Complex feline disease mapping using a dense genotyping array

Isabel Hernandez[1¶], Jessica J. Hayward[2¶*], Jeff A. Brockman[3], Michelle E. White[4], Lara Mouttham[5], Elizabeth A. Wilcox[5], Susan Garrison[5], Marta G. Castelhano[5], John P. Loftus[1], Filipe Espinheira Gomes[1], Cheryl Balkman[1], Marjory B. Brooks[6], Nadine Fiani[1], Marnin Forman[7], Tom Kern[1], Bruce Kornreich[1], Eric Ledbetter[1], Santiago Peralta[1], Angela M. Struble[1], Lisa Caligiuri[1], Elizabeth Corey[5], Lin Lin[5], Julie Jordan[5], Danny Sack[1], Adam R. Boyko[2], Leslie A. Lyons[8], Rory J. Todhunter[1]

[1]Department of Clinical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, New York, United States of America

[2]Department of Biomedical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, New York, United States of America

[3]Pet Nutrition Center, Hill's Pet Nutrition, Topeka, Kansas, United States of America

[4]Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, Massachusetts, United States of America and Vertebrate Genomics Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America

[5]Cornell Veterinary Biobank, College of Veterinary Medicine, Cornell University, Ithaca, New York, United States of America

[6]Department of Population Medicine and Diagnostic Services, College of Veterinary Medicine, Cornell University, Ithaca, New York, United States of America

[7]Cornell University Veterinary Specialists, Stamford, Connecticut, United States of America

[8]Department of Veterinary Medicine & Surgery, College of Veterinary Medicine, University of Missouri, Columbia, Missouri, United States of America

*Corresponding author:

Email: jessica.hayward@cornell.edu

35 ¶These authors contributed equally to this work.

36

37 Short title: Feline complex disease mapping

38

39 Keywords: feline complex disease, genome-wide association study, biobank

40

Abstract

The current feline genotyping array of 63k single nucleotide polymorphisms has proven its utility within breeds, and its use has led to the identification of variants associated with Mendelian traits in purebred cats. However, compared to single gene disorders, association studies of complex diseases, especially with the inclusion of random bred cats with relatively low linkage disequilibrium, require a denser genotyping array and an increased sample size to provide statistically significant associations. Here, we undertook a multi-breed study of 1,122 cats, most of which were admitted and phenotyped for nine common complex feline diseases at the Cornell University Hospital for Animals. Using a proprietary 340k single nucleotide polymorphism mapping array, we identified significant genome-wide associations with hyperthyroidism, diabetes mellitus, and eosinophilic keratoconjunctivitis. These results provide genomic locations for variant discovery and candidate gene screening for these important complex feline diseases, which are relevant not only to feline health, but also to the development of disease models for comparative studies.

Introduction

There are 365 hereditary disorders of cats listed on OMIA (Online Mendelian Inheritance in Animals, https://omia.org/home/ accessed June 7th, 2021), of which only 119 (32.6%) are Mendelian traits and only 136 (37.3%) have likely causal variants. Clearly, there are a large number of feline diseases whose genetic basis is still unknown. Moreover, 230 of these hereditary feline disorders are potentially good models for human disease.

Random bred cats are the most common cats in American households, accounting for 84% of the population in the United States [1]. Random bred cats comprised 89% of cats admitted the Cornell University Hospital for Animals (CUHA) in the last 15 years, thus providing an important spontaneous source of DNA for increasing sample sizes of genetic mapping studies.

Compared to purebreds, random bred cats have shorter linkage disequilibrium, due to the large number of generations since the origin of the random bred cat population, with archaeological evidence of a human and cat burial site as old as 9,500 years [2]. The genetic heterogeneity of random bred cats, the additive effect of many genes, and their environmental interaction makes discovering variants contributing to complex diseases more challenging than for Mendelian traits [3]. At least a few Mendelian traits have been mapped in random bred cats, including spongy encephalopathy, Glanzmann thrombasthenia, and inflammatory linear verrucous epidermal nevus [4–6]. Additional factors that make the discovery of complex disease genetic mechanisms difficult include sample size, phenotyping accuracy, mapping array marker density, and access to whole genome sequences for variant discovery [7].

The current 63k Illumina feline single nucleotide polymorphism (SNP) mapping array has been used successfully within breeds to map variants for Mendelian diseases. Examples include the discovery of the *WNK4* variant that causes hypokalemia in Burmese cats [8], a region on chromosome E1 associated with progressive retinal atrophy in Persian cats [9], a causal variant in *COLQ* for hereditary myopathy in Devon Rex and Sphynx cats [10], refinement of the region on chromosome B4 associated with craniofacial structure and frontonasal dysplasia in Burmese cats [11], a region on chromosome A3 associated with an inherited neurologic syndrome in a family of

Oriental cats [12], and a dominant channelopathy variant causing osteochondrodysplasia in Scottish Fold cats [13]. This array has also been used in a limited number of within-breed genome wide association studies (GWAS) for complex disease [14,15], but there are no reports of GWAS performed with an across-breed design.

Here, we genotyped 1,122 cats using a one-time proprietary Illumina high density 340k SNP mapping array designed by Hill's Pet Nutrition, in an effort to identify genetic underpinnings for nine complex diseases. Our samples consisted of a mix of 31 purebreds and 905 random bred cats, the majority of which were domestic shorthairs. This array improves upon the density of the current commercial 63k array by a factor of >5. As quality control and to validate the accuracy of the 340k array, we performed a GWAS for the *Orange* coat color locus and for Factor XII deficiency, which are known to be associated with a region on chromosomes X and A1, respectively [16–19].

The complex diseases included in this study were hypertrophic cardiomyopathy (HCM), hyperthyroidism, diabetes mellitus (DM), chronic kidney disease (CKD), chronic enteropathy, inflammatory bowel disease (IBD), small cell alimentary lymphoma (SCAL), hypercalcemia, and feline eosinophilic keratoconjunctivitis (FEK). These diseases are among the most common complex diseases of cats admitted to CUHA and are some of the most common and important feline diseases in clinical veterinary practice [20].

We used both a linear mixed model (LMM) and a multi-locus method called Fixed and random model Circulating Probability Unification (FarmCPU) to perform GWAS, and together identified loci significantly associated with hyperthyroidism, DM, FEK, and IBD. Additionally, we

identified suggestive loci for the diseases HCM and hypercalcemia. Here, we describe the largest genetic mapping study of feline complex diseases with the densest mapping array ever performed.

Results

Validation of array

Principal component analysis (PCA) was performed using all genotyped cats that passed quality control, and showed that there was no batch effect due to the 11 sequential plates used for genotyping (Fig. 1A). The first two components, principal component (PC)1 and PC2, explained 31.3% of the total genetic variation. The cluster that separates on PC2 in this PCA includes 40 cats from a closed colony of domestic shorthair (DSH) cats from a local breeding facility, genotyped mainly on plates 7 and 11. Principal component analysis of the genotypes of all 221 purebred cats showed that PC1 separates western breeds, like Manx and Persian, from eastern breeds, like Tonkinese and Burmese (Fig. 1B). This eastern-western distribution of breeds is also seen on PC1 of the PCA of all cats (Fig. 1A) and has been shown previously using the 63k genotyping array [16,21–23]. PC2 of the purebred cat PCA separates the Devon Rex cats from the other breeds. The first two components of the purebred cat PCA explained only 16.4% of the total genetic variation, much less than the 38.4% explained by the first two components of a PCA using the 63k array [23].

Figure 1. **Principal Component Analysis of cat genetic structure.** Dimensions PC1 and PC2 are shown. (A) All 1,122 cats that passed QC, color-coded by genotyping plate (1 to 11), showing the absence of a batch effect. PC1 shows the eastern-western breed distribution. The cluster of cats that separate on PC2 is from a local colony that were genotyped on plates 2 (dark blue), 7 (brown),

and 11 (orange). (B) 221 purebred cats color-coded by breed, showing the eastern-western breed distribution on PC1. The Devon Rex breed (dark green) separates on PC2.

GWAS positive controls

As a positive control, we performed a GWAS on the presence of orange fur in random bred cats (90 orange fur, 121 black/brown fur). Using the linear mixed model in GEMMA, we identified 25 significant associations on a region of chromosome X between 102,884,842 bp and 112,136,902 bp (Fig. 2A; S1 Table). The most significantly associated SNP in both the LMM and FarmCPU GWAS is at 110,230,748 bp (P=1.8×10^{-102} and P=2.2×10^{-97}, respectively), located within an intron in the gene *Ecto-NOX Disulfide-Thiol Exchanger 2 (ENOX2)*. This region is known to contain the *Orange* cat coloration locus [16–18] and the most significant SNP is within the 1.5 Mb haplotype block identified by Gandolfi *et al.* (2018). A linkage disequilibrium (LD) plot of this region showed that the 340k array has very few markers between 105-110 Mb on chromosome X, and only 4 markers within the 1.5 Mb haplotype block remain after minor allele frequency (MAF) and missingness filters, preventing the refinement of this region (S1A Fig.).

Figure 2. Manhattan and quantile-quantile (QQ) plots for GWAS positive controls. X axis represents the chromosomal SNP position and Y axis represents the -log$_{10}$(P-value). The QQ plots show observed versus expected P-values for each SNP. (A) *Orange* coat color locus, showing the significant association on chromosome X (P=1.8×10^{-102}). (B) Factor XII deficiency, showing the significant associations on chromosomes A1 and C2. The red line on the Manhattan plots shows the Bonferroni-corrected significance threshold and the blue line on the Manhattan plot in B shows

154 the Bonferroni-corrected significance threshold calculated using unlinked SNPs. The genomic
155 inflation factor (**λ**) is shown on each QQ plot.
156
157 As a second positive disease control, we performed a GWAS for factor XII deficiency, using 19
158 affecteds and 34 controls. The LMM in GEMMA identified four significant associations on
159 chromosome A1, between 175,333,103 bp and 175,445,463 bp, which reside within 63 kb of the
160 gene *Coagulation Factor XII (F12)* (Fig. 2B, S1B Fig.). The most significant association using the
161 FarmCPU method was the same A1 association at 175,445,463 bp (P=1.4×10-19). Two high-
162 frequency variants in the gene *F12* have previously been reported in cats with factor XII deficiency
163 [19]. Other significant associations were also identified in the factor XII GWAS by both models,
164 on chromosomes C2, C1, D2, F1 and D3.
165

166 Disease GWAS

167 Across-breed case/control GWAS was conducted for the diseases HCM, hyperthyroidism,
168 DM, CKD, chronic enteropathy, IBD, SCAL, FEK, hypercalcemia, and all gastrointestinal
169 phenotypes (chronic enteropathy, IBD, and SCAL) merged together. Significance thresholds were
170 calculated using the Bonferroni correction on all SNPs included in each GWAS, while suggestive
171 thresholds were calculated using the Bonferroni correction on a pruned set of unlinked SNPs.
172
173 Three significant and two suggestive associations were identified above the genome-wide
174 thresholds by the LMM GWAS in GEMMA (Table 1). The FarmCPU GWAS showed very similar
175 results to the LMM GWAS, with significant associations for DM and hyperthyroidism (Table 1;
176 S2 Table). However, the FEK association was not significant in the FarmCPU GWAS while the

177 IBD association was significant (Table 1; S2 Table). Since the results from the two methods were
178 so similar, we have chosen to focus illustrating the results of the LMM GWAS. Genomic inflation
179 factors, □, are all <1.07 (range of 0.997-1.052, average 1.016 for LMM; range of 1.013-1.062,
180 average 1.033 for FarmCPU), showing successful control for underlying population structure.
181
182 **Table 1: Significant and suggestive associations identified for complex diseases using an**
183 **across-breed GWAS design. Results are shown for both the LMM and FarmCPU GWAS.**

Disease (number cases, number controls)	**GWAS model**	**Chr: bp**	***P*-value**	**Allele frequency (cases, controls)**	**Candidate genes**
Hyperthyroidism	LMM	B2: 121,565,607	1.25×10^{-7}	0.037, 0.127	*ARG1, MED23*
(310, 134)	FarmCPU	B2: 121,565,607	1.36×10^{-7}		
DM	LMM	D4: 83,583,678	1.62×10^{-7}	0.366, 0.172	olfactory receptors,
(67, 366)	FarmCPU	D4: 83,583,678	2.55×10^{-7}		*PTGS1*
FEK	LMM	E3: 34,663,327	1.62×10^{-7}	0.100, 0.638	*TNFRSF17*
(15, 40)	FarmCPU	E3: 34,663,327	1.79×10^{-6} #		
IBD	LMM	B4: 10,941,073	2.75×10^{-6} #	0.216, 0.533	N/A
(47, 33)	FarmCPU	B4: 10,941,073	9.54×10^{-8}		
HCM	LMM	E3: 3,583,882	2.76×10^{-7} ^	0.295, 0.604	*SDK1*
(85, 53)	FarmCPU	E3: 3,583,882	1.00×10^{-6} #		
Hypercalcemia	LMM	C1: 19,508,050	6.81×10^{-7} ^	0.300, 0.085	N/A

(25, 443)	FarmCPU	C1: 19,508,050	3.05×10^{-7} ^

^ suggestive association based on unlinked SNPs

not significant

Hyperthyroidism. For hyperthyroidism, we found a solitary significant association on chromosome B2 (P=1.25×10^{-7} in LMM, P=1.36×10^{-7} in FarmCPU), located in the gene *Arginase 1 (ARG1)* and 5.5 kb downstream of, although not in LD with, the gene *Mediator Complex Subunit 23 (MED23)* (Fig. 3A). The B2 locus increases hyperthyroidism risk in DSH cats (S3 Table).

Figure 3. **Manhattan, quantile-quantile (QQ), and LD plots for case-control disease significant associations, using the LMM GWAS results.** X axis represents the chromosomal SNP position and Y axis represents the $-\log_{10}$(P-value). The QQ plots show observed versus expected P-values for each SNP. (A) Hyperthyroidism, showing the significant association on chr B2. (B) DM, showing the significant association on chr D4. (C) FEK, showing the significant association on chr E3. On Manhattan plots, the red line is Bonferroni-corrected significance threshold, and the blue line is Bonferroni-corrected significance threshold calculated using unlinked SNPs. Inflation factors (λ) are shown on QQ plots. On LD plots, the colors indicate the amount of LD (r^2) with the most significant SNP, ranging from black (r^2<0.2) to red (r^2>0.8).

Diabetes mellitus. Diabetes mellitus was significantly associated with a SNP on chromosome D4 (P=1.62×10^{-7} in LMM, P=2.55×10^{-7} in FarmCPU) (Fig. 3B). The LD region includes many members of the olfactory receptor gene family, such as *OR1J*, *OR1N*, *OR1K*, and *OR5C*, among other genes. The gene *PTGS1*, (*prostaglandin synthase G/H isoform 1),* also known as *COX1*

(*cyclooxygenase-1*), is located within 123 kb downstream of, although not in LD with, our significant association. This locus on D4 affects the risk of DM in DSH and Maine Coon cats, but not in cats of other breeds and DLH cats (S3 Table).

Feline eosinophilic keratoconjunctivitis. We identified a significant association for FEK (P=1.62×10^{-7}) in the LMM GWAS, with a marker on chromosome E3, located 10.5 kb from the gene *TNFRSF17* (*tumor necrosis factor receptor superfamily, member 17*) (Fig. 3C). The second most significant association with this disease did not reach significance (P=3.1×10^{-6}) but is located within the gene *TNFRSF21* (*tumor necrosis factor superfamily, member 21*). Both *TNFRSF17* and *TNFRSF21* belong to the tumor necrosis factor receptor superfamily, and *TNFRSF21* is expressed in the eye [24]. The E3 locus affects the risk for FEK in DSH cats (S3 Table).

IBD. A significant association (P=9.54×10^{-8}) for IBD was identified using the FarmCPU GWAS. The marker is on chromosome B4 near the genes *ECHDC3 (enoyl-CoA hydratase domain containing 3)* and *USP6NL (ubiquitin-specific protease 6 N-terminal like)* (S2 Fig.). *ECHDC3* has a role in fatty acid biosynthesis and has been found to have an increased expression in the brains of Alzheimer's patients [25] while *USP6NL* is a GTPase-activating protein for Rabs and is up-regulated in several cancers, including breast and colorectal cancers [26,27]. The B4 significant locus affects risk for IBD in DSH and DLH cats (S3 Table).

HCM. The LMM GWAS for HCM reached suggestive significance (P=2.76×10^{-7}) with a marker on chromosome E3, located within the gene *SDK1* (*sidekick cell adhesion molecule 1*) (Fig. 4A),

228 which is expressed especially in the kidney and retina [28,29] but has also been associated with
229 hypertension [30]. This suggestive E3 locus affects risk for HCM in DSH cats (S3 Table).
230
231 **Figure 4**. **Manhattan, quantile-quantile (QQ), and LD plots for case-control disease**
232 **suggestive associations, using the LMM GWAS results.** X axis represents the chromosomal
233 SNP position and Y axis represents the $-\log_{10}$(*P*-value). The QQ plots show observed versus
234 expected *P*-values for each SNP. (A) HCM, showing the suggestive association on chr E3. (B)
235 Hypercalcemia, showing the suggestive association on chr C1. On Manhattan plots, the red line is
236 Bonferroni-corrected significance threshold, and the blue line is Bonferroni-corrected significance
237 threshold calculated using unlinked SNPs. Inflation factors (**λ**) are shown on QQ plots. On LD
238 plots, the colors indicate the amount of LD (r^2) with the most significant SNP, ranging from black
239 ($r^2<0.2$) to red ($r^2>0.8$).
240
241 **Hypercalcemia.** The hypercalcemia GWAS produced a suggestive association ($P=6.81\times10^{-7}$ in
242 LMM, $P=3.05\times10^{-7}$ in FarmCPU) on chromosome C1, located in the gene *PAFAH2* (*Platelet*
243 *Activating Factor Acetylhydrolase 2*) and within LD of the gene *STMN1* (*Stathmin 1*) (Fig. 4B).
244 The enzyme encoded by *PAFAH2* acts to protect the cell from oxidative cytotoxicity [31], while
245 the protein encoded by *STMN1* is involved in regulating the microtubule cytoskeleton, including
246 mitotic spindle formation [32]. The C1 locus affects risk for hypercalcemia in DSH cats (S3 Table).
247
248 Genome-wide association studies of the other complex diseases, CKD, SCAL, chronic
249 enteropathy, and merged GI phenotypes did not produce a significant or suggestive association
250 using either the LMM or FarmCPU GWAS (S3 Fig., S2 Table, S4 Table).

Discussion

In this study, we identified significant associations for common, clinically relevant, complex diseases in a population of 1,122 random and purebred cats, using a dense genotyping array. While a similar study was previously performed in dogs [33], this is the largest GWAS disease study in cats reported to date, conducted in a heterogeneous natural population including 80% random bred cats. Further advantages of the current study design were the careful phenotyping of aged control cats, accurate phenotyping of diseased participants by specialists performed in an academic clinical setting, and a mapping array approximately 5-fold denser than the current 63k array. Additionally, the quality of the biospecimens used and its associated data demonstrate the importance of using an accredited resource such as the Cornell Veterinary Biobank.

As a positive control, we identified significant associations for the *Orange* coat color locus and factor XII deficiency at the *F12* gene locus. Although the *F12* locus was the most significant association using both LMM and FarmCPU models, three and five other significant SNPs were identified in the factor XII deficiency GWAS, respectively. A BLAT [34] search showed that the flanking region of the three SNP maps to many places in the feline genome, including chromosome A1: 175 Mb, the location of the gene *F12*. Thus, it appears that there may be some non-specific binding with the A1 probe. However, factor XII deficiency is affected by several different loci across the genome, as shown in humans [35].

The majority of the cats included in our analyses are random bred cats, which generally have shorter LD than purebreds [3], because they have not been subject to selective breeding for specific traits. Further, we are mapping complex diseases, which usually consist of many variants each contributing a small effect, and have not been subjected to artificial selection, resulting in shorter LD surrounding the causal variant. As a result of investigating complex diseases in a predominantly random bred cat population, we do not expect to see the stacking of SNPs that are seen in GWAS studies of morphological traits, especially in purebred cats.

Using a case/control approach, we performed GWAS with both a LMM and FarmCPU, and found very similar results. Both methods identified significant associations for hyperthyroidism and DM, and the FEK association was significant in the LMM GWAS while the IBD association was significant in the FarmCPU GWAS. Furthermore, the same SNPs were identified as the most significant associations by both models.

For hyperthyroidism, the candidate gene *ARG1* encodes Arginase 1, a cytosolic enzyme that participates in the urea cycle and is expressed in the liver [36]. Another nearby candidate gene, *MED23,* encodes a component of the thyroid hormone receptor (TR) associated protein complex. As such, it interacts with, and facilitates, TR function. Variants in the TR have been associated with thyroid hormone resistance, for which the clinical presentation is very similar to thyrotoxicosis [37]. This is the first GWAS for feline hyperthyroidism reported and our finding represents a novel locus. Somatic variants in the *thyroid-stimulating hormone receptor (TSHR)* gene have been previously reported, but those variants were identified in DNA extracted from the affected thyroid glands of hyperthyroid cats [38].

The significant DM locus includes many olfactory receptor genes. Genetic and epigenetic variation, and the resulting functional changes, in olfactory receptors have been associated with taste, food intake, and satiety [39,40]. These differences may contribute to obesity risk and risk of DM. Mouse olfactory receptor gene OLFR15 has been shown to be expressed in pancreatic beta-cells and to regulate the secretion of insulin [41]. The other interesting gene near our significant D4 association, although not quite within the LD region of interest, *PTGS1,* encodes an enzyme that converts arachinodate into prostaglandin, which is involved in glucose homeostasis [42]. This gene has been associated with human DM [43,44]. Our findings constitute a novel locus associated with DM. Previous studies have identified several loci associated with DM in Australian Burmese cats [15,45] and a polymorphism in *melanocortin receptor 4 (MCR4)* associated with DM in obese domestic shorthair cats [46].

For eosinophilic keratoconjunctivitis, we identified a significant association in the LMM GWAS near the gene *TNFRSF17*. This is especially promising, and warrants further investigation because of its role in the innate and adaptive immune response. In patients with allergic asthma, eosinophils infiltrate the bronchial wall and lumen, and the bronchial epithelium is often damaged [47]. These pathological findings are associated with aberrant T helper 2 (Th2) cell-mediated immune responses. Interleukin-5, which is produced by Th2 cells, and the chemokine eotaxin are key players for the proliferation, differentiation, activation and mobilization of eosinophils [48,49]. In knockout mice studies, NF-kappa-B, a transcription factor that is activated by the *TNFRSF17* and *TNFRSF21* genes, was found to play an important role in Th2 cell differentiation and is therefore required for induction of allergic airway inflammation [48,50]. Similar to knockout mice with

allergic asthma, it is possible that animals affected with FEK have an abnormal NF-kappa-B activation due to defective expression of *TNFRSF17* and *TNFRSF21* genes, as suggested by the current GWAS study.

The IBD significant association, as identified by the FarmCPU method only, was located near the genes *ECHDC3* and *USP6NL*. Neither of these genes are good candidates for a gastroenteropathy phenotype.

The first of the two suggestive associations, HCM, was located in the gene *SDK1*. A polymorphism in *SDK1* was found to be associated with hypertension in a study of over 5,000 Japanese individuals [30] but the function of this gene related to hypertension has not been described. Finally, the LD region surrounding the suggestive association for hypercalcemia contained the genes *PAFAH2* and *STMN1*, neither of which have been associated with hypercalcemia previously.

Despite the use of a dense genotyping array, across-breed GWAS for CKD, SCAL, chronic enteropathy, and merged gastrointestinal phenotypes did not reach statistical genome-wide significance using either single-locus or multi-locus models. We believe that larger cohorts may be needed due to the genetic architecture of these diseases, especially chronic enteropathy for which we had fewer than 50 cases in the respective GWAS.

By not restricting our analyses to a single breed, we were able to include a relatively large sample size for some of our phenotypes, thereby increasing statistical power to identify significant loci. In

a study of this kind, especially if the majority of cats are randomly bred, LD is shorter, resulting in smaller regions of interest and narrowing the list of potential candidate genes.

Nevertheless, for some other phenotypes, we had an unbalanced proportion of cases and controls. This is due to the fact that accumulation of samples takes a long time, in part because donating samples is an opt-in process in our hospital, and because of the difficulty of recruiting universal controls. It is an unanswered question how many samples are required for a robust across-breed, complex disease GWAS study in cats, but canine simulation studies indicate that 500-1000 cases and controls, plus a further increase in array marker density, would substantially increase loci discovery in dogs [33].

Follow-up analyses using an independent cohort of phenotyped cats are needed to validate the associations we identified in this group of genotyped cats. Further studies involving investigation of the regions surrounding the significant associations are needed to determine causal variants for these complex diseases. Use of the >300 whole genome sequences provided by the 99Lives Feline Genome Consortium will allow variant discovery within candidate genes in the intervals of interest. Finally, functional studies will be required to confirm causal variants.

In this research, we used an across-breed GWAS design with a ~5-fold denser genotyping array than currently available, to identify significant associations with important common feline complex diseases. We demonstrated that a well-curated, hospital-sourced population can be used effectively for mapping studies. We also demonstrated the benefit of such a dense mapping array, propelling the field of complex feline disease genetics forward. Further, these results can be used

364 to develop new diagnostic tests to assist veterinarians in identifying diseases earlier and allowing
365 the implementation of early preventative measures. Breeders could improve their practices by
366 identifying cats with optimum genetic value and owners could make informed decisions regarding
367 the health of cats. This is particularly important in this era of personalized medicine. The shared
368 environment of cats and their owners further enhances the value of domestic cats as models of
369 lifestyle disease common to both species.
370

371 Materials and methods

372 Banking biospecimens and associated data

373 The 1,122 feline biospecimens used for this project were selected from the Cornell
374 Veterinary Biobank (CVB), a core resource at the Cornell University College of Veterinary
375 Medicine, which has been collecting and processing whole blood samples from feline patients
376 admitted to the Cornell University Hospital for Animals (CUHA) since 2006. Biospecimens from
377 participants consented at our satellite clinic, the Cornell University Veterinary Specialist in
378 Stamford, Connecticut, were also included.
379

380 Out of the 1,122 cats, 57 were recruited through the Senior Feline Health Screening program from
381 2014 to 2018. The program was created to build a biobank of DNA and associated clinical data
382 from healthy senior cats to serve as universal controls for mapping studies. In order to participate,
383 feline candidates had to be at least 9.5 years of age and in good health. Privately owned cats that
384 participated in the screening had a general physical examination and were examined accordingly
385 by board certified specialists: cardiac auscultation and echocardiogram, dental examination, body

386 condition scoring, body mapping (used by oncologists to record any masses found), ocular
387 examination, and an orthopedic examination. A complete blood count, serum chemistry panel,
388 coagulation panel, feline immunodeficiency virus (FIV) and feline leukemia virus (FeLV) test,
389 baseline serum thyroxine (T4) level, and urinalysis were performed.
390

391 ## Sample Processing, Storage and Distribution

392 Samples were collected according to the Cornell University Institutional Animal Care and
393 Use Committee (IACUC) protocol #2005-0151. Following owner informed consent, whole blood
394 samples were collected in EDTA tubes and refrigerated at 4℃ until DNA extraction. Formalin
395 fixed, paraffin embedded (FFPE) scrolls of splenic tissue were acquired from a collaborating
396 pathologist and used for DNA extraction when necessary. Genomic DNA was extracted from
397 blood samples using a standard salt precipitation. Genomic DNA was extracted from FFPE
398 samples using the E.Z.N.A. Tissue DNA kit (Omega Bio-Tek) following the manufacturer's
399 instructions. DNA concentration and purity were determined by spectrophotometry on a
400 NanoDrop ND1000 (Thermo Scientific). DNA samples were stored at ≤-20°C until distribution
401 for genotyping.
402

403 ## Inclusion criteria

404 Participants with a disease of interest could simultaneously be used as controls for other
405 traits/diseases, as long as these traits were ruled out. Phenotypes included cases and controls from
406 any breed, unless specified. Controls were at least 9.5 years of age, while cases could be of any
407 age. Numbers of purebred and random bred cats included as cases and controls for each GWAS

408 are shown in Table 2 and numbers of individuals from each breed are shown in S5 Table. The
409 distribution of cases and controls by age is shown for each phenotype in S4 Fig.

410

411 **Table 2: Distribution of cases and controls that are purebred and random bred cats for each**
412 **disease.**

Disease	PUREBRED Number cases, number controls	RANDOM BRED^ Number cases, number controls	Total number cases, controls
HCM	14, 20	69, 33	83, 53
Hyperthyroidism	24, 38	286, 96	310, 134
DM	13, 62	54, 304	67, 366
CKD	29, 10	125, 52	154, 62
Chronic enteropathy	9, 10	32, 18	41, 28
IBD	9, 15	38, 18	47, 33
SCAL	12, 16	66, 18	78, 34
GI combined	33, 16	133, 18	166, 34
FEK	2, 13	13, 27	15, 40
Hypercalcemia	5, 78	20, 365	25, 443

413 ^ includes domestic shorthair (DSH), domestic longhair (DLH), domestic medium hair (DMH), as
414 well as cats identified as breed mixes (e.g., Siamese mix)

415

Hypertrophic cardiomyopathy. HCM is the most common cardiac disease in cats, affecting around 15% of the feline population [51,52]. Similarly to humans, familial HCM has been described in purebred cats, and in Maine coon and Ragdolls is caused by mutations in myosin binding protein C gene (MYBPC3) [53]. Some Maine coon and Ragdolls cats develop HCM in the absence of this mutation, indicating that other mutations are yet to be identified [53]. Diagnosis was based on echocardiography. Phenotypic criteria for controls included normal left ventricular wall thickness measurements: left ventricular free wall (LVFW) and interventricular septum (IVS) in diastole ≤ 6 mm by M-mode (motion mode). Phenotypic criteria for cases included LVFW and IVS wall thickness >6 mm. Additionally, affected cats must have had normal baseline T4 and be normotensive and normally hydrated in order to rule out other causes of cardiac hypertrophy.

Hyperthyroidism. Hyperthyroidism is one of the most common endocrine disorders affecting senior cats. The disease most often results from benign adenomatous thyroid nodules similar to human toxic nodular goiter [54]. Hyperthyroidism is believed to be a multifactorial disease, with nutritional, environmental, and genetic factors postulated as interacting causes [54]. The diagnosis of cases and controls was based on the following criteria: control cats had low-normal thyroxine (T4; <3 □g/dL; normal range 2-5 □g/dL). Cases had T4 >5 □g/dL or normal T4 with increased free T4. Radioiodinated thyroid scan results confirming the diagnosis were recorded, if available.

Diabetes mellitus. DM is also one of the most common endocrine diseases of cats with the majority of the cats resembling Type 2 (adult onset) DM in humans. The disease is caused by a combination of decreased β-cell function, insulin resistance, and environmental and genetic factors [55]. Diagnosis of DM was based on the following criteria: control cats had blood glucose values

<200 mg/dL (reference values: 71-182 mg/dL) and no glucosuria. Cases had elevated blood glucose (>250 mg/dL) and glucosuria in at least two consecutive visits. Also, fructosamine, if evaluated, had to be above normal range (174-294 μmol/L). Of the 67 cases in the GWAS, 39 had fructosamine tests and all had elevated levels. Fifty-three diabetic cases and 339 controls had body weight recorded. Although the body weights of cases were spread throughout the range of 1.8kg to 10kg, a greater proportion of cases (12 of 53, or 22.6%) had weights >7kg, compared to controls (17 of 339, or 5.0%) (S5 Fig.).

Chronic kidney disease. CKD is highly prevalent in both humans and cats with approximately 10% of cats >10 years of age reported to be affected. Cats with CKD experience a progressive loss of functional renal mass. CKD is considered a heterogeneous syndrome, rather than a single entity [56]. CKD was diagnosed by evaluating the level of blood urea nitrogen (BUN) and creatinine, in conjunction with the urine specific gravity (USG). Symmetric dimethylarginine (SDMA), a natural occurring indicator for kidney function, was measured in the blood of some cases to determine if early renal disease was occurring. The diagnosis was established according to the following criteria: controls cats had creatinine <1.6 mg/dL (normal range 0.6-2 mg/dL), BUN within normal range (16-36 mg/dL) and USG >1.035 (preferably performed on the same day as creatinine was measured). Cases had to be azotemic (elevated BUN and creatinine values) with concurrent isosthenuria (failure of the kidney to dilute or concentrate urine) and increased SDMA, diagnosed by a board-certified veterinary internist.

Chronic enteropathy/inflammatory bowel disease/small cell alimentary lymphoma. Chronic enteropathies, which include Inflammatory Bowel Disease (IBD) and

462 Small Cell Alimentary Lymphoma (SCAL), are common forms of primary gastrointestinal disease
463 in cats. Although the cause of feline IBD is unknown, it has been hypothesized that, similar to
464 canines and humans, feline IBD is caused by several factors such as intestinal microbial
465 imbalances, diet, and defects in the mucosal immune system [57]. SCAL is the most frequent
466 digestive neoplasia in cats, accounting for 60-75% of gastrointestinal lymphoma cases [58].
467
468 For this study, cats were assigned as chronic enteropathy cases if gastrointestinal (GI) clinical signs
469 such as chronic vomiting, diarrhea, or weight loss were present, non-GI causes of their clinical
470 signs were excluded, thus highly suggestive of either IBD or SCAL, but no histologic diagnosis
471 was performed. IBD and SCAL were considered separate diagnoses that required histological
472 confirmation. Distinguishing between IBD and SCAL can be difficult, so in addition to histologic
473 assessment, immunophenotyping and polymerase chain reaction (PCR) for antigen receptor
474 rearrangements (PARR) were used in some cases to confirm the SCAL diagnosis. Phenotypic
475 criteria for affected cats included persistent clinical GI signs and histopathology performed by a
476 board-certified veterinary pathologist confirming either IBD or SCAL. Control cats were
477 examined by a board-certified oncologist and had an absence of any GI signs. We performed a
478 separate GWAS for each of IBD and SCAL, and then chronic enteropathy, which includes cats
479 that were not formally diagnosed but could be either IBD or SCAL. Finally, we performed a
480 GWAS including all GI cases in an attempt to increase statistical power, and since IBD, SCAL,
481 and chronic enteropathy can be considered a different manifestation of the same disorder [59].
482 There is also evidence that IBD leads to SCAL [60].
483

Feline eosinophilic keratoconjunctivitis. FEK is a corneal/conjunctival disease characterized by vascularized white-to-pink plaques on the cornea and bulbar conjunctiva. In the majority of cats, previous corneal ulceration has been diagnosed and an association with feline herpesvirus type 1 (FHV-1) infection has been proposed [61]. The diagnosis of FEK was made according to the following criteria: affected cats had signs of the disease during ophthalmologic exam performed by a board-certified veterinary ophthalmologist, including proliferative vascularized lesions affecting peripheral corneal/bulbar conjunctiva and the presence of eosinophils in the ocular cytology. Control cats had a normal ophthalmologic exam.

Hypercalcemia. Hypercalcemia is a common condition of cats defined by an increase in both total and ionized serum calcium. It may be caused by many conditions such as neoplasia, renal failure, primary hyperparathyroidism, hypoadrenocorticism, ingestion of cholecalciferol-containing rodenticides, or granulomatous disease. In cats, hypercalcemia can also be idiopathic [62], which is the phenotype we are investigating here. The diagnosis of hypercalcemia was determined as follows: control cats had total serum calcium values within the normal range (9.1-10.9 mg/dL); affected cats had elevated total serum calcium and ionized calcium values (reference interval 1.11-1.38 mmol/L). Parathyroid hormone (PTH) and PTH related peptide (PTHrP) were recorded if available, and were used to differentiate between causes of hypercalcemia.

Design of array

Genotyping was performed on an Illumina Infinium iSelect Custom BeadChip. These arrays contain 340,000 attempted beadtypes for genotyping single nucleotide polymorphisms

selected across the entire cat genome, using feline genome assembly felCat5. Of the 340,000 markers included on the array, 297,034 (87%) provided a reliable call.

SNPs for the array were selected from whole genome sequencing of 6 genetically diverse female DSH cats. These 6 cats were sequenced on a HiSeq2500 (Illumina, San Diego, CA) to generate 100bp paired-end reads. Following GATK best practices pipeline [63], reads were mapped to the feline reference genome using BWA mem [64], then duplicate reads were tagged by PICARD MarkDuplicates, and indels were realigned and quality scores were recalibrated using GATK. Variants were called and filtered using GATK HaplotypeCaller and VCFtools [65]. The full list of variants was thinned randomly using PLINK and then protein-coding variants with moderate and high impact as defined by SnpEff [66] were added back in.

Genotyping

In total, 1,200 feline DNA samples were genotyped on the Hill's custom Illumina feline high density mapping array. Genotyping was performed in 11 batches, or plates, by Neogen GeneSeek Operations (Lincoln, NE). Raw data files were converted to PLINK format and quality control was performed in PLINK v1.9 (www.cog-genomics.org/plink/1.0/) [67,68].

Quality control

Genotyping data from the 11 batches were merged together using PLINK's --bmerge command and a sex check of all samples was performed using PLINK's --check-sex command. Seventy samples were removed due to missingness >80%, including 53 samples from the same batch.

529

530 SNPs were converted to the genome assembly felCat9 [69] and SNPs with missingness >95% in
531 the 1,130 cats were removed, leaving 252,987 SNPs. Eight cats were genotyped on two different
532 plates each as internal controls. The SNPs that were discordant between these eight duplicates
533 were identified and removed. Finally, duplicate samples were removed, leaving a dataset of 1,122
534 individuals and 251,978 SNPs for GWAS.

535

536 A Principal Component Analysis (PCA) was performed using the program EIGENSTRAT in the
537 EIGENSOFT package [70]. For this, linked SNPs were pruned using PLINK's --indep 50 5 2
538 command, leaving 91,556 SNPs. PCA was performed using all cats to look for batch effects, and
539 all purebred cats to ensure individuals of the same breed clustered together. PCA was also
540 performed using only the cats included in each phenotype to identify and remove outliers before
541 GWAS analysis. An outlier is an individual that is located separately from the main cluster of cats
542 on either the PC1 or PC2 axis. Further, in order to reduce the effects of genetically distinct
543 individuals in our GWAS, we also removed any purebred cat that was located separately from the
544 main cluster of random bred individuals on either PC1 or PC2.

545

546 For the DM and HCM phenotypes, a further two and 16 cats, respectively, were genotyped on the
547 same 340k custom Illumina array by external coauthors (MEW and JAB, respectively). For these
548 cats, the genotype files were merged with the sample set before the QC was performed, as
549 described above. The genotype and phenotype data for all three of these datasets are available as
550 PLINK files, and include the SNP information (chromosome, bp location, alleles).

551

Genome-wide association study

Both a single-locus linear mixed model (LMM) and a multi-locus model were used to perform a GWAS for each disease phenotype. The LMM was performed in the program GEMMA v 0.98.1 [71], which includes a relatedness matrix as a random effect. The multi-locus method performed was FarmCPU (Fixed and random model Circulating Probability Unification) [72] run using rMVP [73] in R. FarmCPU is designed to help control for false positives by including associated markers as covariates, while also reducing false negatives by removing the confounding between the population structure and kinship and the markers to be tested. We used the default parameters, with a maximum of 10 iterations. For each phenotype, we included the relatedness matrix calculated by GEMMA and a covariate file consisting of the first four PCs from a PCA run on the genotypes of the cats included in the phenotype only.

For both models, the Wald test was used to calculate *P*-values, and the Bonferroni correction (p_{genome}=0.05) was used to calculate the genome-wide significance threshold. A suggestive threshold was calculated using the Bonferroni correction on unlinked SNPs (pruned using the --indep 50 5 2 option in PLINK).

For each phenotype, PCA outliers and related cats (pihat>0.40) were excluded. Single nucleotide polymorphisms with a minor allele frequency (MAF) <5% and a genotyping call rate < 90% were removed from each analysis. SNPs are provided in genome assembly felCat9.

Manhattan and quantile-quantile (QQ) plots were created using the package qqman [74] in R v4.0.2 [75]. Lambda values, as a quantification for genomic inflation, were calculated in R. Linkage disequilibrium plots were created using matplolib [76] in jupyter notebook [77].

GWAS positive controls

As a positive control for the 340k array, we performed GWAS on the presence of orange fur. The *Orange* locus has been refined to a 1.5Mb region on the X chromosome, although the causal variant is unknown [16–18]. We used 211 random bred cats in the orange GWAS: 90 cats that had a coat color description of orange (including solid orange, orange and white, and orange tabby), and 121 cats that had a coat color description of black, brown or brown tabby.

We also performed a positive control GWAS of factor XII deficiency, a common hereditary coagulation factor deficiency in cats that does not cause a bleeding diathesis. For this phenotype, affected cats were classified based on severe factor XII deficiency (factor XII coagulant activity < 10% of normal), whereas control cats had values above 60%. Nineteen affected cats and 34 controls were included in the GWAS.

Acknowledgements

We especially thank the faculty and staff of the Cornell University Hospital for Animals for sample collection and phenotyping of control cats. We would like to acknowledge Gregory Acland and John Schimenti for instigation of the Cornell Veterinary Biobank, and support from Baker Institute for Animal Health, Cornell University Center for Vertebrate Genomics, and

595 Department of Clinical Sciences. Finally, we thank numerous pet owners for donating their pets'
596 samples, and collaborators for sample collection and phenotyping.

597

598 References

599 1. AVMA Pet Ownership and Demographics Sourcebook. 2017-2018 edition. Schaumburg,IL:
600 American Veterinary Medical Association, 2018.

601 2. Driscoll CA, Clutton-Brock J, Kitchener AC, O'Brien SJ. The Taming of the cat. Genetic
602 and archaeological findings hint that wildcats became housecats earlier--and in a different
603 place--than previously thought. Sci Am. 2009;300: 68–75.

604 3. Alhaddad H, Khan R, Grahn RA, Gandolfi B, Mullikin JC, Cole SA, et al. Extent of
605 Linkage Disequilibrium in the Domestic Cat, Felis silvestris catus, and Its Breeds. PLoS
606 ONE. 2013;8. doi:10.1371/journal.pone.0053537

607 4. Takaichi Y, Chambers JK, Shiroma-Kohyama M, Haritani M, Une Y, Yamato O, et al.
608 Feline Spongy Encephalopathy With a Mutation in the ASPA Gene. Vet Pathol. 2021;
609 3009858211002176. doi:10.1177/03009858211002176

610 5. Li RHL, Ontiveros E, Nguyen N, Stern JA, Lee E, Hardy BT, et al. Precision medicine
611 identifies a pathogenic variant of the ITGA2B gene responsible for Glanzmann's
612 thrombasthenia in a cat. J Vet Intern Med. 2020;34: 2438–2446. doi:10.1111/jvim.15886

613 6. Lucia MD, Bauer A, Spycher M, Jagannathan V, Romano E, Welle M, et al. Genetic
614 variant in the NSDHL gene in a cat with multiple congenital lesions resembling

615 inflammatory linear verrucous epidermal nevi. Vet Dermatol. 2019;30: 64-e18.
616 doi:10.1111/vde.12699

617 7. Baker L, Muir P, Sample SJ. Genome-wide association studies and genetic testing:
618 Understanding the science, success, and future of a rapidly developing field. Journal of the
619 American Veterinary Medical Association. American Veterinary Medical Association;
620 2019. pp. 1126–1136. doi:10.2460/javma.255.10.1126

621 8. Gandolfi B, Gruffydd-Jones TJ, Malik R, Cortes A, Jones BR, Helps CR, et al. First
622 WNK4-Hypokalemia Animal Model Identified by Genome-Wide Association in Burmese
623 Cats. PLoS ONE. 2012;7. doi:10.1371/journal.pone.0053173

624 9. Alhaddad H, Gandolfi B, Grahn RA, Rah HC, Peterson CB, Maggs DJ, et al. Genome-wide
625 association and linkage analyses localize a progressive retinal atrophy locus in Persian cats.
626 Mamm Genome. 2014;25: 354–362. doi:10.1007/s00335-014-9517-z

627 10. Gandolfi B, Grahn RA, Creighton EK, Williams DC, Dickinson PJ, Sturges BK, et al.
628 COLQ variant associated with Devon Rex and Sphynx feline hereditary myopathy. Anim
629 Genet. 2015;46: 711–715. doi:10.1111/age.12350

630 11. Lyons LA, Erdman CA, Grahn RA, Hamilton MJ, Carter MJ, Helps CR, et al. Aristaless-
631 Like Homeobox protein 1 (ALX1) variant associated with craniofacial structure and
632 frontonasal dysplasia in Burmese cats. Dev Biol. 2016;409: 451–458.
633 doi:10.1016/j.ydbio.2015.11.015

12. Yu Y, Creighton EK, Buckley RM, Lyons LA. A deletion in GDF7 is associated with a heritable forebrain commissural malformation concurrent with ventriculomegaly and interhemispheric cysts in cats. Genes. 2020;11: 1–15. doi:10.3390/genes11060672

13. Gandolfi B, Alamri S, Darby WG, Adhikari B, Lattimer JC, Malik R, et al. A dominant TRPV4 variant underlies osteochondrodysplasia in Scottish fold cats. Osteoarthritis Cartilage. 2016;24: 1441–1450. doi:10.1016/j.joca.2016.03.019

14. Golovko L, Lyons LA, Liu H, Sørensen A, Wehnert S, Pedersen NC. Genetic susceptibility to feline infectious peritonitis in Birman cats. Virus Res. 2013;175: 58–63. doi:10.1016/j.virusres.2013.04.006

15. Samaha G, Wade CM, Beatty J, Lyons LA, Fleeman LM, Haase B. Mapping the genetic basis of diabetes mellitus in the Australian Burmese cat (Felis catus). Sci Rep. 2020;10. doi:10.1038/s41598-020-76166-3

16. Gandolfi B, Alhaddad H, Abdi M, Bach LH, Creighton EK, Davis BW, et al. Applications and efficiencies of the first cat 63K DNA array. Sci Rep. 2018;8. doi:10.1038/s41598-018-25438-0

17. Schmidt-Küntzel A, Nelson G, David VA, Schäffer AA, Eizirik E, Roelke ME, et al. A domestic cat X chromosome linkage map and the sex-linked orange locus: Mapping of orange, multiple origins and epistasis over nonagouti. Genetics. 2009;181: 1415–1425. doi:10.1534/genetics.108.095240

18. Grahn RA, Lemesch BM, Millon L V., Matise T, Rogers QR, Morris JG, et al. Localizing the X-linked orange colour phenotype using feline resource families. Anim Genet. 2005;36: 67–70. doi:10.1111/j.1365-2052.2005.01239.x

19. Maruyama H, Brooks MB, Stablein A, Frye A. Factor XII deficiency is common in domestic cats and associated with two high frequency F12 mutations. Gene. 2019;706: 6–12. doi:10.1016/j.gene.2019.04.053

20. O'Neill DG, Church DB, McGreevy PD, Thomson PC, Brodbelt DC. Prevalence of disorders recorded in cats attending primary-care veterinary practices in England. Vet J. 2014;202: 286–291. doi:10.1016/j.tvjl.2014.08.004

21. Kurushima JD, Lipinski MJ, Gandolfi B, Froenicke L, Grahn JC, Grahn RA, et al. Variation of cats under domestication: Genetic assignment of domestic cats to breeds and worldwide random-bred populations. Anim Genet. 2013;44: 311–324. doi:10.1111/age.12008

22. Lipinski MJ, Froenicke L, Baysac KC, Billings NC, Leutenegger CM, Levy AM, et al. The ascent of cat breeds: Genetic evaluations of breeds and worldwide random-bred populations. Genomics. 2008;91: 12–21. doi:10.1016/j.ygeno.2007.10.009

23. Alhaddad H, Abdi M, Lyons LA. Patterns of allele frequency differences among domestic cat breeds assessed by a 63K SNP array. Barsh GS, editor. PLOS ONE. 2021;16: e0247092. doi:10.1371/journal.pone.0247092

24. Pan H, Wu S, Wang J, Zhu T, Li T, Wan B, et al. TNFRSF21 mutations cause high myopia. J Med Genet. 2019;56: 671–677. doi:10.1136/jmedgenet-2018-105684

673 25. Desikan RS, Schork AJ, Wang Y, Thompson WK, Dehghan A, Ridker PM, et al. Polygenic
674 Overlap Between C-Reactive Protein, Plasma Lipids, and Alzheimer Disease. Circulation.
675 2015;131: 2061–2069. doi:10.1161/CIRCULATIONAHA.115.015489

676 26. Sun K, He S-B, Yao Y-Z, Qu J-G, Xie R, Ma Y-Q, et al. Tre2 (USP6NL) promotes
677 colorectal cancer cell proliferation via Wnt/β-catenin pathway. Cancer Cell Int. 2019;19:
678 102. doi:10.1186/s12935-019-0823-0

679 27. Avanzato D, Pupo E, Ducano N, Isella C, Bertalot G, Luise C, et al. High USP6NL Levels
680 in Breast Cancer Sustain Chronic AKT Phosphorylation and GLUT1 Stability Fueling
681 Aerobic Glycolysis. Cancer Res. 2018;78: 3432–3444. doi:10.1158/0008-5472.CAN-17-
682 3018

683 28. Yamagata M, Weiner JA, Sanes JR. Sidekicks: synaptic adhesion molecules that promote
684 lamina-specific connectivity in the retina. Cell. 2002;110: 649–660. doi:10.1016/s0092-
685 8674(02)00910-8

686 29. Kaufman L, Potla U, Coleman S, Dikiy S, Hata Y, Kurihara H, et al. Up-regulation of the
687 homophilic adhesion molecule sidekick-1 in podocytes contributes to glomerulosclerosis. J
688 Biol Chem. 2010;285: 25677–25685. doi:10.1074/jbc.M110.133959

689 30. Oguri M, Kato K, Yokoi K, Yoshida T, Watanabe S, Metoki N, et al. Assessment of a
690 Polymorphism of SDK1 With Hypertension in Japanese Individuals. Am J Hypertens.
691 2010;23: 70–77. doi:10.1038/ajh.2009.190

31. Tjoelker LW, Stafforini DM. Platelet-activating factor acetylhydrolases in health and disease. Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids. Biochim Biophys Acta; 2000. pp. 102–123. doi:10.1016/S1388-1981(00)00114-1

32. Rubin CI, Atweh GF. The role of stathmin in the regulation of the cell cycle. J Cell Biochem. 2004;93: 242–250. doi:10.1002/jcb.20187

33. Hayward JJ, Castelhano MG, Oliveira KC, Corey E, Balkman C, Baxter TL, et al. Complex disease and phenotype mapping in the domestic dog. Nat Commun. 2016;7. doi:10.1038/ncomms10460

34. Kent WJ. BLAT---The BLAST-Like Alignment Tool. Genome Res. 2002;12: 656–664. doi:10.1101/gr.229202

35. Soria JM, Almasy L, Souto JC, Bacq D, Buil A, Faure A, et al. A Quantitative-Trait Locus in the Human Factor XII Gene Influences Both Plasma Factor XII Levels and Susceptibility to Thrombotic Disease. Am J Hum Genet. 2002;70: 567–574.

36. You J, Chen W, Chen J, Zheng Q, Dong J, Zhu Y. The oncogenic role of ARG1 in progression and metastasis of hepatocellular carcinoma. BioMed Res Int. 2018;2018. doi:10.1155/2018/2109865

37. Olateju TO, Vanderpump MPJ. Thyroid hormone resistance. Annals of Clinical Biochemistry. Ann Clin Biochem; 2006. pp. 431–440. doi:10.1258/000456306778904678

710 38. Watson SG, Radford AD, Kipar A, Ibarrola P, Blackwood L. Somatic mutations of the
711 thyroid-stimulating hormone receptor gene in feline hyperthyroidism: Parallels with human
712 hyperthyroidism. J Endocrinol. 2005;186: 523–537. doi:10.1677/joe.1.06277

713 39. Precone V, Beccari T, Stuppia L, Baglivo M, Paolacci S, Manara E, et al. Taste, olfactory
714 and texture related genes and food choices: Implications on health status. Eur Rev Med
715 Pharmacol Sci. 2019;23: 1305–1321. doi:10.26355/eurrev_201902_17026

716 40. Ramos-Lopez O, Riezu-Boj JI, Milagro FI, Zulet MA, Santos JL, Martinez JA, et al.
717 Associations between olfactory pathway gene methylation marks, obesity features and
718 dietary intakes. Genes Nutr. 2019;14. doi:10.1186/s12263-019-0635-9

719 41. Munakata Y, Yamada T, Imai J, Takahashi K, Tsukita S, Shirai Y, et al. Olfactory receptors
720 are expressed in pancreatic β-cells and promote glucose-stimulated insulin secretion. Sci
721 Rep. 2018;8. doi:10.1038/s41598-018-19765-5

722 42. Robertson RP. Prostaglandins, glucose homeostasis, and diabetes mellitus. Annual review
723 of medicine. Annu Rev Med; 1983. pp. 1–12. doi:10.1146/annurev.me.34.020183.000245

724 43. Attia HRM, Kamel SA, Ibrahim MH, Farouk HA, Rahman AHA, Sayed GH, et al. Open-
725 array analysis of genetic variants in Egyptian patients with type 2 diabetes and obesity.
726 Egypt J Med Hum Genet. 2017;18: 341–348. doi:10.1016/j.ejmhg.2017.03.002

727 44. Tang H, Wei P, Duell EJ, Risch HA, Olson SH, Bueno-De-Mesquita HB, et al. Genes-
728 environment interactions in obesity- and diabetes-associated pancreatic cancer: A GWAS
729 data analysis. Cancer Epidemiol Biomarkers Prev. 2014;23: 98–106. doi:10.1158/1055-
730 9965.EPI-13-0437-T

731 45. Balmer L, O'leary CA, Menotti-Raymond M, David V, O'brien S, Penglis B, et al.
732 Mapping of diabetes susceptibility LOCI in a domestic cat breed with an unusually high
733 incidence of diabetes mellitus. Genes. 2020;11: 1–11. doi:10.3390/genes11111369

734 46. Forcada Y, Holder A, Church DB, Catchpole B. A Polymorphism in the melanocortin 4
735 receptor gene (MC4R: C.92C>T) is associated with diabetes mellitus in overweight
736 domestic shorthaired cats. J Vet Intern Med. 2014;28: 458–464. doi:10.1111/jvim.12275

737 47. Holgate ST. Epithelial damage and response. Clinical and Experimental Allergy,
738 Supplement. Clin Exp Allergy; 2000. pp. 37–41. doi:10.1046/j.1365-2222.2000.00095.x

739 48. Yang L, Cohn L, Zhang DH, Homer R, Ray A, Ray P. Essential role of nuclear factor κB in
740 the induction of eosinophilia in allergic airway inflammation. J Exp Med. 1998;188: 1739–
741 1750. doi:10.1084/jem.188.9.1739

742 49. Collins PD, Marleau S, Griffiths-Johnson DA, Jose PJ, Williams TJ. Cooperation between
743 interleukin-5 and the chemokine eotaxin to induce eosinophil accumulation in vivo. J Exp
744 Med. 1995;182: 1169–1174. doi:10.1084/jem.182.4.1169

745 50. Das J, Chen CH, Yang L, Cohn L, Ray P, Ray A. A critical role for NF-κB in Gata3
746 expression and TH2 differentiation in allergic airway inflammation. Nat Immunol. 2001;2:
747 45–50. doi:10.1038/83158

748 51. Payne JR, Brodbelt DC, Luis Fuentes V. Cardiomyopathy prevalence in 780 apparently
749 healthy cats in rehoming centres (the CatScan study). J Vet Cardiol. 2015;17: S244–S257.
750 doi:10.1016/j.jvc.2015.03.008

751 52. Gil-Ortuño C, Sebastián-Marcos P, Sabater-Molina M, Nicolas-Rocamora E, Gimeno-
752 Blanes JR, Fernández del Palacio MJ. Genetics of feline hypertrophic cardiomyopathy.
753 Clinical Genetics. Blackwell Publishing Ltd; 2020. pp. 203–214. doi:10.1111/cge.13743

754 53. Côté E, MacDonald KA, Meurs KM, Sleeper MM. Hypertrophic Cardiomyopathy. In:
755 Feline cardiology. Chichester, UK: John Wiley & Sons; 2011. pp. 103–75.

756 54. Peterson ME. Feline hyperthyroidism: An animal model for toxic nodular goiter. Journal of
757 Endocrinology. BioScientifica Ltd.; 2014. pp. T97–T114. doi:10.1530/JOE-14-0461

758 55. O'Neill DG, Gostelow R, Orme C, Church DB, Niessen SJM, Verheyen K, et al.
759 Epidemiology of Diabetes Mellitus among 193,435 Cats Attending Primary-Care
760 Veterinary Practices in England. J Vet Intern Med. 2016;30: 964–972.
761 doi:10.1111/jvim.14365

762 56. Finch NC, Syme HM, Elliott J. Risk Factors for Development of Chronic Kidney Disease in
763 Cats. J Vet Intern Med. 2016;30: 602–610. doi:10.1111/jvim.13917

764 57. Jergens AE. Feline Idiopathic Inflammatory Bowel Disease: What we know and what
765 remains to be unraveled. J Feline Med Surg. 2012;14: 445–458.
766 doi:10.1177/1098612X12451548

767 58. Paulin M V., Couronné L, Beguin J, Le Poder S, Delverdier M, Semin MO, et al. Feline
768 low-grade alimentary lymphoma: An emerging entity and a potential animal model for
769 human disease. BMC Veterinary Research. BioMed Central Ltd.; 2018.
770 doi:10.1186/s12917-018-1635-5

59. Marsilio S. Feline chronic enteropathy. J Small Anim Pract. 2021;62: 409–419. doi:10.1111/jsap.13332

60. Moore PF, Woo JC, Vernau W, Kosten S, Graham PS. Characterization of feline T cell receptor gamma (TCRG) variable region genes for the molecular diagnosis of feline intestinal T cell lymphoma. Vet Immunol Immunopathol. 2005;106: 167–178. doi:10.1016/j.vetimm.2005.02.014

61. Lucyshyn DR, Good KL, Knickelbein KE, Chang MW, Strøm AR, Hollingsworth SR, et al. Subcutaneous administration of triamcinolone as part of the management of feline eosinophilic keratoconjunctivitis. J Feline Med Surg. 2020. doi:10.1177/1098612X20968660

62. Midkiff AM, Chew DJ, Randolph JF, Center SA, DiBartola SP. Idiopathic Hypercalcemia in Cats. J Vet Intern Med. 2000;14: 619. doi:10.1892/0891-6640(2000)014<0619:ihic>2.3.co;2

63. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. Curr Protoc Bioinforma. 2013;43. doi:10.1002/0471250953.bi1110s43

64. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25: 1754–1760. doi:10.1093/bioinformatics/btp324

65. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27: 2156–2158. doi:10.1093/bioinformatics/btr330

66. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012;6: 80–92. doi:10.4161/fly.19695

67. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: Rising to the challenge of larger and richer datasets. GigaScience. 2015;4. doi:10.1186/s13742-015-0047-8

68. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81: 559–575. doi:10.1086/519795

69. Buckley RM, Davis BW, Brashear WA, Farias FHG, Kuroki K, Graves T, et al. A new domestic cat genome assembly based on long sequence reads empowers feline genomic medicine and identifies a novel gene for dwarfism. PLoS Genet. 2020;16. doi:10.1371/journal.pgen.1008926

70. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38: 904–909. doi:10.1038/ng1847

71. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012;44: 821–824. doi:10.1038/ng.2310

811 72. Liu X, Huang M, Fan B, Buckler ES, Zhang Z. Iterative Usage of Fixed and Random Effect
812 Models for Powerful and Efficient Genome-Wide Association Studies. PLoS Genet.
813 2016;12: e1005767. doi:10.1371/journal.pgen.1005767

814 73. Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, et al. rMVP: A Memory-efficient,
815 Visualization-enhanced, and Parallel-accelerated tool for Genome-Wide Association Study.
816 Genomics Proteomics Bioinformatics. 2021; S1672-0229(21)00050–4.
817 doi:10.1016/j.gpb.2020.10.007

818 74. D. Turner S. qqman: an R package for visualizing GWAS results using Q-Q and manhattan
819 plots. J Open Source Softw. 2018;3: 731. doi:10.21105/joss.00731

820 75. R Core Team. R: A language and environment for statistical computing. R Found Stat
821 Comput. 2020.

822 76. Hunter JD. Matplotlib: A 2D graphics environment. Comput Sci Eng. 2007;9: 90–95.
823 doi:10.1109/MCSE.2007.55

824 77. Pérez F, Granger BE. IPython: A system for interactive scientific computing. Comput Sci
825 Eng. 2007;9: 21–29. doi:10.1109/MCSE.2007.53

826

Supporting information

S1 Fig. **LD plots for the GWAS positive controls, using the LMM GWAS results.** (A) *Orange* locus GWAS, showing the significant association on chromosome X. (B) Factor XII deficiency GWAS, showing the A1 significant association in the vicinity of the gene *F12*. Colors indicate the amount of LD (r^2) with the most significant SNP, ranging from black ($r^2<0.2$) to red ($r^2>0.8$).

S2 Fig. Manhattan, QQ and LD plots for IBD using the FarmCPU GWAS results. On the Manhattan plot, the red dashed line is the Bonferroni-corrected significance threshold. Inflation factor (**λ**) is shown on the QQ plot.

S3 Fig. **Manhattan and QQ plots for diseases that did not reach genome-wide significance in the LMM GWAS.** On Manhattan plots, the red line is Bonferroni-corrected significance threshold, and the blue line is Bonferroni-corrected significance threshold calculated using unlinked SNPs. Inflation factors (**λ**) are shown on QQ plots.

S4 Fig. Age distribution of cases and controls. (A) HCM, (B) hyperthyroidism, (C) DM, (D) CKD, (E) chronic enteropathy, (F) IBD, (G) SCAL, (H) all GI, (I) FEK, (J) hypercalcemia. Age (in months) is shown on the X axis and number of cats is shown on the Y axis. Cases are the pink bars and controls are the blue bars.

S5 Fig. Weight distribution of cases and controls for DM. Weight (in kg) is shown on the X axis and number of cats is shown on the Y axis. Cases are the pink bars and controls are the blue

bars. Note that weight data was only available for 53 of the 67 cases, and 339 of the 366 controls used in the GWAS.

S1 Table. Significant association LMM GWAS results on chromosome X for the presence of orange fur in random bred cats.

S2 Table. Three most significant associations for each disease from the FarmCPU GWAS results.

S3 Table. Frequencies, odds ratios, and *P*-values for SNPs associated with complex diseases in different breeds and random bred cats.

S4 Table. Non-significant associations identified for complex diseases using a LMM GWAS design.

S5 Table. Numbers of individual cats from each breed and random bred group for each phenotype.

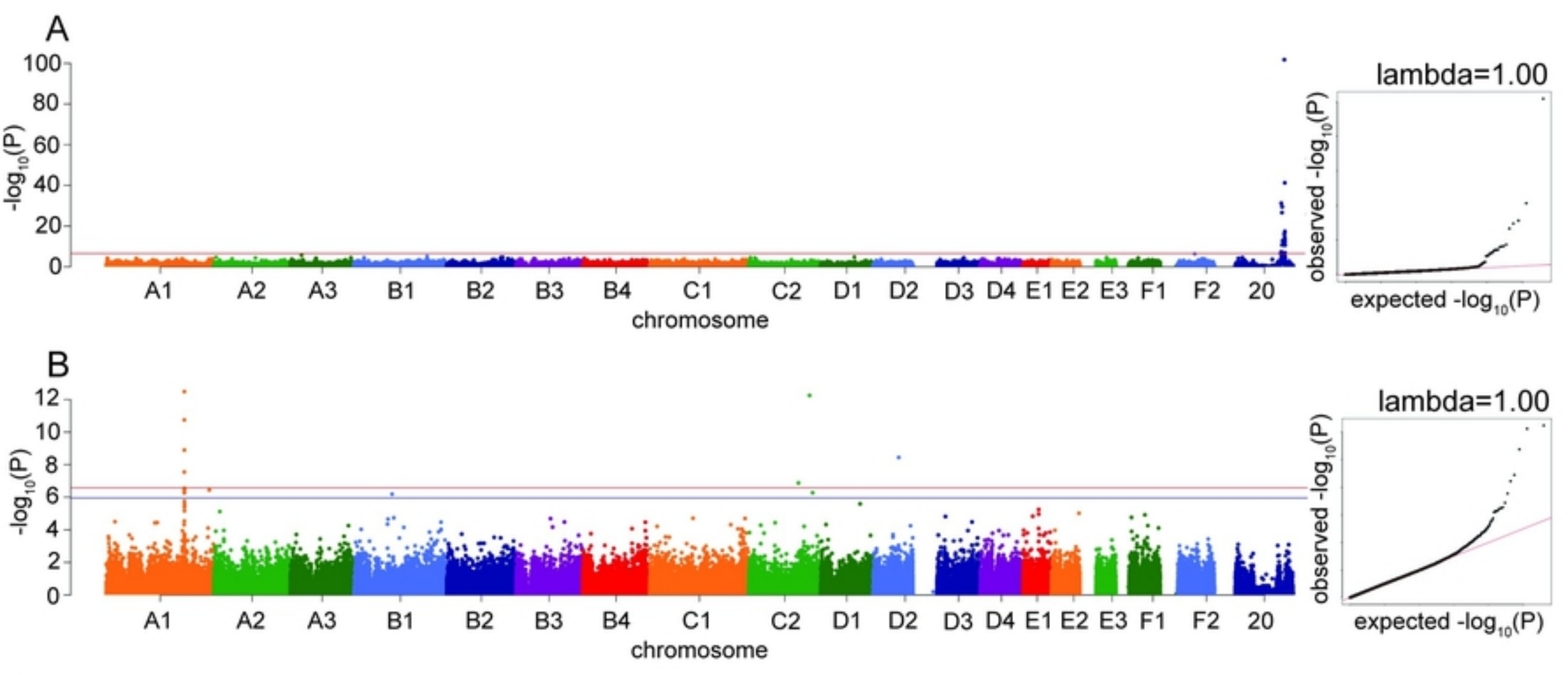

A
100
80
60
40
20
0
-log10(P)
A1
A2
A3
B1
B2
B3
B4
C1
C2
D1
D2
D3
D4
E1
E2
E3
F1
F2
20
chromosome
lambda=1.00
observed -log10(P)
expected -log10(P)
B
12
10
8
6
4
2
0
-log10(P)
A1
A2
A3
B1
B2
B3
B4
C1
C2
D1
D2
D3
D4
E1
E2
E3
F1
F2
20
chromosome
lambda=1.00
observed -log10(P)
expected -log10(P)

Figure 2

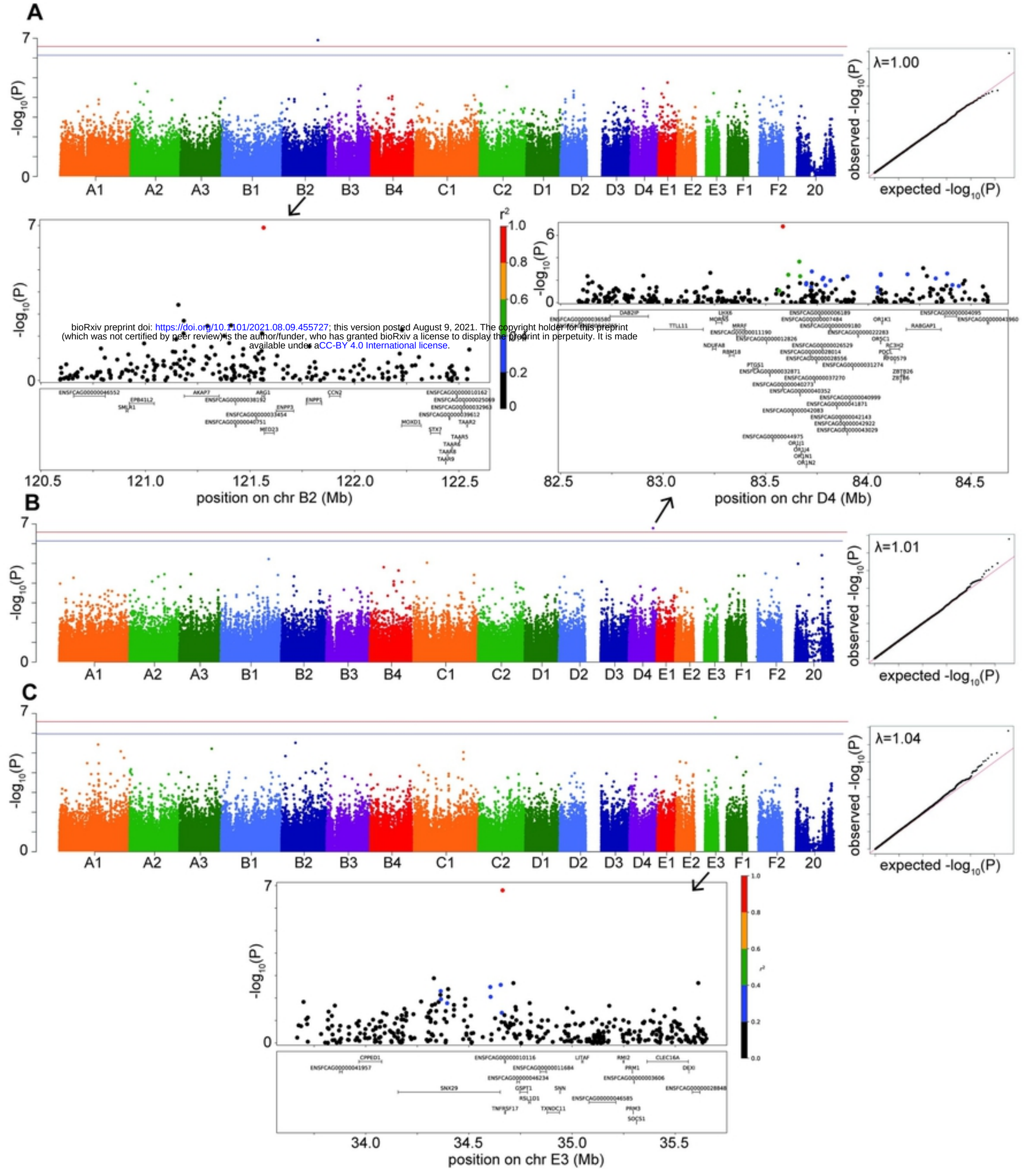


Figure 3

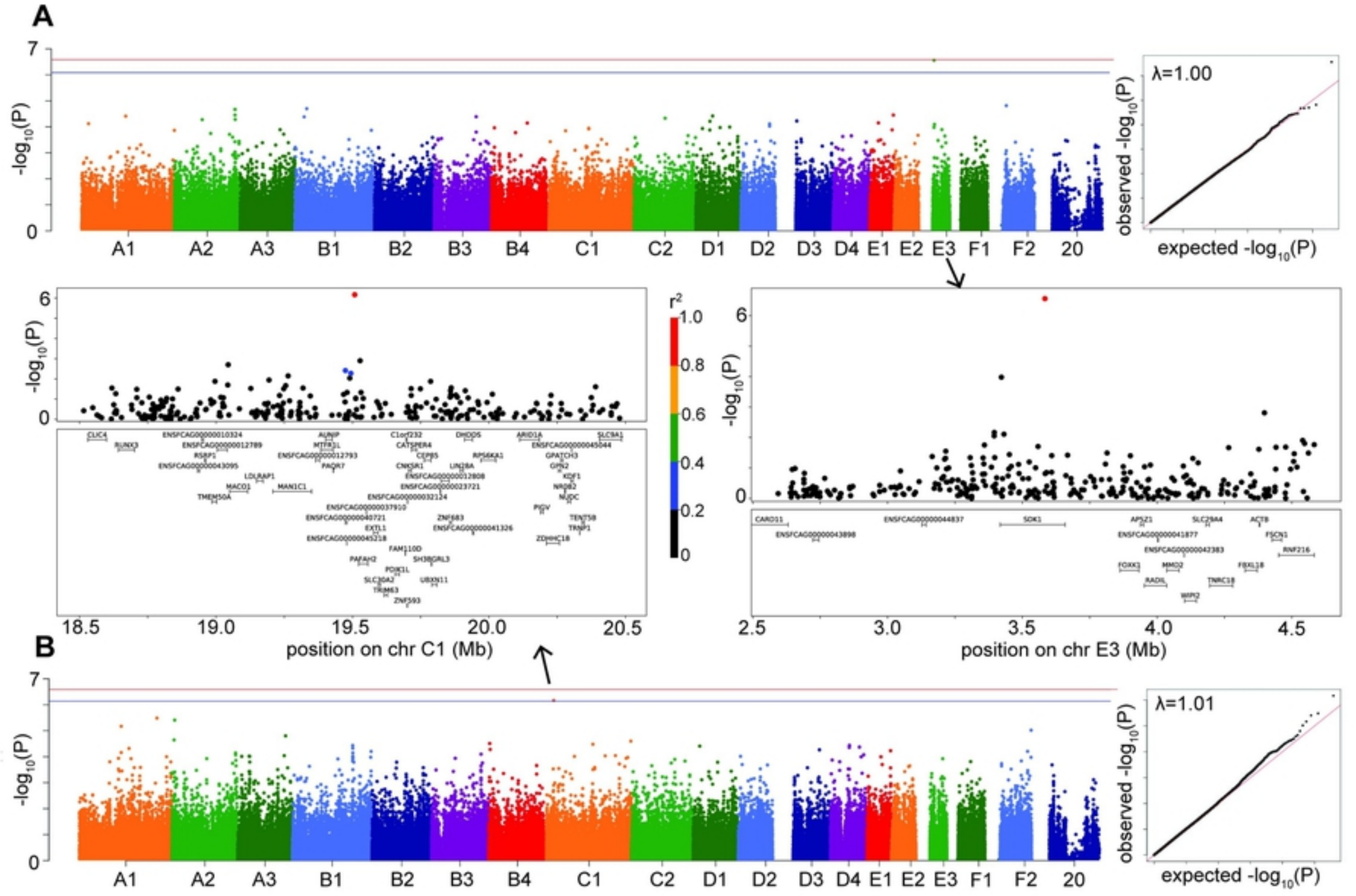
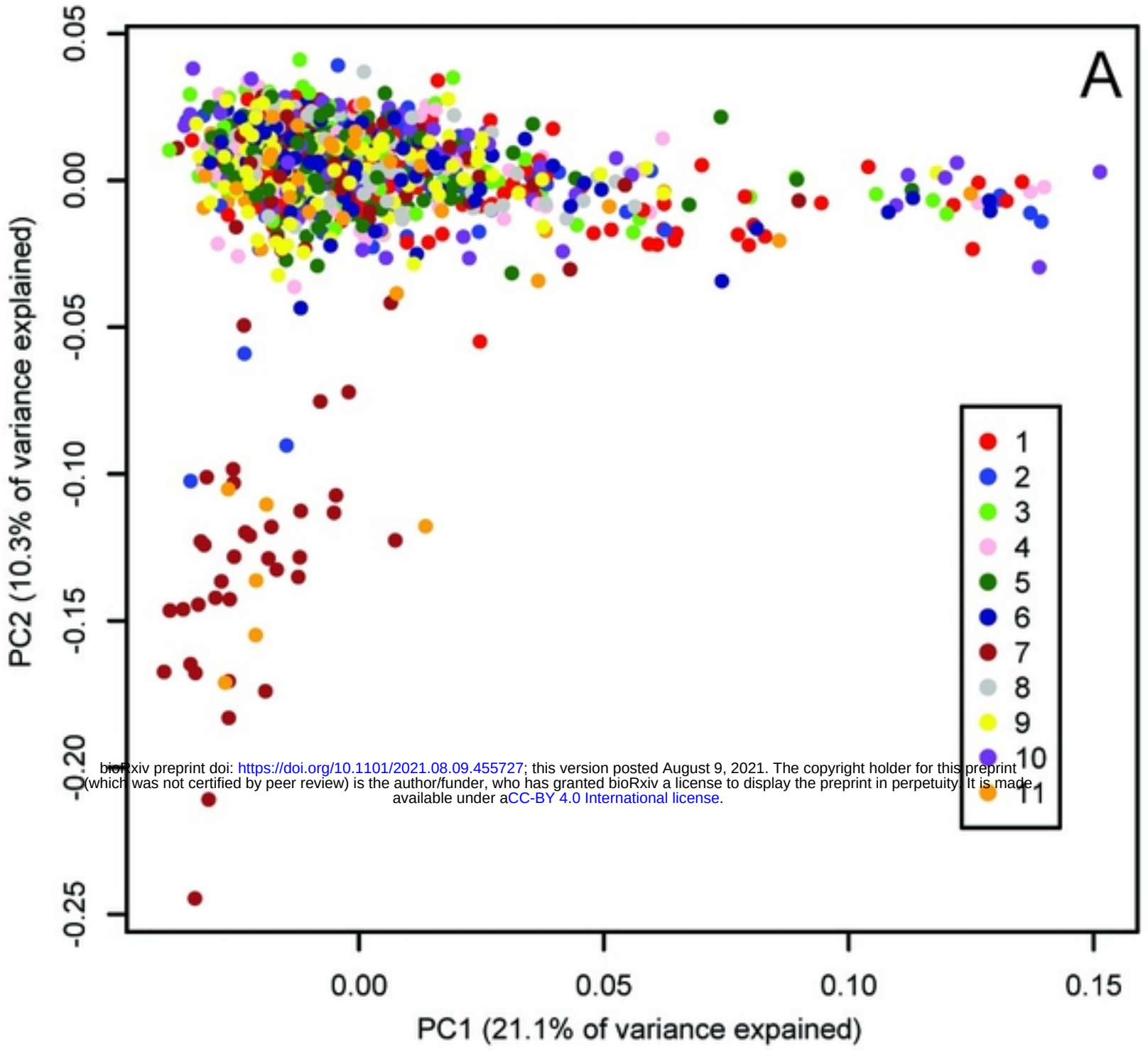

A
B
-log10(P)
A1
A2
A3
B1
B2
B3
B4
C1
C2
D1
D2
D3
D4
E1
E2
E3
F1
F2
20
λ=1.00
λ=1.01
observed -log10(P)
expected -log10(P)
r2
1.0
0.8
0.6
0.4
0.2
0
18.5
19.0
19.5
20.0
20.5
position on chr C1 (Mb)
2.5
3.0
3.5
4.0
4.5
position on chr E3 (Mb)

Figure 4



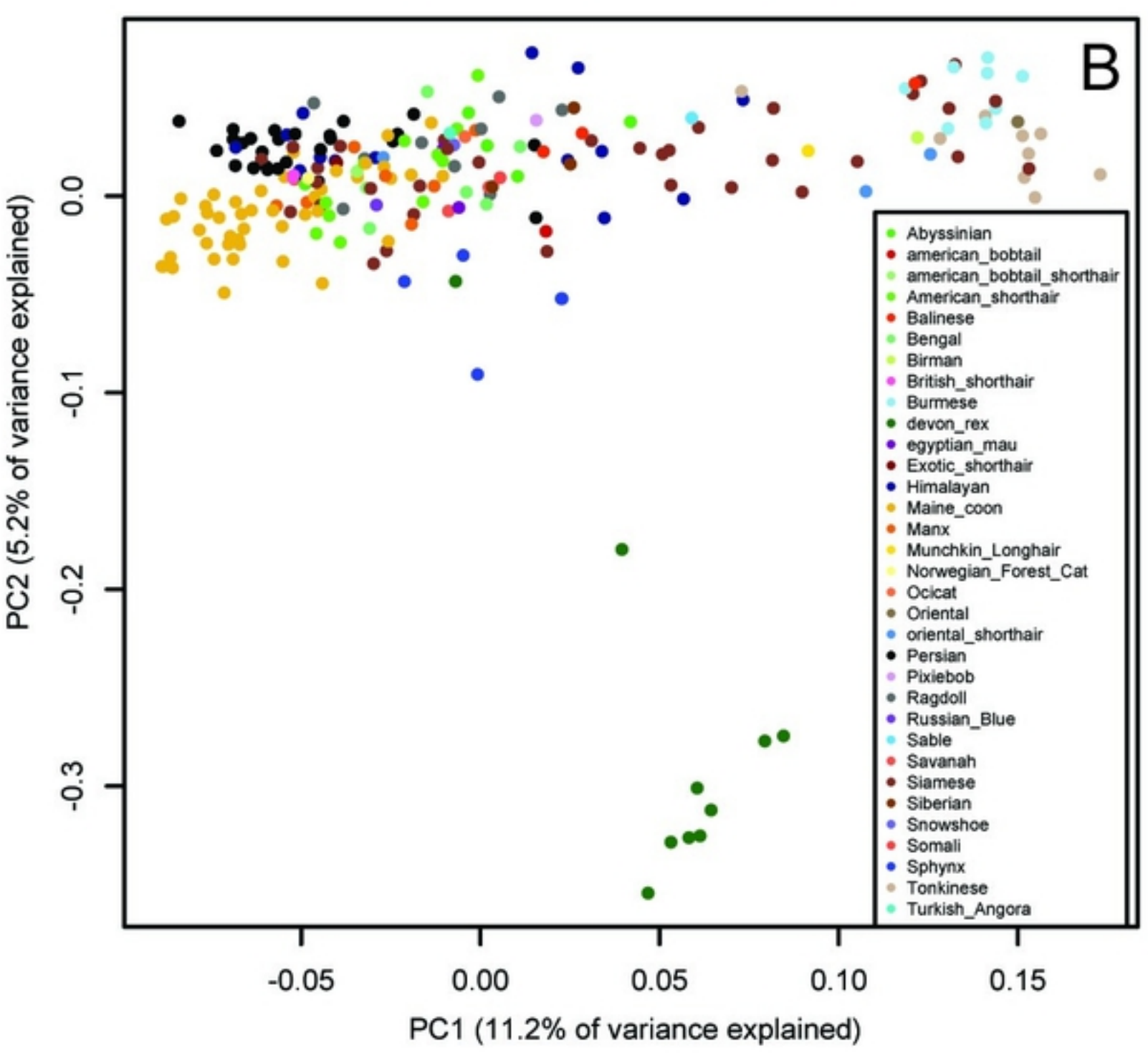


Figure 1