

## Genomic approaches to identify hybrids and estimate admixture times in European wildcat populations

Mattucci, Federica; Galaverni, Marco; Lyons, Leslie A; Alves, Paulo C; Randi, Ettore; Velli, Edoardo; Pagani, Luca; Caniglia, Romolo

*Published in:*  
Scientific Reports

*DOI (link to publication from Publisher):*  
[10.1038/s41598-019-48002-w](https://doi.org/10.1038/s41598-019-48002-w)

*Creative Commons License*  
CC BY 4.0

*Publication date:*  
2019

*Document Version*  
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

### *Citation for published version (APA):*

Mattucci, F., Galaverni, M., Lyons, L. A., Alves, P. C., Randi, E., Velli, E., Pagani, L., & Caniglia, R. (2019). Genomic approaches to identify hybrids and estimate admixture times in European wildcat populations. *Scientific Reports*, 9(1), Article 11612. <https://doi.org/10.1038/s41598-019-48002-w>

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

### **Take down policy**

If you believe that this document breaches copyright please contact us at [vbn@aub.aau.dk](mailto:vbn@aub.aau.dk) providing details, and we will remove access to the work immediately and investigate your claim.



OPEN

# Genomic approaches to identify hybrids and estimate admixture times in European wildcat populations

Federica Mattucci<sup>1</sup>, Marco Galaverni<sup>2</sup>, Leslie A. Lyons<sup>3</sup>, Paulo C. Alves<sup>4,5,6</sup>, Ettore Randi<sup>7,8</sup>, Edoardo Velli<sup>1</sup>, Luca Pagani<sup>9,10</sup> & Romolo Caniglia<sup>1</sup>

The survival of indigenous European wildcat (*Felis silvestris silvestris*) populations can be locally threatened by introgressive hybridization with free-ranging domestic cats. Identifying pure wildcats and investigating the ancestry of admixed individuals becomes thus a conservation priority. We analyzed 63k cat Single Nucleotide Polymorphisms (SNPs) with multivariate, Bayesian and gene-search tools to better evaluate admixture levels between domestic and wild cats collected in Europe, timing and ancestry proportions of their hybrids and backcrosses, and track the origin (wild or domestic) of the genomic blocks carried by admixed cats, also looking for possible deviations from neutrality in their inheritance patterns. Small domestic ancestry blocks were detected in the genomes of most admixed cats, which likely originated from hybridization events occurring from 6 to 22 generations in the past. We identified about 1,900 outlier coding genes with excess of wild or domestic ancestry compared to random expectations in the admixed individuals. More than 600 outlier genes were significantly enriched for Gene Ontology (GO) categories mainly related to social behavior, functional and metabolic adaptive processes (wild-like genes), involved in cognition and neural crest development (domestic-like genes), or associated with immune system functions and lipid metabolism (parental-like genes). These kinds of genomic ancestry analyses could be reliably applied to unravel the admixture dynamics in European wildcats, as well as in other hybridizing populations, in order to design more efficient conservation plans.

Anthropogenic hybridization, the cross-breeding of genetically differentiated *taxa* due to human alterations of habitats and populations, is one of the major threats to the conservation of native plants and animals<sup>1–3</sup>. Hybridization between free-ranging domestic animals and their wild conspecifics may spread artificially-selected maladaptive variants causing fitness declines, outbreeding depression and gradual alterations of locally adapted gene complexes, thus increasing the risk of extinction of wild populations or entire species<sup>2,4–8</sup>. However, recent studies documented cases of beneficial introgression of domestic mutations in wild populations of North American wolves (a melanic deletion at the  $\beta$ -defensin-103 locus<sup>9</sup>) and Alpine ibex (a domestic goat MHC haplotype<sup>10</sup>).

Cross-breeding between wild and domestic cats, intensified by the human-mediated worldwide dispersal of domestic cats (*Felis silvestris catus*), together with the demographic decline and fragmentation of European

<sup>1</sup>Area per la Genetica della Conservazione (BIO-CGE), Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA), Ozzano dell'Emilia, Italy. <sup>2</sup>Area Conservazione, WWF Italia, Rome, Italy. <sup>3</sup>Department of Veterinary Medicine and Surgery, College of Veterinary Medicine, University of Missouri, Columbia, USA. <sup>4</sup>Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO), InBio - Laboratório Associado, Campus Agrário de Vairão, Vairão, Portugal. <sup>5</sup>Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Porto, Portugal. <sup>6</sup>Wildlife Biology Program, Department of Ecosystem and Conservation Sciences, University of Montana, Missoula, USA. <sup>7</sup>Department of Biological, Geological and Environmental Sciences, University of Bologna, Bologna, Italy. <sup>8</sup>Department of Chemistry and Bioscience, Faculty of Engineering and Science, University of Aalborg, Aalborg, Denmark. <sup>9</sup>Dipartimento di Biologia, Università degli Studi di Padova, Padua, Italy. <sup>10</sup>Estonian Biocentre, Institute of Genomics, University of Tartu, Tartu, Estonia. Correspondence and requests for materials should be addressed to F.M. (email: [federica.mattucci@isprambiente.it](mailto:federica.mattucci@isprambiente.it))

wildcat populations (*F. s. silvestris*<sup>11</sup>), offers a remarkable case-study of anthropogenic hybridization<sup>12,13</sup>. The widespread diffusion of stray or feral cat populations<sup>14</sup> likely promoted reproductive interaction between the two subspecies. Moreover, the full fertility of their hybrid offspring<sup>15</sup> due to the recent origin of the domesticated cat<sup>16,17</sup> has likely increased the risk of genetic introgression. Despite the active role of ecological barriers found to limit hybridization in some Mediterranean regions<sup>18,19</sup>, variable degrees of admixture have been detected throughout the European continent and across habitat types as a consequence of human pressures<sup>2,20–26</sup>, leading to a complete hybrid swarm in wild-living cats in Scotland<sup>12</sup>. Such geographical heterogeneity in admixture levels might be explained by different environmental conditions and ecological barriers<sup>19</sup>, population histories and proportions<sup>20,27</sup>, or by the choice of markers and sampling design<sup>26</sup>.

However, the fitness consequences of introgressive hybridization in wildcats are still unknown. In other species, specific genes or gene complexes of domestic origin could either show selective advantages or, in contrast, reduce fitness or induce outbreeding depression<sup>4,7,28</sup>. Therefore, the accurate detection of hybrids, the quantification of introgression in hybridizing populations and the identification of their demographic and ecological determinants are needed for developing appropriate wildcat conservation plans and correctly allocate resources for their application<sup>29,30</sup>.

Variations in coat color patterns and morphological traits between wildcats and domestic cats are not always diagnostic<sup>31,32</sup>, thus their hybrids and backcrosses are not easily identifiable through the analysis of morphological features. Hence, hybridization has been more reliably assessed using molecular markers, mainly small panels of hypervariable microsatellites (short tandem repeats, STRs) and short mitochondrial DNA (mtDNA) sequences. The high variability of these markers, analyzed using Bayesian and phylogenetic statistical tools, has radically improved our knowledge of the European wildcat population genetic structure<sup>20,24,27,33–37</sup>, but showed a limited power to investigate the ancestry of admixed individuals. The wild and domestic cats used as parental references for multivariate and Bayesian assignment analyses were often regionally sampled, and the different applied marker panels were seldom comparable<sup>12,13,21–24,31,32,34,38,39</sup>. Consequently, standardized and more powerful panels of molecular markers to be applied at a large scale are required to lower the risk of underestimating the prevalence of introgressive hybridization in natural wildcat populations.

Recent next-generation sequencing platforms can offer solutions allowing the assemblage of extensive and cost-effective panels of ancestry-informative markers (AIMs) constituted by Single Nucleotide Polymorphisms (SNPs), which represent the most widespread source of genome-wide variation<sup>40</sup>.

A promising set of 96 nuclear and mitochondrial SNPs has been recently selected by Nussberger *et al.*<sup>25</sup> because of their fixed allelic differences between domestic and wild reference cats and applied for European wildcat admixture analyses. Furthermore, an increased set of SNPs ( $n = 158$ ) has been identified by Oliveira *et al.*<sup>30</sup> from the 1.9 × genome sequence of an Abyssinian domestic cat<sup>41</sup> because of their informativeness and variability in domestic cat breeds<sup>42</sup>. Both marker panels were applied on a wide cat sampling and proved to be able to successfully assign hybrid categories up to the second admixture generation (with a mean error rate value of 2–12% in category assignment<sup>26,30</sup>). However, these SNPs showed a limited power to detect old-generation backcrosses resulting from a repeated cross-breeding between admixed individuals and parental species.

Recent studies showed how the employment of thousands of markers might help to reveal previously undetectable backcrosses (older than two–three generations in the past) and estimate the timing from the admixture events<sup>43,44</sup>. Additionally, the availability of efficient AIMs widely distributed across the entire genome can allow to identify patterns of introgressed linkage blocks hosting candidate genes that may underlie introgressed functional traits that are still unknown<sup>45,46</sup>, further help to disentangle historical and contemporary admixture<sup>47</sup> by analyzing the distribution of haplotype block lengths<sup>46,48</sup> and associate anomalous phenotypes with their genetic bases<sup>44</sup>. These approaches allow researchers to better understand the dynamics and consequences of anthropogenic hybridization compared to previous studies, helping to face specific management and conservation issues<sup>46</sup>.

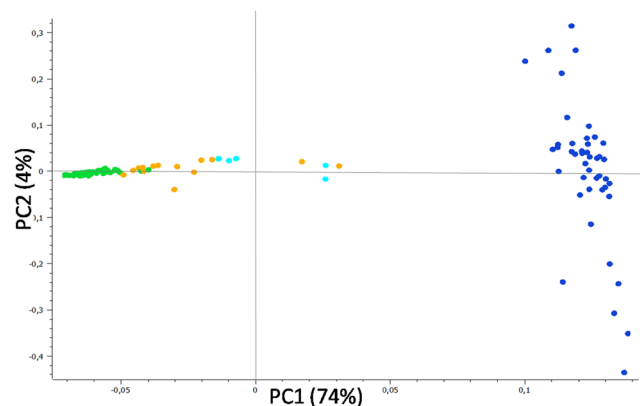
The recently released Illumina Infinium iSelect 63k DNA cat array contains 62,897 variants that are mostly polymorphic within the domestic cats and includes 4,240 wildcat-specific markers<sup>49</sup>. This array offers a suitable molecular tool to further investigate the ancestry of European wildcat populations in conservation and monitoring projects<sup>49</sup>.

Here we genotyped a wide sampling of European wildcats, domestic cats and known or putative admixed cats from a large part of the European wildcat home range distribution with the Illumina Infinium iSelect 63k DNA cat array by applying multivariate, Bayesian and gene-search analysis tools to: (1) improve the identification of admixed genotypes older than the first few generations of backcrossing, (2) estimate their times of admixture<sup>50–52</sup>, (3) quantify and localize domestic and wildcat-derived genomic regions, (4) search for genes significantly deviating from random inheritance patterns, possibly due to selective pressures, (5) define a reduced panel of AIMs to routinely apply for population structure and hybridization monitoring projects.

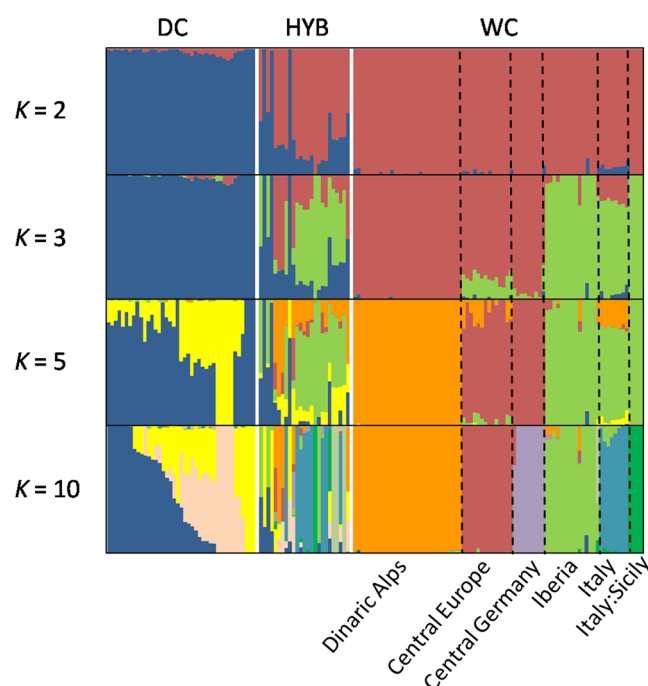
## Results

**Data filtering and marker selection.** Quality-control and filtering procedures yielded a final sample set consisting of 80 presumed European wildcats (WC), 44 domestic cats (DC) and 22 known or presumed WC × DC admixed individuals (Supplementary Fig. S1), previously identified from STR assignment and multivariate analyses<sup>20,21,27,37</sup>, successfully typed for 57,302 autosomal SNPs, hereafter referred to as the 57k SNP panel set (35,228 after linkage disequilibrium pruning, hereafter referred to as the 35k LD-pruned SNP panel set).

**Assignment and admixture analyses of the sampled cats.** More than 78% of the genetic variability of the sampled cats was explained by the first two components of a preliminary Principal Component Analysis (PCA) performed in SVS using the 57k SNP panel set (Fig. 1), which clearly distinguished domestic from wildcats. Putative admixed cats (referred to hybrids and backcrosses), genetically identified through previous STR



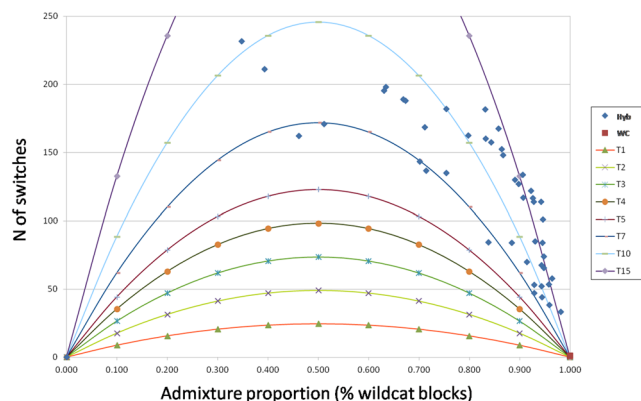
**Figure 1.** PC1 versus PC2 results from an exploratory principal component analysis (PCA) computed in SVS on the 57k SNP panel set and including domestic cats (blue dots), putatively admixed wildcats (orange dots), known captive hybrids previously genetically identified with STR data (light blue dots) and European wildcats (green dots). The two axes are not to scale, in order to better distinguish individuals along PC2.



**Figure 2.** ADMIXTURE results from the 35k LD-pruned SNP panel set at  $K=2$ ,  $K=3$ ,  $K=5$  and  $K=10$ .  $K=2$  clearly separates wild from domestic cats with admixed individuals showing intermediated assignment values. From  $K > 2$  the genetic substructure of European wildcats progressively takes shape. At  $K=3$  the Northern European wildcat populations (including the Dinaric, the Central European and the Central Germany areas) split from the Southern ones (including the Italian and the Iberian Peninsulas). At  $K=5$  the Dinaric wildcat population groups apart from the Central European and the Southern ones and domestic cats form two distinct sub-clusters. At  $K=10$  the five biogeographic macro-populations already identified in Europe through STR analyses (Iberian, Italian, Central European, Dinaric and Central Germany populations<sup>27</sup>) are confirmed, with an additional cluster including the Sicilian samples.

analyses<sup>20,21,27,37</sup>, were scattered along the first axis (74%) between the parental cats, mainly closer to the wildcat group, except for the known hybrids (referred to F1-F2 individuals), which were intermediate (Fig. 1).

Multivariate analyses were clearly confirmed by the assignment values obtained from the ADMIXTURE tests performed with the 35k LD-pruned SNP panel set that showed the main decrease in the cross validation error (CV) at the optimal genetic clusters  $K=2$  (Supplementary Fig. S2) and clearly separated domestic cats from wildcats (Fig. 2). Based on the distribution of individual assignment values, we preliminarily identified the parental reference cats and the admixed individuals that we further investigated in the subsequent ancestry analyses. All domestic cats ( $n=44$ ) showed an individual assignment value  $q_w < 0.090$  and were considered as reference



**Figure 3.** Timing since the admixture event for each admixed individual (Hyb = 45), deduced from the empirical distribution of the number of chromosomal switches inferred from PCADMIX, in relation to the individual assignment values (proportion of wildcat blocks). The colored lines indicate the expected distributions at increasing generations since admixture.

domestic sources, whereas for the putative wildcats, whose individual membership values  $q_w$  ranged from 0.870 to 1.000, we defined a strictly conservative  $q$ -threshold that retained as reference wild sources only individuals with a  $q_w = 1.000$  ( $n = 57$ ). Therefore, we considered as admixed all cats showing an intermediate assignment value  $0.090 > q_w < 1.000$  ( $n = 45$ ). For  $K > 2$ , the genetic substructure of European wildcats<sup>31</sup> progressively took shape, with the initial split of the southern European wildcat populations (Italian and Iberian Peninsulas) from the Central-Northern cats (including the Dinaric and the Central European areas) observed at  $K = 3$  (Fig. 2), followed by the subsequent isolation of the Dinaric population at  $K = 5$  (Fig. 2), and by the final split of five main biogeographic clusters (Iberian, Italian, Central European, Dinaric and Central Germany populations) as previously identified in Europe<sup>27</sup>. An additional cluster represented by Sicilian samples was identified at  $K = 10$  (Fig. 2).

A highly significant admixture rate between domestic cats and European wildcats was corroborated by the THREEPOP results for the F3 tests<sup>53</sup> computed in all the putative admixed individuals detected with ADMIXTURE (z-score =  $-113.04$ ) and confirmed by analyzing the five European wildcat macro-areas separately (z-scores ranging from  $-8.22$  of the Central Germany to  $-35.70$  of the Central Europe, from  $-113.11$  and  $-49.19$  of the Iberian and Italian Peninsulas to  $-91.78$  of the Dinaric macro-area).

The observed widespread signals of genomic admixture were individually estimated with PCADMIX which identified an average 17% of domestic regions (with only slight variations across chromosomes, ranging from 16% in Chr5 to 19% in Chr9, see Supplementary Fig. S3) in the genome of the putative admixed cats. The proportions of domestic blocks within individuals ranged from 1.8% to 65.3%, significantly correlated ( $R^2 \geq 0.95$ ; P-values =  $1.72 \times 10^{-3}$ ; t-test) with those estimated in ADMIXTURE at  $K = 2$  (mean  $q_d = 0.141$ ). None of the 40 randomly-selected reference individuals reanalyzed as hybrids for comparison in PCADMIX showed any switch from domestic to wildcat blocks (or *viceversa*) along their genomes, confirming the reliability of the reference populations selected for the admixture timing analyses and the possibility to exclude any ascertainment bias from the SNP array.

**Time of admixture.** We inferred the time in generations during which the admixture events between domestic and European wildcat populations took place by analyzing patterns of linkage disequilibrium decay<sup>54</sup> in ALDER. Significant admixture was detected in the five European wildcat biogeographic areas identified with ADMIXTURE (P-values  $< 3.5 \times 10^{-8}$ ), although with inconsistent decay rates in all the cohorts considered, except for the Dinaric macro-area. The admixture midpoint in the European wildcats was generally estimated in ALDER to have occurred about  $5.02 \pm 0.37$  generations before sampling, which correspond to about ten years considering a cat generation time of two years. The most ancient hybridization events were detected in the Italian Peninsula ( $6.62 \pm 0.58$  generations) and in the Central European area ( $8.60 \pm 1.56$  generations), respectively corresponding to about 14 and 18 years before sampling, whereas a more recent admixture time of about six years was estimated in the Dinaric region ( $3.15 \pm 0.24$  generations).

The local ancestry inferred with PCADMIX in single individuals identified a number of switches from the reference European wildcat ancestry blocks to domestic cat blocks ranging from 31 to 231 (mean value  $122 \pm 9$ ), revealing that all the admixture events within the European wildcats occurred at least six generations before sampling (Fig. 3 and Supplementary Fig. S4), with the oldest timing estimated up to 22 generations before sampling. Coherently with ALDER, the most ancient hybridization events were traced in the Italian Peninsula (mean generation value  $13 \pm 9$ ), while the most recent events in the Dinaric Alpine populations (mean generation value  $8 \pm 9$ ). This pattern dated the first case of hybridization in the Italian Peninsula to 1962 (corresponding to 44 years before sampling), whereas the last admixture event likely dated to 1994 (*ca.* 14 years before sampling) in the Dinaric region (Supplementary Fig. S4). Although significantly (P-values =  $2.69 \times 10^{-3}$ ; t-test) correlated, the average admixture timing estimated with PCADMIX resulted approximately twofold more ancient compared to the midpoints estimated by ALDER.

Ancestry	Association	Gene name
FCA-like	Fear of human	<i>GPRC5A</i> <sup>85</sup>
	Stress responses	<i>CAPZA1</i> <sup>86</sup>
	Neural crest development	<i>CTNNB1</i> <sup>87,88</sup>
	Tumor suppressor	<i>PHLDB2</i> <sup>89</sup> ; <i>TRIM29</i> <sup>90</sup> ; <i>PTPRT</i> <sup>91</sup>
	Immune system (IAV severity)	<i>ZC3HAV1</i> <sup>92</sup>
	Immune system (H5N1 virus)	<i>SLC16A1</i> <sup>93</sup>
	Lipid metabolism	<i>PLPP3</i> <sup>94</sup>
FSI-like	Memory performance	<i>ADCY8</i> <sup>95</sup> ; <i>CUX2</i> <sup>96</sup>
	Fear of human	<i>DYNLRB2</i> <sup>85</sup> ; <i>MRPS18A</i> <sup>97</sup>
	Tameness and aggressiveness	<i>SLC6A4</i> <sup>98</sup> ; <i>SPTLC3</i> <sup>99</sup>
	Sociability	<i>MAD2L1BP</i> <sup>100</sup> ; <i>ATP8A1</i> <sup>101</sup> ; <i>ADCY7</i> <sup>102</sup>
	Brain development	<i>DCTN2</i> <sup>103</sup>
	Brain development, white coats pattern	<i>KIT</i> <sup>104,105</sup>
	Development processes	<i>EDN1</i> <sup>106</sup> ; <i>KIF2B</i> <sup>107</sup> ; <i>BHMT</i> <sup>108</sup>
	Development processes (milk yielding)	<i>ACP6</i> <sup>109</sup> ; <i>ARHGEF4</i> <sup>110</sup> ; <i>MTX3</i> <sup>110</sup>
	Fertility (spermatogenesis regulation)	<i>MORN3</i> <sup>111</sup> ; <i>NKDI</i> <sup>112</sup>
	Fertility (maintenance pregnancy)	<i>MCL1</i> <sup>113</sup> ; <i>LNPEP</i> <sup>114</sup>
	Fertility (likelihood of survival)	<i>RNPEP</i> <sup>115</sup>
	Morphology	<i>NPR2</i> <sup>116</sup> ; <i>PIK3R4</i> <sup>117</sup> ; <i>STX17</i> <sup>118</sup>
	Muscle development	<i>MYLPF</i> <sup>119</sup> ; <i>COX6A2</i> <sup>120,121</sup>
	Tumor suppressor	<i>BIN1</i> <sup>122</sup> ; <i>CUL9</i> <sup>123</sup> ; <i>GSDMC</i> <sup>124</sup> ; <i>NOL7</i> <sup>125</sup>
	Immune system (FeLV virus)	<i>TRIM25</i> <sup>126</sup>
	Immune system (ASFV virus)	<i>NFKBIE</i> <sup>127</sup>
	DNA repair	<i>CMPK1</i> <sup>128</sup> ; <i>DDIT3</i> <sup>129</sup> ; <i>CCDC92</i> <sup>130</sup>
	Lipid metabolism	<i>ALDH2</i> <sup>131</sup>
	Disease	<i>BCL9</i> <sup>132</sup> ; <i>BFSP2</i> <sup>133</sup>
	Disease (Feline Spinal Muscular Atrophy)	<i>LIX1</i> <sup>134</sup>
	Disease (Hypertrophic Cardiomyopathy Mutations)	<i>MYL2</i> <sup>135</sup>
	Disease (hearing disorders)	<i>MYO1A</i> <sup>136</sup>

**Table 1.** Subset of significantly enriched GO domestic-like (a) and wildcat-like (b) outlier genes detected in the domestic  $\times$  wildcat admixed cats, identified with Bayesian analysis in ADMIXTURE and PCADMIX, which have been previously described in literature.

**Regions of genomic differentiation between domestic and European wildcats.** The admixed cats revealed a complex genomic mosaic of wild and domestic ancestry, as reconstructed by PCADMIX. Therefore, 138 regions with high frequency of wildcat alleles (wildcat-like regions) were identified, including 1,045 annotated genes, 577 of which were significantly enriched for specific Gene Ontology (GO) categories and are known to be involved in several biological and cognitive processes related to communication and elusive behaviors (Table 1 and Supplementary Table S1c,d). In particular, we observed genes belonging to significantly enriched Cellular Component (CC) categories playing important roles in memory performance and sociability, or being related to development processes, key morphological features and fertility (Table 1 and Supplementary Text S1).

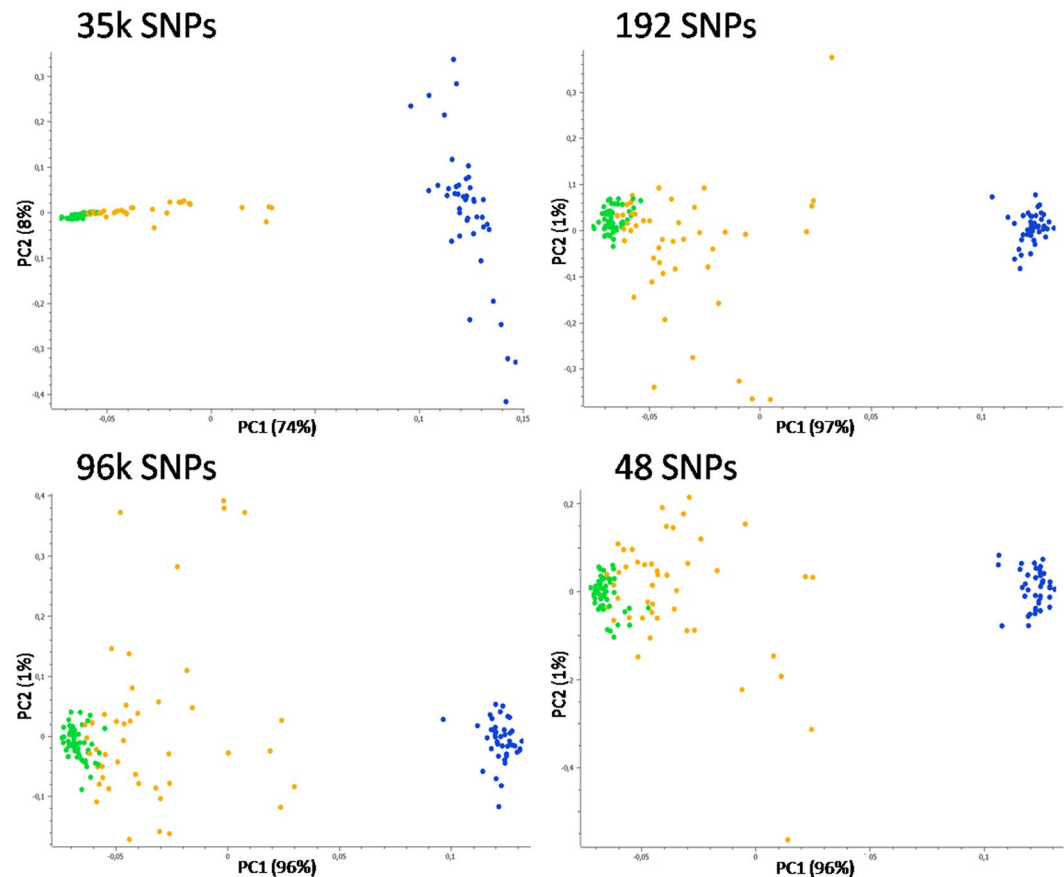
Moreover, 138 segments with high frequency of domestic alleles were identified, containing 902 annotated genes, 39 of which were significantly enriched for Human Phenotype (HP) and Molecular Function (MF) categories correlated to cell adhesion molecular binding (crucial for maintaining tissue structure and function; Supplementary Table S1a,b). Interestingly, we found domestic-like genes significantly enriched in GO categories mainly associated with neural crest development cognition and behavior, or related to biological immune system responses and physiological adaptations (Table 1 and Supplementary Text S1).

Both wild- and domestic-like regions hosted a number of significantly enriched GO genes implicated in muscle development, lipid and energy metabolism or known to be involved in immune functions, tumor suppressor and DNA repair functions (Table 1 and Supplementary Text S1). Another set of wild- and domestic-like enriched GO genes were described as associated with diseases or infections, some of which were feline-specific (Table 1 and Supplementary Text S1).

Conversely, none of the  $F_{ST}$  outlier SNPs showed a significantly positive P-value in BAYESCAN, suggesting no evidence of selection signatures neither comparing admixed individuals *versus* wildcats nor comparing admixed individuals *versus* domestic cats.

**Selection of informative SNPs.** A reduced panel of SNPs was selected based on estimates of WC-DC divergence ( $F_{ST}$  and  $I_N$ ) that were highly correlated one another (Spearman's  $r$   $F_{ST} - I_N = 0.99$ ;  $P < 0.0005$ ), and even to the  $H_E$  values (differences not significant at  $X^2 = 1$ ;  $P > 0.25$  Chi-test; Supplementary Fig. S5). Thus, for instance, SNPs showing the lowest  $H_E$  values (0.010) had also the lowest average  $F_{ST}$  (0.002) and average  $I_N$  (0.003).





**Figure 4.** Principal component analysis (PCA) computed in SVS on the 35k LD-pruned SNP panel set and on the 192, 96, 48 SNPs showing the highest wild-domestic cats  $F_{ST}$  and  $I_N$  values. For each dataset, PC1 versus PC2 are indicated (axes are not to scale). Domestic and European wildcats are represented in green and blue dots, respectively, known hybrids and putatively admixed wildcats ( $ADMIXTURE$   $q_w < 0.999$ ) in orange. The power of the top 48 SNPs is comparable to that reached with 35k SNPs, indicating that they can be used as reliable ancestry-informative-markers (AIMs), although no clear subdivision could be traced between some of the admixed and the non-admixed wildcats.

values. Based on these results, we selected the top 192, 96 and 48 SNPs showing the highest  $F_{ST}$  values and evaluated their performance using a PCA visual summary of their observed genetic variation.

Results were highly concordant and all the reduced panels, which included from 6 to 23 wildcat-specific variants previously described in Gandolfi *et al.*<sup>49</sup>, well differentiated domestic cats and wildcats (192 top SNPs:  $F_{ST} = 0.90$ ,  $H_{O-WC} = 0.060$ ,  $H_{O-DC} = 0.092$ ; 96 top SNPs:  $F_{ST} = 0.93$ ,  $H_{O-WC} = 0.074$ ,  $H_{O-DC} = 0.039$ ; 48 top SNPs:  $F_{ST} = 0.95$ ,  $H_{O-WC} = 0.029$ ,  $H_{O-DC} = 0.057$ ), grouping most admixed individuals intermediately, with the exceptions of about ten genotypes that plotted more closely to the wildcat group (Fig. 4). The assignment values from the  $ADMIXTURE$  run at  $K = 2$  on the 192, 96, and 48 SNPs were not significantly different from values obtained with the 35k LD-pruned SNP panel set ( $P$ -values  $> 0.5$  in all cases;  $t$ -test), although a portion of putative admixed cats ( $q_w < 1.000$  in  $PCADMIX$ ), ranging from 9% to 22% (with the 192 and 48 SNP panel sets, respectively), were misclassified and confused as parental wildcats (Table 2).

The genetic variability of the five main bio-geographic wildcat groups, summarized using the top 96 ( $F_{ST} = 0.93$ ,  $H_O = 0.103$ ) and 48 ( $F_{ST} = 0.93$ ,  $H_O = 0.077$ ) informative SNPs, selected based on WC divergence ( $F_{ST}$ ) and graphically plotted in a PCA (Supplementary Fig. S6), was concordant with the  $ADMIXTURE$  results previously described (Fig. 2).

The combined panel of SNPs, selected based on both WC-DC and WC divergence ( $F_{ST}$ ), mostly confirmed  $ADMIXTURE$  results (Table 2). However, 16% (using 96 + 96 top SNPs,  $PID_{WC} = 5.6 \times 10^{-33}$ ;  $PID_{SibWC} = 3.7 \times 10^{-17}$ ) and 22% (using 48 + 48 top SNPs,  $PID_{WC} = 1.1 \times 10^{-14}$ ;  $PID_{SibWC} = 8.2 \times 10^{-8}$ ) putative admixed individuals were misassigned and confused as reference wildcats, despite their high  $PCADMIX$   $q_w$  values ranging from 0.929 to 0.982 (Table 2) and their ancient origin estimated from 8 to 19 generations before sampling (Table 2). Interestingly, the combined SNP panel sets did not reduce the assignment power to the parental clusters and admixed  $q_i$  values were strictly correlated with those obtained from the 35k LD-pruned SNP panel set ( $R^2 = 0.971$ ;  $P < 0.0001$  for 35k – 96 + 96 SNPs; and  $R^2 = 0.962$ ;  $P < 0.0001$  considering 35k – 48 + 48 SNPs), see Fig. 5.



	192 SNPs	96 SNPs	48 SNPs	96 + 96 SNPs	48 + 48 SNPs
WC Average $q_w$ value	0.990	0.991	0.988	0.991	0.989
WC min $q_w$ value	0.953	0.948	0.936	0.949	0.944
WC misclassified	32 WC $q_w < 1.000$	28 WC $q_w < 1.000$	28 WC $q_w < 1.000$	27 WC $q_w < 1.000$	26 WC $q_w < 1.000$
Hyb misclassified	4 Hyb $q_w > 0.953$	6 Hyb $q_w > 0.948$	10 Hyb $q_w > 0.936$	7 Hyb $q_w > 0.949$	10 Hyb $q_w > 0.944$
% Hyb identified	9	13	22	16	22
Hyb generation	8–19	8–19	8–19	8–19	8–19
Hyb PCADMIX $q_w$ value	0.929–0.982	0.929–0.982	0.929–0.982	0.929–0.982	0.929–0.982

**Table 2.** Performance of the five reduced SNP panel sets in assignment procedures performed in ADMIXTURE on wildcat ( $n = 57$ ), domestic ( $n = 44$ ) and admixed individuals ( $n = 45$ ), previously identified using the 35k LD-pruned SNP panel set in clustering analyses (see Results). Average and minimum membership proportions to the wild cluster of the reference wildcats (WC), number of misclassified WC showing assignment values  $q_w < 1.000$  compared to the 35k-LD pruned SNP panel set results ( $q_w = 1.000$ ), number of misclassified admixed cats (Hyb) showing assignment values  $q_w \geq$  minimum WC  $q_w$ , percentage of hybrids correctly detected (assignment values  $q_w <$  minimum WC  $q_w$ ), generation time and individual membership proportions of admixed individuals retained by PCADMIX analyses on the 35k LD-pruned SNP panel set, are shown for each AIM SNP panel set.

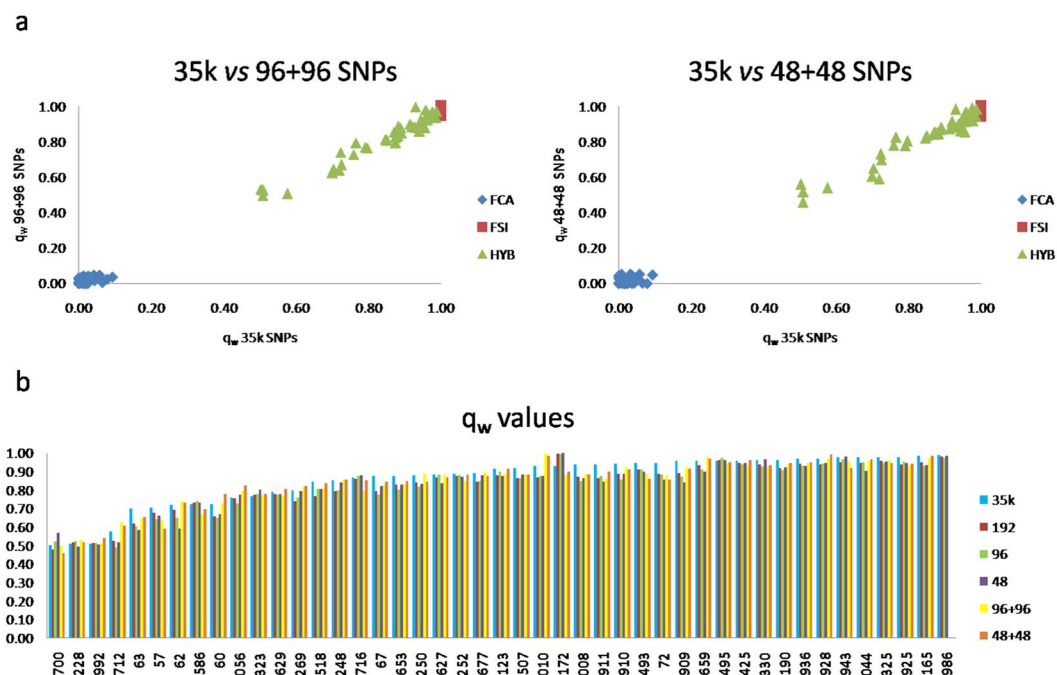
## Discussion

Human-mediated processes, such as habitat destruction, human persecution and anthropogenic hybridization, can directly or indirectly threaten global biodiversity because of their unpredictable consequences on the fitness of natural populations<sup>4,7</sup>. In this study, thanks to the availability of a well-annotated reference genome (FelCat8 – *Felis\_catus\_8.0*<sup>55</sup>), we performed a genome-wide assessment of admixture patterns and timing in a number of European wildcat populations. A preliminary genomic screening, based on pairwise  $F_{ST}$  values, multivariate and assignment procedures, showed that wild and domestic cats remain highly differentiated and well-distinguished. All the analyzed putative admixed cats confirmed to bear admixture signals and were well-identifiable, though some of them were very close to the wildcat group, ranging from *c.* 50% domestic-derived ancestry to almost complete wildcat assignments. Considering a neutral perspective, these patterns clearly indicate that ~40% of the analyzed admixed individuals would fall within the first three hybrid generations, whereas ~60% could represent more ancient backcrosses in which the domestic legacy would have been diluted through time. Genome-wide assignment procedures were also highly efficient in detecting population substructure, since five main biogeographic European wildcat macro-populations were clearly identified, consistent with previous findings based on data generated using 31 autosomal microsatellite loci<sup>27</sup>.

PCADMIX results showed that admixed animals mostly originated from 7 to 14 generations in the past, with some individuals older than the twentieth generation of backcrossing, thus detecting hybridization events occurred between 1962 and 1994 considering a generation time of two years. In particular, the most ancient admixture traces were detected in individuals which had been misclassified as pure in previous microsatellite-based analyses, confirming the deeper diagnostic power of genomic data in detecting past backcrossing events<sup>56–58</sup>. ALDER results estimated the midpoint of hybridization at approximately five generations in the past, compared to the mean value of 12 generations extrapolated from ancestry switches. Such discrepancies between these admixture-dating methods might be attributable to PCADMIX algorithms that are more efficient in detecting more ancient hybridization traces in the genome of the hybrids by identifying their residual domestic blocks. Conversely, ALDER algorithms are more suitable to identify the major admixture event (if a main one occurred), the midpoint (in case of continuous admixture events) or the latest events (if these were punctuated). Such timing patterns, even if preliminary, can provide additional information about the context and the period during which hybridization occurred in our analyzed samples, and thus can be useful to better understand causes and dynamics of the phenomenon at local scale. However, we cannot exclude that more ancient hybridization events have remained undetected in the European cat populations we analyzed since (1) the 63k SNPs still offer a moderate snapshot on the whole *Felis* genome, (2) our sampling design was not homogeneous neither in time nor in space, and (3) data about the collection year were not available for all the analyzed individuals. Future analyses based on a wider sampling and the comparison of entire genomes might shed more light on patterns and histories of admixture.

Interestingly, the six known hybrids, coherently with their possible F2 origins declared in previous analyses<sup>20,21</sup>, had assignment values ranging from 0.347 to 0.673, although they actually showed admixture traces dating back from 9 to 11 generations in the past. Such findings would suggest these animals might represent the product of repeated crosses among F1 or F2 individuals, rather than true second generation hybrids, whose domestic components could be detected only through the haplotype block analyses. These results highlight the power of such methods in improving the assessment of admixture proportions and timing from genomic data by detecting domestic-like and wild-like local genome ancestry better than the assignment procedures or morphology alone<sup>58,59</sup>.

The employment of thousands of SNPs, in fact, allowed us to distinguish backcrossed individuals with small proportions of domestic genome introgressed from wildcat parental populations, by accounting for the number of generations since secondary contact, which can occur when two (or more) species that have been in allopatry come back into sympatry.



**Figure 5.** (a) Scatterplot of individual proportions of membership to the wild clusters ( $q_w$ ) of 146 sampled cats (including domestic, European wild and known/putatively admixed cats), according to the assignment analyses performed in ADMIXTURE. Individual's wild memberships estimated with the combined reduced SNP panel set (96 + 96 and 48 + 48 AIM SNPs, informative for both admixture and genetic structure analyses) were strictly correlated with those obtained with the 35k LD-pruned SNP panel set. (b) Wildcat ancestry proportions of known hybrids and putative admixed cats ( $n = 45$ ) inferred in Admixture with the initial 35k LD-pruned SNP panel set in addition to the reduced (192, 96 and 48) and combined (96 + 96 and 48 + 48) SNP panel sets. All the reduced marker panels did not reduce the assignment power to the parental clusters.

Even if some studies hypothesized that ecological barriers can play a key role in limiting hybridization in some areas (as occurred in some Mediterranean regions<sup>18,19</sup>), variable levels of admixture were detected in all the sampling pools representative of five European wildcat macro-populations we analyzed. Such evidence suggests that, in absence of strong ecological barriers, hybridization can potentially threaten some of the extant wildcat populations, including those living in the Italian Peninsula, as previously described not only for the wildcat<sup>37</sup> but also for other mammalian species such as the wolf<sup>44</sup>, the roe deer<sup>60</sup> and the wild boar<sup>61</sup>.

Since the samples we used in our analyses were not randomly selected, but mostly based on their DNA quality, and included animals collected from previously known areas of suspected hybridization<sup>27,30</sup>, this study did not allow to estimate hybrid prevalence nor the origin and spread of introgression in the local European wild cat populations.

Therefore, we suggest that dynamics and prevalence of hybridization in the European wildcat populations should be better estimated (1) through extensive country-wide sampling programs and genetic analyses of wounded or found-dead wildcats and (2) by well-planned local intensive non-invasive genetic and camera-trapping monitoring projects in hot spots of known or suspected hybridization throughout the entire wildcat distribution range. This approach should avoid the risk that carcasses of introgressed individuals might be confused with feral domestic cats and thus not analyzed, and it would allow to simultaneously obtain detailed phenotypical and genetic information at a European scale, as well as potential capture-recapture data to provide more reliable estimates of population size and hybrid prevalence.

The reduced panels of AIMs we selected from the Illumina Infinium iSelect 63k DNA cat array, based on multiple criteria of genetic differentiation and linkage independence, appeared to be particularly suitable for future large-scale monitoring projects in territories with high conservation priority and in areas of supposed or documented hybridization since (1) they allowed to distinguish individuals, even strictly related, without ambiguity (probabilities of identity  $< 0.001$ ), (2) clearly identified the geographic and genetic European wildcat macro-population structure and variability, (3) were highly concordant with the 35k LD-pruned SNP panel set and more accurate than the small number of previously used microsatellite loci in the identification of admixed individuals.

These reduced SNP panel sets could be easily integrated with other AIMs previously identified in cat admixture studies<sup>25,30</sup> and with markers that will hopefully emerge from coordinated ongoing genomic studies<sup>26</sup>. Routinely, SNP genotyping of both invasively and non-invasively collected samples could be carried out through innovative analysis methods such as quantitative PCR<sup>62</sup> or microfluidic<sup>63–66</sup> techniques, which allow the cost-effective genome-wide characterization of dozens of samples and markers at a time, even starting from low DNA quality or quantity materials. Such approaches turned out to be highly reliable both for multilocus DNA

fingerprinting reconstructions and for the correct identification of admixed individuals until the second backcross generation for a number of *taxa* such as the brown bear<sup>64,66</sup>, the wolf<sup>62,63</sup> and the wildcat<sup>25,26,65</sup>.

We also capitalized the availability of the domestic cat SNP array dataset remapped on the *Felis catus*\_8.0 genome assembly<sup>49</sup> to search for specific genes hosted in both domestic and wildcat-inherited regions, which might be associated with specific biological and phenotypical ontology processes as adaptive response to selective pressures.

The significantly enriched CC wildcat-like genes we identified in the admixed cats were mainly related to brain development and cognitive processes which also regulate aggressiveness and elusive behaviors typical of wild species. A variety of wildcat-like genes included in enriched CC categories were further found to be related with morphological features (such as genes regulating body size and hair coats) and development processes (such as muscle development and the energy metabolism), which might influence the body growth and composition resulting from adaptive pressures. Interestingly, we also identified a few significantly enriched CC genes correlated with fertility (*MORN3*, *NKD1*), the maintenance of pregnancy (*MCL1*, *LNPEP*) and the likelihood of survival (*RNPEP*), that might contribute to increase the fitness of the admixed individuals living sympatrically with wildcats.

Nonetheless, we also found four domestic-derived genes significantly enriched for GO categories mostly related to cognition and behavior, physiological adaptations and neural crest development, whose cellular deficit during embryonic development has been demonstrated to directly or indirectly modify several morphological and physiological traits, as well as to influence tameness during cat domestication, in agreement with the domestication syndrome hypothesis<sup>67</sup>. Such genes might have been maintained in the genome of the introgressed cats thanks to their possible adaptive roles in human-dominated landscapes, where a number of variables are highly modified by the human presence and actions (high density of domesticated *taxa* and their pathogens, habitat fragmentation and perturbation, modified circadian rhythms of prey, etc.) or even in quasi-natural contexts, as demonstrated for the domestic goat MHC haplotypes in the Alpine ibex<sup>10</sup> and the dog-derived black coloration in wolves<sup>9</sup>. The possibility that similar patterns result from an ascertainment bias linked to the original SNP chip design, mostly based on the domestic cat variation, is very unlikely for closely related *taxa* diverging less than one million years<sup>68</sup>.

Cats have experienced a self-domestication history<sup>69,70</sup> in which strong pressures operated by breeding strategies selecting for specific physical features occurred only recently and with limited effects on behavioral traits. Therefore, their gene pool has been poorly isolated from their wild counterparts, and the number of genomic regions with strong signals of selection and differentiation since cat domestication appeared modest<sup>71</sup> compared with those reported in another domesticated species, the domestic dog<sup>72</sup>.

However, though gene enrichment analyses can provide a broad sense of the type of functions that are common to a significant number of genes (in this case, the ones hosted in regions found to be outlier for domestic or wild ancestry), our gene-search approach only allowed us to gain a preliminary insight on the inheritance patterns of domestic and wild ancestry blocks. Indeed, no significant evidences of selection signatures were detected by tests based on  $F_{ST}$  outliers, which are a more direct estimate of deviations from neutrality at a given marker. Such lack of selection could rely on the limited samples analyzed, or reflect the actual absence of differential selection for wild-derived or domestic-derived alleles in the admixed individuals. Therefore, all these data will need to be integrated in the future with systematic studies on fitness of the admixed individuals, including survival and breeding rates, in order to better understand the adaptive patterns of wild-living admixed individuals.

In conclusion, in this study we provide a comprehensive genome-wide approach to detect the occurrence and infer the timing of admixture events in the European wildcat populations investigated, improving the reliability of old-generation backcross identification and pinpointing a number of outlier genes possibly influenced by natural and artificial selection in samples collected from the main genetic macro-populations in Europe<sup>27</sup>. On average 17% of domestic ancestry were detected in the genomes of most analyzed putative admixed cats, which were all classified as backcrosses more ancient than six generations in the past.

Obtaining additional information on the admixture levels, timing and inheritance patterns can improve our understanding on the underlying factors favoring hybridization and its possible consequences, thus supporting the identification of the most appropriate conservation needs.

Consequently, management actions should be mainly aimed at reducing the high number of free-ranging cats within the current wildcat distribution, deserving particular attention to those areas where ecological barriers are not so strong to limit hybridization. Furthermore, priority management actions, such as captivation or sterilization, should be primarily addressed to recent generation hybrids, which carry significant portions of domestic genome ancestry, and eventually further extended toward more ancient backcrosses when they locally occur at high prevalence, thus increasing the probability of interbreeding and retaining domestic variants.

Future genome-wide scanning of a larger number of individuals from the whole European wildcat distribution range and the application of the optimized small marker panels in non-invasive genetic monitoring projects will contribute to (1) assess the hybrid frequencies and the current rates of domestic introgression in the wild populations, (2) provide information on the health status of wild-living individuals (through the analysis of genes related to illness, immune response, reproductive patterns or adaptation to specific ecological pressure), (3) identify areas with high conservational priority where try to limit the occurrence of hybridization and support appropriate local management practices.

## Materials and Methods

**Ethical statements.** No ethics permit was required for this study, and no animal research ethics committee prospectively was needed to approve this research or grant a formal waiver of ethics approval.

**Sampling.** DNA was extracted from blood or muscular tissue samples collected from 100 presumed European wildcats (WC), 46 domestic cats (DC) and 36 known or presumed WC x DC admixed (HY; Table 3). Samples

Subspecies	Acronyms	Macro-populations	No.
Domestic cats	DC	Italy	14
		Dinaric Alps	4
		Central Europe	5
		Iberia	23
		<b>Sub-total</b>	46
Putative and known admixed cats	Hyb	Italy: captivity (known)	8
		Italy	16
		Dinaric Alps	3
		Central Europe	3
		Iberia	6
		<b>Sub-total</b>	36
European wildcats	WC	Italy	15
		Dinaric Alps	35
		Central Germany	10
		Central Europe	21
		Iberia	19
		<b>Sub-total</b>	100
		<b>Total (all cats)</b>	182

**Table 3.** Origin and sample size of the genotyped domestic cats (*Felis silvestris catus*), European wildcats (*F. s. silvestris*) and their putative admixed cats. All cats were preliminarily assigned to their most likely subspecies/macro-population of origin by Bayesian clustering and admixture analyses performed in Mattucci *et al.*<sup>27</sup>.

were collected from a large part of the wildcat distribution range in Europe, including the five main genetic clusters identified by Mattucci *et al.*<sup>27</sup>: Iberian, Italian, Central European, Dinaric and Central Germany populations (Supplementary Fig. S1). All cats were previously analyzed with a few tens of microsatellites<sup>20,21,27,37</sup>.

The vast majority (97%) of the samples used in this study were collected from found-dead cats by specialized technician personnel for scientific purposes. The blood samples ( $n = 5$ ) were collected with permission from owners from domestic cats by veterinarians during their routine health examinations. Additionally no anesthesia, euthanasia, or any kind of animal sacrifice was applied for this study and all blood samples were obtained aiming at minimizing the animal suffering. No ethics permit was required for this study, and no animal research ethics committee prospectively was needed to approve this research or grant a formal waiver of ethics approval.

**Quality-control of the DNA samples and SNP genotyping.** Genomic DNA was extracted using the Qiagen DNAeasy Blood and Tissue kits (Qiagen Inc, Hilden, Germany) according to the manufacturer's instructions, quantified using the Infinite200 PRO NanoQuant (Tecan System Inc, San Jose, USA) and visually-controlled for DNA degradation by standard 1.5% agarose gel electrophoresis. An initial panel of 182 samples, showing no DNA degradation and at least 50 ng/ul DNA, was genotyped using the Infinium iSelect 63k Cat DNA Array (Illumina Inc., San Diego, CA) including 62,897 SNP positions of which 4,240 were wildcat-specific<sup>49</sup>. Considering the alignment of the markers to *Felis\_catus\_8.0*<sup>55</sup> (ICGSC; [https://www.ncbi.nlm.nih.gov/assembly/GCF\\_000181335.2/](https://www.ncbi.nlm.nih.gov/assembly/GCF_000181335.2/)), 704 SNPs (including 7 insertion/deletion makers) did not map to any chromosomes or anchored contigs and were excluded from the analyses. X-linked SNPs ( $n = 2,724$  SNPs) were further excluded. The remaining 59,469 autosomal SNP genotypes were then filtered for individual missingness rates ( $GENO > 0.2$ ), individual missing call rate ( $MIND > 0.2$ ) and number of invariant SNPs in PLINK<sup>73</sup>, resulting in a starting dataset of 146 samples genotyped at 57,302 SNPs (the 57k SNP panel set). This panel included 92% of the wildcat-specific variants<sup>38</sup> ( $n = 3,885$ ). The initial dataset was also pruned for Linkage Disequilibrium (LD), filtering for  $r^2 > 0.5$  in a 50-SNP sliding windows, shifted and recalculated every five SNPs. LD-filtered loci resulted in a dataset of 146 samples genotyped for 35,228 SNPs (the 35k LD-pruned SNP panel set). Based on the analysis undertaken, the most appropriate SNP panel set was utilized.

**Admixture analyses and assignment of the individual genotypes.** Patterns of genetic differentiation among samples was explored by a preliminary non-model Principal Components Analysis<sup>74</sup> (PCA) in the SNP&Variant Suite v.8.0.1 (SVS, Golden Helix Inc., Bozeman, MT) using the 57k SNP panel set. Each sample was then reassigned to its population of origin running the 35k LD-pruned SNP panel set in ADMIXTURE v.1.23<sup>75</sup> assuming K values from 1 to 20. The most likely number of clusters was identified based on the lowest cross validation error<sup>75</sup> and results were plotted in R v.3.5.0 ([www.r-project.org](http://www.r-project.org), last accessed April 23, 2018). Individual ancestry components assessed with ADMIXTURE (at  $K = 2$ ) were then used to select the reference wildcats, reference domestic cats and admixed individuals for all the subsequent analyses (see results).

The occurrence of admixture events in the European wildcat populations was formally tested on the 57k SNP panel set with the F3-statistics running the THREEPOP program implemented in TREEMIX v.1.12, using blocks of 20 adjacent SNPs to estimate standard errors, and Z-score values  $< -3$  to significantly indicate admixture in the target population<sup>53</sup>.



The 57k SNP panel set was further used to infer local ancestry along individual chromosomes and to calculate genome-wide proportions of admixture, through a PCA-based approach implemented in the PCAdmix v.1.0<sup>56</sup>. Each chromosome was analyzed independently, running blocks of 20 consecutive, non-overlapping SNPs, and local-ancestry assignment was based on loadings from principal-component (PC) analysis on the two putative ancestral populations' panels (the reference wildcats and the reference domestic cats). For each admixed individual, we then calculated the average genome-wide proportion of blocks assigned to each reference population.

The reliability of the selected reference populations to detect admixture signals and the absence of any possible ascertainment bias linked to the original SNP chip design based on the domestic cat variation were tested by re-analyzing 20 baseline wildcats and 20 baseline domestic cats, randomly chosen, as putative hybrids in PCAdmix.

**Time of admixture.** We reconstructed chromosomal haplotypes in SHAPEIT v.2.837<sup>76</sup> using the 57k SNP panel set with default parameter settings and considering domestic cat recombination maps<sup>55</sup>. The phased haplotypes were then used to estimate the average time of admixture events between the reference populations (domestic and wild cats) in ALDER v.1.03<sup>54</sup>, which models the signature of decay in Linkage Disequilibrium (LD) between a pair of sites located on the same chromosome as the distance between these sites increases. The putative admixed individuals were first analyzed as a unique cohort and then grouped into cohorts representative of each European wildcat macro-population. Significant admixture events were assessed at P-values < 0.01 and then compared to those estimated using the number of ancestry switches inferred with PCAdmix with the formula developed by Johnson *et al.*<sup>77</sup>, and converted into years since sampling assuming a generation time of two years<sup>78</sup>.

**Local genome ancestry, gene search and gene ontology.** The admixture mapping reconstructed by PCAdmix was finally used to adaptively search for introgressed alleles in the domestic  $\times$  wildcat admixed individuals. We first selected the genome-wide regions showing an excess of domestic or wild cat contributions in the admixed individuals identified by ADMIXTURE. Chromosomal haplotype blocks of 20 SNPs were thus ranked according to their relative proportion of "domestic cat" or "wildcat" assignment detected by PCAdmix (corresponding to 100% domestic or 100% wild cat ancestry, respectively) in order to subsequently identify within them only the top and bottom 1% of the genome-wide frequency distribution that is expected to be enriched for genes bearing signature of positive selection after admixture<sup>79</sup>.

We additionally identified  $F_{ST}$  outlier SNPs at a significant  $P < 0.05$  in BAYESCAN<sup>80</sup> for evidence of selection signatures by comparing admixed individuals *versus* wildcats and admixed individuals *versus* domestic cats. The analysis was performed using default parameter values of 100,000 iterations after an initial burn-in of 50,000 steps, setting a maximum False Discovery Rate (FDR) = 0.05 (the allowed proportion of false positives), and a  $q$  value = 10% (the minimum FDR at which a locus may become significant). Outlier regions were obtained including 100 Kb on each side of the  $F_{ST}$  outlier SNPs detected by BAYESCAN, assuming an average LD in domestic cats of 96 Kb<sup>81</sup>.

Finally, we recovered the genes included in each domestic-like and wildcat-like outlier region obtained from both methods based on the Ensembl gene annotation 92 in BIOMART<sup>82</sup> (<http://www.ensembl.org/biomart/mart-view/>), and checked them for their possible enrichment for any Gene Ontology (GO), Biological Processes (BP) and Human Phenotypes (HP) categories available in G-profiler<sup>83</sup>. Enrichment was tested retaining those categories that were significant at  $P < 0.05$  after Benjamini–Hockberg correction.

### Selection of informative autosomal SNPs for both ancestry detection and population structure.

The 35k LD-pruned SNP panel set was examined to identify a reduced set of ancestry-informative SNPs (AIMs). We first identified the most divergent SNPs between domestic cat and wildcat genotypes (showing ADMIXTURE  $q_w = 1.000$ , and confirmed by PCAdmix, see Results) and ranked the SNPs for decreasing wild  $\times$  domestic cat  $F_{ST}$  values in SVS and for informativeness of the assignment index ( $I_N$ ) in INFOCALC<sup>84</sup>. The Spearman Rank correlation,  $r$ , was estimated among  $F_{ST}$  and  $I_N$  ranks and its significance was tested with a Student's  $t$  test. Three panel sets of 192, 96, and 48 SNPs were finally selected for assignment procedures performed in ADMIXTURE and PCA analyses in SVS to estimate their power to clearly identify reference parental populations (wild and domestic cats) and their admixed individuals.

Subsequently, we identified the most informative SNPs to distinguish the main European wildcat macro-populations<sup>27</sup>, ranking the markers for decreasing macro-pop average  $F_{ST}$  values in SVS and selecting two reduced panel sets of 96 and 48 SNPs that were reanalyzed in a PCA plot.

Finally, all the reduced panel sets of markers were combined to develop the most affordable panel set informative for both admixture and genetic structure analyses.

### Data Availability

The majority of the data generated and analyzed during the current study are presented within the published article or in Supplementary information files. The raw data are available from the corresponding author on reasonable request.

### References

1. Allendorf, F. W., Leary, R. F., Spruell, P. & Wenburg, J. K. The problems with hybrids: setting conservation guidelines. *Trends Ecol. Evol.* **16**, 613–622 (2001).
2. Randi, E. Detecting hybridization between wild species and their domesticated relatives. *Mol. Ecol.* **17**, 285–293 (2008).
3. Laikre, L., Schwartz, M. K., Waples, R. S. & Ryman, N. Compromising genetic diversity in the wild: unmonitored large-scale release of plants and animals. *Trends Ecol. Evol.* **25**, 520–529 (2010).
4. Rhymer, J. M. & Simberloff, D. Extinction by hybridization and introgression. *Annu. Rev. Ecol. Syst.* **27**, 83–109 (1996).
5. Randler, C. Hybrid Wildfowl in Central Europe - an Overview. *Waterbirds* **31**, 143–146 (2008).

6. Goedbloed, D. J. *et al.* Genome-wide single nucleotide polymorphism analysis reveals recent genetic introgression from domestic pigs into Northwest European wild boar populations. *Mol. Ecol.* **22**, 856–866 (2012).
7. Todesco, M. *et al.* Hybridization and extinction. *Evol. Appl.* **9**, 892–908, <https://doi.org/10.1111/eva.12367> (2016).
8. Turek, K. C., Pegg, M. A. & Pope, K. L. Review of the negative influences of non-native salmonids on native fish species. *Great Plains Res.* **23**, 39–49 (2013).
9. Coulson, T. *et al.* Modeling effects of environmental change on wolf population dynamics, trait evolution, and life history. *Science* **334**, 1275–1278, <https://doi.org/10.1126/science.1209441> (2011).
10. Grossen, C., Keller, L., Biebach, I., International, T. & Genome, G. Introgression from domestic goat generated variation at the major histocompatibility complex of Alpine ibex. *PLoS Genet.* **10**(6), e1004438, <https://doi.org/10.1371/journal.pgen.1004438> (2014).
11. McOrist, S. & Kitchener, A. C. Current threats to the European wildcat, *Felis silvestris*, in Scotland. *Ambio*. **23**, 243–245 (1994).
12. Beaumont, M. *et al.* Genetic diversity and introgression in the Scottish wildcat. *Mol. Ecol.* **10**, 319–36 (2001).
13. Randi, E., Pierpaoli, M., Beaumont, M., Ragni, B. & Sforzi, A. Genetic identification of wild and domestic cats (*Felis silvestris*) and their hybrids using Bayesian clustering methods. *Mol. Biol. Evol.* **18**, 1679–93 (2001).
14. Sunquist, M. E. & Sunquist, F. C. *Wild Cats of the World*. The University of Chicago Press, Chicago, USA, 1–452 (2002).
15. Ragni, B. Status and conservation of the wildcat in Italy. *Counc. Eur. Environ. Encount. Ser.* **16**, 40–41 (1993).
16. Vigne, J. D. *et al.* First wave of cultivators spread to Cyprus at least 10,600 y ago. *Proc. Natl. Acad. Sci. USA* **109**, 8445–8449 (2012).
17. Ottoni, C. *et al.* The palaeogenetics of cat dispersal in the ancient world. *Nat. Ecol. Evol.* **1**, 0139 (2017).
18. Lozano, J. & Malo, A. F. Conservation of the European wildcat (*Felis silvestris*) in Mediterranean environments: a reassessment of current threats. In: Williams, G.S. (Ed.), *Mediterranean Ecosystems: Dynamics, Management and Conservation*. Nova Science Publishers Inc., Hauppauge, USA, 1–31 (2012).
19. Gil-Sánchez, J. M., Barea-Azcón, J. M. & Jaramillo, J. Strong spatial segregation between wildcats and domestic cats may explain low hybridization rates on the Iberian Peninsula. *Zoology* **118**, 377–385 (2015).
20. Pierpaoli, M. *et al.* Genetic distinction of wildcat (*Felis silvestris*) populations in Europe, and hybridization with domestic cats in Hungary. *Mol. Ecol.* **12**, 2585–2598 (2003).
21. Lecis, R. *et al.* Bayesian analyses of admixture in wild and domestic cats (*Felis silvestris*) using linked microsatellite loci. *Mol. Ecol.* **15**, 119–131 (2006).
22. Oliveira, R., Godinho, R., Randi, E., Ferrand, N. & Alves, P. C. Molecular analysis of hybridisation between wild and domestic cats (*Felis silvestris*) in Portugal: implications for conservation. *Conserv. Genet.* **9**, 1–11 (2008a).
23. Oliveira, R., Godinho, R., Randi, E. & Alves, P. C. Hybridization versus conservation: are domestic cats threatening the genetic integrity of wildcats (*Felis silvestris silvestris*) in Iberian Peninsula? *Philos. Trans. R. Soc. L. B. Biol. Sci.* **363**, 2953–2961 (2008b).
24. Say, L., Devillard, S., Léger, F., Pontier, D. & Ruetten, S. Distribution and spatial genetic structure of European wildcat in France. *Anim. Conserv.* **15**, 18–27 (2012).
25. Nussberger, B., Wandeler, P., Weber, D. & Keller, L. F. Monitoring introgression in European wildcats in the Swiss Jura. *Conserv. Genet.* **15**(5), 1219–1230 (2014).
26. Steyer, K., Tiesmeyer, A., Muñoz-Fuentes, V. & Nowak, C. Low rates of hybridization between European wildcats and domestic cats in a human-dominated landscape. *Ecol. Evol.* **8**, 2290–2304, <https://doi.org/10.1002/ece3.3650> (2018).
27. Mattucci, F., Oliveira, R., Lyons, L. A., Alves, P. C. & Randi, E. European wildcat populations are subdivided into five main biogeographic groups: Consequences of Pleistocene climate changes or recent anthropogenic fragmentation? *Ecol. Evol.* **6**, 3–22 (2016).
28. Leroy, G. *et al.* Generation metrics for monitoring genetic erosion within populations of conservation concern. *Evol. Appl.* **11**, 1066–1083, <https://doi.org/10.1111/eva.12564> (2017).
29. Nussberger, B., Greminger, M. P., Grossen, C., Keller, L. F. & Wandeler, P. Development of SNP markers identifying European wildcats, domestic cats, and their admixed progeny. *Mol. Ecol. Resour.* **13**, 447–460 (2013).
30. Oliveira, R. *et al.* Toward a genome-wide approach for detecting hybrids: informative SNPs to detect introgression between domestic cats and European wildcats (*Felis silvestris*). *Heredity* **15**, 195–205, <https://doi.org/10.1038/hdy.2015.25> (2015).
31. Daniels, M. J. *et al.* Ecology and genetics of wild-living cats in the north-east of Scotland and the implications for the conservation of the wildcat. *J. Appl. Ecol.* **38**, 146–161 (2001).
32. Devillard, S. *et al.* How reliable are morphological and anatomical characters to distinguish European wildcats, domestic cats and their hybrids in France? *J. Zool. Syst. Evol. Res.* **52**, 154–162 (2014).
33. Driscoll, C. A. *et al.* The Near Eastern origin of cat domestication. *Science* **317**, 519–523 (2007).
34. Hertwig, S. T. *et al.* Regionally high rates of hybridization and introgression in German wildcat populations (*Felis silvestris*, Carnivora, Felidae). *J. Zool. Syst. Evol. Res.* **47**, 283–297 (2009).
35. Eckert, I., Suchentrunk, F., Markov, G. & Hartl, G. B. Genetic diversity and integrity of German wildcat (*Felis silvestris*) populations as revealed by microsatellites, allozymes, and mitochondrial DNA sequences. *Mamm. Biol. - Zeitschrift für Säugetierkd.* **75**, 160–174 (2010).
36. Hartmann, S. A., Steyer, K., Kraus, R. H. S., Segelbacher, G. & Nowak, C. Potential barriers to gene flow in the endangered European wildcat (*Felis silvestris*). *Conserv. Genet.* **14**, 413–426 (2013).
37. Mattucci, F. *et al.* Genetic structure of wildcat (*Felis silvestris*) populations in Italy. *Ecol. Evol.* **3**, 2443–2458 (2013).
38. O'Brien, J. *et al.* Preserving genetic integrity in a hybridising world: are European Wildcats (*Felis silvestris silvestris*) in eastern France distinct from sympatric feral domestic cats? *Biodivers. Conserv.* **18**, 2351–2360 (2009).
39. Steyer, K. *et al.* Large-scale genetic census of an elusive carnivore, the European wildcat (*Felis s. silvestris*). *Conserv. Genet.* **17**(5), 1183–1199, <https://doi.org/10.1007/s10592-016-0853-2> (2016).
40. Davey, J. W. *et al.* Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* **12**, 499–510 (2011).
41. Pontius, J. *et al.* Agencourt Sequencing Team; NISC Comparative Sequencing Program. Initial sequence and comparative analysis of the cat genome. *Genome Res.* **17**(11), 1675–1689 (2007).
42. Kurushima, J. D. *et al.* Variation of cats under domestication: genetic assignment of domestic cats to breeds and worldwide random-bred populations. *Anim. Genet.* **44**(3), 311–324, <https://doi.org/10.1111/age.12008> (2012).
43. Hohenlohe, P. A. *et al.* Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. *Mol. Ecol.* **22**, 3002–3013 (2013).
44. Galaverni, M. *et al.* Disentangling timing of admixture, patterns of introgression, and phenotypic indicators in a hybridizing wolf population. *Mol. Biol. Evol.* **34**, 2324–2339 (2017).
45. Twyford, A. D. & Ennos, R. A. Next-generation hybridization and introgression. *Heredity (Edinb.)* **108**, 179–189 (2012).
46. McFarlane, S. E. & Pemberton, J. M. Detecting the true extent of introgression during anthropogenic hybridization. *Trends Ecol. Evol.* **xx**:1–12, <https://doi.org/10.1016/j.tree.2018.12.013> (2019).
47. Payseur, B. A. & Rieseberg, L. H. A genomic perspective on hybridization and speciation. *Mol. Ecol.* **25**, 2337–2360 (2016).
48. Palamara, P. F. *et al.* Length distributions of identity by descent reveal fine-scale demographic history. *Am. J. Hum. Genet.* **91**, 809–822 (2012).
49. Gandolfi, B. *et al.* Applications and efficiencies of the first cat 63K DNA array. *Sci. Rep.* **8**, 7024, <https://doi.org/10.1038/s41598-018-25438-0> (2018).

50. Lindblad-Toh, K. *et al.* Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803–819 (2005).
51. VonHoldt, B. M. *et al.* A novel assessment of population structure and gene flow in grey wolf populations of the Northern Rocky Mountains of the United States. *Mol. Ecol.* **19**, 4412–4427 (2010).
52. Patterson, N. *et al.* Ancient Admixture in Human History. *Genetics* **192**, 1065–1093 (2012).
53. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**(11), e1002967, <https://doi.org/10.1371/journal.pgen.1002967> (2012).
54. Loh, P., Lipson, M., Patterson, N., Moorjani, P. & Pickrell, J. K. Inferring admixture histories of human populations. *Genetics* **193**, 1233–1254 (2013).
55. Li, G. *et al.* A high-resolution SNP array-based linkage map anchors a new domestic cat draft genome assembly and provides detailed patterns of recombination. *G3* **6**, 1607–1616 (2016).
56. Brisbin, A. *et al.* PCAdmix: Principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. *Hum. Biol.* **84**, 343–364, <https://doi.org/10.3378/027.084.0401> (2012).
57. Pagani, L. *et al.* Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. *Am. J. Hum. Genet.* **91**, 83–96 (2012).
58. Lawson, D. J. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nat. Commun.* **9**, 3258, <https://doi.org/10.1038/s41467-018-05257-7> (2018).
59. Caniglia, R. *et al.* Wolf outside, dog inside? The genomic make-up of the Czechoslovakian Wolfhound. *BMC Genomics* **19**, 533 (2018).
60. Mucci, N., Mattucci, F. & Randi, E. Conservation of threatened local gene pools: landscape genetics of the Italian roe deer (*Capreolus c. italicus*) populations. *Evol. Ecol. Res.* **14**, 897–920 (2012).
61. Scandura, M. *et al.* Genetic diversity in the European wild boar *Sus scrofa*: phylogeography, population structure and wild x domestic hybridization. *Mamm. Rev.* **41**(2), 125–137, <https://doi.org/10.1111/j.1365-2907.2010.00182.x> (2011).
62. vonHoldt, B. M. *et al.* Identification of recent hybridization between gray wolves and domesticated dogs by SNP genotyping. *Mamm. Genome* **24**, 80–88, <https://doi.org/10.1007/s00335-012-9432-0> (2013).
63. Kraus, R. H. *et al.* Single-nucleotide polymorphism-based approach for rapid and cost-effective genetic wolf monitoring in Europe based on noninvasively collected samples. *Mol. Ecol. Resour.* **15**(2), 295–305 (2015).
64. Norman, A. J. & Spong, G. Single nucleotide polymorphism-based dispersal estimates using non invasive sampling. *Ecol. Evol.* **5**, 3056–3065, <https://doi.org/10.1002/ece3.1588> (2015).
65. von Thaden, A. *et al.* Assessing SNP genotyping of non invasively collected wildlife samples using microfluidic arrays. *Sci. Rep.* **7**, 10768, <https://doi.org/10.1038/s41598-017-10647-w> (2017).
66. Giangregorio, P., Norman, A. J., Davoli, F. & Spong, G. Testing a new SNP-chip on the Alpine and Apennine brown bear (*Ursus arctos*) populations using non-invasive samples. *Conserv. Genet. Resour.* <https://doi.org/10.1007/s12686-018-1017-0> (2018).
67. Wilkins, A. S., Wrangham, R. W. & Fitch, W. T. The “Domestication Syndrome” in mammals: a unified explanation based on neural crest cell behavior and genetics. *Genetics* **197**, 795–808 (2014).
68. vonHoldt, B. M. *et al.* A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. *Genome Res.* **21**, 1294–1305 (2011).
69. Cameron-Beaumont, C., Lowe, S. E. & Bradshaw, J. W. S. Evidence suggesting pre adaptation to domestication throughout the small Felidae. *Biol. J. Linn. Soc. Lond.* **75**(3), 361–366 (2002).
70. Driscoll, C. A., Macdonald, D. W. & O’Brien, S. J. From wild animals to domestic pets, an evolutionary view of domestication. *Proc. Natl. Acad. Sci. USA* **106**(Suppl), 9971–9978 (2009).
71. Montague, M. J. *et al.* Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. *Proc. Natl. Acad. Sci.* **111**, 1–6 (2014).
72. Axelsson, E. *et al.* The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* **495**, 360 (2013).
73. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**(3), 559–75 (2007).
74. Novembre, J. & Stephens, M. Interpreting principal component analyses of spatial population genetic variation. *Nat. Genet.* **40**, 646–649 (2008).
75. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664, <https://doi.org/10.1101/gr.094052.109> (2009).
76. Delaneau, O., Marchini, J. & Zagury, J. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181, <https://doi.org/10.1038/nmeth.1785> (2012).
77. Johnson, N. A. *et al.* Ancestral components of admixed genomes in a Mexican cohort. *PLoS Genet.* **7**, e1002410 (2011).
78. Nowak, R. M. *Walker’s Mammals of the World*. Johns Hopkins University Press, Baltimore, USA, 1166–1170 (1999).
79. Barbato, M. *et al.* Genomic signatures of adaptive introgression from European mouflon into domestic sheep. *Sci. Rep.* **7**, 7623, <https://doi.org/10.1038/s41598-017-07382-7> (2017).
80. Foll, M. & Gaggiotti, O. A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics* **180**(2), 977–993, <https://doi.org/10.1534/genetics.108.092221> (2008).
81. Alhaddad, H. *et al.* Extent of linkage disequilibrium in the domestic cat, *Felis silvestris catus* and its breeds. *PLoS One* **8**(1), e53537, <https://doi.org/10.1371/journal.pone.0053537> (2013).
82. Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Res.* **46**(1), 754–761, <https://doi.org/10.1093/nar/gkx1098> (2018).
83. Reimand, J. *et al.* g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.* **44**, 83–89 (2016).
84. Rosenberg, N. A., Li, L. M., Ward, R. & Pritchard, J. K. Informativeness of Genetic Markers for Inference of Ancestry. *Am. J. Hum. Genet.* **73**, 1402–1422 (2003).
85. Bélteki, J., Agnall, B., Johnsson, M., Wright, D. & Jensen, P. Domestication and tameness: brain gene expression in red junglefowl selected for less fear of humans suggests effects on reproduction and immunology. *Royal. Soc. Open Sci.* **3**, 160033, <https://doi.org/10.1098/rsos.160033> (2016).
86. Park, W. *et al.* Investigation of de novo unique differentially expressed genes related to evolution in exercise response during domestication in thoroughbred race horses. *PLoS One* **9**(3), e91418, <https://doi.org/10.1371/journal.pone.0091418> (2014).
87. Davis, S. W. *et al.*  $\beta$ -catenin is required in the neural crest and mesencephalon for pituitary gland organogenesis. *BMC Dev. Biol.* **16**(1), <https://doi.org/10.1186/s12861-016-0118-9> (2016).
88. Leung, A. W., Murdoch, B., Salem, A. F., Prasad, M. S. & Gomez, G. A. Stem cells and regeneration wnt/ $\beta$ -catenin signalling mediates human neural crest induction via a pre-neural border intermediate. *Development* **143**(3), 398–410, <https://doi.org/10.1242/dev.130849> (2016).
89. Chen, G., Zhou, T., Li, Y., Yu, Z. & Sun, L. p53 target miR-29c-3p suppresses colon cancer cell invasion and migration through inhibition of PHLDDB2. *Biochem. Biophys. Res. Commun.* **487**(1), 90–95, <https://doi.org/10.1016/j.bbrc.2017.04.023> (2017).
90. Ai, L., Kim, W. & Alpay, M. TRIM29 Suppresses TWIST1 and invasive breast cancer Behavior. *Cancer Res.* **74**(17), 4875–4887, <https://doi.org/10.1158/0008-5472.CAN-13-3579> (2014).
91. Bermingham, M. L. *et al.* Genome-wide association study identifies novel loci associated with resistance to bovine tuberculosis. *Heredity* **112**(5), 543–551, <https://doi.org/10.1038/hdy.2013.137> (2014).



92. Powell, J. D. & Waters, K. M. Influenza-omics and the host response: recent advances and future prospects. *Pathogens* **6**(2), 25, <https://doi.org/10.3390/pathogens6020025> (2017).
93. Ranaware, P. B., Mishra, A. & Vijayakumar, P. Genome wide host gene expression analysis in chicken lungs infected with avian influenza viruses. *PLoS One* **11**(4), <https://doi.org/10.1371/journal.pone.0153671> (2016).
94. Lan, D. *et al.* Genetic diversity, molecular phylogeny, and selection evidence of Jinchuan yak revealed by whole-genome resequencing. *G3* **8**, 945–952 (2018).
95. De Quervain, D. J. & Papassotiropoulos, A. Identification of a genetic cluster influencing memory performance and hippocampal activity in humans. *PLoS One* **103**, 4270–4274 (2006).
96. Cubelos, B. *et al.* Regulate dendritic branching, spine morphology, and synapses of the upper layer neurons of the cortex. *Neuron* **66**, 523–535 (2010).
97. Bélteki, J., Agnval, B. & Jensen, P. Gene expression of behaviorally relevant genes in the cerebral hemisphere changes after selection for tameness in red junglefowl. *PLoS One* **12**, e0177004 (2017).
98. Van den Berg, L. *et al.* Evaluation of the serotonergic genes htr1A, htr1B, htr2A, and slc6A4 in aggressive behavior of Golden Retriever dogs. *Behav. Genet.* **38**, 55–66 (2008).
99. Pavlov, K. A., Chistiakov, D. A. & Chekhonin, V. P. Genetic determinants of aggression and impulsivity in humans. *J. Appl. Genet.* **53**(1), 61–82, <https://doi.org/10.1007/s13353-011-0069-6> (2012).
100. Greenwood, A. K. & Peichel, C. L. Social regulation of gene expression in threespine sticklebacks. *PLoS One* **10**(9), e0137726, <https://doi.org/10.1371/journal.pone.0137726> (2015).
101. Kerr, D. J. *et al.* Aberrant hippocampal Atp8a1 levels are associated with altered synaptic strength, electrical activity, and autistic-like behavior. *BBA - Mol. Basis Dis.* **1862**, 1755–1765 (2016).
102. Joeyen-Waldorf, J. *et al.* Adenylate Cyclase 7 is implicated in the biology of depression and modulation of affective neural circuitry. *BPS* **71**, 627–632 (2012).
103. Uetake, Y., Terada, Y., Matulien, J. & Kuriyama, R. Interaction of Cep135 With a p50 dynactin subunit in mammalian centrosomes. *Cell. Motil. Cytoskeleton* **66**, 53–66 (2004).
104. Fleischman, R. A., David, L., Stastny, V. & Zneimer, S. Deletion of the c-kit protooncogene the human developmental defect piebald trait. *Proc. Natl. Acad. Sci. USA* **88**, 10885–10889 (1991).
105. Pulos, W. L. & Hutt, F. B. Lethal Dominant White in Horses. *J. Hered.* **60**, 59–63 (1969).
106. Kurihara, Y. *et al.* Elevated blood pressure and craniofacial abnormalities in mice deficient in endothelin-1. *Nature* **368**, 703 (1994).
107. Ai, H. *et al.* Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat. Genet.* **47**, 217 (2015).
108. Ponsuksili, S., Murani, E. & Schellander, K. Identification of functional candidate genes for body composition by expression analyses and evidencing impact by association analysis and mapping. *Biochim. Biophys. Acta.* **1730**, 31–40 (2005).
109. Moio, B., Andrea, M. D. & Pilla, F. Candidate genes affecting sheep and goat milk quality. *Small Rumin. Res.* **68**, 179–192 (2007).
110. Chen, Z., Yao, Y., Ma, P., Wang, Q. & Pan, Y. Haplotype-based genome-wide association study identifies loci and candidate genes for milk yield in Holsteins. *PLoS One* **13**(2), e0192695, <https://doi.org/10.1371/journal.pone.0192695>.
111. Zhang, L. *et al.* Characterization of membrane occupation and recognition nexus repeat containing 3, meiosis expressed gene 1 binding partner, in mouse male germ cells. *Asian J. Androl.* **17**(1), 86–93, <https://doi.org/10.4103/1008-682X.138186> (2015).
112. Li, Q., Ishikawa, T., Miyoshi, H., Oshima, M. & Taketo, M. M. A targeted mutation of Nkd1 impairs mouse spermatogenesis. *J. Biol. Chem.* **280**(4), 2831–2839, <https://doi.org/10.1074/jbc.m405680200> (2005).
113. Boumela, I. *et al.* Involvement of BCL2 family members in the regulation of human oocyte and early embryo survival and death: gene expression and beyond. *Reproduction* **141**, 549–561 (2011).
114. Kim, J. *et al.* Sequence variants in oxytocin pathway genes and preterm birth: a candidate gene association study. *BMC Med. Genet.* **14**(1), <https://doi.org/10.1186/1471-2350-14-77> (2013).
115. Cawthon, M. C., Kerber, R. A., Hasstet, S. J. & O'Brien, E. Methods and kits for determining biological age and longevity based on gene expression profiles. *U.S. Patent Application* **13**(28), 910 (2011).
116. Chase, K., Jones, P., Martin, A., Ostrander, E. A. & Lark, K. G. Genetic mapping of fixed phenotypes: disease frequency as a breed characteristic. *J. Hered.* **100**, S37–S41 (2009).
117. Wang, Z. *et al.* Genome-Wide Association study for wool production traits in a Chinese Merino sheep population. *PLoS One* **9**(9), e107101, <https://doi.org/10.1371/journal.pone.0107101> (2014).
118. Pielberg, R. G. *et al.* A cis-acting regulatory mutation causes premature hair graying and susceptibility to melanoma in the horse. *Nat. Genet.* **40**, 1004 (2008).
119. Ghosh, M. *et al.* An integrated in silico approach for functional and structural impact of non-synonymous SNPs in the MYH1 gene in Jeju Native Pigs. *BMC Genet.* **17**(1), <https://doi.org/10.1186/s12863-016-0341-1> (2016).
120. Li, Y. *et al.* Full Paper A survey of transcriptome complexity in *Sus scrofa* using single-molecule long-read sequencing. *DNA Res.* **25**(4), 421–437, <https://doi.org/10.1093/dnares/dsy014> (2018).
121. Quintens, R. *et al.* Mice deficient in the respiratory chain gene Cox6a2 are protected against high-fat diet-induced obesity and insulin resistance. *PLoS One* **8**(2), e56719, <https://doi.org/10.1371/journal.pone.0056719> (2013).
122. Muller, A. J. *et al.* Targeted disruption of the murine Bin1/Amphiphysin II gene does not disable endocytosis but results in embryonic cardiomyopathy with aberrant myofibril formation. *Mol. Cell. Biol.* **23**(12), 4295–4306 (2003).
123. Sell-Kubiak, E. *et al.* Genome-wide association study reveals novel loci for litter size and its variability in a Large White pig population. *BMC Genomics* **16**, 1049 (2015).
124. Xiaolong, W. *Identification and Characterization of Candidate Genes for Complex Traits in Cattle*. PhD Thesis. Technische Universität München (2013).
125. Hasina, R. *et al.* NOL7 is a nucleolar candidate tumor suppressor gene in cervical cancer that modulates the angiogenic phenotype. *Oncogene* **25**(4), 588–598, <https://doi.org/10.1038/sj.onc.1209070> (2006).
126. Koba, R., Oguma, K. & Sentsui, H. Overexpression of feline tripartite motif-containing 25 interferes with the late stage of feline leukemia virus replication. *Virus Res.* **204**, 88–94 (2015).
127. Palgrave, C. J. *et al.* Species-Specific variation in RELA underlies differences in NF- $\kappa$ B activity: a potential role in African swine fever pathogenesis. *J. Virol.* **85**(12), 6008–6014, <https://doi.org/10.1128/jvi.00331-11> (2011).
128. Tsao, N., Lee, M. H., Zhang, W., Cheng, Y. C. & Chang, Z. F. The contribution of CMP kinase to the efficiency of DNA repair. *Cell Cycle* **14**, 354–363 (2015).
129. Vattem, K. M. & Wek, R. C. Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proc. Natl. Acad. Sci. USA* **101**, 11269–11274, <https://doi.org/10.1073/pnas.0400541101> (2004).
130. Haas, A. V. & McDonnell, M. E. Pathogenesis of Cardiovascular Disease in Diabetes. *Endocrinol. Metab. Clin. NA.* **47**, 51–63 (2018).
131. Dunner, S. *et al.* Genes involved in muscle lipid composition in 15 European *Bos taurus* breeds. *Animal Genet.* **44**(5), 493–501, <https://doi.org/10.1111/age.12044> (2013).
132. Cantù, C. *et al.* Mutations in Bcl9 and Pygo genes cause congenital heart defects by tissue-specific perturbation of Wnt/ $\beta$ -catenin signaling. *Genes & Development*, <https://doi.org/10.1101/gad.315531.118> (2018).
133. Philipp, U., Steinmetz, A. & Distl, O. Development of feline microsatellites and SNPs for evaluating primary cataract candidate genes as cause for cataract in Angolan lions (*Panthera leo bleyenberghi*). *J. Hered.* **101**, 633–638 (2010).

134. Fyfe, J. C. *et al.* An similar to 140-kb deletion associated with feline spinal muscular atrophy implies an essential LIX1 function for motor neuron survival. *Genome Res.* **16**, 1084–1090 (2006).
135. Flavigny, J. *et al.* Identification of two novel mutations in the ventricular regulatory myosin light chain gene (MYL2) associated with familial and classical forms of hypertrophic cardiomyopathy. *J. Mol. Med.* **76**(3–4), 208–214 (1998).
136. Andrzej, J., Magdalena, G. & Beata, H. SNP genetic diversity within a fragment of the gene myo15a responsible for the hearing process in a population of farmed and free-living animals of the canidae family. *Acta Vet.* **64**, 358–366 (2014).

## Acknowledgements

This project was supported by grants from Italian Ministry of Environment (MATTM) and the Italian Institute for Environmental Protection and Research (ISPRA). Partial funding was provided by the Cat Health Network (D12FE-505) and the Winn Feline Foundation (W10-014) (LAL) with “Genetic Estimation of Introgression Between Domestic Cat and Wildcat Populations” and “Cat Phenotypic Health and Information Registry (Cat PHIR)”. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We kindly thank all the people who have contributed to the collection of cat samples, in particular the Portuguese National Tissue Bank/National Institute for Nature and Biodiversity Conservation (BTVS/ICNF), the Tissue Collection at the Zoological Museum of ISPRA (Ozzano dell’Emilia, Bologna, Italy), the Italian Ministry of Environment (MATTM), the Italian Forestry Authority (CFS) and the Italian Institute for Environmental Protection and Research (ISPRA), M. Herrman, F. Suchentrunk, M. Liberek, A. Kitchener, M. Beaumont, B. Szolt, L. Szemethy, A. Sforzi, B. Ragni, L. Lapini, A. De Faveri, K. Hupe, I. Eckert, H. Potocnik, M. Moes, F. Vercillo, L. Bizzarri, J. Godoy, M. Malsaña, M. Mejias, J.M. Fernández, J.L. Robles, G.D. Penafiel, E.B. Duperón, M. Moleón, P. Monterroso, F. Álvares, J.C. Brito, J. Rodrigues, P. Lyberakis, and their collaborators. We are particularly grateful to N. Mucci head of the Conservation Genetic Laboratory of ISPRA and to N. Cappai and C. Pedrazzoli from the National Park of the Foreste Casentinesi, Monte Falterona e Campigna (PNFC). We are also grateful to all anonymous veterinarians and biologists that assisted in samples collection.

## Author Contributions

F.M., M.G., E.R. and R.C. conceived, designed and planned the experiments. F.M., L.A.L., P.C.A., E.R. and R.C. contributed reagents/materials/analysis tools. F.M. performed laboratory experiments. F.M. and R.C. analyzed the data. L.P., M.G., E.R. and R.C. developed the analysis pipeline and provided the experimental supplies. F.M. and R.C. wrote the manuscript, prepared figures and/or tables, reviewed drafts of the paper and performed all the elaborations. L.A.L., P.C.A. and E.R. provided samples and/or data. F.M., M.G. and R.C. shared ideas to realize the manuscript and revised drafts of the paper. F.M., M.G., L.A.L., P.A.C., E.R., E.V., L.P. and R.C. read, reviewed and approved the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-48002-w>.

**Competing Interests:** The authors declare no competing interests.

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



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

© The Author(s) 2019