNOG00000050209	ENSRNOG00000048329	ENSRNOG00000013085	ENSRNOG00000009640	ENSRNOG00000013563
ENSRNOG00000048967	ENSRNOG00000001379	ENSRNOG00000002164	ENSRNOG00000032656	ENSRNOG00000028345
ENSRNOG00000028896	ENSRNOG00000028930	ENSRNOG00000010744	ENSRNOG00000005028	ENSRNOG00000046439
ENSRNOG00000018886	ENSRNOG00000016196	ENSRNOG00000017099	ENSRNOG00000028858	ENSRNOG00000046414
ENSRNOG00000018857	ENSRNOG00000015894	ENSRNOG00000002711	ENSRNOG00000037267	ENSRNOG00000028783
ENSRNOG00000014726	ENSRNOG00000024945	ENSRNOG00000027963	ENSRNOG00000015594	ENSRNOG00000028077
ENSRNOG00000026033	ENSRNOG00000025298	ENSRNOG00000017362	ENSRNOG00000019489	ENSRNOG00000017224
ENSRNOG00000023333	ENSRNOG00000010434	ENSRNOG00000007348	ENSRNOG00000014028	ENSRNOG00000037227
ENSRNOG00000019549	ENSRNOG00000011168	ENSRNOG00000032512	ENSRNOG00000004112	ENSRNOG00000010084
ENSRNOG00000050117	ENSRNOG00000019175	ENSRNOG00000050124	ENSRNOG00000046833	ENSRNOG00000019544
ENSRNOG00000011245	ENSRNOG00000028359	ENSRNOG00000033009	ENSRNOG00000037352	ENSRNOG00000001966
ENSRNOG00000017065	ENSRNOG00000013223	ENSRNOG00000046510	ENSRNOG00000028963	ENSRNOG00000011024
ENSRNOG00000037273	ENSRNOG00000012683	ENSRNOG00000032371	ENSRNOG00000026361	ENSRNOG00000013506
ENSRNOG00000013017	ENSRNOG00000007167	ENSRNOG00000049512	ENSRNOG00000005022	ENSRNOG00000017140
ENSRNOG00000005036	ENSRNOG00000005429	ENSRNOG00000050901	ENSRNOG00000009941	ENSRNOG00000023197
ENSRNOG00000017798	ENSRNOG00000016845	ENSRNOG00000012522	ENSRNOG00000002730	ENSRNOG00000008734
ENSRNOG00000037270	ENSRNOG00000016533	ENSRNOG00000012515	ENSRNOG00000050121	ENSRNOG00000017294
ENSRNOG00000019053	ENSRNOG00000016884	ENSRNOG00000048628	ENSRNOG00000019501	ENSRNOG00000024653
ENSRNOG00000017440	ENSRNOG00000050990	ENSRNOG00000033813	ENSRNOG00000009662	ENSRNOG00000046438
ENSRNOG00000002041	ENSRNOG00000050088	ENSRNOG00000050959	ENSRNOG00000010732	ENSRNOG00000029533
ENSRNOG00000013859	ENSRNOG00000005268	ENSRNOG00000027423	ENSRNOG00000004936	ENSRNOG00000028519
ENSRNOG00000013540	ENSRNOG00000011411	ENSRNOG00000017341	ENSRNOG00000021847	ENSRNOG00000048417
ENSRNOG00000036870	ENSRNOG00000011335	ENSRNOG00000017344	ENSRNOG00000002271	ENSRNOG00000042169
ENSRNOG00000013778	ENSRNOG00000049237	ENSRNOG00000049981	ENSRNOG00000004168	ENSRNOG00000027086
ENSRNOG00000013908	ENSRNOG00000037264	ENSRNOG00000017351	ENSRNOG00000032701	ENSRNOG00000032607
ENSRNOG00000043258	ENSRNOG00000019096	ENSRNOG00000049717	ENSRNOG00000010296	ENSRNOG00000027256
ENSRNOG00000015468	ENSRNOG00000014525	ENSRNOG00000032077	ENSRNOG00000030538	ENSRNOG00000039282
ENSRNOG00000034174	ENSRNOG00000014868	ENSRNOG00000017357	ENSRNOG00000009740	ENSRNOG00000002726
ENSRNOG00000042084	ENSRNOG00000007066	ENSRNOG00000017366	ENSRNOG00000009209	ENSRNOG00000021843
ENSRNOG00000005094	ENSRNOG00000001465	ENSRNOG00000030947	ENSRNOG00000027773	ENSRNOG00000013082
ENSRNOG00000013736	ENSRNOG00000009278	ENSRNOG00000046579	ENSRNOG00000046874	ENSRNOG00000048906
ENSRNOG00000024280	ENSRNOG00000005701	ENSRNOG00000045648	ENSRNOG00000017462	ENSRNOG00000049488
ENSRNOG00000002141	ENSRNOG00000037263	ENSRNOG00000047136	ENSRNOG00000015095	ENSRNOG00000029608
ENSRNOG00000002046	ENSRNOG00000010966	ENSRNOG00000049621	ENSRNOG00000007307	ENSRNOG00000033861
ENSRNOG00000001964	ENSRNOG00000046829	ENSRNOG00000019599	ENSRNOG00000017114	ENSRNOG00000046392
ENSRNOG00000020872	ENSRNOG00000005371	ENSRNOG00000019600	ENSRNOG00000050637	ENSRNOG00000046221
ENSRNOG00000007909	ENSRNOG00000031551	ENSRNOG00000049748	ENSRNOG00000030752	ENSRNOG00000047422
ENSRNOG00000017991	ENSRNOG00000009477	ENSRNOG00000046869	ENSRNOG00000021366	ENSRNOG00000050529
ENSRNOG00000015696	ENSRNOG00000009509	ENSRNOG00000050054	ENSRNOG00000046088	ENSRNOG00000027336
ENSRNOG00000009258	ENSRNOG00000043234	ENSRNOG00000046609	ENSRNOG00000021397	ENSRNOG00000049587
ENSRNOG00000017059	ENSRNOG00000037064	ENSRNOG00000050559	ENSRNOG00000031107	ENSRNOG00000047093
ENSRNOG00000048220	ENSRNOG00000009638	ENSRNOG00000029233	ENSRNOG00000004280	ENSRNOG00000045873
ENSRNOG00000037070	ENSRNOG00000011134	ENSRNOG00000049938	ENSRNOG00000014946	ENSRNOG00000045772
ENSRNOG00000008259	ENSRNOG00000006865	ENSRNOG00000037271	ENSRNOG00000002199	ENSRNOG00000042228
ENSRNOG00000008107	ENSRNOG00000046046	ENSRNOG00000009713	ENSRNOG00000046442	ENSRNOG00000046201
ENSRNOG00000037069	ENSRNOG00000033699	ENSRNOG00000008970	ENSRNOG00000022255	ENSRNOG00000049318
ENSRNOG00000002175	ENSRNOG00000031258	ENSRNOG00000012367	ENSRNOG00000049517	ENSRNOG00000033199
ENSRNOG00000030119	ENSRNOG00000047141	ENSRNOG00000004515	ENSRNOG00000019262	ENSRNOG00000048354
ENSRNOG00000010334	ENSRNOG00000017055	ENSRNOG00000047279	ENSRNOG00000027158	ENSRNOG00000048296
ENSRNOG00000031539	ENSRNOG00000010176	ENSRNOG00000002217	ENSRNOG00000049088	ENSRNOG00000049220
ENSRNOG00000023650	ENSRNOG00000007388	ENSRNOG00000016011	ENSRNOG00000037225	ENSRNOG00000031435
ENSRNOG00000014456	ENSRNOG00000020625	ENSRNOG00000017923	ENSRNOG00000019271	ENSRNOG00000014541
ENSRNOG00000005931	ENSRNOG00000031036	ENSRNOG00000025999	ENSRNOG00000011058	
ENSRNOG00000008989	ENSRNOG00000037262	ENSRNOG00000026014	ENSRNOG00000008372	
ENSRNOG00000017326	ENSRNOG00000039990	ENSRNOG00000013237	ENSRNOG00000018808	
ENSRNOG00000020647	ENSRNOG00000014980	ENSRNOG00000008839	ENSRNOG00000025894	
ENSRNOG00000018566	ENSRNOG00000010855	ENSRNOG00000037230	ENSRNOG00000017726	

ENSRNOG00000017946	ENSRNOG00000037188	ENSRNOG00000016368	ENSRNOG00000048963	
ENSRNOG00000017888	ENSRNOG00000050843	ENSRNOG00000043179	ENSRNOG00000046319	
ENSRNOG00000015076	ENSRNOG00000029373	ENSRNOG00000045663	ENSRNOG00000043159	

Supplementary Table S3. Functional categories of genes with high level differentiation between wild brown rats and laboratory rats detected by F_{ST}

P-value	N	GO ID	Term ID	Description
4.02E-04	59	GO:0003008	BP	system process
2.11E-05	**54**	**GO:0050877**	**BP**	**neurological system process**
9.52E-07	**50**	**GO:0007600**	**BP**	**sensory perception**
1.97E-08	**47**	**GO:0007606**	**BP**	**sensory perception of chemical stimulus**
8.85E-06	**40**	**GO:0007608**	**BP**	**sensory perception of smell**
4.50E-03	**7**	**GO:0050909**	**BP**	**sensory perception of taste**
6.86E-06	**43**	**GO:0051606**	**BP**	**detection of stimulus**
2.24E-06	**42**	**GO:0009593**	**BP**	**detection of chemical stimulus**
9.19E-06	**41**	**GO:0050906**	**BP**	**detection of stimulus involved in sensory perception**
3.42E-06	**41**	**GO:0050907**	**BP**	**detection of chemical stimulus involved in sensory perception**
3.61E-05	**38**	**GO:0050911**	**BP**	**detection of chemical stimulus involved in sensory perception of smell**
1.15E-03	107	GO:0023052	BP	signaling
1.15E-03	107	GO:0044700	BP	single organism signaling
5.00E-02	109	GO:0051716	BP	cellular response to stimulus
1.41E-03	108	GO:0007154	BP	cell communication
1.44E-04	104	GO:0007165	BP	signal transduction
1.64E-05	81	GO:0007166	BP	cell surface receptor signaling pathway
5.74E-09	64	GO:0007186	BP	G-protein coupled receptor signaling pathway
5.57E-03	130	GO:0016020	CC	membrane
1.58E-04	113	GO:0044425	CC	membrane part
2.32E-04	98	GO:0031224	CC	intrinsic component of membrane
9.36E-05	98	GO:0016021	CC	integral component of membrane
1.54E-06	99	GO:0071944	CC	cell periphery
8.51E-07	98	GO:0005886	CC	plasma membrane
1.53E-08	69	GO:0060089	MF	molecular transducer activity
1.53E-08	69	GO:0004871	MF	signal transducer activity
2.97E-09	69	GO:0004872	MF	receptor activity
7.86E-09	65	GO:0038023	MF	signaling receptor activity
1.00E-08	63	GO:0004888	MF	transmembrane signaling receptor activity
2.36E-08	57	GO:0004930	MF	G-protein coupled receptor activity
3.61E-05	38	GO:0004984	MF	olfactory receptor activity
5.31E-06	17	GO:0030246	MF	carbohydrate binding
5.00E-02	**7**	**HP:0006919**	**hp**	**Abnormal aggressive, impulsive or violent behavior**
1.84E-03	3	HP:0001955	hp	Unexplained fevers
1.57E-04	14	HP:0010985	hp	Gonosomal inheritance
1.48E-04	14	HP:0001417	hp	X-linked inheritance
1.78E-05	12	HP:0001419	hp	X-linked recessive inheritance
1.30E-02	7	KEGG:04742	ke	Taste transduction
5.00E-02	30	KEGG:04740	ke	Olfactory transduction

Supplementary Table S4. Differentiated genes ID between wild brown rats and laboratory rats detected by XPEHH

ENSRNOG00000024560	ENSRNOG00000014128	ENSRNOG00000002339	ENSRNOG00000032656	ENSRNOG00000016665
ENSRNOG00000026675	ENSRNOG00000012954	ENSRNOG00000002339	ENSRNOG00000003253	ENSRNOG00000016665
ENSRNOG00000026675	ENSRNOG00000012954	ENSRNOG00000016316	ENSRNOG00000019887	ENSRNOG00000046227
ENSRNOG00000033195	ENSRNOG00000012954	ENSRNOG00000016316	ENSRNOG00000003923	ENSRNOG00000009075
ENSRNOG00000028781	ENSRNOG00000012954	ENSRNOG00000005059	ENSRNOG00000003923	ENSRNOG00000020714
ENSRNOG00000028781	ENSRNOG00000012954	ENSRNOG00000005059	ENSRNOG00000002353	ENSRNOG00000048433
ENSRNOG00000021130	ENSRNOG00000012954	ENSRNOG00000000156	ENSRNOG00000042647	ENSRNOG00000004411
ENSRNOG00000021130	ENSRNOG00000011168	ENSRNOG00000014508	ENSRNOG00000018767	ENSRNOG00000024526
ENSRNOG00000046958	ENSRNOG00000011168	ENSRNOG00000014508	ENSRNOG00000018767	ENSRNOG00000002926
ENSRNOG00000015762	ENSRNOG00000011168	ENSRNOG00000014508	ENSRNOG00000018767	ENSRNOG00000002926
ENSRNOG00000015762	ENSRNOG00000011168	ENSRNOG00000014508	ENSRNOG00000025767	ENSRNOG00000002926
ENSRNOG00000048164	ENSRNOG00000000640	ENSRNOG00000019763	ENSRNOG00000000645	ENSRNOG00000008794
ENSRNOG00000048164	ENSRNOG00000000640	ENSRNOG00000019763	ENSRNOG00000019808	ENSRNOG00000032690
ENSRNOG00000048164	ENSRNOG00000013752	ENSRNOG00000012593	ENSRNOG00000027756	ENSRNOG00000032690
ENSRNOG00000000639	ENSRNOG00000013752	ENSRNOG00000004863	ENSRNOG00000049097	ENSRNOG00000032690
ENSRNOG00000019217	ENSRNOG00000021104	ENSRNOG00000004863	ENSRNOG00000049097	ENSRNOG00000020442
ENSRNOG00000019783	ENSRNOG00000012619	ENSRNOG00000004863	ENSRNOG00000026979	ENSRNOG00000020442
ENSRNOG00000019783	ENSRNOG00000010719	ENSRNOG00000004863	ENSRNOG00000037219	ENSRNOG00000002941
ENSRNOG00000026319	ENSRNOG00000016627	ENSRNOG00000004863	ENSRNOG00000026361	ENSRNOG00000002941
ENSRNOG00000026319	ENSRNOG00000030367	ENSRNOG00000004863	ENSRNOG00000032148	ENSRNOG00000002941
ENSRNOG00000001989	ENSRNOG00000030367	ENSRNOG00000010855	ENSRNOG00000032148	ENSRNOG00000021149
ENSRNOG00000001989	ENSRNOG00000030367	ENSRNOG00000010855	ENSRNOG00000013027	ENSRNOG00000021149
ENSRNOG00000008683	ENSRNOG00000013468	ENSRNOG00000007899	ENSRNOG00000013027	ENSRNOG00000011989
ENSRNOG00000009589	ENSRNOG00000001962	ENSRNOG00000009615	ENSRNOG00000002773	ENSRNOG00000011989
ENSRNOG00000017065	ENSRNOG00000020368	ENSRNOG00000001963	ENSRNOG00000002730	ENSRNOG00000030026
ENSRNOG00000017065	ENSRNOG00000020368	ENSRNOG00000011306	ENSRNOG00000013389	ENSRNOG00000026073
ENSRNOG00000004891	ENSRNOG00000013875	ENSRNOG00000011306	ENSRNOG00000029456	ENSRNOG00000048947
ENSRNOG00000004891	ENSRNOG00000013875	ENSRNOG00000009849	ENSRNOG00000046345	ENSRNOG00000047235
ENSRNOG00000004891	ENSRNOG00000013875	ENSRNOG00000009849	ENSRNOG00000013195	ENSRNOG00000047235
ENSRNOG00000012196	ENSRNOG00000013875	ENSRNOG00000008932	ENSRNOG00000013195	ENSRNOG00000023566
ENSRNOG00000038420	ENSRNOG00000011471	ENSRNOG00000008932	ENSRNOG00000014509	ENSRNOG00000023566
ENSRNOG00000007142	ENSRNOG00000011471	ENSRNOG00000008932	ENSRNOG00000010732	ENSRNOG00000014805
ENSRNOG00000007142	ENSRNOG00000011471	ENSRNOG00000026930	ENSRNOG00000010732	ENSRNOG00000012030
ENSRNOG00000007142	ENSRNOG00000003095	ENSRNOG00000017765	ENSRNOG00000046051	ENSRNOG00000012937
ENSRNOG00000015189	ENSRNOG00000017800	ENSRNOG00000021118	ENSRNOG00000026110	ENSRNOG00000010137
ENSRNOG00000015189	ENSRNOG00000004898	ENSRNOG00000010744	ENSRNOG00000026110	ENSRNOG00000010137
ENSRNOG00000015189	ENSRNOG00000012347	ENSRNOG00000010744	ENSRNOG00000006396	ENSRNOG00000010137
ENSRNOG00000015189	ENSRNOG00000012347	ENSRNOG00000010744	ENSRNOG00000013743	ENSRNOG00000010137
ENSRNOG00000001971	ENSRNOG00000008855	ENSRNOG00000002711	ENSRNOG00000013743	ENSRNOG00000011024
ENSRNOG00000032461	ENSRNOG00000022116	ENSRNOG00000002711	ENSRNOG00000013743	ENSRNOG00000011024
ENSRNOG00000032461	ENSRNOG00000022116	ENSRNOG00000020519	ENSRNOG00000000314	ENSRNOG00000011024
ENSRNOG00000005094	ENSRNOG00000007561	ENSRNOG00000020519	ENSRNOG00000031599	ENSRNOG00000011024
ENSRNOG00000005094	ENSRNOG00000007561	ENSRNOG00000020519	ENSRNOG00000038432	ENSRNOG00000004106
ENSRNOG00000028576	ENSRNOG00000007561	ENSRNOG00000046552	ENSRNOG00000038432	ENSRNOG00000004106
ENSRNOG00000031955	ENSRNOG00000031469	ENSRNOG00000049774	ENSRNOG00000019305	ENSRNOG00000016719
ENSRNOG00000031955	ENSRNOG00000031469	ENSRNOG00000050929	ENSRNOG00000019305	ENSRNOG00000025937
ENSRNOG00000016322	ENSRNOG00000031469	ENSRNOG00000030569	ENSRNOG00000002013	ENSRNOG00000027234
ENSRNOG00000034015	ENSRNOG00000017605	ENSRNOG00000029097	ENSRNOG00000015275	ENSRNOG00000014661
ENSRNOG00000003468	ENSRNOG00000002309	ENSRNOG00000047228	ENSRNOG00000021847	ENSRNOG00000014661
ENSRNOG00000003468	ENSRNOG00000020719	ENSRNOG00000032325	ENSRNOG00000021847	ENSRNOG00000033624
ENSRNOG00000003468	ENSRNOG00000019096	ENSRNOG00000050176	ENSRNOG00000027229	ENSRNOG00000034037

ENSRNOG00000001982	ENSRNOG00000019096	ENSRNOG00000047710	ENSRNOG00000025220	ENSRNOG00000034037
ENSRNOG00000001982	ENSRNOG00000032834	ENSRNOG00000010983	ENSRNOG00000028924	ENSRNOG00000020356
ENSRNOG00000001982	ENSRNOG00000032834	ENSRNOG00000010983	ENSRNOG00000028924	ENSRNOG00000020356
ENSRNOG00000001982	ENSRNOG00000009278	ENSRNOG00000008970	ENSRNOG00000004821	ENSRNOG00000014065
ENSRNOG00000021109	ENSRNOG00000009278	ENSRNOG00000008970	ENSRNOG00000004821	ENSRNOG00000014407
ENSRNOG00000021109	ENSRNOG00000047647	ENSRNOG00000018911	ENSRNOG00000015095	ENSRNOG00000014407
ENSRNOG00000013263	ENSRNOG00000047647	ENSRNOG00000018911	ENSRNOG00000015095	ENSRNOG00000050529
ENSRNOG00000013263	ENSRNOG00000017660	ENSRNOG00000018911	ENSRNOG00000014720	ENSRNOG00000048026
ENSRNOG00000028005	ENSRNOG00000017660	ENSRNOG00000034228	ENSRNOG00000011078	ENSRNOG00000050529
ENSRNOG00000007909	ENSRNOG00000009149	ENSRNOG00000034228	ENSRNOG00000005247	ENSRNOG00000046083
ENSRNOG00000014371	ENSRNOG00000000648	ENSRNOG00000034228	ENSRNOG00000005247	ENSRNOG00000039997
ENSRNOG00000040257	ENSRNOG00000014498	ENSRNOG00000034228	ENSRNOG00000047187	ENSRNOG00000049148
ENSRNOG00000040257	ENSRNOG00000014498	ENSRNOG00000034228	ENSRNOG00000047187	ENSRNOG00000048044
ENSRNOG00000012238	ENSRNOG00000014498	ENSRNOG00000034228	ENSRNOG00000003562	ENSRNOG00000042129
ENSRNOG00000027195	ENSRNOG00000047736	ENSRNOG00000020474	ENSRNOG00000003562	ENSRNOG00000046083
ENSRNOG00000017582	ENSRNOG00000011380	ENSRNOG00000020474	ENSRNOG00000012604	ENSRNOG00000043129
ENSRNOG00000004909	ENSRNOG00000011380	ENSRNOG00000016255	ENSRNOG00000021093	ENSRNOG00000023905
ENSRNOG00000004281	ENSRNOG00000011380	ENSRNOG00000016255	ENSRNOG00000005853	ENSRNOG00000050014
ENSRNOG00000016281	ENSRNOG00000022297	ENSRNOG00000016255	ENSRNOG00000032448	ENSRNOG00000050908
ENSRNOG00000023972	ENSRNOG00000022297	ENSRNOG00000008902	ENSRNOG00000005664	ENSRNOG00000006713
ENSRNOG00000020525	ENSRNOG00000021831	ENSRNOG00000008902	ENSRNOG00000031185	ENSRNOG00000050720
ENSRNOG00000020525	ENSRNOG00000031515	ENSRNOG00000009096	ENSRNOG00000005692	ENSRNOG00000029847
ENSRNOG00000019648	ENSRNOG00000017237	ENSRNOG00000000715	ENSRNOG00000003889	ENSRNOG00000042281
ENSRNOG00000019648	ENSRNOG00000017237	ENSRNOG00000013237	ENSRNOG00000002695	ENSRNOG00000047151
ENSRNOG00000008989	ENSRNOG00000017237	ENSRNOG00000020731	ENSRNOG00000003993	ENSRNOG00000045883
ENSRNOG00000008989	ENSRNOG00000006865	ENSRNOG00000008869	ENSRNOG00000006649	ENSRNOG00000045883
ENSRNOG00000008989	ENSRNOG00000010544	ENSRNOG00000015603	ENSRNOG00000006649	ENSRNOG00000047151
ENSRNOG00000005480	ENSRNOG00000010544	ENSRNOG00000013360	ENSRNOG00000006649	ENSRNOG00000048546
ENSRNOG00000014614	ENSRNOG00000010544	ENSRNOG00000013360	ENSRNOG00000047053	ENSRNOG00000015605
ENSRNOG00000022178	ENSRNOG00000039902	ENSRNOG00000019871	ENSRNOG00000047053	ENSRNOG00000034129
ENSRNOG00000022178	ENSRNOG00000004221	ENSRNOG00000003098	ENSRNOG00000022067	ENSRNOG00000039627
ENSRNOG00000003635	ENSRNOG00000004221	ENSRNOG00000003098	ENSRNOG00000022067	ENSRNOG00000047345
ENSRNOG00000003635	ENSRNOG00000004221	ENSRNOG00000003098	ENSRNOG00000021096	ENSRNOG00000034129
ENSRNOG00000003635	ENSRNOG00000004221	ENSRNOG00000014097	ENSRNOG00000003985	ENSRNOG00000030250
ENSRNOG00000003635	ENSRNOG00000011236	ENSRNOG00000004840	ENSRNOG00000003985	ENSRNOG00000021843
ENSRNOG00000034088	ENSRNOG00000012164	ENSRNOG00000004840	ENSRNOG00000001976	ENSRNOG00000039627
ENSRNOG00000034088	ENSRNOG00000030389	ENSRNOG00000004483	ENSRNOG00000016575	
ENSRNOG00000024774	ENSRNOG00000049370	ENSRNOG00000004483	ENSRNOG00000016782	
ENSRNOG00000024774	ENSRNOG00000048253	ENSRNOG00000032656	ENSRNOG00000024707	

Supplementary Table S5. Functional categories of genes with high level differentiation between wild brown rats and laboratory rats detected by XPEHH

P-value	N	GO ID	term ID	Description
0.00757	13	GO:0048638	BP	regulation of developmental growth
0.0156	7	GO:0046620	BP	regulation of organ growth
0.0481	3	GO:0015925	MF	galactosidase activity
0.014	3	GO:0004565	MF	beta-galactosidase activity
0.05	11	HP:0030177	hp	Abnormality of peripheral nervous system electrophysiology
0.0469	3	KEGG:03018	keg	RNA degradation
0.0492	7	KEGG:04151	keg	PI3K-Akt signaling pathway
0.0469	3	KEGG:04512	keg	ECM-receptor interaction
0.0257	5	KEGG:04141	keg	Protein processing in endoplasmic reticulum
0.0127	4	KEGG:04742	keg	Taste transduction
0.0104	5	KEGG:05162	keg	Measles
0.0235	2	KEGG:03060	keg	Protein export

Supplementary Table S6. Functional categories of differential expression genes enriched by g:profiler
(please see the attached file)

Supplementary Table S7. Functional categories of differential expression genes enriched by DAVID
(please see the attached file)

Reference

Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. 2002. Interrogating a high-density SNP map for signatures of natural selection. *Genome research* **12**(12): 1805-1814.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome research* **19**(9): 1655-1664.

Atanur SS, Diaz AG, Maratou K, Sarkis A, Rotival M, Game L, Tschannen MR, Kaisaki PJ, Otto GW, Ma MC et al. 2013. Genome sequencing reveals loci under artificial selection that underlie disease phenotypes in the laboratory rat. *Cell* **154**(3): 691-703.

Bruno WJ, Socci ND, Halpern AL. 2000. Weighted neighbor joining: a likelihood-based approach to distance-based phylogeny reconstruction. *Mol Biol Evol* **17**(1): 189-197.

Delaneau O, Zagury JF, Marchini J. 2013. Improved whole-chromosome phasing for disease and population genetic studies. *Nature methods* **10**(1): 5-6.

Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome biology* **4**(5): P3.

Gibbs RA Weinstock GM Metzker ML Muzny DM Sodergren EJ Scherer S Scott G Steffen D Worley KC Burch PE et al. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**(6982): 493-521.

Kong Y. 2011. Btrim: a fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. *Genomics* **98**(2): 152-153.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *Bioinformatics*: arXiv:1303.3997v1302.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* **17**: 10-12.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* **20**(9): 1297-1303.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient admixture in human history. *Genetics* **192**(3): 1065-1093.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**(3): 559-575.

Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, Rasmussen S, Stafford TW, Jr., Orlando L, Metspalu E et al. 2014. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**(7481): 87-91.

Reimand J, Arak T, Vilo J. 2011. g:Profiler--a web server for functional interpretation of gene lists (2011 update). *Nucleic acids research* **39**(Web Server issue): W307-315.

Sabeti PC Varilly P Fry B Lohmueller J Hostetter E Cotsapas C Xie X Byrne EH McCarroll SA Gaudet R et al. 2007. Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**(7164): 913-918.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**(9): 1105-1111.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols* **7**(3): 562-578.

Yang J, Lee SH, Goddard ME, Visscher PM. 2011. GCTA: a tool for genome-wide complex trait analysis. *American journal of human genetics* **88**(1): 76-82.