


RESEARCH ARTICLE

Open Access



# Comparison of village dog and wolf genomes highlights the role of the neural crest in dog domestication

Amanda L. Pendleton<sup>1</sup>, Feichen Shen<sup>1</sup>, Angela M. Taravella<sup>1</sup>, Sarah Emery<sup>1</sup>, Krishna R. Veeramah<sup>2</sup>, Adam R. Boyko<sup>3</sup> and Jeffrey M. Kidd<sup>1,4\*</sup> 

## Abstract

**Background:** Domesticated from gray wolves between 10 and 40 kya in Eurasia, dogs display a vast array of phenotypes that differ from their ancestors, yet mirror other domesticated animal species, a phenomenon known as the domestication syndrome. Here, we use signatures persisting in dog genomes to identify genes and pathways possibly altered by the selective pressures of domestication.

**Results:** Whole-genome SNP analyses of 43 globally distributed village dogs and 10 wolves differentiated signatures resulting from domestication rather than breed formation. We identified 246 candidate domestication regions containing 10.8 Mb of genome sequence and 429 genes. The regions share haplotypes with ancient dogs, suggesting that the detected signals are not the result of recent selection. Gene enrichments highlight numerous genes linked to neural crest and central nervous system development as well as neurological function. Read depth analysis suggests that copy number variation played a minor role in dog domestication.

**Conclusions:** Our results identify genes that act early in embryogenesis and can confer phenotypes distinguishing domesticated dogs from wolves, such as tameness, smaller jaws, floppy ears, and diminished craniofacial development as the targets of selection during domestication. These differences reflect the phenotypes of the domestication syndrome, which can be explained by alterations in the migration or activity of neural crest cells during development. We propose that initial selection during early dog domestication was for behavior, a trait influenced by genes which act in the neural crest, which secondarily gave rise to the phenotypes of modern dogs.

**Keywords:** Domestication, Canine, Selection scan, Neural crest, Retinoic acid

## Background

The process of animal domestication by humans was complex and multi-staged, resulting in disparate appearances and behaviors of domesticates relative to their wild ancestors [1–3]. In 1868, Darwin noted that numerous traits are shared among domesticated animals, an observation that has since been classified as the domestication syndrome [4]. This syndrome describes the phenomenon

where diverse phenotypes are shared among phylogenetically distinct domesticated species but absent in their wild progenitors. Such traits include increased tameness, shorter muzzles/snouts, smaller teeth, more frequent estrous cycles, floppy ears, reduced brain size, depigmentation of skin or fur, and loss of hair.

During the domestication process, the most desired traits are subject to selection. This selection process may result in detectable genetic signatures such as alterations in allele frequencies [5–11], amino acid substitution patterns [12–14], and linkage disequilibrium patterns [15, 16]. Numerous genome selection scans have been performed within a variety of domesticated

\* Correspondence: [jmkidd@umich.edu](mailto:jmkidd@umich.edu)

<sup>1</sup>Department of Human Genetics, University of Michigan, Ann Arbor, MI 48109, USA

<sup>4</sup>Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA

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



animal taxa [5–11, 17], and several genes are highlighted as likely associated with the domestication syndrome. This is not unexpected given that more than a dozen diverse behavioral and complex physical traits fall under the syndrome, making it likely that numerous genes with pleiotropic effects contribute through mechanisms which act early in organismal development [18, 19]. For this reason, the putative role of the neural crest in domestication has gained traction [18, 20, 21]. Alterations in neural crest cells number and function can also influence behavior. For example, the adrenal and pituitary systems, which are derived from neural crest cells, influence aggression and the “fight or flight” behavioral reactions, two responses which are lessened in domesticates [22].

No domestic animal has shared more of its evolutionary history in direct contact with humans than the dog (*Canis lupus familiaris*, also referred to as *Canis familiaris*), living alongside humans for more than ten thousand years since domestication from its ancestor the gray wolf (*Canis lupus*). Despite numerous studies, vigorous debate still persist regarding the location, timing, and number of dog domestication events [23–27]. Several studies [5, 8, 26, 28, 29] using related approaches have attempted to identify genomic regions which are highly differentiated between dogs and wolves, with the goal of identifying candidate targets of selection during domestications (candidate domestication regions, CDRs [5]). In these studies, breed dogs either fully or partially represented dog genetic diversity. Most modern breeds arose ~300 years ago [30] and contain only a small portion of the genetic diversity found among the vast majority of extant dogs. Instead, semi-feral village dogs are the most abundant and genetically diverse modern dog populations and have undergone limited targeted selection by humans since initial domestication [24, 31]. These two dog groups represent products of two bottlenecks in the evolution of the domestic dog, the first resulting from the initial domestication of gray wolves, and the second from modern breed formation [32, 33]. Selection scans including breed dog genetic data may therefore confound signatures associated with these two events. Indeed, we recently reported [34] that neither ancient nor modern village dogs could be genetically distinguished from wolves at 18 of 30 previously identified autosomal CDRs [5, 8]. Furthermore, most of these studies employed empirical outlier approaches wherein the extreme tail of differentiated loci is assumed to differ due to the action of selection [35]. Freedman et al. [29] extended these studies through the use of a simulated demographic history to identify loci whose variability is unlikely to result from a neutral population history of bottlenecks and migration. When compared to previous outlier-based studies, most of the regions identified in

[29] were novel, and harbored genes in neurological, behavioral, and metabolic pathways.

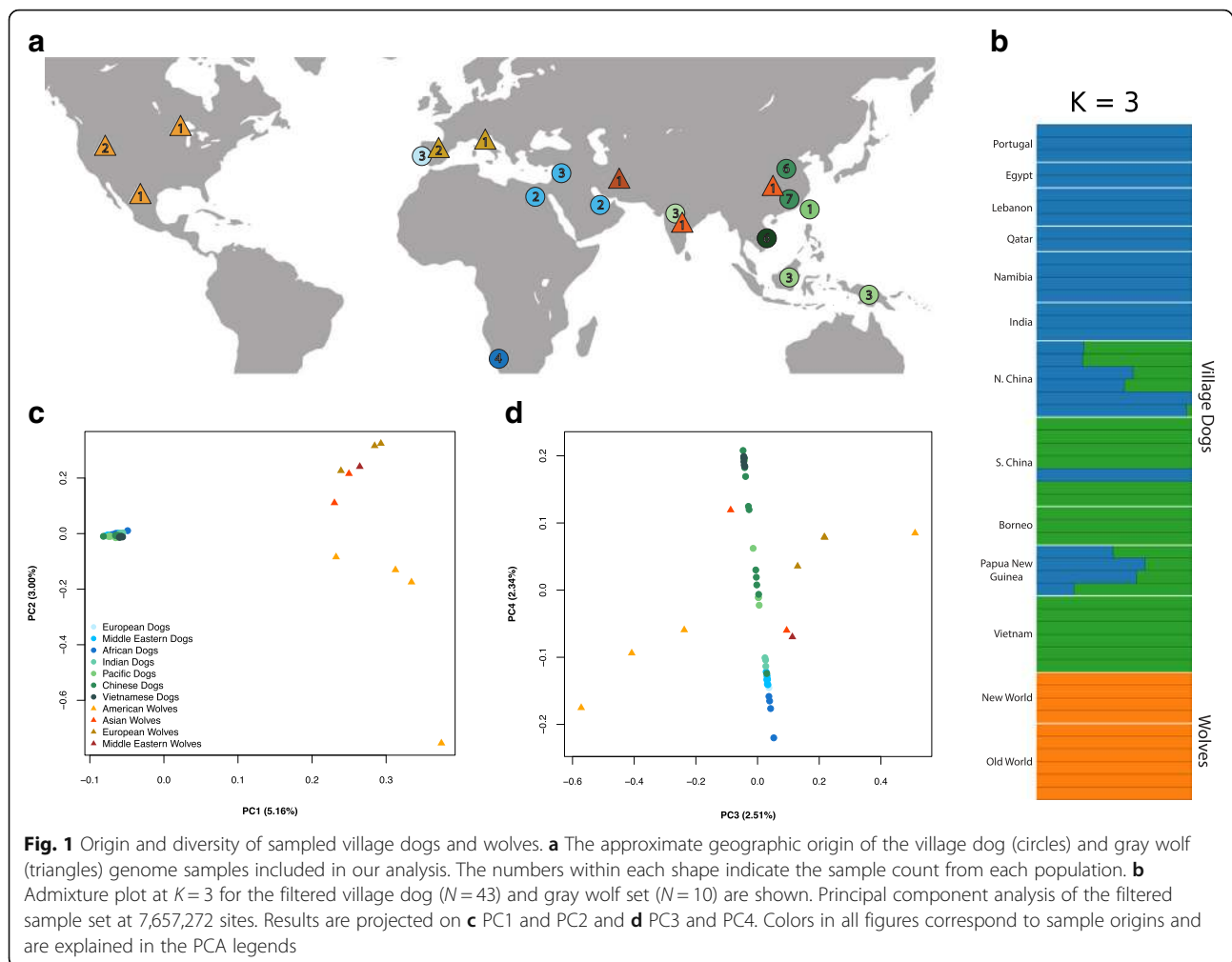
In this study, we reassess candidate domestication regions in dogs using genome sequence data from a globally diverse collection of village dogs and wolves. First, using methods previously applied to breed dog samples, we show that the use of semi-feral village dogs better captures dog genetic diversity and identifies loci more likely to be truly associated with domestication. Next, we perform a scan for CDRs in village dogs utilizing the XP-CLR statistic, refine our results by requiring shared haplotypes with ancient dogs (>5000 years old) and present a revised set of pathways altered during dog domestication. Finally, we perform a scan for copy number differences between village dogs and wolves, and identify additional copy number variation at the starch-metabolizing gene *amylase-2b* (*AMY2B*) that is independent of the *AMY2B* tandem expansion previously found in dogs [5, 36–38].

## Results

### Use of village dogs eliminates bias in domestication scans associated with breed formation

#### Comparison using $F_{ST}$ outlier approaches

Utilizing pooled  $F_{ST}$  calculations in sliding windows along the genome, two previous studies [5, 8] isolated candidate domestication regions from sample sets consisting of mostly breed dogs and wolves. These loci were classified as statistical outliers based on empirical thresholds (arbitrary  $Z$  score cutoffs). In order to demonstrate the impact of sample choice (i.e., breed vs village dogs) on the detection of selective signatures associated with early domestication pressures, rather than breed formation, we adapted the methods from these studies and identified outlier loci empirically [5, 8]. First, through ADMIXTURE [39] and identity-by-state (IBS) analyses, we identified a collection of 43 village dog and 10 gray wolf samples (Additional file 1: Table S1) that have less than 5% dog-wolf admixed ancestry and excludes close relatives (Fig. 1a, b; see the “Methods” section). Principal component analysis (PCA) illustrates the genetic separation between village dogs and wolves along PC’s 1 and 2 (Fig. 1c), while positions along PC4 reflect the east-west geographic distribution of the village dog populations (Fig. 1d). To compare directly with previous studies, we calculated average  $F_{ST}$  values in overlapping 200 kb sliding windows with a step-size of 50 kb across the genome using a pooled approach. As in [5, 8], we performed a  $Z$  transformation of  $F_{ST}$  values to normalize the resulting values and identified windows with a  $ZF_{ST}$  score greater than 5 (autosomes) or 3 (X chromosome) as candidate domestication regions. Following merging, this outlier procedure identified 31 CDRs encompassing 12.3 Mb of sequence (Additional file 1: Table S2). As in previous studies, a 550 kb region on chromosome 6



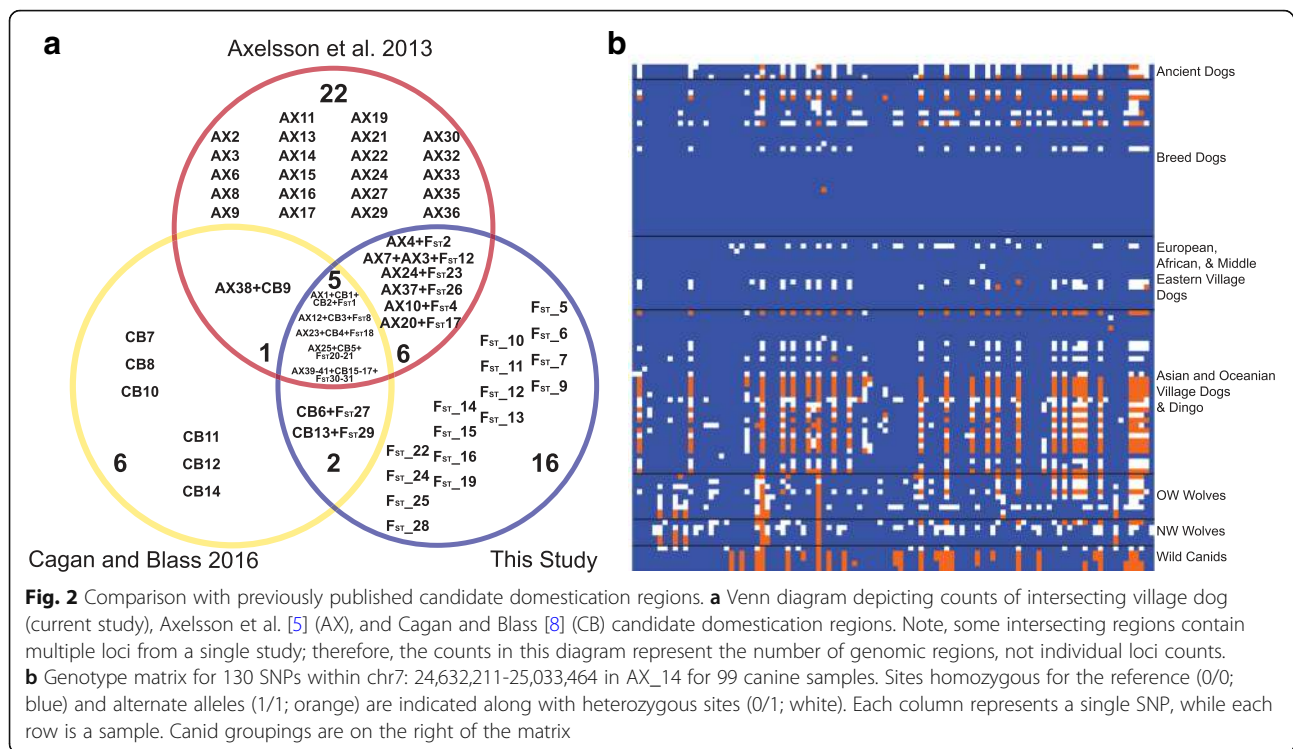
(46.80–47.35 Mb) that contains the *pancreatic amylase 2B* (*AMY2B*) and *RNA Binding Region Containing 3* (*RNPC3*) genes had the highest observed average  $ZF_{ST}$  score ( $ZF_{ST} = 7.67$ ).

Only 15 of these 31 regions intersect with those reported in [5] and [8] (Fig. 2a). To further explore this discrepancy, we visually assessed whether the dog or wolf haplotype is present at the loci reported in these earlier studies in 46 additional canine samples, including three ancient European dogs ranging in age from 5000 to 7000 years old (see the “Methods” section; [23, 34]). Likely due to the absence of village dogs in their study, some loci identified in Axelsson et al. [5] appear to contain selective sweeps associated with breed formation, as evidenced by the presence of the wild haplotype in ancient and village dogs (example in Fig. 2b). Although all autosomal sweeps identified by [8] intersected with CDRs from our study, seven of their X chromosome windows did not meet the thresholds of significance from our SNP sets (example in Additional file 2: Figure S1). Unlike [8], we performed  $F_{ST}$  scans and  $Z$  transformations

for windows on autosomes and the X chromosome separately, which may limit false inflation of  $F_{ST}$  signals on the X that arise due to smaller effective population sizes and correspondingly higher expected levels of genetic drift on the X chromosome. More detailed analysis of the loci highlighted in these two earlier studies [5, 8] will be elaborated in the following section.

#### Refined assessment of previously identified candidate differentiated loci using demographic models and ancient genomes

The above results suggest that the use of village dogs, rather than breed dogs, in selection scans identifies novel candidate domestication regions that are not confounded by breed formation. We developed a statistical filtering strategy to systematically further explore the impact of sample choice on  $F_{ST}$ -based scans. First, rather than setting an empirical threshold at a  $ZF_{ST}$  score of 5, we created a neutral null model that captures key aspects of dog and wolf demographic history (Additional file 1: Table S3; Additional file 2: Figure S2;



[34, 40]). We identified 443 autosomal sliding windows with  $F_{ST}$  values that exceed the 99th percentile of the neutral simulations ( $F_{ST} = 0.308$ ; Additional file 2: Figure S3a). Second, reasoning that a true domestication sweep will be largely fixed among extant dogs with no recent wolf admixture, we calculated pooled heterozygosity ( $H_p$ ) in village dogs within the same window boundaries and retained windows with a  $H_p$  lower than the 0.1th percentile observed in our simulations (Additional file 2: Figure S3b). This heterozygosity filter removed 199 of the 443 windows. Finally, we excluded regions where the putatively selected haplotype is not found in ancient dog samples. To do this, we calculated the difference in dog  $H_p$  ( $\Delta H_p$ ) with and without the inclusion of two ancient dog samples HXH, a 7-ky-old dog from Herxheim, Germany [34] and NGD, a 5-ky-old dog from Newgrange, Ireland [23]; see the “Methods” section). Windows with  $\Delta H_p$  greater than the 5th percentile of all windows genome-wide ( $\Delta H_p = -0.0036$ ) were removed (Additional file 2: Figures S3c, d and S4). Remaining overlapping windows were merged, resulting in 58 autosomal  $F_{ST}$  CDRs that encompass 18.65 Mbp of the genome and are within 50 kb of 248 Ensembl gene models (Fig. 3; Additional file 1: Table S4).

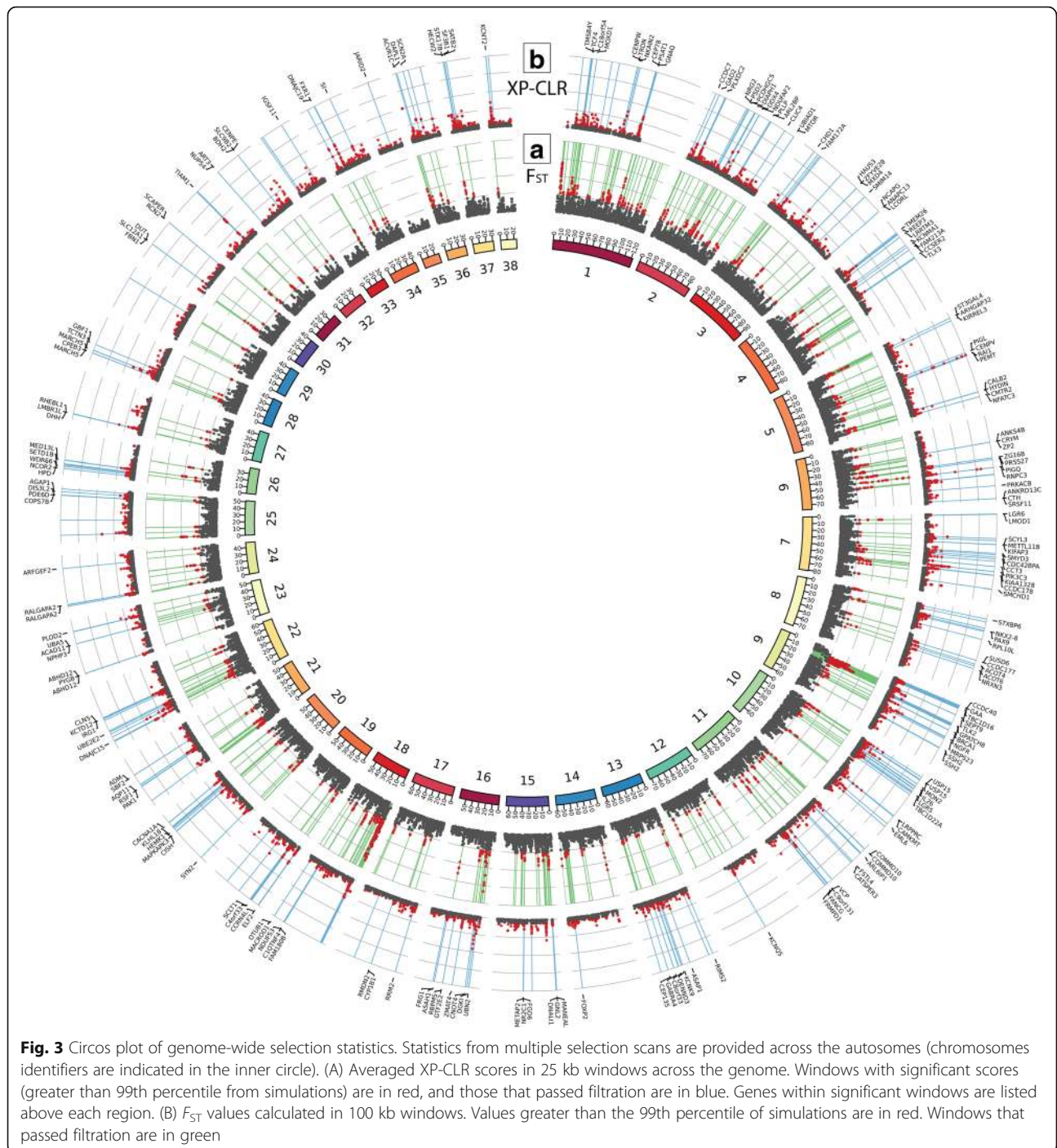
We applied the same filtration parameters to the candidate domestication regions identified on the autosomes in Axelsson et al. ( $N = 30$ ; [5]) and Cagan and Blass ( $N = 5$ ; [8]) (Additional file 2: Figure S5a and b). Since window coordinates of these studies may not

precisely match our own, we selected the maximum  $F_{ST}$  value per locus from our village dog and wolf data. We then removed any locus with  $F_{ST}$ ,  $H_p$  and  $\Delta H_p$  levels not passing our thresholds. Following these three filtration steps, only 14 Axelsson and 4 Cagan and Blass loci remained. In addition, we separately assessed the overlap of our  $F_{ST}$ -based regions with the 349 loci identified by [29] using various statistics and a simulation-based significance threshold which is more comparable to our approach. We found that only 41 of the 349 loci from [29] loci passed our filtrations (Additional file 2: Figure S5c). In total, 25/58 loci identified using  $F_{ST}$  in village dogs intersected with a putative sweep identified from at least one previous study (for specific overlaps, see Additional file 1: Table S4). The fact that the majority of the previously reported CDRs fail our thresholds when examined in village dogs and ancient dogs suggest that these CDRs reflect selection events that occurred in breeds after dog domestication, rather than true domestication sweeps which should be present in all dogs.

#### A scan for the targets of selection during domestication using cross-population haplotype comparisons

To gain a better picture of the targets of selection during dog domestication, we conducted a search for domestication regions in village dogs using XP-CLR, a statistic developed to identify loci under selection based on patterns of correlated multilocus allele frequency differences between two populations [41]. XP-CLR has several





advantages over other methods used to identify selection signatures, as it is less biased by demographic history, by uncertainty in recombination rates, and does not maintain strict window boundaries [41]. Instead, the method considers patterns of contiguous SNPs to isolate loci that, based on the size of the affected region, had more rapid correlated changes in allele frequency than expected by genetic drift [41]. Since we are searching for regions under selection in the dog genome, wolves were

set as our reference population and XP-CLR was run on both simulated and real SNP datasets with a spacing of 2 kb, and a window size of 50 kb. Average XP-CLR values were calculated within 25 kb sliding windows (10 kb step size) for both datasets, and we retained 889 windows with scores greater than the 99th percentile obtained from simulations (XP-CLR = 19.78; Additional file 2: Figure S6a). Using methods similar to those employed for the  $F_{ST}$  scans described

above, windows with village dog  $H_P$  values less than the 0.1st simulation percentile ( $H_P = 0.0598$ ) or where the ancient dog samples carried a different haplotype ( $\Delta H_P$  filtration threshold at 5th percentile =  $-0.0066$ ) were eliminated (Additional file 2: Figures S6b–d and S3c). This resulted in 598 autosomal windows which we merged into 246 candidate loci, encompassing 10.81 Mb of genomic sequence and within 50 kb of 429 unique genes (Fig. 3b; Additional file 1: Table S5). Of these windows, 178 are located within 50 kb of at least one an Ensembl gene model. No SNPs with high  $F_{ST}$  within these intervals had predicted deleterious effects on coding sequence. (Additional file 1: Table S6; [42]). The vast majority of the XP-CLR regions (204/246) were not found in previous studies [5, 8, 29], with 4 also found in Axelsson et al. [5] only, 33 in Freedman et al. [29] only and 5 in both Axelsson et al. [5] and Freedman et al. [29]. No loci intersected with the Cagan and Blass [8] findings. Thirty four XP-CLR regions overlap with 21 of the 58 loci we identified using  $F_{ST}$ -based approaches, indicating that XP-CLR often identifies selection signatures within narrower regions.

#### Gene content of 246 candidate domestication regions

We sought to identify gene sets and pathways enriched within our candidate domestication regions. Based on 1000 randomized permutations (see the “Methods” section), we found that the XP-CLR regions are not more likely to localize near genes than expected ( $p = 0.07$ ), though the loci are near a greater total number of genes than random permutations ( $p = 0.003$ ; Additional file 2: Figure S7a and b). We observed that our candidate loci contain genes of the similar average length as found in the randomized set ( $p > 0.05$ ; Additional file 2: Figure S7c). The biological functions of numerous genes near the candidate domestication regions are consistent with the neural crest hypothesis, linking this critical embryonic development pathway to the domestication syndrome (Table 1; [18, 20, 21]). Multiple genes are also involved in retinoic acid signaling, neurotransmission, and RNA splicing.

#### Candidate genes influencing retinoic acid signaling

Retinoic acid (RA) is a signaling molecule that has numerous critical roles in development at the embryonic level, continuing into adult stages with roles such as maintaining stem cell proliferation, tissue regeneration, and regulation of circadian rhythm [43, 44]. The highest scoring XP-CLR locus centers upon *RAI1* (*retinoic acid-induced 1*; XP 52; Fig. 4), a gene that has not been identified in previous domestication scans. *RAI1* has numerous developmental functions in the RA pathway, and mutations in this gene are responsible for Smith-Magenis and Potocki-Lupski syndromes in humans [45, 46]. Other genes with related functions include *NR2C1* (XP 143),

essential for the development of early retina cells through regulation of early transcription factors that govern retinal progenitor cells such as RA receptors [47] and calreticulin, a protein involved in inhibition of both androgen and RA transcriptional activities [47, 48]. *Ncor2* (XP 209) increases cell sensitivity to RA when knocked out in mice [49], and *CYP11B1* (XP 152) is a pathway component that can direct embryonic patterning by RA [50].

#### Candidate genes regulating brain development and behavior

Twelve XP-CLR candidate genes related to neurotransmitter function include the serotonin transporter *SLC6A4* (XP 101) and dopamine signaling members *GNAQ* (XP 16) and *ADCY6* (XP 215). Genes associated with glutamate, the excitatory neurotransmitter, include *DGKI* (ranked 6th by XP-CLR; XP 145), which regulates presynaptic release in glutamate receptors [51], and *GRIK3* (XP 141), a glutamate receptor [52]. Other genes include *UNC13B*, which is essential for competence of glutamatergic synaptic vesicles [53], and *CACNA1A* (XP 176) influences glutamatergic synaptic transmission [54]. In contrast to glutamate, GABA is the nervous system’s inhibitory neurotransmitter and has been linked to the response to and memory of fear [55, 56]. Genes in our XP-CLR loci relating to GABA include one of the two mammalian GABA biosynthetic enzymes *GAD2* (or *GAD65*; ranked 20th), the GABA receptor *GABRA4*, auxiliary subunit of GABA-B receptors *KCTD12* ([57]), and the GABA inhibitor osteocalcin (or *BGLAP*; [58]). Lastly, *TLX3* (XP 48) is a key switch between glutamatergic and GABAergic cell fates [59].

#### Candidate genes related to RNA splicing

We also observe numerous candidate genes involved in splicing of transcripts by both the major and minor splicing pathways. The eighth highest XP-CLR region (XP 57) harbors the gene *RNPC3*, the 65 KDa subunit of the U12 minor spliceosome, which is located ~ 55 kb downstream of pancreatic amylase *AMY2B* (Fig. 5). Another core subunit, *SF3B1*, belongs to both the minor and major (U2) spliceosome. Additional XP-CLR genes related to splicing and/or spliceosome function include *FRG1* [60], *DDX23* (alias *PRP28*; [61]), *CELF1* [62], *NSR1* (alias *NSrp70*; [63, 64]), and *SRSF11* (alias *P54*; [65]).

#### Survey of copy number variation between dogs and wolves

Copy number variants have also been associated with population-specific selection and domestication in a number of species [5, 66, 67]. Since regions showing extensive copy number variation may not be uniquely localized in the genome reference and may have a deficit of SNPs passing our coverage thresholds, we directly estimated

**Table 1** XP-CLR CDR genes with evidenced or putative roles in nervous system and neural crest pathways

Gene	XP-CLR locus (rank)	System	Phenotypes/effects
<i>RAI1</i>	XP 52 (1st)	Xenopus	Mutants display craniofacial defects, improper migration of neural crest cells, decrease in facial cartilage components, axonal defects, and altered forebrain ventricle sizes [119].
<i>NKAIN2 (TCBA1)</i>	XP 9 (4th)	Human	Neurocristopathy-like phenotypes observed in patients with translocation breakpoint in <i>NKAIN2</i> such as hair hypopigmentation, craniofacial and limb malformation, misdevelopment of eyes, and macrocephaly [172].
<i>RNPC3</i>	XP 57 (8th)	Human	Mutations linked to isolated growth hormone deficiency and pituitary hypoplasia [128].
<i>NPR2</i>	XP 127 (14th)	Human, mouse	Mutants exhibit dwarfism and impacted skeletal growth during embryogenesis [173, 174].
<i>NPHP3</i>	XP 197 (24th)	Mouse, xenopus	Left/right asymmetry, shortened body axes, and neural folds fail to close in mutants. Interacts with non-canonical <i>Wnt</i> pathway [84].
<i>LIMCH1</i>	XP 135 (30th)	Human	Significantly altered methylation patterns in Chinese Han pedigrees exhibiting neural tube defects [175]. <i>LIMCH1</i> depletion increased cell migration by spatiotemporally regulating non-muscular myosin II activity [176].
<i>CCDC65</i>	XP 215 (45th)	Zebrafish	Critical for cilia and dynein function. Knockdowns cause left-right asymmetry and axis curvature embryos [177].
<i>DAND5 (cerberus-like)</i>	XP 177 (51st)	Mouse	Prevents signaling of the Nodal pathway on the right side of the developing mouse embryo, establishing left/right asymmetry during early somitogenesis [178].
<i>GBF1</i>	XP 220 (67th)	Fly	Expressed in embryogenesis, contributes to cell polarity in tubular organs and chemotaxis of neutrophils [179, 180].
<i>GDPD5</i>	XP 181 (102nd)	Zebrafish	Regulator of the notch signaling pathway, essential for neural crest pathway, linked to body axis determination [181], is induced by retinoids and drives motor neuron differentiation [182].
<i>HAUS3</i>	XP 38 (111th)	Zebrafish	Essential regulator of embryonic hematopoietic stem/progenitor cell maintenance and cell cycle progression [183].
<i>PAX9</i>	XP 80 (114th)	Mouse	Mutants displayed improper craniofacial development, lacked organs deriving from pharyngeal pouches, no teeth [184].
<i>DIAPH1</i>	XP 21 (117th)	Human	Expressed in neural progenitors, linked to microcephaly in humans [185], impacts migration of glioma cells [186].
<i>TCF4</i>	XP 3 (127th)	Mouse	Myelinates oligodendrocytes, antagonizes the <i>Wnt</i> signaling pathway, and interacts with <i>SOX10</i> (a known neural crest gene [187]) to promote oligodendrocytic maturation gene expression [188].
<i>TSPAN14</i>	XP 46 (129th)	Human	Promotes the activity of notch receptors and the expression of <i>ADAM10</i> [189], both players in the neural crest signaling pathway [190].
<i>SATB2</i>	XP 244 (131st)	Mouse	Mutants exhibit craniofacial abnormalities (e.g., cleft palate, dental misgrowth) and disrupted osteoblast differentiation [99].
<i>FOXI1</i>	XP 49 (136th)	Zebrafish	Regulates inner ear and jaw development in embryogenesis, and hypothesized to influence neural crest cell migration and/or separation in the brachial arches [191].
<i>PRKCAB</i>	XP 61 (138th)	Mouse	Mutations yield improper development of the neural tube and spina bifida in mice, asymmetric expansion of hedgehog signaling in the neural tube, impact neuronal cell survival [192].
<i>GNAQ</i>	XP 16 (141st)	Mouse	Mutants exhibit heart malformations and shortened jaws [193].
<i>Tlx3</i>	XP 48 (146th)	Mouse	Dorsal spinal cord development, specification of glutamatergic neurons [194], and is a target of <i>Wnt</i> signaling pathway [59].
<i>SEMA4A</i>	XP 70 (150th)	Xenopus	Expressed in neurogenic placodes in the developing neural tube, which along with neural crest cells, migrate to final cell locations [195].
<i>TIAM1</i>	XP 225 (154th)	Mouse	With <i>PAR3</i> , gene is responsible for determination of front-rear and apical-basal polarity in migratory keratinocyte cells [196].
<i>PITX1</i>	XP 124 (157th)	Mouse, anolis	Transcription factor whose binding sites are near key neural crest signaling members ( <i>Wnt</i> , Hedgehog, <i>BMP</i> ) [197]. Mutants have improper hind limb development and patterning as well as craniofacial abnormalities [100].

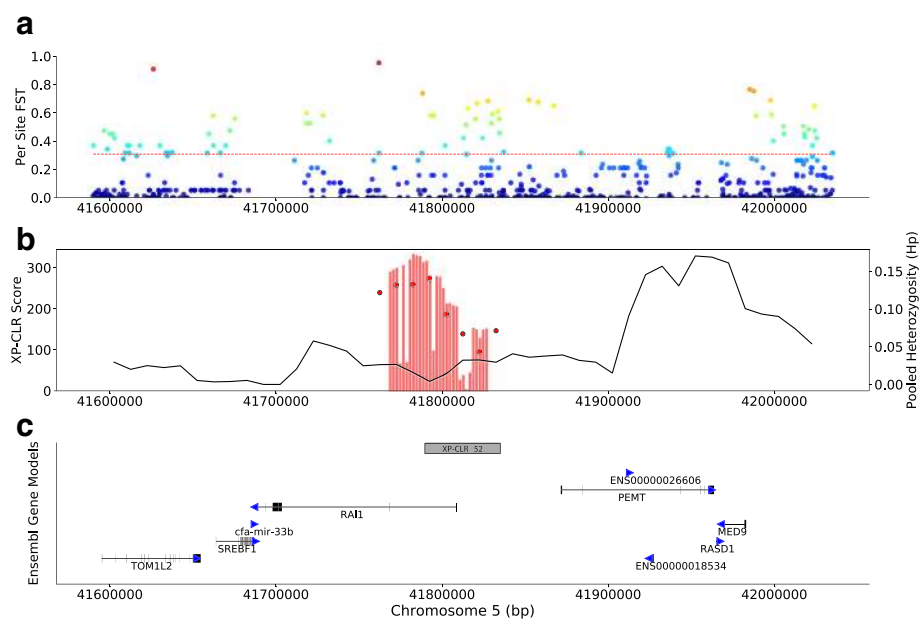
**Table 1** XP-CLR CDR genes with evidenced or putative roles in nervous system and neural crest pathways (Continued)

Gene	XP-CLR locus (rank)	System	Phenotypes/effects
<i>FKBP8</i>	XP 175 (167th)	Mouse	Critical for development of the neural tube, establishes dorso-ventral patterning, and prevents apoptosis in embryonic cells in the neural tube [198].
<i>AMBRA1</i>	XP 161 (169th)	Mouse	Mutants show disrupted embryonic development, neural tube defects, cell cycle perturbations (unbalanced proliferation and high apoptosis) [199].
<i>SCUBE1</i>	XP 115 (202nd)	Mouse	Required for proper development of the central nervous system, neural tube, brain regions, and the cranial vault formation [200].
<i>CYP11B1</i>	XP 152 (229th)	Zebrafish	Mutants showed disrupted neural crest migration [201] and is associated with retinoic acid synthesis during the patterning of the developing embryo [50].

Genes within XP-CLR candidate domestication regions (with rank) that have experimental or clinical evidence that illustrate roles in early embryonic pathways, especially in the developing central nervous system and components of the neural crest and its signaling pathways

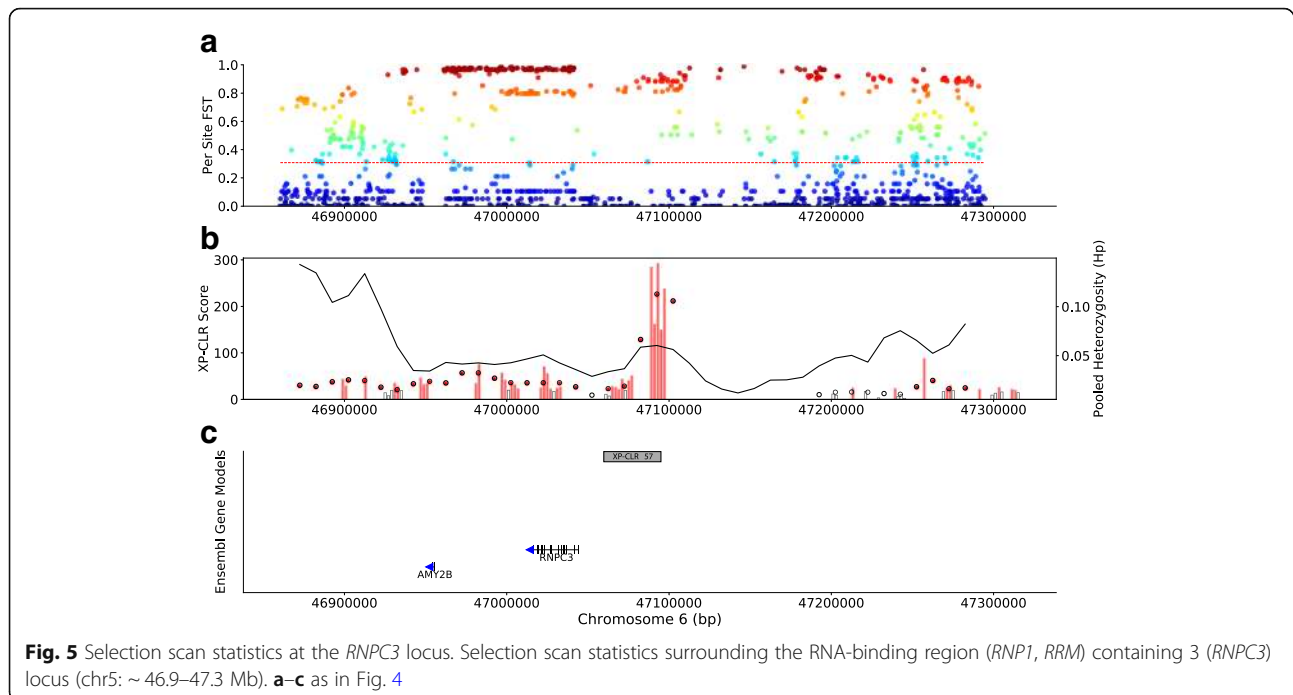
copy number along the reference assembly and searched for regions of extreme copy number differences (see the “Methods” section). Using  $V_{ST}$ , a statistic analogous to  $F_{ST}$  [66], we identified 67 regions of extreme copy number difference between village dogs and wolves which are within 50 kb of 89 unique genes (Additional file 1: Table S7). There was no overlap of these copy number outliers with regions identified through  $F_{ST}$  or XP-CLR. Relative to randomly permuted intervals, the 67  $V_{ST}$  outliers are more likely to be near genes ( $p < 0.01$ ; Additional file 2: Figure S8a) but do not encompass more total genes than expected ( $p > 0.05$ ; Additional file 2: Figure S8b).

The top locus identified through  $V_{ST}$  analysis encompasses the *AMY2B* gene, which at increased copy number confers greater starch metabolism efficiency due to higher pancreatic amylase enzyme levels [5, 37]. Quantitative PCR results have suggested an ancient origin for the *AMY2B* copy number expansion, as 7-ky-old Romanian dogs exhibit elevated *AMY2B* copy number [38]. However, read-depth analysis shows that the *AMY2B* tandem expansion is absent in 5–7-ky-old ancient European dogs [34]. We identified two large duplications, one of 1.9 Mb and the other of 2.0 Mb, that encompass *AMY2B* (Additional file 2: Figure S9). We quantified copy number



**Fig. 4** Selection scan statistics at the *RAI1* Locus. Selection scan statistics surrounding the *retinoic acid-induced 1* (*RAI1*) locus (chr5; ~41.6–41.2 Mb). **a** Per site  $F_{ST}$  scores for all SNPs are indicated along with the  $F_{ST}$  significance threshold determined by the 99th percentile of simulations (red dashed line). **b** Bars represent raw XP-CLR grid scores. Circles indicate the mean XP-CLR score calculated from averaging grid scores within 25 kb windows and are positioned within the center point window. Red bars and circles indicate that the score is significant (above the 99th percentile significance threshold determined through simulations). The black line indicates the average pooled heterozygosity ( $H_p$ ) values for the same window boundaries. **c** The significant XP-CLR locus (gray box) is presented relative to Ensembl gene models (black). Direction of each gene is indicated with blue arrows





at *AMY2B* itself and regions which discriminate the two segmental duplications in 90 dogs using digital droplet PCR (ddPCR). Copy number estimated through read depth strongly correlated with estimates from ddPCR (Additional file 2: Figure S10) confirming the presence of standing copy number variation of *AMY2B* in dogs (range of  $2n_{AMY2B} = 2–18$ ) and distinguishing the two large-scale duplications (Additional file 2: Figure S11). The extreme *AMY2B* copy number expansion appears to be independent of the large-scale duplications, as ddPCR results show that some dogs without the large duplications still have very high *AMY2B* copy number. Read-depth patterns at the duplication breakpoints indicated that NGD, the ancient Irish dog, harbored the 2.0 Mb duplication resulting in increased *AMY2B* copy number.

#### Gene ontology enrichment analysis

We performed enrichment tests using the parent-child model [68] in the topGO R package [69] with the intersecting 429 unique genes as the test set. To control for biasing factors such as gene size, function, and colocalization, we calculated permutation-based  $p$  values ( $p_{perm}$ ) for each GO term by comparing the observed parent-child significance score for each GO term with the distribution obtained by applying the parent-child test to gene sets identified by 1000 randomly permuted genome intervals (see the “Methods” section). We identified 636 enriched GO terms ( $p_{perm} < 0.05$ ) including 327 GO terms represented by more than one gene and more than one XP-CLR locus (Additional file 1: Table S8). The set supported by multiple loci includes several categories related to the process noted above

including the regulation of retinoic acid receptors ( $p_{perm} = 0.028$ ), retinol metabolism ( $p_{perm} = 0.014$ ), the secretion ( $p_{perm} = 0.01$ ), transport ( $p_{perm} = 0.01$ ), and signaling of GABA ( $p_{perm} = 0.03$ ), dopamine receptor signaling ( $p_{perm} = 0.04$ ), and cell maturation ( $p_{perm} = 0.012$ ). Similar enrichment results were also observed using EMBL-EBI ontology annotations (see the “Methods” section; Additional file 1: Table S9). Seventy-one enriched ( $p_{perm} < 0.05$ ) categories were identified using the same methods for the 89 genes intersecting the  $V_{ST}$  (copy number) candidate loci (Additional file 1: Table S10). However, these enrichments were largely driven by a handful of genes with broad biological functions. No enrichments for either XP-CLR or copy number results remain statistically significant if one corrects for the 19,408 tests representing all of the possible GO terms in our gene set, although there are limitations to the application of multiple testing corrections to correlated GO terms.

#### Discussion

Genetic and archaeological data indicate that the dog was first domesticated from Eurasian gray wolves well over 10 kya [23, 27, 34, 40]. Evidence suggests that the domestication process was complex and may have spanned thousands of years [3, 23]. Through multiple analyses, we have identified regions that are strongly differentiated between modern village dogs and wolves and which may represent targets of selection during domestication. Our approach differs from previous studies in several ways including the use of village dogs rather than breed dogs, using neutral simulations to set statistical

cut-offs, and filtering candidate loci based on ancient dog DNA data. Most (83%) of the 246 candidate domestication regions we identified are novel to our study, which we largely attribute to reduced signals associated with post-domestication breed formation. We argue that swept haplotypes identified in modern village dogs and also present in Neolithic dogs more likely represent signals of ancient selection events. Although the 43 village dogs sampled here do not represent the full spectrum of genetic diversity of modern dogs, these samples largely reflect the diversity found in an extensive panel of canids sampled by SNP array and represent populations estimated to have split over 15 kya (European vs Asian) [34]. We expect true targets of selection associated with domestication to be found across all dogs. Signals restricted to breed dogs, although unlikely to reflect selective pressures during domestication, identify genes and pathways important for understanding the genetic basis of modern dog biology and disease. Deeper sampling of village dog diversity may reveal that the CDRs we identified are unique to the studied samples, perhaps as a potential result of geographically restricted selection. As more village dogs are sequenced, it is likely that these candidate domestication regions will be refined and narrowed.

While the use of neutral simulations accounts for genetic diversity in both wild and domestic sampled populations, and better controls false positive rates than arbitrary empirical thresholds [29, 70], several limitations are still apparent in our approach. The demographic model we used does not capture all aspects of dog history, does not include the X chromosome, and does not fit all aspects of the observed data equally well. This likely represents unaccounted for features of the data, such as unmodeled population structure, as well as technical issues such as reduced ascertainment of low frequency alleles due to sequencing depth. Although previous studies have identified detectable jackal admixture ranging from 1 to 2% in the ancestral dog population, we did not include the jackal in our demographic model. Since this gene flow occurred in the ancestral lineage of both modern dogs and wolves (> 20 kya) [32, 34, 40] the jackal ancestry is expected to be similarly represented in all of our samples. This assumption may not hold if the ancestral population had a high degree of population structure, but suitable data to model such complexities is not available.

Although the inclusion of ancient samples allows for the removal of candidate domestication regions that are unique to modern dogs, this approach is limited by the narrow temporal (5–7 kya) and geographic (restricted to Europe) sampling offered by the available data. Even though most selected alleles likely preexisted in the

ancestral wolf population, our approach identifies regions where modern village dogs share the same haplotype. However, even when selection acts on preexisting mutation, a single haplotype often reaches fixation [71], consistent with the variation patterns we identify across village dog populations. As the amount of ancient dogs with genome data increase, it will become possible to apply sophisticated tests that make direct use of ancient genomes to discover sites of selection [72, 73].

Our gene annotations were obtained directly through established BLAST2GO pipelines [74]. Similar results, although with fewer gene-function links, were obtained when using the Ensembl Release 92 of the EMBL-EBI GO gene annotations (Additional file 1: Table S10). After correcting for a total of 19,408 possible tests, none of our enrichments would be significant, even if the raw parent-child  $p$  values were used. However, several factors complicate these gene set enrichment tests. First, the nature of the GO ontology relationships introduces non-independence among related GO terms and genes, a problem partially ameliorated by the parent-child model [68]. Second, the underlying statistical tests assume that every gene is equally likely to be a member of the test set under the null hypothesis, an assumption that may be reasonable for studies of gene expression. Our permutation strategy attempts to control for the non-random correlation between gene size, colocalization, and gene function. However, since no GO term survives a global multiple testing correction, these enrichments must be viewed as tentative.

#### The role of the neural crest in dog domestication

Our XP-CLR candidate domestication regions include 52 genes that were also identified in analyses of other domesticated or self-domesticated animals [9, 11, 17, 75–79], including four genes (*RNPC3*, *CUEDC1*, *GBA2*, *NPR2*) in our top 20 XP-CLR loci. No gene was found in more than three species, consistent with the hypothesis that no single domestication gene exists [19]. Although the overlap of specific genes across species is modest, there are many enriched gene pathways and ontologies shared in domesticates including neurological and nervous system development, behavior, reproduction, metabolism, and pigmentation [10, 11, 17, 73, 75, 80]. We attribute these patterns to the domestication syndrome, a phenomenon where diverse traits, manifested in vastly different anatomical zones, appear seemingly disconnected, yet are maintained across domesticates. Two possible modes of action could generate the domestication syndrome phenotypes while still displaying the genome-wide distribution of sweeps. The first would require independent selection events for distinct traits at numerous loci. Alternatively, selection could have acted on considerably fewer genes that are members of

early-acting developmental pathways with broad phenotypic effects.

For these reasons, the role of the neural crest in animal domestication has gained support from researchers over recent years [18, 20, 21] (Table 1). In 2014, Wilkins et al. [18] established that the vast array of phenotypes displayed in the animal domestication syndrome mirror those exhibited in mild human neurocristopathies, whose pathology stems from aberrant differentiation, division, survival, and altered migration of neural crest cells (NCCs). These cells are multipotent, transient, embryonic stem cells that are initially located at the crest (or dorsal border) of the neural tube. The initiation and regulation of neural crest development is a multi-stage process requiring the actions of many early-expressed genes including the fibroblast growth factor (*Fgf*), bone morphogenic protein (*Bmp*), wingless (*Wnt*), and *Zic* gene families [81]. Several of the genes identified in our XP-CLR analysis are involved in this transition including members of the *Fgf* (*Fgf1*) family as well as a transcription factor (*TCF4*; [82]), inhibitors (*RRM2*; *NPHP3*; [83, 84]), and regulators (*LGR5*; [85]) of the Wnt signaling pathways.

Following induction, NCCs migrate along defined pathways to various sites in the developing embryo. Assignment of identity and the determination of migration routes rely on positional information provided by external signaling cues [86, 87]. *KCTD12*, *CLIC4*, *PAK1*, *NCOR2*, *DOCK2*, and *EXOC7* are all examples of such genes found in our candidate loci that are linked to the determination of symmetry, polarity, and/or axis specification [88–92]. Together, our results suggest that early selection may have acted on genes essential to the initiation of the neural crest and the definition of migration routes for NCCs.

#### ***NCC-derived tissues linked to domestication syndrome phenotypes***

Once in their final destinations, NCC further differentiates as the precursors to many tissues in the developing embryo. Most of the head, for example, arises from NCCs including craniofacial bones, cartilage, and teeth [93, 94]. Ancient dog remains indicate that body size, snout lengths, and cranial proportions of dogs considerably decreased compared to the wolf ancestral state following early domestication [95]. Further, these remains indicate jaw size reduction also occurred, as evidenced by tooth crowding [95]. Such alterations are consistent with the domestication syndrome and implicate aberrant NCC migration since decreases in the number of NCCs in facial primordia are directly correlated with reductions in mid-face and jaw sizes [18, 96]. Genes associated with both craniofacial and tooth development in vertebrates are

found in our candidate loci including *SCUBE1* (XP 115), which is essential in craniofacial development of mice, and *SATB2* (XP 244), which has roles in patterning of the developing branchial arches, palate fusion, and regulation of *HOXA2* in the developing neural crest [97–99]. Lastly, when knocked out in mice, *Bicoid*-related homeodomain factor *PITX1* (XP 124) not only affected hindlimb growth, but also displayed craniofacial abnormalities such as cleft palate and branchial arch defects [100], and influences vertebrate tooth development [101].

Insufficient cartilage, a NCC-derived tissue [94] that consists of chondrocytes and collagen, in the outer ear of humans results in a drooping ear phenotype linked to numerous NC-associated neurocristopathies (e.g., Treacher Collins and Mowat-Wilson) [102]. Analogously, compared to the pricked ears of wolves, dogs predominantly have “floppy” ears [103], a hallmark feature of domesticates [18]. Ablation of *SERPINH1* (XP 181), a collagen-binding protein found in our list of CDRs, is embryonically lethal in ablated in mice [104] and appears to be required for chondrocyte maturation [105]. Alterations of activity by genes such as *SERPINH1* and those regulating NCC migration may have reduced the numbers of NCCs in dog ears, contributing to the floppy phenotype [18].

#### ***Genes associated with neurological signaling, circadian rhythms, and behavior***

Tameness or reduced fear toward humans was likely the earliest trait selected for by humans during domestication [3, 106, 107]. Recapitulating such selection, numerous physiological and morphological characteristics, including domestication syndrome phenotypes (i.e., floppy ears, altered craniofacial proportions, and unseasonal timing for mating), appeared within 20 generations when researchers selected only for tameness in a silver fox breeding population [1, 108]. As the progenitors for the adrenal medulla, which produces hormones associated with the “fight-or-flight” response, hypofunction of NCCs can lead to changes in the tameness of animals [18]. The link between tameness and the NC suggests that changes in neural crest development could have arisen first, either through direct selection by humans for desired behaviors or via the “self-domestication” [109, 110] of wolves that were more docile around humans. Genes contributing to neurological function and behavioral responses were observed in our XP-CLR candidate loci, suggesting these genes may influence chemical and morphological differences associated with tameness. Numerous candidate loci contain genes influencing neurological function and behavioral responses including genes in the dopamine, serotonin, glutamate, and GABA neurotransmission pathways, as well as genes

contributing to the connectivity and development of synapses and dendrites.

In addition to changes in behavior, alterations in sleep patterns would also likely have occurred early in the domestication process due to the shift from the ancestral nocturnal state of wolves, to that of the diurnal lifestyle also exhibited by humans. Evidenced by this, levels of circadian rhythm determinants (e.g., melatonin and serotonin) were significantly altered in domesticated silver foxes selected for tameness compared to wild foxes [111–113]. We hypothesize that early selection on genes influencing behavior have additional functions in the establishment of circadian rhythms, and that both can be explained by impaired NC function. The Smith-Magenis syndrome is caused by disrupted function of *RAII* [114], the gene with the highest XP-CLR score in our study. Humans with Smith-Magenis syndrome display increased aggression and altered circadian rhythms, as well as craniofacial and skeletal deformations, developmental delays, and intellectual disabilities [115]. Similarly, Williams-Beuren syndrome, another neurodevelopmental disorder, affects sleep patterns as well as contributes to hypersociability in humans [116]. A recent study in canines linked behavioral changes in breed dogs to structural variants near *WBSCR17*, a Williams-Beuren syndrome gene [117]. Both syndromes display multiple features associated with improper NCC development, resembling phenotypes of neurocristopathies [115, 118]. For example, disruption of the transcription factors *RAII* and *WSTF* in xenopus (also disrupted in Williams-Beuren syndrome) negatively impacts proper NCC migration, recapitulating the human craniofacial defects associated with the syndromes [119, 120]. *RAII* also regulates circadian rhythms [121–124], a pathway within which other XP-CLR candidate loci genes also exhibit possible (*RNPC3*; [125, 126]) and experimentally verified (*FBLX3*; [127]) roles. Altogether, the top scoring locus, as well as others, indicate overlap of gene functions in influencing behavior and circadian rhythms, and were likely early genetic components of the domestication syndrome.

#### **Misregulation of gene expression may contribute to domestication syndrome phenotypes**

Similar to other domestication scans [6, 9, 19], we did not find SNPs deleteriously altering protein sequence in our predicted sweeps, indicating that gene loss did not have a significant role in dog domestication. Instead, we hypothesize that alterations in gene regulatory pathways or the regulation of transcriptional activity could contribute to broad domestication syndrome phenotypes. Our gene list includes two components of the minor spliceosome; *RNPC3* and *Sf3b1*. *RNPC3*, which affects early development and is linked to dwarfism (isolated growth hormone deficiency; [128]), is also under selection in cats and humans [17, 77]. Absence of *Sf3b1*

disrupts proper NCC specification, survival, and migration [129]. A further example of the role of splicing in NC development is that mutations in *U4atac*, a U12 snRNA subunit gene missing in the current dog annotation, causes Taybi-Lindner syndrome (TALS) in humans. Phenotypes of this syndrome resemble those of the domestication syndrome including craniofacial, brain, and skeletal abnormalities [130]. Thus, proper splicing, particularly for transcripts processed by the minor spliceosome, is required for proper NC function and development.

#### **Copy number variation was likely not a major driver during dog domestication**

Our scan for differentiated copy number states identified few regions that differentiate village dogs and wolves. A previous study found that dogs and wolves have a similar proportion of CNV loci [131]. This suggests that copy number expansion or contraction may not have made as significant contributions to the phenotypic changes associated with domestication. The quantification of wolf copy number using a dog genome reference limits the accuracy of the estimates and prevents detection of wolf-specific insertions. Therefore, reassessment of population-specific copy number changes would be improved by the use of a wolf genome reference [132]. Of note, the top hit from the copy number selection scan corresponded to the *AMY2B*, a gene linked to increased efficiency of starch digestion in dogs [5, 36, 37]. Previous studies have concluded that the increase in *AMY2B* copy number occurred post-domestication, since the timing of domestication (> 10 kya) predates the introduction of starch-rich diets in both humans and dogs [32, 34, 36]. However, this study utilizes previously implemented copy number estimation techniques [34, 36] to identify two independent large-scale duplications (1.9 and 2.0 Mb) that are at least the age of the oldest sampled dog genome (7 ky old). Significant selection signatures from XP-CLR are distal to *AMY2B*, instead centered on *RNPC3* (discussed above) which also lies within the boundaries of both large duplications. Since these large duplications are not fixed in dogs, yet the *RNPC3* selected haplotypes are, we speculate that the initial target of selection may have been on *RNPC3* which could have global effects on expression and phenotype (body size).

#### **Conclusions**

By comparing village dogs and wolves, we identified 246 candidate domestication regions in the dog genome. Analysis of gene function in these regions suggests that perturbation of crucial neural crest signaling pathways could result in the broad phenotypes associated with the



domestication syndrome. Additionally, these findings suggest links between transcriptional regulation and splicing to alterations in cell differentiation, migration, and neural crest development. Altogether, we conclude that while primary selection during domestication likely targeted tameness, genes that contribute to determination of this behavioral change are also involved in critical, far-reaching pathways that conferred drastic phenotypic changes in dogs relative to their wild counterparts.

## Methods

### Sample processing and population structure analysis

The primary selection scans in this paper are based on 43 village dog and 10 gray wolf samples selected from a larger sample set as described below. Additional analysis of candidate genomic regions is based on genotype data from two ancient European samples. For visualization purposes, Fig. 1 also includes genotype data from a larger collection of breed dogs and wild canid out groups. Canid genomes (Additional file 1: Table S1) were processed using the pipeline outlined in [34] to produce a data set of single nucleotide polymorphisms (SNPs) using GATK [133]. From this larger sample set, 37 breed dogs, 45 village dogs, and 12 wolves were selected from the samples described in [34], and ADMIXTURE [39] was utilized to estimate the levels of wolf-dog admixture within this subset. This sample set includes three New Guinea Singing Dogs sequenced as described in [134]. To account for LD, the data was thinned with PLINK v1.07 (–indep-pairwise 50 10 0.1; [135]), where SNPs with an  $R^2$  value over 0.1 were removed in 50 kb windows, sliding 10 sites at a time. The remaining 1,030,234 SNPs were used in five independent ADMIXTURE runs using different seeds, for up to five ancestral populations ( $K = 1-5$ ).  $K = 3$  had the lowest average cross validation error (0.0373) from the five runs and was therefore the best fit for the data (Additional file 2: Figure S12). To eliminate noise in subsequent analyses, we removed all village dogs with greater than 5% wolf ancestry and wolves with greater than 5% dog ancestry. Fifty-four samples remained following this filtration.

Following elimination of admixed samples, we called SNPs in 43 village dogs and 11 gray wolves (Additional file 1: Table S1) using GATK (v. 3.4-46; [133]). Using the GATK VQSR procedure, we identified a high quality variant set such that 99% of positions on the Illumina canine HD array were retained. VQSR filtration was performed separately for the autosomes + chrX pseudoautosomal region (PAR) and the non-PAR region. SNPs within 5 bp of an indel identified by GATK were also removed. We further excluded sites with missing genotype calls in any sample, triallelic sites, and X-nonPAR positions where any male sample was called as heterozygous. The final SNP set contained 7,657,272 sites.

Using these SNPs, we removed samples that exhibited over 30% relatedness following identity by state (IBS) analysis with PLINK v1.90 (–min 0.05; [135]). Only one sample (*mxh*) was removed from the sample set, a sample known to be related to another Mexican wolf in the dataset. Principal component analyses were completed on the remaining 53 samples (43 dogs and 10 wolves) using smartpca, a component of Eigensoft package version 3.0 [136] after randomly thinning the total SNP set to 500,000 sites using PLINK v1.90 [135]. Once PCA confirmed clear genetic distinctions between these dogs and wolves, this final sample set was used for subsequent analyses. For visualization of the final sample set used in selection scans, a further ADMIXTURE plot was generated for this filtered set of 53 samples (Fig. 1b). The SNP set was further filtered for the selection scans to remove rare alleles (minor allele frequencies < 3 out of possible 106 alleles or 0.028). Finally, village dog and wolf allele frequencies were calculated separately using VCFtools [137].

### Demographic model and simulations

Simulations of dog and wolf demographic history were performed using msprime v.0.4.0 [138]. For each autosome, 75 independent simulations were performed using independent random seeds and a pedigree-based genetic map [139]. A mutation rate of  $4 \times 10^{-9}$  per site per generation with a generation time of 3 years was assumed. The 53 samples were modeled as coming from 10 lineages with population histories adapted from [34, 40] (Additional file 1: Table S3; Additional file 2: Figure S2). The simulation is designed to capture key aspects impacting dog and wolf diversity, rather than a definitive depiction of their demography. Resulting simulated SNP sets were filtered for minor allele frequency and randomly thinned to have the same number of SNPs per chromosome as the real SNP datasets used in  $F_{ST}$ , XP-CLR, and  $H_p$  calculations.

### $F_{ST}$ selection scans

Dog and wolf allele counts generated above were used to calculate the fixation index ( $F_{ST}$ ) using the Hudson estimator derived in [140] with the following formula:  $F_{ST} = (p_1 - p_2) - (p_1(1 - p_1)/n_1 - 1) - (p_2(1 - p_2)/n_2 - 1) / (p_1(1 - p_2) + p_2(1 - p_1))$  where  $p_x$  is the allele frequency in population  $x$ , and  $n_x$  is the number of individuals in population  $x$ , with village dogs and wolves treated as separate populations. With this equation, the X chromosome could be included in  $F_{ST}$  calculations. A custom script [141] calculated the per site  $F_{ST}$  across the genome for both the real and 75 simulated SNP sets. Due to differences in effective population size and corresponding expected levels of genetic drift, analyses were performed separately for the chromosome X non-pseudoautosomal region (PAR). Ratio of averages

for the resulting  $F_{ST}$  values were calculated in 200 kb sliding windows with 50 kb step sizes, and we required each window to contain at least 10 SNPs. Additionally, we calculated per site  $F_{ST}$  for each SNP that did not have missing data in any sample.

$F_{ST}$  loci filtration was completed differently for the outlier and non-outlier approach. For the outlier  $F_{ST}$  approach, the windows were  $Z$ -transformed and only windows with  $Z$  scores  $\geq 5$  standard deviations were deemed significant for autosomal and X-PAR loci, and  $\geq 3$  for the X-NonPAR. Significance thresholds for the non-outlier approach were determined as the 99th percentile from  $F_{ST}$  score distributions from the simulated genomes. Overlapping windows passing these thresholds were merged.

#### Pooled heterozygosity ( $H_p$ ) and $\Delta H_p$ calculations

Per window, dog allele frequencies were used to calculate pooled heterozygosity ( $H_p$ ) using the following formula from [6]:  $2\Sigma n_{MAJ}\Sigma n_{MIN}/(\Sigma n_{MAJ} + \Sigma n_{MIN})^2$ , where  $\Sigma n_{MAJ}$  is the sum of major and  $\Sigma n_{MIN}$  minor dog alleles, respectively, for all sites in the window. Significance threshold for window filtration was set as the 0.1th percentile of the  $H_p$  distribution from the simulated genomes. The change in  $H_p$  (or  $\Delta H_p$ ) was calculated as the difference in  $\Delta H_p$  with and without the inclusion of the two ancient dog samples (HXH and NGD). Importantly, genotypes in the ancient samples were determined for the sites variable among the modern samples using an approach that accounts for post-mortem ancient DNA damage [34]. The 5-ky-old German dog (CTC) was not included in this analysis due to known wolf admixture [34]. Windows with  $\Delta H_p$  greater than the 5th percentile observed genome-wide were removed.

#### XP-CLR selection scans

Cross-population comparative likelihood ratio (XP-CLR; [41]) scores were calculated using pooled dog and wolf allele frequencies at sites described above. This analysis requires separate genotype files for each population, and a single SNP file with positions of each SNP and their genetic distance (in Morgans), which were determined through linear extrapolation from the pedigree-based recombination map from [139]. Wolves were set as the reference population, and XP-CLR was run on both the real and simulated SNP sets with a grid size of 2 kb and a window size of 50 kb. Windows that did not return a value (failed) or did not have at least five grids were removed. Average XP-CLR scores from passing grids were calculated in 25 kb windows (step size = 10 kb). Filtration of real windows with averages less than the 99th percentile of averaged simulation scores was performed. Remaining adjacent windows were merged if

they were within 50 kb distance (i.e., one sliding window apart).

#### Visualization of candidate domestication regions

Forty-six additional canines (e.g., dog breeds, jackals, coyotes; Additional file 1: Table S1) were genotyped at candidate loci identified in this study, as well as those from [5, 8, 29] using autosomal SNPs previously called in [34]. SNPs within CDRs of interest were extracted from the SNP dataset using the PLINK make-bed tool with no missing data filter. Per sample, each SNP was classified as 0/0, 0/1, or 1/1 at all loci (1 representing the non-reference allele), and this genotype data was stored in Eigenstrat genotype files, which were generated per window using convertf (Eigensoft package; [136]). A custom script [141] then converted the Eigenstrat genotype files into matrices for visualization using matrix2png [142].

#### Gene enrichment and variant annotation

Coordinates and annotations of dog gene models were obtained from Ensembl ([143, 144], respectively), and a non-redundant annotation set was determined. The sequence of each Ensembl protein was BLASTed against the NCBI non-redundant database (blastp -outfmt 5 -evalue 1e-3 -word\_size 3 -show\_gis -max\_hsps\_per\_subject 20 -num\_threads 5 -max\_target\_seqs 20) and all blastp outputs were processed through BLAST2GO [74] with the following parameters: minimum annotation cut-off of 55, GO weight equal to 5, BLASTp cut-off equal to  $1e^{-6}$ , HSP-hit cut-off of 0, and a hit filter equal to 55. Of the 19,017 autosomal genes in our non-redundant gene set, 16,927 received BLAST2GO annotations representing a total of 19,958 GO terms. To account effects from differential annotations, we also obtained GO annotations from EMBL-EBI (Ensembl Release 92) for the 19,017 gene models above. Predicted effects of SNP variants were obtained by the processing of the total variant VCF file of all canine samples by variant effect predictor (VEP; [42]).

Positions of predicted domestication regions (XP-CLR or  $V_{ST}$ ) were intersected using BEDtools [145] (within a window of 50 kb) with the coordinates of the annotated Ensembl dog gene set to isolate genes within the putatively swept regions, and we defined these as the observed gene set. We performed 1000 randomized shuffles of the loci of interest and, again, identified gene models intersecting within 50 kb, and defined these as the permuted gene sets. Gene enrichment analyses were separately performed on the observed and permuted gene sets using the parent-child model [68] in the topGO R package [69]. Permutation-based  $p$  values ( $p_{perm}$ ) were produced for all GO terms by comparing the observed parent-child test score with the results of the 1000 permutations using the formula  $p_{perm} = (X_{perm} + 1)/(N+1)$ , where  $X_{perm}$  is the number of instances where a permutation obtained a parent-child  $p$

value less than or equal to the observed  $p$  value, and  $N$  is the number of permutations ( $N = 1000$ ). One was added to both the numerator and denominator in this equation to avoid adjusted  $p$  values of 1.0. GO terms with  $p_{\text{perm}}$  values less than 0.05 were further filtered to produce our final enriched GO set. First, terms that were not represented by more than one locus (XP-CLR or  $V_{\text{ST}}$ ) were removed, as these could have arisen due to clustering of genes belonging to a given gene ontology. Finally, terms were removed if they were represented by only one gene. This occurs when one gene may be spanned by more than one XP-CLR or  $V_{\text{ST}}$  locus. Remaining GO terms are considered the enriched set. This approach was performed separately for BLAST2GO and EMBL-EBI go annotation sets.

### Copy number estimation using Quick-mer and fastCN

We implemented two copy number estimation pipelines to assess copy number in village dogs and wolves using the depth of sequencing reads. The first, fastCN, is a modified version of existing pipelines that considers multi-mapping reads to calculate copy number within 3 kb windows (Additional file 3: Note 1; [5, 23, 24, 32, 34, 36–38, 66, 145–171]). By considering multi-mapping reads, copy number profiles will be shared among related gene paralogs, making it difficult to identify specific sequences that are potentially variable. The second pipeline we employed, Quick-mer, a map-free approach based on k-mer counting which can accurately assess copy number in a paralog-sensitive manner (Additional file 3: Note 2; Additional file 4). Both pipelines analyze sequencing read-depth within predefined windows, apply GC-correction and other normalizations, and are able to convert read depth to a copy-number estimate for each window (Additional file 3: Note 3.1). The signal-to-noise ratio (SNR), defined as the mean depth in autosomal control windows divided by the standard deviation, was calculated for each sample (Additional file 3: Note 3.2). The copy number states called by both the Quick-mer and fastCN pipelines were validated through comparison with aCGH data from [170] (Additional file 3: Note 3.3; Additional file 5). Regions with copy number variation between samples in the aCGH or WGS data were selected for correlation analysis.

### $V_{\text{ST}}$ selection scans

Treating village dogs and wolves as separate populations,  $V_{\text{ST}}$  values [66] were calculated for genomic windows with evidence of copy number variation.  $V_{\text{ST}}$  values were Z-transformed and we identified outlier regions as windows exhibiting at least a 1.5 copy number range across all samples, and  $ZV_{\text{ST}}$  scores greater than 5 on the autosomes and the X-PAR, or greater than 3 in the X-nonPAR. Prior to analysis, estimated copy numbers for male samples on the non-PAR region of the X were

doubled. Outlier regions spanning more than one window were then classified as copy number outlier regions (Additional file 1: Table S7). A similar analysis was performed for the unplaced chromosomal contigs in the CanFam3.1 assembly (Additional file 1: Table S11). See Additional file 3: Note 3.4 for additional methods and details.

### Amylase structural variant analysis

We estimated copy number using short-read sequencing data from each canine listed in Additional file 1: Table S1. Copy number estimates for the *AMY2B* gene using fastCN were based on a single window located at chrUn\_AAEX03020568: 4873-8379. See Supplementary Methods: Note 3.5.1 (Additional file 3) for further methods and results. Digital droplet PCR (ddPCR) primers were designed targeting overlapping 1.9 and 2.0 Mb duplications, the *AMY2B* gene and a copy number control region (chr18: 27,529,623-27,535,395) found to have a copy number of two in all sampled canines by Quick-mer and fastCN. Copy number for each target was determined from ddPCR results from a single replication for 30 village dogs, 3 New Guinea singing dogs, and 5 breed dogs (Additional file 1: Table S12), and averaged from two replicates for 48 breed dogs (Additional file 1: Table S13). For more details on primer design, methods, and results for the characterization of the *AMY2B* locus, see Additional file 3: Note 3.5.

### Additional files

**Additional file 1: Table S1.** Description and accession numbers for canine genomes processed in this study. Indications for whether or not a sample was used in the ADMIXTURE analysis, selection scans ( $F_{\text{ST}}$ , XP-CLR, or  $V_{\text{ST}}$ ) are provided. **Table S2.** Coordinates and annotations of outlier  $F_{\text{ST}}$  loci identified through the empirical approach, including intersecting Axelsson, Cagan and Blass, and Freedman loci [5, 8, 29]. **Table S3.** Parameters incorporated into demographic model for neutral evolution in village dog and wolf populations. **Table S4.** Coordinates and annotations of outlier  $F_{\text{ST}}$  loci identified through the simulation informed (non-empirical) approach, including genes as well as intersecting Axelsson, Cagan and Blass, and Freedman loci [5, 8, 29]. **Table S5.** Coordinates and annotations of outlier XP-CLR candidate domestication regions identified through simulation approach, including intersecting genes as well as intersecting Axelsson, Cagan and Blass, and Freedman loci [5, 8, 29]. **Table S6.** Predicted SNP effects (per variant effect predictor [42]) for sites in XP-CLR candidate domestication regions. **Table S7.** Coordinates of  $V_{\text{ST}}$  copy number outliers on autosomes and chromosome X. **Table S8.** Gene enrichment results for XP-CLR candidate domestication regions following  $p$  value adjustment (BLAST2GO). **Table S9.** Gene enrichment for XP-CLR candidate domestication regions following  $p$  value adjustment (EMBL-EBI). **Table S10.** Gene enrichment results for  $V_{\text{ST}}$  copy number outliers following  $p$  value adjustment. **Table S11.** Coordinates of  $V_{\text{ST}}$  copy number outliers on chromosome unknown (chrUn). **Table S12.** ddPCR results from 30 village, 3 New Guinea singing, and 5 breed dog samples of *AMY2B* segmental duplications. **Table S13.** ddPCR results from 48 breed dog samples of *AMY2B* segmental duplications. (XLSX 328 kb)

**Additional file 2: Figure S1.** Z-transformed  $F_{\text{ST}}$  scores for Cagan and Blass locus. **Figure S2.** Demographic model for village dog and wolf populations used in neutral simulations. **Figure S3.** Filtration pipeline implemented for  $F_{\text{ST}}$  and XP-CLR windows. **Figure S4.** Distribution of selection scan statistics for real and simulated  $F_{\text{ST}}$  windows. **Figure S5.**

Filtration pipeline implemented for Axelsson, Cagan and Blass, and Freedman CDRs. **Figure S6.** Distribution of selection scan statistics for real and simulated XP-CLR windows. **Figure S7.** Gene intersect statistics from randomized permutations of XP-CLR gene positions. **Figure S8.** Gene intersect statistics from randomized permutations of  $V_{ST}$  gene positions.

**Figure S9.** Read-depth profiles at the AMY2B locus highlights large-scale structural variant. **Figure S10.** Correlations between the ddPCR and read-depth estimated copy number for AMY2B and associated segmental duplications. **Figure S11.** ddPCR results for the AMY2B gene, 1.9 Mb duplication, and the 2.0 Mb duplication.

**Figure S12.** Admixture plot for K 2-5 for the full assayed canines for sample filtration. This includes breed dogs, village dogs, as well as gray wolves. (DOCX 3397 kb)

**Additional file 3:** Notes 1–3 providing supplementary methods and results of copy number analysis. (DOCX 1850 kb)

**Additional file 4:** Supplemental Quick-mer validation figures. (PDF 1024 kb)

**Additional file 5:** Plots displaying aCGH probe intensity correlations with in silico copy number estimates. (PDF 1916 kb)

## Abbreviations

aCGH: Array comparative genomic hybridization; CDR: Candidate domestication region; chrUn: Chromosome unknown; ddPCR: Droplet digital polymerase chain reaction; GO: Gene ontology;  $H_p$ : Pooled heterozygosity; NC: Neural crest; NCC: Neural crest cell; qPCR: Quantitative polymerase chain reaction; SNP: Single-nucleotide polymorphism; XP-CLR: Cross-population composite likelihood ratio

## Acknowledgements

We thank Shiya Song for advice and assistance in the processing of canid variation data and Laura Botigue for discussion of results utilizing ancient DNA.

## Funding

This work was supported by National Institutes of Health (R01GM103961 to ARB and JMK, and T32HG00040 AP). DNA samples and associated phenotypic data were provided by the Cornell Veterinary Biobank, a resource built with the support of NIH grant R24GM082910 and the Cornell University College of Veterinary Medicine.

## Availability of data and materials

Custom scripts and datasets supporting the conclusions of this article are available in the article and its additional files, as well as a custom UCSC track hub ([https://raw.githubusercontent.com/KiddLab/Pendleton\\_2018\\_Selection\\_Scan/master/Selection\\_track\\_hub.txt](https://raw.githubusercontent.com/KiddLab/Pendleton_2018_Selection_Scan/master/Selection_track_hub.txt)). Software (fastCN and Quick-mer) implemented in this article are available for download in a GitHub repository (<https://github.com/KiddLab/>). Pre-computed 30-mers from the dog, human, mouse, and chimpanzee genomes can be downloaded from <http://kiddlabshare.umms.med.umich.edu/public-data/Quick-mer/Ref/> for Quick-mer processing. Genome sequence data for three New Guinea singing dogs was published under project ID SRP034749 in the Short Read Archive.

## Authors' contributions

JMK, ALP, and FS designed the study. JMK oversaw the study. Selection scans were performed by ALP, AT, and FS. AT and ALP assessed population structure. Copy number variation was estimated by FS and JMK. Functional annotations and enrichment analyses were performed by ALP. FS processed aCGH data. KV processed ancient dog samples. Samples and genome sequences were provided by KV, ARB, and JMK. SE performed the DNA extractions, library generation, and ddPCR analyses. ALP, FS, and JMK wrote the paper with input from the other authors. All authors read and approved the final manuscript.

## Ethics approval

Not applicable.

## Competing interests

ARB is a cofounder and officer of Embark Veterinary, Inc., a canine genetics testing company.

## Publisher's Note

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

## Author details

<sup>1</sup>Department of Human Genetics, University of Michigan, Ann Arbor, MI 48109, USA. <sup>2</sup>Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794, USA. <sup>3</sup>Department of Biomedical Sciences, Cornell University, Ithaca, New York 14853, USA. <sup>4</sup>Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA.

Received: 27 February 2018 Accepted: 23 May 2018

Published online: 28 June 2018

## References

- Trut LN. Early Canid Domestication: The Farm-Fox Experiment: foxes bred for tamability in a 40-year experiment exhibit remarkable transformations that suggest an interplay between behavioral genetics and development. *Am Sci.* 1999;87(2):160–9.
- Germonpré M, Sablin MV, Lázničková-Galetová M, Després V, Stevens RE, Stiller M, Hofreiter M. Palaeolithic dogs and Pleistocene wolves revisited: a reply to Morey (2014). *J Archaeol Sci.* 2015;54:210–6.
- Larson G, Burger J. A population genetics view of animal domestication. *Trends Genet.* 2013;29(4):197–205.
- Harlan JR. Crops and man. Foundation for modern Crop Science. Madison: American Society of Agronomy; 1975.
- Axelsson E, Ratnakumar A, Arendt ML, Maqbool K, Webster MT, Perloski M, Liberg O, Arnemo JM, Hedhammar A, Lindblad-Toh K. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature.* 2013;495(7441):360–4.
- Rubin CJ, Zody MC, Eriksson J, Meadows JR, Sherwood E, Webster MT, Jiang L, Ingman M, Sharpe T, Ka S, et al. Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature.* 2010; 464(7288):587–91.
- Li M, Tian S, Jin L, Zhou G, Li Y, Zhang Y, Wang T, Yeung CK, Chen L, Ma J, et al. Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat Genet.* 2013;45(12):1431–8.
- Cagan A, Blass T. Identification of genomic variants putatively targeted by selection during dog domestication. *BMC Evol Biol.* 2016;16:10.
- Rubin C-J, Megens H-J, Barrio AM, Maqbool K, Sayyab S, Schwochow D, Wang C, Carlborg Ö, Jern P, Jørgensen CB. Strong signatures of selection in the domestic pig genome. *Proc Natl Acad Sci.* 2012;109(48):19529–36.
- Qiu Q, Wang L, Wang K, Yang Y, Ma T, Wang Z, Zhang X, Ni Z, Hou F, Long R, et al. Yak whole-genome resequencing reveals domestication signatures and prehistoric population expansions. *Nat Commun.* 2015;6:10283.
- Fariello MI, Servin B, Tosser-Klopp G, Rupp R, Moreno C, International Sheep Genomics Consortium, San Cristobal M, Boitard S. Selection signatures in worldwide sheep populations. *PLoS One.* 2014;9(8):e103813.
- Fang M, Larson G, Ribeiro HS, Li N, Andersson L. Contrasting mode of evolution at a coat color locus in wild and domestic pigs. *PLoS Genet.* 2009; 5(1):e1000341.
- Wang Z, Yonezawa T, Liu B, Ma T, Shen X, Su J, Guo S, Hasegawa M, Liu J. Domestication relaxed selective constraints on the yak mitochondrial genome. *Mol Biol Evol.* 2010;28(5):1553–6.
- Cheng T, Fu B, Wu Y, Long R, Liu C, Xia Q. Transcriptome sequencing and positive selected genes analysis of *Bombyx mandarina*. *PLoS One.* 2015; 10(3):e0122837.
- Gray MM, Granka JM, Bustamante CD, Sutter NB, Boyko AR, Zhu L, Ostrander EA, Wayne RK. Linkage disequilibrium and demographic history of wild and domestic canids. *Genetics.* 2009;181(4):1493–505.
- Amaral AJ, Megens H-J, Crooijmans RP, Heuven HC, Groenen MA. Linkage disequilibrium decay and haplotype block structure in the pig. *Genetics.* 2008;179(1):569–79.
- Montague MJ, Li G, Gandolfi B, Khan R, Aken BL, Searle SM, Minx P, Hillier LW, Koboldt DC, Davis BW, et al. Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline



- biology and domestication. *Proc Natl Acad Sci U S A*. 2014;111(48):17230–5.
18. Wilkins AS, Wrangham RW, Fitch WT. The “domestication syndrome” in mammals: a unified explanation based on neural crest cell behavior and genetics. *Genetics*. 2014;197(3):795–808.
  19. Carneiro M, Rubin C-J, Di Palma F, Albert FW, Alföldi J, Barrio AM, Pielberg G, Rafati N, Sayyab S, Turner-Maier J. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science*. 2014;345(6200):1074–9.
  20. Wilkins AS. Revisiting two hypotheses on the “domestication syndrome” in light of genomic data. *Vavilov J Genet Breed*. 2017;21(4):435–42.
  21. Sánchez-Villagra MR, Geiger M, Schneider RA. The taming of the neural crest: a developmental perspective on the origins of morphological covariation in domesticated mammals. *R Soc Open Sci*. 2016;3(6):160107.
  22. Fallahsharoudi A, De Kock N, Johnsson M, Ubhayasekera SKA, Bergquist J, Wright D, Jensen P. Domestication effects on stress induced steroid secretion and adrenal gene expression in chickens. *Sci Rep*. 2015;5:15345.
  23. Frantz LA, Mullin VE, Pionnier-Capitan M, Lebrasseur O, Ollivier M, Perri A, Linderholm A, Mattiangeli V, Teasdale MD, Dimopoulos EA, et al. Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science*. 2016;352(6290):1228–31.
  24. Shannon LM, Boyko RH, Castelano M, Corey E, Hayward JJ, McLean C, White ME, Abi Said M, Anita BA, Bondjengo NI, et al. Genetic structure in village dogs reveals a Central Asian domestication origin. *Proc Natl Acad Sci U S A*. 2015;112(44):13639–44.
  25. Wang GD, Zhai W, Yang HC, Wang L, Zhong L, Liu YH, Fan RX, Yin TT, Zhu CL, Poyarkov AD, et al. Out of southern East Asia: the natural history of domestic dogs across the world. *Cell Res*. 2016;26(1):21–33.
  26. Vonholdt BM, Pollinger JP, Lohmueller KE, Han E, Parker HG, Quignon P, Degenhardt JD, Boyko AR, Earl DA, Auton A, et al. Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature*. 2010;464(7290):898–902.
  27. Skoglund P, Ersmark E, Palkopoulou E, Dalen L. Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. *Curr Biol*. 2015;25(11):1515–9.
  28. Wang G-D, Zhai W, Yang H-C, Fan R-X, Cao X, Zhong L, Wang L, Liu F, Wu H, Cheng L-G. The genomics of selection in dogs and the parallel evolution between dogs and humans. *Nat Commun*. 2013;4:1860.
  29. Freedman AH, Schweizer RM, Ortega-Del Vecchyo D, Han E, Davis BW, Gronau I, Silva PM, Galaverni M, Fan Z, Marx P, et al. Demographically-based evaluation of genomic regions under selection in domestic dogs. *PLoS Genet*. 2016;12(3):e1005851.
  30. Boyko AR. The domestic dog: man’s best friend in the genomic era. *Genome Biol*. 2011;12(2):216.
  31. Pilot M, Malewski T, Moura AE, Grzybowski T, Oleński K, Ruśc A, Kamiński S, Fadel FR, Mills DS, Alagaili AN. On the origin of mongrels: evolutionary history of free-breeding dogs in Eurasia. *Proc R Soc B*. 2015;282(1820):20152189.
  32. Freedman AH, Gronau I, Schweizer RM, Ortega-Del Vecchyo D, Han E, Silva PM, Galaverni M, Fan Z, Marx P, Lorente-Galdos B. Genome sequencing highlights the dynamic early history of dogs. *PLoS Genet*. 2014;10(1):e1004016.
  33. Marsden CD, Ortega-Del Vecchyo D, O’Brien DP, Taylor JF, Ramirez O, Vilà C, Marques-Bonet T, Schnabel RD, Wayne RK, Lohmueller KE. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proc Natl Acad Sci*. 2016;113(1):152–7.
  34. Botigue LR, Song S, Scheu A, Gopalan S, Pendleton AL, Oetjens M, Taravella AM, Seregely T, Zeeb-Lanz A, Arbogast RM, et al. Ancient European dog genomes reveal continuity since the Early Neolithic. *Nat Commun*. 2017;8:16082.
  35. Kelley JL, Madeoy J, Calhoun JC, Swanson W, Akey JM. Genomic signatures of positive selection in humans and the limits of outlier approaches. *Genome Res*. 2006;16(8):980–9.
  36. Arendt M, Cairns KM, Ballard JW, Savolainen P, Axelsson E. Diet adaptation in dog reflects spread of prehistoric agriculture. *Heredity (Edinb)*. 2016;117(5):301–6.
  37. Arendt M, Fall T, Lindblad-Toh K, Axelsson E. Amylase activity is associated with AMY2B copy numbers in dog: implications for dog domestication, diet and diabetes. *Anim Genet*. 2014;45(5):716–22.
  38. Ollivier M, Tresset A, Bastian F, Lagoutte L, Axelsson E, Arendt ML, Balasescu A, Marshour M, Sablin MV, Salanova L, et al. Amy2B copy number variation reveals starch diet adaptations in ancient European dogs. *R Soc Open Sci*. 2016;3(11):160449.
  39. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19(9):1655–64.
  40. Fan Z, Silva P, Gronau I, Wang S, Armero AS, Schweizer RM, Ramirez O, Pollinger J, Galaverni M, Del-Vecchyo DO. Worldwide patterns of genomic variation and admixture in gray wolves. *Genome Res*. 2016;26(2):163–73.
  41. Chen H, Patterson N, Reich D. Population differentiation as a test for selective sweeps. *Genome Res*. 2010;20(3):393–402.
  42. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. The ensembl variant effect predictor. *Genome Biol*. 2016;17(1):122.
  43. Maden M. Retinoic acid in the development, regeneration and maintenance of the nervous system. *Nat Rev Neurosci*. 2007;8(10):755.
  44. Shirai H, Oishi K, Ishida N. Bidirectional CLOCK/BMAL1-dependent circadian gene regulation by retinoic acid in vitro. *Biochem Biophys Res Commun*. 2006;351(2):387–91.
  45. Slager RE, Newton TL, Vlangos CN, Finucane B, Elsea SH. Mutations in RAI1 associated with Smith–Magenis syndrome. *Nat Genet*. 2003;33(4):466.
  46. Potocki L, Bi W, Treadwell-Deering D, Carvalho CM, Eifert A, Friedman EM, Glaze D, Krull K, Lee JA, Lewis RA. Characterization of Potocki-Lupski syndrome (dup (17)(p11. 2p11. 2)) and delineation of a dosage-sensitive critical interval that can convey an autism phenotype. *Am J Hum Genet*. 2007;80(4):633–49.
  47. Olivares AM, Han Y, Soto D, Flattery K, Marini J, Mollema N, Haider A, Escher RG, Bruchoovsky N, Cheng H, Matusik RJ. Inhibition of nuclear hormone receptor activity by calreticulin. *Nature*. 1994;367(6462):480.
  48. Dedhar S, Rennie PS, Shago M, Hagesteijn C-YL, Yang H, Filmus J, Hawley RG, Bruchoovsky N, Cheng H, Matusik RJ. Inhibition of nuclear hormone receptor activity by calreticulin. *Nature*. 1994;367(6462):480.
  49. Takahashi H, Kanno T, Nakayama S, Hirahara K, Sciumè G, Muljo SA, Kuchen S, Casellas R, Wei L, Kanno Y. TGF- $\beta$  and retinoic acid induce the microRNA miR-10a, which targets Bcl-6 and constrains the plasticity of helper T cells. *Nat Immunol*. 2012;13(6):587.
  50. Chambers D, Wilson L, Maden M, Lumsden A. RALDH-independent generation of retinoic acid during vertebrate embryogenesis by CYP1B1. *Development*. 2007;134(7):1369–83.
  51. Yang J, Seo J, Nair R, Han S, Jang S, Kim K, Han K, Paik SK, Choi J, Lee S. DGKi regulates presynaptic release during mGluR-dependent LTD. *EMBO J*. 2011;30(1):165–80.
  52. Puranam RS, Eubanks JH, Heinemann SF, McNamara JO. Chromosomal localization of gene for human glutamate receptor subunit-7. *Somat Cell Mol Genet*. 1993;19(6):581–8.
  53. Augustin I, Rosenmund C, Südhof TC, Brose N. Munc13-1 is essential for fusion competence of glutamatergic synaptic vesicles. *Nature*. 1999;400(6743):457.
  54. Caddick SJ, Wang C, Fletcher CF, Jenkins NA, Copeland NG, Hosford DA. Excitatory but not inhibitory synaptic transmission is reduced in lethargic (Cacnb4 lh) and tottering (Cacna1a tg) mouse thalami. *J Neurophysiol*. 1999;81(5):2066–74.
  55. Harris JA, Westbrook RF. Evidence that GABA transmission mediates context-specific extinction of learned fear. *Psychopharmacology*. 1998;140(1):105–15.
  56. Stork O, Ji F-Y, Obata K. Reduction of extracellular GABA in the mouse amygdala during and following confrontation with a conditioned fear stimulus. *Neurosci Lett*. 2002;327(2):138–42.
  57. Gassmann M, Bettler B. Regulation of neuronal GABA B receptor functions by subunit composition. *Nat Rev Neurosci*. 2012;13(6):380.
  58. Oury F, Khrimian L, Denny CA, Gardin A, Chamouni A, Goeden N, Huang Y-Y, Lee H, Srinivas P, Gao X-B. Maternal and offspring pools of osteocalcin influence brain development and functions. *Cell*. 2013;155(1):228–41.
  59. Cheng L, Arata A, Mizuguchi R, Qian Y, Karunarathne A, Gray PA, Arata S, Shirasawa S, Bouchard M, Luo P. Tlx3 and Tlx1 are post-mitotic selector genes determining glutamatergic over GABAergic cell fates. *Nat Neurosci*. 2004;7(5):510.
  60. van Koningsbruggen S, Straasheijm KR, Sterrenburg E, de Graaf N, Dauwense HG, Frants RR, van der Maarel SM. FRG1P-mediated aggregation of proteins involved in pre-mRNA processing. *Chromosoma*. 2007;116(1):53–64.

61. Mathew R, Hartmuth K, Möhlmann S, Urlaub H, Ficner R, Lührmann R. Phosphorylation of human PRP28 by SRPK2 is required for integration of the U4/U6-U5 tri-snRNP into the spliceosome. *Nat Struct Mol Biol.* 2008; 15(5):435.
62. Xia H, Chen D, Wu Q, Wu G, Zhou Y, Zhang Y, Zhang L. CELF1 preferentially binds to exon-intron boundary and regulates alternative splicing in HeLa cells. *Biochim Biophys Acta.* 2017;1860(9):911–21.
63. Kim Y-D, Lee J-Y, Oh K-M, Araki M, Araki K, Yamamura K-i, Jun C-D. NSrp70 is a novel nuclear speckle-related protein that modulates alternative pre-mRNA splicing in vivo. *Nucleic Acids Res.* 2011;39(10):4300–14.
64. Kim C-H, Kim Y-D, Choi E-K, Kim H-R, Na B-R, Im S-H, Jun C-D. Nuclear speckle-related protein 70 binds to serine/arginine-rich splicing factors 1 and 2 via an arginine/serine-like region and counteracts their alternative splicing activity. *J Biol Chem.* 2016;291(12):6169–81.
65. Straub T, Grue P, Uhse A, Lisby M, Knudsen BR, Tange TØ, Westergaard O, Boege F. The RNA-splicing factor PSF/p54 controls DNA-topoisomerase I activity by a direct interaction. *J Biol Chem.* 1998;273(41):26261–4.
66. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W. Global variation in copy number in the human genome. *nature.* 2006;444(7118):444.
67. Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, Li W, Yu Y, Shu L, Zhao Y, Ma Y. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat Biotechnol.* 2015;33(4):408.
68. Grossmann S, Bauer S, Robinson PN, Vingron M. Improved detection of overrepresentation of Gene-Ontology annotations with parent-child analysis. *Bioinformatics.* 2007;23(22):3024–31.
69. Alexa A, Rahnenfuhrer J. topGO: enrichment analysis for gene ontology. *R Packag Ver.* 2010;2(0):1–26.
70. Schlamp F, Made J, Stambler R, Chesebrough L, Boyko AR, Messer PW. Evaluating the performance of selection scans to detect selective sweeps in domestic dogs. *Mol Ecol.* 2016;25(1):342–56.
71. Jensen JD. On the unfounded enthusiasm for soft selective sweeps. *Nat Commun.* 2014;5:5281.
72. Librado P, Gamba C, Gaultier C, Der Sarkissian C, Pruvost M, Albrechtsen A, Fages A, Khan N, Schubert M, Jagannathan V, et al. Ancient genomic changes associated with domestication of the horse. *Science.* 2017; 356(6336):442–5.
73. Schubert M, Jonsson H, Chang D, Der Sarkissian C, Ermini L, Ginolhac A, Albrechtsen A, Dupanloup I, Foucal A, Petersen B, et al. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc Natl Acad Sci U S A.* 2014;111(52):E5661–9.
74. Gotz S, Garcia-Gomez J, Terol J, Williams T, Nagaraj S, Nueda M, Robles M, Talon M, Dopazo J, Conesa A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 2008;36(10):3420–35.
75. Qanbari S, Pausch H, Jansen S, Somel M, Strom TM, Fries R, Nielsen R, Simianer H. Classic selective sweeps revealed by massive sequencing in cattle. *PLoS Genet.* 2014;10(2):e1004148.
76. Kijas JW. Haplotype-based analysis of selective sweeps in sheep. *Genome.* 2014;57(8):433–7.
77. Peyrégne S, Boyle MJ, Dannemann M, Prüfer K. Detecting ancient positive selection in humans using extended lineage sorting. *Genome Res.* 2017; 27(9):1563–72.
78. Prüfer K, Racimo F, Patterson N, Jay F, Sankararam S, Sawyer S, Heinze A, Renaud G, Sudmant PH, De Filippo C. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature.* 2014;505(7481):43.
79. Racimo F. Testing for ancient selection using cross-population allele frequency differentiation. *Genetics.* 2016;202(2):733–50.
80. Lin R, Du X, Peng S, Yang L, Ma Y, Gong Y, Li S. Discovering all transcriptome single-nucleotide polymorphisms and scanning for selection signatures in ducks (*Anas platyrhynchos*). *Evol Bioinformatics Online.* 2015; 11(Suppl 1):67–76.
81. Bronner ME, LeDouarin NM. Development and evolution of the neural crest: an overview. *Dev Biol.* 2012;366(1):2–9.
82. van Es JH, Haegebarth A, Kujala P, Itzkovitz S, Koo B-K, Boj SF, Korving J, van den Born M, van Oudenaarden A, Robine S. A critical role for the Wnt effector Tcf4 in adult intestinal homeostatic self-renewal. *Mol Cell Biol.* 2012;32(10):1918–27.
83. Tang L-Y, Deng N, Wang L-S, Dai J, Wang Z-L, Jiang X-S, Li S-J, Li L, Sheng Q-H, Wu D-Q. Quantitative phosphoproteome profiling of Wnt3a-mediated signaling network indicating the involvement of ribonucleoside-diphosphate reductase M2 subunit phosphorylation at residue serine 20 in canonical Wnt signal transduction. *Mol Cell Proteomics.* 2007;6(11):1952–67.
84. Bergmann C, Fliegau M, Bröchle NO, Frank V, Olbrich H, Kirschner J, Schermer B, Schmedding I, Kispert A, Kränzlin B. Loss of nephrocystin-3 function can cause embryonic lethality, Meckel-Gruber-like syndrome, situs inversus, and renal-hepatic-pancreatic dysplasia. *Am J Hum Genet.* 2008; 82(4):959–70.
85. Carmon KS, Gong X, Lin Q, Thomas A, Liu Q. R-spondins function as ligands of the orphan receptors LGR4 and LGR5 to regulate Wnt/ $\beta$ -catenin signaling. *Proc Natl Acad Sci.* 2011;108(28):11452–7.
86. Bronner-Fraser M. Neural crest cell formation and migration in the developing embryo. *FASEB J.* 1994;8(10):699–706.
87. Santagati F, Rijli FM. Cranial neural crest and the building of the vertebrate head. *Nat Rev Neurosci.* 2003;4(10):806.
88. Lagadec R, Laguerre L, Menuet A, Amara A, Rocancourt C, Péricard P, Godard BG, Rodicio MC, Rodriguez-Moldes I, Mayeur H. The ancestral role of nodal signalling in breaking L/R symmetry in the vertebrate forebrain. *Nat Commun.* 2015;6:6686.
89. Berryman MA, Goldenring JR. CLIC4 is enriched at cell-cell junctions and colocalizes with AKAP350 at the centrosome and midbody of cultured mammalian cells. *Cytoskeleton.* 2003;56(3):159–72.
90. de la Torre-Ubieta L, Gaudillière B, Yang Y, Ikeuchi Y, Yamada T, DiBacco S, Stegmüller J, Schüller U, Salih DA, Rowitch D. A FOXO-Pak1 transcriptional pathway controls neuronal polarity. *Genes Dev.* 2010;24(8):799–813.
91. Kunisaki Y, Nishikimi A, Tanaka Y, Takii R, Noda M, Inayoshi A, Watanabe K-i, Sanematsu F, Sasazuki T, Sasaki T. DOCK2 is a Rac activator that regulates motility and polarity during neutrophil chemotaxis. *J Cell Biol.* 2006;174(5):647–52.
92. Dupraz S, Grassi D, Bernis ME, Sosa L, Bisbal M, Gastaldi L, Jausoro I, Cáceres A, Pfenninger KH, Quiroga S. The TC10-Exo70 complex is essential for membrane expansion and axonal specification in developing neurons. *J Neurosci.* 2009;29(42):13292–301.
93. Le Douarin NM, Dupin E. The neural crest in vertebrate evolution. *Curr Opin Genet Dev.* 2012;22(4):381–9.
94. Minoux M, Rijli FM. Molecular mechanisms of cranial neural crest cell migration and patterning in craniofacial development. *Development.* 2010; 137(16):2605–21.
95. Morey DF. Size, shape and development in the evolution of the domestic dog. *J Archaeol Sci.* 1992;19(2):181–204.
96. Etchevers HC, Couly G, Vincent C, Le Douarin NM. Anterior cephalic neural crest is required for forebrain viability. *Development.* 1999;126(16):3533–43.
97. Xavier GM, Sharpe PT, Cobourne MT. Scube1 is expressed during facial development in the mouse. *J Exp Zool B Mol Dev Evol.* 2009;312(5):518–24.
98. FitzPatrick DR, Carr IM, McLaren L, Leek JP, Wightman P, Williamson K, Gautier P, McGill N, Hayward C, Firth H. Identification of SATB2 as the cleft palate gene on 2q32–q33. *Hum Mol Genet.* 2003;12(19):2491–501.
99. Dobrev G, Chahrouh M, Dautzenberg M, Chirivella L, Kanzler B, Fariñas I, Karsenty G, Grosschedl R. SATB2 is a multifunctional determinant of craniofacial patterning and osteoblast differentiation. *Cell.* 2006;125(5):971–86.
100. Szeto DP, Rodriguez-Esteban C, Ryan AK, O'Connell SM, Liu F, Kioussi C, Gleiberman AS, Izpisua-Belmonte JC, Rosenfeld MG. Role of the Bicoid-related homeodomain factor Pitx1 in specifying hindlimb morphogenesis and pituitary development. *Genes Dev.* 1999;13(4):484–94.
101. Amand TRS, Zhang Y, Semina EV, Zhao X, Hu Y, Nguyen L, Murray JC, Chen Y. Antagonistic signals between BMP4 and FGF8 define the expression of Pitx1 and Pitx2 in mouse tooth-forming anlage. *Dev Biol.* 2000;217(2):323–32.
102. Sandell L. Neural crest cells in ear development. In: *Neural Crest Cells*. 1st ed. Cambridge: Academic Press, Elsevier; 2014. p. 167–87.
103. Boyko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, Lohmueller KE, Zhao K, Brisbin A, Parker HG, Cargill M. A simple genetic architecture underlies morphological variation in dogs. *PLoS Biol.* 2010;8(8):e1000451.
104. Nagai N, Hosokawa M, Itoharu S, Adachi E, Matsushita T, Hosokawa N, Nagata K. Embryonic lethality of molecular chaperone hsp47 knockout mice is associated with defects in collagen biosynthesis. *J Cell Biol.* 2000;150(6): 1499–506.
105. Wilson R, Norris EL, Brachvogel B, Angelucci C, Zivkovic S, Gordon L, Bernardo BC, Stermann J, Sekiguchi K, Gorman JJ. Changes in the chondrocyte and extracellular matrix proteome during post-natal mouse cartilage development. *Mol Cell Proteomics.* 2012;11(1):M111. 014159.

106. Agnvall B, Jöngren M, Strandberg E, Jensen P. Heritability and genetic correlations of fear-related behaviour in red junglefowl—possible implications for early domestication. *PLoS One*. 2012;7(4):e35162.
107. Lindberg J, Björnerfeldt S, Saetre P, Svartberg K, Seehus B, Bakken M, Vilà C, Jazin E. Selection for tameness has changed brain gene expression in silver foxes. *Curr Biol*. 2005;15(22):R915–6.
108. Trut LN, Plyusnina I, Oskina I. An experiment on fox domestication and debatable issues of evolution of the dog. *Russ J Genet*. 2004;40(6):644–55.
109. Hare B, Wobber V, Wrangham R. The self-domestication hypothesis: evolution of bonobo psychology is due to selection against aggression. *Anim Behav*. 2012;83(3):573–85.
110. Morey DF, Jeger R. Paleolithic dogs: why sustained domestication then? *J Archaeol Sci Rep*. 2015;3:420–8.
111. Popova N, Voitenko N, Kulikov A, Avgustinovich D. Evidence for the involvement of central serotonin in mechanism of domestication of silver foxes. *Pharmacol Biochem Behav*. 1991;40(4):751–6.
112. Popova N, Kulikov A, Avgustinovich D, Voitenko N, Trut L. Effect of domestication of the silver fox on the main enzymes of serotonin metabolism and serotonin receptors. *Genetika*. 1997;33(3):370–4.
113. Kolesnikova L, Trut L, Beliaev D. Changes in the morphology of the epiphysis of silver foxes during domestication. *Zh Obshch Biol*. 1988; 49(4):487.
114. Truong HT, Solaymani-Kohal S, Baker KR, Girirajan S, Williams SR, Vlangos CN, Smith AC, Bunyan DJ, Roffey PE, Blanchard CL. Diagnosing Smith–Magenis syndrome and duplication 17p11. 2 syndrome by RAI1 gene copy number variation using quantitative real-time PCR. *Genet Test*. 2008;12(1):67–73.
115. Elsea SH, Girirajan S. Smith–Magenis syndrome. *Eur J Hum Genet*. 2008; 16(4):412.
116. Jones W, Bellugi U, Lai Z, Chiles M, Reilly J, Lincoln A, Adolphs R. II. Hypersociability in Williams syndrome. *J Cogn Neurosci*. 2000; 12(Supplement 1):30–46.
117. Shuldiner E, Koch U, Kartzinell RY, Hogan A, Brubaker L, Wanser S, Stahler D, Wynne CD, Ostrander EA, Sinsheimer JS. Structural variants in genes associated with human Williams–Beuren syndrome underlie stereotypical hypersociability in domestic dogs. *Sci Adv*. 2017;3(7):e1700398.
118. Adams MS, Gammill LS, Bronner-Fraser M. Discovery of transcription factors and other candidate regulators of neural crest development. *Dev Dyn*. 2008; 237(4):1021–33.
119. Tahir R, Kennedy A, Elsea SH, Dickinson AJ. Retinoic acid induced-1 (Rai1) regulates craniofacial and brain development in *Xenopus*. *Mech Dev*. 2014; 133:91–104.
120. Barnett C, Yazgan O, Kuo H-C, Malakar S, Thomas T, Fitzgerald A, Harbour W, Henry JJ, Krebs JE. Williams Syndrome Transcription Factor is critical for neural crest cell function in *Xenopus laevis*. *Mech Dev*. 2012;129(9):324–38.
121. Goldman S, Malow B, Newman K, Roof E, Dykens E. Sleep patterns and daytime sleepiness in adolescents and young adults with Williams syndrome. *J Intellect Disabil Res*. 2009;53(2):182–8.
122. Sniecinska-Cooper AM, Iles RK, Butler SA, Jones H, Bayford R, Dimitriou D. Abnormal secretion of melatonin and cortisol in relation to sleep disturbances in children with Williams syndrome. *Sleep Med*. 2015;16(1):94–100.
123. Williams SR, Zies D, Mullegama SV, Grotewiel MS, Elsea SH. Smith–Magenis syndrome results in disruption of CLOCK gene transcription and reveals an integral role for RAI1 in the maintenance of circadian rhythmicity. *Am J Hum Genet*. 2012;90(6):941–9.
124. De Leersnyder H, de Blois M-C, Claustrat B, Romana S, Albrecht U, von Kleist-Retzow J-C, Delobel B, Viot G, Lyonnet S, Vekemans M. Inversion of the circadian rhythm of melatonin in the Smith–Magenis syndrome. *J Pediatr*. 2001;139(1):111–6.
125. Tian C, Liu D, Sun Q-L, Chen C, Xu Y, Wang H, Xiang W, Kretschmar HA, Li W, Chen C. Comparative analysis of gene expression profiles between cortex and thalamus in Chinese fatal familial insomnia patients. *Mol Neurobiol*. 2013;48(1):36–48.
126. Nikonova EV, Gilliland JD, Tanis KQ, Podtelezchnikov AA, Rigby AM, Galante RJ, Finney EM, Stone DJ, Renger JJ, Pack AI. Transcriptional profiling of cholinergic neurons from basal forebrain identifies changes in expression of genes between sleep and wake. *Sleep*. 2017;40(6):zsx059. <https://doi.org/10.1093/sleep/zsx059>.
127. Godinho SI, Maywood ES, Shaw L, Tucci V, Barnard AR, Busino L, Pagano M, Kendall R, Quwaillid MM, Romero MR. The after-hours mutant reveals a role for Fbxl3 in determining mammalian circadian period. *Science*. 2007; 316(5826):897–900.
128. Argente J, Flores R, Gutierrez-Arumi A, Verma B, Martos-Moreno GA, Cusco I, Oghabian A, Chowen JA, Frilander MJ, Perez-Jurado LA. Defective minor spliceosome mRNA processing results in isolated familial growth hormone deficiency. *EMBO Mol Med*. 2014;6(3):299–306.
129. An M, Henion PD. The zebrafish sf3b1b460 mutant reveals differential requirements for the sf3b1 pre-mRNA processing gene during neural crest development. *Int J Dev Biol*. 2012;56(4):223.
130. Ederly P, Marcaillou C, Sahbatou M, Labalme A, Chastang J, Touraine R, Tubacher E, Senni F, Bober MB, Nampoothiri S. Association of TALS developmental disorder with defect in minor splicing component U4atac snRNA. *Science*. 2011;332(6026):240–3.
131. Serres-Armero A, Povolotskaya IS, Quilez J, Ramirez O, Santpere G, Kuderna LF, Hernandez-Rodriguez J, Fernandez-Callejo M, Gomez-Sanchez D, Freedman AH. Similar genomic proportions of copy number variation within gray wolves and modern dog breeds inferred from whole genome sequencing. *BMC Genomics*. 2017;18(1):977.
132. Gopalakrishnan S, Castruita JAS, Sinding M-HS, Kuderna LF, Rääkkönen J, Petersen B, Sicheritz-Ponten T, Larson G, Orlando L, Marques-Bonet T. The wolf reference genome sequence (*Canis lupus lupus*) and its implications for *Canis* spp. population genomics. *BMC Genomics*. 2017;18(1):495.
133. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, Garimella K, Altshuler D, Gabriel S, Daly M. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–303.
134. Auton A, Li YR, Kidd J, Oliveira K, Nadel J, Holloway JK, Hayward JJ, Cohen PE, Grealley JM, Wang J. Genetic recombination is targeted towards gene promoter regions in dogs. *PLoS Genet*. 2013;9(12):e1003984.
135. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–75.
136. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet*. 2006;2(12):e190.
137. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156–8.
138. Kelleher J, Etheridge AM, McVean G. Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLoS Comput Biol*. 2016;12(5): e1004842.
139. Campbell CL, Bhéer C, Morrow BE, Boyko AR, Auton A. A pedigree-based map of recombination in the domestic dog genome. *G3 (Bethesda)*. 2016; 6(11):3517–24.
140. Bhatia G, Patterson N, Sankararaman S, Price AL. Estimating and interpreting FST: the impact of rare variants. *Genome Res*. 2013;23(9):1514–21.
141. Kidd Lab - Selection Scan GitHub Repository. [https://github.com/KiddLab/Pendleton\\_2018\\_Selection\\_Scan/](https://github.com/KiddLab/Pendleton_2018_Selection_Scan/).
142. Pavlidis P, Noble WS. Matrix2png: a utility for visualizing matrix data. *Bioinformatics*. 2003;19(2):295–6.
143. Ensembl *Canis familiaris* database. <ftp://ftp.ensembl.org/pub/release-81/>. Accessed 1 Mar 2016.
144. Ensembl *Canis familiaris* Index - Release 81. [http://www.ensembl.org/Canis\\_familiaris/Info/Index](http://www.ensembl.org/Canis_familiaris/Info/Index). Accessed 1 Mar 2016.
145. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2.
146. Sudmant PH, Mallick S, Nelson BJ, Hormozdiari F, Krumm N, Huddleston J, Coe BP, Baker C, Nordenfelt S, Bamshad M. Global diversity, population stratification, and selection of human copy-number variation. *Science*. 2015; 349(6253):aab3761.
147. Sudmant PH, Huddleston J, Catacchio CR, Malig M, Hillier LW, Baker C, Mohajeri K, Kondova I, Bontrop RE, Persengiev S. Evolution and diversity of copy number variation in the great ape lineage. *Genome Res*. 2013;23(9): 1373–82.
148. Sudmant PH, Kitzman JO, Antonacci F, Alkan C, Malig M, Tsalenko A, Sampas N, Bruhn L, Shendure J, Eichler EE. Diversity of human copy number variation and multicopy genes. *Science*. 2010;330(6004):641–6.
149. Alkan C, Kidd JM, Marques-Bonet T, Aksay G, Antonacci F, Hormozdiari F, Kitzman JO, Baker C, Malig M, Mutlu O. Personalized copy number and segmental duplication maps using next-generation sequencing. *Nat Genet*. 2009;41(10):1061.
150. Hach F, Hormozdiari F, Alkan C, Hormozdiari F, Birol I, Eichler EE, Sahinalp SC. mrsFAST: a cache-oblivious algorithm for short-read mapping. *Nat Methods*. 2010;7(8):576.

151. fastCN. <https://github.com/KiddLab/fastCN>.
152. Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. *Nat Rev Genet*. 2011;12(5):363.
153. Handsaker RE, Van Doren V, Berman JR, Genovese G, Kashin S, Boettger LM, McCarroll SA. Large multiallelic copy number variations in humans. *Nat Genet*. 2015;47(3):296.
154. Zhang Z, Wang W. RNA-Skim: a rapid method for RNA-Seq quantification at transcript level. *Bioinformatics*. 2014;30(12):i283–92.
155. Patro R, Mount SM, Kingsford C. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nat Biotechnol*. 2014;32(5):462.
156. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. 2016;34(5):525.
157. QuicK-mer. <https://github.com/KiddLab/QuicK-mer>.
158. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*. 2011;27(6):764–70.
159. Consortium GP. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467(7319):1061.
160. Consortium GP. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491(7422):56.
161. Consortium GP. A global reference for human genetic variation. *Nature*. 2015;526(7571):68.
162. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008;456(7218):53.
163. Song S, Sliwerska E, Emery S, Kidd JM. Modeling human population separation history using physically phased genomes. *Genetics*. 2016; <https://doi.org/10.1534/genetics.116.192963>.
164. Hughes JF, Skaletsky H, Pynnikova T, Graves TA, van Daalen SK, Minx PJ, Fulton RS, McGrath SD, Locke DP, Friedman C. Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature*. 2010;463(7280):536.
165. Oetjens MT, Shen F, Emery SB, Zou Z, Kidd JM. Y-chromosome structural diversity in the bonobo and chimpanzee lineages. *Genome Biol Evol*. 2016; 8(7):2231–40.
166. QuicK-mer Precomputed K-mers. <http://kiddlabshare.umms.med.umich.edu/public-data/QuicK-mer/Ref/>.
167. Nicholas TJ, Baker C, Eichler EE, Akey JM. A high-resolution integrated map of copy number polymorphisms within and between breeds of the modern domesticated dog. *BMC Genomics*. 2011;12(1):414.
168. Nicholas TJ, Cheng Z, Ventura M, Mealey K, Eichler EE, Akey JM. The genomic architecture of segmental duplications and associated copy number variants in dogs. *Genome Res*. 2009;19(3):491–9.
169. Chen W-K, Swartz JD, Rush LJ, Alvarez CE. Mapping DNA structural variation in dogs. *Genome Res*. 2009;19(3):500–9.
170. Ramirez O, Olalde I, Berglund J, Lorente-Galdos B, Hernandez-Rodriguez J, Quilez J, Webster MT, Wayne RK, Lalueza-Fox C, Vilà C. Analysis of structural diversity in wolf-like canids reveals post-domestication variants. *BMC Genomics*. 2014;15(1):465.
171. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. Primer3—new capabilities and interfaces. *Nucleic Acids Res*. 2012;40(15):e115.
172. Bocciardi R, Giorda R, Marigo V, Zordan P, Montanaro D, Gimelli S, Seri M, Lerone M, Ravazzolo R, Gimelli G. Molecular characterization of a (2; 6) balanced translocation that is associated with a complex phenotype and leads to truncation of the TCBA1 gene. *Hum Mutat*. 2005;26(5):426–36.
173. Bartels CF, Bükülmez H, Padayatti P, Rhee DK, van Ravenswaaij-Arts C, Pauli RM, Mundlos S, Chitayat D, Shih L-Y, Al-Gazali LI. Mutations in the transmembrane natriuretic peptide receptor NPR-B impair skeletal growth and cause acromesomelic dysplasia, type Maroteaux. *Am J Hum Genet*. 2004;75(1):27–34.
174. Tsuji T, Kunieda T. A loss-of-function mutation in natriuretic peptide receptor 2 (Npr2) gene is responsible for disproportionate dwarfism in *cn/cn* mouse. *J Biol Chem*. 2005;280(14):14288–92.
175. Zhang R, Cao L, Wang Y, Fang Y, Zhao L, Li W, Shi O-Y, Cai C-Q. A unique methylation pattern co-segregates with neural tube defect statuses in Han Chinese pedigrees. *Neuro Sci*. 2017;38(12):2153–64.
176. Lin Y-H, Zhen Y-Y, Chien K-Y, Lee I-C, Lin W-C, Chen M-Y, Pai L-M. LIMCH1 regulates nonmuscle myosin-II activity and suppresses cell migration. *Mol Biol Cell*. 2017;28(8):1054–65.
177. Austin-Tse C, Halbritter J, Zariwala MA, Gilberti RM, Gee HY, Hellman N, Pathak N, Liu Y, Panizzi JR, Patel-King RS. Zebrafish ciliopathy screen plus human mutational analysis identifies C21orf59 and CCDC65 defects as causing primary ciliary dyskinesia. *Am J Hum Genet*. 2013;93(4):672–86.
178. Marques S, Borges AC, Silva AC, Freitas S, Cordenonsi M, Belo JA. The activity of the Nodal antagonist Cerl-2 in the mouse node is required for correct L/R body axis. *Genes Dev*. 2004;18(19):2342–7.
179. Wang S, Meyer H, Ochoa-Espinosa A, Buchwald U, Önel S, Altenhein B, Heinisch JJ, Affolter M, Paululat A. GBF1 (Gartenzweig)-dependent secretion is required for *Drosophila* tubulogenesis. *J Cell Sci*. 2012;125(2):461–72.
180. Mazaki Y, Nishimura Y, Sabe H. GBF1 bears a novel phosphatidylinositol-phosphate binding module, BP3K, to link PI3Ky activity with Arf1 activation involved in GPCR-mediated neutrophil chemotaxis and superoxide production. *Mol Biol Cell*. 2012;23(13):2457–67.
181. van Veen M, Mans LA, Matas-Rico E, van Pelt J, Perrakis A, Moolenaar WH, Haramis A-PG. Glycerophosphodiesterase GDE2/GDPD5 affects pancreas differentiation in zebrafish. *Int J Biochem Cell Biol*. 2018;94:71–8.
182. Rao M, Sockanathan S. Transmembrane protein GDE2 induces motor neuron differentiation in vivo. *Science*. 2005;309(5744):2212–5.
183. Du L, Xu J, Li X, Ma N, Liu Y, Peng J, Osato M, Zhang W, Wen Z. Rumba and Haus3 are essential factors for the maintenance of hematopoietic stem/progenitor cells during zebrafish hematopoiesis. *Development*. 2011;138(4):619–29.
184. Peters H, Neubüser A, Kratochwil K, Balling R. Pax9-deficient mice lack pharyngeal pouch derivatives and teeth and exhibit craniofacial and limb abnormalities. *Genes Dev*. 1998;12(17):2735–47.
185. Ercan-Sencicek AG, Jambi S, Franjic D, Nishimura S, Li M, El-Fishawy P, Morgan TM, Sanders SJ, Bilguvar K, Suri M. Homozygous loss of DIAPH1 is a novel cause of microcephaly in humans. *Eur J Hum Genet*. 2015;23(2):165.
186. Bai SW, Herrera-Abreu MT, Rohn JL, Racine V, Tajadura V, Suryavanshi N, Bechtel S, Wiemann S, Baum B, Ridley AJ. Identification and characterization of a set of conserved and new regulators of cytoskeletal organization, cell morphology and migration. *BMC Biol*. 2011;9(1):54.
187. Alkobtawi M, Ray H, Barriga EH, Moreno M, Kerney R, Monsoro-Burq A-H, Saint-Jeannet J-P, Mayor R. Characterization of Pax3 and Sox10 transgenic *Xenopus laevis* embryos as tools to study neural crest development. *Dev Biol*. 2018. <https://doi.org/10.1016/j.ydbio.2018.02.020>.
188. Zhao C, Deng Y, Liu L, Yu K, Zhang L, Wang H, He X, Wang J, Lu C, Wu LN. Dual regulatory switch through interactions of Tcf712/Tcf4 with stage-specific partners propels oligodendroglial maturation. *Nat Commun*. 2016;7:10883.
189. Dornier E, Coumilleau F, Ottavi J-F, Moretti J, Boucheix C, Mauduit P, Schweisguth F, Rubinstein E. TspanC8 tetraspanins regulate ADAM10/Kuzbanian trafficking and promote Notch activation in flies and mammals. *J Cell Biol*. 2012;199(3):481–96.
190. Theveneau E, Mayor R. Neural crest delamination and migration: from epithelium-to-mesenchyme transition to collective cell migration. *Dev Biol*. 2012;366(1):34–54.
191. Solomon KS, Kudoh T, Dawid IB, Fritz A. Zebrafish foxi1 mediates otic placode formation and jaw development. *Development*. 2003;130(5):929–40.
192. Huang Y, Roelink H, McKnight GS. Protein kinase A deficiency causes axially localized neural tube defects in mice. *J Biol Chem*. 2002;277(22):19889–96.
193. Offermanns S, Zhao LP, Gohla A, Sarosi I, Simon MI, Wilkie TM. Embryonic cardiomyocyte hypoplasia and craniofacial defects in *Gaq*. *Ga11*-mutant mice. *EMBO J*. 1998;17(15):4304–12.
194. Kondo T, Matsuoka AJ, Shimomura A, Koehler KR, Chan RJ, Miller JM, Srour EF, Hashino E. Wnt signaling promotes neuronal differentiation from mesenchymal stem cells through activation of Tlx3. *Stem Cells*. 2011;29(5):836–46.
195. Koestner U, Shnitsar I, Linnemannstöns K, Hufton AL, Borchers A. Semaphorin and neuropilin expression during early morphogenesis of *Xenopus laevis*. *Dev Dyn*. 2008;237(12):3853–63.
196. Pegtel DM, Ellenbroek SI, Mertens AE, van der Kammen RA, de Rooij J, Collard JG. The Par-Tiam1 complex controls persistent migration by stabilizing microtubule-dependent front-rear polarity. *Curr Biol*. 2007;17(19):1623–34.
197. Wang JS, Infante CR, Park S, Menke DB. PITX1 promotes chondrogenesis and myogenesis in mouse hindlimbs through conserved regulatory targets. *Dev Biol*. 2017;434(1):186–95.
198. Wong RLY, Włodarczyk BJ, Min KS, Scott ML, Kartiko S, Yu W, Merriweather MY, Vogel P, Zambrowicz BP, Finnell RH. Mouse Fkbp8 activity is required to inhibit cell death and establish dorso-ventral patterning in the posterior neural tube. *Hum Mol Genet*. 2007;17(4):587–601.



199. Fimia GM, Stoykova A, Romagnoli A, Giunta L, Di Bartolomeo S, Nardacci R, Corazzari M, Fuoco C, Ucar A, Schwartz P. Ambra1 regulates autophagy and development of the nervous system. *Nature*. 2007;447(7148):1121.
200. Tu C-F, Yan Y-T, Wu S-Y, Djoko B, Tsai M-T, Cheng C-J, Yang R-B. Domain and functional analysis of a novel platelet-endothelial cell surface protein, SCUBE1. *J Biol Chem*. 2008;283(18):12478–88.
201. Williams AL, Eason J, Chawla B, Bohnsack BL. Cyp1b1 regulates ocular fissure closure through a retinoic acid-independent pathway. *Invest Ophthalmol Vis Sci*. 2017;58(2):1084–97.

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

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

At BMC, research is always in progress.

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

