


RESEARCH ARTICLE

Open Access



# Genome-wide analysis reveals molecular convergence underlying domestication in 7 bird and mammals

Yali Hou<sup>1,2\*</sup> , Furong Qi<sup>1,2,3†</sup>, Xue Bai<sup>1,2†</sup>, Tong Ren<sup>1</sup>, Xu Shen<sup>1</sup>, Qin Chu<sup>4</sup>, Xiquan Zhang<sup>5\*</sup> and Xuemei Lu<sup>1,3,6\*</sup>

## Abstract

**Background:** In response to ecological niche of domestication, domesticated mammals and birds developed adaptively phenotypic homoplasy in behavior modifications like fearlessness, altered sociability, exploration and cognition, which partly or indirectly result in consequences for economic productivity. Such independent adaptations provide an excellent model to investigate molecular mechanisms and patterns of evolutionary convergence driven by artificial selection.

**Results:** First performing population genomic and brain transcriptional comparisons in 68 wild and domesticated chickens, we revealed evolutionary trajectories, genetic architectures and physiologic bases of adaptively behavioral alterations. To extensively decipher molecular convergence on behavioral changes thanks to domestication, we investigated selection signatures in hundreds of genomes and brain transcriptomes across chicken and 6 other domesticated mammals. Although no shared substitution was detected, a common enrichment of the adaptive mutations in regulatory sequences was observed, presenting significance to drive adaptations. Strong convergent pattern emerged at levels of gene, gene family, pathway and network. Genes implicated in neurotransmission, semaphorin, tectonic protein and modules regulating neuroplasticity were central focus of selection, supporting molecular repeatability of homoplastic behavior reshapes. Genes at nodal positions in trans-regulatory networks were preferably targeted. Consistent down-regulation of majority brain genes may be correlated with reduced brain size during domestication. Up-regulation of splicesome genes in chicken rather mammals highlights splicing as an efficient way to evolve since avian-specific genomic contraction of introns and intergenics. Genetic burden of domestication elicits a general hallmark. The commonly selected genes were relatively evolutionary conserved and associated with analogous neuropsychiatric disorders in human, revealing trade-off between adaption to life with human at the cost of neural changes affecting fitness in wild.

(Continued on next page)

\* Correspondence: [houyl@big.ac.cn](mailto:houyl@big.ac.cn); [xqzhang@scau.edu.cn](mailto:xqzhang@scau.edu.cn); [xuemeilu@mail.kiz.ac.cn](mailto:xuemeilu@mail.kiz.ac.cn)

†Yali Hou, Furong Qi and Xue Bai contributed equally to this work.

<sup>1</sup>Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, People's Republic of China

<sup>5</sup>Guangdong Provincial Key Lab of Agro-Animal Genomics and Molecular Breeding, and Key Lab of Chicken Genetics, Breeding and Reproduction, Ministry of Agriculture, South China Agricultural University, Guangzhou, People's Republic of China

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

**Conclusions:** After a comprehensive investigation on genomic diversity and evolutionary trajectories in chickens, we revealed basis, pattern and evolutionary significance of molecular convergence in domesticated bird and mammals, highlighted the genetic basis of a compromise on utmost adaptation to the lives with human at the cost of high risk of neurophysiological changes affecting animals' fitness in wild.

**Keywords:** Convergent evolution, Domestication and adaptation, Artificial selection, Behavioral modification

## Background

Since the days of Darwin [1] it has been recognized that a succession of species tends to possess resembling modifications in behavior, morphology and physiology in response to domestication as compared with their wild ancestors, referred to as “domestication syndromes” [2]. It hitherto remains elusive how and when domestication originally started, but regardless, a diminished fear of humans is assumed to be a crucial prerequisite during the early phase of domestication. It has also been implicated, based on the experiments conducted on fox [3], rat [4] and chicken [5], that selection on reduced fearfulness initiates the acquisition of other domestication-related behavioral modifications by means of correlated responses, encompassing less aggression, elevated stress and social tolerances, altered activity, explorative tendency, and cognition, which partly or indirectly result in consequences for economic productivity [2, 5, 6]. One of the important aspects of further research in domestication is investigation of the basis for genetic selection that leads to tameness and other behavioral traits [7], which might reflect the mechanisms and consequences in shaping behavioral alteration during the early process of domestication. So far, the investigation has rarely been done. Only the tameness-related QTLs have been independently pinpointed in domesticated rat and fox based on genetic mapping [4, 8]. However, few orthologous regions have been identified.

Adaptively phenotypic homoplasy in divergent organisms may provide sources for identifying genetic architectures underlying complex traits. Domesticated animals share the set of adaptively behavioral alterations in response to human cohabitation, representing a typical convergent system and reasonably raising the key questions of whether molecular homoplasy exhibit and, if any, what are the characteristics and mechanisms, which remain largely undetermined. Repeated phenotypic adaptation was historically thought to occur primarily by evolution of divergent genetic mechanisms, whereas recent compelling evidences suggest that it is powerful enough to drive similar molecular mechanisms in independent lineages of the nature species [9–13]. So far, application of population genomic-based methods for detection of the genomic footprints under artificial selection has unraveled that genes associated with brain development in mammals have been often targeted [14–

21], hinting towards a potential molecular similarity underlying parallel behavioral adaptations.

Parallelism of the molecular basis of convergent evolution may vary according to taxonomic scale of investigation and genetic distance of species [10, 12]. In addition to mammals, the traits in fear, aggression, activity level, explorative tendency and cognitive capability are also altered in domesticated birds [2, 22]. Such independent adaptations across genetically distant species provide an opportunity to comprehensively investigate the molecular parallelisms that may link to adaptive modification under domestication. So far, genome-wide scan for genetic basis of domestication hasn't been implemented in all species like chicken. Although pioneering studies have been performed [23, 24], they are constrained by specific Tibetan chickens or pooled samples lacking individual-based genotype information to fully profile population architectures and genomic trajectories. In this study, we first carried out whole-genome re-sequencing and RNA sequencing (RNA-Seq) of 2 brain tissues from 6 representatively domesticated breeds and 3 wild red jungle fowl (RJF) populations, characterized the genomic footprints and brain transcriptional alterations thanks to domestication, revealed genetic bases of behavioral alterations. Integrating the genomic signatures targeted by positive selection in chicken and 6 other commonly domesticated farm or accompany mammals whose genome sequencing data or genome-wide mutations or the positively selected genes due to domestication are publically available, such as cattle, pig, rabbit, horse, dog, and cat, we further identified the convergent molecular signatures at hierarchical levels of mutation, gene, gene family, pathway and network, which were significantly involved in neuroplasticity, behavior modifications and human neuropsychiatric disorders. We verified a general hallmark of genetic burden in animal domestication. Our findings provide the first genomic evidence of convergent mechanisms underlying common neurophysiological/behavioral reshapes thanks to domestication across bird and mammals.

## Results

### Genetic diversity and introgression within and between wild and domesticated chicken breeds

Based on the fact that parallelism of molecular basis of convergent evolution may vary according to taxonomic

scale of investigation and genetic distance of species [10, 12], we first characterized the genomic and transcriptional signature of domestication in chickens besides of mammals. A panel of 68 individuals from 3 RJF populations and 6 domesticated breeds that were not only specialized for meat production, egg laying, ornamental and medical purposes, but ranged from Chinese indigenous breeds to highly commercialized lines, were re-sequenced. An individual coverage of  $\sim 10X$  was achieved, followed by read mapping, variant calling and filtering. A total of 21,475,107 SNPs were identified, which was 1.88 times more than the SNPs that had been previously identified in pooled samples [23], and provided 61% more variants than those in the current variant repository (Ensembl Release 78). The details regarding samples and statistics of sequencing data and SNPs are summarized in Additional file 1: Table S1.

Understanding population architectures is fundamental for revelation of genetic basis under domestication. We first examined the genetic diversity ( $\pi$ ), divergence and introgression within and between wild and domesticated chicken populations. The level of  $\pi$  was 1.2% in RJF, which was significantly higher than those ( $0.86\% \pm 0.04\%$ ) in domesticated breeds ( $P < 0.05$ ), indicative of reduced genetic diversity during domestication. The  $\pi$  was the lowest in White Leghorn (WL) and Recessive White Rock (RW) ( $0.76\% \pm 0.03\%$ ), reflecting inbreeding of highly selected commercial strains with heightened egg and meat productions. Chinese indigenous breeds remained larger diversity ( $0.91\% \pm 0.07\%$ ).

Phylogenetic relationship and principle component analysis (PCA) congruously illustrated the divergence of the wild and domesticated chicken populations (Fig. 1). The RJFs separated from the domesticated breeds. The WL and RW first clustered together. Beijing You (YOU) and Silkie (SILK) were closer to the two commercialized breeds than other two Chinese native breeds, Xinghua (XH) and Luxi Dou (LXD). The top 2 components (PC1 and PC2), explaining 12% of the total variance, showed a strong correlation with chicken origin, breeding history and characteristic (Fig. 1b). According to PC1, the WL, a Europe-developed breed [25], showed larger difference from the Chinese native breeds than the RW. The genetic admixture of European and southeast Asian ancestors during the breeding of RW has been previously reported [26]. PC2 separated the domesticated chickens from the RJFs.

To interpret the population stratification and genetic admixture in wild and domesticated chickens, we performed a Bayesian clustering inference in ADMIXTURE that inferred the optimal number of genetic clusters to be  $K = 2$  [27, 28]. We found that WL and RW belonged to one cluster, indicating the genetic differentiation from Chinese native breeds and RJFs due to origin rather than domestication (Fig. 1c). Slight genetic admixture was

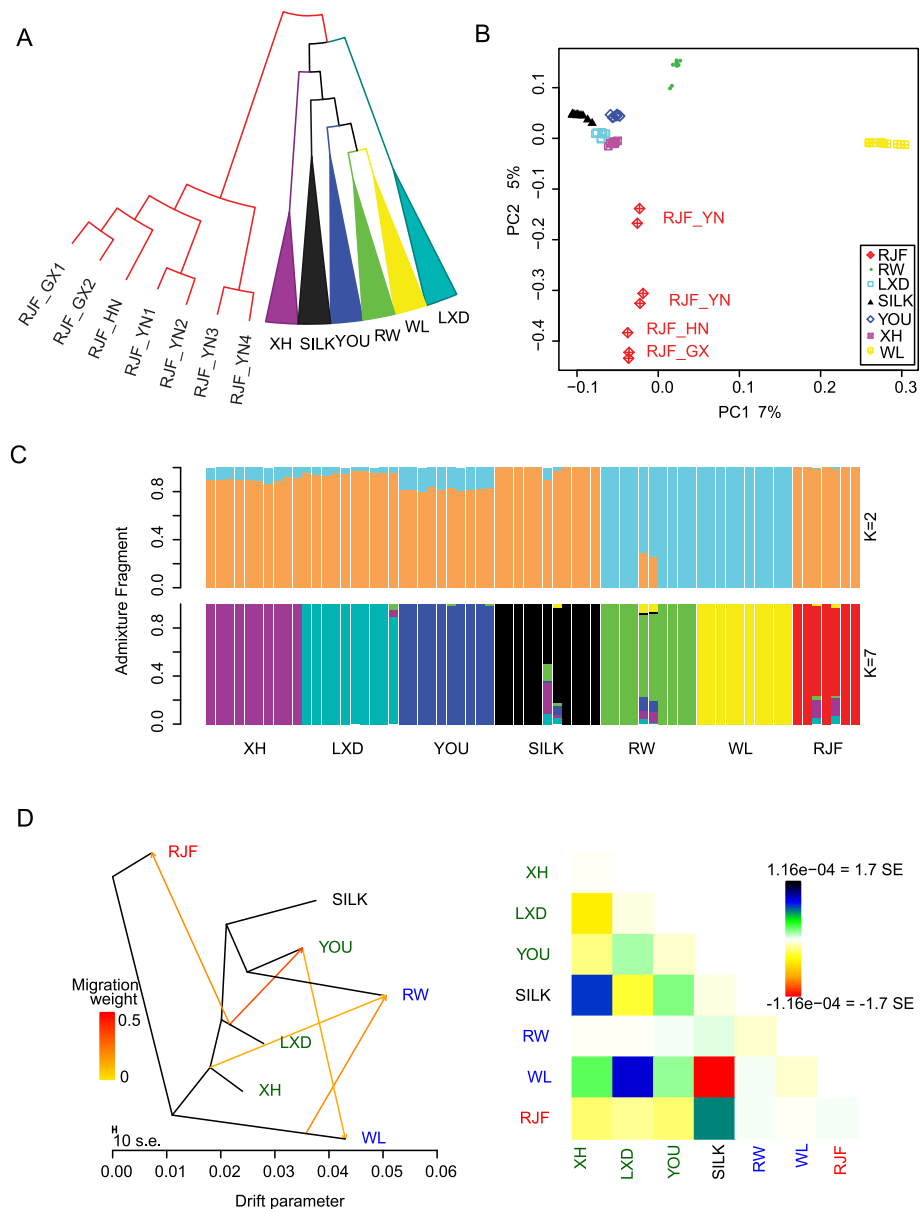
inferred by the other  $K$  values (Additional file 2: Figure S1). We used TreeMix [29] and a less-parameterized model of four-population test [30] to examine the topology of relationships and the history of genetic introgression among populations (Fig. 1d and Additional file 3: Figure S2, Additional file 1: Table S2). The inferred gene flow from LXD to RJF indicated the genetic introgression from domesticated to wild populations. Within the domesticated breeds, considerable gene flow between RW and WL ( $15\sim 20\%$ ,  $P < 2.23e-308$ ), between YOU and WL ( $13\sim 33\%$ ,  $P < 5.66e-15$ ), and between YOU and LXD ( $7\sim 30\%$ ,  $P < 1.46e-09$ ) was indicated, representing considerably genetic exchange or hybridization between distinct populations under intensive breeding programs for breed development and improvement.

### Selective sweeps and genes with selection signals in domesticated chickens

Population genomic analysis has revealed a complex genetic architecture in modern chickens. To identify genomic variations constrained by domestication rather than due to demography, introgression and specialization for various purposes, we examined selective sweeps and genes with selection signals in all domesticated breeds, which were not only used for varied purposes, but genetically divergent based on abovementioned results. We performed 5 different statistical methods, comprising the site frequency spectrum (SFS) - (heterozygosity, Tajima's  $D$ , Fay & Wu's  $H$  and SweepFinder) and linkage disequilibrium (LD) - based methods (integrated haplotype score,  $iHS$ ). To be conservative, genomic regions that were supported by at least two of the tests were identified as sweeps and used in the following analysis (Additional file 4: Figure S3).

A total of 281 selective sweeps were identified on autosomes, representing 2% of the genome. Within the sweeps, 244 of 390 Ensembl genes presented significant signals of selection in their gene body or promoter regions. Based on the Ingenuity pathway analysis (IPA), the 244 positively selected genes (PSGs) exhibited an over-representation on the biological categories of nervous system function, ophthalmic disease, developmental disorder, energy production and immunological disease (Additional file 1: Table S3), reflecting the reshapes of the nervous system, metabolism and immune system under selective regimen of domestication.

The conspicuous signals of selection were related to the nervous system which is the most enriched term of the networks. According to IPA,  $\sim 30\%$  of the PSGs (65/216) were associated with the developmental processes of nervous system and brain, including genesis, projection, branching and extension of neurons, neurotransmission, and synaptic plasticity ( $P < 0.001$ , Additional file 1: Table S4). Re-analyzing the expression pattern of these genes across 6 tissues based on the published RNA-Seq data in



**Fig. 1** Population characteristics of wild and domesticated chickens. **a** Neighbor-joining phylogeny among wild and domesticated chickens based on genome-wide SNP data. The wild progenitors include red jungle fowls from the Guangxi (RJF\_GX), Yunnan (RJF\_YN) and Hainan (RJF\_HN) provinces in China. The domesticated chickens consist of Chinese native breeds, including Guangdong Xinghua (XH), Luxi Dou (LXD), Beijing You (YOU), Silkie (SILK), and commercial strains such as Recessive White Rock (RW) and White Leghorn (WL). **b** The principle component analysis plot based on the first two principle components. **c** Population stratification and individual admixture with genetic cluster (*K*) equivalent to 2 and 7, respectively; colors in each column represent the individual ancestry proportions. **d** The maximum likelihood tree among chicken populations with 5 deduced introgression events and the residual matrix of the fitted model. The introgression events are highlighted as the arrows with colors from yellow to red, which represent the various weights of introgression (left panel). The arrow direction indicates the introgression direction. The residuals of the fitted model are illustrated in the right panel, where, the color in each cell  $[i, j]$  proportionally reflects the scaled residual covariance between population  $i$  and  $j$ , i.e. the residual covariance divides the average standard error (SE) of the observed covariances across pairs of population. The color scale bar is described in the palette on the right. Small residuals represent a well-fitted model. The fraction of the variance in relatedness among populations interpreted by the fitted model is 99.96%

RJFs [31], we found that 92% of the 65 genes were exclusively or predominantly expressed in cerebrum and cerebellum (Additional file 5: Figure S4). Intriguingly, behavioral features involved by these genes interrogated

fear response, anxiety, exploratory and hyperactive behaviors, locomotion, learning, cognition and conditioning ( $P < 0.001$ , Additional file 1: Table S4), which congruently matched the observed behavioral modifications under

animal domestication, i.e. reduced fear and anxiety, decreased explorative tendency, altered activity, locomotion, learning and memory capability [2, 5, 22].

### Gene expression changes in brain tissues of domesticated chickens compared to RJFs

The analysis of the selective sweep suggests that neurological functions and related behavioral alterations might be the main targets of selection during the early phase of chicken domestication. Performing RNA-Seq for 2 brain tissues of 11 domesticated chickens from 4 modern breeds and downloading the published RNA-Seq data in 2 RJFs [31] (Additional file 1: Table S5), we detected the gene expression difference in brain tissues between wild and domesticated chickens to investigate whether the PSGs may play a role in the gene expression changes, resulting in the traits related to brain functions and behavioral alterations in domesticated chickens.

Gene expression profiles of cerebrum and cerebellum in domesticated individuals clustered together, unambiguously separating from RJFs (Additional file 6: Figure S5). One thousand six hundred eighteen and eight hundred forty two differentially expressed genes (DEGs) in cerebrum and cerebellum between wild and domesticated samples ( $P_{adj} < 0.01$  and fold change  $> 1.5$ ) were detected. According to the gene ontology analysis, neurological processes in which the DEGs were significantly enriched ( $P < 0.01$ , Additional file 1: Table S6), were congruent with the functional classifications of PSGs. Intriguingly, genes associated with the functions of cerebrum and cerebellum, such as neurological system processes and sensory (visual and auditory) perception, were consistently down-regulated in domesticated chickens. This may imply substantial functional alteration of brains in domesticated samples, or be correlated to their decreased brain size relative to body size [32]. In contrast, genes participating in RNA splicing and translation were significantly up-regulated in domesticated animals, indicative of significance of post-transcriptional regulation during the short-term evolutionary process.

Up to 22% (53/244) of the PSGs displayed expression difference in brain tissues between wild and domesticated chickens, being a set of genes adapted to domestication at both genomic and transcriptional levels. These genes possess the potential of altering neurological functions according to their enrichment in the processes of neuritogenesis, synaptic transmission, neuron and neuroglia development, and long-term potentiation ( $P < 0.05$ , Additional file 1: Table S7). Eight of these genes have been verified to be responsible for behavioral changes in mutant mice, consisting of fear response (*GRIK2*, *TRPC5*), social interaction (*TRPC5*), learning and memory (*NF1*, *RELN*, *CSMD1*, *AQP4*), exploration and locomotion (*ERC2*, *OMG*), which have been regarded as the

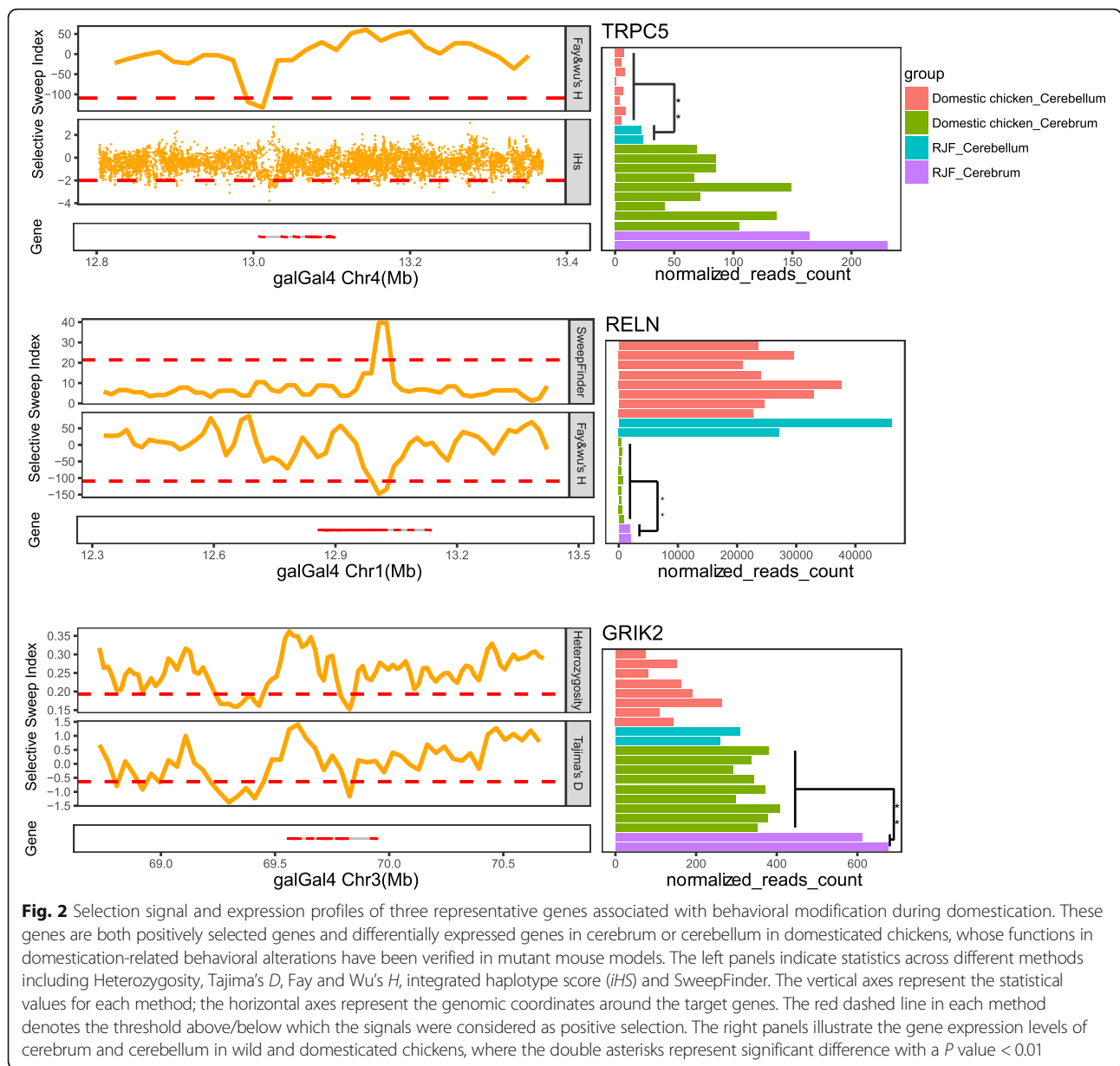
consequent behavioral phenotypes of chicken domestication [33–40] (Fig. 2 and Additional file 7: Figure S6).

To further investigate the potentially regulatory effects of the 53 PSGs on gene expression changes, we searched for the interactive genes of the PSGs in a chicken protein-protein interaction network (<https://string-db.org/>). Thirty of the 53 PSGs were present as part of the network. By implementing Wilcoxon rank-sum test, the 30 PSGs tend to be hub nodes in comparison with 30 randomly selected genes ( $P < 10^{-8}$ ), supporting the genes at nodal positions in regulatory networks are preferred targets of evolutionary changes [41]. In addition, the 30 PSGs interacted with 15% more DEGs compared to other DEGs (Chi-square test,  $P = 0.018$ ), suggesting the expression changes of these PSGs more likely resulted in regulation alteration of the DEGs in brain. Functionally, these PSGs primarily regulate the DEGs through calcium signaling, long-term potentiation and depression, CREB and nNOS signaling in neurons, axonal guidance and glutamate receptor signaling (Additional file 8: Figure S7). All these findings propose that PSGs play a vital role in regulating gene expression changes that may result in the neurological alterations and behavioral modifications that underlie chicken domestication.

### Common targets of selection involved in behavior modification among domesticated species

Modification in behavior is believed to be a key aspect of the early phase in animal domestication [2, 22]. Several previous studies have identified various genomic regions with selection signals that might be associated with domestication-related behavioral traits in different mammals [14–20]. In this study, the genes that we identified under positive selection are also involved in the neurological processes and resembling behavioral traits in chicken. Given the similarity of the consequences of domestication, it raises whether there are unique genomic properties on which strong selection prefers to act in all domesticated animals. To answer it, we profiled the genomic signals of selection and transcriptional modifications if available in 7 domesticated species to decipher the potential targets that commonly appear in a variety of animals.

Using the same methods and PSG filtering criteria that we used in chicken, we identified the selective sweeps and PSGs based on the SNPs called from whole-genome sequence data in 55 dogs and 7 wolves from DoGSD (<http://dogsd.big.ac.cn>), and 100 domesticated and 34 wild pigs in Asia and Europe from PigVar (<http://res.xaut.edu.cn/pigspis>). In total, we detected 206 and 353 PSGs in dogs and pigs, respectively. Since the sequence data and the identified SNPs in cat, cattle, horse and rabbit cannot be accessed in public database, we directly used the list of selected genes that have been reported in



the literatures for these mammals [16, 18, 20, 21]. Totally, 291, 83, 101 and 100 PSGs, from cat, cattle, horse and rabbit, respectively, were identified.

According to IPA for PSGs in each of the species, we found that the network of nervous system development and function was commonly enriched in all domesticated animals (Additional file 1: Table S8), suggesting that evolution of parallel molecular mechanisms might lead to the phenotypic convergence in nervous system and behaviors in these divergent animals. Evidences that parallel evolution may result from common variants at specific nucleotide sites, in homologous genes, pathways, and networks have been reported in many taxa [9–13]. Therefore, we integrated the PSGs involved in nervous

system functions (Additional file 1: Table S9) to search for common molecular mechanisms at multiple levels: mutations, genes, gene families, pathways and networks.

### Mutations and genes

No shared amino acid substitution was observed across domesticated species. None PSG was shared by at least 4 species. Twenty six genes were recurrently detected as PSGs in 2 domesticated species, and 1 in 3 species (Additional file 1: Table S10), whose orthology across species was confirmed by performing synteny analysis (Additional file 1: Table S11). *TCTN3*, the only gene detected in 3 species that encodes tectonic protein, functions in neural tube patterning, and causes holoprosencephaly

and neural tube defects that are the most common risks of anomalies in the central nervous system [42]. IPA showed that the shared 27 genes appeared to be responsible for domestication-related behavioral aspects like fearlessness (*GRIK2*) and learning deficit (*NPAS3*) ( $P < 0.05$ , Additional file 1: Table S12).

### Gene families

A total of 71 gene families were identified in more than one domesticated species. Based on IPA, top 2 enriched physiological processes of PSGs in these families were neurological functions and behaviors like anxiety, social exploration, cognition, exploratory behavior ( $P < 0.01$ , Additional file 1: Table S12). Based on Fisher's exact test and FDR (false discovery rate) correction, 3 families including glutamate ionotropic receptors (GluIRs), semaphorins, and tectonic proteins, were significantly targeted by selection in at least 4 species ( $P_{adj} < 0.05$ , Additional file 1: Table S10, Additional file 9: Figure S8).

Genes encoding GluIRs were selected in 6 domesticated animals, including *GRIK2* in chicken and rabbit, *GRIK3* in cattle and dog, *GRIA1* in dog and cat, *GRIA2* in cat, and *GRID1* in horse. Semaphorin genes that act as axon guidance cues like *SEMA3A*, *SEMA3D*, *SEMA3E* and *SEMA3F*, were targeted in chicken, dog, and pig, respectively, while *SEMA6A* in cattle. Besides of the tectonic gene *TCTN3* that was simultaneously selected in 3 domesticated animals (chicken, dog, and pig), *TCTN1* presented signals of selection in horse.

Gene expression changes caused by mutations on these genes may be a major determinant of phenotypic variability. Supportably, our RNA-Seq data in cerebrum and cerebellum showed that expression of genes in these families like *GRIK2* and *SEMA3A* was significantly decreased in domesticated chickens than that in RIFs (Additional file 10: Figure S9). Re-analyzing the RNA-Seq data in frontal cortex from 3 pairs of Rongchang pigs and wild boars [43], and those in cerebellum and hypothalamus from pairs of 3 dogs and 1 wolf (personal communication), we observed a consistent decrease in gene expression of *GRIK2*, *SEMA3A*, *SEMA3D* and *SEMA3E* (declined by 0.10–0.86 fold) in any brain tissues in at least 2 domesticated animals although not all of them were significant DEGs (Additional file 10: Figure S9).

### Pathways and networks

Selection may ultimately target the functional units of pathway and network [44]. To investigate if a set of PSGs as a whole to share function in common pathway or network, we mapped the PSGs relevant to neuronal functions onto KEGG (Kyoto encyclopedia of genes and genomes) pathways and performed Fisher's exact test and FDR correction to identify the significantly enriched pathways by PSGs of species. A total of 38 enriched pathways were identified in 5 or 7 domesticated species ( $P_{adj} < 0.05$ ,

Additional file 1: Table S10), among which, 7 were related to neurotransmission and 8 were diverse signal transduction pathways.

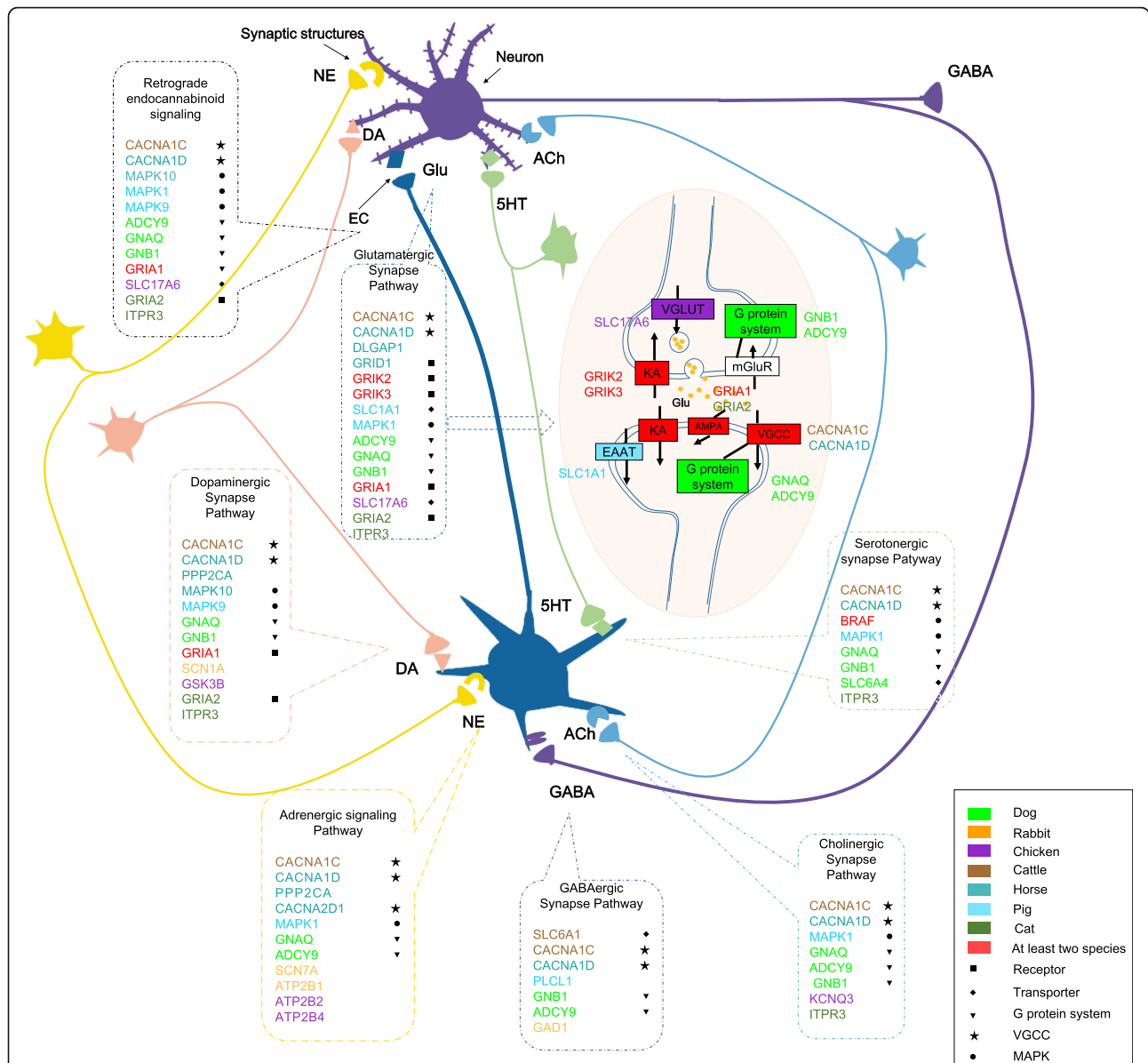
The 7 neurotransmission pathways involve transmissions of a variety of neurotransmitters of glutamate, (nor) epinephrine, dopamine, serotonin, endocannabinoid,  $\gamma$ -aminobutyric acid, and acetylcholine across presynaptic and postsynaptic neurons, which function in concert to build a complicated neuronal circuit in CNS [45]. According to adjust  $P$  values, glutamatergic synapse was the most significant neurotransmission pathway ( $P_{adj} = 4.56 \times 10^{-7}$ ), in which genes of GluIRs, voltage-gated calcium channels (VGCC), vesicular and postsynaptic transporters, and G-protein system presented significant signals of selection (Fig. 3). Within the 7 pathways, it was found that, adapted to domestication, although different genes tended to be positively selected in distinct species, there existed a repeated target on a limited set of genes encoding neurotransmitter receptors, transporters, G-protein system, VGCCs and mitogen-activated protein kinases (MAPKs), with proportions of the PSGs that involved these processes in each pathway as 54.54–91% (Fig. 3).

During neurotransmission, neurotransmitters bind to synaptic receptors on pre-/post-synaptic neurons and result in long-term excitatory or inhibitory consequences through activation of a series of signal transduction cascades. The 8 signal transduction pathways that we identified involved cAMP, cGMP-PKG, MAPK, Ras, Rap1, ErbB, calcium and PI3K-Akt signaling. To reveal whether these pathways coordinate as signaling cascades in neurotransmission, we constructed their interactive network based on the interaction relationships annotated in KEGG database (Fig. 4). It was observed that these pathways substantially interplayed with each other, forming a network-based signaling cascade and ultimately converging on CREB and MEK/ERK systems to mediate expression of genes that may regulate functions of neurons and plastic changes necessary for domestication-related behaviors.

### Genetic burden of domestication across animals

It has recently been proposed as an intriguing evolutionary concept that there exists genetic cost of domestication in most plants [46, 47] and animals like dog [48–50] and horse [16]. It is reasonably wondering would it be a generally genetic burden on animal domestication. Herein, we systematically profile this feature across taxa of dog, pig and chicken.

Similar to dog [48–50], we first introduced human and zebra finch as outgroups of pig and chicken, respectively, based on which, we determined the allele states (ancestral or derived) of mutations in wild and domestic populations. We calculated the relative occurrence of nonsynonymous versus synonymous mutations ( $d_N/d_S$  ratio or  $\omega$ ) for each individual. A significantly increased



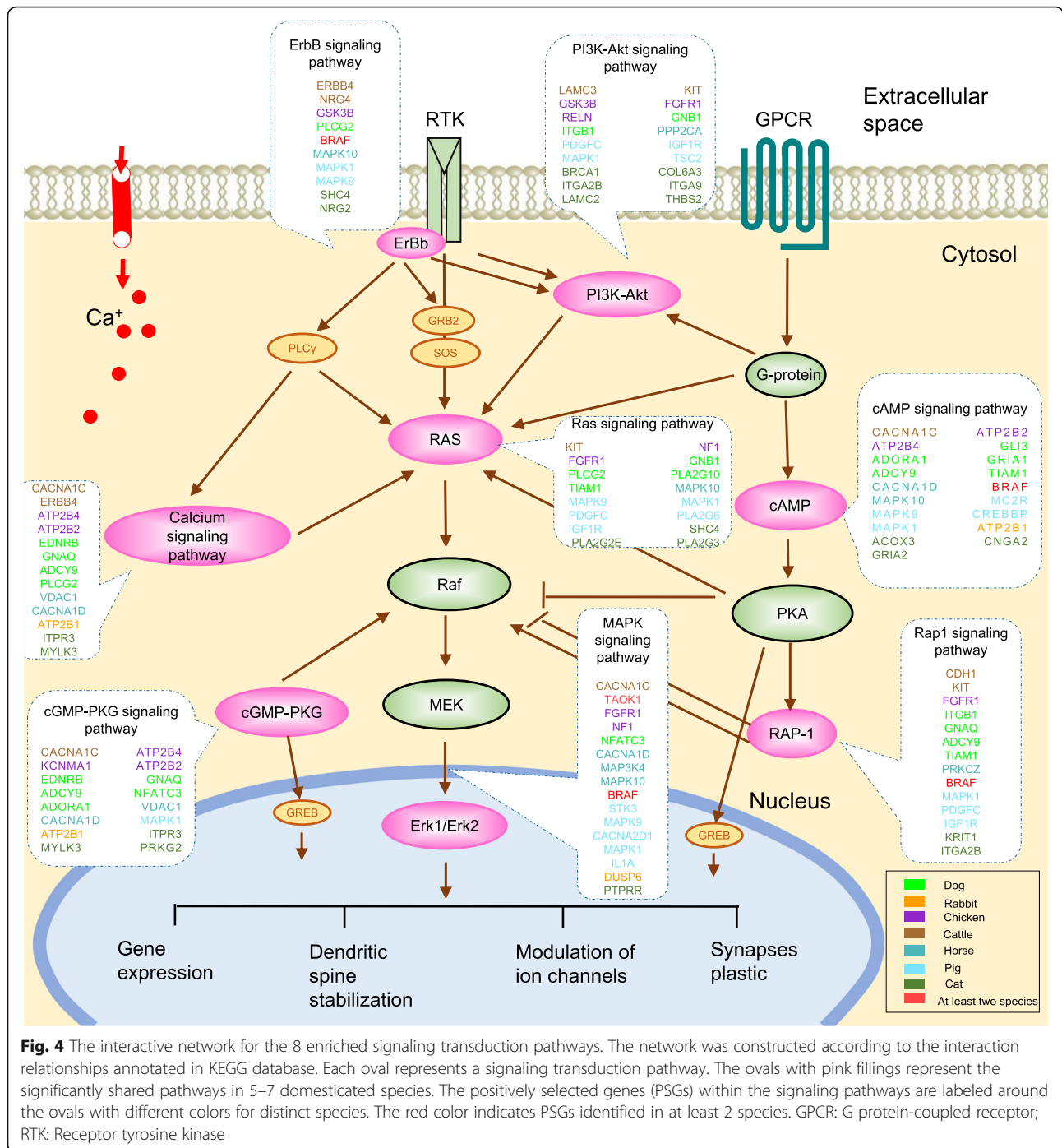
**Fig. 3** A conceptualized network for the 7 enriched neurotransmission pathways. The conceptualized figure is cited from the book of Stephen M. Stahl (2013). The 7 pathways involve transmissions of neurotransmitters of glutamate (Glu), (nor) epinephrine (NE), dopamine (DA), endocannabinoid (EC),  $\gamma$ -aminobutyric acid (GABA), serotonin (5HT) and acetylcholine (ACh) across presynaptic and postsynaptic neurons. The synaptic structures (pre- and post-synaptic terminals, and cleft) for each neurotransmitter signaling are marked on the neuron surfaces, around which, the positively selected genes (PSGs) within the signaling pathway are labeled with different colors for distinct species (chicken, dog, pig, cattle, horse and rabbit, cat). Glutamatergic synapse pathway is interpreted in detail in the middle, in which genes with signals of selection are marked

proportion of nonsynonymous SNPs (greater  $\omega$ ) was observed in domesticated chickens compared with the wild ancestors (ANOVA test,  $P < 0.01$ , Additional file 11: Figure S10), highlighting the cost of domestication, i.e. the excess accumulation of nonsynonymous mutants since domestication. The pattern was also observed in modern European pig compared with wild counterparts although the difference was not significant.

Further, based on a referred-to wild individual, we defined the allele states of mutations in domestic populations,

and predicted the functional effect of nonsynonymous mutations using the GERP algorithm that assesses the position-specific estimates of evolutionary constraint using maximum likelihood evolutionary rate estimation [51]. It was excavated that 26.7, 25.3 and 20.5% of the nonsynonymous substitutions are predicted as functionally deleterious for pig, dog and chicken, respectively, by taking GERP score greater than 2 as applied in Henn et al. [52]. These deleterious variants consistently showed the evidence of domestication cost, with a significantly elevated proportion





**Fig. 4** The interactive network for the 8 enriched signaling transduction pathways. The network was constructed according to the interaction relationships annotated in KEGG database. Each oval represents a signaling transduction pathway. The ovals with pink fillings represent the significantly shared pathways in 5–7 domesticated species. The positively selected genes (PSGs) within the signaling pathways are labeled around the ovals with different colors for distinct species. The red color indicates PSGs identified in at least 2 species. GPCR: G protein-coupled receptor; RTK: Receptor tyrosine kinase

and frequency in domestic breeds in comparison with the wild progenitors ( $\chi^2$  test,  $P < 0.005$ , Additional file 1: Table S13, Additional file 12: Figure S11). Furthermore, we unraveled the frequencies of deleterious mutation in domestic population present strong positive and negative correlations with LD magnitude (Pearson's  $r = 0.80$ ,  $P = 0.0564$ ) and nucleotide diversity (Pearson's  $r = -0.90$ ,  $P = 0.0158$ ), respectively, indicative of reduced efficacy of purify selection.

**Common targets in domesticated species are evolutionary conserved and associated with neurological disorders in human**

The convergent analysis among domesticated species indicates that genes and functional units relevant to neurological functions and behavior modifications have been identified under positive selection in both birds and mammals. Nevertheless, genes associated with brain functions are traditionally thought to be subjected to

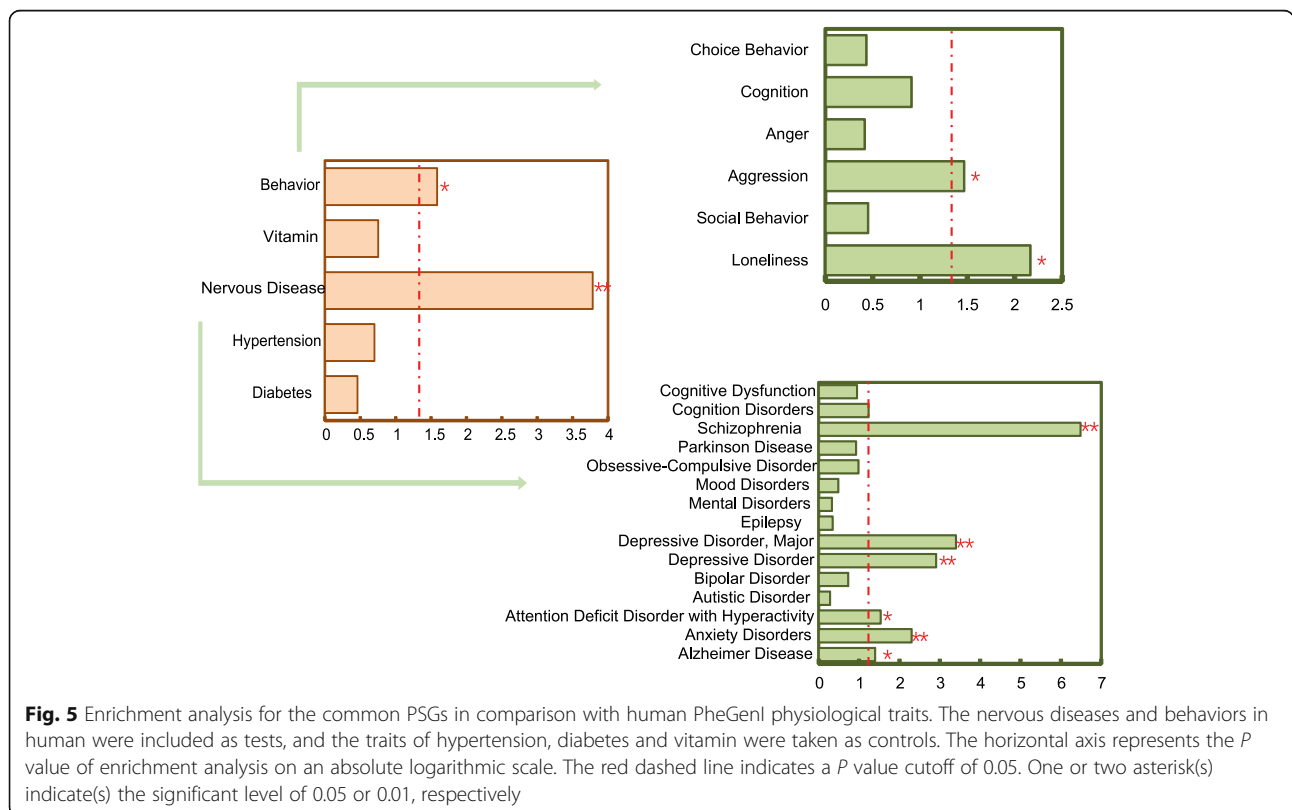
strong constraint and evolve slowly in mammals [53]. It is reasonably wondering what's the functional significance or clinical implication for these commonly targeted PSGs. Evidence has suggested that there exist conserved molecular mechanisms between animal behaviors and human mental health disorders [54, 55]. We herein deduced that the common molecular targets of selection in different animals might be associated with neuropsychiatric disorders in human.

We first defined the common molecular targets of positive selection as PSGs that showed signatures of selection in at least two species and were members in the common gene families and pathways that were shared by at least two species. We investigated degree of evolutionary conservation for the common PSGs. The phastCons and phyloP scores of the common PSGs among 100 vertebrates were significantly higher than those of background genes (Additional file 13: Figure S12), indicative of a high degree of conservation. We scanned the common targets against the vulnerability loci of various human physiological traits retrieved from genome-wide association studies from the PheGenI project [56], and revealed a significant enrichment of the common target genes in neurological diseases and behavioral disorders relative to control traits such as diabetes, hypertension, and vitamin traits ( $P < 0.05$ , Fig. 5, Additional file 1: Table S14). The overrepresented diseases and disorders

by these common genes, such as anxiety, aggression, attention deficit hyperactivity, schizophrenia, depression and loneliness, displayed phenotypic analogy with domestication-related behavioral alterations. Apart from the genes that already existed in PheGenI database as putative risk factors of human neuropsychiatric diseases, the common target genes provided more novel candidates potentially related to the overrepresented diseases through functional interaction with the known risk factors in the same pathways (Additional file 14: Figure S13). The results suggest that domesticated animals could be natural disease models for understanding the role of genes implicated in neuropsychiatric disorders in human.

### Discussion

A fundamental question in evolutionary genetics concerns the extent to which parallel phenotypic adaptation in divergent organisms is attributed to convergent or parallel changes at molecular levels [12]. Recent evidence has suggested that the adaptively genetic evolution might be more predictable than was previously appreciated in nature species [9–13, 57–60]. However, basis and patterns of molecular homoplasy across domesticated species in response to intensified selection during a short period of time rather than nature selections in a long-term run remain largely elusive.



Parallelism of the molecular basis of convergent evolution may vary according to taxonomic scale of investigation and genetic distance of species [10, 12]. To investigate the general rule of adaptive convergence under domestication in divergent birds and mammals, we first profiled the signatures of chicken domestication at both genomic and transcriptional levels. Genetic diversity substantially reflects chicken breed origin, demography and specialization for various purposes. There exist massive introgressions from domesticated to wild chickens and among distinct breeds. Continuous gene flow to wild populations during or after domestication has also been reported in dog, pig and horse [61–63], whether the immigrant alleles are neutral, detrimental, or beneficial needs detailed spatio-temporal factors for further exploitation [64]. Considerable gene flow among breeds was attributed to intensive breeding programs for breed development and improvement. These results suggest introgressions need to be carefully considered when studying genetics and evolutions among wild and domesticated animals.

Consistent with domesticated mammals, chicken PSGs exhibit enrichment in neurological processes and brain functions, suggesting of parallel molecular mechanisms that may result in phenotypic convergence in behavior alterations. We have revealed novel genetic architectures and physiologic bases of behavioral alterations and vision degeneration during early phase of chicken domestication. Part of the identified PSGs have been suggested association with reduced fear, anxiety and explorative tendency, altered activity, locomotion, learning and memory capability using mouse model [33–40]. Intriguingly, genes relevant to brain functions in cerebrum and cerebellum were consistently down-regulated in domesticated chickens, which may imply substantial functional alteration, or be correlated to their decreased brain size relative to body size [32]. Consistently, knockout of *SLC2A1*, a down-regulated gene, has been verified to be associated with neurodegeneration, behavioral deficits and microcephaly, i.e. abnormally small head in mice [65]. Similar pattern was also observed in dog, where majority of DEGs (80.44–92.82%) were down-regulated in comparison with wolf across brain tissues of cerebellum and hypothalamus. These results may be correlated to the fact that domesticated animals of the same species evolved significantly reduced brain sizes compared to their wild relatives, relative to body size, in the process of adaptation to ecological niche of domestication [32]. Contrastly, although domesticated chickens were also found to have relatively reduced brain mass compared to RJF, they underwent an absolutely enlarged cerebellum independent of increased body size that was supposed to occur in light of domestication [66]. Due to the complexity, genetic mechanism by which domestication affects brain size or composition warrants further

investigation. In contrast, genes participating in RNA splicing, e.g. the spliceosome snRNA or proteins and serine/arginine proteins, were significantly up-regulated, highlighting significance of post-transcriptional mechanism, i.e. RNA splicing, in response to chicken domestication. Whereas, this pattern is ambiguous in dog and pig by additionally re-analyzing the RNA-Seq data in frontal cortex from pairs of Rongchang pigs and wild boars [43], which may be attributed to the fact that avian genomes experienced extensive shorting of introns and reduced intergenic distances compared to mammals [67] and alternative splicing might be the mostly efficient mechanism to increase proteome diversity to adapt to evolutionary transition.

Of note, from the perspective of network for the PSGs in chicken, we have observed that the genes at nodal positions in a *trans*-regulatory network are preferred targets of domestication. To test whether this is also applicable to other domesticated animals, we scrutinized the PSGs in dog and pig, respectively. By performing Wilcoxon rank-sum test, the PSGs tend to be hub nodes in comparison with randomly selected genes with comparable gene numbers ( $P < 10^{-8}$ , Additional file 15: Figure S14), supporting the hypothesis that genes at nodal positions in regulatory networks are favored targets of evolutionary changes [41].

By integrating the selection signals in the genomes of chicken and 6 domesticated mammals, we profiled the genetic basis and patterns of parallel evolution at different hierarchical levels. No common amino acid substitution is observed, which indicates adaptively convergent phenotypes may not always arise from the same substitution, revealing a potential role of contingency in the context of epistatic interactions in divergent species [10, 68]. In contrast, mutations in adaptation to domestication presented a convergent enrichment in non-coding regions in chicken, pig and dog (Additional file 16: Figure S15), where abundant *cis*-regulatory elements reside. These support that mutations in regulatory sequences may be important for driving phenotypic adaptation in response to domestication. Non-coding region like introns have been studied as essential mediators of cell response to starvation [69].

A convergence of domestication occurs at the level of gene and gene family. A limited gene set, especially those encoding GluIRs, semaphorins and tectonic proteins, was repeatedly targeted across species. GluIRs respond to glutamate via receptor-type-dependent neurotransmission, regulating neuronal excitability and synaptic plasticity. *GRIK2* and *GRIA1* have been confirmed as molecular basis of neurodegeneration and domestication-related behavioral alterations like fearfulness, anxiety, social exploration, learning and memory [34, 70]. *GRIK3* and *GRIA1* have also been identified as overlapped genes under selection in domesticated dogs, cattle, horses and cats [71]. Semaphorins act as axon guidance molecules by specifying cell

morphology or inducing cell collapse as a result of change in cytoskeleton and cell adhesion, mediating neural circuit developments including, but not limited to, neurite extension, dendrite specification, axon sorting and synaptic specificity [72]. Association of *SEMA3F* with fear and *SEMA6A* with exploration and learning has been reported in knockout mice [73, 74]. Tectonic protein gene *TCTN3* functions in neural tube patterning and exhibits holoprosencephaly and neural tube defects in the mutant mice [42]. Intriguingly, we have noticed that five shared positively selected genes or gene family members, including *BRAF*, *SEMA3A*, *SEMA3F*, *SEMA3D*, and *MAPK1*, are known to be involved in neural crest cell development and migration [75, 76], which appears to support the neural crest hypothesis of domestication syndrome [77].

A striking common pattern targeting pathways of neurotransmission and signal transduction cascades was uncovered across species, among which glutamatergic synapse is the mostly shared. Both neurotransmissions and signal transductions are known to regulate synaptic plasticity and neural circuits that underlie emotional, social and cognitive function, in which many genes have been demonstrated association with domestication-related behavioral alterations such as reduced fear and explorative tendency, altered learning and memory capability in a variety of studies in knockout mice [78–81]. Although different genes within the same pathway tend to be selected in distinct species, there exists a repeated target on genes encoding neurotransmitter receptor, transporter, G-protein system, VGCC, MAPK and signaling cascades. GluIR and ERK/MAPK signaling pathways has previously been highlighted in multiple domesticated animals [71]. These results highlight that domesticated species repeatedly evolved similar behavioral adaptations through certain molecular repeatability at the levels of pathways and networks, but having more than one solution, i.e. diverse selected genes in distinct species within the same pathway. In spite of molecular convergence, different domesticated animals experienced significantly varied domestication routes, for instances, dog, cat, chicken and pig for the commensal pathway with diverged purposes, cattle and rabbit for the prey pathway, and horse for the directed pathway [82]. Besides, changes in social ecology of animals (i.e., both feeding niche and social organization), along with other parameters such as interaction into environments, have also been recently suggested to contribute to domestication process [83]. How different routes of domestication give rise to different biological processes will be targeted in the future study.

The predominant factor accounting for the genetic burden is rationally attributed to the Hill-Robertson effect that selection on linked sites hinders the purging of deleterious variants and impinges the overall effectiveness of

selection in finite populations in the absence of recombination [84]. Intensive artificial selection on beneficial sites during domestication and breed improvement decreases locus-specific  $N_e$ , thereby increases the retain probability of linked deleterious mutations due to hitchhiking, and hinders their efficient purging under purifying selection [46, 48, 49, 84]. Demographic history such as severe bottlenecks during domestication and breed formation that leads to relaxation of purifying selection on slightly pernicious mutations also contributes to the accumulation [47, 48]. Summarily, accumulation of detrimental mutations is widely acknowledged as the genetic load of domestication across a wide variety of domestic species, eliciting a general hallmark of domestication.

The commonly selected genes associated with domestication-related behavioral alterations were a set of conserved genes, which benefitted the fitness of species, i.e. ability to survive and reproduce in wild environment [85]. By contrast, these genes that experienced positive selection during domestication may give rise to high fitness of domesticated animals under circumstances of human beings. The adaptive changes caused by these genes in reduced fear, aggression and anxiety, increased social tolerance, altered activity and cognition, were all preferable for successful domestication [2, 5, 6]. The commonly selected genes displayed enrichment in analogous neuropsychiatric disorders in human, consisting of anxiety, aggression, attention deficit hyperactivity, schizophrenia, depression and loneliness, and provided potential novel risk factors underlying these diseases. Reports on couple positively selected genes in multiple domesticated species that are strong candidates for neurodevelopmental diseases and syndromes support the finding [71]. These results render domesticated animals as natural disease models in understanding the roles of genes implicated in neurodegeneration and neuropsychiatric disorders in human.

## Conclusions

In this paper, we provide one of the genome-wide views of the evolutionary trajectories, including population hierarchy and introgressions, among 6 domesticated chicken breeds and 3 wild RJF populations. Our study reveals strong signatures of selection around genes associated with the nervous system and behavior modification, similar to mammals adapted to domestication. We have systematically recapitulated the molecular mechanisms of parallel evolution underlying behavior modifications across 7 domesticated species, and reveal an appreciated genetic bias during the evolution. We disclose the repertoire of convergent selection molecules that used to be constrained by purify selection has experienced positive selection after transitions from wild to domesticated constraints, and propose conservation between these genes and risk factors underlying human

neuropsychiatric disorders. These findings implicate a compromise between utmost adaptation to human cohabitation and higher risks of neurophysiological changes that influence the fitness of the animals in wild. Our analyses provide valuable theories for the underlying convergent evolution across birds and mammals.

## Methods

### Chicken breeds, samples, and tissues

A total of 68 individuals from 7 typical populations composed of 4 Chinese indigenous breeds (XH from the Guangdong province; LXD from the Shandong province; YOU from Beijing; SILK from the Jiangxi province), 2 highly commercialized breeds (RW and WL), and 3 wild red jungle fowl populations (RJF) were investigated (Additional file 1: Table S1). For each breed, 6 males and 4 females were sampled for whole-genome sequencing (WGS). The blood samples were taken from the wing vein by the standard venepuncture procedure. Chickens were kept alive after blood collection. A Taiwan Silkie with accession number of PRJNA202483 was also included [86]. For RJF, 1 male and female from the Guangxi province were sampled for WGS. In addition, 3 males and 2 females from the Yunnan and Hainan provinces with accession number PRJNA241474 were obtained [87].

Eleven adult chickens were anesthetized with phenobarbital sodium solution (30 mg/kg) and immediately euthanized by rapid decapitation. Seventeen cerebrum and cerebellum samples from eleven LXD, SILK, RW, and WL males were dissected and collected for RNA-Seq (Additional file 1: Table S5). Additionally, 4 transcripts of cerebrum and cerebellum for RJFs (One male and one female for each tissue) with accession numbers SRR306710–306713 were also analyzed [31].

### DNA extraction, library construction, and WGS

Genomic DNA was extracted from blood with QIAamp® DNA Mini Kit (QIAGEN, Duesseldorf, Germany). DNA quality and concentration were monitored using agarose gel electrophoresis, Bioanalyzer 2100 system (Agilent Technologies, CA, USA), and Qubit® 2.0 Fluorometer (Life Technologies, CA, USA). Sequencing libraries were constructed by following IlluminaTruSeq™ DNA preparation kit, and sequenced on Hiseq2000 platform. Approximately 92~140 millions of clean paired-end reads with a length of 100 bp were generated for each individual (Additional file 1: Table S1).

### RNA extraction, library construction, and RNA-Seq

Total RNA was extracted from tissues using TRIzol reagent (Invitrogen, CA, USA). RNA quality and concentration were determined using NanoPhotometer® spectrophotometer (Implen, CA, USA), Qubit® 2.0 Fluorometer (Life Technologies, CA, USA), agarose gel electrophoresis, and

Bioanalyzer 2100 system (Agilent Technologies, CA, USA). Sequencing libraries with poly(A) enrichment were constructed by following IlluminaTruSeq™ RNA sample preparation kit, and sequenced on Hiseq2500 platform, achieving an average of 21.86 million clean paired-end reads with a length of 125 bp for each sample (Additional file 1: Table S5).

### Detection of genomic variations

Clean paired-end reads were aligned to the indexed chicken reference genome (galGal4) by running the maximal exact matches algorithm in BWA [88]. After alignment, multi-sample calling based on Bayesian inference algorithm [89] was implemented to identify and genotype SNPs using SAMtools and BCFtools [90]. PCR duplicates were removed by applying rmdup. Mpileup was used to pile up the base-pair alignment information for reads, and -C50 was used to downgrade the mapping quality for reads containing excessive mismatches. A site was considered to be a SNP when the following post-filters were applicable: 1) read depth less than 100; 2) base quality score  $\geq 20$  and mapping quality score  $\geq 10$ ; 3) variant occurred with a probability of 99%, i.e., the variant/reference quality score  $\geq 20$ ; 4) alternative allele supporting reads  $\geq 3$ .

### Calculation of genetic diversity and introgression within and between chicken populations

The genetic diversity ( $\pi$ ) was calculated as normalized SNP number by the genome size, which was compared across populations using ANOVA and Shaffer's multiple comparison tests.

Neighbor-joining tree was constructed using PHYLIP [91] based on 1,809,179 independent SNPs. SNPs in strong LD, i.e., having  $r^2 > 0.1$  with neighboring SNPs, were excluded using a 100-SNP sliding window (stepped by 20 SNPs) implemented in PLINK [92]. Genetic divergence among chicken populations was investigated using PCA in EIGENSOFT program [93] based on all SNPs.

Population stratification and genetic admixture in wild and domesticated chickens were inferred based on independent SNPs by implementing ADMIXTURE [27]. Various genetic clusters ( $K$ ) (ranging from 2 to 7) were applied. The optimal  $K$  was determined through a cross validation procedure [28].

Introgression was derived using TreeMix [29]. Maximum likelihood bifurcating tree was constructed based on the independent SNPs. Introgression events were subsequently interrogated using "-m 8" based on all SNPs, where TreeMix accounted for LD by grouping SNPs in a window size ( $-k$ ) of 100,000, corresponding to 6.26 Mb, which was comparable to the window size (10 Mb) that was used for humans [94]. The fraction of variance that can be interpreted by the fitted model was

calculated as described in Pickrell and Pritchard (2012) [29].

Four-population test was executed in TreeMix to verify the derived introgression events. The  $f_4$  (A, B; C, D) was calculated as described in Reich et al. (2009) [30], and its standard error (SE) was estimated using a Block Jackknife procedure with a block size of 100,000 SNPs. To assess significance of introgression, Z-score was calculated as  $\frac{f_4(A,B;C,D)}{SE_{f_4(A,B;C,D)}}$ , which normally distributed with mean = 0 and variance = 1 under the null hypothesis [30]. Significantly positive Z-score (Z-score > 4) indicates gene flow between A and C or between B and D, whereas, negative (Z-score < -4) implies introgression between A and D or between B and C [95].

### Identification of putative selective sweeps (PSS) and genes with selection signals

Based on previous studies [15, 17, 23], we determined the window size for selective sweep screening as 40 kb for chicken and 100 kb for pig and dog, which far exceed the LD extent to avoid confusion between selection and demography. Window-sliding step was set as half of the window size.

Heterozygosity ( $H_p$ ) [23] was calculated as  $H_p = 2 \sum n_{\max} \sum n_{\min} / (\sum n_{\max} + \sum n_{\min})^2$ , where  $n_{\max}$  and  $n_{\min}$  are numbers of the most and least alleles within populations at a specific SNP, and  $\Sigma$  is the sum of SNPs within windows. Threshold to claim a PSS was determined based on a randomly shuffled method [96]. Genome was split into 1-kb windows with estimates of  $\sum n_{\max}$  and  $\sum n_{\min}$ , which were randomly shuffled for 1000 times to disturb LD. For each shuffled genome, we calculated  $H_p$  within windows and recorded the minimum. This resulted in a set of minimums with a dimension 1000. We chose the 50th lowest value as cutoff, which indicated a significant level of 0.05.

Tajima's  $D$  [97] was deduced as  $\frac{\hat{\theta}_\pi - \hat{\theta}_w}{\sqrt{\widehat{Var}(\hat{\theta}_\pi - \hat{\theta}_w)}}$ , where  $\hat{\theta}_\pi$  indicates the average heterozygosity within populations, equivalent to  $\sum_{i=1}^{n-1} \frac{2S_i i(n-i)}{n(n-1)}$ ;  $S_i$  is the number of derived variants found  $i$  times in a sample of  $n$  chromosomes;  $\hat{\theta}_w$  is derived from the number of segregating sites equal to  $(\sum_{i=1}^{n-1} \frac{1}{i})^{-1} \sum_{i=1}^{n-1} S_i$ . The calculation was implemented using VCFtools [98]. In contrast, Fay & Wu's  $H$  [99] was calculated as  $\frac{\hat{\theta}_\pi - \hat{\theta}_H}{\sqrt{\widehat{Var}(\hat{\theta}_\pi - \hat{\theta}_H)}}$ , where  $\hat{\theta}_H = \sum_{i=1}^{n-1} \frac{2S_i i^2}{n(n-1)}$ . The deduction was performed using VariS-

can [100]. For both statistics, the bottommost 1% windows were considered as PSSs.

Haplotypes were constructed using fastPHASE [101] based on SNPs with minor allele frequency > 5%.  $iHS$  [102] for each core SNP was derived as  $\frac{\ln(\frac{iHH_A}{iHH_D}) - E_p[\ln(\frac{iHH_A}{iHH_D})]}{SD_p[\ln(\frac{iHH_A}{iHH_D})]}$ , where  $iHH_A$  indicates the integral of decay of extended haplotype homozygosity ( $EHH$ ) away from the ancestral allele until  $EHH$  reaches 0.05;  $iHH_D$  is the integrated  $EHH$  with respect to the derived allele;  $E_p()$  and  $SD_p()$  are the expectation and standard deviation of  $\ln(\frac{iHH_A}{iHH_D})$ , estimated from the empirical distribution at SNPs whose derived allele frequencies ( $p$ ) match that of the core SNP. Further, a statistic of a window was defined as proportion of SNP with  $iHS < -2$  [102, 103]. The corresponding empirical  $P$  value was assigned as follows [103]: 1) binned the windows by similar SNP numbers with an increment of 10 and 2) for each bin, the empirical  $P$  value for a specific window was assigned as the fraction of windows owning relatively larger statistics. During the process, the bins with windows less than 20 were removed. The windows with the lowest 1% of the empirical  $P$  value were considered as PSSs.

SweepFinder was implemented to determine PSSs based on spatial pattern of SFS around the focal SNP and the algorithm of composite likelihood ratio (CLR), where null hypothesis is neutral SFS pattern and alternative hypothesis allows selection, conferring robustness to complex demography and recombination rate [104]. Grid size was assigned as number of SNPs on each chromosome since the method is single point based [105]. For the focal SNP, surrounding sites were taken into account with a cut-off of  $\alpha d = 12$ , where  $d$  is the distance to the focal SNP;  $\alpha = r \ln(2N)/s$ ;  $N$  is the population size;  $r$  is the recombination rate per base pair;  $s$  is the selection coefficient. The cut-off corresponds to a probability of escaping a sweep of 0.999994 for surrounding SNPs. The maximum of CLR within a window was presented as the statistic. Eventually, the windows with the topmost 1% of the statistics were considered as PSSs.

Final selective sweeps were ascertained as the conserved set that were identified by at least two of the methods so as to limit method-specific bias. Additionally, selective sweeps do not necessarily mean that the overlapped genes have selection signals. We further filtered the genes with selection signals using the following procedures: 1) for each method, with the exception of  $iHS$ , we calculated 1-kb window statistics within the selective sweeps. The 1-kb windows with 2% bottommost statistics (topmost for SweepFinder) that were assumed to have selection signals were selected, while the windows with SNPs less than 2 were excluded; 2) the

sweep-overlapped genes were eliminated when they or their promoter regions (2 kb upstream gene start site) were barely supported by any of the selected 1 kb-windows; 3) since the selected 1 kb-windows for Tajima's  $D$  and heterozygosity are bias to have less SNPs (Additional file 4: Figure S3B), the genes with selection signals were required to encompass at least 10 selected 1-kb windows to avoid Type I errors, where the filter strategy of 10 was determined based on the frequency of selected 1-kb windows that supported the genes (Additional file 4: Figure S3C); and 4) the genes containing any SNP with  $iHS < -2$  were included since  $iHS$  considers not only the focal SNP but also the surrounding mutations. Specifically, the pigs in Asia and Europe were independently analyzed to account for distinct origins and the final PSGs were combined.

#### Gene expression pattern of the 65 PSGs related to neurodevelopment

Gene expression pattern of the 65 PSGs is based on the published RNA-Seq data of cerebrum, cerebellum, testis, liver, kidney and heart in RJF with accession number SRR306710-SRR306713 [31]. The expression level for each gene was characterized by reads per kilobase of exon model per million mapped reads ( $RPKM$ ), and was calculated based on the  $G$ -values provided by Brawand et al. [31]. Formula for calculation referred to Qi et al. [106].

#### Identification of differentially expressed genes

Clean paired-end reads from RNA-Seq data for each individual were aligned to the reference genome using TopHat v2.0.13 [107] with default settings. The normalized count value for each gene was calculated by the relative log expression method of DESeq [108] based on raw count provided by HTSeq [109]. The phylogenetic relationships among transcriptomes were profiled using heatmap under R environment with a distance algorithm of Pearson correlation and a hierarchy method of complete Hclust. DEGs between groups were identified using DESeq with filters of  $P_{adj} < 0.01$  and fold change (FC)  $> 1.5$ .

Six transcriptomes of frontal cortex from 3 pairs of Rongchang pigs and wild boars with accession numbers SRR5470021~5,470,026 [43] were downloaded and re-analyzed.

#### Determining the predisposition to being hub nodes for the 30 PSGs

A chicken protein-protein interaction network was accessed from STRING database (version 10.5, <https://string-db.org/>), encompassing 7108 genes and 399,271 interactions if interrogating interaction types of activation, binding, catalysis, inhibition, post-translational modification, and reaction. To determine whether the 30 PSGs are

predisposed to being hub nodes in the network, we compared the degrees of the 30 genes and those of 30 randomly selected genes in corresponding sub-networks. Degree of a gene was defined as number of interactions that the gene had with the other genes in the network. For the 30 PSGs, we extracted a sub-interaction network, comprising 1800 genes and 3166 interactions. Within the sub-network, we calculated degrees and degree ranks of the 30 PSGs. For comparison, we randomly selected 30 genes for 1000 times. For each time, we extracted the sub-network and calculated the degree ranks of these genes. The median degree ranks of the 1000 samplings were assigned as those for 30 randomly selected genes. We compared the ranks of the 30 PSGs and those of 30 randomly selected genes using Wilcoxon rank-sum test. The PSGs presenting a significantly higher rank of degree were considered to be predisposed to being hub nodes.

#### Definition of gene family, pathway, and network for PSGs

We intersected the PSGs with gene family that was approved by HUGO gene nomenclature committee and curated in an online repository (<https://www.genenames.org>). We mapped the PSGs onto the annotated KEGG pathways (<https://www.genome.jp/kegg/pathway.html>). A high degree of conservation in the functional properties of proteins encoded by orthologous genes across domesticated species and human, especially in the context of pathway, is assumed based on the previous evidence and the feature of protein-sequence similarity [110, 111]. Fisher's exact tests and false discovery rate corrections of Bonferroni correction were applied to test the enrichment of the PSGs in annotated gene families or pathways in HUGO and KEGG datasets. We constructed the network of 8 signal transduction cascades by interrogating interaction relationships in KEGG pathway database.

#### Association of the commonly targeted genes of positive selection with human neuropsychiatric disorders

We downloaded the human Phenotype-Genotype dataset from PheGenI project [56]. The commonly targeted genes by positive selection were compared to the vulnerability loci of human nervous system diseases (including Alzheimer Disease, Autism Spectrum Disorder, Bipolar Disorder and other 19 terms) and behaviors (including Loneliness, Aggression, Cognition and other 12 terms), with those of Hypertension, Diabetes, and Vitamin as controls. Details of the human PheGenI physiological traits are presented in Additional file 1: Table S14. Fisher's exact test was applied to determine the enrichment of the commonly targeted genes on the vulnerability genes of human physiological traits.

### Enrichment analysis of the putatively adaptive mutations

We defined SNPs within sweeps with  $iHS < -2$  in domesticated animals as putatively adaptive mutations in response to domestication. We downloaded gene annotation via ANNOVAR (Wang et al., 2010), and calculated the percentages of exon, intron, upstream, downstream, UTR3, UTR5, splicing and intergenic regions across selective sweeps. Upstream and downstream regions were defined as 1 kb upstream and downstream the transcription start and end site, respectively. For each region, SNPs with  $iHS < -2$  were counted as observed number, and the expected number was calculated as the total observed number multiplied by the percentages of this region. Based on the observed and expected numbers for each region, significant level was determined by Chi-square test and further corrected by Bonferroni adjustment.

### Calculation of $d_N/d_S$ ratios in wild and domestic populations across species

We utilized zebra finch as outgroup to unbiasedly deduce the allele states (ancestral or derived) of mutations in wild and domestic chickens. We firstly downloaded the chained BlastZ pairwise alignments of zebra finch (taeGut2) and chicken genomes (galGal4) from the University of California, Santa Cruz Genome Browser (<http://hgdownload.soe.ucsc.edu/goldenPath/galGal4/vsTaeGut2/>). The alignments were further trimmed into protein-coding regions according to the chicken genome annotation (galGal4), comprising a total of 17,041 Ensembl genes and 23.43 Mb in length (1% of the total). The identified SNPs for each chicken were then assigned onto the trimmed alignments, where the allele states were exquisitely determined. When a site represented distinguished alleles between wild and domestic chickens and one of those matched the zebra finch sequence, the base from zebra finch was considered as ancestral state. Cases where a site was endowed with at least three types of alleles across zebra finch, wild and domestic chickens were ignored. The trimmed alignments were concatenated and further analyzed using CODEML in PAML to calculate the  $d_N/d_S$  ratio in wild and domestic chickens. ANOVA was implemented for population comparison. Similar processes were implemented for pig by using outgroup of human.

### Identification of deleterious nonsynonymous mutations

According to the well-clustered wild individuals and domesticated individuals, as well as the substantial genetic divergences between them, principally, any of the wild sample could be used to define the allele states of mutations in domestic populations despite of few improperly determined mutations. For chicken, the reference genome is indeed a wild bird, which was used to infer the

allele states of mutations. For dog, we used LUPWCHN00001 (Grey wolf, Haerbin, China) as the reference. For pig, SRS387323 (Wild boar-2) was used as reference of Asian Pig, and WB25U11 (WB25U11\_Wild-Boar) of European pig. We computed the genomic evolutionary rate profiling (GERP) scores using the GERP++ code for all nonsynonymous SNPs, which reflect whether an amino acid substitution is likely to significantly affect protein function based on sequence homology and physical property of amino acids. We downloaded FASTA alignments of 45 vertebrate genomes aligned to hg19 (<http://hgdownload.cse.ucsc.edu/goldenPath/hg19/multiz46way/alignments/>), and ran each exon of gene through GERP++, omitting the dog and pig sequence, respectively, where the phylogenetic tree “46way.nh” was used. We then converted the CanMap3.1 coordinates of dog SNPs and susScr3 coordinates of pig SNPs to hg19 coordinates using LiftOver to annotate them with a GERP score. Sites with  $> 2$  rejected substitution score per site were considered deleterious.

### Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12864-020-6613-1>.

#### Additional file 1. Supplementary Tables

**Additional file 2: Figure S1.** The inferred population stratification and individual genetic admixture in wild and domesticated chickens. A. The cross-validation procedure to infer the optimal number of genetic clusters ( $K$ ) that presents the minimum cross-validation error. B. The delivered population stratification and individual genetic admixture with varying  $K$  (ranging from 2 to 7); colors in each column represent the individual ancestry proportions. XH: Guangdong Xinghua, LX: Luxi Dou, YOU: Beijing You, SILK: Silkie, RW: Recessive White Rock, WL: White Leghorn, RJF: Red jungle fowls.

**Additional file 3: Figure S2.** The deduced introgression events among chicken populations. RJFs are integrated into one outgroup in A and B, and are subdivided into distinct populations based on their geographic locations (RJF\_HN from the Hainan province, RJF\_GX from Guangxi, RJF\_YN from Yunnan, among which, RJF\_YN\_mix indicates two admixed individuals) in C and D. A and C. (Left panel) The maximum likelihood tree (the breed symbol in red color represents red jungle fowls, blue: the commercial breeds, green: Chinese native breeds, black: the Silkie) and (Right panel) the residual matrix of the fitted model, where, the color in each cell  $[i, j]$  proportionally reflects the scaled residual covariance between population  $i$  and  $j$ , i.e. the residual covariance divides the average standard error (SE) of the observed covariances across pairs of population. The color scale bar is described in the palette on the right. Residuals above zero represent populations that are more closely related to each other in the data than in the fitted tree, and thus are candidates for introgression events. The fitted tree accounts for 87.43% of the variance in relatedness among populations in A, and 98.50% in C. B and D. (Left panel) The maximum likelihood tree with 5 and 8 deduced introgression events and the residual matrix of the improved model. The introgression events are highlighted as the arrows with colors from yellow to red, which represent the various weights of introgression. The arrow direction indicates the introgression direction. (Right panel) The residuals of the improved model are illustrated, where the color scale is the same as that in A. The fraction of the variance in relatedness among populations interpreted by the improved model rises up to 99.96% in B and 99.97% in D.



**Additional file 4: Figure S3.** The selective sweeps derived from multiple statistics and the filtering strategy for the positively selected genes. A. Venn diagram of the identified selective sweeps from five methods including Heterozygosity, Tajima's *D*, Fay and Wu's *H*, integrated haplotype score (*iHS*) and SweepFinder. B. The density of SNPs with selection signals in 1-kb windows that locate in the identified selective sweeps and have 2% bottommost/topmost statistics for each method, reflecting the selection bias against smaller SNP number for Tajima's *D* and Heterozygosity. C. The frequency of selected 1-kb windows supporting selective sweep genes for each method, determining the filter strategy of selective sweep genes as 10 1 kb-windows for Tajima's *D* and Heterozygosity.

**Additional file 5: Figure S4.** The expression pattern across 6 chicken tissues for the 65 selective sweep genes related to neurodevelopment. The gene expression pattern is based on the published RNA-Seq data of RJFs [31]. The gene expression level is calibrated on the logarithmic scale. The tissues include cerebrum, cerebellum, testis, liver, kidney and heart. For each tissue, one male and one female were examined, with the exception of testis where two males were examined.

**Additional file 6: Figure S5.** Clustering analyses of gene expression in cerebrum and cerebellum across chicken populations.

**Additional file 7: Figure S6.** Selection signal and expression profiles of five representative genes associated with behavioral modification during chicken domestication. These genes are both positively selected genes and differentially expressed genes in cerebrum or cerebellum in domesticated chickens, whose functions in domestication-related behavioral alterations have been verified in mutant mouse models. The left panels indicate the statistics across different methods including Heterozygosity, Tajima's *D*, Fay and Wu's *H*, integrated haplotype score (*iHS*) and SweepFinder. The vertical axes represent the statistical values for each method; the horizontal axes represent the genomic coordinates around the target genes. The red dashed line in each method denotes the threshold above/below which signals were considered as positive selection. The right panels illustrate the gene expression levels of cerebrum and cerebellum in wild and domesticated chickens, where the double asterisks represent significant difference with a *P* value < 0.01.

**Additional file 8: Figure S7.** The regulatory network between the 53 positively selected genes (PSGs) and the differentially expressed genes (DEGs) of cerebrum and cerebellum in domesticated chickens compared with their wild counterparts. The molecules filled by orange color indicate the overlapped genes between PSGs and DEGs. The molecules filled by red or green color represent the up- or down-regulated DEGs, respectively, where the shades of filled colors represent the extent of the alteration at gene expression level. The solid lines imply direct interactions between molecules. The ovals filled by yellow color at outer layer indicate the canonical pathways related to neurological functions, in which, both PSGs and DEGs are known to participate.

**Additional file 9: Figure S8.** Selective signals on genes in convergent family of glutamate ionotropic receptors, semaphorins and tectonic proteins. The species include chicken (galGal), dog (canFam) and pig (susScr). The methods encompass Heterozygosity, Tajima's *D*, and Fay and Wu's *H*, integrated haplotype score (*iHS*) and SweepFinder. The vertical axes represent the statistical values for each method; the horizontal axes represent the genomic coordinates around the target genes. The red dashed line in each method denotes the threshold, above/below which signals were considered as positive selection. The red bars in genes indicate the exon regions of gene.

**Additional file 10: Figure S9.** Expression level of genes in convergent family of glutamate ionotropic receptors, semaphorins and tectonic proteins in paired wild and domesticated animals. The species include chicken (galGal), dog (canFam) and pig (susScr). Brain tissues include cerebrum and cerebellum for chicken, frontal cortex for pig, and cerebellum and hypothalamus for dog. The vertical axes represent the gene expression levels; the horizontal axes indicate the genes in families of glutamate ionotropic receptors, semaphorins and tectonic proteins.

**Additional file 11: Figure S10.** The comparison of nonsynonymous versus synonymous mutations ( $d_N/d_S$  ratio) between wild and domestic breeds across chicken (A), pig (B) and dog (C). Chicken population

includes XH (Guangdong Xinghua), LX (Luxi Dou), YOU (Beijing You), SILK (Silkie), RW (Recessive White Rock), WL (White Leghorn) and RJF (Red jungle fowls). One-way ANOVA is used for the significance test.

**Additional file 12: Figure S11.** The comparison of frequency of deleterious variants between wild and domestic breeds across pig, dog and chicken.

**Additional file 13: Figure S12.** Conservation degree of the commonly targeted genes of selection. The cumulative distributions of phastCons and phyloP scores for the commonly targeted genes of selection (red) in comparison with other background genes (blue). Wilcoxon rank-sum test was implemented to check the significance of difference.

**Additional file 14: Figure S13.** Comparison of the commonly selected genes in domesticated species with the risk factors of enriched human neurological diseases and behavioral disorders annotated in the PheGen project. The commonly selected genes, were defined as those that showed signatures of selection in at least 2 domesticated species and were members in the common gene families and pathways that were shared by at least two species. The enriched neurological diseases and behavioral disorders in humans were anxiety, aggression, attention deficit hyperactivity, schizophrenia, depression and loneliness. The risk factors of these diseases were obtained from the PheGen project. The conserved genes between the commonly selected genes during domestication and the risk factors of human neurological diseases and behavioral disorders were marked by circles and filled with different colors for different species. Red color indicates genes selected in at least 2 species. Besides of the conserved genes, other commonly selected genes may provide more risk factors (grey color) that may serve as novel candidates for these diseases, which were potentially related to the overrepresented neuropsychiatric diseases through functional interaction with the existed risk factors in the same pathways.

**Additional file 15: Figure S14.** The rank of degree of the PSGs and the randomly selected genes from dog and pig gene-gene interaction network. The orange line is the rank of degree of the PSGs, calculated from their subnetwork and sorted by increasing x-coordinate. The blue line is the median rank of degree of randomly selected genes (Equal number with PSGs) in their subnetwork after 1000 times sampling.

**Additional file 16: Figure S15.** Enrichment of the putatively adaptive mutations in different genomic regions across pig, dog and chicken. The putatively adaptive mutations in response to domestication were defined as SNPs within selective sweeps with  $iHS < -2$  in domesticated animals. Genomic regions of exon, intron, upstream, downstream, UTR3, UTR5, splicing, and intergenic regions were considered. Upstream and downstream regions were defined as 1 kb upstream and downstream the transcription start and end site, respectively. The pigs in Asia and Europe were independently analyzed to account for their distinct origins. The significantly enriched genomic regions were circumscribed. Double asterisks represent significant difference with a *P* value < 0.01.

## Abbreviations

CLR: Composite likelihood ratio; DEG: Differentially expressed gene; FDR: False discovery rate; GERP: Genomic evolutionary rate profiling;  $H_p$ : Heterozygosity; IPA: Ingenuity pathway analysis; KEGG: Kyoto encyclopedia of genes and genomes; LD: Linkage disequilibrium; LX: Luxi Dou chicken; PCA: Principle component analysis; PheGen: Phenotype-Genotype Integrator; PSG: Positively selected genes; PSS: Putative selective sweeps; QTL: Quantitative trait loci; RJF: Red jungle fowl; RW: Recessive white rock chicken; SE: Standard error; SFS: Site frequency spectrum; SILK: Silkie chicken; WGS: Whole-genome sequencing; WL: White Leghorn chicken; XH: Xinghua chicken; YOU: Beijing You chicken

## Acknowledgments

We acknowledge Yu Ying in College of Animal Science and Technology, Chinese Agricultural University for the guidance of brain tissue dissection of chickens. We also acknowledge Zhao Guiping in Institute of Animal Science, Chinese Academy of Agricultural Sciences, Chinese game chicken breeding farm and Beijing Huadu Yukou poultry industry Co., Ltd. for kindly donation of part of chicken samples for the purpose of scientific research.

**Authors' contributions**

XL, YH and XZ conceived and designed the experiments. XS, TR and QC performed blood collection, tissue sample dissection, DNA and RNA extraction, and library construction. YH, FQ, and XB performed the data analysis. YH, XL, FQ, XB and XZ wrote the manuscript. QC took part in the interpretation of the results. XZ, YH, XL and QC provided and collected the samples. All authors have read and approved the manuscript.

**Funding**

This study was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB13040300), National Key R&D Program of China (2018YFC0910402), National Natural Science Foundation of China (31301963), the China Agriculture Research System (CARS-41-G03), Youth Innovation Promotion Association, Chinese Academy of Sciences.

**Availability of data and materials**

The 62 genome sequencing data and 17 transcriptome data in chicken have been deposited in the archive of Beijing Institute of Genomics (<https://bigd.big.ac.cn/gsa/>) under the accession number CRA000005.

**Ethics approval and consent to participate**

The study was approved by the Animal Experimental ethics board in Beijing Institute of Genomics, Chinese Academy of Science, in South China Agricultural University, and in Institute of Animal Husbandry and Veterinary Medicine, Beijing Academy of Agriculture and Forestry Sciences. The informed consents were written. Additionally, Institute of Animal Science, Chinese Academy of Agricultural Sciences, Chinese game chicken breeding farm in Shandong province, and Beijing Huadu Yukou poultry industry Co., Ltd. provided informed consents for donation of part of chicken samples for the purpose of scientific research.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

<sup>1</sup>Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, People's Republic of China. <sup>2</sup>China National Center for Bioinformation, Beijing, People's Republic of China. <sup>3</sup>University of Chinese Academy of Sciences, Beijing, People's Republic of China. <sup>4</sup>Institute of Animal Husbandry and Veterinary Medicine, Beijing Academy of Agriculture and Forestry Sciences, Beijing, People's Republic of China. <sup>5</sup>Guangdong Provincial Key Lab of Agro-Animal Genomics and Molecular Breeding, and Key Lab of Chicken Genetics, Breeding and Reproduction, Ministry of Agriculture, South China Agricultural University, Guangzhou, People's Republic of China. <sup>6</sup>Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, People's Republic of China.

Received: 25 May 2019 Accepted: 24 February 2020

Published online: 04 March 2020

**References**

- Darwin CR. The variation of animals and plants under domestication. First edit. London: John Murray; 1868.
- Jensen P. Behavior genetics and the domestication of animals. *Ann Rev Anim Biosci.* 2014;2:85–104.
- Belyaev D. Destabilizing selection as a factor in domestication. *J Hered.* 1979;70:301–8.
- Albert FW, Shchepina O, Winter C, Römpler H, Teupser D, Palme R, et al. Phenotypic differences in behavior, physiology and neurochemistry between rats selected for tameness and for defensive aggression towards humans. *Horm Behav.* 2008;53:413–21.
- Campler M, Jöngren M, Jensen P. Fearfulness in red junglefowl and domesticated white Leghorn chickens. *Behav Processes.* 2009;81:39–43.
- Schütz KE, Forkman B, Jensen P. Domestication effects on foraging strategy, social behaviour and different fear responses: a comparison between the red junglefowl (*Gallus gallus*) and a modern layer strain. *Appl Anim Behav Sci.* 2001;74:1–14.
- Wiener P, Wilkinson S. Deciphering the genetic basis of animal domestication. *Proc Biol Sci.* 2011;278:3161–70.
- Kukekova A, Trut L, Chase K, Kharamova A, Johnson J, Temnykh S, et al. Mapping loci for fox domestication: deconstruction/reconstruction of a behavioral phenotype. *Behav Genet.* 2011;41:593–606.
- Footo AD, Liu Y, Thomas GWC, Vinař T, Alföldi J, Deng J, et al. Convergent evolution of the genomes of marine mammals. *Nat Genet.* 2015;47:272–5.
- Bridgham JT. Predicting the basis of convergent evolution. *Science.* 2016;354:289.
- Kelley JL, Brown AP, Therkildsen NO, Footo AD. The life aquatic: advances in marine vertebrate genomics. *Nat Rev Genet.* 2016;17:523–34.
- Natarajan C, Hoffmann FG, Weber RE, Fago A, Witt CC, Storz JF. Predictable convergence in hemoglobin function has unpredictable molecular underpinnings. *Science.* 2016;354:336–9.
- Stern DL. The genetic causes of convergent evolution. *Nat Rev Genet.* 2013;14:751–64.
- Axelsson E, Ratnakumar A, Arendt ML, Maqbool K, Webster MT, Perloski M, et al. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature.* 2013;495:360–4.
- Wang G, Zhai W, Yang H, Fan R, Cao X, Zhong L, et al. The genomics of selection in dogs and the parallel evolution between dogs and humans. *Nat Commun.* 2013;4:1860.
- Schubert M, Jónsson H, Chang D, Der Sarkissian C, Ermini L, Ginolhac A, et al. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc Natl Acad Sci.* 2014;111:E5661–9.
- Li M, Tian S, Yeung CKL, Meng X, Tang Q, Niu L, et al. Whole-genome sequencing of Berkshire (European native pig) provides insights into its origin and domestication. *Sci Rep.* 2014;4:4678.
- Carneiro M, Rubin CJ, Di Palma F, Albert FW, Alföldi J, Martínez Barrio A, et al. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science.* 2014;345:1074–9.
- Frantz LAF, Schraiber JG, Madsen O, Megens H, Cagan A, Bosse M, et al. Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nat Genet.* 2015;47:1141–8.
- Qanbari S, Pausch H, Jansen S, Somel M, Strom TM, Fries R, et al. Classic selective sweeps revealed by massive sequencing in cattle. *PLoS Genet.* 2014;10:e1004148.
- Montague MJ, Li G, Gandolfi B, Khan R, Aken BL, Searle SMJ, et al. Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. *Proc Natl Acad Sci U S A.* 2014;111:17230–5.
- Wright D. The genetic architecture of domestication in animals. *Bioinform Biol Insights.* 2015;9(Suppl 4):11–20.
- Rubin C-J, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT, et al. Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature.* 2010;464:587–91.
- Wang M, Li Y, Peng M, Zhong L, Li Q, Tu X, et al. Genomic analyses reveal potential independent adaptation to high altitude in Tibetan chickens. *Mol Biol Evol.* 2015;32:1–31.
- Crawford RD. Poultry breeding and genetics Amsterdam. New York: Elsevier; 1990.
- Dohner JV. The encyclopedia of historic and endangered livestock and poultry breeds. First edit. New Haven and London: Yale University Press; 2001.
- Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19:1655–64.
- Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics.* 2011;12:246.
- Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 2012;8:e1002967.
- Reich D, Thangaraj K, Patterson N, Price AL, Singh L. Reconstructing Indian population history. *Nature.* 2009;461:489–94.
- Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, et al. The evolution of gene expression levels in mammalian organs. *Nature.* 2011;478:343–8.
- Callaway E. When chickens go wild. *Nature.* 2016;529:270–3.
- Costa RM, Yang T, Huynh DP, Pulst SM, Viskochil DH, Silva AJ, et al. Learning deficits, but normal development and tumor predisposition, in mice lacking exon 23a of Nf1. *Nat Genet.* 2001;27:399–405.
- Ko S, Zhao M-G, Toyoda H, Qiu C-S, Zhuo M. Altered behavioral responses to noxious stimuli and fear in glutamate receptor 5 (GluR5)- or GluR6-deficient mice. *J Neurosci.* 2005;25:977–84.

35. Kaeser PS, Deng L, Chávez AE, Liu X, Castillo PE, Südhof TC. ELKS2a/CAST deletion selectively increases neurotransmitter release at inhibitory synapses. *Neuron*. 2009;64:227–39.
36. Riccio A, Li Y, Moon J, Kim K-S, Smith KS, Rudolph U, et al. Essential role for TRPC5 in amygdala function and fear-related behavior. *Cell*. 2009;137:761–72.
37. Cafferty WBJ, Duffy P, Huebner E, Strittmatter SM. MAG and OMgp synergize with Nogo- $\alpha$  to restrict axonal growth and neurological recovery after spinal cord trauma. *J Neurosci*. 2010;30:6825–37.
38. Rogers JT, Rusiana I, Trotter J, Zhao L, Donaldson E, Pak DTS, et al. Reelin supplementation enhances cognitive ability, synaptic plasticity, and dendritic spine density. *Learning Memory (Cold Spring Harbor, NY)*. 2011;18:558–64.
39. Drgonova J, Walther D, Singhal S, Johnson K, Kessler B, Troncoso J, et al. Altered CSMD1 expression alters cocaine-conditioned place preference: mutual support for a complex locus from human and mouse models. *PLoS One*. 2015;10:e0120908.
40. Yang J, Li M-X, Luo Y, Chen T, Liu J, Fang P, et al. Chronic ceftriaxone treatment rescues hippocampal memory deficit in AQP4 knockout mice via activation of GLT-1. *Neuropharmacology*. 2013;75:213–22.
41. Gompel N, Prud'homme B. The causes of repeated genetic evolution. *Dev Biol*. 2009;332:36–47.
42. Wang C, Li J, Meng Q, Wang B. Three Tctn proteins are functionally conserved in the regulation of neural tube patterning and Gli3 processing but not ciliogenesis and hedgehog signaling in the mouse. *Dev Biol*. 2017;430:156–65.
43. Long K, Mao K, Che T, Zhang J, Qiu W, Wang Y, et al. Transcriptome differences in frontal cortex between wild boar and domesticated pig. *Anim Sci J*. 2018;89:848–57.
44. Tenaillon O, Rodriguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD, et al. The molecular diversity of adaptive convergence. *Science*. 2012;335:457–61.
45. Stahl SM, Stahl SM. *Stahl's essential psychopharmacology: neuroscientific basis and practical applications*. UK: Cambridge University Press; 2013.
46. Lu J, Tang T, Tang H, Huang J, Shi S, Wu C-I. The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. *Trends Genet*. 2006;22:126–31.
47. Renault S, Rieseberg LH. The accumulation of deleterious mutations as a consequence of domestication and improvement in sunflowers and other Compositae crops. *Mol Biol Evol*. 2015;32:2273–83.
48. Cruz F, Vilà C, Webster MT. The legacy of domestication: accumulation of deleterious mutations in the dog genome. *Mol Biol Evol*. 2008;25:2331–6.
49. Marsden CD, Ortega-Del Vecchyo D, O'Brien DP, Taylor JF, Ramirez O, Vilà C, et al. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proc Natl Acad Sci*. 2016;113:152–7.
50. Freedman AH, Lohmueller KE, Wayne RK. Evolutionary history, selective sweeps, and deleterious variation in the dog. *Annu Rev Ecol Syst*. 2016;47:73–96.
51. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol*. 2010;6:e1001025. <https://doi.org/10.1371/journal.pcbi.1001025>.
52. Henn BM, Botigué LR, Bustamante CD, Clark AG, Gravel S. Estimating the mutation load in human genomes. *Nat Rev Genet*. 2015;16:333–43.
53. Wang HY, Chien HC, Osada N, Hashimoto K, Sugano S, Gojobori T, et al. Rate of evolution in brain-expressed genes in humans and other primates. *PLoS Biol*. 2007;5:e13.
54. Vaz RL, Outeiro TF, Ferreira JJ. Zebrafish as an animal model for drug discovery in Parkinson's disease and other movement disorders: a systematic review. *Front Neurol*. 2018;9:347.
55. Shpigler HY, Saul MC, Corona F, Block L, Cash Ahmed A, Zhao SD, et al. Deep evolutionary conservation of autism-related genes. *Proc Natl Acad Sci*. 2017;114:9653–8.
56. Ramos EM, Hoffman D, Junkins HA, Maglott D, Phan L, Sherry ST, et al. Phenotype-genotype integrator (PheGenI): synthesizing genome-wide association study (GWAS) data with existing genomic resources. *Eur J Human Genet*. 2014;22:144–7.
57. Roth G. Convergent evolution of complex brains and high intelligence. *Philos Trans R Soc Lond B Biol Sci*. 2015;370:20150049.
58. Shang Y, Xiao G, Zheng P, Cen K, Zhan S, Wang C. Divergent and convergent evolution of fungal pathogenicity. *Genome Biol Evol*. 2016;8:1374–87.
59. Storz JF. Causes of molecular convergence and parallelism in protein evolution. *Nat Rev Genet*. 2016;17:239–50.
60. Zhang Z, Xu D, Wang L, Hao J, Wang J, Zhou X, et al. Convergent evolution of rumen microbiomes in high-altitude mammals. *Curr Biol*. 2016;26:1873–9.
61. Anderson TM, VonHoldt BM, Candille SI, Musiani M, Greco C, Stahler DR, et al. Molecular and evolutionary history of melanism in north American gray wolves. *Science*. 2009;323:1339–43.
62. Maselli V, Polese G, Larson G, Raia P, Forte N, Ripa D, et al. A dysfunctional sense of smell: the irreversibility of olfactory evolution in free-living pigs. *Evol Biol*. 2014;41:229–39.
63. Der Sarkissian C, Ermini L, Schubert M, Yang MA, Librado P, Fumagalli M, et al. Evolutionary genomics and conservation of the endangered Przewalski's horse. *Curr Biol*. 2015;25:2577–83.
64. Ellstrand NC, Prentice HC, Hancock JF. Gene flow and introgression from domesticated plants into their wild relatives. *Annu Rev Ecol Syst*. 1999;30:539–63.
65. Winkler EA, Nishida Y, Sagare AP, Rege SV, Bell RD, Perlmutter D, et al. GLUT1 reductions exacerbate Alzheimer's disease vasculoneuronal dysfunction and degeneration. *Nat Neurosci*. 2015;18:521–30.
66. Henriksen R, Johnsson M, Andersson L, Jensen P, Wright D. The domesticated brain: genetics of brain mass and brain structure in an avian species. *Sci Rep*. 2016;6:34031.
67. Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, et al. Avian genome evolution and adaptation. *Science (New York, NY)*. 2014;346:1311–21.
68. Feigin CY, Newton AH, Doronina L, Schmitz J, Hipsley CA, Mitchell KJ, et al. Genome of the Tasmanian tiger provides insights into the evolution and demography of an extinct marsupial carnivore. *Nat Ecol Evol*. 2018;2:182–92.
69. Parenteau J, Maignon L, Berthoumieux M, Catala M, Gagnon V, Abou ES. Introns are mediators of cell response to starvation. *Nature*. 2019;565:612–7.
70. Bannerman DM, Sprengel R, Sanderson DJ, McHugh SB, Rawlins JNP, Monyer H, et al. Hippocampal synaptic plasticity, spatial memory and anxiety. *Nat Rev Neurosci*. 2014;15:181–92.
71. Theofanopoulou C, Gastaldon S, O'Rourke T, Samuels BD, Martins PT, Delogu F, et al. Self-domestication in *Homo sapiens*: insights from comparative genomics. *PLoS One*. 2017;12:e0185306.
72. Pasterkamp RJ. Getting neural circuits into shape with semaphorins. *Nat Rev Neurosci*. 2012;13:605–18.
73. Matsuda I, Shoji H, Yamasaki N, Miyakawa T, Aiba A. Comprehensive behavioral phenotyping of a new Semaphorin 3 F mutant mouse. *Mol Brain*. 2016;9:15.
74. Håkansson K, Runker AE, O'Sullivan GJ, Mitchell KJ, Waddington JL, O'Tuathaigh CM. Semaphorin 6A knockout mice display abnormalities across ethologically-based topographies of exploration and in motor learning. *Neurosci Lett*. 2017;641:70–6.
75. Kumar SM, Dai J, Li S, Yang R, Yu H, Nathanson KL, et al. Human skin neural crest progenitor cells are susceptible to BRAF(V600E)-induced transformation. *Oncogene*. 2014;33:832–41.
76. Gaudet P, Livstone MS, Lewis SE, Thomas PD. Phylogenetic-based propagation of functional annotations within the gene ontology consortium. *Brief Bioinform*. 2011;12:449–62.
77. Wilkins AS, Wrangham RW, Fitch WT. The “domestication syndrome” in mammals: a unified explanation based on neural crest cell behavior and genetics. *Genetics*. 2014;197:795–808.
78. Schafe GE, Nader K, Blair HT, LeDoux JE. Memory consolidation of Pavlovian fear conditioning: a cellular and molecular perspective. *Trends Neurosci*. 2001;24:540–6.
79. Sheng M, Kim MJ. Postsynaptic signaling and plasticity mechanisms. *Science*. 2002;298:776–80.
80. Sweatt JD. Mitogen-activated protein kinases in synaptic plasticity and memory. *Curr Opin Neurobiol*. 2004;14:311–7.
81. Forrest MP, Parnell E, Penzes P. Dendritic structural plasticity and neuropsychiatric disease. *Nat Rev Neurosci*. 2018;19:215–34.
82. Zeder MA. The domestication of animals. *J Anthropol Res*. 2012;68:161–90.
83. Marshall-Pescini S, Cafazzo S, Virányi Z, Range F. Integrating social ecology in explanations of wolf–dog behavioral differences. *Curr Opin Behav Sci*. 2017;16:80–6.
84. Hill W, Robertson A. The effect of linkage on limits to artificial selection. *Genet Res*. 1966;8:269–94.
85. Orr HA. Fitness and its role in evolutionary genetics. *Nat Rev Genet*. 2009;10:531–9.
86. Fan W, Ng CS, Chen C, Lu MJ, Chen Y, Liu C, et al. Genome-wide patterns of genetic variation in two domestic chickens. *Genome Biol Evol*. 2013;5:1376–92.

87. Wang M-S, Zhang R, Su L-Y, Li Y, Peng M-S, Liu H-Q, et al. Positive selection rather than relaxation of functional constraint drives the evolution of vision during chicken domestication. *Cell Res.* 2016;26:556–73.
88. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics.* 2009;25:1754–60.
89. Nielsen R, Paul JS, Albrechtsen A, Song YS. Genotype and SNP calling from next-generation sequencing data. *Nat Rev Genet.* 2011;12:443–51.
90. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25:2078–9.
91. Felsenstein J. PHYLIP (phylogeny inference package) version 3.6; 2005.
92. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–75.
93. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet.* 2006;2:e190.
94. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000;155(2):945–59.
95. Wang G-D, Zhai W, Yang H-C, Wang L, Zhong L, Liu Y-H, et al. Out of southern East Asia: the natural history of domestic dogs across the world. *Cell Res.* 2015;26:21.
96. He Z, Zhai W, Wen H, Tang T, Wang Y, Lu X, et al. Two evolutionary histories in the genome of rice: the roles of domestication genes. *PLoS Genet.* 2011; 7:e1002100.
97. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics.* 1989;123(3):585–95.
98. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics.* 2011;27:2156–8.
99. Fay JC, Wu C. Hitchhiking under positive darwinian selection. *Genetics.* 2000;155(3):1405–13.
100. Vilella AJ, Blanco-Garcia A, Hutter S, Rozas J. VariScan: analysis of evolutionary patterns from large-scale DNA sequence polymorphism data. *Bioinformatics.* 2005;21:2791–3.
101. Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data : applications to inferring missing genotypes and Haplotypic phase. *Am J Hum Genet.* 2006;78:629–44.
102. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol.* 2006;4:e72.
103. Pickrell JK, Coop G, Novembre J, Kudaravalli S, Li JZ, Absher D, et al. Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.* 2009;19:826–37.
104. Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. Genomic scans for selective sweeps using SNP data. *Genome Res.* 2005;15:1566–75.
105. Huber CD, DeGiorgio M, Hellmann I, Nielsen R. Detecting recent selective sweeps while controlling for mutation rate and background selection. *Mol Ecol.* 2016;25:142–56.
106. Qi F, Yang A, Ambreen S, Bai X, Hou Y, Lu X. Birth and death of mx genes and the presence/absence of genes regulating mx transcription are correlated with the diversity of anti-pathogenicity in vertebrate species. *Mol Genet Genomics.* 2019;294:121–33. <https://doi.org/10.1007/s00438-018-1490-x>.
107. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics.* 2009;25:1105–11.
108. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol.* 2010;11:R106.
109. Anders S, Pyl PT, Huber W. HTSeq - a python framework to work with high-throughput sequencing data. *Bioinformatics.* 2014;31:166–9.
110. Robinson PN, Webber C. Phenotype ontologies and cross-species analysis for translational research. *PLoS Genet.* 2014;10:e1004268.
111. Bandyopadhyay S, Sharan R, Ideker T. Systematic identification of functional orthologs based on protein network comparison. *Genome Res.* 2006;16: 428–35.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

