Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication

Michael J. Montague[a,1], Gang Li[b,1], Barbara Gandolfi[c], Razib Khan[d], Bronwen L. Aken[e], Steven M. J. Searle[e], Patrick Minx[a], LaDeana W. Hillier[a], Daniel C. Koboldt[a], Brian W. Davis[b], Carlos A. Driscoll[f], Christina S. Barr[f], Kevin Blackistone[f], Javier Quilez[g], Belen Lorente-Galdos[g], Tomas Marques-Bonet[g,h], Can Alkan[i], Gregg W. C. Thomas[j], Matthew W. Hahn[j], Marilyn Menotti-Raymond[k], Stephen J. O'Brien[l,m], Richard K. Wilson[a], Leslie A. Lyons[c,2], William J. Murphy[b,2], and Wesley C. Warren[a,2]

[a]The Genome Institute, Washington University School of Medicine, St. Louis, MO 63108; [b]Department of Veterinary Integrative Biosciences, College of Veterinary Medicine, Texas A&M University, College Station, TX 77843; [c]Department of Veterinary Medicine & Surgery, College of Veterinary Medicine, University of Missouri, Columbia, MO 65201; [d]Population Health & Reproduction, School of Veterinary Medicine, University of California, Davis, CA 95616; [e]Wellcome Trust Sanger Institute, Hinxton CB10 1SA, United Kingdom; [f]National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda, MD 20886; [g]Catalan Institution for Research and Advanced Studies, Institute of Evolutionary Biology, Pompeu Fabra University, 08003 Barcelona, Spain; [h]Centro de Analisis Genomico 08028, Barcelona, Spain; [i]Department of Computer Engineering, Bilkent University, Ankara 06800, Turkey; [j]Department of Biology, Indiana University, Bloomington, IN 47405; [k]Laboratory of Genomic Diversity, Center for Cancer Research, Frederick, MD 21702; [l]Dobzhansky Center for Genome Bioinformatics, St. Petersburg State University, St. Petersburg 199178, Russia; and [m]Oceanographic Center, Nova Southeastern University, Fort Lauderdale, FL 33314

Edited by James E. Womack, Texas A&M University, College Station, TX, and approved October 3, 2014 (received for review June 2, 2014)

Little is known about the genetic changes that distinguish domestic cat populations from their wild progenitors. Here we describe a high-quality domestic cat reference genome assembly and comparative inferences made with other cat breeds, wildcats, and other mammals. Based upon these comparisons, we identified positively selected genes enriched for genes involved in lipid metabolism that underpin adaptations to a hypercarnivorous diet. We also found positive selection signals within genes underlying sensory processes, especially those affecting vision and hearing in the carnivore lineage. We observed an evolutionary tradeoff between functional olfactory and vomeronasal receptor gene repertoires in the cat and dog genomes, with an expansion of the feline chemosensory system for detecting pheromones at the expense of odorant detection. Genomic regions harboring signatures of natural selection that distinguish domestic cats from their wild congeners are enriched in neural crest-related genes associated with behavior and reward in mouse models, as predicted by the domestication syndrome hypothesis. Our description of a previously unidentified allele for the gloving pigmentation pattern found in the Birman breed supports the hypothesis that cat breeds experienced strong selection on specific mutations drawn from random bred populations. Collectively, these findings provide insight into how the process of domestication altered the ancestral wildcat genome and build a resource for future disease mapping and phylogenomic studies across all members of the Felidae.

Felis catus | domestication | genome

The domestic cat (*Felis silvestris catus*) is a popular pet species, with as many as 600 million individuals worldwide (1). Cats and other members of Carnivora last shared a common ancestor with humans ~92 million years ago (2, 3). The cat family Felidae includes ~38 species that are widely distributed across the world, inhabiting diverse ecological niches that have resulted in divergent morphological and behavioral adaptations (4). The earliest archaeological evidence for human coexistence with cats dates to ~9.5 kya in Cyprus and ~5 kya in central China (5, 6), during periods when human populations adopted more agricultural lifestyles. Given their sustained beneficial role surrounding vermin control since the human transition to agriculture, any selective forces acting on cats may have been minimal subsequent to their domestication. Unlike many other domesticated mammals bred for food, herding, hunting, or security, most of the 30–40 cat breeds originated recently, within the past 150 y, largely due to selection for aesthetic rather than functional traits.

Previous studies have assessed breed differentiation (6, 7), phylogenetic origins of the domestic cat (8), and the extent of recent introgression between domestic cats and wildcats (9, 10). However, little is known regarding the impact of the domestication process within the genomes of modern cats and how this compares with genetic changes accompanying selection identified in other domesticated companion animal species. Here we describe, to our knowledge, the first high-quality annotation of the complete

Significance

We present highlights of the first complete domestic cat reference genome, to our knowledge. We provide evolutionary assessments of the feline protein-coding genome, population genetic discoveries surrounding domestication, and a resource of domestic cat genetic variants. These analyses span broadly, from carnivore adaptations for hunting behavior to comparative odorant and chemical detection abilities between cats and dogs. We describe how segregating genetic variation in pigmentation phenotypes has reached fixation within a single breed, and also highlight the genomic differences between domestic cats and wildcats. Specifically, the signatures of selection in the domestic cat genome are linked to genes associated with gene knockout models affecting memory, fear-conditioning behavior, and stimulus-reward learning, and potentially point to the processes by which cats became domesticated.

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

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. GU270865.1, KJ923925–KJ924979, SRX026946, SRX026943, SRX026929, SRX027004, SRX026944, SRX026941, SRX026909, SRX026901, SRX026955, SRX026947, SRX026911, SRX026910, SRX026948, SRX026928, SRX026912, SRX026942, SRX026930, SRX026913, SRX019549, SRX019524, SRX026956, SRX026945, and SRX026960).

[1]M.J.M. and G.L. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: wwarren@genome.wustl.edu, wmurphy@cvm.tamu.edu, or lyonsla@missouri.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1410083111/-/DCSupplemental.

domestic cat genome and a comparative genomic analysis including whole-genome sequences from other felids and mammals to identify the molecular footprints of the domestication process within cats.

Results and Discussion

To identify molecular signatures underlying felid phenotypic innovations, we developed a higher-quality reference assembly for the domestic cat genome using whole-genome shotgun sequences (*Materials and Methods* and *SI Materials and Methods*). The assembly (FelCat5) comprises 2.35 gigabases (Gb) assigned to all 18 autosomes and the X chromosome relying on physical and linkage maps (11) with a further 11 megabases (Mb) in unplaced scaffolds. The assembly is represented by an N50 contig length of 20.6 kb and a scaffold N50 of 4.7 Mb, both of which show substantial improvement over previous light-coverage genome survey sequences that included only 60% of the genome (12, 13). The *Felis catus* genome is predicted to contain 19,493 protein-coding genes and 1,855 noncoding RNAs, similar to dog (14). Hundreds of feline traits and disease pathologies (15) offer novel opportunities to explore the genetic basis of simple and complex traits, host susceptibility to infectious diseases, as well as the distinctive genetic changes accompanying the evolution of carnivorans from other mammals.

To identify signatures of natural selection along the lineages leading to the domestic cat, we identified rates of evolution using genome-wide analyses of the ratio of divergence at nonsynonymous and synonymous sites (d_N/d_S) (16) (*Materials and Methods* and *SI Materials and Methods*). We used the annotated gene set (19,493 protein-coding genes) to compare unambiguous mammalian gene orthologs shared between cat, tiger, dog, cow, and human ($n = 10{,}317$). Two-branch and branch-site models (17) collectively identified 467, 331, and 281 genes that were putatively shaped by positive selection in the carnivore, felid, and domestic cat (subfamily Felinae) ancestral lineages, respectively (S1.1–S1.3 in Dataset S1). We assessed the potential impact of amino acid changes using TreeSAAP (18) and PROVEAN (19). The majority of identified genes possess substitutions with significant predicted structural or biochemical effects based on one or both tests (Fig. S1 and S1.4 in Dataset S1). Although the inferences produced by our methods call for additional functional analyses, we highlight several positively selected genes to illustrate their importance to carnivore and feline biology.

Carnivores are endowed with extremely acute sensory adaptations, allowing them to effectively locate potential prey before being discovered (20). Within carnivores, cats have the broadest hearing range, allowing them to detect both ultrasonic communication by prey as well as their movement (21). We identified six positively selected genes (Fig. 1) that conceivably evolved to increase auditory acuity over a wider range of frequencies in the carnivore ancestor and within Felidae, as mutations within each gene have been associated with autosomal, nonsyndromic deafness or hearing loss (22, 23). Visual acuity is adaptive for hunting and catching prey, especially for crepuscular predators such as the cat and other carnivores. Accordingly, we identified elevated d_N/d_S values for 20 carnivoran genes that, when mutated in humans, have well-described roles in a spectrum of visual pathologies (Fig. 1). For example, truncating mutations in human *CHM* cause the progressive disease choroideremia (24), beginning with a loss of night vision and peripheral vision and later a loss of central vision. Many carnivores have excellent night vision (20, 25), and we postulate that the acquisition of selectively advantageous amino acid substitutions within several genes increased visual acuity under low-light conditions. In one interesting dual-role example, *MYO7A* encodes a protein involved in the maintenance of both auditory and visual systems that, when mutated, results in loss of hearing and vision (26).

Cats differ from most other carnivores as a result of being obligately carnivorous. One outcome of this adaptive process is that cats are unable to synthesize certain essential fatty acids, specifically arachidonic acid, due to low Delta-6-desaturase activity (27). This has led to suggestions that cats use an alternate (yet unknown) pathway to generate this essential fatty acid for normal health and reproduction. Furthermore, cats fed a diet rich in saturated and polyunsaturated fatty acids showed no effects on plasma lipid concentrations that in humans are risk factors for coronary heart disease and atherosclerosis (28). These aspects of feline biology are reflected in our positive selection results, where the notable classes of genes overrepresented in the Felinae list

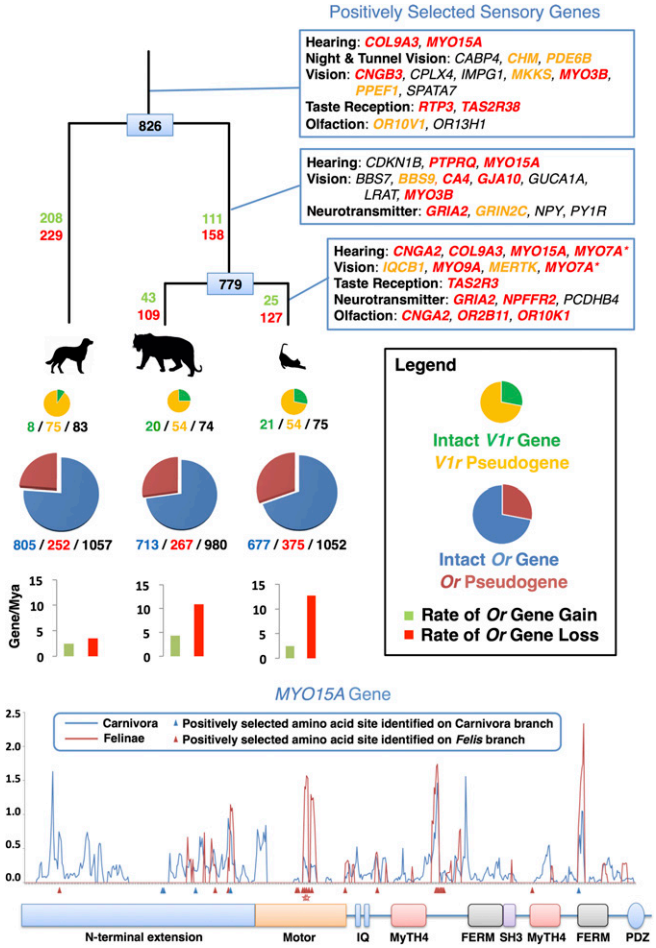

Fig. 1. Dynamic evolution of feline sensory repertoires (*Upper*). The phylogenetic tree depicts relationships scaled to time between dog, tiger, and domestic cat. Positively selected genes are listed (*Top Right*), with lines indicating genes identified on the ancestral branch of Carnivora (*Top*), Felidae (*Middle*), and Felinae (*Bottom*). Genes highlighted in red and orange were identified with significant structural or biochemical effects by two tests or one test, respectively (S1.4 in Dataset S1). *MYO7A* (*) expression is associated with hearing and vision. Numbers at each tree node represent the reconstructed ancestral functional olfactory receptor gene (*Or*) repertoire for carnivores and felids. Numbers labeling each branch are estimated *Or* gene gain (green) and loss (red). The pie charts refer to functional and nonfunctional (pseudogenic) vomeronasal (*V1r*; *Top*) and *Or* (*Bottom*) gene repertoires, with circles drawn in proportion to the size of each gene repertoire. *Or* genes are depicted in blue (functional) and red (nonfunctional), and *V1r* genes are depicted in green (functional) and yellow (nonfunctional). Beneath each pie chart are numbers of functional/nonfunctional/total genes identified in the current genome annotations of the three species. Bar graphs depict rates of *Or* gene gain and loss. Location of signatures of positive selection (*Lower*). Several genes encode members of the myosin gene family of mechanochemical proteins, with *MYO15A* notably under selection in all three branches tested. Curved lines represent the estimated d_N/d_S values (*y* axis) calculated in 90-bp sliding windows (step size of 18 bp) along the length of the gene alignment (*x* axis) for dog, cat, and tiger. Colored boxes indicate known functional domains. Arrowheads indicate the location of positively selected amino acid sites based on the results of the branch-site test. Stars indicate deleterious mutations in the domestic cat (*Materials and Methods*). Motifs and domains include the IQ calmodulin-binding motif (IQ); the myosin tail homology 4 domain (MyTH4); the FERM domain (FERM); the SRC homology 3 domain (SH3); and the PDZ domain (PDZ).

are related to lipid metabolism (S1.5 in Dataset S1). For example, one of the positively selected genes, *ACOX2*, is critical for metabolism of branch-chain fatty acids and has been suggested to regulate triglyceride levels (29), whereas mutations in *PAFAH2* have been associated with risk for coronary heart disease and ischemia (30). The enrichment of genes related to lipid metabolism is likely a signature of adaptation for accommodating the hypercarnivorous diet of felids (31), and mirrors similar signs of selection on lipid metabolic pathways in the genomes of polar bears (32).

Gene duplication and gene loss events often play substantial roles in phenotypic differences between species. To identify protein families that rapidly evolved in the domestic cat, either by contraction or expansion, we examined gene family expansion along an established species tree (33) using tree orthology (34). Two extensive chemosensory gene families, coding for olfactory (*Or*) and vomeronasal (*V1r*) receptors, are responsible for small-molecule detection of odorants and other chemicals for mediating pheromone perception, respectively. Cats rely less on smell to hunt and locate prey in comparison with dogs, which are well-known for their olfactory prowess (35). These observations are confirmed by our analysis of the complete *Or* gene repertoires for cat, tiger, and dog (Fig. 1), illustrating smaller functional repertoires in felids relative to dogs (~700 genes versus >800, respectively). By contrast, the *V1r* gene repertoire is markedly reduced in dogs but expanded in the ancestor of the cat family (8 versus 21 functional genes, respectively), with evidence for species-specific gene loss in different felids (Fig. 1 and Figs. S2 and S3). A growing body of evidence cataloging *Or* gene repertoires in diverse mammals demonstrates common tradeoffs between functional *Or* repertoire size and other sensory systems involved in ecological niche specialization, such as loss of *Or* genes coinciding with gains in trichromatic color vision in primates (36) and chemosensation in platypus (37). These results add further evidence supporting cats' extensive reliance on pheromones for sociochemical communication (38), which is consistent with a genomic tradeoff between functional *Or* and *V1r* repertoires in response to uniquely evolved ecological strategies in the canid and felid lineages (4).

Cats are considered only a semidomesticated species, because many populations are not isolated from wildcats and humans do not control their food supply or breeding (39, 40). We therefore predicted a relatively modest effect of domestication on the cat genome based on recent divergence from and ongoing admixture with wildcats (8–10), a relatively short human cohabitation time compared with dogs (5, 6), and the lack of clear morphological and behavioral differences from wildcats, with docility, gracility, and pigmentation being the exceptions. To identify genomic regions showing signatures of selection influenced by the domestication process, we used whole-genome analyses of cats from different domestic breeds and wildcats (i.e., other *F. silvestris* subspecies) using pooling methods that control for genetic drift (41). Detecting the genomic regions under putative selection during cat domestication can be complicated by random fixation due to genetic drift during the formation of breeds. We mitigated this effect by combining sequence data from a collection of 22 cats (~58× coverage) from six phylogenetically and geographically dispersed domestic breeds (42) before variant detection and performed selection analyses relative to variants detected within a pool of European (*F. silvestris silvestris*) and Near Eastern (*F. silvestris lybica*) wildcats (~7× coverage; Figs. S4 and S5 and S2.1 in Dataset S2).

After stringent filtering of resequencing data, we aligned sequences to the cat reference genome and identified 8,676,486 and 5,190,430 high-quality single-nucleotide variants (SNVs) among domestic breeds and wildcats, respectively, at a total of 10,975,197 sites (Fig. S3). We next identified 130 regions along cat autosomes with either pooled heterozygosity (H_p) 4 SDs below the mean or divergence (F_{ST}) greater than 4 SDs from the mean (Figs. S4 and S6, *SI Materials and Methods*, and S2.2 and S2.3 in Dataset S2). After parsing regions of high confidence displaying both low domestic H_p and high F_{ST}, we found 13 genes underlying five chromosomal regions (Fig. 2, Fig. S4, and S2.4 in Dataset S2). Genes within each of these regions play important roles in neural processes, notably pathways related to synaptic circuitry that influence behavior and contextual clues related to reward.

One putative region of selection along chromosome A1 (chrA1) (Fig. 3) is denoted by a pair of protocadherin genes (*PCDHA1* and *PCDHB4*), which establish and maintain specific neuronal connections and have implications for synaptic specificity, serotonergic innervation of the brain, and fear conditioning (43). *PCDHB4* was also identified in the d_N/d_S analyses. A second region, also on chrA1 (Fig. 3), overlaps with a glutamate receptor gene, *GRIA1*. Glutamate receptors are the predominant excitatory neurotransmitter receptors in the mammalian brain and play an important role in the expression of long-term potentiation and memory formation (44). *GRIA1* knockout mice exhibit defects in stimulus-reward learning, notably those related to food rewards (45). Two additional glutamate receptor genes,

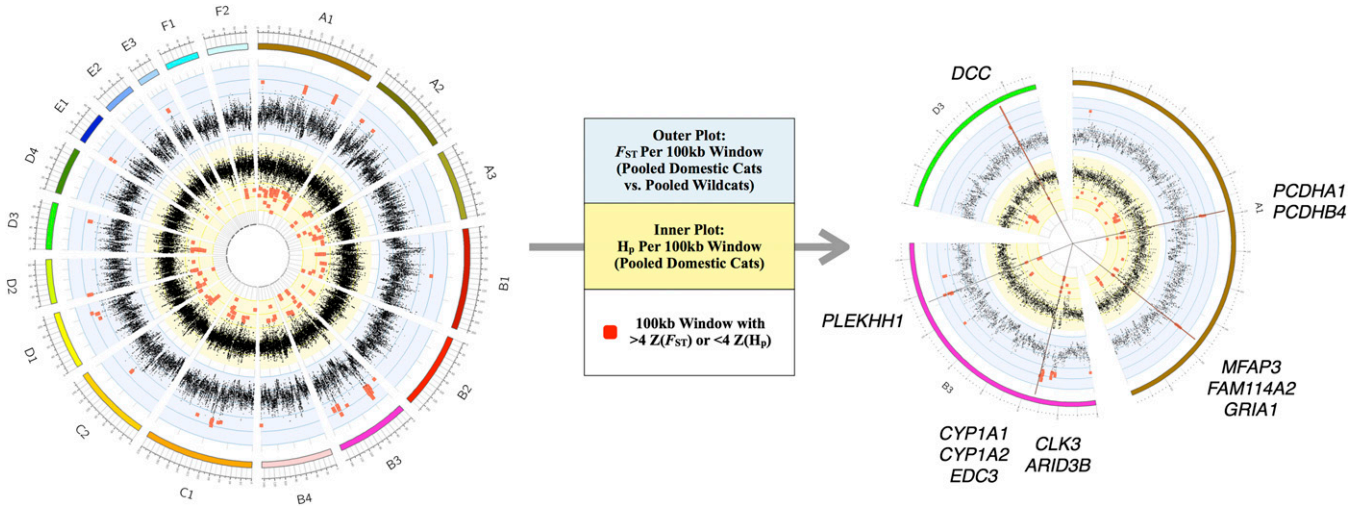

Fig. 2. Sliding window analyses identify five regions of putative selection in the domestic cat genome. Measurements of Z-transformed pooled heterozygosity in cat [inner plot; Z(H_p)] and the Z-transformed fixation index between pooled domestic cat and pooled wildcat [outer plot; Z(F_{ST})] for autosomal 100-kb windows across all 18 autosomes (*Left*). Red points indicate windows that passed the threshold for elevated divergence [>4 Z(F_{ST})] or low diversity [<4 Z(H_P)]. The five regions of putative selection are represented by the straight lines and include contiguous windows that passed both thresholds for elevated divergence and low diversity (*Right*). These regions, across cat autosomes A1, B3, and D3, contain 12 known genes.

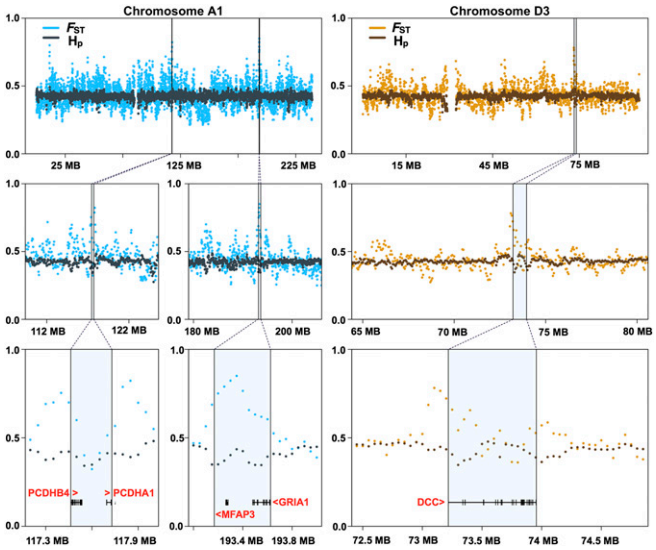

Fig. 3. Comparison between domestic cats and wildcats identifying genes within putative regions of selection in the domestic cat genome that are associated with pathways related to synaptic circuitry and contextual clues related to reward. We identified 130 regions along cat autosomes with either pooled domestic $Z(H_p) < -4$ or $Z(F_{ST}) > 4$, and 5 annotated regions met both criteria. A total of 12 genes was found within these regions, many of which are implicated in neural processes; for instance, genes within regions along chromosomes A1 and D3 are highlighted.

GRIA2 and *NPFFR2*, have elevated d_N/d_S rates within the domestic cat branch of the felid tree (Fig. 1). A third region on chromosome D3 (Fig. 3) encompasses a single gene, *DCC*, encoding the netrin receptor. This gene shows abundant expression in dopaminergic neurons, and behavioral studies of *DCC*-deficient mice show altered dopaminergic system organization, culminating in impaired memory, behavior, and reward responses (46, 47). Two additional regions on chromosome B3 harbor strong signatures of selection (Fig. S7). The first contains three genes, including *ARID3B* (AT rich interactive domain 3B), which plays a critical role in neural crest cell survival (48). The second region contains a single gene, *PLEKHH1*, which encodes a plekstrin homology domain expressed predominantly in human brain. Human genome-wide association studies link variants in *PLEKHH1* with sphingolipid concentrations that, when altered, lead to neurological and psychiatric disease (49).

The genetic signals from this analysis fall in line with the predictions of the domestication syndrome hypothesis (50), which posits that the morphological and physiological traits modified by mammalian domestication are explained by direct and indirect consequences of mild neural crest cell deficits during embryonic development. *ARID3B*, *DCC*, *PLEKHH1*, and protocadherins are all implicated in neural crest cell migration. *ARID3B* is induced in developing mouse embryos during the differentiation of neural crest cells to mature sympathetic ganglia cells (51). *DCC* directly interacts with the Myosin Tail Homology 4 (MyTH4) domain of *MYO10* (myosin X) (52), a gene critical for the migratory ability of neural crest cells. In this way, *DCC* regulates the function of *MYO10* to stimulate the formation and elongation of axons and cranial neural crest cells in developing mouse (53) and frog embryos (54). Like *MYO10*, *PLEKHH1* contains a MyTH4 domain and interacts with the transcription factor *MYC*, a regulator of neural crest cells, to activate transcription of growth-related genes (55). Taken together, we propose that changes in these neural crest-related genes underlie the evolution of tameness during cat domestication, in agreement with analyses of other domesticated genomes (56–58).

We also examined regions of high genetic differentiation between domestic cats and wildcats and observed enrichment in several Wiki and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (S2.5 in Dataset S2), including homologous recombination and axon guidance. Divergence in regions harboring homologous recombination genes (*RAD51B*, *ZFYVE26*, *BRCA2*) may contribute to the high recombination rate reported for domestic cats relative to other mammals (59). Previous studies have suggested that domestication may select for an increase in recombination as a mechanism to generate diversity (60). Specifically, selection for a recombination driver allele may be favored when it is tightly linked to two or more genes with alleles under selection (61). We hypothesize that the close proximity (<350 kb) of two adjacent genes that regulate homologous recombination (*ZFYVE26* and *RAD51B*, which directly interact with *BRCA2*), two visual genes (*RDH11* and *RDH12*) related to retinol metabolism and dark adaptation (62), and one of our candidate domestication genes, *PLEKHH1* (S2.4 in Dataset S2), represents such a case of adaptive linkage.

Aesthetic qualities such as hair color, texture, and pattern strongly differentiate wildcats from domesticated populations and breeds; however, unlike other domesticated species, less than 30–40 genetically distinct breeds exist (63). At the beginning of the cat fancy ~200 y ago, only five different cat "breeds" were recognized, with each being akin to geographical isolates (64). Long hair and the Siamese coloration of "points" were the only diagnostic breed characteristics. Although most breeds were developed recently, following different breeding strategies and selection pressures, much of the color variation in cats developed during domestication, before breed development, and thus is known as "natural" or "ancient" mutations by cat fanciers.

White-spotting phenotypes are a hallmark of domestication, and in cats can range from a complete lack of pigmentation (white) to intermediate bicolor spotting phenotypes (spotting) to white at only the extremities (gloving). For instance, the Birman breed is characterized by point coloration, long hair, and gloving (Fig. 4). A recent study in several white-spotted cats localized the mutation responsible for the spotting pigmentation phenotype within *KIT* intron 1 (65). The KIT gene, located on cat chromosome B1 (66), is primarily involved in melanocyte migration, proliferation, and survival (67). Surprisingly, direct PCR and sequencing excluded the published dominant allele as being associated with the white coloration pattern in Birman (*SI Materials and Methods*). At the same time, whole-genome resequencing data from a pooled sample of Birman cats ($n = 4$; *SI Materials and Methods* and S2.6 in Dataset S2) identified the genomic region containing *KIT* as an outlier exhibiting unusually low genetic diversity (Fig. 4). We therefore resequenced *KIT* exons in a large cohort of domestic cats with various white-spotting phenotypes to genotype candidate SNVs (409 from 21 breeds, 5 Birman outcrosses, and 315 random bred cats). We identified just two adjacent missense mutations that were concordant with the gloving pattern in Birman cats (Fig. 4 and S2.7 in Dataset S2). Genotyping these SNPs in a larger sample including 150 Birman cats and 729 additional cats confirmed that all Birman cats were homozygous for both SNPs and that all first-generation outcrossed Birman cats with no gloving were carriers of the polymorphisms (S2.8 in Dataset S2).

Several lines of evidence indicate that the gloving phenotype in the Birman breed is the result of these two recessive mutations in *KIT*. Both mutations affect the fourth Ig domain of *KIT*, and mutations in this motif near the dimerization site have been shown to result in accelerated ligand dissociation and reduced downstream signal transduction events (68). Interestingly, the frequency of the Birman gloving haplotype in the Ragdoll breed, which shares an extremely similar white-spotting phenotype, was only 12.3%. We suggest that other genetic variants, including the endogenous retrovirus insertion in *KIT* intron 1 (65), likely contribute to the white-spotting phenotype in the Ragdoll breed. The frequency of the Birman gloving haplotype is just 10% in the random nonbreed population, thus illustrating a case where segregating genetic variation in ancestral nonbred populations has reached fixation within Birman cats through strong artificial selection in a remarkably short time frame.

In conclusion, our analyses have identified genetic signatures within feline genomes that match their unique biology and sensory skills. The number of genomic regions with strong signals of selection since cat domestication appears modest compared with those in

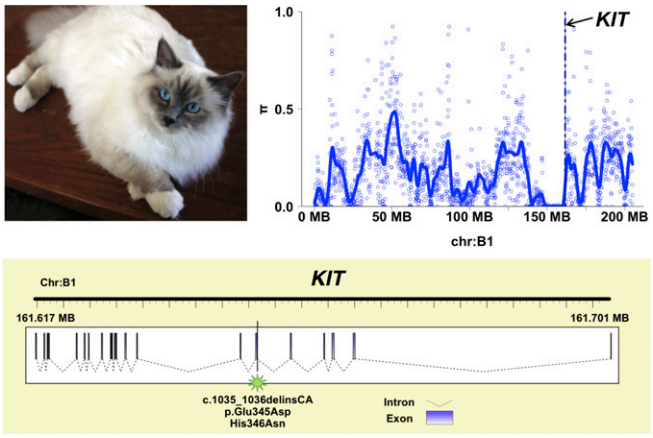

Fig. 4. Genetics of the gloving pigmentation pattern in the Birman cat. The paws of the Birman breed (*Top Left*) are distinguished by white gloving. The average nucleotide diversity adjacent to *KIT* was low (*Top Right*). Sequencing experiments identified two adjacent missense mutations within exon 6 of *KIT* that were concordant with the gloving pattern in Birman cats (*Bottom*).

the domestic dog (41), which is concordant with a more recent domestication history, the absence of strong selection for specific physical characteristics, as well as limited isolation from wild populations. Our results suggest that selection for docility, as a result of becoming accustomed to humans for food rewards, was most likely the major force that altered the first domesticated cat genomes.

Materials and Methods

A female Abyssinian cat, named Cinnamon, served as the DNA source for all sequencing reads (12). From this source we generated ~14× whole-genome shotgun coverage with Sanger and 454 technology. A BAC library was also constructed and all BACs were end-sequenced. We assembled the combined sequences using CABOG software (69) (*SI Materials and Methods*).

We estimated nonsynonymous and synonymous substitution rates using the software PAML 4.0 (17). The following pipeline was used to perform genome-wide selection analyses. (*i*) We identified 10,317 sets of 1:1:1:1:1 orthologs from the whole-genome annotations of human (GRCh37), cow (UMD3.1), dog (CanFam3.1), tiger (tigergenome.org), and domestic cat using the Ensembl pipeline (70). We tested for signatures of natural selection assuming the species tree topology (((cat, tiger), dog), cow, human). (*ii*) We aligned the translated amino acid sequence of the coding region of each gene using MAFFT (71) with the slow and most accurate parameter settings. A locally developed Perl script pipeline was applied that removed poorly aligned or incorrectly annotated amino acid residues caused by obvious gene annotation errors within the domestic cat and tiger genome assemblies. Aligned amino acid sequences were used for guiding nucleotide-coding sequences by adding insertion gaps and removing poorly aligned regions. (*iii*) Model testing and likelihood ratio tests (LRTs) were performed using PAML 4.0. Paired models representing different hypotheses consisted of branch tests and branch-site tests (fixed $\omega = 1$ vs. variable ω). For the branch-specific tests, free ratio vs. one-ratio tests were used to identify putatively positively selected genes. These genes were subsequently tested by two-ratio and one-ratio models to identify genes with significant positive selection of one branch versus all other branches (two-branch test). Significance of LRT results used a threshold of $P < 0.05$. We also report the mean synonymous rates along the ancestral felid lineage as well as the tiger, cat, and dog lineages (Fig. S1). We assessed enrichment of gene functional clusters under positive natural selection using WebGestalt (72) (S1.5–S1.7 in Dataset S1). Entrez Gene IDs were input as gene symbols, with the organism of interest set to *Homo sapiens* using the genome as the reference set. Significant Gene Ontology categories (73), Pathway Commons categories (74), WikiPathways (75), and KEGG Pathways (76) were reported using a hypergeometric test, and the significance level was set at 0.05. We implemented the Benjamini and Hochberg multiple test adjustment (77) to control for false discovery.

Using the whole-genome assembly of domestic cat (FelCat5) as a reference, we mapped Illumina raw sequences from a pool of four wildcat individuals [two European wildcats (*F. s. silvestris*) and two Eastern wildcats (*F. s. lybica*)]. Six additional domestic cat breeds from different worldwide regional populations were sequenced using the Illumina platform (*SI Materials and Methods*). Before sequencing, we pooled samples by breed for the following individuals: Maine Coon ($n = 5$), Norwegian Forest ($n = 4$), Birman ($n = 4$), Japanese Bobtail ($n = 4$), and Turkish Van ($n = 4$). Whole-genome sequencing was also performed on an Egyptian Mau cat ($n = 1$) and on the Abyssinian reference individual ($n = 1$).

We combined the raw reads from the following breed sequencing experiments (described above) before alignment and variant calling: Egyptian Mau, Maine Coon, Norwegian Forest, Birman, Japanese Bobtail, and Turkish Van. The domestic cat pool ($n = 22$) was sequenced to a genome coverage depth of ~58-fold, whereas the wildcat pool was sequenced to a depth of ~7-fold (S2.1 in Dataset S2). Base position differences were called using the convergent outcomes of the software SAMtools (78) and VarScan 2 (79). Parameters included a *P* value of 0.1, a map quality of 10, and parameters for filtering by false positives. A clustered variant filter was implemented to allow for a maximum of five variant sites in any 500-bp window. Variants were finally filtered using PoPoolation2 (80) to yield a high-confidence set of SNVs ($n = 6{,}534{,}957$; filtering steps included a minimum coverage of 8, a minimum variant count of 6, a maximum coverage of 500 for the domestic cat pool, and a maximum coverage of 200 for the wildcat pool).

We screened for positively selected candidate genes during cat domestication by parsing specific 100-kb windows that showed low diversity [low pooled heterozygosity (H_p)] in domestic cat breeds and had high divergence [a high fixation index (F_{ST})] between domestic cats and wildcats (41, 81). F_{ST} was calculated using PoPoolation2, and measurements of H_p were calculated using a custom script. A total of 6,534,957 high-quality SNV sites were used to calculate F_{ST} and H_p at each 100-kb window, and a step size of 50 kb was incorporated. All windows containing less than 10 variant sites were removed from the analysis, resulting in $n = 46{,}906$ 100-kb windows along cat autosomes, as represented in the FelCat5 assembly. We Z-transformed the autosomal H_p [$Z(H_p)$] and F_{ST} [$Z(F_{ST})$] distributions and designated as putatively selected regions those that fell at least 4 SDs away from the mean [$Z(H_p) < -4$ and $Z(F_{ST}) > 4$]. We applied a threshold of $Z(H_p) \leq -4$ and $Z(F_{ST}) \geq 4$ for putative selective sweeps, because windows below or above these thresholds represent the extreme lower and extreme upper ends of the respective distributions (Fig. S4). Windows with elevated F_{ST} or depressed H_p were annotated for gene content using the intersect tool in BEDTools (82). Enrichment analysis of underlying gene content was carried out using WebGestalt (72) using the same methods as described above, except only significant WikiPathways (75) and KEGG Pathways (76) were reported (S2.5 and S2.10–S2.11 in Dataset S2).

Primers to amplify *KIT* exons (ENSFCAG00000003112) were designed using Primer3Plus (83) and annealed to intronic regions flanking each exon. A PCR assay was performed to determine the presence or absence of the dominant, white-spotting retroviral insertion in *KIT* (65). An allele-specific PCR assay was designed for genotyping exon 6 SNPs (S2.9 in Dataset S2). See *SI Materials and Methods* for additional details.

ACKNOWLEDGMENTS. We thank The Genome Institute members Kim Kyung, Dave Larson, Karyn Meltz Steinburg, and Chad Tomlinson for providing assistance and advice on data analysis, and Tom Nicholas for manuscript review. We also thank NIH/National Institute on Alcohol Abuse and Alcoholism members David Goldman and Qiaoping Yuan. The cat genome and transcriptome sequencing was funded by NIH/National Human Genome Research Institute Grant U54HG003079 (to R.K.W.). Further research support included grants to M.W.H. (National Science Foundation Grant DBI-0845494), W.J.M. (Morris Animal Foundation Grants D06FE-063 and D12FE-019), T.M.-B. (European Research Council Starting Grant 260372 and Spanish Government Grant BFU2011-28549), and L.A.L. [National Center for Research Resources (R24 RR016094), Office of Research Infrastructure Programs/Office of the Director (R24 OD010928), and Winn Feline Foundation (W10-014, W09-009)].

1. American Pet Product Manufacturing Association (2008) *National Pet Owner's Survey* (Am Pet Prod Manuf Assoc, Greenwich, CT).
2. Meredith RW, et al. (2011) Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science* 334(6055):521–524.
3. Hedges SB, Dudley J, Kumar S (2006) TimeTree: A public knowledge-base of divergence times among organisms. *Bioinformatics* 22(23):2971–2972.
4. Sunquist M, Sunquist F (2002) *Wild Cats of the World* (Univ of Chicago Press, Chicago).
5. Vigne J-D, Guilaine J, Debue K, Haye L, Gérard P (2004) Early taming of the cat in Cyprus. *Science* 304(5668):259.
6. Hu Y, et al. (2014) Earliest evidence for commensal processes of cat domestication. *Proc Natl Acad Sci USA* 111(1):116–120.

7. Menotti-Raymond M, et al. (2008) Patterns of molecular genetic variation among cat breeds. *Genomics* 91(1):1–11.
8. Driscoll CA, et al. (2007) The Near Eastern origin of cat domestication. *Science* 317(5837):519–523.
9. Nussberger B, Greminger MP, Grossen C, Keller LF, Wandeler P (2013) Development of SNP markers identifying European wildcats, domestic cats, and their admixed progeny. *Mol Ecol Resour* 13(3):447–460.
10. Beaumont M, et al. (2001) Genetic diversity and introgression in the Scottish wildcat. *Mol Ecol* 10(2):319–336.
11. Bach LH, et al. (2012) A high-resolution $15{,}000_{Rad}$ radiation hybrid panel for the domestic cat. *Cytogenet Genome Res* 137(1):7–14.
12. Pontius JU, et al.; Agencourt Sequencing Team; NISC Comparative Sequencing Program (2007) Initial sequence and comparative analysis of the cat genome. *Genome Res* 17(11):1675–1689.
13. Mullikin JC, et al.; NISC Comparative Sequencing Program (2010) Light whole genome sequence for SNP discovery across domestic cat breeds. *BMC Genomics* 11(1):406.
14. Lindblad-Toh K, et al. (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438(7069):803–819.
15. Nicholas FW (2003) Online Mendelian Inheritance in Animals (OMIA): A comparative knowledgebase of genetic disorders and other familial traits in non-laboratory animals. *Nucleic Acids Res* 31(1):275–277.
16. Hill RE, Hastie ND (1987) Accelerated evolution in the reactive centre regions of serine protease inhibitors. *Nature* 326(6108):96–99.
17. Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8):1586–1591.
18. Woolley S, Johnson J, Smith MJ, Crandall KA, McClellan DA (2003) TreeSAAP: Selection on amino acid properties using phylogenetic trees. *Bioinformatics* 19(5):671–672.
19. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP (2012) Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE* 7(10):e46688.
20. Savage RJG (1977) Evolution in carnivorous mammals. *Palaeontology* 20:237–271.
21. Heffner RS, Heffner HE (1985) Hearing range of the domestic cat. *Hear Res* 19(1):85–88.
22. Riazuddin S, et al. (2006) Tricellulin is a tight-junction protein necessary for hearing. *Am J Hum Genet* 79(6):1040–1051.
23. Su C-C, et al. (2013) Mechanism of two novel human *GJC3* missense mutations in causing non-syndromic hearing loss. *Cell Biochem Biophys* 66(2):277–286.
24. Huang AS, Kim LA, Fawzi AA (2012) Clinical characteristics of a large choroideremia pedigree carrying a novel *CHM* mutation. *Arch Ophthalmol* 130(9):1184–1189.
25. Ewer RF (1973) *The Carnivores* (Cornell Univ Press, New York).
26. Miller KA, et al. (2012) Inner ear morphology is perturbed in two novel mouse models of recessive deafness. *PLoS ONE* 7(12):e51284.
27. Bauer JE (2006) Metabolic basis for the essential nature of fatty acids and the unique dietary fatty acid requirements of cats. *J Am Vet Med Assoc* 229(11):1729–1732.
28. Butterwick RF, Salt C, Watson TDG (2012) Effects of increases in dietary fat intake on plasma lipid and lipoprotein cholesterol concentrations and associated enzyme activities in cats. *Am J Vet Res* 73(1):62–67.
29. Johansson A, et al. (2011) Identification of *ACOX2* as a shared genetic risk factor for preeclampsia and cardiovascular disease. *Eur J Hum Genet* 19(7):796–800.
30. Unno N, et al. (2006) A single nucleotide polymorphism in the plasma PAF acetylhydrolase gene and risk of atherosclerosis in Japanese patients with peripheral artery occlusive disease. *J Surg Res* 134(1):36–43.
31. Cho YS, et al. (2013) The tiger genome and comparative analysis with lion and snow leopard genomes. *Nat Commun* 4:2433.
32. Liu S, et al. (2014) Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell* 157(4):785–794.
33. Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW (2013) Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol Biol Evol* 30(8):1987–1997.
34. De Bie T, Cristianini N, Demuth JP, Hahn MW (2006) CAFE: A computational tool for the study of gene family evolution. *Bioinformatics* 22(10):1269–1271.
35. Kitchener AC (1991) *The Natural History of the Wild Cats* (Cornell Univ Press, New York).
36. Gilad Y, Przeworski M, Lancet D, Lancet D, Pääbo S (2004) Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. *PLoS Biol* 2(1):E5.
37. Warren WC, et al. (2008) Genome analysis of the platypus reveals unique signatures of evolution. *Nature* 453(7192):175–183.
38. Li G, Janecka JE, Murphy WJ (2011) Accelerated evolution of *CES7*, a gene encoding a novel major urinary protein in the cat family. *Mol Biol Evol* 28(2):911–920.
39. Cameron-Beaumont C, Lowe SE, Bradshaw CJA (2002) Evidence suggesting preadaptation to domestication throughout the small Felidae. *Biol J Linn Soc Lond* 75(3):361–366.
40. Driscoll CA, Macdonald DW, O'Brien SJ (2009) From wild animals to domestic pets, an evolutionary view of domestication. *Proc Natl Acad Sci USA* 106(Suppl 1):9971–9978.
41. Axelsson E, et al. (2013) The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495(7441):360–364.
42. Alhaddad H, et al. (2013) Extent of linkage disequilibrium in the domestic cat, *Felis silvestris catus*, and its breeds. *PLoS ONE* 8(1):e53537.
43. Fukuda E, et al. (2008) Down-regulation of protocadherin-α A isoforms in mice changes contextual fear conditioning and spatial working memory. *Eur J Neurosci* 28(7):1362–1376.
44. Mead AN, Stephens DN (2003) Selective disruption of stimulus-reward learning in glutamate receptor *gria1* knock-out mice. *J Neurosci* 23(3):1041–1048.
45. Mead AN, Brown G, Le Merrer J, Stephens DN (2005) Effects of deletion of *gria1* or *gria2* genes encoding glutamatergic AMPA-receptor subunits on place preference conditioning in mice. *Psychopharmacology (Berl)* 179(1):164–171.
46. Horn KE, et al. (2013) *DCC* expression by neurons regulates synaptic plasticity in the adult brain. *Cell Reports* 3(1):173–185.
47. Yetnikoff L, Almey A, Arvanitogiannis A, Flores C (2011) Abolition of the behavioral phenotype of adult netrin-1 receptor deficient mice by exposure to amphetamine during the juvenile period. *Psychopharmacology (Berl)* 217(4):505–514.
48. Takebe A, et al. (2006) Microarray analysis of *PDGFRα*$^{+}$ populations in ES cell differentiation culture identifies genes involved in differentiation of mesoderm and mesenchyme including *ARID3b* that is essential for development of embryonic mesenchymal cells. *Dev Biol* 293(1):25–37.
49. Demirkan A, et al.; DIAGRAM Consortium; CARDIoGRAM Consortium; CHARGE Consortium; EUROSPAN Consortium (2012) Genome-wide association study identifies novel loci associated with circulating phospho- and sphingolipid concentrations. *PLoS Genet* 8(2):e1002490.
50. Wilkins AS, Wrangham RW, Fitch WT (2014) The "domestication syndrome" in mammals: A unified explanation based on neural crest cell behavior and genetics. *Genetics* 197(3):795–808.
51. Kobayashi K, Jakt LM, Nishikawa SI (2013) Epigenetic regulation of the neuroblastoma genes, *Arid3b* and *Mycn*. *Oncogene* 32(21):2640–2648.
52. Wei Z, Yan J, Lu Q, Pan L, Zhang M (2011) Cargo recognition mechanism of myosin X revealed by the structure of its tail MyTH4-FERM tandem in complex with the *DCC* P3 domain. *Proc Natl Acad Sci USA* 108(9):3572–3577.
53. Zhu X-J, et al. (2007) Myosin X regulates netrin receptors and functions in axonal path-finding. *Nat Cell Biol* 9(2):184–192.
54. Hwang Y-S, Luo T, Xu Y, Sargent TD (2009) Myosin-X is required for cranial neural crest cell migration in *Xenopus laevis*. *Dev Dyn* 238(10):2522–2529.
55. Brown KR, Jurisica I (2007) Unequal evolutionary conservation of human protein interactions in interologous networks. *Genome Biol* 8(5):R95.
56. Hauswirth R, et al. (2012) Mutations in MITF and PAX3 cause "splashed white" and other white spotting phenotypes in horses. *PLoS Genet* 8(4):e1002653.
57. Rubin CJ, et al. (2012) Strong signatures of selection in the domestic pig genome. *Proc Natl Acad Sci* 109(48):19529–19536.
58. Reissmann M, Ludwig A (2013) Pleiotropic effects of coat colour-associated mutations in humans, mice and other mammals. *Semin Cell Dev Biol* 24(6-7):576–586.
59. Menotti-Raymond M, et al. (2009) An autosomal genetic linkage map of the domestic cat, *Felis silvestris catus*. *Genomics* 93(4):305–313.
60. Ross-Ibarra J (2004) The evolution of recombination under domestication: A test of two hypotheses. *Am Nat* 163(1):105–112.
61. Coop G, Przeworski M (2007) An evolutionary view of human recombination. *Nat Rev Genet* 8(1):23–34.
62. Kanan Y, Wicker LD, Al-Ubaidi MR, Mandal NA, Kasus-Jacobi A (2008) Retinol dehydrogenases *RDH11* and *RDH12* in the mouse retina: Expression levels during development and regulation by oxidative stress. *Invest Ophthalmol Vis Sci* 49(3):1071–1078.
63. Kurushima JD, et al. (2013) Variation of cats under domestication: Genetic assignment of domestic cats to breeds and worldwide random-bred populations. *Anim Genet* 44(3):311–324.
64. Anonymous (July 22, 1871). Crystal Palace - Summer concert today, Cat Show on July 13. *Penny Illustrated Paper.* p 16.
65. David VA, et al. (2014) Endogenous retrovirus insertion in the *KIT* oncogene determines white and white spotting in domestic cats. *G3 (Bethesda)*, 10.1534/g3.114.013425.
66. Cooper MP, Fretwell N, Bailey SJ, Lyons LA (2006) White spotting in the domestic cat (*Felis catus*) maps near *KIT* on feline chromosome B1. *Anim Genet* 37(2):163–165.
67. Geissler EN, Ryan MA, Housman DE (1988) The dominant-white spotting (W) locus of the mouse encodes the c-kit proto-oncogene. *Cell* 55(1):185–192.
68. Blechman JM, et al. (1995) The fourth immunoglobulin domain of the stem cell factor receptor couples ligand binding to signal transduction. *Cell* 80(1):103–113.
69. Miller JR, et al. (2008) Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics* 24(24):2818–2824.
70. Flicek P, et al. (2012) Ensembl 2012. *Nucleic Acids Res* 40(database issue):D84–D90.
71. Katoh K, Toh H (2010) Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics* 26(15):1899–1900.
72. Wang J, Duncan D, Shi Z, Zhang B (2013) WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): Update 2013. *Nucleic Acids Res* 41(web server issue):W77–W83.
73. Ashburner M, et al.; The Gene Ontology Consortium (2000) Gene Ontology: Tool for the unification of biology. *Nat Genet* 25(1):25–29.
74. Cerami EG, et al. (2011) Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res* 39(database issue):D685–D690.
75. Kelder T, et al. (2012) WikiPathways: Building research communities on biological pathways. *Nucleic Acids Res* 40(database issue):D1301–D1307.
76. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 40(database issue): D109–D114.
77. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 57(1):289–300.
78. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
79. Koboldt DC, et al. (2012) VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 22(3):568–576.
80. Kofler R, Pandey RV, Schlötterer C (2011) PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* 27(24):3435–3436.
81. Rubin C-J, et al. (2010) Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* 464(7288):587–591.
82. Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842.
83. Untergasser A, et al. (2012) Primer3—New capabilities and interfaces. *Nucleic Acids Res* 40(15):e115.

Supporting Information

Montague et al. 10.1073/pnas.1410083111

SI Materials and Methods

Genome Assembly. The current draft assembly is referred to as FelCat5 or *Felis catus* 6.2. There are ∼2.35 Gb (including Ns in gaps) on ordered/oriented chromosomes, ∼15.4 Mb on the chr*_random, and ∼11.74 Mb on chromosome Un. Initially, we ran CABOG 6.1 (1) with default parameters (2). To evaluate changes in contiguity, we altered a small set of the default parameters to obtain the best assembly possible. CABOG settings, including parameters used, are available upon request.

To create an initial chromosomal version of the assembly, we aligned marker sequences associated with a radiation hybrid (RH) map (3) to the assembled genome sequence. The chromosomal index file (.agp) contains the ordered/oriented bases for each chromosome (named after the respective linkage group).

Once scaffolds were ordered and oriented along the cat chromosomes using the RH map marker content (3), the assembled cat genome was broken into 1-kb segments and aligned against the dog genome (CanFam2) and human genome (hg19) using BLASTZ (4) to align and score nonrepetitive cat regions against repeat-masked dog and human sequences, respectively. BLASTZ (4) and BLAT (5) alignments with the dog and human genomes were then used to refine the order and orientation information as well as to insert additional scaffolds into the conditional scaffold framework provided by the marker assignments. Alignment chains differentiated all orthologous and paralogous alignments, and breakpoint identification confirmed a false join within the genome assembly. Only "reciprocal best" alignments were retained in the alignment set. Finally, satellite sequences were identified in the genome, and centromeres were placed along each chromosome using localization data (3) in combination with the localization of the satellite sequences. In the last step, finished cat BACs (n = 86; totaling ∼14.92 Mb) were integrated into the assembly, using the BLAT (5) aligner for accurate coordinates.

Gene Family Expansions and Contractions. To explore gene family expansions and contractions, we obtained peptides from cat, dog, ferret, panda, cow, pig, horse, human, and elephant from Ensembl (6). We clustered these into protein families by performing an all-against-all BLAST (7) search using the OrthoMCL clustering program (8). The clusters were converted to CAFE (9) format, and families were filtered out based on the following groupings: (*i*) at least one protein must be present in (elephant, human, horse), and (*ii*) at least one protein must be present in (cat, dog, ferret, panda, cow, pig), or the family is filtered out. We used www.timetree.org to obtain divergence times for all species to construct the following tree: (elephant:101.7, (human:94.2, (horse:82.4, ((pig:63.1, cow:63.1):14.3, (cat:55.1, (dog:42.6, (ferret:38, panda:38):4.6):12.5):22.3):5):11.8):7.5).

From phylogenetic inference, we found 50 expanded gene families in the cat genome, of which 28 have known homologs in other mammals (Fig. S3 and S1.8 in Dataset S1). Analyses using CAFE 3.0 (9) confirmed contraction in multiple *Or* gene families and expansion in the *V1r* gene family in the ancestor of modern felids (S1.8 in Dataset S1), with differential gene gain and loss within the cat family (Fig. S2).

In addition to the chemosensory *Or* and *V1r* gene families mentioned, we found evidence for expansion of genes related to processes of mechanotransduction (10) (*PIEZO2*), T-cell receptors (*TRAV8*), melanocyte development (11) (*SOX10*), and meiotic processes (*SYCP1*). Four gene families were complete losses along the cat lineage; the annotations for these entries for other species include gene families related to reproduction (spermatogenesis-associated protein 31D1 and precursor acrosomal vesicle 1), secretory proteins (precursor lipophilin), and hair fibers (high sulfur keratin associated).

Segmental Duplication, Copy Number (CNV) Discovery, and Structural Variation.

Sequencing data. For the domestic cat (Abyssinian), sequenced with Illumina technology, bam files resulting from mapping 100-bp reads were used to recover the original fastq reads, which were clipped into 36-bp reads after trimming the first 10 bp to avoid lower-quality positions. That is, we used a total of 1,485,609,004 reads for mapping (coverage 21.8×).

Reference assembly. We downloaded the FelCat5 assembly from the UCSC Genome Browser (12). The 5,480 scaffolds either unplaced or labeled as random were concatenated into a single artificial chromosome. In addition to the repeats already masked in FelCat5 with RepeatMasker (www.repeatmasker.org) and Tandem repeats finder (13), we sought to identify and mask potential hidden repeats in the assembly. To do so, chromosomes were partitioned into 36-bp K-mers (with adjacent K-mers overlapping 5 bp), and these were mapped against FelCat5 using mrsFAST (14). Next, we masked positions in the assembly mapped by K-mers with more than 20 placements in the genome, resulting in 5,942,755 bp additionally masked compared with the original masked assembly (Fig. S3).

Mapping and copy number estimation from read depth. In the domestic cat, the 36-bp reads resulting from clipping the original fastq reads (see above) were mapped to the prepared reference assembly using mrFAST (15). mrCaNaVaR (version 0.41) (15) was used to estimate the copy number along the genome from the mapping read depth. Briefly, mean read depth per base pair is calculated in 1-kbp nonoverlapping windows of nonmasked sequence (that is, the size of a window will include any repeat or gap, and thus the real window size may be larger than 1 kbp). Importantly, because reads will not map to positions covering regions masked in the reference assembly, read depth will be lower at the edges of these regions, which could underestimate the copy number in the subsequent step. To avoid this, the 36 bp flanking any masked region or gap were masked as well and thus are not included within the defined windows. In addition, gaps >10 kbp were not included within the defined windows. A read depth distribution was obtained through iteratively excluding windows with extreme read depth values relative to the normal distribution, and the remaining windows were defined as control regions (Fig. S3 and S1.9 in Dataset S1). The mean read depth in these control regions was considered to correspond to a copy number equal to two and was used to convert the read depth value in each window into a GC-corrected absolute copy number. Note that the control/noncontrol status was determined based on the read depth distribution, making this step critical for further copy number calls. Of the 993,102 control windows, none aligned to the artificial chromosome (see above), and 37,123 (3.7%) were on chromosome X in the sample.

Calling of duplications and deletions. The copy number distribution in the control regions was used to define specific gain/loss cutoffs as the mean copy number plus/minus three units of SD (calculated not considering those windows exceeding the 1% highest copy number value). Note that because the mean copy number in the control regions was equal to two (by definition), the gain/loss cutoffs were largely influenced by the SD.

We used two methods to call duplications: M1, the circular binary segmentation (CBS) method (16), was used to combine 1-kbp windows that represent segments with significantly the same copy number. Segments with copy number (defined as the median copy number of the 1-kbp windows comprising the segment) exceeding the gain/loss cutoffs defined above (but lower than 100 copies in the case of duplications) were merged and called as duplications or deletions if comprising more than 10 1-kbp windows (~10 kbp); finally, only duplications with >85% of their size not overlapping with repeats were retained for the analyses.

As a second method (M2), we also called duplications avoiding the segmentation step with the CBS method by merging 1-kbp windows with copy number larger than sample-specific gain cutoff (but lower than 100 copies) and then selecting those regions comprising at least five 1-kbp windows and >10 kbp; similarly, only duplications with >85% of their size not overlapping with repeats were retained for the analyses.

In M1, the copy number distribution in the control regions was used to define sample-specific gain/loss cutoffs as the mean copy number plus/minus three units of SD (calculated not considering those windows exceeding the 1% highest copy number value). Note that because the mean copy number in the control regions is equal to two by definition, the gain/loss cutoffs will be largely influenced by the SD. Then, we merged 1-kbp windows with copy number larger than sample-specific gain cutoff (but lower than 100 copies) and identified as duplications the regions that comprised at least five 1-kbp windows and >10 kbp. Finally, only duplications with >85% of their size not overlapping with repeats were retained. This method is highly restrictive (conservative), so we used an alternative method (M2) similar to what had been previously done with Sanger capillary reads (17). We performed a 5-kbp sliding window approach and required six out of seven windows with a significantly higher read depth, relative to the control regions, to consider a region as duplicated.

Several categories were significantly overrepresented in regions of expanded CNV (Fig. S3 and S1.10–S1.12 in Dataset S1), some of which overlap those identified in other CNV studies for other taxa (18–22). In the cat, we note that an expanded CNV region on chromosome B2 contained a pair of genes that transcribe an MHC class I antigen and an MHC class I antigen precursor. The MHC class I molecules present self-antigens to cytotoxic $CD8^+$ T lymphocytes and regulate natural killer cell activity. Investigations of MHC genes in other domesticated animals, including pig (23), sheep (24), and cow (25, 26), have shown that MHCs in these groups are affected by CNV. These results suggest that CNV is an additional common source of disease resistance or susceptibility variability in the MHC of the cat as well.

***V1r/Or* Identification and Annotation.** Published *V1r* and *Or* sequences from human, mouse, rat, cow, dog, and opossum were used as the query sequences for BLAST (7) searches against the domestic cat genome. All query sequences were previously shown as belonging to *V1r* (27, 28) and *Or* (29) subfamilies, thus ensuring identification of the most complete gene repertoires. We enforced an *E*-value threshold of 10^{-5} for filtering BLAST results. All identified sequences were extended 1.5 kb on either side for open reading identification and assessment of functionality. If multiple start codons were found, the alignment results of known intact mammalian *V1r* and *Or* amino acid sequences were used as guidance for determining the most appropriate one. Any putative genes containing early stop codons, frameshift mutations, and/or incomplete gene structure (i.e., not containing three extracellular regions, seven transmembrane regions, and three intracellular regions) were designated as pseudogenes. To confirm orthology, we aligned all members of the *V1r* and *Or* gene families and constructed maximum likelihood trees rooted with appropriate outgroup taxa, such as *V2r* and taste receptor gene families. Assembled whole-sequencing data were obtained from the Ensembl database (6) [domestic cat: vFelCat5; domestic dog: vCanFam; domestic horse: vEquCab2; human: vGRCh39; domestic cow: vBosTau7; great panda: vAilMel1; and tiger (tigergenome.org)]. *V1r* gene clusters were defined as all identified functional genes and pseudogenes within a 2-Mb window. Synteny blocks of different mammals were identified using the software SyntenyTracker (30).

Felid V1r sequencing. The following felid taxa were used for *V1r* PCR and sequencing: *Felis catus* (domestic cat; FCA), *Felis nigripes* (Black-footed cat; FNI), *Prionailurus bengalensis* (Leopard cat; PBE), *Prionailurus viverrinus* (Fishing cat; PVI), *Puma concolor* (Cougar; PCO), *Puma yagouaroundi* (Jaguarundi; PYA), *Acinonyx jubatus* (Cheetah; AJU), *Lynx canadensis* (Canadian Lynx; LCA), *Lynx lynx* (Eurasian Lynx; LLY), *Lynx pardinus* (Iberian Lynx; LPA), *Lynx rufus* (Bobcat; LRU), *Leopardus pardalis* (Ocelot; LPA), *Leopardus wiedii* (Margay; LWI), *Leopardus geoffroyi* (Geoffroy's cat; LGE), *Leopardus colocolo* (Pampas cat; LCO), *Leopardus tigrinus* (Tiger cat; LTI), *Profelis serval* (Serval; PSE), *Profelis caracal* (Caracal; PCL), *Pardofelis temminckii* (Asian Golden cat; PTE), *Pardofelis marmorata* (Marbled cat; PMA), *Neofelis nebulosa* (Clouded Leopard; NNE), *Panthera leo* (Lion; PLE), *Panthera onca* (Jaguar; PON), *Panthera pardus* (Leopard; PPA), *Panthera tigris* (Tiger; PTI), and *Panthera uncia* (Snow Leopard; PUN). Forty-three pairs of primers for *V1r* amplification were designed using several versions of the domestic cat whole-genome assembly (FelCat1–FelCat5). Target amplicons were designed to be longer than 1.1 kb to ensure amplification of the complete coding region sequence. PCR was performed using PlatinumTaq DNA polymerase using a touchdown profile of 60–55 °C, as described (31). All amplicons were sequenced using Sanger sequencing on an ABI 3700 (Applied Biosystems). A total of 1,055 sequences of intact *V1r* genes and pseudogenes from 27 cat species were submitted to GenBank under accession numbers KJ923925–KJ924979.

Sequence alignment and phylogenetic reconstruction. We aligned our previously unidentified *V1r* sequences with known published *V1r* sequences using MAFFT (32) with stringent parameter settings. Coding sequences were aligned under the guidance of the translated amino acid alignment results. Poorly aligned 5′ and 3′ flanking regions were trimmed before tree building. MODELTEST (33) was used to estimate the best nucleotide substitution models and parameters for sequence data. Maximum likelihood trees (with 500 bootstrap replicates) were constructed with RAxML7.0.0 (34).

Estimation of gene gain and loss within V1r and Or gene families. We compared the *Or* and *V1r* gene trees generated above with a mammalian species tree (35) to estimate gene gain and loss using the software NOTUNG (36). We examined variation in *V1r* and *Or* gene family repertoire size among different domestic cat breeds by aligning Illumina reads to the cat assembly using BWA (37). Mapping results were analyzed with CNVnator (38). We reestimated the tiger *Or* and *V1r* repertoires by remapping all of the raw tiger Illumina reads to the Siberian tiger assembly (tigergenome.org) as well as the current domestic cat version 6.2 assembly.

Natural Selection Tests.

Phylogenetic analyses by maximum likelihood. Four sets of models were applied for null hypothesis and alternative hypothesis comparisons. Set 1 involved a comparison between the free-ratio model and the one-ratio model, whereas set 2 compared the two-ratio model with the one-ratio model. These two comparisons are classified as branch-specific tests, which were used to identify accelerated rates of genes on specific branches of an evolutionary tree. In addition, we performed site-specific tests, which detected natural selection acting on specific amino acid sites of the protein. For this step, we performed model tests within sets 3 and 4, which involved model 1a (nearly neutral) versus model 2a (positive selection) and model 7 (gamma) versus model 8

(gamma and ω) to evaluate and identify specific amino acid sites that were potentially under positive selection.

To evaluate the structural influence of domestic cat nonsynonymous substitutions from the common ancestor of felids, we used TreeSAAP (39) to measure 31 structural and biochemical amino acid properties while applying the tree topology (human, (cow, (dog, (cat, tiger)))). We used a significance threshold of $P < 0.001$ to report structural or biochemical properties of amino acid substitutions likely to affect protein function. We also used PROVEAN (40) to predict the potential functional impact of domestic cat-specific amino acid substitutions and indels. We considered amino acid substitutions as "deleterious" if the PROVEAN score was ≤−2.5. We considered amino acid substitutions as "neutral replacements" if the PROVEAN score was >−2.5 (Fig. S1).

To explore the heterogeneous selection pressure across positively selected genes, peaks of high d_N/d_S were visualized using sliding window analyses performed across alignments of the full coding sequence. Sliding windows of ω values were estimated using the Nei and Gojobori method (41) with a default window size of 90 bp and a step size of 18 bp.

Many positively selected genes appear to have played a role in the sensory evolution of felines, as highlighted above. For instance, chemosensory genes with significant signatures of positive selection in the Felinae include two gustducin-coupled bitter taste receptors, *TAS2R1* and *TAS2R3*, as well as a cofactor, *RTP3* (S1.3 in Dataset S1). We speculate that selection at these loci increased sensitivity to and avoidance of toxic prey items in the hypercarnivorous ancestor of cats (42). Other positively selected genes appear to have played a role in the morphological evolution of carnivores. For instance, all carnivores have robust claws (except where they are secondarily lost) that serve as critical adaptations to capture and disarticulate prey. The *RSPO4* gene (S1.1 in Dataset S1) plays a crucial role in nail morphogenesis across mammals, and its expression is restricted to the developing nail mesenchyme (43). Further, the recessive human disorders anonychia/hyponychia congenita result from mutations in *RPSO4* (44), and are characterized by absence of or severe reduction in fingernails and toenails. Evidence of positive selection within the *RSPO4* gene in the ancestral carnivore lineage likely reflects molecular adaptations driving enhanced nail morphology.

Genome mapping and variant analysis. We next performed whole-genome analyses of cats from different domestic breeds [Maine Coon (SRX026946, SRX026943, SRX026929), Norwegian Forest (SRX027004, SRX026944, SRX026941, SRX026909, SRX026901), Birman (SRX026955, SRX026947, SRX026911, SRX026910), Japanese Bobtail (SRX026948, SRX026928, SRX026912), Turkish Van (SRX026942, SRX026930, SRX026913), and Egyptian Mau (SRX019549, SRX019524, SRX026956, SRX026945)] and wildcats [i.e., other *F. silvestris* subspecies (SRX026960)] using pooling methods that control for genetic drift (45). All reads were preprocessed by removing duplicate reads and only properly paired reads were aligned to the FelCat5 reference using BWA (37) (n = 2,332,398,473 reads from the pooled domestic cats combined; n = 189,543,907 reads from pooled wildcats). A total of 8,676,486 and 5,190,430 high-quality single-nucleotide variants (SNVs) among domestic breeds and wildcats, respectively, at a total of 10,975,197 sites, passed the thresholds using our initial variant-calling methods with SAMtools (46) and VarScan (47). Because SNVs for the domestic and wildcat pools were called separately, variants ascertained in one may not be present in the other. This can be due to homozygosity for the reference allele or inadequate data at the locus. We therefore implemented a consensus-calling analysis for the combined variant set to categorize each SNV as high-quality passing, low-quality failure, or no sequence coverage within each pool for all 10,975,197 passing sites. To do this, we generated a two-sample mpileup using SAMtools (46) for every site that was called a variant. We next implemented the mpileup2cns command in VarScan (47) with the minimum read depth set to three. Because every site in the mpileup passed the initial false positive filtering in at least one pool, we were able to determine the percentage of variant overlap between the pool of domestic cats and the pool of wildcats. This revealed 9,010,197 shared variant alleles between the domestic cats and wildcats, indicating that 1.7% and 10.3% of sites with variant alleles were unique to domestic cats and wildcats, respectively (Fig. S4). As expected, due to the coverage differences between the pools, a total of 3,121 and 745,091 sites, in the pooled domestic cats and pooled wildcats, respectively, contained low coverage (fewer than three aligned reads) or missing coverage.

We next used VCFtools (48) to explore the extent of overlap between the different variant callers. For the domestic cat pool, SAMtools (46) called 11,119,091 variants and VarScan (47) called 10,138,788 variants. A total of 9,683,549 variants overlapped, revealing that 4.5% and 12.9% of the original VarScan (47) and original SAMtools (46) calls, respectively, were undetected by the other variant caller. For the wildcat pool, SAMtools (46) called 9,860,972 variants and VarScan (47) called 9,098,242 variants. A total of 7,848,268 variants overlapped, revealing that 13.7% and 20.4% of the original VarScan (47) and original SAMtools (46) calls, respectively, were undetected by the other variant caller.

SNV validation. We verified our high-quality set of SNVs by comparing the list of markers with those of an SNP array developed previously (49). To accomplish this, we used BLAST (7) to locate the best-hit coordinates along the *F. catus* 6.2 reference assembly for each of the array variants. We then parsed our pooled domestic cat variant file for matching coordinates and discovered 184 out of 384 variants (47.9%). The calls made by our pipeline matched the variant on the chip in 183 out 184 cases (99.5%).

Breed differentiation. We verified the genetic relationships among the breeds using multidimensional scaling (MDS) and a population stratification analysis. Seven populations of 26 domestic cats were analyzed, including the breeds described above as well as a population of Eastern Random Bred cats (n = 4; SRX026993). Genome mapping and variant calling was performed on a per-breed basis using described variant-calling methods (above). After aligning the short reads to FelCat5, we identified 77,749 high-quality variants that were shared among all seven breeds. The pedigree genotype file was quality-controlled with PLINK (50) to remove all individuals with more than 80% missing genotype data, all SNVs missing in more than 5% of cases, and all SNVs with less than 5% minor allele frequency (MAF). Following quality control filtering, a total of 44,377 autosomal SNVs remained. MDS was implemented using PLINK (50) to produce an output file with identity by state values, and genetic distances of the first four principal coordinates were visualized (Fig. S5). Model-based clustering was performed with ADMIXTURE (51). A total of 20 replicates of K = 2 to K = 20 was run in unsupervised mode, each with random seeds and fivefold cross-validation. The replicates of each Q file for each K were merged using the LargeKGreedy method (with random input orders) using the program CLUMPP (52). The merged Q files were then visualized in DISTRUCT (53) to output plots of estimated membership coefficients for each individual according to each K, with K = 5 offering the highest support (Fig. S5).

Discovering putative regions of selection in the domestic cat genome. As a quality control assessment, the average H_p and F_{ST} of all autosomal 100-kb windows were plotted against the corresponding number of segregating sites per window. In line with our expectations, H_p was positively correlated with the number of segregating sites (rho = 0.021, $P < 0.001$, Spearman; Fig. S5) whereas F_{ST} was negatively correlated (rho = 0.225, $P < 0.001$, Spearman; Fig. S5), suggesting that the number of variants per window was lower in our putative regions of selection due to the loss of linked variation following an adaptive sweep. We also compared the depth of coverage at variant sites within the putative regions of selection with the depth of coverage at variants

found within all other genomic regions. The average read depth among the 3,265 variant positions for pooled domestic cats within the five regions of putative selection was relatively equivalent to the average read depth of all 8,676,486 variants across all autosomes for pooled domestic cats (53.82 versus 53.65, respectively).

Although accurate detection of heterozygosity is dependent on coverage, similar depths among the breed pools and members of each pool were not obtained for this study. Further, individual cats were not indexed when pooling by breed. Although equal numbers of samples among pools and subsets were difficult to obtain, we tested whether unequal representation between domestic cat and wildcat contributed to variance of the divergence statistics across the genome by reperforming the F_{ST} analysis based on a random subsample of the domestic cat data where the average coverage (6.81×) approximated the original coverage for the wildcat pool (6.84×), with ∼1.1× coverage contributed by each domestic breed pool. First, the variant-calling pipeline identified 3,494,488 total variants in the subsampled data. A final variant set consisting of 1,274,175 autosomal variants was then used for a sliding window analysis of F_{ST} using the same methods as the original analysis. When analyzing the subsampled data, all windows that passed the threshold under the original analysis were found within the 99th percentile of highest F_{ST} using the subsampled domestic cat data (Fig. S5). All of the original windows were thus identified as windows with high divergence using the subsampled data. These results suggest that the unequal sample sizes of domestic cats and wildcats likely had little effect on the overall results of our sliding window analyses.

Analysis of the X chromosome. To not confound the results of the autosomal analyses, we analyzed the X chromosome separately, using the method as described previously for the autosomes. We found that the average pooled heterozygosity, H_p, is higher (H_pX: 0.496 vs. H_pA: 0.385) and the average fixation index, F_{ST}, is higher (F_{ST}X: 0.674 vs. F_{ST}A: 0.429) on X compared with on autosomes. We also note that the SDs of the H_p (σX: 0.049 vs. σA: 0.029) and F_{ST} (σX: 0.183 vs. σA: 0.074) distributions are larger on the X chromosome relative to the autosomes. No windows passed the thresholds of significance [$Z(H_p) < -4$ or $Z(F_{ST}) > 4$] used for the autosomal analyses. We instead applied a lower threshold of 1.5 SDs from the mean of both the H_p and F_{ST} distributions. A total of 54 windows, representing 36 unique regions, passed this cutoff in the F_{ST} analysis (S2.12 in Dataset S2). A total of 210 windows representing 72 unique regions passed this threshold for the H_p analysis (S2.13 in Dataset S2). Known genes underlying regions of low domestic H_p and high F_{ST} (Fig. S6 and S2.14 in Dataset S2) include cyclin B3 (*CCNB3*), Cdc42 guanine nucleotide exchange factor 9 (*ARHGEF9*), zinc finger C4H2 domain containing (*ZC4H2*), family with sequence similarity 155, member B (*FAM155B*), protocadherin 19 (*PCDH19*), annexin A2 (*ANXA2*), and brain expressed X-linked 5 (*BEX5*). Our sliding window analysis along the autosomes revealed a strong trend associating genomic signatures of selection in domestic cats with genes influencing memory, fear-conditioning behavior, and stimulus-reward learning, particularly those predicted to underlie the evolution of tameness (54). This analysis of the X chromosome reveals similar functional trends, with four of six regions containing genes associated with neurological diseases and aberrant synaptic activity, including an additional protocadherin locus.

The Z-transformation technique, outlined above, resulted in a skewed (i.e., not normal) distribution (Fig. S6), so the conclusions must be viewed cautiously. By applying a percentile approach, we found that no genes underlie windows that met thresholds for either the 99th percentile or the 95th percentile for both F_{ST} and domestic H_p. Only a single window met the 99th percentile for F_{ST} and the 99th percentile for domestic H_p. This window (X:23800000–23900000), although noncoding, is within the X-linked *MAGE* gene family complex. The protocadherin gene that we highlighted above (*PCDH19*) was found within the 95th percentile threshold for F_{ST} and the 90th percentile for domestic H_p. The annexin gene (*ANXA2*), which is located within an adjacent window to *PCDH19*, met the 95th percentile threshold for F_{ST} and the 85th percentile for domestic H_p. Only one other gene displayed a higher F_{ST} value than *PCDH19* and *ANXA2*: *BEX5*, also highlighted by our Z-transformation analysis, met the 99th percentile threshold for F_{ST} and the 90th percentile for domestic H_p.

Pigmentation Patterns in Domestic Cat Breeds. Several breeds represent random bred populations of cats that do not have strong selection on a specific trait, such as Maine Coon and Norwegian Forest; however, the vast majority of cat breeds, including Japanese Bobtail, Birman, Egyptian Mau, and Turkish Van, likely experienced strong selection on novel and specific mutations (i.e., morphological traits and pigmentation patterns), as individuals were selected from random bred populations.

The genomic sequence data from the pooled Birman breed revealed an ∼10-Mb homozygous block located directly upstream of *KIT*. The average nucleotide diversity for 100-kb windows adjacent to *KIT* was lower (ChrB1: 161.5–161.9 Mb; pi = 0.0011) than the average nucleotide diversity for 100-kb windows across all autosomes (pi = 0.2185) or the average nucleotide diversity for 100-kb windows across ChrB1 (pi = 0.1762) (Fig. 4). An additional analysis of 63K single-nucleotide variants in individual Birman cats revealed an ∼5-Mb homozygous block located directly upstream of *KIT*. This loss of variation could be explained by genetic drift (e.g., inbreeding, the small founder population of the breed) or as a consequence of selection (e.g., the white gloving trait is fixed and recessive). We hypothesized that an extensive homozygous block is a measure of the selection on the gloving trait because Birman is highly selected for coat color, and we discovered a unique pair of fixed SNVs within the Birman breed that are associated with amino acid changes in *KIT*.

Samples and genotyping. We noninvasively collected DNA samples from all domestic cats by buccal swabs using a cytological brush or cotton tip applicator. DNA was isolated using the QIAamp DNA Mini Kit (Qiagen). The previous linkage analysis pedigree from the Waltham Centre for Pet Nutrition (55) was extended from 114 to 147 cats to refine the linkage region. Phenotypes were determined as in the previous study (55). Two previously published short tandem repeats (STRs) (56) (*FCA097* and *FCA149*) and four previously unidentified feline-derived STRs (*UCDC259b*, *UCDC443*, *UCDC487*, and *UCDC489*) (S2.9 in Dataset S2), flanking *KIT* on feline chromosome B1, were genotyped. Genotyping for the markers and two-point linkage between the microsatellite genotypes and the spotting phenotype was conducted using the LINKAGE (57) and FASTLINK (58) programs as in previous studies (59).

Genomic analysis of KIT. To identify *KIT* exons, publicly available (in GenBank) sequences from various species were aligned, including *Homo sapiens* (NM_000222.2), *Canis familiaris* (NM_001003181.1), *Mus musculus* (NM_021099.3), and *Equus caballus* (NM_001163866.1) and a partial sequence for the domestic cat, *F. catus* (NM_001009837.3), because *F. catus KIT* was located on the previous version of the assembly (60) (GeneScaffold_3098:168,162–233,592). Primers (Operon) were tested for efficient product amplification, and the final magnesium and temperature conditions for each primer pair are presented in Dataset S2 (S2.9). PCR and thermocycling conditions were conducted as previously described (61). The PCR products with the appropriate lengths were purified using the ExoSap (USB) enzyme per the manufacturer's recommendations. Purified genomic products were sequenced using BigDye Terminator Sequencing Kit version 3.1 (Applied Biosystems), purified with Illustra Sephadex G-50 (GE Healthcare) according to the manufacturer's recommendations, and electrophoretically separated on an ABI 3730 DNA Analyzer (Applied Biosystems). Sequences were verified and aligned using the software Sequencher version 4.8 (Gene Codes). The complete coding sequence for *F. catus KIT* was submitted to GenBank under accession number GU270865.1.

KIT mRNA analysis. RNA from a nonwhite control cat was isolated from whole blood using the PAXgene Blood RNA Kit (Qiagen) following the manufacturer's directions. The 5′ UTR amplification and the PCR analysis were conducted as previously described (62). The 5′ RACE used the cDNA pool generated by the KIT-specific primers (S2.9 in Dataset S2). The 5′ RACE PCR products were cloned using the TOPO TA Cloning Kit (Invitrogen) before sequencing. Five 5′ RACE cDNA clones from the control cat were selected and sequenced. Genomic primers (S2.9 in Dataset S2) were then designed in the 5′ UTR region to sequence the cats used for the genomic analysis of *KIT* (S2.7 in Dataset S2).

KIT SNP genotyping. An allele-specific PCR (AS-PCR) assay was designed for genotyping exon 6 SNPs (S2.9 in Dataset S2). Both allele-specific primer pairs annealed at the 2-nt primer–template mismatch (c.1035_1036delinsCA; p.Glu345Asp; His346Asn; S2.9 in Dataset S2). The AS-PCR assay used 1× buffer, 1.5 mM $MgCl_2$, 200 μM each dNTP, and 0.1 U Taq (Denville), per 15 μL of reaction mixture. The primer concentrations in each PCR were 0.67 μM KITgloA-FAM, 0.67 μM KITgloB-VIC, and 0.67 μM KITR. PCR conditions were: initial denaturation at 95 °C for 5 min, followed by 35 cycles of 95 °C for 30 s, 60 °C for 30 s, and 72 °C for 45 s, and a final extension step of 72 °C for 7 min. The amplified products were separated on an ABI 3730 DNA Analyzer (Applied Biosystems). The genotypes were scored based on fluorescence intensity using the software STRand (63). The variants and exons within *KIT* were schematically presented with FancyGene (64).

Exploring other potential regulatory variation within KIT. We initially planned to investigate only the exonic regions of *KIT*, even though flanking or intronic regions often regulate gene expression. Along this line of reasoning, an ∼7-kb retroviral (FERV1) insertion within *KIT* intron 1 was recently identified as the causative factor for white spotting among different cat breeds (65); however, the Birman breed was not surveyed for the insertion. We therefore searched for the dominant, white-spotting FERV1 insertion sequence in the pooled Birman genomic sequence data (with estimated 4× coverage). To do this, we aligned all ∼190 million 50-bp reads from the Birman pool to the 7,296-bp FERV1 insertion sequence and generated a consensus to compare with the FERV1 reference. A total of 778 reads aligned using BWA (37), but the result was ambiguous due to the following observations: (*i*) 1,169 bp (16%) of the FERV1 reference were missing across 23 regions, with an average of 50.8 bp missing per region; and (*ii*) there were regions of 275, 173, and 296 bp within the FERV1 reference with no read coverage. Instead, we designed a long-range PCR experiment (for primers and conditions, see ref. 65) to capture the white-spotting alleles in mitted and bicolor Ragdoll ($n = 10$), Birman ($n = 10$), and other white-spotted ($n = 5$) and solid ($n = 5$) cats. Whereas the FERV1 insertion was confirmed in spotted Ragdolls and other spotted cats, we found no evidence for the insertion in all Birman cats and solid cats. These results demonstrate a second mechanism for white spotting in the Birman breed while also confirming a separate mode of inheritance. Future experiments will investigate how the fixed mutations in *KIT* exon 6 interact with *KIT* regulatory elements during expression.

1. Miller JR, et al. (2008) Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics* 24(24):2818–2824.
2. Salzberg SL, et al. (2012) GAGE: A critical evaluation of genome assemblies and assembly algorithms. *Genome Res* 22(3):557–567.
3. Davis BW, et al. (2009) A high-resolution cat radiation hybrid and integrated FISH mapping resource for phylogenomic studies across Felidae. *Genomics* 93(4): 299–304.
4. Schwartz S, et al. (2003) Human-mouse alignments with BLASTZ. *Genome Res* 13(1): 103–107.
5. Kent WJ (2002) BLAT—The BLAST-like alignment tool. *Genome Res* 12(4):656–664.
6. Flicek P, et al. (2012) Ensembl 2012. *Nucleic Acids Res* 40(database issue):D84–D90.
7. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389–3402.
8. Li L, Stoeckert CJ, Jr, Roos DS (2003) OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res* 13(9):2178–2189.
9. Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW (2013) Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol Biol Evol* 30(8):1987–1997.
10. Coste B, et al. (2010) *Piezo1* and *Piezo2* are essential components of distinct mechanically activated cation channels. *Science* 330(6000):55–60.
11. Hou L, Arnheiter H, Pavan WJ (2006) Interspecies difference in the regulation of melanocyte development by *SOX10* and *MITF*. *Proc Natl Acad Sci USA* 103(24): 9081–9085.
12. Karolchik D, et al. (2004) The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* 32(Suppl 1):D493–D496.
13. Benson G (1999) Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res* 27(2):573–580.
14. Hach F, et al. (2010) mrsFAST: A cache-oblivious algorithm for short-read mapping. *Nat Methods* 7(8):576–577.
15. Alkan C, et al. (2009) Personalized copy number and segmental duplication maps using next-generation sequencing. *Nat Genet* 41(10):1061–1067.
16. Olshen AB, Venkatraman ES, Lucito R, Wigler M (2004) Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 5(4): 557–572.
17. Bailey JA, et al. (2002) Recent segmental duplications in the human genome. *Science* 297(5583):1003–1007.
18. Fontanesi L, et al. (2012) Exploring copy number variation in the rabbit (*Oryctolagus cuniculus*) genome by array comparative genome hybridization. *Genomics* 100(4):245–251.
19. Graubert TA, et al. (2007) A high-resolution map of segmental DNA copy number variation in the mouse genome. *PLoS Genet* 3(1):e3.
20. Chen W-K, Swartz JD, Rush LJ, Alvarez CE (2009) Mapping DNA structural variation in dogs. *Genome Res* 19(3):500–509.
21. Nicholas TJ, et al. (2009) The genomic architecture of segmental duplications and associated copy number variants in dogs. *Genome Res* 19(3):491–499.
22. Fontanesi L, et al. (2010) An initial comparative map of copy number variations in the goat (*Capra hircus*) genome. *BMC Genomics* 11:639.
23. Tanaka-Matsuda M, Ando A, Rogel-Gaillard C, Chardon P, Uenishi H (2009) Difference in number of loci of swine leukocyte antigen classical class I genes among haplotypes. *Genomics* 93(3):261–273.
24. Fontanesi L, et al. (2011) A first comparative map of copy number variations in the sheep genome. *Genomics* 97(3):158–165.
25. Liu GE, et al. (2010) Analysis of copy number variations among diverse cattle breeds. *Genome Res* 20(5):693–703.
26. Fadista J, Thomsen B, Holm L-E, Bendixen C (2010) Copy number variation in the bovine genome. *BMC Genomics* 11:284.
27. Shi P, Bielawski JP, Yang H, Zhang Y-P (2005) Adaptive diversification of vomeronasal receptor 1 genes in rodents. *J Mol Evol* 60(5):566–576.
28. Young JM, Massa HF, Hsu L, Trask BJ (2010) Extreme variability among mammalian *V1R* gene families. *Genome Res* 20(1):10–18.
29. Niimura Y, Nei M (2007) Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLoS ONE* 2(8):e708.
30. Donthu R, Lewin HA, Larkin DM (2009) SyntenyTracker: A tool for defining homologous synteny blocks using radiation hybrid maps and whole-genome sequence. *BMC Res Notes* 2:148.
31. Murphy WJ, O'Brien SJ (2007) Designing and optimizing comparative anchor primers for comparative gene mapping and phylogenetic inference. *Nat Protoc* 2(11): 3022–3030.
32. Katoh K, Toh H (2010) Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics* 26(15):1899–1900.
33. Posada D, Crandall KA (1998) MODELTEST: Testing the model of DNA substitution. *Bioinformatics* 14(9):817–818.
34. Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21):2688–2690.
35. Murphy WJ, et al. (2001) Molecular phylogenetics and the origins of placental mammals. *Nature* 409(6820):614–618.
36. Chen K, Durand D, Farach-Colton M (2000) NOTUNG: A program for dating gene duplications and optimizing gene family trees. *J Comput Biol* 7(3-4):429–447.
37. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14):1754–1760.
38. Abyzov A, Urban AE, Snyder M, Gerstein M (2011) CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res* 21(6):974–984.
39. Woolley S, Johnson J, Smith MJ, Crandall KA, McClellan DA (2003) TreeSAAP: Selection on amino acid properties using phylogenetic trees. *Bioinformatics* 19(5):671–672.
40. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP (2012) Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE* 7(10):e46688.
41. Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3(5):418–426.
42. Shi P, Zhang J (2006) Contrasting modes of evolution between vertebrate sweet/-umami receptor genes and bitter receptor genes. *Mol Biol Evol* 23(2):292–300.
43. Ishii Y, et al. (2008) Mutations in R-spondin 4 (*RSPO4*) underlie inherited anonychia. *J Invest Dermatol* 128(4):867–870.

44. Khan TN, et al. (2012) Novel missense mutation in the *RSPO4* gene in congenital hyponychia and evidence for a polymorphic initiation codon (p.M1I). *BMC Med Genet* 13:120.
45. Axelsson E, et al. (2013) The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495(7441):360–364.
46. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
47. Koboldt DC, et al. (2012) VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 22(3):568–576.
48. Danecek P, et al.; 1000 Genomes Project Analysis Group (2011) The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158.
49. Alhaddad H, et al. (2013) Extent of linkage disequilibrium in the domestic cat, *Felis silvestris catus*, and its breeds. *PLoS ONE* 8(1):e53537.
50. Purcell S, et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81(3):559–575.
51. Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19(9):1655–1664.
52. Jakobsson M, Rosenberg NA (2007) CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23(14):1801–1806.
53. Rosenberg NA (2003) DISTRUCT: A program for the graphical display of population structure. *Mol Ecol Notes* 4(1):137–138.
54. Albert FW, et al. (2009) Genetic architecture of tameness in a rat model of animal domestication. *Genetics* 182(2):541–554.
55. Cooper MP, Fretwell N, Bailey SJ, Lyons LA (2006) White spotting in the domestic cat (*Felis catus*) maps near *KIT* on feline chromosome B1. *Anim Genet* 37(2): 163–165.
56. Menotti-Raymond M, et al. (1999) A genetic linkage map of microsatellites in the domestic cat (*Felis catus*). *Genomics* 57(1):9–23.
57. Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci USA* 81(11):3443–3446.
58. Schäffer AA (1996) Faster linkage analysis computations for pedigrees with loops or unused alleles. *Hum Hered* 46(4):226–235.
59. Young AE, Biller DS, Herrgesell EJ, Roberts HR, Lyons LA (2005) Feline polycystic kidney disease is linked to the *PKD1* region. *Mamm Genome* 16(1):59–65.
60. Pontius JU, et al.; Agencourt Sequencing Team; NISC Comparative Sequencing Program (2007) Initial sequence and comparative analysis of the cat genome. *Genome Res* 17(11):1675–1689.
61. Bighignoli B, et al. (2007) Cytidine monophospho-N-acetylneuraminic acid hydroxylase (CMAH) mutations associated with the domestic cat AB blood group. *BMC Genet* 8:27.
62. Gandolfi B, et al. (2012) First *WNK4*-hypokalemia animal model identified by genome-wide association in Burmese cats. *PLoS ONE* 7(12):e53173.
63. Toonen RJ, Hughes S (2001) Increased throughput for fragment analysis on an ABI PRISM 377 automated sequencer using a membrane comb and STRand software. *Biotechniques* 31(6):1320–1324.
64. Rambaldi D, Ciccarelli FD (2009) FancyGene: Dynamic visualization of gene structures and protein domain architectures on genomic loci. *Bioinformatics* 25(17):2281–2282.
65. David VA, et al. (2014) Endogenous retrovirus insertion in the *KIT* oncogene determines white and white spotting in domestic cats. *G3 (Bethesda)*, 10.1534/g3.114.013425.

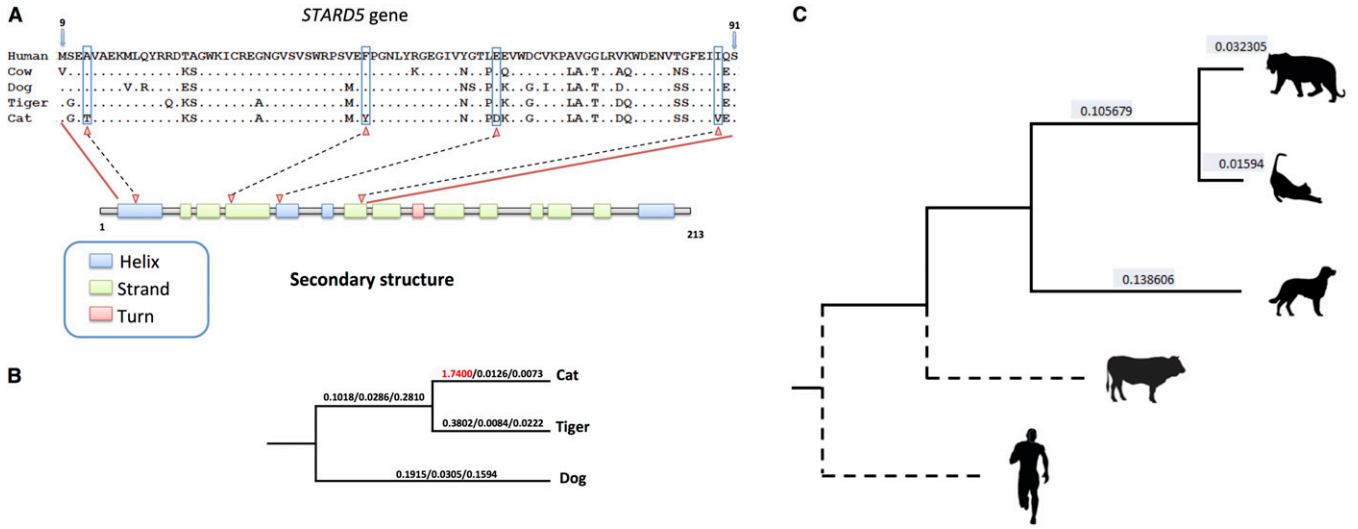

Fig. S1. (*A*) Predicted structure of the domestic cat *STARD5* gene. Positively selected amino acid sites are indicated with red arrowheads. (*B*) Results of the d_N/d_S test suggest an accelerated evolutionary rate of the *STARD5* gene on the domestic cat branch. Numbers on each branch are scores of the estimated d_N/d_S, d_N, and d_S. (*C*) Average synonymous mutation rates along branches used for assessments of positive selection. Dashed lines indicate relationships since rates are not reported for cow and human.

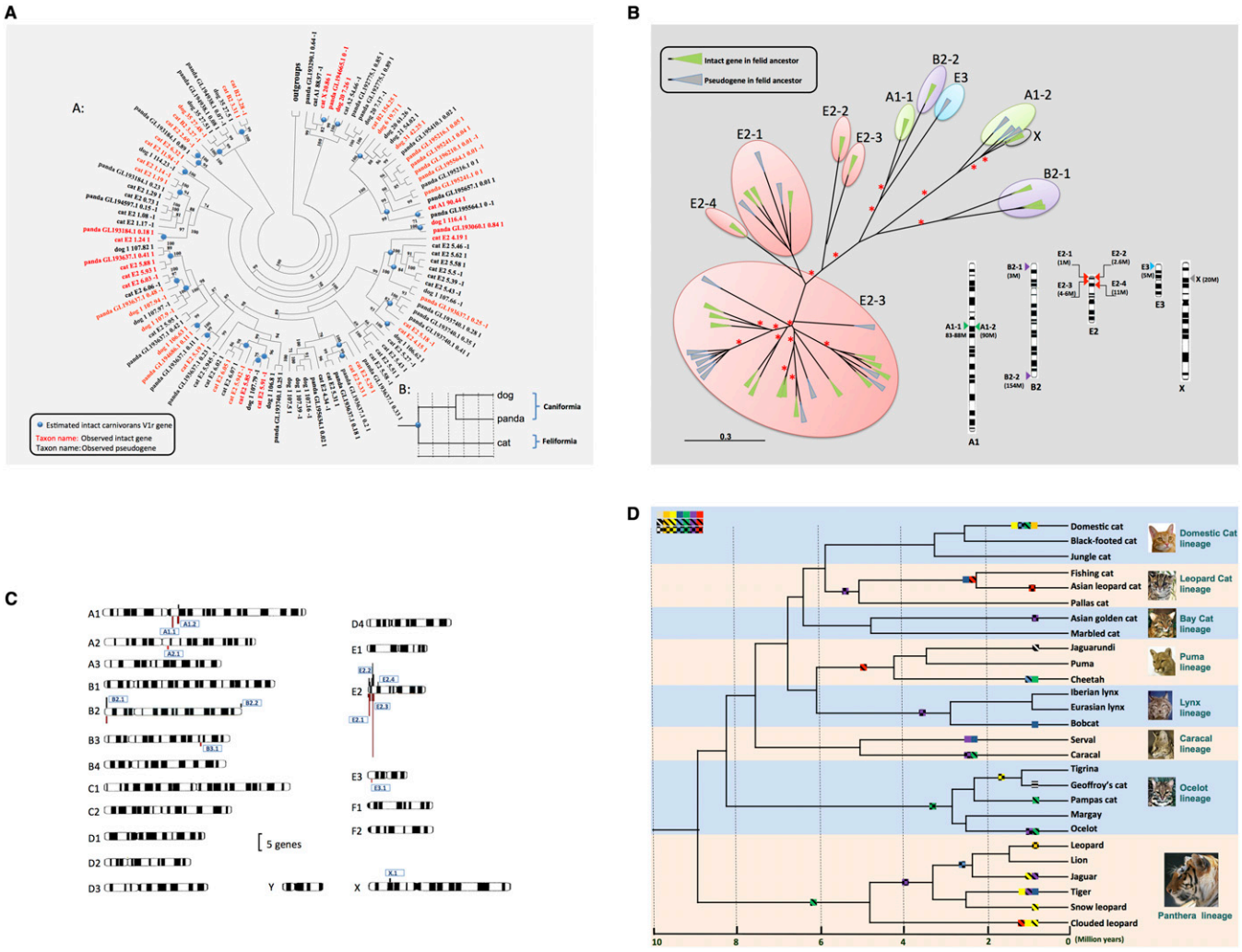

Fig. S2. (*A*) Maximum likelihood phylogenetic tree of functional *V1r* genes and long pseudogenes from the domestic cat, giant panda, and dog genomes. (*B*) Gene tree of the *V1r* supergene family determined for 35 feline species suggesting the early expansion of *V1r* genes in the common ancestor of felids. (*C*) Distribution of the *V1r* gene family in the domestic cat genome. Intact genes are denoted in black; pseudogenes are denoted in red. (*D*) Detected *V1r* gene loss among different lineages of the Felidae. Colored boxes in the tree indicate gene loss events based on 48 putatively intact *V1r* genes present in the common ancestor of all current felids.

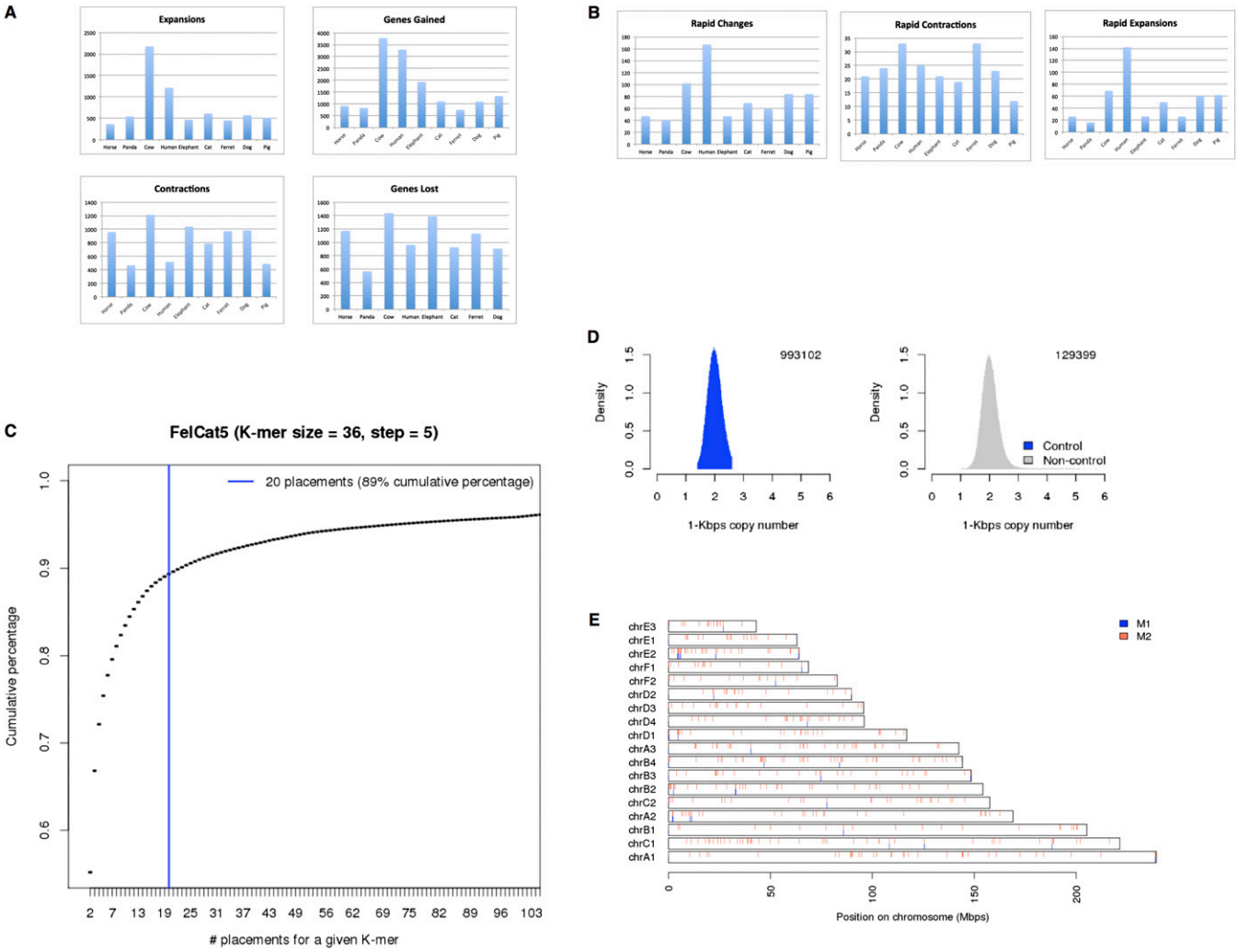

Fig. S3. (*A*) Raw counts for the number of gene family expansions, number of genes gained, number of gene family contractions, and number of genes lost among horse, panda, cow, human, elephant, cat, ferret, dog, and pig. (*B*) Significant results for the number of rapid gene family expansions and the number of rapid gene family contractions among horse, panda, cow, human, elephant, cat, ferret, dog, and pig. (*C*) Cumulative distribution of additional masking achieved by masking overrepresented K-mers in Fca 6.2 (FelCat5 in UCSC). (*D*) Distribution of 1-kbp copy number values in control and noncontrol regions. The number of windows in each distribution is indicated. (*E*) CNV map of expansions on domestic cat autosomes.

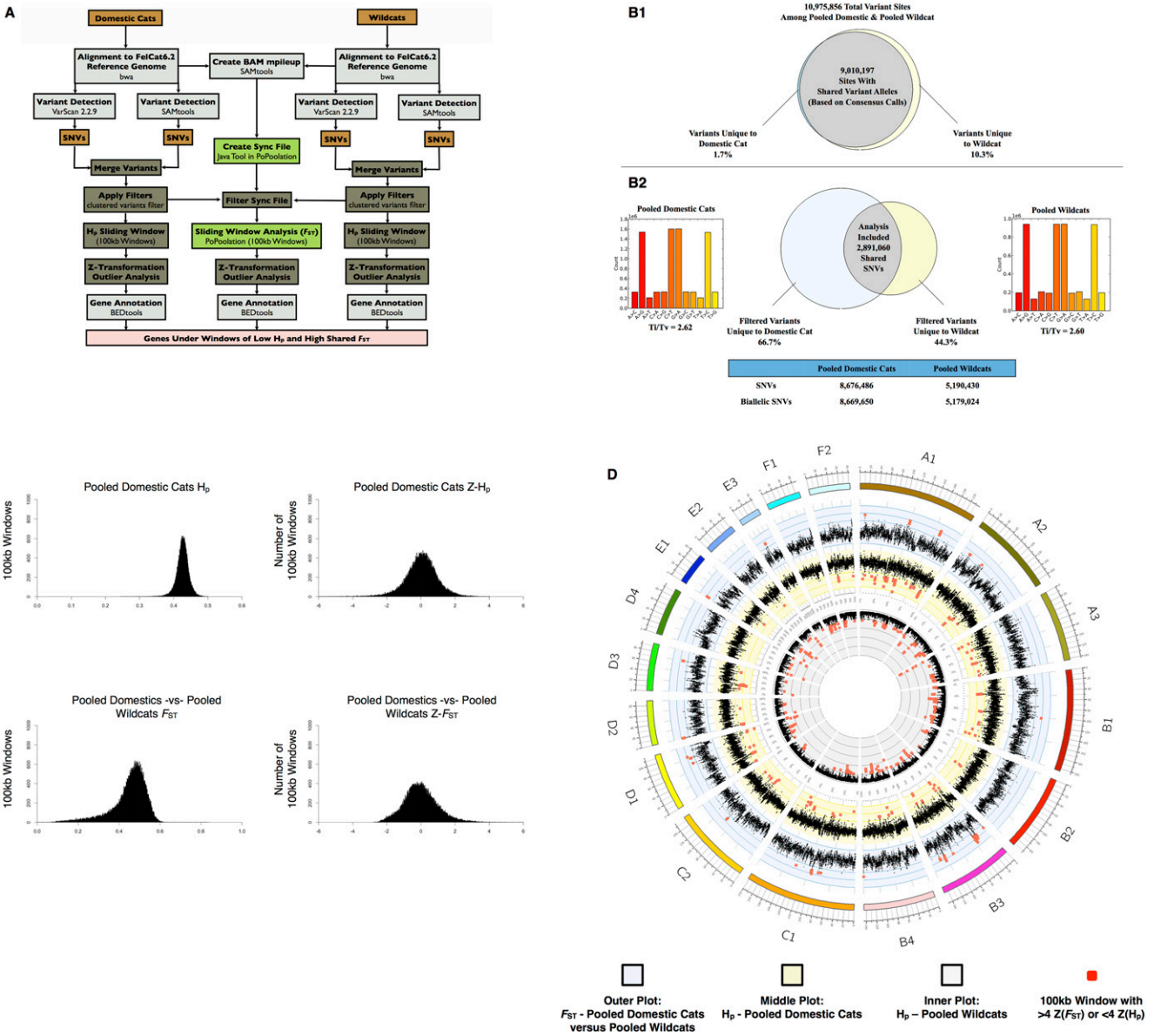

Fig. S4. (*A*) Analysis pipeline for determining putative regions of selection using variant data. (*B*) Summary results of variant calling for pooled domestic cats ($n = 22$) and pooled wildcats ($n = 4$). (*B1*) Percentage of overlapping variant alleles at each of the sites where a high-quality variant was detected. (*B2*) Percentage of unique and overlapping variant sites included in the sliding window analysis comparing domestic cats with wildcats based on stringent filtering parameters. Also included are transition:transversion ratios per pool as well as counts of variant types per pool. (*C*) Distribution of pooled heterozygosity, H_P, and average fixation index, F_{ST}, and corresponding Z transformations, $Z(H_P)$ and $Z(F_{ST})$, estimated in 100-kb windows across all cat autosomes. (*D*) Circos plot of (*i*) pooled domestic cat versus pooled wildcat F_{ST}, (*ii*) pooled domestic cat H_p, and (*iii*) pooled wildcat H_p results for each 100-kb window (with a step size of 50 kb) along each chromosome. Windows with elevated F_{ST} or depressed H_p are depicted as red dots, whereas all other windows are depicted as black dots.

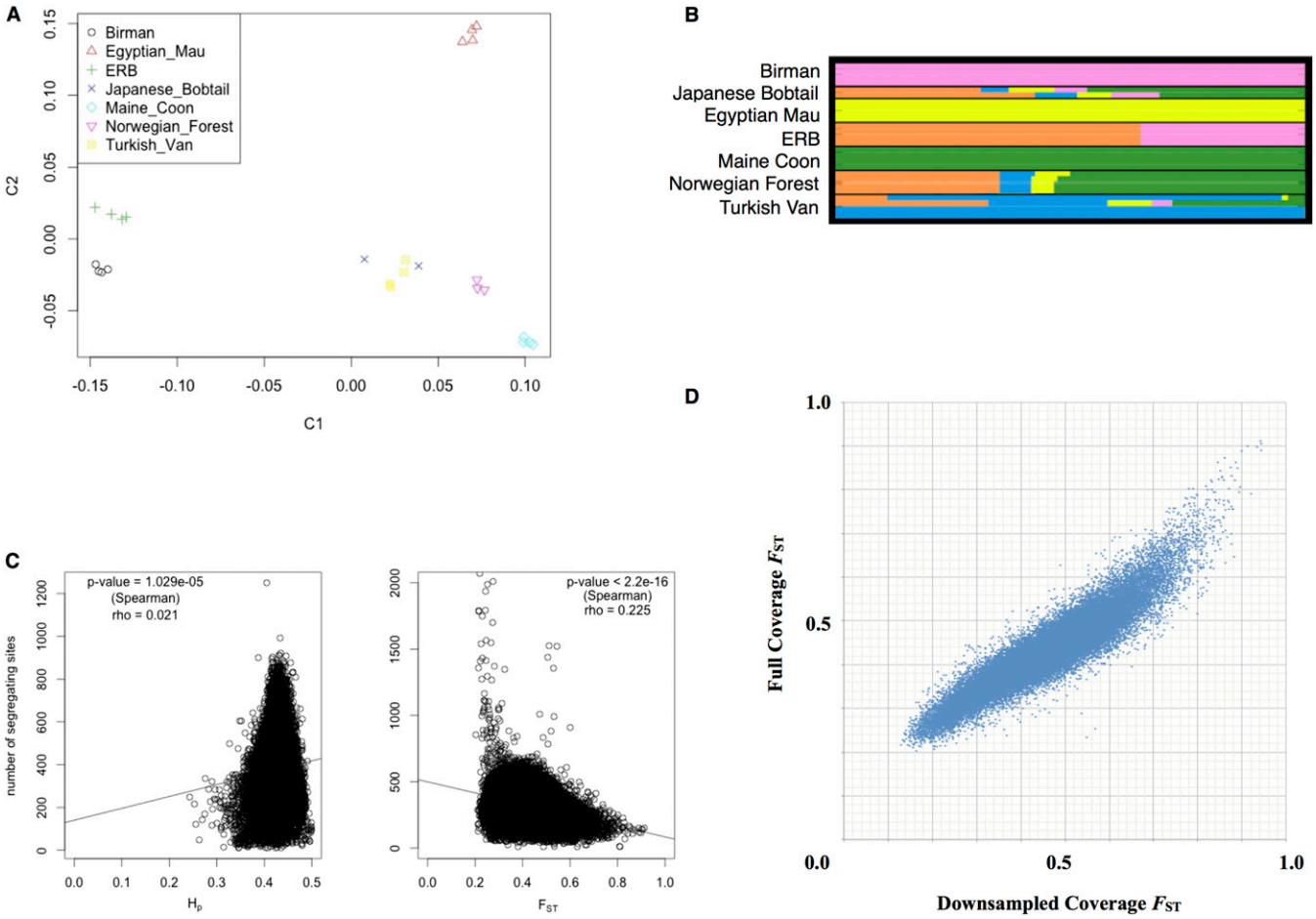

Fig. S5. (*A*) MDS plot depicting the relationship between individuals within the seven domestic cat pools used for the analysis of breed differentiation. (*B*) Admixture results for K = 5 showing genetic differentiation between eastern (Birman) and western (Maine Coon) populations, with moderate admixture between other breeds, including eastern random bred (ERB) individuals. (*C*) The average H_P and F_{ST} of all autosomal 100-kb windows plotted against the corresponding number of segregating sites per window. H_P is positively correlated with the number of segregating sites, whereas F_{ST} is negatively correlated. (*D*) The F_{ST} results for all autosomal 100-kb windows for the full coverage (~55×) pooled domestic analysis (*x* axis) are plotted against the F_{ST} results for all autosomal 100-kb windows for the subsampled (~7×) pooled domestic analysis (*y* axis).

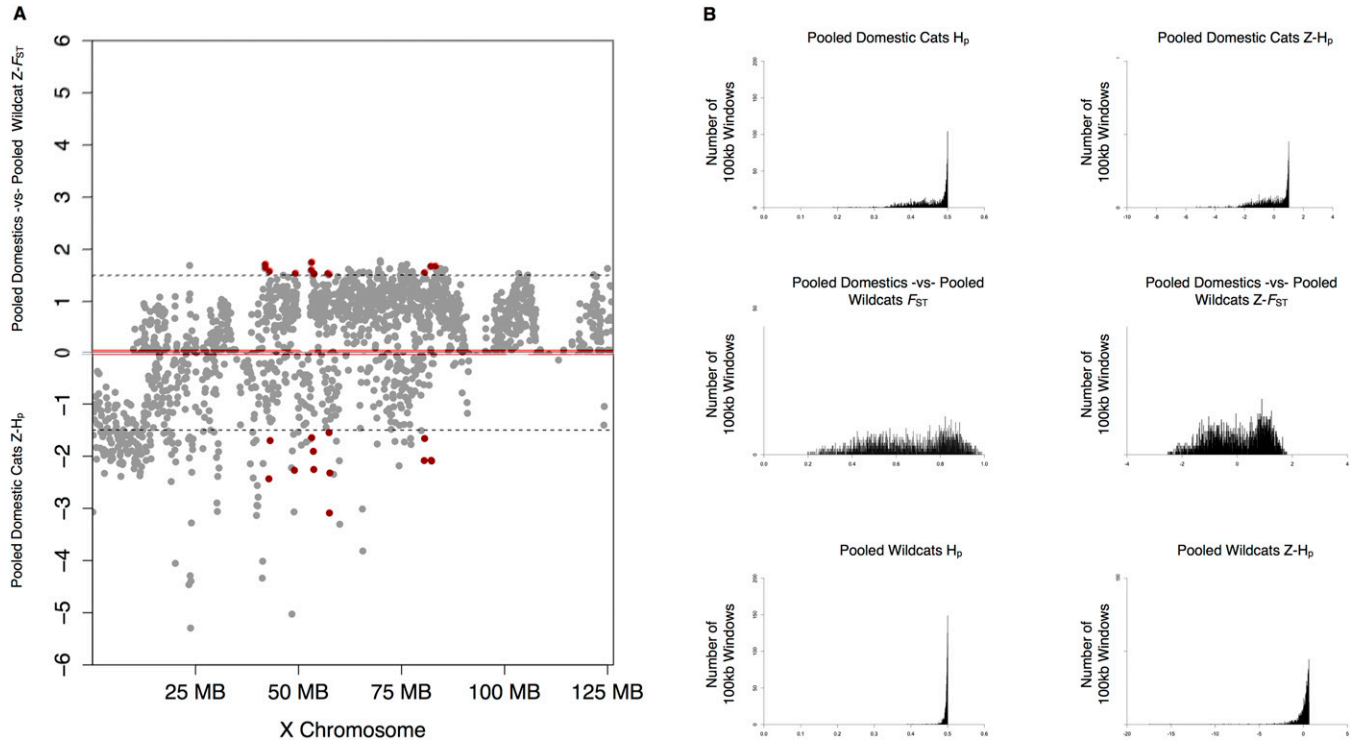

Fig. S6. (*A*) Z-transformed average fixation index (only positive values are shown), Z(F_{ST}), and pooled heterozygosity (only negative values are shown), Z(H_p), in 100-kb windows across chromosome X. Red dots indicate windows with (*i*) high F_{ST} and low H_p along with (*ii*) underlying gene content. (*B*) Distribution of pooled heterozygosity, H_p, and average fixation index, F_{ST}, and corresponding Z transformations, Z(H_p) and Z(F_{ST}), estimated in 100-kb windows across chromosome X.

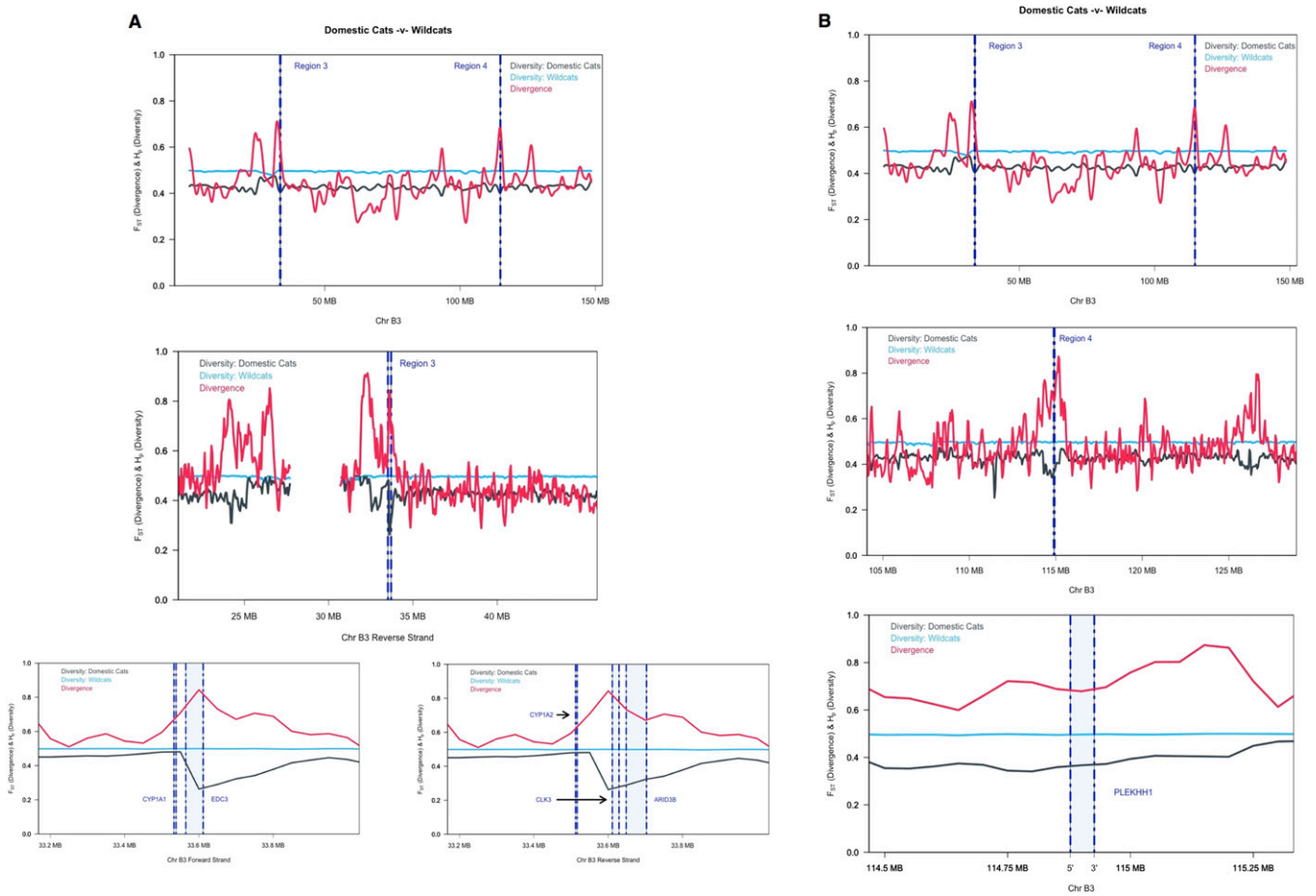

Fig. S7. (*A*) Plots of (*i*) pooled domestic cat versus pooled wildcat F_{ST} results (red), (*ii*) pooled domestic cat H_p results (light blue), and (*iii*) pooled wildcat H_p results (dark blue) for each 100-kb window (with a step size of 50 kb) along chromosome B3 with increasing resolution to the genes underlying region 3. (*B*) Plots of (*i*) pooled domestic cat versus pooled wildcat F_{ST} results (red), (*ii*) pooled domestic cat H_p results (light blue), and (*iii*) pooled wildcat H_p results (dark blue) for each 100-kb window (with a step size of 50 kb) along chromosome B3 with increasing resolution to the genes underlying region 4.

Other Supporting Information Files

Dataset S1 (PDF)
Dataset S2 (PDF)

Dataset S1.1(a). List of positively selected genes identified by the two-branch and branch-site tests in the Carnivore lineage

A4GALT	APH1B	C9orf123	CD84	COX7A1	DNHD1	FAM195B	GPRC5A	IWS1	LRRC43
AASDH	APOB	C9orf89	CDC26	CPD	DOK1	FAM83C	GTF3A	JAK2	LRRC63
ABCG2	APOBEC1	CA14	CDH16	CPLX4	DPH2	FAN1	GUCA1B	JARID2	LRRK2
ACR	ARHGAP18	CABP4	CDH17	CRTAP	DPM2	FECH	HAUS8	KARS	LSS
ADAM28	ARHGAP31	CACNB2	CDH24	CTSA	DQX1	FGF23	HEATR2	KCNJ15	LTF
ADH4	ARMC3	CAPZA3	CDSN	CX3CR1	DSG2	FGFBP1	HEPHL1	KCNK7	LY6E
ADH4	ARMCX1	CARD6	CECR1	CXADR	DYNLRB2	FMO3	HJURP	KCNN4	LYVE1
AEN	ATF6	CATSPER2	CELA1	CXCR6	EDC4	FN3KRP	HMGCL	KCTD3	MANBA
AGA	ATP2C2	CATSPER4	CEP152	CXorf30	EDN1	FOXRED1	HOOK1	KIAA1143	MANEA
AGFG2	AXIN1	CC2D1B	CEP250	CYGB	EIF2B3	FSD1	IGLON5	KIAA1984	MAP1B
AHSG	B4GALT4	CC2D2A	CHCHD2	CYP27A1	EIF3I	FSD2	IL1R1	KLF4	MARCH8
AKAP1	BCAS1	CCBL2	CHM	DAG1	EMR1	FTO	IL4I1	KLHL22	MARCO
AKAP12	BCL3	CCDC108	CHRM5	DBNDD2	ENAH	GALNT14	IL7	KRTAP11-1	MAT2B
AMPD1	BLZF1	CCDC117	CKAP2L	DCAF5	ENOSF1	GALNT5	ILF2	LAMA3	MBD1
ANGPT2	BPI	CCDC137	CLCC1	DCDC2B	ENPP4	GALT	IMPG1	LAX1	MDM1
ANGPTL2	BRCA2	CCDC15	CLDN8	DCK	ENPP7	GAPDHS	INPP5A	LBR	MED17
ANK3	BRIP1	CCDC40	CMYA5	DCLRE1B	ERAL1	GCAT	INSL6	LCA5	MKKS
ANKS1A	BRWD1	CCDC67	CNGB	DDRGK1	F10	GGCT	IQGAP3	LCLAT1	MLH1
ANKS6	C15orf52	CCKBR	CNGB3	DGCR2	F5	GGH	IRF8	LETM1	MMP24
ANPEP	C18orf32	CCNF	COG7	DHRS9	FAM101B	GLB1	ITGA8	LGR6	MORN1
ANXA1	C18orf54	CCR6	COL15A1	DIAPH3	FAM111B	GLB1L	ITGAM	LPXN	MRPL2
ANXA13	C2orf62	CCR8	COL23A1	DIRAS1	FAM129B	GOLM1	ITGB1BP2	LRP2	MRPL28
ANXA5	C4orf47	CD47	COL6A1	DLEC1	FAM149A	GPR111	ITGB2	LRRC15	MRPL41
ANXA7	C9	CD63	COL9A3	DNAJC12	FAM188B	GPR179	ITPR2	LRRC32	MRPL51

Dataset S1.1(b). List of positively selected genes identified by the two-branch and branch-site tests in the Carnivore lineage

MRPS12	NMI	PDE4C	PPP1R3A	RELL2	SHARPIN	SNAPC4	THADA	TP53BP1	VRK2
MSMP	NODAL	PDE6B	PRC1	RGS4	SHC4	SNAPC5	THBS3	TPX2	WFDC11
MSR1	NOL7	PDILT	PRF1	RHBDD1	SIGIRR	SNCAIP	THSD1	TRA2A	YIPF3
MUM1L1	NOV	PEX6	PRICKLE2	RHCG	SKAP2	SNRNP25	TK2	TRAF7	ZC3HAV1
MUTYH	NSD1	PGLYRP1	PSD3	RHEBL1	SLC10A4	SOAT1	TLR2	TRDMT1	ZMYM3
MYD88	NSUN5	PGM2	PSMD6	RHOT2	SLC12A4	SOCS6	TLR4	TREM2	ZNF331
MYH8	NUP153	PHF15	PTCD2	ROS1	SLC13A2	SPATA7	TLR6	TRIM25	ZNF398
MYO15A	OAZ3	PIBF1	PTGR1	RRAGD	SLC15A1	SPTA1	TLR8	TRIM33	ZNF473
MYO3B	OAZ3	PIGQ	PTPRC	RRS1	SLC16A5	SRGN	TMCO6	TSEN2	ZNF687
MYO7A	OBFC1	PIK3CB	PTPRH	RSPO4	SLC22A23	SS18L1	TMEM109	TSHZ2	ZNF777
NANOS3	OIT3	PITPNA	PXN	RTBDN	SLC22A8	ST3GAL1	TMEM116	TSPAN10	ZSWIM5
NBEAL1	OLFM1	PITX1	PYCARD	RTN4	SLC25A23	STK24	TMEM150B	TSPYL4	
NDOR1	OR10V1	PKHD1	QSER1	RTP3	SLC25A42	SUCLG1	TMEM156	TTC34	
NDST3	OR13H1	PKMYT1	RAB11FIP5	SCGB1C1	SLC29A2	TAL2	TMEM167B	TTF2	
NDUFA6	OSGEP	PLA2G2F	RAB18	SCML2	SLC2A10	TAS2R38	TMEM176A	TUBGCP3	
NDUFAF4	OSMR	PLBD1	RAB19	SCN3B	SLC44A4	TBC1D21	TMEM182	UBA7	
NDUFV2	OTOF	POC1B	RAD52	SCNM1	SLC47A1	TBL3	TMEM215	UBE2L6	
NFAM1	OTUB2	PPA2	RAG1	SDR39U1	SLC4A1	TBXAS1	TMIGD1	UGT2A3	
NGLY1	OVCA2	PPEF1	RANGAP1	SEC61A2	SLC4A5	TCN2	TMOD1	UMOD	
NGRN	OXCT1	PPID	RASAL1	SEL1L2	SLC6A4	TEP1	TMX1	UNC13D	
NID1	PADI2	PPL	RASSF5	SELP	SLC7A1	TFAM	TNFRSF13B	UPRT	
NIN	PCDH12	PPM1E	RCSD1	SEPT12	SLC7A4	TFB2M	TNKS1BP1	UTP11L	
NKTR	PCNXL2	PPM1K	RECQL4	SERPINB9	SLCO1C1	TG	TOM1	VAMP8	
NLRP14	PDCL3	PPP1R2	REEP1	SH2D2A	SMPDL3B	TGM6	TOX4	VPS13C	

Dataset S1.2(a). List of positively selected genes identified by the two-branch and branch-site tests in the Felidae lineage

ABCA5	BAK1	C5	CHRD	DNAJB4	GMIP	IPO7	LRRC14B	N4BP2	PRX
ACADL	BBS7	CA4	CLEC5A	DNHD1	GPA33	IQCH	LRRC6	NIF3L1	PSMB8
ACCS	BBS9	CALML5	CLUL1	DUSP2	GPAA1	IRF8	LRRTM2	NOLC1	PSME3
ACSF3	BCAT2	CAPN13	CMTM2	DYSF	GPRIN3	IRS4	LTA4H	NPNT	PSMG3
ADAM22	BCL2L14	CATSPER3	CMYA5	E2F7	GRIA2	ISG15	LY9	NPY	PTPRC
ADC	BCL2L15	CCDC107	CNKSR1	ECM2	GRIA2	ITGAE	MAMDC2	NPY1R	PTPRH
AHSP	BIN1	CCDC112	CRTAM	EFCAB2	GRIN2C	ITGB7	MAP7D3	NTRK1	PTPRN
AIM1L	BIRC3	CCDC113	CST7	EHBP1L1	GSDMC	ITIH4	MAPK8IP2	NUBP2	PTPRQ
AKAP9	BMF	CCDC150	CTSZ	EHHADH	GTPBP8	ITPR3	MAPKBP1	NUDCD3	RAB20
AKNA	BMP15	CCNE2	CTTNBP2NL	EPHX1	GUCA1A	KIF1A	MCM7	NUDT22	RASGRP1
AKNAD1	BMPR2	CD200	CXorf23	FA2H	HAUS5	KREMEN2	MECR	OAZ3	RBM28
ALB	BPI	CD244	CXorf57	FAIM3	HCFC2	L1CAM	MEIS1	OMA1	RELA
ALDH1A2	BRD7	CD274	CYP17A1	FAM161B	HDGF	LARS2	MKNK2	OOEP	RIBC1
ALG3	BRIP1	CD48	CYP1A2	FAM181A	HFM1	LAT2	MORN3	PARP2	RNF141
ALPK2	BRWD1	CD8B	DCST2	FBN3	HHIPL2	LAX1	MRPL50	PC	RNF217
ANGPT2	C10orf137	CD97	DDO	FIGF	HSD17B14	LCNL1	MRPL55	PFN2	RNPEP
ANKS4B	C12orf56	CDC25B	DDX49	FKBP3	HSPBP1	LGALS2	MSGN1	PIK3C2G	ROS1
ANPEP	C14orf166B	CDH1	DEPDC1	FKBP4	IFNK	LIMS2	MTRR	PLAC8L1	RPUSD4
AP3B2	C17orf64	CDH17	DEPDC7	FKBP7	IGF1	LIN28B	MUTYH	PNLIP	RSPH6A
ARF4	C1orf146	CDH5	DHRS1	GAP43	IGFBP5	LMBRD2	MVK	PPAPDC1A	S100A12
ARMCX2	C1orf194	CDKN1B	DHX32	GGA3	IL17RB	LMF1	MYH8	PPP1R13L	SACS
ATE1	C1QB	CEACAM18	DLGAP5	GJA10	IL22	LONRF3	MYO15A	PRF1	SCAMP2
ATP2B3	C2orf43	CENPE	DMP1	GJA5	INHBB	LPAR5	MYO1F	PROM1	SCAP
AZGP1	C3orf62	CES2	DNAH8	GLIPR1L2	INVS	LRAT	MYO7A	PRRG3	SCD5

Dataset S1.2(b). List of positively selected genes identified by the two-branch and branch-site tests in the Felidae lineage

SCG2	SLC16A5	SNCG	STARD13	TDG	TMEM176A	TRAT1	UBE2S	WRN	ZSWIM2
SCGB1A1	SLC1A7	SNRNP70	STARD3	TEX11	TMEM19	TRMU	UCHL1	XKR7	
SDC2	SLC27A1	SPATA7	STK31	TFAP2A	TMEM190	TRPM4	UMPS	XRCC5	
SELL	SLC2A4	SPEF2	STOX1	TGM7	TMEM211	TRPS1	UNC93B1	YIF1B	
SERHL2	SLC38A8	SPHK1	SVEP1	THTPA	TMX3	TSPAN8	USHBP1	ZFYVE16	
SF3A2	SLC43A3	SPTBN4	SYDE2	TM6SF2	TNFAIP3	TSSK4	VTI1A	ZNF304	
SFTPB	SLC7A11	SPTBN5	SYNM	TMEM140	TNIP2	TTC34	WDFY4	ZNF408	
SH2D2A	SLCO2B1	SPTLC3	SYTL3	TMEM150B	TNIP3	TTC39C	WFDC8	ZNF780B	
SIPA1L2	SMOC2	SRCRB4D	TAPBPL	TMEM156	TRA2A	TTYH1	WHAMM	ZNF804B	
SIT1	SNAPC3	STAM	TCF3	TMEM161B	TRAF3IP2	TUSC5	WIPF2	ZSCAN29	

Dataset S1.3(a). List of positively selected genes identified by the two-branch and branch-site tests in the Felinae lineage

ABHD1	BRAF	CENPM	ENKUR	GPR174	ITGA9	MIIP	OR2B11	PSPH	SERINC3
ACOT11	BRCA1	CEP68	ENTPD7	GPRASP2	ITGBL1	MORC1	OR4D6	PSTK	SH2D5
ACOT13	C11orf54	CEP97	EPHB4	GPRC5A	ITPR3	MRPL11	OTOF	PTPRR	SHC4
ACOT8	C11orf63	CHMP4B	ETV4	GPRIN2	JMJD1C	MRPL52	PAFAH2	PTPRS	SIAE
ACOX2	C16orf71	CIB4	FAIM3	GRHL3	KIAA0226	MTIF2	PARVG	PUSL1	SLC22A13
ACOX3	C1orf109	CLDN17	FAM131B	GRIA2	KIF1C	MTRF1	PCDHB4	RABL3	SLC22A18
ADAMDEC1	C22orf31	CLEC5A	FAM179A	HADH	KIF22	MURC	PHLDB3	RBM11	SLC25A38
ADAMTS13	C2orf40	CNGA2	FAM69A	HEATR5B	KIF27	MVK	PITRM1	RBP5	SLC35F5
ADAMTSL3	C2orf62	COL6A3	FANCB	HECA	KIRREL2	MYLK3	PJA2	RCSD1	SLC39A7
AK1	C3orf62	COL9A3	FAT4	HEPACAM2	KRIT1	MYO15A	PLA2G2E	RELL1	SLC39A8
ALDH16A1	C8B	CROCC	FBN3	HEPH	KYNU	MYO9A	PLA2G3	RHBDD1	SLC46A1
ALS2CR12	C9orf96	CSPP1	FBXL22	HMMR	LAMC2	NAPRT1	PLAC1	RIMKLA	SLCO1A2
AMACR	CAGE1	CTTN	FBXO28	HPS5	LAP3	NEK1	PLAC8L1	RNASE6	SMG1
ANKRD2	CASP7	CYB5R1	FER	HSD3B7	LATS2	NEK4	PLIN3	RNPC3	SMG6
ANKRD49	CBX2	CYP27B1	FGA	HSPA13	LCAT	NFAM1	PML	RSL1D1	SPATA21
ANKRD50	CCDC38	DACT1	FN3K	IFT81	LIAS	NFKBIZ	PPAP2A	RTP3	SPATA7
APEH	CCDC64B	DAPK1	FRMD7	IGHMBP2	LIMD1	NOLC1	PPAPDC1B	S100A12	SPERT
APOBEC4	CCDC70	DNAJB9	GCNT7	INHBC	LRRC32	NOSTRIN	PPFIBP1	SCN9A	SPHKAP
ARHGAP26	CD27	DNTTIP2	GEMIN7	INPP4B	LRRC36	NOTCH2	PPP1R13B	SCRIB	SPINT1
ASB11	CD48	DPEP3	GGT6	INPP5J	LSM3	NPFFR2	PRICKLE4	SDK2	SPTBN5
ATXN7L1	CD93	DUSP19	GOLGA1	IPO4	MAP7D2	NRG2	PRKAG1	SEC24A	SREBF1
BARD1	CDH6	ECHDC1	GPATCH8	IQCB1	MARVELD3	NUDT15	PRKG2	SENP5	SRRM2
BCAP31	CELA1	EDC3	GPR133	ISG15	MERTK	OPTC	PRR11	SENP7	STARD5
BPI	CENPE	EHBP1L1	GPR15	ITGA2B	METTL8	OR10K1	PRX	SEPT10	STK11IP

Dataset S1.3(b). List of positively selected genes identified by the two-branch and branch-site tests in the Felinae lineage

STS	TAS2R1	THUMPD1	TRPV6	UBXN10	WDR62	WWC1	ZMYND10	ZZEF1
SUN3	TAS2R3	TMEM59L	TSTD2	USP45	WDR90	XCR1	ZNF436	
SURF2	TEX14	TMEM71	TXN2	UVRAG	WFDC8	XPC	ZNF555	
SYNM	TF	TOE1	TXNRD2	VEZT	WIPF2	ZFAT	ZNF622	
SYTL1	THBS2	TP53BP1	TYK2	WDR17	WIPF3	ZFYVE19	ZNF780B	

Dataset S1.4(a). Predicted structural/functional influence of the domestic cat nonsynonymous substitutions for positively selected sensory and lipid metabolism genes

Gene Name	Number of Significant Amino Acid Properties	Identified Categories	Intense Protein Functional Changes	Number of Suggested Deleterious Amino Acid Substitutions
ABHD1	2	9,22	negative	0
ACOT11	3	12,17,26	negative	0
ACOT8	1	9	negative	0
ACOX2	5	7,10,12,15,31	negative	0
ACOX3	**6**	**4,10,12,15,17,21**	**positive**	**2**
AMACR	**5**	**10,17,24,30,31**	**positive**	**4**
BARD1	2	9,12	negative	0
BBS7	0		negative	0
BBS9	2	3,7	negative	0
BRAF	**2**	**9,22**	**positive**	**1**
BRCA1	**28**	**1-9,11-15,17-26,28-31**	**positive**	**11**
CA4	**2**	**13,27**	**positive**	**1**
CABP4	0		negative	0
CDKN1B	0		negative	0
CHM	4	1,4,11,12	negative	0
CNGA2	**2**	**10,15**	**positive**	**1**
CNGB3	**1**	**31**	**positive**	**1**
COL6A3	**31**	**1-31**	**positive**	**1**
COL9A3	**4**	**13,15,17,31**	**positive**	**1**
CPLX4	0		negative	0
CYP27B1	1	17	negative	0
GJA10	0		positive	1
GRIA2	**1**	**2**	**positive**	**2**
GRIN2C	2	9,17	negative	0
GUCA1A	0		negative	0
GUCA1B	0		negative	0
HADH	1	28	negative	0
HMMR	**5**	**1,7,9,15,17**	**positive**	**2**
HSD3B7	2	9,15	negative	0
IMPG1	0		negative	0
INPP5J	4	1,2,4,12	negative	0
IQCB1	2	15,22	negative	0
ITGA2B	14	3,6,7,8,10,12,16,17,19,22,24,28,29,31	negative	0
ITGA9	3	1,3,15	negative	0
LAMC2	**14**	**3,7-10,12,13,15,16,17,19,26,29,31**	**positive**	**1**
LCAT	0		negative	0
LRAT	0		negative	0

Significant genes common to both approaches are highlighted in red.
a- TreeSAAP is used to measure structural and biochemical properties of amino acid replacement using a threshold of P<0.001. 31 categories are tested as follows: 1. Alpha-helical tendencies, 2. Average number of surrounding residues, 3. Beta-structure tendencies, 4. Bulkiness, 5. Buriedness, 6. Chromatographic index, 7. Coil tendencies, 8. Composition, 9. Compressibility, 10. Equilibrium constant, 11. Helical contact area, 12. Hydropathy, 13. Isoelectric point, 14. Long-range non-bonded energy, 15. Mean r.m.s. fluctuation displacement, 16. Molecular volume, 17. Molecular weight, 18. Normalized consensus hydrophobicity, 19. Partial specific volume, 20. Polar requirement, 21. Polarity, 22. Power to be at the C-terminal, 23. Power to be at the middle of alpha-helix, 24. Power to be at the N-terminal, 25. Refractive index, 26. Short and medium range non-bonded energy, 27. Solvent accessible reduction ratio, 28. Surrounding hydrophobicity, 29. Thermodynamic transfer hydrohphobicity, 30. Total non-bonded energy, 31. Turn tendencies
b - Amino acid substitutions labeled as "deleterious" based on Provean.

Dataset S1.4(b). Predicted structural/functional influence of the domestic cat nonsynonymous substitutions for positively selected sensory and lipid metabolism genes

Gene Name	Number of Significant Amino Acid Properties	Identified Categories	Intense Protein Functional Changes	Number of Suggested Deleterious Amino Acid Substitutions
MERTK	2	9,17	negative	0
MKKS	3	3,4,25	negative	0
MVK	27	1-6,8-19,20,21,22,25-30	negative	0
MYLK3	0		negative	0
MYO15A	**17**	**1,4,5,9,12,14-17,19,20-23,26,30,31**	**positive**	**2**
MYO3B	**0**		**positive**	**1**
MYO7A	**5**	**7,9,12,17,26**	**positive**	**2**
MYO9A	**13**	**1,3,10-13,15,16,17,19,20,22,31**	**positive**	**3**
NPFFR2	**4**	**11,16,19,26**	**positive**	**1**
NPY	0		negative	0
NPY1R	0		negative	0
OR10K1	**2**	**3,22**	**positive**	**2**
OR10V1	2	9,17	negative	0
OR13H1	0		negative	0
OR2B11	**1**	**15**	**positive**	**1**
PAFAH2	2	12,15	negative	0
PARVG	1	12	negative	0
PCDH4B	0		negative	0
PDE6B	2	15,26	negative	0
PLA2G2E	4	9,17,26,27	negative	0
PLA2G3	**1**	**17**	**positive**	**2**
PPAP2A	**2**	**5,26**	**positive**	**4**
PPAPDC1B	1	9	negative	0
PPEF1	6	4,11,15,22,23,28	negative	0
PRKAG1	0		negative	0
PRKG2	0		negative	0
PROM1	**16**	**3,4,6,7,8,11,14-17,20,22,23,28,30,31**	**positive**	**6**
PTPRQ	**25**	**2-17,19,21,22,23,27~31**	**positive**	**4**
RTP3	**1**	**1**	**positive**	**1**
SHC4	0		negative	0
SIAE	**2**	**9,24**	**positive**	**1**
SLCO1A2	5	7,10,11,16,23	negative	0
SMG1	12	3,10,11,12,14-17,22,23,24,29	negative	0
STARD5	1	9	negative	0
TAS2R3	**10**	**2,3,5,8,10,12,18,19,25,30**	**positive**	**3**
TAS2R38	**11**	**3,5,6,7,8,10,11,15,26,30,31**	**positive**	**1**
THBS2	12	2,3,7-11,15,17,22,26,31	negative	0

Significant genes common to both approaches are highlighted in red.
a- TreeSAAP is used to measure structural and biochemical properties of amino acid replacement using a threshold of P<0.001. 31 categories are tested as follows: 1. Alpha-helical tendencies, 2. Average number of surrounding residues, 3. Beta-structure tendencies, 4. Bulkiness, 5. Buriedness, 6. Chromatographic index, 7. Coil tendencies, 8. Composition, 9. Compressibility, 10. Equilibrium constant, 11. Helical contact area, 12. Hydropathy, 13. Isoelectric point, 14. Long-range non-bonded energy, 15. Mean r.m.s. fluctuation displacement, 16. Molecular volume, 17. Molecular weight, 18. Normalized consensus hydrophobicity, 19. Partial specific volume, 20. Polar requirement, 21. Polarity, 22. Power to be at the C-terminal, 23. Power to be at the middle of alpha-helix, 24. Power to be at the N-terminal, 25. Refractive index, 26. Short and medium range non-bonded energy, 27. Solvent accessible reduction ratio, 28. Surrounding hydrophobicity, 29. Thermodynamic transfer hydrohphobicity, 30. Total non-bonded energy, 31. Turn tendencies
b - Amino acid substitutions labeled as "deleterious" based on Provean.

Dataset S1.5. Enriched pathways among genes under positive selection in the domestic cat (Felinae) lineage

PATHWAY COMMONS CATEGORY	C	O	E	GENES
BETA-OXIDATION OF PRISTANOYL-COA	8	4	0.11	ACOX2, AMACR, ACOX3, ACOT8
BILE ACID AND BILE SALT METABOLISM	27	5	0.37	SLCO1A2, ACOX2, AMACR, HSD3B7, ACOT8
SYNTHESIS OF BILE ACIDS AND BILE SALTS VIA 7ALPHA-HYDROXYCHOLESTEROL	15	4	0.21	ACOX2, AMACR, HSD3B7, ACOT8
PEROXISOMAL LIPID METABOLISM	20	4	0.28	ACOX2, AMACR, ACOX3, ACOT8
METABOLISM OF LIPIDS AND LIPOPROTEINS	258	12	3.57	LCAT, CYP27B1, PPAP2A, SLCO1A2, MVK, HADH, STARD5, ACOX2, AMACR, ACOX3, HSD3B7, ACOT8
KEGG CATEGORY				
ECM-RECEPTOR INTERACTION	85	6	1.18	HMMR, ITGA9, THBS2, LAMC2, ITGA2B, COL6A3
LONG-TERM DEPRESSION	70	6	0.97	PRKG2, PLA2G2E, BRAF, GRIA2, ITPR3, PLA2G3
PRIMARY BILE ACID BIOSYNTHESIS	16	3	0.22	ACOX2, AMACR, HSD3B7
ETHER LIPID METABOLISM	36	4	0.5	PLA2G2E, PPAP2A, PAFAH2, PLA2G3
FOCAL ADHESION	200	9	2.77	SHC4, BRAF, PARVG, MYLK3, ITGA9, THBS2, LAMC2, ITGA2B, COL6A3
ALPHA-LINOLENIC ACID METABOLISM	20	3	0.28	PLA2G2E, ACOX3, PLA2G3
PEROXISOME	79	5	1.09	ACOX2, MVK, AMACR, ACOX3, ACOT8
GO CATEGORY				
LIPID MODIFICATION	143	11	2.16	LCAT, PPAP2A, HADH, PRKAG1, ACOX2, AMACR, INPP5J, ACOX3, SMG1, PPAPDC1B, ACOT8
FATTY ACID BETA-OXIDATION USING ACYL-COA OXIDASE	11	4	0.17	ACOX2, AMACR, ACOX3, ACOT8
CARBOXYLIC ESTER HYDROLASE ACTIVITY	116	8	1.71	LCAT, PAFAH2, ACOT11, PLA2G2E, SIAE, ABHD1, PLA2G3, ACOT8
PRISTANOYL-COA OXIDASE ACTIVITY	2	2	0.03	ACOX2, ACOX3
BRCA1-BARD1 COMPLEX	2	2	0.03	BRCA1, BARD1

USER DATA & PARAMETERS - N = 281 genes submitted, Genes mapped to unique Entrez Gene IDs: 281, Organism: hsapiens, Id Type: gene_symbol, Ref Set: entrezgene_protein-coding, Significance Level: .05, Statistics Test: Hypergeometric, MTC: BH, Minimum: 2

COLUMN HEADINGS - number of reference genes in the category (C), number of genes in the gene set and also in the category (O), expected number in the category (E).

Dataset S1.6. Enriched gene ontology categories among genes under positive selection in Carnivora

GO CATEGORY	PATHWAY ID	C	O	E	R	rawP	adjP
pattern recognition receptor activity	GO:0008329	15	4	0.39	10.29	0.0005	0.0482
glycosaminoglycan binding	GO:0005539	174	13	4.51	2.88	0.0006	0.0482
diacyl lipopeptide binding	GO:0042498	2	2	0.05	38.60	0.0007	0.0482
secondary active oligopeptide transmembrane transporter activity	GO:0015322	2	2	0.05	38.60	0.0007	0.0482
bacterial cell surface binding	GO:0051635	17	4	0.44	9.08	0.0008	0.0482
proton-dependent oligopeptide secondary active transmembrane transporter activity	GO:0005427	2	2	0.05	38.60	0.0007	0.0482
carbohydrate derivative binding	GO:0097367	189	14	4.90	2.86	0.0004	0.0482
cytoplasmic part	GO:0044444	6728	210	170	1.23	5.38E-05	0.0157
plasma membrane part	GO:0044459	1908	72	48.38	1.49	0.0003	0.0292
intrinsic to plasma membrane	GO:0031226	1255	53	31.82	1.67	0.0002	0.0292
integral to plasma membrane	GO:0005887	1214	49	30.78	1.59	0.0008	0.0389
Toll-like receptor 2-Toll-like receptor 6 protein	GO:0035355	2	2	0.05	2.41	0.0008	0.0389
mitochondrial matrix	GO:0005759	278	17	7.05	2.41	0.0008	0.0389
cytoplasm	GO:0005737	9051	261	229.5	1.14	0.001	0.0417
membrane	GO:0016020	7631	224	193.5	1.16	0.0015	0.0487
cell periphery	GO:0071944	4286	136	108.68	1.25	0.0015	0.0487

USER DATA & PARAMETERS - N = 467 genes submitted, Genes mapped to unique Entrez Gene IDs: 466, Organism: hsapiens, Id Type: gene_symbol, Ref Set: entrezgene_protein-coding, Significance Level: .05, Statistics Test: Hypergeometric, MTC: BH, Minimum: 2

COLUMN HEADINGS - number of reference genes in the category (C), number of genes in the gene set and also in the category (O), expected number in the category (E), Ratio of enrichment (R), p value from hypergeometric test (rawP), and p value adjusted by the multiple test adjustment (adjP).

Dataset S1.7. Enriched gene ontology categories among genes under positive selection in Felidae

PATHWAY COMMONS	Pathway ID	C	O	E	R	rawP	adjP
AlphaE beta7 integrin cell surface interactions	1632	3	3	0.5	61.46	4.27E-06	0.0012
Adaptive Immune System	515	237	14	3	3.63	3.64E-05	0.0049
Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	1098	52	6	0.82	7.09	0.0002	0.0180
Immune System	522	522	20	8.49	2.35	0.0004	0.027
Interaction between L1 and Ankyrins	45	12	3	0.2	15.37	0.0008	0.0432
KEGG CATEGORY							
Cell adhesion molecules (CAMs)	4514	133	9	2.16	4.16	0.0003	0.0246
GO CATEGORY							
epoxide hydrolase activity	GO:0004301	5	3	0.09	33.78	5.40E-05	0.0155
ether hydrolase activity	GO:0016803	7	3	0.12	24.13	0.0002	0.0287
external side of plasma membrane	GO:0009897	199	12	3.42	3.51	0.0002	0.0442

USER DATA & PARAMETERS - N = 331 genes submitted, Genes mapped to unique Entrez Gene IDs: 331, Organism: hsapiens, Id Type: gene_symbol, Ref Set: entrezgene_protein-coding, Significance Level: .05, Statistics Test: Hypergeometric, MTC: BH, Minimum: 2

COLUMN HEADINGS - number of reference genes in the category (C), number of genes in the gene set and also in the category (O), expected number in the category (E), Ratio of enrichment (R), p value from hypergeometric test (rawP), and p value adjusted by the multiple test adjustment (adjP).

Dataset S1.8(a). *Felis catus* gene members that underwent rapid gene family expansions along the Felidae lineage.

Ensembl Gene ID	Ensembl Family Description	Ensembl Protein Family ID(s)
CAFE Family ID:48		
ENSFCAG00000025324	IG HEAVY CHAIN V REGION	ENSFM00670001235368
ENSFCAG00000027432	"	ENSFM00670001235368
ENSFCAG00000028301	"	ENSFM00670001235368
ENSFCAG00000028921	"	ENSFM00670001235368
ENSFCAG00000023488	"	ENSFM00670001235368
ENSFCAG00000026432	"	ENSFM00670001235368
ENSFCAG00000029107	"	ENSFM00670001235368
ENSFCAG00000028889	"	ENSFM00670001235368
ENSFCAG00000029901	"	ENSFM00670001235368
ENSFCAG00000028814	"	ENSFM00670001235368
ENSFCAG00000023635	"	ENSFM00670001235368
ENSFCAG00000027142	"	ENSFM00670001235368
ENSFCAG00000028661	"	ENSFM00670001235368
ENSFCAG00000023737	"	ENSFM00670001235368
ENSFCAG00000026570	"	ENSFM00670001235368
ENSFCAG00000025755	"	ENSFM00670001235368
ENSFCAG00000027760	"	ENSFM00670001235368
ENSFCAG00000026585	"	ENSFM00670001235368
ENSFCAG00000031242	"	ENSFM00670001235368
ENSFCAG00000023332	"	ENSFM00670001235368
ENSFCAG00000023265	"	ENSFM00670001235368
ENSFCAG00000023423	"	ENSFM00670001235368
ENSFCAG00000023729	"	ENSFM00670001235368
ENSFCAG00000030410	"	ENSFM00670001235368
ENSFCAG00000029880	"	ENSFM00670001235368
ENSFCAG00000022386	"	ENSFM00670001235368
ENSFCAG00000026880	"	ENSFM00670001235368
ENSFCAG00000022825	HEAVY V	ENSFM00670001235643
ENSFCAG00000024010	IG HEAVY CHAIN V I REGION	ENSFM00670001235685
ENSFCAG00000022778	"	ENSFM00670001235685
ENSFCAG00000023329	"	ENSFM00670001235685
ENSFCAG00000022071	UNKNOWN	ENSFM00700001406400

Dataset S1.8(b). *Felis catus* gene members that underwent rapid gene family expansions along the Felidae lineage.

CAFE Family ID:60		
ENSFCAG00000001090	PEPTIDYL PROLYL CIS TRANS ISOMERASE 1 PPIASE EC_5.2.1.8 ROTAMASE	ENSFM00500000270856
ENSFCAG00000005373	"	ENSFM00500000271254
ENSFCAG00000028008	"	ENSFM00500000272090
ENSFCAG00000004182	"	ENSFM00500000269861
ENSFCAG00000008910	"	ENSFM00500000269861
ENSFCAG00000006027	PEPTIDYL PROLYL CIS TRANS ISOMERASE PPIASE EC_5.2.1.8 CYCLOPHILIN CYCLOSPORIN A BINDING ROTAMASE	ENSFM00600000921134
ENSFCAG00000009159	"	ENSFM00600000921134
ENSFCAG00000023140	"	ENSFM00600000921134
ENSFCAG00000028094	"	ENSFM00600000921134
ENSFCAG00000028260	"	ENSFM00600000921134
ENSFCAG00000030523	"	ENSFM00600000921134
ENSFCAG00000027344	"	ENSFM00600000921134
ENSFCAG00000028314	"	ENSFM00600000921134
ENSFCAG00000029878	"	ENSFM00600000921134
ENSFCAG00000030193	"	ENSFM00600000921134
ENSFCAG00000028578	"	ENSFM00600000921134
ENSFCAG00000012326	"	ENSFM00600000921134
ENSFCAG00000026216	"	ENSFM00600000921134
ENSFCAG00000025543	"	ENSFM00600000921134
ENSFCAG00000022115	"	ENSFM00600000921134
ENSFCAG00000000811	PEPTIDYL PROLYL CIS TRANS ISOMERASE PPIASE EC_5.2.1.8 ROTAMASE	ENSFM00710001441744
ENSFCAG00000022870	"	ENSFM00710001441744
ENSFCAG00000028615	"	ENSFM00710001441744
ENSFCAG00000028926	"	ENSFM00710001441744
ENSFCAG00000028168	"	ENSFM00710001441744
ENSFCAG00000009056	PEPTIDYLPROLYL ISOMERASE DOMAIN AND WD REPEAT CONTAINING 1 EC_5.2.1.8	ENSFM00500000270357
ENSFCAG00000027963	RANBP2 AND GRIP DOMAIN CONTAINING RAN BINDING 2 RANBP2 RANB	ENSFM00500000270422

Dataset S1.8(c). *Felis catus* gene members that underwent rapid gene family expansions along the Felidae lineage.

CAFE Family ID:77		
ENSFCAG00000000768	DYNEIN HEAVY CHAIN AXONEMAL AXONEMAL BETA DYNEIN HEAVY CHAIN CILIARY DYNEIN HEAVY CHAIN	ENSFM00710001441583
ENSFCAG00000002303	"	ENSFM00710001441583
ENSFCAG00000003487	"	ENSFM00710001441583
ENSFCAG00000008938	"	ENSFM00710001441583
ENSFCAG00000009626	"	ENSFM00710001441583
ENSFCAG00000011050	"	ENSFM00710001441583
ENSFCAG00000011062	"	ENSFM00710001441583
ENSFCAG00000011997	"	ENSFM00710001441583
ENSFCAG00000014410	"	ENSFM00710001441583
ENSFCAG00000015163	"	ENSFM00710001441583
ENSFCAG00000015341	"	ENSFM00710001441583
ENSFCAG00000015710	"	ENSFM00710001441583
ENSFCAG00000024375	"	ENSFM00710001441583
ENSFCAG00000031892	"	ENSFM00710001441583
ENSFCAG00000028573	"	ENSFM00710001441583
ENSFCAG00000030613	"	ENSFM00710001441583
ENSFCAG00000022940	"	ENSFM00710001441583
ENSFCAG00000030413	"	ENSFM00710001441583
ENSFCAG00000023988	"	ENSFM00710001441583
ENSFCAG00000027018	"	ENSFM00710001441583
ENSFCAG00000025942	"	ENSFM00710001441583
ENSFCAG00000029696	"	ENSFM00710001441583
ENSFCAG00000025884	"	ENSFM00710001441583
ENSFCAG00000029772	DYNEIN HEAVY CHAIN 14 AXONEMAL AXONEMAL BETA DYNEIN HEAVY CHAIN 14 CILIARY DYNEIN HEAVY CHAIN 14	ENSFM00250000013821
ENSFCAG00000027160	UNKNOWN	ENSFM00700001403725
ENSFCAG00000003480	UNKNOWN	ENSFM00700001395909

CAFE Family ID:96		
ENSFCAG00000024916	T CELL RECEPTOR ALPHA CHAIN V REGION PY14 PRECURSOR	ENSFM00670001239630
ENSFCAG00000025340	UNCHARACTERIZED FRAGMENT	ENSFM00670001238133
ENSFCAG00000028266	"	ENSFM00670001240217
ENSFCAG00000026476	"	ENSFM00670001257046
ENSFCAG00000030178	"	ENSFM00670001244595
ENSFCAG00000030540	UNKNOWN	ENSFM00670001238314
ENSFCAG00000023413	"	ENSFM00700001403106
ENSFCAG00000028123	"	ENSFM00670001238314

Dataset S1.8(d). *Felis catus* gene members that underwent rapid gene family expansions along the Felidae lineage.

CAFE Family ID:107		
ENSFCAG00000007354	VOMERONASAL TYPE 1 RECEPTOR V1R RECEPTOR	ENSFM00420000140525
ENSFCAG00000013746	"	ENSFM00420000140525
ENSFCAG00000026501	"	ENSFM00390000126342
ENSFCAG00000023564	"	ENSFM00420000140525
ENSFCAG00000030794	"	ENSFM00420000140525
ENSFCAG00000030467	"	ENSFM00420000140525
ENSFCAG00000029349	"	ENSFM00420000140525
ENSFCAG00000028999	"	ENSFM00500000270777
ENSFCAG00000025970	"	ENSFM00500000270777
ENSFCAG00000031455	"	ENSFM00420000140525
ENSFCAG00000029994	"	ENSFM00420000140525
ENSFCAG00000028171	"	ENSFM00500000269919
ENSFCAG00000025619	"	ENSFM00420000140525
ENSFCAG00000022751	"	ENSFM00420000140525
ENSFCAG00000030971	"	ENSFM00420000140525
ENSFCAG00000031841	"	ENSFM00570000851064
ENSFCAG00000026750	"	ENSFM00420000140525
ENSFCAG00000022670	"	ENSFM00420000140525
ENSFCAG00000000122	"	ENSFM00420000140525
ENSFCAG00000022668	"	ENSFM00500000270777
ENSFCAG00000029277	"	ENSFM00420000140525
ENSFCAG00000031101	"	ENSFM00500000269919

CAFE Family ID:159		
ENSFCAG00000002807	TRANSCRIPTION FACTOR SOX	ENSFM00500000269754
ENSFCAG00000004219	"	ENSFM00500000269754
ENSFCAG00000015685	"	ENSFM00500000269754
ENSFCAG00000022613	"	ENSFM00670001235710
ENSFCAG00000009619	SOX 15	ENSFM00500000274021

CAFE Family ID:323		
ENSFCAG00000001958	COLLAGEN ALPHA CHAIN PRECURSOR	ENSFM00250000000231
ENSFCAG00000004005	"	ENSFM00250000000231
ENSFCAG00000009383	"	ENSFM00250000000231
ENSFCAG00000026038	"	ENSFM00250000000231
ENSFCAG00000029007	"	ENSFM00250000000231
ENSFCAG00000025042	"	ENSFM00250000000231
ENSFCAG00000031671	UNKNOWN	ENSFM00700001406119
ENSFCAG00000030864	"	ENSFM00700001407229
ENSFCAG00000025023	"	ENSFM00700001406121

Dataset S1.8(e). *Felis catus* gene members that underwent rapid gene family expansions along the Felidae lineage.

CAFE Family ID:415		
ENSFCAG00000003805	HISTONE H1	ENSFM00670001235652
ENSFCAG00000005967	"	ENSFM00670001235652
ENSFCAG00000006079	"	ENSFM00670001235652
ENSFCAG00000006768	"	ENSFM00670001235652
ENSFCAG00000015177	"	ENSFM00670001235652
ENSFCAG00000005970	HISTONE H1T TESTICULAR H1 HISTONE	ENSFM00600000922221
ENSFCAG00000005962	HISTONE H1 1 HISTONE H1A	ENSFM00670001237167
ENSFCAG00000023697	UNKNOWN	ENSFM00700001402417

CAFE Family ID:494		
ENSFCAG00000027832	DIAPHANOUS HOMOLOG DIAPHANOUS RELATED FORMIN	ENSFM00260000050429
ENSFCAG00000031011	"	ENSFM00260000050429
ENSFCAG00000025687	"	ENSFM00260000050429
ENSFCAG00000027130	UNKNOWN	ENSFM00700001407483
ENSFCAG00000031194	"	ENSFM00700001406364
ENSFCAG00000029496	"	ENSFM00700001403036
ENSFCAG00000029591	"	ENSFM00700001406366
ENSFCAG00000027080	"	ENSFM00700001406365

CAFE Family ID:507		
ENSFCAG00000009104	PARTITIONING DEFECTIVE 3 HOMOLOG B AMYOTROPHIC LATERAL SCLEROSIS 2 CHROMOSOMAL REGION CANDIDATE GENE 19 PAR3 BETA PARTITIONING DEFECTIVE 3 PAR3 L	ENSFM00610000952891
ENSFCAG00000026384	"	ENSFM00610000952891
ENSFCAG00000025827	"	ENSFM00610000952891
ENSFCAG00000022493	"	ENSFM00610000952891
ENSFCAG00000026773	"	ENSFM00610000952891
ENSFCAG00000025870	"	ENSFM00610000952891
ENSFCAG00000024471	UNKNOWN	ENSFM00700001403157
ENSFCAG00000027778	"	ENSFM00700001404186

Dataset S1.8(f). *Felis catus* gene members that underwent rapid gene family expansions along the Felidae lineage.

CAFE Family ID:598		
ENSFCAG00000015547	MYOTILIN MYOFIBRILLAR TITIN IG DOMAINS TITIN IMMUNOGLOBULIN DOMAIN	ENSFM00570000851448
ENSFCAG00000026398	"	ENSFM00570000851448
ENSFCAG00000023378	MYOSIN LIGHT CHAIN KINASE SMOOTH MUSCLE FRAGMENT MLCK EC_2.7.11.18	ENSFM00710001444534
ENSFCAG00000023052	MYOSIN LIGHT CHAIN KINASE EC_2.7.11.18	ENSFM00550000743135
ENSFCAG00000031720	PALLADIN	ENSFM00690001356798
ENSFCAG00000027262	"	ENSFM00570000851711
ENSFCAG00000024606	"	ENSFM00570000851711
ENSFCAG00000003424	"	ENSFM00570000851679
ENSFCAG00000031314	UNKNOWN	ENSFM00700001403043

CAFE Family ID:757		
ENSFCAG00000030661	MYOMEGALIN PHOSPHODIESTERASE 4D INTERACTING	ENSFM00250000001701
ENSFCAG00000026455	"	ENSFM00250000001701
ENSFCAG00000028868	"	ENSFM00250000001701
ENSFCAG00000023617	"	ENSFM00250000001701
ENSFCAG00000031847	"	ENSFM00250000001701
ENSFCAG00000027086	"	ENSFM00250000001701
ENSFCAG00000031642	NEUROBLASTOMA BREAKPOINT FAMILY MEMBER 6	ENSFM00500000284802
ENSFCAG00000031382	UNKNOWN	ENSFM00700001403283

CAFE Family ID:764		
ENSFCAG00000005848	SET PHOSPHATASE 2A INHIBITOR I2PP2A I 2PP2A TEMPLATE ACTIVATING FACTOR I TAF I	ENSFM00500000270208
ENSFCAG00000005959	"	ENSFM00500000270208
ENSFCAG00000021897	"	ENSFM00500000270208
ENSFCAG00000024431	"	ENSFM00500000270208
ENSFCAG00000031705	"	ENSFM00500000270208
ENSFCAG00000024762	"	ENSFM00500000270208

Dataset S1.8(g). *Felis catus* gene members that underwent rapid gene family expansions along the Felidae lineage.

CAFE Family ID:950		
ENSFCAG00000006215	PIEZO TYPE MECHANOSENSITIVE ION CHANNEL COMPONENT 1 MEMBRANE INDUCED BY BETA AMYLOID TREATMENT MIB FAM38A	ENSFM00250000000782
ENSFCAG00000031105	"	ENSFM00250000000782
ENSFCAG00000026991	"	ENSFM00250000000782
ENSFCAG00000022624	"	ENSFM00250000000782
ENSFCAG00000027894	UNKNOWN	ENSFM00700001403972
ENSFCAG00000029479	"	ENSFM00700001407161

CAFE Family ID:1069		
ENSFCAG00000009998	NUCLEAR RECEPTOR COREPRESSOR 1 N COR N COR1	ENSFM00250000001120
ENSFCAG00000026816	"	ENSFM00250000001120
ENSFCAG00000030219	"	ENSFM00250000001120
ENSFCAG00000025434	"	ENSFM00250000001120
ENSFCAG00000027575	"	ENSFM00250000001120
ENSFCAG00000029502	"	ENSFM00250000001120
ENSFCAG00000031193	"	ENSFM00250000001120

CAFE Family ID:1264		
ENSFCAG00000000945	E3 UBIQUITIN LIGASE RNF213 EC_6.3.2.-	ENSFM00440000236907
ENSFCAG00000024632	"	ENSFM00440000236907
ENSFCAG00000022783	"	ENSFM00440000236907
ENSFCAG00000026190	"	ENSFM00440000236907
ENSFCAG00000022578	"	ENSFM00440000236907
ENSFCAG00000023279	"	ENSFM00440000236907
ENSFCAG00000022915	"	ENSFM00440000236907
ENSFCAG00000024960	UNKNOWN	ENSFM00700001402418

CAFE Family ID:1785		
ENSFCAG00000027649	HERV R_7Q21 2 PROVIRUS ANCESTRAL ENV POLYPROTEIN PRECURSOR ERV 3 ENVELOPE ERV3 ENVELOPE ERV3 1 ENVELOPE ENVELOPE POLYPROTEIN HERV R ENVELOPE ERV R ENVELOPE [CONTAINS SURFACE SU ; TRANSMEMBRANE TM]	ENSFM00250000016078
ENSFCAG00000025816	"	ENSFM00250000016078
ENSFCAG00000027786	"	ENSFM00250000016078
ENSFCAG00000029632	"	ENSFM00250000016078
ENSFCAG00000026669	"	ENSFM00250000016078
ENSFCAG00000031768	"	ENSFM00250000016078

Dataset S1.8(h). *Felis catus* gene members that underwent rapid gene family expansions along the Felidae lineage.

CAFE Family ID:2247		
ENSFCAG00000018707	EYES SHUT HOMOLOG FRAGMENT EPIDERMAL GROWTH FACTOR 10 EGF 10 EPIDERMAL GROWTH FACTOR 11 EGF 11 SPACEMAKER HOMOLOG	ENSFM00570000852066
ENSFCAG00000024318	"	ENSFM00690001356937
ENSFCAG00000022442	"	ENSFM00690001356937
ENSFCAG00000030693	EYES SHUT	ENSFM00570000851871
ENSFCAG00000024482	"	ENSFM00570000851871
ENSFCAG00000030936	UNKNOWN	ENSFM00700001405307

CAFE Family ID:2388		
ENSFCAG00000013038	GLYCINE CLEAVAGE SYSTEM H PROTEIN MITOCHONDRIAL PRECURSOR	ENSFM00500000271167
ENSFCAG00000025063	"	ENSFM00500000271167
ENSFCAG00000024721	UNKNOWN	ENSFM00700001404018

CAFE Family ID:2487		
ENSFCAG00000014455	10 KDA HEAT SHOCK PROTEIN MITOCHONDRIAL HSP10.10 KDA CHAPERONIN CHAPERONIN 10 CPN10	ENSFM00670001235755
ENSFCAG00000028974	"	ENSFM00670001235755
ENSFCAG00000025005	UNKNOWN	ENSFM00700001401236
ENSFCAG00000024093	"	ENSFM00700001401235
ENSFCAG00000023339	"	ENSFM00700001401234

CAFE Family ID:2491		
ENSFCAG00000001973	LEUCINE RICH REPEAT SERINE/THREONINE KINASE 1 EC_2.7.11.1	ENSFM00250000001794
ENSFCAG00000030173	UNKNOWN	ENSFM00700001403499
ENSFCAG00000027327	"	ENSFM00700001401791

CAFE Family ID:2587		
ENSFCAG00000023000	UBIQUITIN CARBOXYL TERMINAL HYDROLASE 40 EC_3.4.19.12 DEUBIQUITINATING ENZYME 40 UBIQUITIN THIOESTERASE 40 UBIQUITIN SPECIFIC PROCESSING PROTEASE 40	ENSFM00250000005400
ENSFCAG00000025764	"	ENSFM00250000005400
ENSFCAG00000028768	"	ENSFM00250000005400

CAFE Family ID:2614		
ENSFCAG00000005785	NONSENSE 2 UP FRAMESHIFT SUPPRESSOR 2	ENSFM00250000002346
ENSFCAG00000030556	"	ENSFM00250000002346
ENSFCAG00000024774	"	ENSFM00250000002346

Dataset S1.8(i). *Felis catus* gene members that underwent rapid gene family expansions along the Felidae lineage.

CAFE Family ID:2825		
ENSFCAG00000023028	SYNAPTONEMAL COMPLEX 1 SCP 1	ENSFM00250000006405
ENSFCAG00000022262	"	ENSFM00250000006405
ENSFCAG00000029189	"	ENSFM00250000006405
CAFE Family ID:2831		
ENSFCAG00000004711	TRANSCRIPTION ELONGATION FACTOR B POLYPEPTIDE 2 ELONGIN 18 KDA SUBUNIT ELONGIN B ELOB RNA POLYMERASE II TRANSCRIPTION FACTOR SIII SUBUNIT B SIII P18	ENSFM00500000273664
ENSFCAG00000029930	"	ENSFM00500000273664
ENSFCAG00000026327	"	ENSFM00500000273664
CAFE Family ID:3042		
ENSFCAG00000026757	COILED COIL DOMAIN CONTAINING 168	ENSFM00570000852061
ENSFCAG00000022657	"	ENSFM00570000852061
ENSFCAG00000023173	UNKNOWN	ENSFM00700001404181
CAFE Family ID:9827		
ENSFCAG00000001617	STELLA FRAGMENT	ENSFM00680001305395
ENSFCAG00000026706	"	ENSFM00680001305395
ENSFCAG00000026149	UNKNOWN	ENSFM00700001402535

Dataset S1.9. Summary of 1-Kbps windows, copy number distribution in control regions and gain/loss cutoffs for the domestic cat (Abyssinian sample)

Sequencing	
Sequencing technology	Illumina
# Reads	1,485,609,004
Coverage	21.8X
1-Kbps windows	
# Total windows	1,122,501
# Control windows	993,102
# Non control windows	129,399
Gain/loss cutoffs	
Mean copy number in control regions	2
StDev copy number in control regions	0.24
(# windows excluded*)	9,932
Gain cutoff	2.71
Loss cutoff	1.29

*1-Kbps windows exceeding the 1% highest copy number value

Dataset S1.10. Summary of duplications and deletions using sample-specific gain/loss cutoffs based on the copy number distribution from the control regions within the domestic cat genome (Abyssinian sample)

Duplications	
# Duplications	85
# Duplications (gaps removed)	1002
# Bps*	9,065,598
% size of autosomes	0.39
# Bps in shared duplications*	4,377,574
% of duplicated bps	48.29
Deletions	
# Deletions	1
# Deletions (gaps removed)	18
# Bps*	54,896
% size of autosomes	<0.01
# Bps in shared deletions*	0
% of deleted bps	0

*All bps are after excluding the size of the gaps (M1 method)

Dataset S1.11(a). Genes underlying regions of segmental duplications in the domestic cat genome

Chromosome Name	Gene Start (bp)	Gene End (bp)	Ensembl Gene ID	Associated Gene Name	Description
A2	1999024	2024218	ENSFCAG00000026539	ZNF77	zinc finger protein 77
A2	2178186	2233965	ENSFCAG00000026952		
A2	4837081	4840788	ENSFCAG00000022669		
A2	4850030	4850775	ENSFCAG00000027538		
A2	4879236	4926614	ENSFCAG00000005041		
A2	4940497	4947843	ENSFCAG00000024220		
A2	4951852	4957168	ENSFCAG00000024591		
A2	10342173	10343114	ENSFCAG00000027467		
A2	10403957	10404904	ENSFCAG00000024728		
A2	10424192	10425154	ENSFCAG00000026934		
A2	10444212	10445147	ENSFCAG00000028346		
A2	10454027	10454932	ENSFCAG00000026032		
A2	10465336	10466262	ENSFCAG00000030313		
A2	10484879	10486405	ENSFCAG00000030477		
A2	10503936	10505982	ENSFCAG00000022920		
A2	10557343	10558290	ENSFCAG00000031912		
A2	10561828	10562778	ENSFCAG00000029420		
A2	10571496	10572431	ENSFCAG00000030284		
A2	10588491	10589447	ENSFCAG00000027325		
A2	10608719	10609642	ENSFCAG00000029342		
A2	10637607	10639184	ENSFCAG00000025226		
A2	10666812	10667738	ENSFCAG00000026535		
A2	10677429	10678364	ENSFCAG00000027309	OR7C1	olfactory receptor, family 7, subfamily C, member 1
A2	10691698	10693042	ENSFCAG00000024926		
A2	10735731	10736645	ENSFCAG00000025361		
A2	10749819	10750754	ENSFCAG00000028413		
A2	10774323	10775967	ENSFCAG00000026904		
A2	10802221	10803189	ENSFCAG00000030343		
A2	10811748	10813240	ENSFCAG00000031709		
A2	11305320	11311224	ENSFCAG00000025482		
A2	11312522	11362773	ENSFCAG00000008623	CYP4F3	cytochrome P450, family 4, subfamily F, polypeptide 3
A2	11380632	11382628	ENSFCAG00000031558		
A2	11390153	11391070	ENSFCAG00000024776		
A2	11408178	11409176	ENSFCAG00000029860		
A2	11427607	11428557	ENSFCAG00000029127		
A2	11444217	11445483	ENSFCAG00000026640		
A2	11454602	11455552	ENSFCAG00000023347		
A2	11469229	11470179	ENSFCAG00000028009		
A2	11478143	11479132	ENSFCAG00000030787		
A2	55532891	55578035	ENSFCAG00000013910	IQSEC1	IQ motif and Sec7 domain 1
A2	58451403	58575137	ENSFCAG00000001776	ALDH1L1	aldehyde dehydrogenase 1 family, member L1
A2	58512851	58512965	ENSFCAG00000020614	5S_rRNA	5S ribosomal RNA
A2	156323262	156324206	ENSFCAG00000025442		
A2	156336015	156336959	ENSFCAG00000024190		
A2	157059766	157060707	ENSFCAG00000026036		
A2	157078181	157079122	ENSFCAG00000025223		
A2	162660127	162661174	ENSFCAG00000003990	GIMAP2	GTPase, IMAP family member 2
A2	162671690	162672605	ENSFCAG00000031058		
A2	162686216	162686593	ENSFCAG00000027719		
A2	162720551	162752217	ENSFCAG00000011443		
A3	30261827	30272148	ENSFCAG00000025532		
A3	30278131	30288600	ENSFCAG00000030407		
A3	30353317	30361064	ENSFCAG00000001879		
A3	30372145	30383790	ENSFCAG00000031263		
A3	40441218	40441327	ENSFCAG00000024869	5S_rRNA	5S ribosomal RNA
B1	40121	43236	ENSFCAG00000019039		
B1	47201	48493	ENSFCAG00000007120	ZNF781	zinc finger protein 781
B1	36295742	36296780	ENSFCAG00000023307		
B2	328785	331268	ENSFCAG00000003660	OR12D2	olfactory receptor, family 12, subfamily D, member 2
B2	713302	714237	ENSFCAG00000011606		
B2	749421	750359	ENSFCAG00000005124		
B2	837074	838335	ENSFCAG00000025316		
B2	884452	885396	ENSFCAG00000030837		

Dataset S1.11(b). Genes underlying regions of segmental duplications in the domestic cat genome

Chromosome Name	Gene Start (bp)	Gene End (bp)	Ensembl Gene ID	Associated Gene Name	Description
B2	906325	906436	ENSFCAG00000027033	5S_rRNA	5S ribosomal RNA
B2	916798	917733	ENSFCAG00000023506		
B2	977120	979761	ENSFCAG00000026047		
B2	1060434	1061724	ENSFCAG00000028748		
B2	1085480	1086424	ENSFCAG00000030186		
B2	1104867	1106243	ENSFCAG00000000501	OR2B3	olfactory receptor, family 2, subfamily B, member 3
B2	1129129	1130064	ENSFCAG00000023353		
B2	1140538	1144027	ENSFCAG00000028271		
B2	1230045	1231165	ENSFCAG00000028202		
B2	1243051	1243163	ENSFCAG00000029309	5S_rRNA	5S ribosomal RNA
B2	1257504	1257616	ENSFCAG00000028666	5S_rRNA	5S ribosomal RNA
B2	1277229	1278164	ENSFCAG00000028799		
B2	1343658	1344622	ENSFCAG00000025334		
B2	1353504	1357463	ENSFCAG00000029693		
B2	1459936	1460865	ENSFCAG00000010634	OR2W1	olfactory receptor, family 2, subfamily W, member 1
B2	2299340	2300272	ENSFCAG00000026697		
B2	2322927	2323862	ENSFCAG00000029062		
B2	2332065	2348095	ENSFCAG00000028781		
B2	2364168	2365106	ENSFCAG00000023042		
B2	2374024	2374136	ENSFCAG00000030916	5S_rRNA	5S ribosomal RNA
B2	2398409	2399350	ENSFCAG00000027883		
B2	2437587	2438525	ENSFCAG00000023744		
B2	2529387	2530319	ENSFCAG00000030548		
B2	2539008	2539946	ENSFCAG00000025071		
B2	32597800	32602576	ENSFCAG00000021900		
B2	32667396	32670292	ENSFCAG00000000629	FLA-Z	MHC class I antigen
B2	32703681	32706770	ENSFCAG00000027223		
B2	32774101	32776385	ENSFCAG00000015379		
B2	32835681	32838540	ENSFCAG00000000877		
B2	32870997	32871109	ENSFCAG00000027024	5S_rRNA	5S ribosomal RNA
B2	32907635	32910483	ENSFCAG00000018113		
B2	32945158	32948534	ENSFCAG00000027242	FLA-I	MHC class I antigen precursor
B2	33007185	33013570	ENSFCAG00000022105		
B3	148227485	148229683	ENSFCAG00000025368		
B3	148232277	148232726	ENSFCAG00000028661		
B3	148259497	148259934	ENSFCAG00000025324		
B3	148322272	148322384	ENSFCAG00000031776	5S_rRNA	5S ribosomal RNA
B4	24696117	24799719	ENSFCAG00000014236	ANKRD26	ankyrin repeat domain 26
B4	24819736	24859130	ENSFCAG00000027161	RAB18	RAB18, member RAS oncogene family
B4	46929808	46931537	ENSFCAG00000013935		
B4	46958758	46960567	ENSFCAG00000030370		
D1	4486594	4486706	ENSFCAG00000022937	5S_rRNA	5S ribosomal RNA
D1	4776114	4776224	ENSFCAG00000030025	5S_rRNA	5S ribosomal RNA
D1	4803021	4803133	ENSFCAG00000025220	5S_rRNA	5S ribosomal RNA
D1	20649778	20650728	ENSFCAG00000029628		
D1	21354029	21355006	ENSFCAG00000002614		
D1	21380576	21381508	ENSFCAG00000008131	OR8B12	olfactory receptor, family 8, subfamily B, member 12
D1	64918054	64918998	ENSFCAG00000025048		
D1	64938061	64939629	ENSFCAG00000028751		
D1	66883484	66884427	ENSFCAG00000024203		
D1	66892866	66893789	ENSFCAG00000000727	OR10A3	olfactory receptor, family 10, subfamily A, member 3
D1	66908368	66909313	ENSFCAG00000028608		
D1	87753796	88005871	ENSFCAG00000030334	ELP4	elongator acetyltransferase complex subunit 4
D1	88012047	88025138	ENSFCAG00000007094	PAX6	paired box 6
D1	102240555	102241478	ENSFCAG00000001814		
D1	102283570	102284514	ENSFCAG00000024648		
D1	102337235	102338164	ENSFCAG00000014680	OR4A47	olfactory receptor, family 4, subfamily A, member 47
D1	103550482	103551426	ENSFCAG00000028369		
D1	103568480	103569423	ENSFCAG00000024411		
D1	113601303	113688582	ENSFCAG00000004765	PPFIA1	protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF), interacting protein (liprin), alpha 1
D2	129076	129188	ENSFCAG00000029175	5S_rRNA	5S ribosomal RNA
D2	8749213	9075406	ENSFCAG00000029000		
D2	8960580	8980244	ENSFCAG00000023704		
D2	20153518	20206464	ENSFCAG00000023459	TTC13	tetratricopeptide repeat domain 13

Dataset S1.11(c). Genes underlying regions of segmental duplications in the domestic cat genome

Chromosome Name	Gene Start (bp)	Gene End (bp)	Ensembl Gene ID	Associated Gene Name	Description
D2	22330902	22331016	ENSFCAG00000027550	5S_rRNA	5S ribosomal RNA
D2	22332106	22332224	ENSFCAG00000029290	5S_rRNA	5S ribosomal RNA
D2	22333313	22333421	ENSFCAG00000017601	5S_rRNA	5S ribosomal RNA
D2	22367720	22368642	ENSFCAG00000028109		
D2	22379806	22380681	ENSFCAG00000011440		
D2	89795938	89805869	ENSFCAG00000013762	CYP2E2	cytochrome P450 2E2
D2	89809103	89814632	ENSFCAG00000031109	ZNF717	zinc finger protein 717
D2	89817222	89822044	ENSFCAG00000013763	SYCE1	synaptonemal complex central element protein 1
D3	80759	80867	ENSFCAG00000021705	5S_rRNA	5S ribosomal RNA
D3	23225438	23225981	ENSFCAG00000025218		
D3	23273316	23273795	ENSFCAG00000026724		
D3	23297965	23298282	ENSFCAG00000025197		
D3	23382135	23382440	ENSFCAG00000029689		
D3	23419057	23419488	ENSFCAG00000023616		
D3	23509526	23510019	ENSFCAG00000031794		
D3	26658713	26658825	ENSFCAG00000027766	5S_rRNA	5S ribosomal RNA
D3	26681356	26709785	ENSFCAG00000006400	PIWIL3	piwi-like RNA-mediated gene silencing 3
D3	28148397	28167337	ENSFCAG00000005999	MED15	mediator complex subunit 15
D3	28167822	28174177	ENSFCAG00000030668		
D3	28235497	28244430	ENSFCAG00000006009	P2RX6	purinergic receptor P2X, ligand-gated ion channel, 6
D3	28280102	28284449	ENSFCAG00000022091	TUBA3E	tubulin, alpha 3e
D4	7198	11817	ENSFCAG00000029042		
D4	88592476	88595451	ENSFCAG00000023879		
D4	88654911	88657790	ENSFCAG00000001496		
D4	88692182	88695035	ENSFCAG00000027840		
D4	88712762	88715537	ENSFCAG00000031788		
D4	95006881	95010017	ENSFCAG00000012216		
D4	95011716	95016678	ENSFCAG00000012219		
E1	2184	3160	ENSFCAG00000008583		
E1	42029674	42032124	ENSFCAG00000030823		
E1	56288309	56290286	ENSFCAG00000023215		
E1	56322164	56324255	ENSFCAG00000030685		
E1	56334152	56334460	ENSFCAG00000001618		
E1	56382117	56385015	ENSFCAG00000029475		
E2	4520136	4522367	ENSFCAG00000023824		
E2	4673182	4688006	ENSFCAG00000028391		
E2	4739276	4739981	ENSFCAG00000016263		
E2	4893549	4895709	ENSFCAG00000025435		
E2	4950706	4951646	ENSFCAG00000030225		
E2	4960706	4962448	ENSFCAG00000024112		
E2	5330162	5331094	ENSFCAG00000025619	FELCATV1R6	vomeronasal 1 receptor felCatV1R6
E2	5360235	5361671	ENSFCAG00000023132		
E2	5412456	5420577	ENSFCAG00000023403		
E2	5480972	5529785	ENSFCAG00000023819		
E2	5486927	5492030	ENSFCAG00000029493		
E2	5537527	5537593	ENSFCAG00000017968		
E2	5564448	5571174	ENSFCAG00000031161		
E2	5604920	5606596	ENSFCAG00000025806		
E2	5641792	5645484	ENSFCAG00000023019		
E2	5712974	5714008	ENSFCAG00000028544		
E2	5880689	5881638	ENSFCAG00000025070		
E2	5887792	5888197	ENSFCAG00000028057		
E2	5918068	5919030	ENSFCAG00000022670	FELCATV1R7	vomeronasal 1 receptor felCatV1R7
E2	8497229	8501013	ENSFCAG00000007363		
E2	8513256	8529986	ENSFCAG00000022344	FUT2	fucosyltransferase 2 (secretor status included)
E2	8515155	8527136	ENSFCAG00000027085		
E2	12316795	12325442	ENSFCAG00000029888	CEACAM21	carcinoembryonic antigen-related cell adhesion molecule 21
E2	13122830	13132919	ENSFCAG00000013094	CYP2S1	cytochrome P450, family 2, subfamily S, polypeptide 1
E3	26876172	26876284	ENSFCAG00000029594	5S_rRNA	5S ribosomal RNA
E3	26994957	27039168	ENSFCAG00000008109	ACSM1	acyl-CoA synthetase medium-chain family member 1
E3	32693517	33115346	ENSFCAG00000010119	SNX29	sorting nexin 29

Dataset S1.12. Pathway enrichment results using all genes underlying regions of segmental duplications in the domestic cat genome (Abyssinian sample)

KEGG Pathway	Pathway ID	C	O	E	R	rawP	adjP	Genes
Olfactory transduction	4740	388	7	0.27	25.94	7.7E-09	1.54E-08	**OR4A47, OR7C1, OR8B12, OR10A3, OR12D2, OR2W1, OR2B3**
Metabolic pathways	1100	1130	3	0.79	3.82	0.0431	0.0431	**FUT2, CYP4F3, ACSM1**
Wikipathways Pathway								
GPCRs, Class A Rhodopsin-like	WP455	259	3	0.18	16.65	0.0008	0.0014	**OR7C1, OR2W1, OR2B3**
cytochrome P450	WP43	65	2	0.05	44.23	0.0009	0.0014	**CYP2S1, CYP4F3**
GO Category (Sub-root)								
olfactory receptor activity (molecular function)	GO:0004984	419	7	0.71	9.82	4.64E-06	0.0003	**OR4A47, OR7C1, OR8B12, OR10A3, OR12D2, OR2W1, OR2B3**
guanyl nucleotide binding (molecular function)	GO:0019001	392	4	0.67	6	0.0041	0.023	**RAB18, TUBA3E, ACSM1, GIMAP2**

USER DATA & PARAMETERS - N = 35 genes submitted, Genes mapped to unique Entrez Gene IDs: 33, Organism: hsapiens, Id Type: gene_symbol, Ref Set: entrezgene, Significance Level: .05, Statistics Test: Hypergeometric, MTC: BH, Minimum: 2

COLUMN HEADINGS - number of reference genes in the category (C), number of genes in the gene set and also in the category (O), expected number in the category (E), Ratio of enrichment (R), p value from hypergeometric test (rawP), and p value adjusted by the multiple test adjustment (adjP).

Dataset S2.1. Coverage statistics per pool.

Breed	Individuals	Sequence Depth (All Chromosomes)	Sequence Depth (Autosomes)	Properly Paired Reads
Abyssinian	1	20.39	20.43	208,102,582
Egyptian Mau	1	4.97	4.96	93,318,282
Maine Coon	5	10.52	10.65	271,512,388
Norwegian Forest	4	14.89	15.02	258,851,848
Birman	4	3.86	3.92	163,585,510
Japanese Bobtail	4	11.09	11.26	384,722,308
Turkish Van	4	9.26	9.38	405,812,058
Pooled Breeds	Total = 22	Mean = 9.1 Pooled = 54.57	Mean = 9.2 Pooled = 55.18	Total = 1,577,802,394
Felis silvestris	4	6.84	7.02	189,543,907

Dataset S2.2(a). Genes underlying regions of low H_p in the pooled domestic cat variant dataset following annotation of 100kb windows that fell below four standard deviations from the mean H_p

Chromosome Name	Gene Start (bp)	Gene End (bp)	Ensembl Gene ID	Associated Gene Name	Description	Overlap With F_{ST} List
A1	8396110	8766675	ENSFCAG00000014322	MTUS2	microtubule associated tumor suppressor candidate 2	
A1	8776308	8797912	ENSFCAG00000014326	SLC7A1	solute carrier family 7 (cationic amino acid transporter, y+ system), member 1	
A1	40039128	40039239	ENSFCAG00000023540	5S_rRNA	5S ribosomal RNA	
A1	52410273	52410385	ENSFCAG00000028097	5S_rRNA	5S ribosomal RNA	
A1	52479995	52528867	ENSFCAG00000000561	RBM26	RNA binding motif protein 26	
A1	52625952	52688765	ENSFCAG00000025797	NDFIP2	Nedd4 family interacting protein 2	
A1	84103052	84104082	ENSFCAG00000015576			
A1	84167672	84168631	ENSFCAG00000022722			
A1	84208569	84209522	ENSFCAG00000031931			
A1	88361081	88362067	ENSFCAG00000026872	OR2G3	olfactory receptor, family 2, subfamily G, member 3	
A1	88391241	88392229	ENSFCAG00000008196			
A1	88521633	88522592	ENSFCAG00000002236	OR2C3	olfactory receptor, family 2, subfamily C, member 3	
A1	88551982	88552917	ENSFCAG00000024148			
A1	88616973	88617956	ENSFCAG00000021910			
A1	88647235	88648721	ENSFCAG00000008976			
A1	88667669	88668640	ENSFCAG00000010456	OR2B11	olfactory receptor, family 2, subfamily B, member 11	
A1	88708791	88709744	ENSFCAG00000031962			
A1	88723599	88774919	ENSFCAG00000009344	NLRP3	NLR family, pyrin domain containing 3	
A1	89064612	89073091	ENSFCAG00000023784			
A1	89091044	89092033	ENSFCAG00000006397	RNF187	ring finger protein 187	
A1	89124037	89124141	ENSFCAG00000024726	5S_rRNA	5S ribosomal RNA	
A1	89136000	89136365	ENSFCAG00000024427			
A1	89140926	89141306	ENSFCAG00000032040	HIST3H2BB	histone cluster 3, H2bb	
A1	89141619	89142011	ENSFCAG00000024530	HIST3H2A	histone cluster 3, H2a	
A1	89163922	89164322	ENSFCAG00000006396			
A1	89176315	89184314	ENSFCAG00000029709	TRIM17	tripartite motif containing 17	
A1	89187378	89192375	ENSFCAG00000028702	TRIM11	tripartite motif containing 11	
A1	89211477	89214157	ENSFCAG00000025004			
A1	89215471	89216301	ENSFCAG00000023698			
A1	89216810	89218801	ENSFCAG00000022817			
A1	89227536	89243984	ENSFCAG00000030432			
A1	89246495	89247270	ENSFCAG00000025517			
A1	95458053	95465888	ENSFCAG00000031517			
A1	95498785	95504740	ENSFCAG00000030457			
A1	95525507	95744069	ENSFCAG00000025195	COMMD10	COMM domain containing 10	
A1	117462646	117535510	ENSFCAG00000022810	PCDHA1	protocadherin alpha 1	X
A1	117574779	117620293	ENSFCAG00000003685	PCDHAC2	protocadherin alpha subfamily C, 2	
A1	117653631	117656087	ENSFCAG00000003687	PCDHB1	protocadherin beta 1	
A1	117675941	117678295	ENSFCAG00000001367	PCDHB2	protocadherin beta 2	
A1	117694636	117729173	ENSFCAG00000013156	PCDHB4	protocadherin beta 4	X
A1	124586618	124647903	ENSFCAG00000025994	SLC38A9	solute carrier family 38, member 9	
A1	124685687	124778148	ENSFCAG00000012560	DDX4	DEAD (Asp-Glu-Ala-Asp) box polypeptide 4	
A1	124828501	124867387	ENSFCAG00000010857	IL31RA	interleukin 31 receptor A	
A1	124892755	124943980	ENSFCAG00000010859	IL6ST	interleukin 6 signal transducer (gp130, oncostatin M receptor)	
A1	125019663	125086857	ENSFCAG00000010860	ANKRD55	ankyrin repeat domain 55	
A1	182239426	182624612	ENSFCAG00000008547	EBF1	early B-cell factor 1	
A1	182675442	182769870	ENSFCAG00000012160			
A1	182768748	182791531	ENSFCAG00000026371	UBLCP1	ubiquitin-like domain containing CTD phosphatase 1	
A1	182815541	182825071	ENSFCAG00000015571	IL12B	Interleukin-12 subunit beta	
A1	192916759	193132695	ENSFCAG00000014984	GALNT10	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 10 (GalNAc-T10)	
A1	193168465	193168577	ENSFCAG00000026163	5S_rRNA	5S ribosomal RNA	X
A1	193262073	193280914	ENSFCAG00000028091	MFAP3	microfibrillar-associated protein 3	X
A1	193286712	193315803	ENSFCAG00000023708	FAM114A2	family with sequence similarity 114, member A2	X
A1	193479446	193624866	ENSFCAG00000005223	GRIA1	glutamate receptor, ionotropic, AMPA 1	X
A1	219080923	219081031	ENSFCAG00000027389	5S_rRNA	5S ribosomal RNA	
A2	110372750	110413021	ENSFCAG00000005018	AHR	aryl hydrocarbon receptor	
A2	110810088	110898682	ENSFCAG00000024117	SNX13	sorting nexin 13	
A3	24313334	24397487	ENSFCAG00000002305	PIGU	phosphatidylinositol glycan anchor biosynthesis, class U	
A3	24398183	24399185	ENSFCAG00000022040	MAP1LC3A	microtubule-associated protein 1 light chain 3 alpha	
A3	24418275	24429758	ENSFCAG00000002304	DYNLRB1	dynein, light chain, roadblock-type 1	
A3	24443756	24584348	ENSFCAG00000008780	ITCH	itchy E3 ubiquitin protein ligase	
A3	24642489	24656544	ENSFCAG00000007301			
A3	24671683	24676154	ENSFCAG00000011037	ASIP	Agouti-signaling protein	
A3	50153794	50156025	ENSFCAG00000030154			
A3	50157809	50158123	ENSFCAG00000027566			
A3	50222873	50229392	ENSFCAG00000022966			
A3	76497901	76527797	ENSFCAG00000003522	CCDC104	coiled-coil domain containing 104	
A3	76529028	76595389	ENSFCAG00000026457	SMEK2	SMEK homolog 2, suppressor of mek1 (Dictyostelium)	
A3	76618801	76664057	ENSFCAG00000013236	PNPT1	polyribonucleotide nucleotidyltransferase 1	
A3	96863689	96865875	ENSFCAG00000024660			

Dataset S2.2(b). Genes underlying regions of low H_p in the pooled domestic cat variant dataset following annotation of 100kb windows that fell below four standard deviations from the mean H_p

Chromosome Name	Gene Start (bp)	Gene End (bp)	Ensembl Gene ID	Associated Gene Name	Description	Overlap With F_{ST} List
A3	96889747	96889859	ENSFCAG00000028765	5S_rRNA	5S ribosomal RNA	
A3	96900604	96902810	ENSFCAG00000027440			
A3	126610231	126619976	ENSFCAG00000028084			
A3	126698895	126761424	ENSFCAG00000027520	PUM2	pumilio homolog 2 (Drosophila)	
A3	126775239	126796393	ENSFCAG00000005354	SDC1	syndecan 1	
A3	141321118	141464325	ENSFCAG00000013984	MYT1L	myelin transcription factor 1-like	
B1	44313905	44415983	ENSFCAG00000031578			
B1	44440898	44568664	ENSFCAG00000000589			
B1	44593908	44639323	ENSFCAG00000030919			
B1	44605069	44605181	ENSFCAG00000031785	5S_rRNA	5S ribosomal RNA	
B1	44691591	44732046	ENSFCAG00000028945			
B1	44746128	44770572	ENSFCAG00000028120			
B1	44774622	44784782	ENSFCAG00000023569			
B1	44792973	44919764	ENSFCAG00000002595	ADAM9	ADAM metallopeptidase domain 9	
B1	44794749	44795624	ENSFCAG00000002701			
B1	44920372	44925151	ENSFCAG00000002593	TM2D2	TM2 domain containing 2	
B1	44928099	44938386	ENSFCAG00000002591	HTRA4	HtrA serine peptidase 4	
B1	44938751	44984295	ENSFCAG00000030276	PLEKHA2	pleckstrin homology domain containing, family A (phosphoinositide binding specific) member 2	
B1	45043315	45101607	ENSFCAG00000024992	TACC1	transforming, acidic coiled-coil containing protein 1	
B1	104880577	104998796	ENSFCAG00000010364	PDE5A	cGMP-specific 3',5'-cyclic phosphodiesterase	
B1	105033096	105036707	ENSFCAG00000026257	FABP2	fatty acid binding protein 2, intestinal	
B1	105069890	105121855	ENSFCAG00000014472	USP53	ubiquitin specific peptidase 53	
B1	105153960	105193563	ENSFCAG00000003007	MYOZ2	myozenin 2	
B1	105252135	105424698	ENSFCAG00000028738	SYNPO2	synaptopodin 2	
B1	143593440	143685242	ENSFCAG00000010427	CCDC158	coiled-coil domain containing 158	
B1	172072490	172152353	ENSFCAG00000003209			
B1	172191715	172262048	ENSFCAG00000025531	UBE2K	ubiquitin-conjugating enzyme E2K	
B1	191450950	191588878	ENSFCAG00000029474	LCORL	ligand dependent nuclear receptor corepressor-like	
B1	191619013	191665923	ENSFCAG00000030778	NCAPG	non-SMC condensin I complex, subunit G	
B2	82062518	82106450	ENSFCAG00000030695			
B2	82129137	82129462	ENSFCAG00000022894			
B2	82190939	82198755	ENSFCAG00000029256	SRSF12	serine/arginine-rich splicing factor 12	
B2	82233859	82249448	ENSFCAG00000014224	PM20D2	peptidase M20 domain containing 2	
B2	82263854	82290954	ENSFCAG00000022082	GABRR1	gamma-aminobutyric acid (GABA) A receptor, rho 1	
B3	18642298	18719149	ENSFCAG00000007543	CERS3	ceramide synthase 3	
B3	33511989	33516370	ENSFCAG00000000344	CYP1A2	Cytochrome P450 1A2	X
B3	33532067	33538470	ENSFCAG00000002016	CYP1A1	Cytochrome P450 1A1	X
B3	33564717	33611587	ENSFCAG00000002014	EDC3	enhancer of mRNA decapping 3 homolog (S. cerevisiae)	X
B3	33611037	33628940	ENSFCAG00000031747	CLK3	CDC-like kinase 3	X
B3	33648267	33702607	ENSFCAG00000002012	ARID3B	AT rich interactive domain 3B (BRIGHT-like)	X
B3	33785369	33799296	ENSFCAG00000027852	UBL7	ubiquitin-like 7 (bone marrow stromal cell-derived)	
B3	33820675	33830309	ENSFCAG00000008142	SEMA7A	semaphorin 7A, GPI membrane anchor (John Milton Hagen blood group)	
B3	93427038	93427596	ENSFCAG00000023471			
B3	111258215	111397325	ENSFCAG00000014176	PPP2R5E	protein phosphatase 2, regulatory subunit B', epsilon isoform	
B3	111427724	111428887	ENSFCAG00000025276	WDR89	WD repeat domain 89	
B3	111434101	111434478	ENSFCAG00000029391			
B3	111477502	111477894	ENSFCAG00000013027			
B3	111505184	111540616	ENSFCAG00000031669	SGPP1	sphingosine-1-phosphate phosphatase 1	
B3	114616480	114684401	ENSFCAG00000011076	MPP5	membrane protein, palmitoylated 5 (MAGUK p55 subfamily member 5)	
B3	114689401	114711272	ENSFCAG00000011078	ATP6V1D	ATPase, H+ transporting, lysosomal 34kDa, V1 subunit D	
B3	114715814	114732784	ENSFCAG00000031891	EIF2S1	eukaryotic translation initiation factor 2, subunit 1 alpha, 35kDa	
B3	114733231	114742902	ENSFCAG00000011080	PLEK2	pleckstrin 2	
B3	114809050	114809553	ENSFCAG00000031183	TMEM229B	transmembrane protein 229B	
B3	114877701	114926189	ENSFCAG00000014084	PLEKHH1	pleckstrin homology domain containing, family H (with MyTH4 domain) member 1	X
B4	39020791	39064360	ENSFCAG00000029824	AKAP3	A kinase (PRKA) anchor protein 3	
B4	39067047	39103662	ENSFCAG00000005951			
B4	39141172	39189805	ENSFCAG00000005953	GALNT8	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 8 (GalNAc-T8)	
B4	39221313	39222899	ENSFCAG00000002340	KCNA6	potassium voltage-gated channel, shaker-related subfamily, member 6	
B4	51667745	51689061	ENSFCAG00000024292	STRAP	serine/threonine kinase receptor associated protein	
B4	51747458	51817466	ENSFCAG00000012551	DERA	deoxyribose-phosphate aldolase (putative)	
B4	83180790	83186313	ENSFCAG00000010156			
B4	83201595	83207215	ENSFCAG00000031316			
B4	83212103	83250831	ENSFCAG00000010157	NCKAP1L	NCK-associated protein 1-like	
B4	83256882	83284307	ENSFCAG00000010158	PDE1B	phosphodiesterase 1B, calmodulin-dependent	
B4	83285929	83289181	ENSFCAG00000010159	PPP1R1A	protein phosphatase 1, regulatory (inhibitor) subunit 1A	
B4	85201642	85213246	ENSFCAG00000012017	ANKRD52	ankyrin repeat domain 52	
B4	85212877	85212953	ENSFCAG00000021736			
B4	85219509	85222879	ENSFCAG00000012018	COQ10A	coenzyme Q10 homolog A (S. cerevisiae)	
B4	85223591	85251393	ENSFCAG00000012019	CS	Citrate synthase	
B4	85256805	85258874	ENSFCAG00000012020			
B4	85261031	85272858	ENSFCAG00000012021	PAN2	PAN2 poly(A) specific ribonuclease subunit homolog (S. cerevisiae)	

Dataset S2.2(c). Genes underlying regions of low H_p in the pooled domestic cat variant dataset following annotation of 100kb windows that fell below four standard deviations from the mean H_p

Chromosome Name	Gene Start (bp)	Gene End (bp)	Ensembl Gene ID	Associated Gene Name	Description	Overlap With F_{ST} List
B4	85277282	85279111	ENSFCAG00000012022	IL23A	interleukin 23, alpha subunit p19	
B4	85281059	85298979	ENSFCAG00000012023	STAT2	signal transducer and activator of transcription 2, 113kDa	
B4	85300402	85302055	ENSFCAG00000028837	APOF	apolipoprotein F	
B4	85313463	85330693	ENSFCAG00000012024	TIMELESS	timeless circadian clock	
B4	85345392	85348942	ENSFCAG00000012025	MIP	major intrinsic protein of lens fiber	
B4	85359827	85362487	ENSFCAG00000023005	SPRYD4	SPRY domain containing 4	
B4	85360205	85376232	ENSFCAG00000012027	GLS2	glutaminase 2 (liver, mitochondrial)	
B4	85400194	85463992	ENSFCAG00000012028	RBMS2	RNA binding motif, single stranded interacting protein 2	
C1	10640571	10657545	ENSFCAG00000011590	FBLIM1	filamin binding LIM protein 1	
C1	10732129	10795582	ENSFCAG00000010165	SPEN	spen homolog, transcriptional regulator (Drosophila)	
C1	10796361	10827960	ENSFCAG00000023421	ZBTB17	zinc finger and BTB domain containing 17	
C1	10837025	10839250	ENSFCAG00000010166	C1orf64	chromosome 1 open reading frame 64	
C1	10848258	10852249	ENSFCAG00000026415	HSPB7	heat shock 27kDa protein family, member 7 (cardiovascular)	
C1	58101938	58102050	ENSFCAG00000031207	5S_rRNA	5S ribosomal RNA	
C1	78698661	78717765	ENSFCAG00000031703	DNTTIP2	deoxynucleotidyltransferase, terminal, interacting protein 2	
C1	78722799	78738417	ENSFCAG00000026377	GCLM	glutamate-cysteine ligase, modifier subunit	
C1	78772870	78906672	ENSFCAG00000015512	ABCA4	ATP-binding cassette, sub-family A (ABC1), member 4	
C1	84943030	84960266	ENSFCAG00000011031	VCAM1	vascular cell adhesion molecule 1	
C1	85015454	85015558	ENSFCAG00000025299	5S_rRNA	5S ribosomal RNA	
C1	85064668	85071066	ENSFCAG00000023298	EXTL2	exostoses (multiple)-like 2	
C1	85087065	85167096	ENSFCAG00000011033	SLC30A7	solute carrier family 30 (zinc transporter), member 7	
C1	85133519	85134071	ENSFCAG00000028505			
C1	85186588	85225405	ENSFCAG00000022542	DPH5	DPH5 homolog (S. cerevisiae)	
C1	153704054	153753962	ENSFCAG00000004699	GRB14	growth factor receptor-bound protein 14	
C1	154302722	154384425	ENSFCAG00000024761	SCN3A	sodium channel, voltage-gated, type III, alpha subunit	
C1	181824781	181826420	ENSFCAG00000010136			
C1	181987901	182031850	ENSFCAG00000029020	SLC39A10	solute carrier family 39 (zinc transporter), member 10	
C2	77632001	77703504	ENSFCAG00000000693	ATP13A5	ATPase type 13A5	
C2	77706977	77717650	ENSFCAG00000000692			
C2	78412932	78673284	ENSFCAG00000025224	FGF12	fibroblast growth factor 12	
C2	108304270	108463279	ENSFCAG00000012834	PLCH1	phospholipase C, eta 1	
C2	108510488	108510600	ENSFCAG00000031412	5S_rRNA	5S ribosomal RNA	
C2	128014862	128022609	ENSFCAG00000009596	RAB6B	RAB6B, member RAS oncogene family	
C2	128027632	128059443	ENSFCAG00000026858	SRPRB	signal recognition particle receptor, B subunit	
C2	128061749	128090487	ENSFCAG00000009592	TF	transferrin	
C2	128112530	128155615	ENSFCAG00000027859			
C2	128189988	128251128	ENSFCAG00000005146	TOPBP1	topoisomerase (DNA) II binding protein 1	
D1	1607316	1636408	ENSFCAG00000029835	DCUN1D5	DCN1, defective in cullin neddylation 1, domain containing 5 (S. cerevisiae)	
D1	1651507	1992762	ENSFCAG00000028573	DYNC2H1	dynein, cytoplasmic 2, heavy chain 1	
D1	30227928	30233580	ENSFCAG00000011024	SPATA19	spermatogenesis associated 19	
D1	30287365	30324514	ENSFCAG00000007138	IGSF9B	immunoglobulin superfamily, member 9B	
D1	53604863	53604975	ENSFCAG00000030335	5S_rRNA	5S ribosomal RNA	
D1	107809557	107915426	ENSFCAG00000008866	UBXN1	UBX domain protein 1 [Source:HGNC Symbol;Acc:18402]	
D1	107850747	107857696	ENSFCAG00000008861	MTA2	metastasis associated 1 family, member 2	
D1	107858491	107867300	ENSFCAG00000008862	EML3	echinoderm microtubule associated protein like 3	
D1	107868913	107870543	ENSFCAG00000008863	ROM1	retinal outer segment membrane protein 1	
D1	107870916	107875681	ENSFCAG00000008864	B3GAT3	beta-1,3-glucuronyltransferase 3 (glucuronosyltransferase I)	
D1	107876904	107892987	ENSFCAG00000008865	GANAB	glucosidase, alpha; neutral AB	
D1	107889470	107890314	ENSFCAG00000030933			
D1	107893127	107897937	ENSFCAG00000022488	INTS5	integrator complex subunit 5	
D1	107901491	107905417	ENSFCAG00000025454			
D1	107906121	107907032	ENSFCAG00000019062	METTL12	methyltransferase like 12	
D1	107910273	107910654	ENSFCAG00000026508	C11orf83	chromosome 11 open reading frame 83	
D1	107918401	107919123	ENSFCAG00000008867	LRRN4CL	LRRN4 C-terminal like	
D1	107920655	107930057	ENSFCAG00000018428	BSCL2	Berardinelli-Seip congenital lipodystrophy 2 (seipin)	
D1	107931091	107931596	ENSFCAG00000006326			
D1	107937649	107947789	ENSFCAG00000014766			
D1	107948981	107954599	ENSFCAG00000014768	TTC9C	tetratricopeptide repeat domain 9C	
D1	107961224	107964011	ENSFCAG00000027690	ZBTB3	zinc finger and BTB domain containing 3	
D1	107973264	107976548	ENSFCAG00000026072	POLR2G	polymerase (RNA) II (DNA directed) polypeptide G	
D1	107984157	107992940	ENSFCAG00000022346	TAF6L	TAF6-like RNA polymerase II, p300/CBP-associated factor (PCAF)-associated factor, 65kDa	
D1	107993688	107994689	ENSFCAG00000014773	TMEM179B	transmembrane protein 179B	
D1	107995077	107996089	ENSFCAG00000025098	TMEM223	transmembrane protein 223	
D1	107996609	108006783	ENSFCAG00000014774	NXF1	nuclear RNA export factor 1	
D1	108008793	108027809	ENSFCAG00000014778	STX5	syntaxin 5	
D1	108029559	108035227	ENSFCAG00000014780	WDR74	WD repeat domain 74	
D1	108066532	108073223	ENSFCAG00000014781	SLC3A2	solute carrier family 3 (activators of dibasic and neutral amino acid transport), member 2	
D1	108091101	108092483	ENSFCAG00000000336	CHRM1	cholinergic receptor, muscarinic 1	
D1	111390904	111406979	ENSFCAG00000007404	SYT12	synaptotagmin XII	
D1	111418646	111423350	ENSFCAG00000007405	RHOD	ras homolog family member D	
D1	111493966	111569206	ENSFCAG00000003383	KDM2A	lysine (K)-specific demethylase 2A	

Dataset S2.2(d). Genes underlying regions of low H_p in the pooled domestic cat variant dataset following annotation of 100kb windows that fell below four standard deviations from the mean H_p

Chromosome Name	Gene Start (bp)	Gene End (bp)	Ensembl Gene ID	Associated Gene Name	Description	Overlap With F_{ST} List
D1	111590001	111597609	ENSFCAG00000003386	ADRBK1	adrenergic, beta, receptor kinase 1	
D1	111598900	111611310	ENSFCAG00000003388	ANKRD13D	ankyrin repeat domain 13 family, member D	
D1	111612360	111620131	ENSFCAG00000003389	SSH3	slingshot homolog 3 (Drosophila)	
D1	111642857	111644862	ENSFCAG00000026317	POLD4	polymerase (DNA-directed), delta 4, accessory subunit	
D2	129076	129188	ENSFCAG00000029175	5S_rRNA	5S ribosomal RNA	
D2	57260928	57261537	ENSFCAG00000031023			
D2	57304321	57346173	ENSFCAG00000014012	ENTPD1	ectonucleoside triphosphate diphosphohydrolase 1	
D2	57377604	57453891	ENSFCAG00000024547			
D3	16753418	16765475	ENSFCAG00000023496	UNC119B	unc-119 homolog B (C. elegans)	
D3	16767557	16779781	ENSFCAG00000002402	ACADS	acyl-CoA dehydrogenase, C-2 to C-3 short chain	
D3	16818395	16846475	ENSFCAG00000030187	SPPL3	signal peptide peptidase like 3	
D3	27486649	27511914	ENSFCAG00000004294	UFD1L	ubiquitin fusion degradation 1 like (yeast)	
D3	27514585	27517737	ENSFCAG00000025374	C22orf39	chromosome 22 open reading frame 39	
D3	27522219	27526181	ENSFCAG00000002752	MRPL40	mitochondrial ribosomal protein L40	
D3	27545550	27619849	ENSFCAG00000002750	HIRA	HIR histone cell cycle regulation defective homolog A (S. cerevisiae)	
D3	27661078	27763200	ENSFCAG00000002747	CLTCL1	clathrin, heavy chain-like 1	
D3	28412397	28523536	ENSFCAG00000006001	PI4KA	phosphatidylinositol 4-kinase, catalytic, alpha	
D3	28460397	28469039	ENSFCAG00000006002	SERPIND1	serpin peptidase inhibitor, clade D (heparin cofactor), member 1	
D3	28532228	28532992	ENSFCAG00000003742			
D3	28549103	28552706	ENSFCAG00000023903	HIC2	hypermethylated in cancer 2	
D3	28587145	28591887	ENSFCAG00000008005			
D3	28601226	28653670	ENSFCAG00000008008	UBE2L3	ubiquitin-conjugating enzyme E2L 3	
D3	28659840	28661374	ENSFCAG00000008011	YDJC	YdjC homolog (bacterial)	
D3	28663783	28666362	ENSFCAG00000027678	CCDC116	coiled-coil domain containing 116	
D3	28672570	28674479	ENSFCAG00000008014	SDF2L1	stromal cell-derived factor 2-like 1	
D3	28677737	28677830	ENSFCAG00000024408			
D3	28678081	28678140	ENSFCAG00000029099			
D3	28680749	28690369	ENSFCAG00000031033			
D3	28692128	28720327	ENSFCAG00000002622	PPIL2	peptidylprolyl isomerase (cyclophilin)-like 2	
D3	28724636	28743087	ENSFCAG00000028941	YPEL1	yippee-like 1 (Drosophila)	
D3	28767507	28877605	ENSFCAG00000023435	MAPK1	mitogen-activated protein kinase 1	
D3	28905826	28929444	ENSFCAG00000002630	PPM1F	protein phosphatase, Mg2+/Mn2+ dependent, 1F	
D3	28941162	28961110	ENSFCAG00000002631	TOP3B	topoisomerase (DNA) III beta	
D3	29050629	29085036	ENSFCAG00000011848			
D3	29119195	29120502	ENSFCAG00000004058	VPREB3	pre-B lymphocyte 3	
D3	29124112	29125589	ENSFCAG00000004065	C22orf15	chromosome 22 open reading frame 15	
D3	29126680	29128333	ENSFCAG00000004059	CHCHD10	coiled-coil-helix-coiled-coil-helix domain containing 10	
D3	29130819	29140423	ENSFCAG00000015309	MMP11	matrix metallopeptidase 11 (stromelysin 3)	
D3	29141586	29176322	ENSFCAG00000004068	SMARCB1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily b, member 1	
D3	29178943	29181128	ENSFCAG00000004069	DERL3	derlin 3	
D3	29194518	29214625	ENSFCAG00000004070			
D3	29230481	29231130	ENSFCAG00000004071	MIF	macrophage migration inhibitory factor (glycosylation-inhibiting factor)	
D3	32440361	32440473	ENSFCAG00000028017	5S_rRNA	5S ribosomal RNA	
D3	73217674	73955407	ENSFCAG00000012953	DCC	deleted in colorectal carcinoma	X
E1	29838089	29850820	ENSFCAG00000018819	PRR11	proline rich 11	
E1	29867217	29867300	ENSFCAG00000022665			
E1	29878598	29878667	ENSFCAG00000023049			
E1	29884263	29908906	ENSFCAG00000022802			
E1	29914732	30045189	ENSFCAG00000013333	TRIM37	tripartite motif containing 37	
E2	45291420	45336043	ENSFCAG00000003492	CTCF	CCCTC-binding factor (zinc finger protein)	
E2	45340148	45352135	ENSFCAG00000009279	RLTPR	RGD motif, leucine rich repeats, tropomodulin domain and proline-rich containing	
E2	45344436	45344645	ENSFCAG00000023723			
E2	45352266	45355166	ENSFCAG00000009280	ACD	adrenocortical dysplasia homolog (mouse)	
E2	45355872	45357527	ENSFCAG00000009281	PARD6A	par-6 partitioning defective 6 homolog alpha (C. elegans)	
E2	45358052	45361200	ENSFCAG00000009282	ENKD1	enkurin domain containing 1	
E2	45361603	45363331	ENSFCAG00000026033	C16orf86	chromosome 16 open reading frame 86	
E2	45369230	45375475	ENSFCAG00000003493	GFOD2	glucose-fructose oxidoreductase domain containing 2	
E2	45408032	45469729	ENSFCAG00000003494	RANBP10	ran-binding protein 10	
E2	45440127	45440219	ENSFCAG00000029801			
E2	45468767	45482371	ENSFCAG00000012848	TSNAXIP1	translin-associated factor X interacting protein 1	
E2	45482659	45488605	ENSFCAG00000012849	CENPT	centromere protein T	
E2	45492005	45492928	ENSFCAG00000012850	THAP11	THAP domain containing 11	
F2	78455289	78455684	ENSFCAG00000023697			
F2	78470381	78470478	ENSFCAG00000031046			

Dataset S2.3(a). Genes underlying regions of high F_{ST} in the pooled domestic cat variant dataset relative to the pooled wildcat variant dataset

Chromosome Name	Gene Start (bp)	Gene End (bp)	Ensembl Gene ID	Associated Gene Name	Description	Overlap With Low Domestic H_p
A1	11398212	11455078	ENSFCAG00000025587	BRCA2	breast cancer type 2 susceptibility protein homolog	
A1	11458331	11465570	ENSFCAG00000024257	N4BP2L1	NEDD4 binding protein 2-like 1	
A1	11495916	11500421	ENSFCAG00000022569	N4BP2L2	NEDD4 binding protein 2-like 2	
A1	11500635	11566192	ENSFCAG00000027199			
A1	117311206	117312333	ENSFCAG00000001278	CD14	CD14 molecule	
A1	117316550	117322265	ENSFCAG00000001280	TMCO6	transmembrane and coiled-coil domains 6	
A1	117322928	117325000	ENSFCAG00000001282	NDUFA2	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 2, 8kDa	
A1	117325316	117335480	ENSFCAG00000001279	IK	IK cytokine, down-regulator of HLA II	
A1	117339751	117346166	ENSFCAG00000031277	WDR55	WD repeat domain 55	
A1	117344253	117346522	ENSFCAG00000001285	DND1	dead end homolog 1 (zebrafish)	
A1	117347249	117360519	ENSFCAG00000001286	HARS	histidyl-tRNA synthetase	
A1	117360312	117367965	ENSFCAG00000001288	HARS2	histidyl-tRNA synthetase 2, mitochondrial	
A1	117368560	117373405	ENSFCAG00000001289	ZMAT2	zinc finger, matrin-type 2	
A1	117462646	117535510	ENSFCAG00000022810	PCDHA1	protocadherin alpha 1	X
A1	117694636	117729173	ENSFCAG00000013156	PCDHB4	protocadherin beta 4	X
A1	117745827	117769414	ENSFCAG00000003467	PCDHB14	protocadherin beta 14	
A1	117799797	117801198	ENSFCAG00000029398	SLC25A2	solute carrier family 25 (mitochondrial carrier; ornithine transporter) member 2	
A1	117809591	117810640	ENSFCAG00000025624	TAF7	TAF7 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 55kDa	
A1	117822381	117986226	ENSFCAG00000027095	PCDHGC4	protocadherin gamma subfamily C, 4	
A1	193168465	193168577	ENSFCAG00000026163	5S_rRNA	5S ribosomal RNA	X
A1	193262073	193280914	ENSFCAG00000028091	MFAP3	microfibrillar-associated protein 3	X
A1	193286712	193315803	ENSFCAG00000023708	FAM114A2	family with sequence similarity 114, member A2	X
A1	193353906	193354042	ENSFCAG00000030338			
A1	193479446	193624866	ENSFCAG00000005223	GRIA1	glutamate receptor, ionotropic, AMPA 1	X
A2	18272574	18306642	ENSFCAG00000031387	BSN	bassoon presynaptic cytomatrix protein	
A2	18315503	18324298	ENSFCAG00000028315	APEH	N-acylaminoacyl-peptide hydrolase	
A2	18324835	18329543	ENSFCAG00000025772	MST1	macrophage stimulating 1 (hepatocyte growth factor-like)	
A2	18331960	18361039	ENSFCAG00000022451	RNF123	ring finger protein 123	
A2	18358000	18359547	ENSFCAG00000010675	AMIGO3	adhesion molecule with Ig-like domain 3	
A2	18361478	18363484	ENSFCAG00000010676	GMPPB	GDP-mannose pyrophosphorylase B	
A2	18367013	18386714	ENSFCAG00000010677	IP6K1	inositol hexakisphosphate kinase 1	
A2	18422363	18430166	ENSFCAG00000029853	CDHR4	cadherin-related family member 4	
A2	18433891	18435481	ENSFCAG00000010679	FAM212A	family with sequence similarity 212, member A	
A2	18435671	18444581	ENSFCAG00000010680	UBA7	ubiquitin-like modifier activating enzyme 7	
A2	18450943	18472637	ENSFCAG00000010682	TRAIP	TRAF interacting protein	
A2	18475364	18478459	ENSFCAG00000010683	CAMKV	CaM kinase-like vesicle-associated	
A2	18489446	18493192	ENSFCAG00000024955			
A2	20299485	20562418	ENSFCAG00000015704	POC1A	POC1 centriolar protein homolog A (Chlamydomonas)	
A2	20524053	20596054	ENSFCAG00000015710	DNAH1	dynein, axonemal, heavy chain 1	
A2	20596545	20605161	ENSFCAG00000015711	BAP1	BRCA1 associated protein-1 (ubiquitin carboxy-terminal hydrolase)	
A2	20605964	20617985	ENSFCAG00000015712	PHF7	PHD finger protein 7	
A2	20628755	20638133	ENSFCAG00000015713	SEMA3G	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3G	
A2	20643522	20646338	ENSFCAG00000015714	TNNC1	troponin C type 1 (slow)	
A2	20648138	20682140	ENSFCAG00000015715	NISCH	nischarin	
A2	20684301	20710994	ENSFCAG00000015716	STAB1	stabilin 1	
A2	56358682	56359374	ENSFCAG00000001739	DNAJB8	DnaJ (Hsp40) homolog, subfamily B, member 8	
A2	56408518	56572045	ENSFCAG00000006153	EEFSEC	eukaryotic elongation factor, selenocysteine-tRNA-specific	
B1	104314018	104319723	ENSFCAG00000023700			
B3	562402	612682	ENSFCAG00000010511	TBC1D2B	TBC1 domain family, member 2B	
B3	614460	645400	ENSFCAG00000019139	ADAMTS7	ADAM metallopeptidase with thrombospondin type 1 motif, 7	
B3	692868	709793	ENSFCAG00000012197			
B3	724131	733861	ENSFCAG00000012198	CTSH	cathepsin H	
B3	752155	849199	ENSFCAG00000012199	RASGRF1	Ras protein-specific guanine nucleotide-releasing factor 1	
B3	23976933	24067525	ENSFCAG00000027111	UBE3A	ubiquitin protein ligase E3A	
B3	26496373	26600819	ENSFCAG00000023805	GABRG3	gamma-aminobutyric acid (GABA) A receptor, gamma 3	

Dataset S2.3(b). Genes underlying regions of high F_{ST} in the pooled domestic cat variant dataset relative to the pooled wildcat variant dataset

Chromosome Name	Gene Start (bp)	Gene End (bp)	Ensembl Gene ID	Associated Gene Name	Description	Overlap With Low Domestic H_p
B3	31880462	31899698	ENSFCAG00000022033	RCN2	reticulocalbin 2, EF-hand calcium binding domain	
B3	31953456	31981469	ENSFCAG00000010953	PSTPIP1	proline-serine-threonine phosphatase interacting protein 1	
B3	31986675	32018232	ENSFCAG00000025746	TSPAN3	tetraspanin 3	
B3	32061239	32135509	ENSFCAG00000026195			
B3	32120379	32121889	ENSFCAG00000031049			
B3	32401108	32428309	ENSFCAG00000022539	HMG20A	high mobility group 20A	
B3	32540693	32558379	ENSFCAG00000013212	LINGO1	leucine rich repeat and Ig domain containing 1	
B3	33511989	33516370	ENSFCAG00000000344	CYP1A2	Cytochrome P450 1A2	X
B3	33532067	33538470	ENSFCAG00000002016	CYP1A1	Cytochrome P450 1A1	X
B3	33564717	33611587	ENSFCAG00000002014	EDC3	enhancer of mRNA decapping 3 homolog (S. cerevisiae)	X
B3	33611037	33628940	ENSFCAG00000031747	CLK3	CDC-like kinase 3	X
B3	33648267	33702607	ENSFCAG00000002012	ARID3B	AT rich interactive domain 3B (BRIGHT-like)	X
B3	113916112	114542162	ENSFCAG00000013487	GPHN	gephyrin	
B3	114877701	114926189	ENSFCAG00000014084	PLEKHH1	pleckstrin homology domain containing, family H (with MyTH4 domain) member 1	X
B3	114924265	114931845	ENSFCAG00000014087	PIGH	phosphatidylinositol glycan anchor biosynthesis, class H	
B3	114944868	114977046	ENSFCAG00000014088	ARG2	arginase, type II	
B3	114976632	115008847	ENSFCAG00000014090			
B3	114997195	114997263	ENSFCAG00000021393			
B3	115012454	115028806	ENSFCAG00000014091	RDH11	retinol dehydrogenase 11 (all-trans/9-cis/11-cis)	
B3	115040768	115049236	ENSFCAG00000014092	RDH12	retinol dehydrogenase 12 (all-trans/9-cis/11-cis)	
B3	115064179	115135168	ENSFCAG00000007653	ZFYVE26	zinc finger, FYVE domain containing 26	
B3	115157242	115190375	ENSFCAG00000007657	RAD51B	RAD51 homolog B (S. cerevisiae)	
B3	126320314	126713029	ENSFCAG00000007824	CEP128	centrosomal protein 128kDa	
B3	126725185	126884461	ENSFCAG00000011083	TSHR	thyrotropin receptor precursor	
B4	143787880	143800787	ENSFCAG00000011914	PLXNB2	plexin B2	
B4	143815714	143824099	ENSFCAG00000011915	DENND6B	DENN/MADD domain containing 6B	
B4	143870997	143920737	ENSFCAG00000004431	PPP6R2	protein phosphatase 6, regulatory subunit 2	
B4	143923211	143944516	ENSFCAG00000004434	SBF1	SET binding factor 1	
B4	143955788	143956879	ENSFCAG00000022957	ADM2	adrenomedullin 2	
B4	143960792	143962760	ENSFCAG00000013030	MIOX	myo-inositol oxygenase	
B4	143974923	143979512	ENSFCAG00000013032	LMF2	lipase maturation factor 2	
B4	143979545	143979666	ENSFCAG00000028233			
B4	143979850	143995006	ENSFCAG00000013034	NCAPH2	non-SMC condensin II complex, subunit H2	
B4	143992569	143993369	ENSFCAG00000013035	SCO2	SCO2 cytochrome c oxidase assembly protein	
B4	143998592	144000511	ENSFCAG00000021933	ODF3B	outer dense fiber of sperm tails 3B	
C1	81116342	81201784	ENSFCAG00000027655	PTBP2	polypyrimidine tract binding protein 2	
C1	87218356	87218468	ENSFCAG00000028148	5S_rRNA	5S ribosomal RNA	
C1	87333358	87358352	ENSFCAG00000024312	RNPC3	RNA-binding region (RNP1, RRM) containing 3	
C1	102252296	102252706	ENSFCAG00000029741	HIST2H3D	histone cluster 2, H3d	
C1	102253768	102254160	ENSFCAG00000026932			
C1	102254475	102254892	ENSFCAG00000027797			
C1	102288465	102288845	ENSFCAG00000023970	HIST2H2BE	histone cluster 2, H2be	
C1	102289173	102289562	ENSFCAG00000024806	HIST2H2AC	histone cluster 2, H2ac	
C1	102289719	102290111	ENSFCAG00000028667	HIST2H2AB	histone cluster 2, H2ab	
C1	102300828	102301241	ENSFCAG00000010103	BOLA1	bolA homolog 1 (E. coli)	
C1	102306044	102313944	ENSFCAG00000010104	SV2A	synaptic vesicle glycoprotein 2A	
C1	102328326	102332748	ENSFCAG00000010106	SF3B4	splicing factor 3b, subunit 4, 49kDa	
C1	102334404	102341438	ENSFCAG00000010108	MTMR11	myotubularin related protein 11	
C1	102347722	102369563	ENSFCAG00000010110	OTUD7B	OTU domain containing 7B	
C1	103510046	103513819	ENSFCAG00000005568	RFX5	regulatory factor X, 5 (influences HLA class II expression)	
C1	103523531	103532221	ENSFCAG00000001859	SELENBP1	selenium binding protein 1	
C1	103572933	103576251	ENSFCAG00000001861	PSMB4	proteasome (prosome, macropain) subunit, beta type, 4	
C1	103579750	103629870	ENSFCAG00000001864	POGZ	pogo transposable element with ZNF domain	
C1	103669485	103693186	ENSFCAG00000004072	CGN	cingulin	
C1	103694358	103735432	ENSFCAG00000004073	TUFT1	tuftelin 1	
C1	104304676	104409375	ENSFCAG00000022223	FLG2	filaggrin family member 2	

Dataset S2.3(c). Genes underlying regions of high F_{ST} in the pooled domestic cat variant dataset relative to the pooled wildcat variant dataset

Chromosome Name	Gene Start (bp)	Gene End (bp)	Ensembl Gene ID	Associated Gene Name	Description	Overlap With Low Domestic H_p
C1	104308093	104310222	ENSFCAG00000025915	HRNR	hornerin	
C1	104341706	104342582	ENSFCAG00000028908	FLG	filaggrin	
C1	104478129	104480640	ENSFCAG00000030196	CRNN	cornulin	
C1	142020467	142037136	ENSFCAG00000027956	TNFAIP6	tumor necrosis factor, alpha-induced protein 6	
C1	142068033	142135438	ENSFCAG00000029742	RIF1	RAP1 interacting factor homolog (yeast)	
C1	142149853	142362875	ENSFCAG00000006778	NEB	nebulin	
D1	87552960	87615211	ENSFCAG00000015003	DCDC1	doublecortin domain containing 1	
D3	73217674	73955407	ENSFCAG00000012953	DCC	deleted in colorectal carcinoma	X
E1	2012450	2018802	ENSFCAG00000001360	ASGR1	asialoglycoprotein receptor 1	
E1	2062948	2084818	ENSFCAG00000002824	DLG4	discs, large homolog 4 (Drosophila)	
E1	2071965	2654649	ENSFCAG00000009629	CHD3	chromodomain helicase DNA binding protein 3	
E1	2085784	2091265	ENSFCAG00000002826	ACADVL	acyl-CoA dehydrogenase, very long chain	
E1	2089360	2089473	ENSFCAG00000020094			
E1	2092041	2099011	ENSFCAG00000024333	DVL2	dishevelled, dsh homolog 2 (Drosophila)	
E1	2099343	2103630	ENSFCAG00000025727	PHF23	PHD finger protein 23	
E1	2105175	2106741	ENSFCAG00000030191	GABARAP	GABA(A) receptor-associated protein	
E1	2108629	2114091	ENSFCAG00000018399	CTDNEP1	CTD nuclear envelope phosphatase 1	
E1	2115255	2120631	ENSFCAG00000002827	ELP5	elongator acetyltransferase complex subunit 5	
E1	2121374	2122689	ENSFCAG00000031173			
E1	2133863	2138934	ENSFCAG00000003061	SLC2A4	solute carrier family 2 (facilitated glucose transporter), member 4	
E1	2140924	2144812	ENSFCAG00000018097	YBX2	Y box binding protein 2	
E1	2160740	2165161	ENSFCAG00000010886			
E1	2166294	2168525	ENSFCAG00000030260			
E1	2169030	2180923	ENSFCAG00000010887			
E1	2182103	2185125	ENSFCAG00000030827			
E1	2186123	2197946	ENSFCAG00000010890	ACAP1	ArfGAP with coiled-coil, ankyrin repeat and PH domains 1	
E1	2202287	2203655	ENSFCAG00000010892	TMEM95	transmembrane protein 95	
E1	2213358	2218182	ENSFCAG00000010893	TNK1	tyrosine kinase, non-receptor, 1	
E1	2219417	2222502	ENSFCAG00000010894	PLSCR3	phospholipid scramblase 3	
E1	2239238	2240164	ENSFCAG00000029899			
E1	2245191	2253571	ENSFCAG00000031923	NLGN2	neuroligin 2	
E3	19523974	19531163	ENSFCAG00000006398	TRIM72	tripartite motif containing 72	
E3	19542074	19543165	ENSFCAG00000002608	PYCARD	PYD and CARD domain containing	
E3	19551867	19563123	ENSFCAG00000027928	FUS	fused in sarcoma	
E3	19590723	19599388	ENSFCAG00000002607	PRSS36	protease, serine, 36	
E3	19602108	19605310	ENSFCAG00000002606	PRSS8	protease, serine, 8	
E3	19606046	19617514	ENSFCAG00000002605	KAT8	K(lysine) acetyltransferase 8	
E3	19619822	19623515	ENSFCAG00000002604	BCKDK	branched chain ketoacid dehydrogenase kinase	
E3	19631084	19634172	ENSFCAG00000029684			
E3	19635291	19640292	ENSFCAG00000002602	PRSS53	protease, serine, 53	
E3	19638036	19647917	ENSFCAG00000021918	ZNF646	zinc finger protein 646	
E3	19656066	19658480	ENSFCAG00000008954	ZNF668	zinc finger protein 668	
E3	19670742	19679273	ENSFCAG00000030459	STX4	syntaxin 4	
E3	19700268	19707021	ENSFCAG00000002598	STX1B	syntaxin 1B	
E3	19710473	19714090	ENSFCAG00000002594	HSD3B7	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 7	
E3	19715271	19736956	ENSFCAG00000002588	SETD1A	SET domain containing 1A	
E3	19740874	19745210	ENSFCAG00000023192	ORAI3	ORAI calcium release-activated calcium modulator 3	
E3	19747142	19766168	ENSFCAG00000002585	FBXL19	F-box and leucine-rich repeat protein 19	

Dataset S2.4. Summary of genes underlying regions of elevated F_{ST} and low H_p in the pooled domestic cats.

Genes Underlying Putative Regions of Selection in the Domestic Cat							
Region	**Chr:Pos**	**Gene ID**	**Gene Name**	**Description**	**Domestic $Z(H_p)$**	**$Z(F_{ST})$**	**Wildcat $Z(H_p)$**
1	A1:117462646-117535510	ENSFCAG00000022810	**PCDHA1**	**protocadherin alpha 1**	**-4.6 to -4.2**	**4.0 to 4.5**	**-2.6 to -1.5**
	A1:117694636-117729173	ENSFCAG00000013156	**PCDHB4**	**protocadherin beta 4**			
2	A1:193168465-193168577	ENSFCAG00000026163		5S ribosomal RNA	**-4.4 to -4.1**	**4.5 to 5.2**	**0 to 0.6**
	A1:193262073-193280914	ENSFCAG00000028091	**MFAP3**	**microfibrillar-associated protein 3**			
	A1:193286712-193315803	ENSFCAG00000023708	**FAM114A2**	**family with sequence similarity 114, member A2**			
	A1:193479446-193624866	ENSFCAG00000005223	**GRIA1**	**glutamate receptor, ionotropic, AMPA 1**			
3	B3: 33511989-33516370	ENSFCAG00000000344	**CYP1A2**	**cytochrome P450 1A2**	**-8.9 to -5.6**	**4.7 to 5.2**	**0.7 to 0.8**
	B3:33532067-33538470	ENSFCAG00000002016	**CYP1A1**	**cytochrome P450 1A1**			
	B3:33564717-33611587	ENSFCAG00000002014	**EDC3**	**enhancer of mRNA decapping 3 homolog**			
	B3:33611037-33628940	ENSFCAG00000031747	**CLK3**	**CDC-like kinase 3**			
	B3:33648267-33702607	ENSFCAG00000002012	**ARID3B**	**AT rich interactive domain 3B (BRIGHT-like)**			
4	B3: 114877701-114926189	ENSFCAG00000014084	**PLEKHH1**	**pleckstrin homology domain containing, family H (with MyTH4 domain) member 1**	**-4.6**	**4.1**	**0.3**
5	D3: 73217674-73955407	ENSFCAG00000012953	**DCC**	**deleted in colorectal carcinoma**	**-4.3**	**4.2**	**-0.8**

Dataset S2.5. Pathway enrichment results using all genes underlying regions of elevated F_{ST} in the pooled domestic cats relative to the pooled wildcats

KEGG Pathway	Pathway ID	C	O	E	R	rawP	adjP	Genes
Retinol metabolism	830	64	4	0.20	20.12	4.90E-05	0.0009	**RDH12, CYP1A1, CYP1A2, RDH11**
Systemic lupus erythematosus	5322	136	4	0.42	9.47	0.0009	0.0081	**HIST2H3D, HIST2H2BE, HIST2H2AC, HIST2H2AB**
Metabolic pathways	1100	1130	10	3.51	2.85	0.0029	0.0153	**ARG2, CYP1A1, NDUFA2, CYP1A2, PIGH, GMPPB, RDH12, ACADVL, HSD3B7, RDH11**
Homologous recombination	3440	28	2	0.09	22.99	0.0034	0.0153	**BRCA2, RAD51B**
Tryptophan metabolism	380	42	2	0.13	15.33	0.0076	0.0198	**CYP1A1, CYP1A2**
SNARE interactions in vesicular transport	4130	36	2	0.11	17.88	0.0056	0.0198	**STX4, STX1B**
Axon guidance	4360	129	3	0.40	7.48	0.0077	0.0198	**SEMA3G, DCC, PLXNB2**
NOD-like receptor signaling pathway	4621	58	2	0.18	11.10	0.0141	0.0317	**PSTPIP1, PYCARD**
Aminoacyl-tRNA biosynthesis	970	63	2	0.20	10.22	0.0165	0.0330	**HARS2, HARS**
Metabolism of xenobiotics by cytochrome P450	980	71	2	0.22	9.07	0.0207	0.0339	**CYP1A1, CYP1A2**
Huntington's disease	5016	183	3	0.57	5.28	0.0196	0.0339	**DLG4, NDUFA2, DNAH1**
Wikipathway								
AhR pathway	WP2100	39	3	0.12	24.76	0.0002	0.0034	**CYP1A1, FLG, CYP1A2**
Tryptophan metabolism	WP465	69	3	0.21	13.99	0.0013	0.0055	**CYP1A1, UBE3A, CYP1A2**
Fatty Acid Omega Oxidation	WP206	15	2	0.05	42.91	0.0010	0.0055	**CYP1A1, CYP1A2**
mRNA processing	WP411	132	4	0.41	9.75	0.0008	0.0055	**FUS, SF3B4, PTBP2, CLK3**
Striated Muscle Contraction	WP383	38	2	0.12	16.94	0.0063	0.016	**TNNC1, NEB**
Tamoxifen metabolism	WP691	38	2	0.12	16.94	0.0063	0.016	**CYP1A1, CYP1A2**
NOD pathway	WP1433	39	2	0.12	16.50	0.0066	0.016	**PYCARD, ACAP1**
Estrogen metabolism	WP697	44	2	0.14	14.63	0.0083	0.0176	**CYP1A1, CYP1A2**
Selenium Metabolism and Selenoproteins	WP28	49	2	0.15	13.14	0.0102	0.0193	**SELENBP1, EEFSEC**
cytochrome P450	WP43	65	2	0.20	9.90	0.0175	0.027	**CYP1A1, CYP1A2**
Proteasome Degradation	WP183	65	2	0.20	9.90	0.0175	0.027	**PSMB4, UBA7**
Integrated Pancreatic Cancer Pathway	WP2256	181	3	0.56	5.33	0.0191	0.0271	**BRCA2, DCC, MST1**

USER DATA & PARAMETERS - N= 137 genes submitted, Genes mapped to unique Entrez Gene IDs: 134, Organism: hsapiens, Id Type: gene_symbol, Ref Set: entrezgene, Significance Level: .05, Statistics Test: Hypergeometric, MTC: BH, Minimum: 2, Enrichment Analyses: KEGG and Wikipathways

COLUMN HEADINGS - number of reference genes in the category (C), number of genes in the gene set and also in the category (O), expected number in the category (E), Ratio of enrichment (R), p value from hypergeometric test (rawP), and p value adjusted by the multiple test adjustment (adjP).

Dataset S2.6. Variant calls per individual breed pools

Breed	High Quality SNVs	Breed Specific SNVs (% of Breed Total)	Breed Specific CNVs
Abyssinian	1,515,266	273,261 (18.0)	25,510
Egyptian Mau	2,843,666	517,787 (18.2)	57,109
Maine Coon	5,057,577	866,564 (17.1)	16,724
Norwegian Forest	6,367,368	1,483,201 (23.3)	9,339
Birman	2,094,270	315,732 (15.1)	20,687
Japanese Bobtail	5,606,127	1,140,768 (20.3)	19,921
Turkish Van	4,929,273	1,008,628 (20.5)	19,241
	Mean = 4,059,078	Mean = 800,849	Mean = 24,076

Dataset S2.7. SNPs analyses of *KIT* in the domestic cat

ID	Breed[1]	Type[2]	*Exon(E) or intron(I) nucleotide site*																					
			10	281	396	522	531	-67	1035	1036	-5	1473	1479	-18	1617	+34	2054	2325	-30	-3	+37	2805	2856	2862
			E1	E2	E3	E3	E3	I3	E6	E6	I8	E9	E9	I9	E10	I10	E14	E16	I17	I17	I18	E20	E21	E21
	Wild-type sequence		**G**	**A**	**G**	**T**	**C**	**A**	**G**	**C**	**T**	**G**	**G**	**G**	**T**	**G**	**G**	**C**	**T**	**C**	**T**	**T**	**A**	**C**
4910	PE	Black	G	M	G	Y	C	A	G	C	T	G	G	R	C	G	G	Y	T	C	W	T	R	C
4406	RB	Orange	G	M	R	C	Y	M	G	C	Y	G	G	G	W	R	G	Y	Y	Y	A	Y	G	C
9299	RG	Seal pt	G	A	G	K	C	M	S	M	T	G	G	G	C	R	G	C	Y	Y	A	T	R	Y
5337	PE	White	R	A	R	Y	Y	M	G	C	Y	G	G	G	Y	R	G	C	Y	Y	W	T	A	C
10630	AC	Bicolor	G	M	G	Y	C	A	G	C	T	G	G	R	C	G	G	Y	T	C	W	T	R	C
H1001[3]	RB	Bicolor	G	A	R	Y	Y	A	G	C	T	G	G	G	Y	G	G	C	T	C	T	T	R	Y
5779	RG	Bicolor	G	A	G	T	C	A	G	C	T	G	G	G	C	G	G	C	T	C	W	T	A	C
11555	RG	Bicolor	G	A	G	T	C	A	G	C	T	G	G	R	C	G	R	C	T	C	W	T	A	C
11556	RG	Bicolor	G	A	G	T	C	A	G	C	T	G	G	R	C	G	G	C	T	C	W	T	A	C
10660	BI	Gloves	G	C	G	C	C	A	C	A	T	G	G	A	T	G	G	C	T	C	A	T	R	T
11558	RG	Mitted	G	A	R	K	Y	A	G	C	Y	G	G	G	Y	G	R	C	T	C	T	T	R	C
H1174[3]	RB	Van	G	A	G	T	C	A	G	C	T	G	G	G	C	G	G	C	T	C	T	T	A	C
8592	RB	Van	G	M	G	Y	C	M	S	M	Y	G	G	G	Y	G	G	C	Y	Y	W	T	R	Y
11608	TV	Van	G	A	G	T	C	A	G	C	T	G	G	G	C	G	G	C	T	C	T	T	A	C
11618	TV	Van	G	A	R	Y	C	M	G	C	T	R	R	G	W	G	G	C	Y	Y	W	T	R	C
NM_001009837.3	/	/	G	A	A	C	T	/	G	C	/	G	G	/	T	/	G	C	/	/	/	T	G	C
ENSFCAT00000003113	AB	Cinnamon	/	A	G	T	C	A	G	C	T	G	G	G	T	G	G	C	T	C	T	T	A	C
	Amino Acid change		**A4T**	**N94T**					**E345D**	**H346N**		**A491T**					**R685K**							

[1]Breed designations: RB = random bred, PE = Persian, RG = Ragdoll, AC = American Curl, BI = Birman, TV = Turkish Van, AB = Abyssinian. [2]Type indicates coloration of the cat, including the white spotting patterns. Some colors have epistasis, for example, dominant white. A cat may be a seal point (pt) or a bicolor, but dominant white will override these colors as melanocytes are absent, preventing the expression of the melanin. Alleles for bicolor and van may be different between breeds and may be additive. Birmans all have “gloves’ and are pointed according to breed standards. Mitted Ragdolls may or may not be pointed.
[3]Cats from the WALTHAM pedigree used for the linkage analysis of the *Spotting* locus.

Dataset S2.8. Frequency of the glove haplotype in cat breed

Observed Genotypes/Phenotypes														
Breed	No.	Gloves/mitted			Solid			Ambiguous			Unknown			Frequency c. 1035_1036delinsCA haplotype
		GG	GN	NN	GG	GN	NN	GG	GN	NN	GG	GN	NN	
Birman	182	177	3	2	0	0	0	0	0	0	0	0	0	0.98
Birman outcrosses	5	0	0	0	0	5	0	0	0	0	0	0	0	0.50
Ragdoll	171	1	7	19	0	11	30	0	7	55	0	15	26	0.12
Random Bred	315	0	0	3	2	15	48	2	10	56	4	22	153	0.10
American Shorthair	11	0	0	0	0	0	5	0	0	1	0	0	5	0.00
American Wirehair	3	0	0	0	0	0	0	0	0	1	0	0	2	0.00
Egyptian Mau	6	0	0	0	0	1	5	0	0	0	0	0	0	0.08
Exotic	10	0	0	0	0	0	0	0	0	0	0	1	9	0.05
Japanese Bobtail	12	0	0	0	0	0	4	0	0	7	0	0	1	0.00
Korat	11	0	0	0	0	0	11	0	0	0	0	0	0	0.00
Maine Coon	10	0	0	0	0	2	1	0	0	1	0	0	6	0.10
Manx	13	0	0	0	0	1	1	0	4	7	0	0	0	0.19
Norwegian Forest Cat	11	0	0	0	0	0	3	0	0	3	0	0	5	0.00
Persian	8	0	0	0	0	0	5	0	0	1	0	0	2	0.00
Russian Blue	11	0	0	0	0	0	11	0	0	0	0	0	0	0.00
Seychellois	2	0	2	0	0	0	0	0	0	0	0	0	0	0.50
Siamese	52	0	0	0	0	3	49	0	0	0	0	0	0	0.03
Siberian	20	0	0	0	1	3	2	1	0	12	0	0	1	0.17
Singapura	8	0	0	0	0	0	8	0	0	0	0	0	0	0.00
Snowshoe	2	0	0	2	0	0	0	0	0	0	0	0	0	0.00
Sphynx	14	0	0	0	0	0	0	1	0	2	3	6	2	0.50
Turkish Angora	14	0	0	0	0	0	0	0	0	12	0	0	2	0.00
Turkish Van	20	0	0	0	0	0	1	0	1	17	0	0	1	0.02
TOTAL	911	178	12	26	3	41	183	4	22	176	7	44	215	/

G (gloves) implies the c.1035_1036delinsCA haplotype, the gloves haplotype. N implies the wildtype allele. 1. Gloves/mitted are cats with white feet. 2. Solid implies a cat with no white spotting pattern. 3. Ambiguous implies a cat with a white spotting pattern that is epistatic and may mask the glove pattern, such as bicolor or dominant white. 4. Cats with no phenotypic description available are listed as unknown. 5.Two random bred cats were included from the WALTHAM pedigree.

Dataset S2.9. Primer sequences and PCR condition for the analysis of feline KIT

Genomic Primers					
Exon	Exon Size (bp)	Product Size (bp)	Forward Primer 5'-3'	Reverse Primer 5'-3'	Tm/[Mg^{2+}] µM
1	154	173	TCTGGGGGCTCGGCTTTGC	GTCCGCGGCGCTCTCCCAC	60/1.75
2	270	366	ATGCTTTATTTCGCCAAGGA	TCCAAAGCATAGCATGAAAGAA	58/2.25
3	282	395	GCAAAGAGAAACGTCGGAGT	CCCAGAAGAACGCGAGAA	58/1.75
4	140	237	AGGCCACCGAATAAGTTGTG	CGGGCTGTTTTCCTTGATCCA	58/2.25
5	169	361	GACAGACTTGTCATGATGCTTTATT	CATTTATAGAGATACGCTTG	58/2.25
6	190	248	TTCATTAACATCTTCCCTATGATGAA	AGGCCCTGGTAAGCCAAG	60/2.00
7	116	245	CAGGCCCTTCACAAGTGATT	CCAACACGAGCCACAACTTA	58/2.25
8	115	212	GGTGAGGTTTTCCAGCAGTC	GTCCTTCCCTTACGCATGTC	58/2.25
9	194	295	CTTTCTGGAGTAAATCGGGTTG	TGACTGATATGGCAGGCAGA	60/1.75
10-11	107-127	394	CTGCCAATAGATTGTGATTCC	AAAGCCCCGGCTTCATAC	58/2.25
12-13	105-111	380	ACACCACCACGTGCTCTCT	TTTGAAAGATAATAAAAGGTAATTTGG	58/2.25
14	151	496	TTGCCAGCAGTGTCAATAGG	TTCTGATTTTGTGCCTCGAA	58/1.75
15	92	259	CTCCCCTTTTTCCCATTTTG	GCACTGTTATTGGGGGCTAC	58/2.25
16	128	245	CCTTGCTTTGAGGTTTAATTGCT	CTCCAAAGTGGGGCTTGG	58/1.75
17	123	263	CGAAACACACATCATTCAGAG	GGGTACTTACGTTTCCTTTG	60/1.75
18-19	112-100	456	CCTCAGCAGGAGCAATGTCT	AGGGGAAGCACTATCTGAAGG	58/1.75
20	106	288	GCCCTGGAATTTGAGATTGT	AAAGGTCTTCACCCCCAGAG	60/2.00
21	132	159	GGTGTAGGGACTGGCATGTAA	GAACCAAAAGAAGAGGGATCG	60/1.75
5'UTR	/	185	GeneRacer 5' Primer (Invitrogen)	GAGCAGGAGGAGCAGGACG	62/1.50
Primer name			Allele Specific PCR primers		
KITgloA-VIC		168	GGCATATCCCAAGCCTGACA	AGGCCCTGGTAAGCCAAG	60/1.50
KITgloB-FAM		168	GGCATATCCCAAGCCTGAGC	AGGCCCTGGTAAGCCAAG	60/1.50
Primer name			Microsatellites primers		
UCDC259b		117	AGACCTTCAGAGTTGCCAGTG	TGTCCTCATTACCGTCCTACC	58/2.00
UCDC489		212	GCTCTGCTCCAACATTGC	GGACCATGCTAATCTAATCGAC	58/2.00
UCDC487		158	CCTCCTCCTCAACAACCTG	CTTGAAGCATTGTAGCTGGAAC	58/2.00
UCDC443		148	GCAACTAGCCAGCTCCAG	ACTCCACTTGTTGACGATCC	58/2.00

Dataset S2.10. Pathway enrichment results using all genes underlying regions of low H_p in the pooled domestic cats

KEGG Pathway	Pathway ID	C	O	E	R	rawP	adjP	Genes
Purine metabolism	230	162	6	0.73	8.23	9.80E-05	0.0023	**PDE5A, ENTPD1, POLR2G, PDE1B, POLD4, PNPT1**
Metabolic pathways	1100	1130	16	5.08	3.15	5.84E-05	0.0023	**ACADS, GCLM, ATP6V1D, CYP1A1, POLR2G, GLS2, CYP1A2, GALNT8, B3GAT3, CS, PI4KA, PIGU, GALNT10, POLD4, GANAB, EXTL2**
Ubiquitin mediated proteolysis	4120	135	5	0.61	8.23	0.0004	0.0061	**TRIM37, UBE2K, ITCH, UBE2L3, PPIL2**
Pyrimidine metabolism	240	99	4	0.45	8.98	0.0011	0.0126	**ENTPD1, POLR2G, POLD4, PNPT1**
Axon guidance	4360	129	4	0.58	6.89	0.0028	0.0222	**SEMA7A, MAPK1, DCC, RHOD**
Regulation of actin cytoskeleton	4810	213	5	0.96	5.22	0.0029	0.0222	**SSH3, MAPK1, NCKAP1L, CHRM1, FGF12**
RNA degradation	3018	71	3	0.32	9.39	0.0041	0.0236	**PNPT1, PAN2, EDC3**
Long-term potentiation	4720	70	3	0.31	9.53	0.0039	0.0236	**GRIA1, MAPK1, PPP1R1A**
Homologous recombination	3440	28	2	0.13	15.88	0.0070	0.0268	**TOP3B, POLD4**
Jak-STAT signaling pathway	4630	155	4	0.70	5.74	0.0054	0.0268	**IL23A, IL6ST, IL12B, STAT2**
Protein processing in endoplasmic reticulum	4141	165	4	0.74	5.39	0.0067	0.0268	**EIF2S1, UFD1L, DERL3, GANAB**
Glycosaminoglycan biosynthesis - heparan sulfate	534	26	2	0.12	17.10	0.0061	0.0268	**EXTL2, B3GAT3**
Mucin type O-Glycan biosynthesis	512	30	2	0.13	14.82	0.0081	0.0287	**GALNT10, GALNT8**
African trypanosomiasis	5143	35	2	0.16	12.70	0.0109	0.0358	**VCAM1, IL12B**
Endocytosis	4144	201	4	0.90	4.42	0.0132	0.0405	**PARD6A, ADRBK1, CLTCL1, ITCH**
Tryptophan metabolism	380	42	2	0.19	10.59	0.0154	0.0443	**CYP1A1, CYP1A2**
Wikipathway								
Adipogenesis	WP236	130	7	0.58	11.97	2.21E-06	7.29E-05	**EBF1, MIF, ASIP, BSCL2, AHR, IL6ST, STAT2**
AhR pathway	WP2100	39	3	0.18	17.10	0.0007	0.0115	**CYP1A1, CYP1A2, AHR**
Fatty Acid Omega Oxidation	WP206	15	2	0.07	29.64	0.002	0.022	**CYP1A1, CYP1A2**
Integrated Breast Cancer Pathway	WP1984	68	3	0.31	9.81	0.0036	0.0297	**MAPK1, AHR, SMEK2**
Regulation of Actin Cytoskeleton	WP51	157	4	0.71	5.66	0.0057	0.0336	**SSH3, MAPK1, FGF12, CHRM1**
Physiological and Pathological Hypertrophy of the Heart	WP1528	26	2	0.12	17.10	0.0061	0.0336	**MAPK1, IL6ST**

USER DATA & PARAMETERS - N= 208 genes submitted, Genes mapped to unique Entrez Gene IDs: 194, Organism: hsapiens, Id Type: gene_symbol, Ref Set: entrezgene, Significance Level: .05, Statistics Test: Hypergeometric, MTC: BH, Minimum: 2, Enrichment Analyses: KEGG and Wikipathways
COLUMN HEADINGS - number of reference genes in the category (C), number of genes in the gene set and also in the category (O), expected number in the category (E), Ratio of enrichment (R), p value from hypergeometric test (rawP), and p value adjusted by the multiple test adjustment (adjP).

Dataset S2.11. Pathway enrichment results using genes underlying regions of elevated F_{ST} in the pooled domestic cats relative to the pooled wildcats and genes underlying regions of low H_p in the pooled domestic cats

KEGG Pathway	Pathway ID	C	O	E	R	rawP	adjP	Genes
Metabolic pathways	1100	1130	24	8.33	2.88	4.40E-06	0.0003	**ACADS, ARG2, GCLM, POLR2G, B3GAT3, PIGH, PI4KA, PIGU, HSD3B7, RDH11, ATP6V1D, CYP1A1, NDUFA2, CYP1A2, GLS2, GALNT8, CS, GMPPB, RDH12, GALNT10, ACADVL, POLD4, GANAB, EXTL2**
Homologous recombination	3440	28	4	0.21	19.37	5.16E-05	0.0015	**TOP3B, BRCA2, POLD4, RAD51B**
Ubiquitin mediated proteolysis	4120	135	7	1	7.03	6.86E-05	0.0015	**UBE3A, TRIM37, UBE2K, UBA7, ITCH, UBE2L3, PPIL2**
Axon guidance	4360	129	6	0.95	6.31	0.0004	0.0066	**SEMA7A, SEMA3G, MAPK1, DCC, RHOD, PLXNB2**
Systemic lupus erythematosus	5322	136	6	1	5.98	0.0005	0.0066	**HIST2H3D, HIST2H2BE, HIST3H2BB, HIST2H2AC, HIST2H2AB, HIST3H2A**
NOD-like receptor signaling pathway	4621	58	4	0.43	9.35	0.0009	0.0099	**MAPK1, NLRP3, PYCARD, PSTPIP1**
Purine metabolism	230	162	6	1.19	5.02	0.0013	0.0107	**PDE5A, ENTPD1, POLR2G, PDE1B, POLD4, PNPT1**
Retinol metabolism	830	64	4	0.47	8.48	0.0013	0.0107	**RDH12, CYP1A1, CYP1A2, RDH11**
SNARE interactions in vesicular transport	4130	36	3	0.27	11.30	0.0024	0.0176	**STX4, STX1B, STX5**
Regulation of actin cytoskeleton	4810	213	6	1.57	3.82	0.0052	0.0343	**SSH3, CD14, MAPK1, NCKAP1L, CHRM1, FGF12**
Pyrimidine metabolism	240	99	4	0.73	5.48	0.0063	0.0378	**ENTPD1, POLR2G, POLD4, PNPT1**
Wikipathway								
Adipogenesis	WP236	130	8	0.96	8.35	5.96E-06	0.0003	**EBF1, MIF, BSCL2, SLC2A4, IL6ST, ASIP, AHR, STAT2**
AhR pathway	WP2100	39	4	0.29	13.91	0.0002	0.0051	**CYP1A1, FLG, CYP1A2, AHR**
Integrated Breast Cancer Pathway	WP1984	68	4	0.50	7.98	0.0016	0.0255	**BRCA2, MAPK1, AHR, SMEK2**
Hypothetical Network for Drug Addiction	WP666	35	3	0.26	11.62	0.0022	0.0255	**GRIA1, MAPK1, NISCH**
mRNA processing	WP411	132	5	0.97	5.14	0.003	0.0255	**NXF1, FUS, SF3B4, PTBP2, CLK3**
NOD pathway	WP1433	39	3	0.29	10.43	0.003	0.0255	**NLRP3, PYCARD, ACAP1**
Regulation of Actin Cytoskeleton	WP51	157	5	1.16	4.32	0.0063	0.0357	**SSH3, CD14, MAPK1, CHRM1, FGF12**
Fatty Acid Omega Oxidation	WP206	15	2	0.11	18.08	0.0053	0.0357	**CYP1A1, CYP1A2**
Mitochondrial LC-Fatty Acid Beta-Oxidation	WP368	16	2	0.12	16.95	0.0061	0.0357	**ACADS, ACADVL**

USER DATA & PARAMETERS - N= 345 genes submitted, Genes mapped to unique Entrez Gene IDs: 378, Organism: hsapiens, Id Type: gene_symbol, Ref Set: entrezgene, Significance Level: .05, Statistics Test: Hypergeometric, MTC: BH, Minimum: 2, Enrichment Analyses: KEGG and Wikipathways

COLUMN HEADINGS - number of reference genes in the category (C), number of genes in the gene set and also in the category (O), expected number in the category (E), Ratio of enrichment (R), p value from hypergeometric test (rawP), and p value adjusted by the multiple test adjustment (adjP).

Dataset S2.12(a). Genes underlying regions of high F_{ST} in the pooled domestic cat X-chromosome variant dataset relative to the pooled wildcat variant dataset

Chromosome Name	Gene Start (bp)	Gene End (bp)	Ensembl Gene ID	Associated Gene Name	Description	Overlap With Low Domestic H_P
X	41887246	41908184	ENSFCAG00000002568	HDAC6	histone deacetylase 6	
X	41911909	41912610	ENSFCAG00000011221	ERAS	ES cell expressed Ras	
X	41913754	41919251	ENSFCAG00000011223	PCSK1N	proprotein convertase subtilisin/kexin type 1 inhibitor	
X	41934479	41938238	ENSFCAG00000002569	TIMM17B	translocase of inner mitochondrial membrane 17 homolog B (yeast)	
X	41938194	41942483	ENSFCAG00000002570	PQBP1	polyglutamine binding protein 1	
X	41942562	41951493	ENSFCAG00000002571	SLC35A2	solute carrier family 35 (UDP-galactose transporter), member A2	
X	41952709	41957926	ENSFCAG00000022467	PIM2	pim-2 oncogene	
X	41961788	41988147	ENSFCAG00000002572	OTUD5	OTU domain containing 5	
X	41991981	41998591	ENSFCAG00000002579	KCND1	potassium voltage-gated channel, Shal-related subfamily, member 1	
X	42002178	42025194	ENSFCAG00000002573	GRIPAP1	GRIP1 associated protein 1	
X	42059037	42068974	ENSFCAG00000002574	TFE3	transcription factor binding to IGHM enhancer 3	
X	42086076	42092788	ENSFCAG00000028936	CCDC120	coiled-coil domain containing 120	
X	42093264	42096118	ENSFCAG00000002576	PRAF2	PRA1 domain family, member 2	
X	42096851	42101817	ENSFCAG00000002577			
X	42129150	42136833	ENSFCAG00000003818	GPKOW	G patch domain and KOW motifs	
X	42149508	42149825	ENSFCAG00000028597			
X	42872212	42934297	ENSFCAG00000015279	CCNB3	cyclin B3	X
X	42969241	43078360	ENSFCAG00000015283			X
X	46413121	46442443	ENSFCAG00000008699	GNL3L	guanine nucleotide binding protein-like 3 (nucleolar)-like	
X	46490447	46490559	ENSFCAG00000022607	5S_rRNA	5S ribosomal RNA	
X	49084155	49256064	ENSFCAG00000008079	ARHGEF9	Cdc42 guanine nucleotide exchange factor (GEF) 9	X
X	49256901	49256994	ENSFCAG00000020519			
X	49353869	49353981	ENSFCAG00000030238	5S_rRNA	5S ribosomal RNA	
X	53226451	53260915	ENSFCAG00000030156	ZC4H2	zinc finger, C4H2 domain containing	X
X	53707791	53708303	ENSFCAG00000022560			X
X	53752218	53752715	ENSFCAG00000026187			X
X	53897284	53962444	ENSFCAG00000013821	MSN	moesin	
X	53999183	53999682	ENSFCAG00000031039			
X	54021127	54021234	ENSFCAG00000030106	5S_rRNA	5S ribosomal RNA	
X	57120311	57122215	ENSFCAG00000013727	PJA1	praja ring finger 1, E3 ubiquitin protein ligase	
X	57461705	57487956	ENSFCAG00000000752	FAM155B	family with sequence similarity 155, member B	X
X	62121214	62344089	ENSFCAG00000014128	ZDHHC15	zinc finger, DHHC-type containing 15	
X	62258432	62259126	ENSFCAG00000030281			
X	62407545	62409116	ENSFCAG00000027322	MAGEE2	melanoma antigen family E, 2	
X	62859874	62859986	ENSFCAG00000027539	5S_rRNA	5S ribosomal RNA	
X	62950563	62950675	ENSFCAG00000027233	5S_rRNA	5S ribosomal RNA	
X	62990509	62991208	ENSFCAG00000025177			
X	64307801	64308823	ENSFCAG00000024786	CYSLTR1	cysteinyl leukotriene receptor 1	
X	64322586	64322696	ENSFCAG00000024593	5S_rRNA	5S ribosomal RNA	
X	64453218	64453330	ENSFCAG00000022019	5S_rRNA	5S ribosomal RNA	
X	64983078	64984100	ENSFCAG00000012066	P2RY10	purinergic receptor P2Y, G-protein coupled, 10	
X	64997854	64997966	ENSFCAG00000027429	5S_rRNA	5S ribosomal RNA	
X	65054928	65056011	ENSFCAG00000022899			
X	65077240	65077352	ENSFCAG00000023385	5S_rRNA	5S ribosomal RNA	
X	65091920	65092786	ENSFCAG00000023690			
X	65852774	65860885	ENSFCAG00000002301	TBX22	T-box 22	
X	69515199	69515308	ENSFCAG00000023760	5S_rRNA	5S ribosomal RNA	
X	69533113	69538158	ENSFCAG00000012662	CYLC1	cylicin, basic protein of sperm head cytoskeleton 1	
X	69578192	69578304	ENSFCAG00000022168	5S_rRNA	5S ribosomal RNA	
X	69687439	69829396	ENSFCAG00000009130	RPS6KA6	ribosomal protein S6 kinase, 90kDa, polypeptide 6	
X	70028313	70186830	ENSFCAG00000005781	HDX	highly divergent homeobox	
X	70466196	70466308	ENSFCAG00000028603	5S_rRNA	5S ribosomal RNA	
X	72446722	72446834	ENSFCAG00000024434	5S_rRNA	5S ribosomal RNA	
X	74987292	74987597	ENSFCAG00000024785			
X	76506764	76565232	ENSFCAG00000025687			
X	76745486	76757291	ENSFCAG00000031194			
X	76768969	76774841	ENSFCAG00000027080			
X	76776829	76779222	ENSFCAG00000029591			
X	76853016	76853919	ENSFCAG00000028074			
X	77857129	77857212	ENSFCAG00000028826			
X	77885307	77888089	ENSFCAG00000031149			
X	78058091	78058203	ENSFCAG00000024847	5S_rRNA	5S ribosomal RNA	
X	80450058	80592747	ENSFCAG00000013438	PCDH19	protocadherin 19	X
X	80616295	80616399	ENSFCAG00000023489	5S_rRNA	5S ribosomal RNA	X
X	80649550	80650477	ENSFCAG00000026700	ANXA2	annexin A2	X
X	80741994	80757092	ENSFCAG00000024939	TNMD	tenomodulin	

Dataset S2.12(b). Genes underlying regions of high F_{ST} in the pooled domestic cat X-chromosome variant dataset relative to the pooled wildcat variant dataset

Chromosome Name	Gene Start (bp)	Gene End (bp)	Ensembl Gene ID	Associated Gene Name	Description	Overlap With Low Domestic H_p
X	82163945	82164473	ENSFCAG00000023241			X
X	82185302	82186069	ENSFCAG00000022838			X
X	82201194	82201535	ENSFCAG00000013057	BEX5	brain expressed, X-linked 5	X
X	82233303	82256641	ENSFCAG00000013704			X
X	83182109	83182975	ENSFCAG00000013517	MORF4L2	mortality factor 4 like 2	
X	83212584	83234142	ENSFCAG00000013518	GLRA4	glycine receptor, alpha 4	
X	83215743	83219091	ENSFCAG00000028808	TMEM31	transmembrane protein 31	
X	83280484	83296694	ENSFCAG00000029818	PLP1	proteolipid protein 1	
X	83331532	83332137	ENSFCAG00000000424	RAB9B	RAB9B, member RAS oncogene family	
X	83491286	83491820	ENSFCAG00000026463			
X	83499961	83528634	ENSFCAG00000006496	FAM199X	family with sequence similarity 199, X-linked	
X	83591368	83596306	ENSFCAG00000007095	ESX1	ESX homeobox 1	
X	83679272	83679384	ENSFCAG00000025445	5S_rRNA	5S ribosomal RNA	
X	84656031	84656372	ENSFCAG00000024474			
X	85860177	85861301	ENSFCAG00000026273			
X	85885076	85885160	ENSFCAG00000024140	5S_rRNA	5S ribosomal RNA	
X	103563484	103563596	ENSFCAG00000029982	5S_rRNA	5S ribosomal RNA	
X	103728395	103728504	ENSFCAG00000031724	5S_rRNA	5S ribosomal RNA	
X	105790212	105795802	ENSFCAG00000022024	APLN	apelin	
X	105873179	105898315	ENSFCAG00000008070	XPNPEP2	X-prolyl aminopeptidase (aminopeptidase P) 2, membrane-bound	
X	105876930	105877062	ENSFCAG00000020596			
X	105908858	105920969	ENSFCAG00000005733	SASH3	SAM and SH3 domain containing 3	
X	105928883	105963235	ENSFCAG00000025956	ZDHHC9	zinc finger, DHHC-type containing 9	
X	121763118	121818915	ENSFCAG00000004331	MAMLD1	mastermind-like domain containing 1	
X	121893254	121967078	ENSFCAG00000004332	MTM1	myotubularin 1	
X	125095537	125102445	ENSFCAG00000011399	FAM50A	family with sequence similarity 50, member A	
X	125116074	125128684	ENSFCAG00000011400	PLXNA3	plexin A3	
X	125134531	125136460	ENSFCAG00000029129	LAGE3	L antigen family, member 3	
X	125140852	125143560	ENSFCAG00000025607	UBL4A	ubiquitin-like 4A	
X	125144336	125145769	ENSFCAG00000022989	SLC10A3	solute carrier family 10 (sodium/bile acid cotransporter family), member 3	
X	125152456	125159000	ENSFCAG00000011403	FAM3A	family with sequence similarity 3, member A	
X	125167043	125177827	ENSFCAG00000011404	G6PD	glucose-6-phosphate dehydrogenase	
X	125182326	125199683	ENSFCAG00000029840	IKBKG	inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase gamma	
X	125244852	125245838	ENSFCAG00000028188			

Dataset S2.13(a). Genes underlying regions of low H_p in the pooled domestic cat X-chromosome variant dataset following annotation of 100kb windows that fell below 1.5 standard deviations from the mean H_p

Chromosome Name	Gene Start (bp)	Gene End (bp)	Ensembl Gene ID	Associated Gene Name	Description	Overlap With High F_{ST} (>1.5)
X	21630	63492	ENSFCAG00000025306			
X	69519	101606	ENSFCAG00000015077	PPP2R3B	protein phosphatase 2, regulatory subunit B'', beta	
X	193651	210731	ENSFCAG00000010308			
X	216285	223908	ENSFCAG00000013263			
X	258750	260256	ENSFCAG00000022728			
X	297649	299085	ENSFCAG00000029324			
X	788985	801017	ENSFCAG00000030897	IL3RA	interleukin 3 receptor, alpha (low affinity)	
X	802775	805108	ENSFCAG00000001211	SLC25A6	solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 6	
X	810005	827797	ENSFCAG00000014304	ASMTL	acetylserotonin O-methyltransferase-like	
X	832290	833503	ENSFCAG00000029147			
X	832943	833542	ENSFCAG00000025999			
X	836793	839803	ENSFCAG00000031505	ASMT	acetylserotonin O-methyltransferase	
X	982130	1057940	ENSFCAG00000025503			
X	1076513	1078597	ENSFCAG00000012522	ZBED1	zinc finger, BED-type containing 1	
X	2583245	2591210	ENSFCAG00000022434			
X	3110376	3385224	ENSFCAG00000000375			
X	4013126	4013238	ENSFCAG00000024922	5S_rRNA	5S ribosomal RNA	
X	4042639	4093472	ENSFCAG00000023323	HDHD1	haloacid dehalogenase-like hydrolase domain containing 1	
X	4206218	4291197	ENSFCAG00000024019	STS	steroid sulfatase (microsomal), isozyme S	
X	4770564	4800388	ENSFCAG00000004082	PNPLA4	patatin-like phospholipase domain containing 4	
X	5336558	5408176	ENSFCAG00000004854	KAL1	Kallmann syndrome 1 sequence	
X	6295756	6440915	ENSFCAG00000011563	TBL1Y	transducin (beta)-like 1, Y-linked	
X	6470618	6495329	ENSFCAG00000021932	GPR143	G protein-coupled receptor 143	
X	6513656	6514431	ENSFCAG00000025783			
X	6596578	6677533	ENSFCAG00000011183	SHROOM2	shroom family member 2	
X	6690557	6691192	ENSFCAG00000028772			
X	6710553	6711562	ENSFCAG00000022520			
X	6789548	6865297	ENSFCAG00000011192	WWC3	WWC family member 3	
X	6780137	6953135	ENSFCAG00000007631	CLCN4	chloride channel, voltage-sensitive 4	
X	7112953	7216047	ENSFCAG00000002582	MID1	midline 1 (Opitz/BBB syndrome)	
X	7699738	7710351	ENSFCAG00000022393	HCCS	holocytochrome c synthase	
X	7720243	7791414	ENSFCAG00000002122	ARHGAP6	Rho GTPase activating protein 6	
X	7830735	7835138	ENSFCAG00000023640	AMELX	Amelogenin	
X	8074048	8074539	ENSFCAG00000030523			
X	8099270	8099635	ENSFCAG00000024727			
X	8259014	8271418	ENSFCAG00000002794	MSL3	male-specific lethal 3 homolog (Drosophila)	
X	8513291	8513384	ENSFCAG00000017961			
X	9076358	9162154	ENSFCAG00000006682	FRMPD4	FERM and PDZ domain containing 4	
X	9305798	9308923	ENSFCAG00000027513	TLR8	Toll-like receptor 8	
X	9354701	9356802	ENSFCAG00000022748			
X	9885613	9886440	ENSFCAG00000027185			
X	9931463	9964979	ENSFCAG00000012437	EGFL6	EGF-like-domain, multiple 6	
X	9981863	9982885	ENSFCAG00000024286			
X	10023119	10023736	ENSFCAG00000023992	RAB9A	RAB9A, member RAS oncogene family	
X	10027827	10039663	ENSFCAG00000023912			
X	10039520	10081402	ENSFCAG00000014870	OFD1	oral-facial-digital syndrome 1	
X	10103285	10143181	ENSFCAG00000014877	GPM6B	glycoprotein M6B	
X	10305715	10316101	ENSFCAG00000014428	GEMIN8	gem (nuclear organelle) associated protein 8	
X	10529330	10529398	ENSFCAG00000025105			
X	10538216	10538293	ENSFCAG00000027438			
X	10784014	10784129	ENSFCAG00000022459	5S_rRNA	5S ribosomal RNA	
X	10741200	10911440	ENSFCAG00000011448	GLRA2	glycine receptor, alpha 2	
X	11018197	11033965	ENSFCAG00000029176	FANCB	Fanconi anemia, complementation group B	
X	11041578	11095832	ENSFCAG00000022388	MOSPD2	motile sperm domain containing 2	
X	11187292	11187374	ENSFCAG00000029611			
X	11286501	11288251	ENSFCAG00000031484	CBX4	chromobox homolog 4	
X	11413677	11441669	ENSFCAG00000010484	ASB9	ankyrin repeat and SOCS box containing 9	

Dataset S2.13(b). Genes underlying regions of low H_p in the pooled domestic cat X-chromosome variant dataset following annotation of 100kb windows that fell below 1.5 standard deviations from the mean H_p

Chromosome Name	Gene Start (bp)	Gene End (bp)	Ensembl Gene ID	Associated Gene Name	Description	Overlap With High F_{ST} (>1.5)
X	11450862	11475292	ENSFCAG00000022236	ASB11	ankyrin repeat and SOCS box containing 11	
X	11478140	11488483	ENSFCAG00000010727	PIGA	phosphatidylinositol glycan anchor biosynthesis, class A	
X	11497779	11531651	ENSFCAG00000010485	FIGF	c-fos induced growth factor (vascular endothelial growth factor D)	
X	11532692	11621888	ENSFCAG00000010486	PIR	pirin (iron-binding nuclear protein)	
X	11631474	11678324	ENSFCAG00000010487	BMX	BMX non-receptor tyrosine kinase	
X	11681716	11720452	ENSFCAG00000009320	ACE2	Angiotensin-converting enzyme 2 Processed angiotensin-converting enzyme 2	
X	11746176	11777544	ENSFCAG00000009328	TMEM27	transmembrane protein 27	
X	11828552	11828664	ENSFCAG00000029736	5S_rRNA	5S ribosomal RNA	
X	11788172	11901542	ENSFCAG00000031216	CA5B	carbonic anhydrase VB, mitochondrial	
X	11906051	11933524	ENSFCAG00000009333			
X	11936136	11962265	ENSFCAG00000024224	AP1S2	adaptor-related protein complex 1, sigma 2 subunit	
X	12787190	12837643	ENSFCAG00000006523	TXLNG	taxilin gamma	
X	12837168	12864998	ENSFCAG00000028919	RBBP7	retinoblastoma binding protein 7	
X	12996604	13134277	ENSFCAG00000006527	REPS2	RALBP1 associated Eps domain containing 2	
X	13935586	13935741	ENSFCAG00000018915			
X	14086539	14119271	ENSFCAG00000006437	BEND2	BEN domain containing 2	
X	14172289	14250913	ENSFCAG00000006438	SCML2	sex comb on midleg-like 2 (Drosophila)	
X	17399416	17399526	ENSFCAG00000017180	5S_rRNA	5S ribosomal RNA	
X	17175933	17401205	ENSFCAG00000026807	CNKSR2	connector enhancer of kinase suppressor of Ras 2	
X	17833482	17848892	ENSFCAG00000023168			
X	17988539	18006475	ENSFCAG00000023221			
X	18029515	18031154	ENSFCAG00000010060	ZNF645	zinc finger protein 645	
X	18790237	18791979	ENSFCAG00000013616	DDX53	DEAD (Asp-Glu-Ala-Asp) box polypeptide 53	
X	19109565	19159869	ENSFCAG00000031503	PTCHD1	patched domain containing 1	
X	19250902	19251014	ENSFCAG00000027204	5S_rRNA	5S ribosomal RNA	
X	19308144	19308597	ENSFCAG00000031952			
X	19393444	19413380	ENSFCAG00000004028	PRDX4	peroxiredoxin 4	
X	19420080	19443002	ENSFCAG00000004030	ACOT9	acyl-CoA thioesterase 9	
X	19476530	19479076	ENSFCAG00000004032	SAT1	spermidine/spermine N1-acetyltransferase 1	
X	19912217	19915290	ENSFCAG00000022717			
X	19931048	19934125	ENSFCAG00000026176			
X	20018240	20085226	ENSFCAG00000002134	PDK3	pyruvate dehydrogenase kinase, isozyme 3	
X	20104464	20200081	ENSFCAG00000002136	PCYT1B	phosphate cytidylyltransferase 1, choline, beta	
X	20249684	20553632	ENSFCAG00000002138	POLA1	polymerase (DNA directed), alpha 1, catalytic subunit	
X	20868777	20869679	ENSFCAG00000031841	FELCATV1R3	vomeronasal 1 receptor felCatV1R3	
X	26713585	26713773	ENSFCAG00000028835			
X	26746658	26767210	ENSFCAG00000022029			
X	26860944	26917110	ENSFCAG00000028265			
X	26971105	26971226	ENSFCAG00000029796			
X	30543947	30544592	ENSFCAG00000030015			
X	30661122	30662084	ENSFCAG00000030986	MAGEB16	melanoma antigen family B, 16	
X	30790586	30849071	ENSFCAG00000018231	CXorf22	chromosome X open reading frame 22	
X	38356463	38359279	ENSFCAG00000007766			
X	38214551	38413159	ENSFCAG00000007765	EFHC2	EF-hand domain (C-terminal) containing 2	
X	38876657	39013854	ENSFCAG00000009445	KDM6A	lysine (K)-specific demethylase 6A	
X	39054031	39099121	ENSFCAG00000009449	CXorf36	chromosome X open reading frame 36	
X	40057175	40084354	ENSFCAG00000027707			
X	40134846	40135831	ENSFCAG00000025461	CHST7	carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 7	
X	40163495	40226463	ENSFCAG00000003541	SLC9A7	solute carrier family 9, subfamily A (NHE7, cation proton antiporter 7), member 7	
X	41093658	41108251	ENSFCAG00000008192	ELK1	ELK1, member of ETS oncogene family	
X	41109294	41118283	ENSFCAG00000028638	UXT	ubiquitously-expressed, prefoldin-like chaperone	
X	41127225	41128098	ENSFCAG00000023763			
X	41165141	41169982	ENSFCAG00000027976			
X	41220128	41306403	ENSFCAG00000022818	ZNF81	zinc finger protein 81	
X	41320923	41321464	ENSFCAG00000026947			
X	41394530	41395849	ENSFCAG00000030362			
X	41428902	41431781	ENSFCAG00000008528			

Dataset S2.13(c). Genes underlying regions of low H_p in the pooled domestic cat X-chromosome variant dataset following annotation of 100kb windows that fell below 1.5 standard deviations from the mean H_p

Chromosome Name	Gene Start (bp)	Gene End (bp)	Ensembl Gene ID	Associated Gene Name	Description	Overlap With High F_{ST} (>1.5)
X	41488558	41490795	ENSFCAG00000027919			
X	42702287	42741226	ENSFCAG00000008619	CLCN5	chloride channel, voltage-sensitive 5	
X	42750538	42750651	ENSFCAG00000017254	5S_rRNA	5S ribosomal RNA	
X	42836383	42851018	ENSFCAG00000031319	AKAP4	A kinase (PRKA) anchor protein 4	
X	42872212	42934297	ENSFCAG00000015279	CCNB3	cyclin B3	X
X	42969241	43078360	ENSFCAG00000015283			X
X	43193426	43193530	ENSFCAG00000026630	5S_rRNA	5S ribosomal RNA	
X	48251045	48253353	ENSFCAG00000005774			
X	48582110	48582823	ENSFCAG00000003939	SPIN4	spindlin family, member 4	
X	48716489	48717304	ENSFCAG00000024136			
X	48773862	48773974	ENSFCAG00000031522	5S_rRNA	5S ribosomal RNA	
X	49049755	49049858	ENSFCAG00000027820	5S_rRNA	5S ribosomal RNA	
X	49084155	49256064	ENSFCAG00000008079	ARHGEF9	Cdc42 guanine nucleotide exchange factor (GEF) 9	X
X	53226451	53260915	ENSFCAG00000030156	ZC4H2	zinc finger, C4H2 domain containing	X
X	53639070	53640170	ENSFCAG00000031009			
X	53707791	53708303	ENSFCAG00000022560			X
X	53752218	53752715	ENSFCAG00000026187			X
X	57343796	57343908	ENSFCAG00000030304	5S_rRNA	5S ribosomal RNA	
X	57461705	57487956	ENSFCAG00000000752	FAM155B	family with sequence similarity 155, member B	X
X	57631690	57632265	ENSFCAG00000028821			
X	58391478	58449666	ENSFCAG00000008636	DLG3	discs, large homolog 3 (Drosophila)	
X	58465920	58707117	ENSFCAG00000029781	TEX11	testis expressed 11	
X	59100432	59101007	ENSFCAG00000021936			
X	59107833	59192395	ENSFCAG00000014809			
X	59250509	59286633	ENSFCAG00000002182	OGT	O-linked N-acetylglucosamine (GlcNAc) transferase	
X	59802421	59806881	ENSFCAG00000023693	CITED1	Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 1	
X	59828592	60051076	ENSFCAG00000005065	HDAC8	histone deacetylase 8	
X	60058404	60229276	ENSFCAG00000007121	PHKA1	phosphorylase kinase, alpha 1 (muscle)	
X	69165483	69166568	ENSFCAG00000011964	POU3F4	POU class 3 homeobox 4	
X	73345566	73346382	ENSFCAG00000023461			
X	73348095	73348790	ENSFCAG00000030624			
X	73490123	73490814	ENSFCAG00000026491			
X	75676071	75676176	ENSFCAG00000022654	5S_rRNA	5S ribosomal RNA	
X	75747036	75747148	ENSFCAG00000026298	5S_rRNA	5S ribosomal RNA	
X	80450058	80592747	ENSFCAG00000013438	PCDH19	protocadherin 19	X
X	80616295	80616399	ENSFCAG00000023489	5S_rRNA	5S ribosomal RNA	X
X	80649550	80650477	ENSFCAG00000026700	ANXA2	annexin A2	X
X	82163945	82164473	ENSFCAG00000023241			X
X	82185302	82186069	ENSFCAG00000022838			X
X	82201194	82201535	ENSFCAG00000013057	BEX5	brain expressed, X-linked 5	X
X	82233303	82256641	ENSFCAG00000013704			X
X	82327996	82335121	ENSFCAG00000010017			

Dataset S2.14. Summary of genes underlying regions of elevated F_{ST} and low H_p along the X-chromosome in domestic cats

Genes Underlying Putative Regions of Selection in the Domestic Cat Along the X-Chromosome							
Region	Chr:Pos	Gene ID	Gene Name	Description	Domestic $Z(H_p)$	$Z(F_{ST})$	Wildcat $Z(H_p)$
1	X:42872212-42934297	ENSFCAG00000015279	CCNB3	cyclin B3	-2.4 to -1.7	1.6 to 1.7	-0.8 to 0.6
	X:42969241-43078360	ENSFCAG00000015283		unknown			
2	X:49084155-49256064	ENSFCAG00000008079	ARHGEF9	Cdc42 guanine nucleotide exchange factor (GEF) 9	-2.3	1.5	0.30
3	X:53226451-53260915	ENSFCAG00000030156	ZC4H2	zinc finger, C4H2 domain containing	-2.3 to -1.6	1.5 to 1.8	0.2 to 0.5
	X:53707791-53708303	ENSFCAG00000022560		unknown			
	X:53752218-53752715	ENSFCAG00000026187		unknown			
4	X:57461705-57487956	ENSFCAG00000000752	FAM155B	family with sequence similarity 155, member B	-3.1 to -1.5	1.5	-0.8 to 0.5
5	X:80450058-80592747	ENSFCAG00000013438	PCDH19	protocadherin 19	-2.1 to -1.7	1.60	0.3 to 0.6
	X:80616295-80616399	ENSFCAG00000023489	5S_rRNA	5S ribosomal RNA			
	X:80649550-80650477	ENSFCAG00000026700	ANXA2	annexin A2			
6	X:82163945-82164473	ENSFCAG00000023241		unknown	-2.1	1.7	0.6
	X:82185302-82186069	ENSFCAG00000022838		unknown			
	X:82201194-82201535	ENSFCAG00000013057	BEX5	brain expressed, X-linked 5			
	X:82233303-82256641	ENSFCAG00000013704		unknown			