

1 **A new domestic cat genome assembly based on long sequence reads empowers feline**
2 **genomic medicine and identifies a novel gene for dwarfism.**

3 Reuben M. Buckley¹, Brian W. Davis², Wesley A. Brashear², Fabiana H. G. Farias³, Kei Kuroki⁴,
4 Tina Graves³, LaDeana W. Hillier³, Milinn Kremitzki³, Gang Li², Rondo Middleton⁵, Patrick Minx³,
5 Chad Tomlinson³, Leslie A. Lyons¹, William J. Murphy², and Wesley C. Warren^{5,6}

6 ¹Department of Veterinary Medicine and Surgery, College of Veterinary Medicine, University of
7 Missouri, Columbia, Missouri 65211, USA.

8 ²Department of Veterinary Integrative Biosciences, Interdisciplinary Program in Genetics,
9 College of Veterinary Medicine, Texas A&M University, College Station, Texas 77843, USA

10 ³McDonnell Genome Institute, Washington University, School of Medicine, St Louis, Missouri
11 63108, USA;

12 ⁴Veterinary Medical Diagnostic Laboratory, College of Veterinary Medicine, University of
13 Missouri, Columbia, Missouri, 65211, USA

14 ⁵Nestlé Purina Research, Saint Louis, Missouri 63164, USA

15 ⁶Division of Animal Sciences, School of Medicine, University of Missouri, Columbia, Missouri
16 65211, USA

17 Corresponding author: warrenwc@missouri.edu

18 Keywords: animal model, chondrodysplasia, dwarfism, *Felis catus*, long-read, Precision
19 Medicine

20

21

1 **Abstract**

2 The domestic cat (*Felis catus*) numbers over 94 million in the USA alone, occupies households
3 as a companion animal, and, like humans, suffers from cancer and common and rare diseases.
4 However, genome-wide sequence variant information is limited for this species. To empower
5 trait analyses, a new cat genome reference assembly was developed from PacBio long
6 sequence reads that significantly improve sequence representation and assembly contiguity.
7 The whole genome sequences of 54 domestic cats were aligned to the reference to identify
8 single nucleotide variants (SNVs) and structural variants (SVs). Across all cats, 16 SNVs
9 predicted to have deleterious impacts and in a singleton state were identified as high priority
10 candidates for causative mutations. One candidate was a stop gain in the tumor suppressor
11 *FBXW7*. The SNV is found in cats segregating for feline mediastinal lymphoma and is a
12 candidate for inherited cancer susceptibility. SV analysis revealed a complex deletion coupled
13 with a nearby potential duplication event that was shared privately across three unrelated
14 dwarfism cats and is found within a known dwarfism associated region on cat chromosome B1.
15 This SV interrupted *UDP-glucose 6-dehydrogenase (UGDH)*, a gene involved in the
16 biosynthesis of glycosaminoglycans. Importantly, *UGDH* has not yet been associated with
17 human dwarfism and should be screened in undiagnosed patients. The new high-quality cat
18 genome reference and the compilation of sequence variation demonstrate the importance of
19 these resources when searching for disease causative alleles in the domestic cat and for
20 identification of feline biomedical models.

1 **Author summary**

2 The practice of genomic medicine is predicated on the availability of a high quality reference
3 genome and an understanding of the impact of genome variation. Such resources have lead to
4 countless discoveries in humans, however by working exclusively within the framework of
5 human genetics, our potential for understanding diseases biology is limited, as similar analyses
6 in other species have often lead to novel insights. The generation of *Felis_catus_9.0*, a new
7 high quality reference genome for the domestic cat, helps facilitate the expansion of genomic
8 medicine into the *felis* lineage. Using *Felis_catus_9.0* we analyze the landscape of genomic
9 variation from a collection of 54 cats within the context of human gene constraint. The
10 distribution of variant impacts in cats is correlated with patterns of gene constraint in humans,
11 indicating the utility of this reference for identifying novel mutations that cause phenotypes
12 relevant to human and cat health. Moreover, structural variant analysis revealed a novel variant
13 for feline dwarfism in *UGDH*, a gene that has not been associated with dwarfism in any other
14 species, suggesting a role for *UGDH* in cases of undiagnosed dwarfism in humans.

1 **Background**

2 In the veterinary clinic, the practice of genomic medicine is impending [1]. With actionable
3 genetic information in hand, companion animal therapeutic interventions are feasible, including
4 treatment of animal patients prior to, or to prevent the appearance of, more severe symptoms
5 and allow therapeutic administration of drugs with higher efficacy and fewer side effects.
6 Genomic information can also alert veterinarians to imminent disease risks for diagnostic
7 consideration. Each of these applications could significantly enhance veterinary medicine,
8 however, none are currently in practice. As in human medicine, formidable challenges exist for
9 the implementation of genomic-based medicine, including accurate annotation of the genome
10 and databases of genetic variation from well-phenotyped individuals that include both single
11 nucleotide variant (SNV) and structural variant (SV) discovery and annotation [2-4]. Targeted
12 individual companion animal genome information is becoming more readily available, cost
13 effective, and tentatively linked to the actionable phenotypes via direct-to-consumer DNA testing.
14 Thus correct interpretation of DNA variants is of the utmost importance for communicating
15 findings to clinicians practicing companion animal genomic medicine [5].

16 Companion animals suffer from many of the same diseases as humans, with over 600 different
17 phenotypes identified as comparative models for human physiology, biology, development, and
18 disease [6, 7]. In domestic cats, at least 70 genes are shown to harbor single and multiple DNA
19 variants that are associated with disease [8] with many more discoveries expected as health
20 care improves.. The genetic and clinical manifestations of most of these known variants are
21 described in the Online Mendelian Inheritance in Animals [6]. Examples include common human
22 diseases such as cardiomyopathy [9], retinal degenerations [10], and polycystic kidney disease
23 [11]. Veterinarians, geneticists and other researchers are actively banking DNA from companion
24 animals and attempting to implement genomic medicine [1]. Once coupled with the quickly
25 advancing sequencing technology and exploitable results, genomic medicine in companion

1 animals promises to expand the comparative knowledge of mechanisms of action across
2 species. Despite the continuing successful discovery of feline disease variants using both
3 candidate gene and whole genome sequencing (WGS) approaches [1, 12-15], the
4 understanding of normal and disease sequence variation in the domestic cat and interrogation
5 of gene structure and function is limited by an incomplete genome assembly.

6 A fundamental hurdle hampering the interpretation of feline disease variant data is the
7 availability of a high-quality, gapless reference genome. The previous domestic cat reference,
8 *Felis_catus_8.0*, contains over 300,000 gaps, compromising its utility for identifying all types of
9 sequence variation [16], in particular SVs. In conjunction with various mapping technologies,
10 such as optical resolved physical maps, recent advances in the use of long-read sequencing
11 and assembly technology has produced a more complete genome representation (i.e., fewer
12 gaps) for many species [17-20].

13 Another hurdle for performing feline genomic medicine is the availability of WGS data from
14 various breeds of cat with sufficient sequencing depth to uncover rare alleles and complex
15 structural variants. Knowledge of variant frequency and uniqueness among domestic cats is
16 very limited and is crucial in the identification of causal alleles. As a result of the paucity of
17 sequence variant data across breeds, the 99 Lives Cat Genome Sequencing Initiative was
18 founded as a centralized resource with genome sequences produced of similar quality and
19 techniques. The resource supports researchers with variant discovery for evolutionary studies
20 and identifying the genetic origin of inherited diseases and can assist in the development of
21 high-density DNA arrays for complex disease studies in domestic cats [1, 21-24].

22 Here we present a new version of the domestic cat genome reference (Cinnamon, an
23 Abyssinian breed), generated from deep sequence coverage of long-reads and scaffolding from
24 an optical map (BioNano) and a high-density genetic linkage map [16]. Published cat genomes
25 from the 99 Lives Cat Genome Consortium [1, 23] were aligned to the *Felis_catus_9.0*

1 reference to discover a plethora of unknown SNVs and SVs (multi-base insertions and
2 deletions), including a newly identified structural variant (SV) for feline disproportionate
3 dwarfism. Our case study of dwarfism demonstrates when disease phenotypes are coupled with
4 revised gene annotation and sequence variation ascertained from diverse breeds, the new cat
5 genome assembly is a powerful resource for trait discovery. This enables the future practice of
6 feline genomic medicine and improved ascertainment of biomedical models relevant to human
7 health.

8 **Results**

9 **Genome assembly.** A female Abyssinian cat (Cinnamon) was sequenced to high-depth (72-
10 fold coverage) using real-time (SMRT; PacBio) sequence data and all sequence reads were
11 used to generate a *de novo* assembly. Two PacBio instruments were used to produce average
12 read insert lengths of 12 kb (RSII) and 9 kb (Sequel). The ungapped assembly size was 2.48
13 Gb and is comparable in size to other assembled carnivores (**Table 1**). There were 4,909 total
14 contigs compared to 367,672 contigs in *Felis_catus_8.0* showing a significant reduction in
15 sequence gaps. The assembly contiguity metric of N50 contig and scaffold lengths were 42 and
16 84 Mb, respectively (**Table 1**). The N50 contig length of other PacBio sequenced carnivore
17 assemblies are less contiguous, ranging from 3.13 Mb to 20.91 Mb (**Table 1**). Across carnivores,
18 RepeatMasker showed consistent measures of total interspersed repeat content, (with 43% in
19 *Felis_catus_9.0*; **S1 Table**) [25]. Due to repetitive and other genome architecture features, 1.8%
20 (46 Mb) of all assembled sequences remained unassigned to any chromosome position. These
21 sequences had an N50 scaffold length of 12,618 bp, demonstrating the assembly challenge of
22 some repeat types in diploid genome assemblies, even of an inbred individual.

1 **Table 1. Representative assembly metrics for various chromosome level assembled carnivore genomes¹.**

Assembly	Species	Breed	Isolate	Release date (MM/DD/YY)	Sequencing technology	Genome coverage	Total ungapped length (Gb)	Scaffold N50 (Mb)	Contig N50 (Mb)	Unplaced length (Mb)
Felis_catus_9.0	<i>Felis catus</i> (domestic cat)	Abyssinian	Cinnamon	11/20/17	PacBio; 454 Titanium; Illumina; Sanger dideoxy sequencing	72x	2.48	83.97	41.92	46.02
Felis_catus_8.0	<i>Felis catus</i> (domestic cat)	Abyssinian	Cinnamon	11/07/14	Sanger; 454 Titanium; Illumina	2x Sanger; 14x 454, 20x Illumina	2.60	18.07	0.05	73.71
mLynCan4_v1.p	<i>Lynx canadensis</i> (Canada lynx)	NA	LIC74	07/26/19	PacBio Sequel I; 10X genome; Bionano Genomics; Arima Genomics Hi-C	72x	2.41	146.11	7.50	6.18
PanLeo1.0	<i>Panthera leo</i> (lion)	NA	Brooke	10/07/19	Illumina; Oxford Nanopore; 10X Genomics	46x	2.39	136.05	0.29	242.29
ASM864105v1	<i>Canis lupus familiaris</i> (dog)	German Shepherd	Nala	09/25/19	PacBio Sequel; Oxford Nanopore PromethION; Illumina (10X Chromium)	30x	2.40	64.35	20.91	22.83
ASM488618v2	<i>Canis lupus familiaris</i> (dog)	Basenji	MU ID 185726	08/16/19	Sequel	45x	2.41	61.09	3.13	120.80
UMICH_Zoey_3.1	<i>Canis lupus familiaris</i> (dog)	Great Dane	Zoey	05/30/19	PacBio RSII	50x	2.34	64.20	4.72	16.82
CanFam3.1	<i>Canis lupus familiaris</i> (dog)	boxer	Tasha	11/02/11	Sanger	7x plus >90Mb finished sequence	2.39	45.88	0.27	75.10

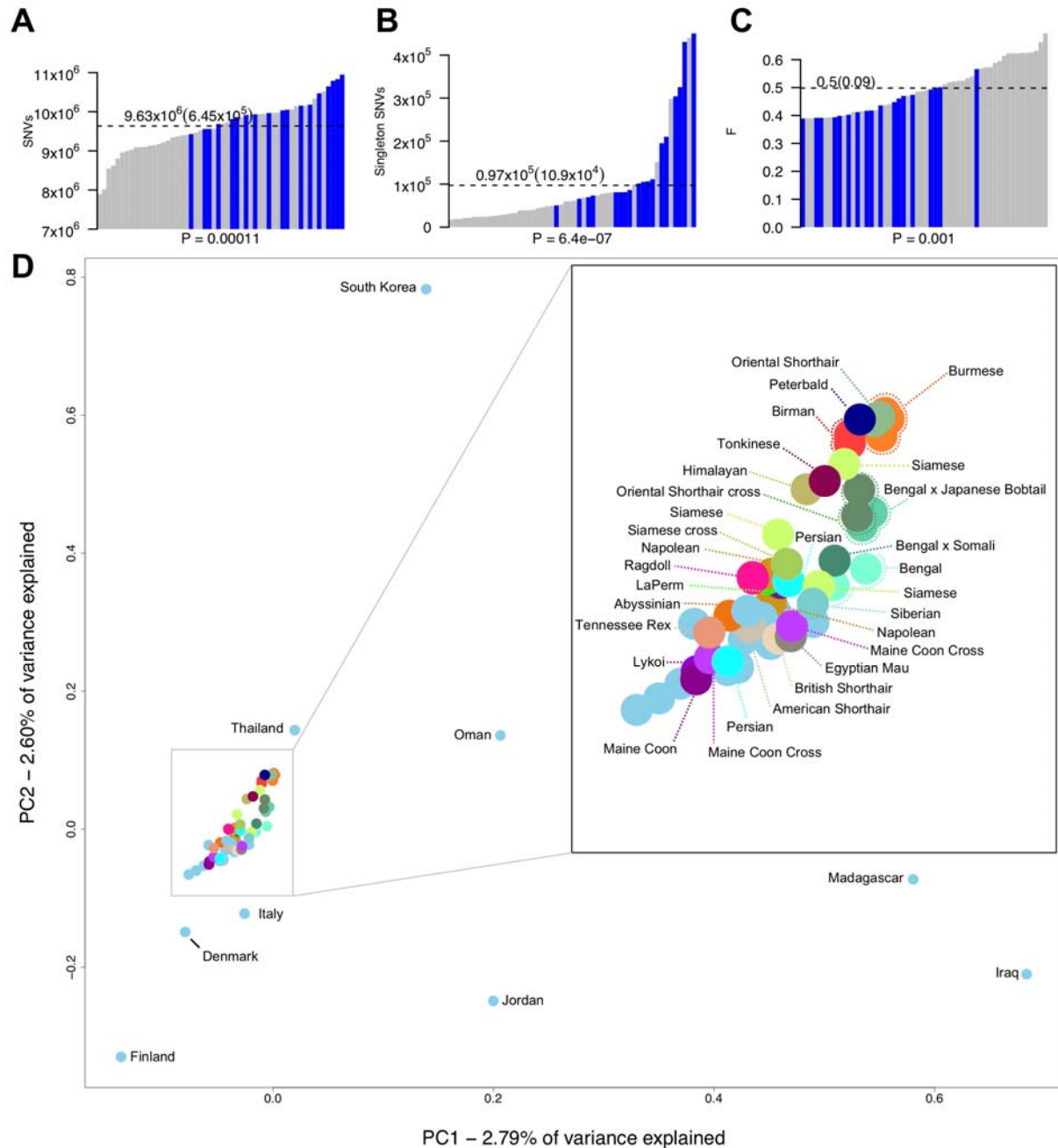
2 ¹All species-specific assembly metrics derived from the NCBI assembly archive.

1 **Sequence accuracy and quality assessment.** Illumina whole-genome sequence data from
2 Cinnamon was used to identify reference sequence errors as homozygous SNVs. These
3 numbered 60,449 in total, indicating a high level of sequence accuracy across assembled
4 contigs (>99.9%). Sequence order and orientation was also highly accurate (>98%), as only
5 1.2% of BAC-end sequence alignments derived from Cinnamon were identified as discordant.
6 Felis_catus_9.0 sequence order and orientation was also supported by high levels of agreement
7 between individual chromosome sequence alignment and ordered sequence markers from the
8 published cat genetic linkage map [16] (**S1 Data**). The raw sequence data, assembled contigs,
9 and sequence order coordinates (AGP) were accessioned and are fully available by searching
10 GCF_000181335.3 in GenBank.

11 **Gene annotation.** The number of annotated protein-coding genes was very similar between the
12 NCBI and Ensembl pipelines at 19,748 and 19,409, respectively. Approximately 376 protein-
13 coding genes (NCBI) were identified as novel with no matching annotations in Felis_catus_8.0
14 (**S2 Data**). Conversely, 178 genes from Felis_catus_8.0 did not map to Felis_catus_9.0, of
15 which the cause is unknown (**S1 Table**). A large portion of genes changed substantially (8.4%)
16 in Felis_catus_9.0 during NCBI annotation (**S2 Data**). Aligned sequence of known same-
17 species RefSeq transcripts (n = 420) to Felis_catus_9.0 is higher (99.5%) than Felis_catus_8.0
18 (97.8%) and the mean coverage of these same translated RefSeq proteins is also improved
19 (90.1% versus 88.3% in Felis_catus_8.0). One important consequence of the less fragmented
20 gene annotation is a 2% increase in aggregate sequence alignments of feline RNA-seq datasets
21 to Felis_catus_9.0. These improvements are largely attributed to fewer assembly gaps. The
22 various reported metrics of gene annotation quality conclusively show the protein-coding genes
23 of the domestic cat are of high quality for all trait discovery studies.

24 **Genetic variation in cats.** To improve variant knowledge of the domestic cat, variants from a
25 diverse set of 74 resequenced cats from the 99 Lives project were analyzed in depth. The

1 average sequence coverage was 38.5x with a mean of 98% reads mapped per cat (**S3 Data**).
2 Approximately 46,600,527 variants were discovered, 39,043,080 were SNVs with 93% as
3 biallelic displaying a Ts/Tv ratio of 2.44, suggesting a relatively high level of specificity for
4 variant detection (**S2 Table**). In addition, 97% of SNV positions from the feline 63K genotyping
5 array mapped to *Felis_catus_9.0* were detected as SNVs in the WGS call set (**S3 Table; S4**
6 **Data**) [26]. Using the variant data to estimate cat relatedness, 13 highly related cats ($r > 0.15$),
7 two cats with poor read quality, four bengals or bengal crosses, and Cinnamon, the reference,
8 were removed from the sequence dataset to obtain a final set of 54 cats for all subsequent
9 analyses (**S3 Data**). The average number of discovered SNVs per cat was 9.6 million (**Fig. 1a**).
10 Differences in SNV numbers varied according to whether cats were from a specific breed or
11 were random bred (P-value < 0.005, Wilcoxon rank sum test), the two cats with the lowest
12 number of SNVs (~8 million) were both Abyssinians, the same breed as Cinnamon, while
13 random bred cats from either the Middle East or Madagascar each carried the highest number
14 (> 10.5 million) (**S5 Data**). Individual singleton frequency and estimated inbreeding coefficients
15 (*F* statistic) showed a similar trend with random bred cats generally having significantly more
16 singletons and higher levels of heterozygosity than breed cats (P-value < 0.005, Wilcoxon rank
17 sum test) (**Fig. 1b, Fig. 1c**). Breed cats with higher levels of variation and heterozygosity were
18 either from newly established breeds or were outcrossed individuals. For example, Napoleon
19 cats, which all had an *F* statistic at least one standard deviation below the mean (**Fig. 1c; S5**
20 **Data**), are frequently outcrossed as their defining trait dwarfism is likely homozygous lethal *in*
21 *utero* [27].



1
2 **Figure 1. Genetic variation in whole genome sequenced cats.** (a) Number of SNVs found in
3 each genome, (b) singletons per genome, and (c) the per sample inbreeding coefficient F . Blue
4 bars indicate each individual random bred cat and grey bars indicate each individual breed cat.
5 P-values underneath each x axes were calculated using Wilcoxon rank-sum test and were used
6 to compare breed cats to random bred cats. Dotted lines indicate the mean for each statistic,
7 which is printed above along with standard deviation in braces. (d) Population structure of all
8 unrelated cats (including Bengal breeds) estimated using principal components analysis.
9 Random bred cats are colored light blue. Those that were sampled globally for diversity regions
10 are named according to their sampling location. All other cats are named according to their
11 breed or breed cross.

1 PCA analysis showed the expected distribution of genetic relatedness among cats when
2 considering their geographical location and genetic origins (**Fig. 1d**). In general, most random
3 bred cats displayed a scattered distribution consistent with previous studies on cat population
4 diversity and origins [26, 28]. Although tightly clustered, breed cats could also be distinguished
5 according to their populations of origin. The Asian-derived breeds, Siamese, Burmese, Birman,
6 and Oriental shorthairs were at one end of the spectrum, clustering closely with random bred
7 cats from Thailand. Conversely, cats derived from western populations, such as Maine Coons
8 and Persians, were at the opposite end of the spectrum grouping with random bred cats from
9 Northern Europe, such as Denmark and Finland.

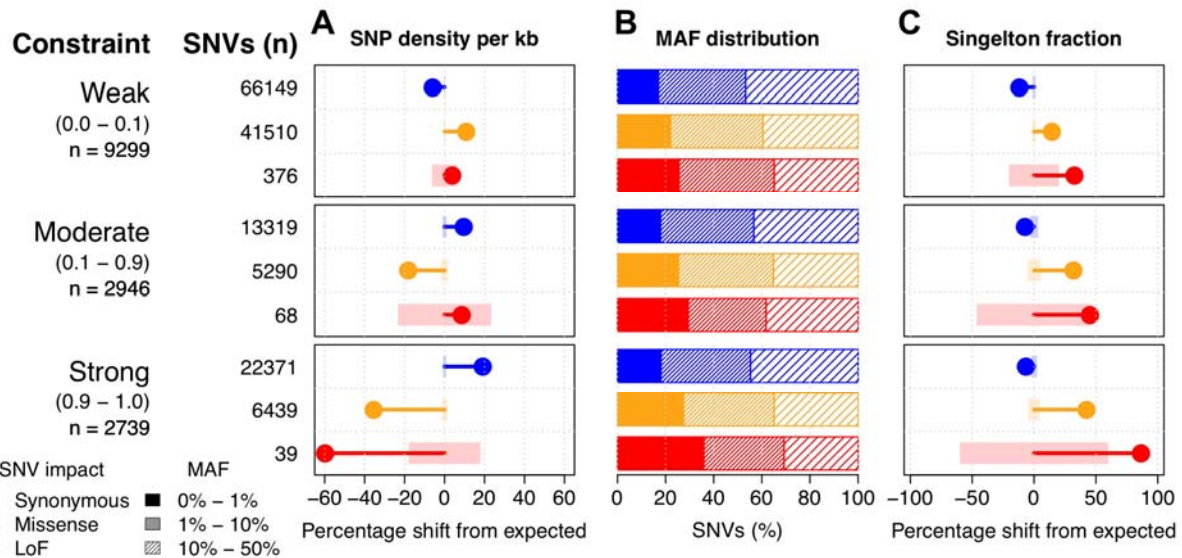
10 **Implications of feline genetic variation on human disease genes.** To characterize feline
11 genetic variation in disease contexts, variant effect predictor (VEP) was used to identify 128,844
12 synonymous, 77,662 missense, and 1,179 loss of function (LoF) SNVs, where SNVs causing a
13 stop gain were the largest contributor to the LoF category (**S4 Table**). In addition to SNP
14 annotation, genes were grouped according to their genetic constraint across human populations,
15 where genetic constraint was expressed as probability of LoF intolerance. In total, 15,962 cat-
16 human orthologs were identified with 14,291 assigned pLI values. Of these, 9,299 were in the
17 weak constraint group ($pLI < 0.1$), 2,946 were in the moderate constraint group ($0.1 < pLI < 0.9$),
18 and 2,739 were in the strong constraint group ($pLI > 0.9$). For genes under weak constraint in
19 humans, feline SNV density within coding sequences, regardless of impact on gene function,
20 was similar to expected SNV densities based on random assignment of SNV impacts.
21 Conversely, the density of SNVs in genes under strong constraint varied significantly according
22 to SNV impact. LoF and missense SNVs, which have potential to deleteriously impact gene
23 function, were depleted by 59.7% and 35.5%, respectively, while synonymous SNVs, which
24 likely have no deleterious impact on gene function, were enriched by 19.2% relative to expected

1 levels (**Fig. 2a**). Similar results, while less pronounced, were also observed for synonymous and
2 missense SNVs in genes under moderate constraint.

3 SNV minor allele frequency (MAF) distributions were compared across each constraint group.
4 MAFs for synonymous SNVs were similarly distributed in all constraint groups. Conversely, the
5 distribution of nonsynonymous SNVs increasingly skewed toward lower MAFs under stronger
6 levels of constraint. For example, 25.5% of LoF SNVs in genes under weak constraint have a
7 MAF < 1%, whereas 35.9% of LoF SNVs in genes under strong constraint have a MAF < 1%
8 (**Fig. 2b**). To determine whether these shifts in MAFs were significant, the fraction of SNVs in a
9 singleton state were compared to expected levels based on random assignment of singleton
10 states. Singleton states of SNVs were significantly enriched for nonsynonymous SNVs with
11 enrichment levels increasing under stronger constraint, indicating many SNVs with functional
12 impacts in genes under constraint are likely rare (**Fig. 2c**). Together, these results show a
13 significant association between selection in human genes and SNV accumulation in cats,
14 suggesting selection pressure within cats and humans is similar across orthologous genes.

15

16



1
2 **Figure 2. Deleterious SNVs in cats are uncommon and depleted from human constrained**
3 **genes.** To the left of the figure panels, constraint groups are labeled with pLI ranges shown
4 below in braces. Below the pLI ranges are the number of genes found in each constraint
5 group. SNV values show the total number of SNVs of a particular impact that belong to each constraint
6 group. **(a)** Observed percentage differences from expected values for per kb SNV density. Light
7 colored rectangles represent the 95% confidence intervals for the expected values. Confidence
8 intervals were calculated using 10,000 permutations (methods). **(b)** The percentage of SNVs of
9 different impacts and constraint groups across various MAF intervals. **(c)** Observed percentage
10 differences from expected values for the fraction of SNVs in a singleton state (allele count of 1).
11 Light colored rectangles represent the 95% confidence intervals for the expected values,
12 generated from 10,000 permutations.

13

14 Overall, 16 LoF singleton SNVs were identified in intolerant orthologs as potential candidate
15 disease causing variants. Since some cats within 99 lives had recorded disease statuses, these
16 SNVs were assessed for their potential role in cat diseases (**S5 Table**). Of the 16 SNVs, four
17 were supported by both Ensembl and NCBI annotations and were in cats segregating for
18 particular diseases (**Table 2**). Of note is a stop gain in the tumor suppressor *F-box and WD*
19 *repeat domain containing 7 (FBXW7)* [29], which was only found in a parent and child
20 segregating for feline mediastinal lymphoma. Other LoF SNVs include stop gains found in
21 *Family With Sequence Similarity 13 Member B (FAM13B)* in a random bred with ectodermal
22 dysplasia, cytoplasmic FMR1 interacting protein 2 (*CYFIP2*) in an Egyptian Mau with urate
23 stones, and *SH3 And PX Domains 2A (SH3PXD2A)* in a random bred cat with feline infectious

1 peritonitis. Most candidates are not likely disease causing, as each cat carried a mean of 10.0
 2 LoF SNVs in strongly constrained genes (**S1 Fig.**). However, while most LoF SNVs had MAFs >
 3 10%, the mean number of LoF SNVs with MAF < 1% in strongly constrained genes was 0.26
 4 per cat (**S2 Fig.**). These results suggest gene intolerance to mutations may provide as a useful
 5 metric for reducing the number of candidate variants for certain diseases.

6 **Table 2. High impact singletons in intolerant orthologs for unrelated cats with disease**
 7 **traits.**

SNV location ^a	Ref/Alt	Consequence	Gene symbol	pLI	Individual ID	Disease	Status
chrA1:116653102	G/A	Stop gained	<i>FAM13B</i>	0.92	felCat.Fcat19194.Pudge	Ectodermal dysplasia ^b	Affected
chrA1:192824838	G/A	Stop gained	<i>CYFIP2</i>	1.00	felCat.Fcat20406.Gannon	Stones	Affected
chrB1:77305060	C/T	Stop gained	<i>FBXW7</i>	1.00	felCat.Fcat5012.Colorado ^c	Lymphoma	Carrier
chrD2:63980875	G/A	Stop gained	<i>SH3PXD2A</i>	1.00	felCat.CR1397.Isabella	Infectious peritonitis ^d	Affected

8 ^a SNV locations were only reported if they were supported by NCBI annotations.

9 ^b A second unrelated cat, felCat.Fcat19197.Kooki, was also affected.

10 ^c Affected offspring removed earlier from analysis inherited the same SNV.

11 ^d A second unrelated cat, felCat.CR1219.Tamborine, was also affected.

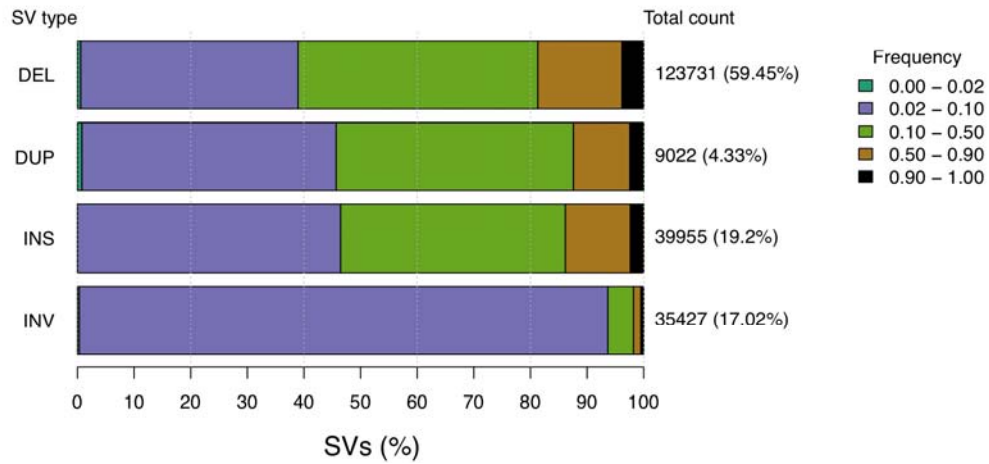
12

13 **Structural variant discovery.** The merging of the two independent SV call sets was performed
 14 across all individuals for variants occurring within 50 bp of the independent variant call position,
 15 with agreement on variant type and strand, and variant size within 500 bp. Per cat, an average
 16 of 44,990 SVs were identified, with variants encompassing 134.3 Mb across all individuals.
 17 Deletions averaged 905 bp, duplications 7,497 bp, insertions 30 bp, and inversions 10,993 bp.
 18 The breed and breed crosses (n = 36) compared to random bred cats (n = 18) showed
 19 comparable SV diversity (t-test p = 0.6) (**S3 Fig.**). In total, 208,135 SVs were discovered, of
 20 which 123,731 (60%) were deletions (**Fig. 3a**). SV population frequencies were similar across
 21 SV types, except for inversions. For deletions, duplications and insertions, 38% to 48% of each
 22 SV type was found at population frequencies of 0.02 – 0.10 and 0.10 – 0.50. Meanwhile, > 90%

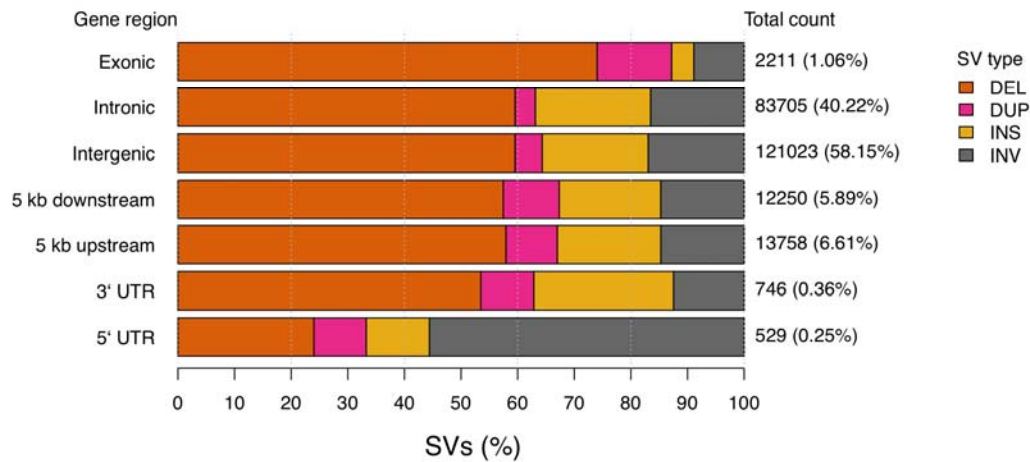
1 of inversions are found at a population frequency of 0.02 – 0.10 (**Fig. 3a**). The majority of SVs
2 identified are common across cats, suggesting their impacts are mostly tolerated.

3 SV density across autosomes was relatively constant with chromosome E1 carrying the largest
4 SV burden at 96.95 SVs per Mb (**S4 Fig.**). Approximately 6,096 SVs (3%, 10.1 Mb) were
5 observed in >90% of the cat genomes (**S6 Data**), indicating the cat used for the *Felis_catus_9.0*
6 assembly, Cinnamon, an Abyssinian, likely carries a minor allele at these positions. SV
7 annotation showed that SV counts per region were consistent with the fraction of the genome
8 occupied by each region type. For example, 58.15% of SVs were intergenic, 40.22% of SVs
9 were intronic, and 1.06% of SVs were exonic, potentially impacting 217 different protein coding
10 genes (**Fig. 3b**) (**S7 Data**). Conversely, the proportion of some SV types found in certain gene
11 regions varied from their genome-wide averages. For example, in regions 5 kb upstream and
12 downstream of genes, duplications were increased approximately two-fold. For exonic regions,
13 74% of SVs were deletions, an increase from the genome wide level of 59.45%. For 5' UTRs,
14 the majority of SVs were inversions, which only represent 17.02% of total SVs. These results
15 suggest an interaction between the impact of SV types and the potential function of the gene
16 regions they are found in.

A



B



1
 2 **Figure 3. The structural variant landscape of cats. (a)** Population frequency of each SV type.
 3 Colored bars represent the proportion of a given SV type found within various population
 4 frequency ranges. Total count values to the right of each bar represent the total number of SVs
 5 of each type, while percentage values in braces represent the proportion of all SVs that belong
 6 to each type. Frequency ranges shown in the legend are ordered along each bar from left to
 7 right. **(b)** The proportion of SVs found in each genomic region. Colored bars represent the
 8 proportion of SVs belonging to a particular type found across different genomic regions. Similar
 9 to above, the total count values to the right of each bar represent the total number of SVs found
 10 in each genomic region. The percentage values in braces represent the genome-wide
 11 proportion of all SVs found in each gene region. These values sum greater than 100 percent as
 12 a single SV can span multiple types of genomic regions. Structural variants are noted as:
 13 deletion (DEL); duplication (DUP); insertion (INS); inversion (INV).

14

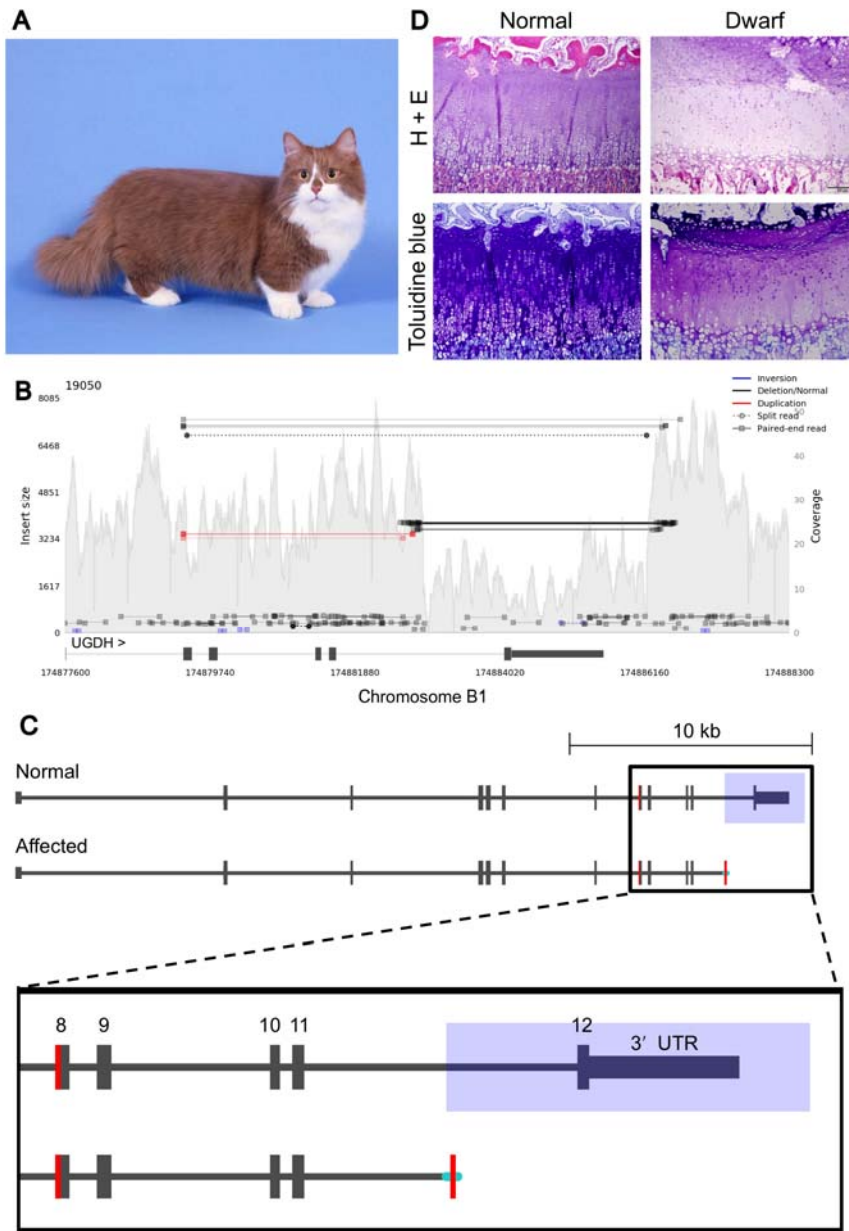
15

1 **Genetics of feline dwarfism.** Dwarfism in cats is the defining feature of the munchkin breed
2 and is characterized by shortened limbs and normal sized torso (**Fig. 4a**) [30]. Similar to
3 analyses with cat assemblies Felis_catus-6.2 and Felis_catus_8.0, previous investigations for
4 SNVs in the new Felis_catus_9.0 assembly did not identify any high priority candidate variants
5 for disproportionate dwarfism [27]. However, SV analysis within the critical region previously
6 identified by linkage and GWAS on chromosome B1:170,786,914-175,975,857 [27] revealed a
7 3.3 kb deletion at position chrB1:174,882,897-174,886,198, overlapping the final exon of *UDP-*
8 *glucose 6-dehydrogenase (UGDH)* (**Fig. 4b**). Upon manual inspection of this SV, a 49 bp
9 segment from exon 8 appeared to be duplicated and inserted 3.5 kb downstream, replacing the
10 deleted sequence. This potentially duplicated segment was flanked by a 37 bp sequence at the
11 5' end and a 20 bp sequence at the 3' end, both of unknown origin (**Fig. 4c**). Discordant reads
12 consistent with the SV were private to all three unrelated WGS dwarf samples (**S5 Fig.**). The
13 breakpoints surrounding the deletion were validated in WGS affected cats with Sanger
14 sequencing. PCR-based genotyping of the 3.3 kb deletion breakpoints was conducted in an
15 additional 108 cats including, 40 normal and 68 affected dwarf cats (**S6 Fig.**). Expected
16 amplicon sizes and phenotypes were concordant across all cats, except for a “munchkin; non-
17 standard (normal legs); Selkirk mix”, which appeared to carry the mutant allele, suggesting an
18 alternate causal gene or sampling error.

19 As *UGDH* has known roles in proteoglycan synthesis in chondrocytes [31, 32], growth plates in
20 two healthy and seven dwarf neonatal kittens were histologically analyzed for structural
21 irregularities and proteoglycan concentrations (Methods). The articular cartilage or bone in the
22 dwarf specimens had no significant histopathologic changes and epiphyseal plate was present
23 in all specimens. In the two normal kittens, chondrocytes in the epiphyseal plate exhibited a
24 regular columnar arrangement with organization into a zone of reserve cells, a zone of
25 proliferation, a zone of hypertrophy and a zone of provisional calcification (**Fig. 4d; S7 Fig.**).

1 Moreover, the articular-epiphyseal cartilage complex and the epiphyseal plate in the distal
2 radius from these two normal kittens also contained abundant proteoglycan as determined by
3 toluidine blue stain. In contrast, in dwarf kittens, the epiphyseal plate had a disorganized
4 columnar arrangement that was frequently coupled with proteoglycan depletion (**Fig. 4d; S7**
5 **Fig.**).

6



7

1 **Figure 4. A complex structural variant is associated with feline dwarfism. (a)** Image of
2 dwarf cat from the munchkin breed. Notice that limbs are short while torso remains normal size.
3 Image was donated courtesy of Terri Harris. **(b)** Samplot output of discordant reads within an
4 affected cat. Decreased coverage over the final exon of *UGDH* represents a deletion.
5 Discordant reads spanning beyond the deleted region show newly inserted sequence sharing
6 homology with *UGDH* exon 8. **(c)** Schematic illustrating the candidate structural variant for
7 dwarfism. Blue rectangle overlaps deleted region, shaded portion of exon 8 shows potential
8 duplication in the affected, and cyan lines indicate inserted sequence of unknown origin. **(d)**
9 Hematoxylin and eosin and toluidine blue histologic samples of the distal radius epiphyseal
10 cartilage plate from a neonatal kitten and an age-matched dwarf kitten. The normal kitten
11 showed regular columnar arrangement of chondrocytes with abundant proteoglycan as
12 determined by strong metachromasia with toluidine blue stain. The dwarf kitten showed
13 disorganized columnar arrangement of chondrocytes with proteoglycan depletion as determined
14 by weak metachromasia with toluidine blue stain.
15

16 **Discussion**

17 Studies of the domestic cat show the amazing features of this obligate carnivore, including their
18 incomplete domestication, wide range of coat color variation, and use as biomedical models [11,
19 33-35]. Using a combination of long-reads, combined with long-range scaffolding methods,
20 *Felis_catus_9.0* was generated as a new cat genome resource, with an N50 contig length (42
21 Mb) surpassing all other carnivore genome assemblies. This sequence contiguity represents a
22 1000-fold increase of ungapped sequence (contigs) length over *Felis_catus_8.0*, as well as a
23 40% reduction in the amount of unplaced sequence. All measures of sequence base and order
24 accuracy, such as the low number of discordant BAC end sequence alignments, suggest this
25 reference will strongly support future resequencing studies in cats. Equally important are the
26 observed improvements in gene annotation including the annotation of new genes, overall more
27 complete gene models and improvements to transcript mappability. Only 178 genes were
28 missing in *Felis_catus_9.0* compared to *Felis_catus_8.0*, which will require further investigation.
29 However, 376 predicted genes are novel to *Felis_catus_9.0*, i.e. not found in *Felis_catus_8.0*.
30 Using this *Felis_catus_9.0* assembly, a vast new repertoire of SNVs and SVs were discovered
31 for the domestic cat. The total number of variants discovered across our diverse collection of
32 domestic cats was substantially higher than previous studies in other mammals, such as, cow

1 [36], dog [37], rat [38-40], sheep [41, 42], pig [43], and horse [44]. Even rhesus macaques, with
2 twice as many variants as human, do not approach the same levels of cat SNV variation [45,
3 46]. The cats had 36.6 million biallelic SNVs with individuals carrying ~9.6 million SNVs each.
4 Conversely, humans roughly carry 4 to 5 million SNVs per individual [47]. One possible
5 explanation for this discrepancy may be the unique process cats underwent for domestication.
6 Rather than undergoing strong selective breeding leading to a severe population bottleneck,
7 cats were instead “self” domesticated, never losing some ancestral traits such as hunting
8 behavior [48-51]. The practice of strong selective breeding in cats only began in recent history
9 and is based almost exclusively on aesthetic traits. Consistent with previous analyses, our
10 evaluation of breeds and random bred cats showed tight clustering between cats from different
11 breeds that suggest cat breeds were likely initiated from local random bred cat populations [28,
12 52]. In regards to per sample numbers of variants, most breed cats had fewer SNVs than
13 random bred cats. Likewise, random bred cats also had higher numbers of singletons and a
14 lower inbreeding coefficient than breed cats, suggesting breed cats may share a common
15 genetic signature distinct from random bred populations.

16 For genetic discovery applications, animals with lower levels of genetic diversity are often more
17 desirable as they reduce experimental variability linked to variation in genetic architecture [53].
18 However, with decreasing costs of genome or exome sequencing leading to exponential growth
19 in variant discovery, animals with higher genetic diversity, such as cats, could enhance
20 discovery of tolerable loss of function or pathogenic missense variants. Similar to cats, rhesus
21 macaques from research colonies, exhibited per sample SNV rates more than two-fold greater
22 than human [46]. Importantly, macaque genetic diversity has been useful for characterizing non-
23 deleterious missense variation in humans [54].

24 A major premise of our feline genomics analysis is that increased discovery of genetic variants
25 in cats will help improve benign/pathogenic variant classification and reveal similarities between

1 human and feline genetic disease phenotypes, aiding disease interpretation in both species. To
2 gain insights into the burden of segregating variants with potentially deleterious impacts on
3 protein function, SNV impacts were classified across 54 unrelated domestic cats. Overall, cat
4 genes identified as being under strong constraint in humans were depleted of nonsynonymous
5 SNVs and enriched for potentially rare variants, showing both species harbor similar landscapes
6 of genetic constraint and indicating the utility of *Felis_catus_9.0* in modeling human genetic
7 disease. An important limitation on the analysis was the available sample size. As there were
8 only 54 cats, true rare variants, defined as variants with allele frequency of at least 1% in the
9 population [55], could not be distinguished from common variants that appear in only one of 54
10 randomly sampled cats by chance. Instead, singleton SNVs were focused on as candidates for
11 rare variants. Altogether, 18.6% of all feline SNVs could be considered candidates for rare
12 variants. Alternatively, larger analyses in human reveal a much higher fraction of low frequency
13 variants [56, 57]. For example, 36% of variants in an Icelandic population of over 2000
14 individuals had a minor allele frequency below 0.1% [57]. Similarly sized analyses in other
15 mammals reported ~20% of variants had allele frequencies < 1% [42, 44, 46]. As the number of
16 cats sequenced increases, the resolution to detect rare variants, as well as the total fraction of
17 low frequency variants, will continue to grow linearly as each individual will contribute a small
18 number of previously undiscovered variants [58].

19 Focusing specifically on potentially rare variants with high impacts in constrained genes
20 identified a potential cause for early onset feline mediastinal lymphoma, a stop gain in tumor
21 suppressor gene *FBXW7* [29]. Feline mediastinal lymphoma is distinct from other feline
22 lymphomas in its early onset and prevalence in Siamese cats and Oriental Shorthairs [59],
23 suggesting a genetic cause for lymphoma susceptibility specific to Siamese related cat breeds
24 [60, 61]. The stop gain was initially observed in a heterozygous state in a single cat identified as
25 a carrier in the unrelated set of cats. Subsequent analysis of the full set of cats revealed the

1 SNV had been inherited in an affected offspring. Despite discordance between the presence of
2 this mutant allele and the affection status of the cats it was found in, the variant still fits the
3 profile as a susceptibility allele for mediastinal lymphoma. For example, homozygous knockout
4 of *Fbxw7* is embryonic lethal in mice [62, 63], while heterozygous knockout mice develop
5 normally [63]. However, irradiation experiments of *Fbxw7*^{+/-} mice and *Fbxw7*^{+/-}*p53*^{+/-} crosses
6 identify *Fbxw7* as a haploinsufficient tumor suppressor gene that requires mutations in other
7 cancer related genes for tumorigenesis [64], a finding supported in subsequent mouse studies
8 [65, 66]. Similarly, in humans, germline variants in *FBXW7* are strongly associated with
9 predisposition to early onset cancers, such as Wilms tumors and Hodgkin's lymphoma [67, 68].
10 Screening of Siamese cats and other related breeds will validate the *FBWX7* stop gain as a
11 causative mutation for lymphoma susceptibility and may eventually aid in the development of a
12 feline cancer model.

13 An important concern with using human constraint metrics for identifying causative variants in
14 cats is the potential for false positives. For example, out of 16 SNVs initially identified as
15 potential feline disease candidates, many belonged to cats with no recorded disease status (**S5**
16 **Table**). However, since a large fraction of healthy humans also carried SNVs matching similar
17 disease causing criteria as used in cats [56], evolutionary distance between humans and cats is
18 unlikely to be a significant contributing factor to false positive candidate disease variant
19 identification. Instead, the frequency of high impact variants in constrained genes is likely due to
20 the limited resolution provided by analyses at the gene level. Recent strategies in humans have
21 confronted this problem by focusing on constraint at the level of gene region. These analyses
22 identified many genes with low pLI values that contained highly constrained gene regions that
23 were also enriched for disease causing variants [69]. Alternatively, another strategy for further
24 refining constrained regions could involve combining genomic variation from multiple species.

1 Many variants observed in cats, along with their impact on genes, are likely unique to cats and
2 could potentially be applied to variant prioritization workflows in humans.

3 Presented here is the first comprehensive genome-wide SV analysis for Felis_catus_9.0.
4 Previous genome-wide SV analyses in the cat were performed in Felis_catus-6.2 and focused
5 solely on copy number variation (CNV) [24]. Approximately 39,955 insertions, 123,731 deletions,
6 35,427 inversions, and 9,022 duplications were identified, far exceeding previous CNV
7 calculations of 521 losses and 68 gains. This large discrepancy is likely due to a number of
8 reasons including filtering stringency, sensitivity, and reference contiguity, where increased
9 numbers of assembly gaps in previous reference genome assemblies hindered detection of
10 larger SV events. An important difference regarding filtering stringency and sensitivity is that the
11 average CNV length was 37.4 kb as compared to 2.7 kb for SVs in Felis_catus_9.0. Since CNV
12 detection depends on read-depth [70, 71], only larger CNVs can be accurately identified, as
13 coverage at small window sizes can be highly variable. The use of both split-read and read-pair
14 information [72, 73], allowed the identification of SV events at much finer resolution than read-
15 depth based tools [74].

16 Improved sensitivity for smaller SV events, while helpful for finding new disease causing
17 mutations, may also lead to increased false positive SV detection. In three human trios
18 sequenced with Illumina technology, LUMPY and Delly were both used to identify 12,067 and
19 5,307 SVs respectively, contributing to a unified call set, along with several other tools, of
20 10,884 SVs per human on average [75]. In cats, the average number of SVs per individual was
21 4 times higher than in humans, with 44,990 SVs per cat, suggesting the total number SVs in
22 cats is likely inflated. However, the majority of the SVs were at population frequencies below 0.5,
23 ruling out poor reference assembly as a contributing factor. Instead, the difference in SV count
24 between cats and humans is likely due to sample specific factors. For example, the majority of
25 samples were sequenced using two separate libraries of 350 bp and 550 bp insert sizes (**S2**

1 **Data**). Despite the potentially high number of false positive SVs, increased SV sensitivity was
2 useful for trait discovery. The deletion associated with dwarfism was only found in the Delly2 call
3 set. If SV filtering were more stringent, such as requiring SVs to be called by both callers, the
4 feline dwarfism SV may not have otherwise been detected. Ultimately, these results highlight
5 the importance of high sensitivity for initial SV discovery and the use of highly specific molecular
6 techniques for downstream validation of candidate causative SVs.

7 The improved contiguity of *Felis_catus_9.0* was particularly beneficial for identifying a causative
8 SV for feline dwarfism. In humans, approximately 70% of cases are caused by spontaneous
9 mutations in *fibroblast growth factor 3* resulting in achondroplasia or a milder form of the
10 condition known as hypochondroplasia [76, 77]. The domestic cat is one of the few species with
11 an autosomal dominant mode of inheritance for dwarfism that does not have other syndromic
12 features. It can therefore provide as a strong model for hypochondroplasia. Previous GWAS and
13 linkage analyses suggested a critical region of association for feline disproportionate dwarfism
14 on cat chromosome B1 that spanned 5.2 Mb [27]. Within the critical region, SV analysis
15 identified a 3.3 kb deletion that had removed the final exon of *UGDH*, which was replaced by a
16 106 bp insertion with partial homology to *UGDH* exon 8, suggesting a potential duplication event.
17 Importantly, *UGDH* likely plays a role in proteoglycan synthesis in articular chondrocytes, as
18 osteoarthritic human and rat cartilage samples have revealed reduced *UGDH* protein
19 expression was associated with disease state [31]. Similarly, in dwarf cat samples, histology of
20 the distal radius showed irregularity of the chondrocyte organization and proteoglycan depletion.
21 Collectively, results suggest a disease model of reduced proteoglycan synthesis in dwarf cat
22 chondrocytes caused by loss of function of *UGDH* resulting in abnormal growth in the long
23 bones of dwarf cats. In ClinVar, only one missense variant in *UGDH* has been documented
24 NM_003359.4(UGDH):c.950G>A (p.Arg317Gln) and is associated with epileptic
25 encephalopathy. Seventeen other variants involving *UGDH* span multiple genes in the region.

1 The SV in the dwarf cats suggest *UGDH* is important in early long bone development. As a
2 novel gene association with dwarfism, *UGDH* should be screened for variants in undiagnosed
3 human dwarf patients.

4 High-quality genomes are a prerequisite for unhindered computational experimentation.
5 *Felis_catus_9.0* is currently the most contiguous genome of a companion animal, with high
6 accuracy and improved gene annotation that serves as a reference point for the discovery of
7 genetic variation associated with many traits. This new genomic resource will provide a
8 foundation for the future practice of genomic medicine in cats and for comparative analyses with
9 other species.

10 **Methods**

11 **Whole genome sequencing.** The same genome reference inbred domestic cat, Cinnamon, the
12 Abyssinian, was used for the long-read sequencing [78] [33]. High molecular weight DNA was
13 isolated using a MagAttract HMW-DNA Kit (Qiagen, Germantown, MD) from cultured fibroblast
14 cells according to the manufacturer's protocol. Single molecule real-time (SMRT) sequencing
15 was completed on the RSII and Sequel instruments (Pacific Biosciences, Menlo Park, CA).

16 **Genome assembly.** All sequences (~72x total sequence coverage) were assembled with the
17 fuzzy Bruijn graph algorithm, WTDBG (<https://github.com/ruanjue/wtdbg>), followed by collective
18 raw read alignment using MINIMAP to the error-prone primary contigs [79]. As a result, contig
19 coverage and graph topology were used to detect base errors that deviated from the majority
20 haplotype branches that are due to long-read error (i.e. chimeric reads) or erroneous graph
21 trajectories (i.e. repeats) as opposed to allelic structural variation, in which case, the sequence
22 of one of the alleles is incorporated into the final consensus bases. As a final step to improve
23 the consensus base quality of the assembly, from the same source DNA (Cinammon), short
24 read sequences (150 bp) were generated from 400 bp fragment size TruSeq libraries to ~60X

1 coverage on the Illumina HiSeqX instrument, which was then used to correct homozygous
2 insertion, deletion and single base differences using PILON [80].

3 **Assembly scaffolding.** To generate the first iteration of scaffolds from assembled contigs, the
4 BioNano Irys technology was used to define order and orientation, as well as, detect chimeric
5 contigs for automated breaks [81]. HMW-DNA in agar plugs was prepared from the same
6 cultured fibroblast cell line using the BioNano recommended protocol for soft tissues, where
7 using the IrysPrep Reagent Kit, a series of enzymatic reactions lysed cells, degraded protein
8 and RNA, and added fluorescent labels to nicked sites. The nicked DNA fragments were labeled
9 with ALEXA Fluor 546 dye and the DNA molecules were counter-stained with YOYO-1 dye.
10 After which, the labeled DNA fragments were electrophoretically elongated and sized on a
11 single IrysChip, with subsequent imaging and data processing to determine the size of each
12 DNA fragment. Finally, a *de novo* assembly was performed by using all labeled fragments >150
13 kb to construct a whole-genome optical map with defined overlap patterns. Individual maps
14 were clustered, scored for pairwise similarity, and Euclidian distance matrices were built.
15 Manual refinements were then performed as previously described [81].

16 **Assembly QC.** The scaffolded assembly was aligned to the latest cat linkage map [16] to detect
17 incorrect linkage between and within scaffolds, as well as discontinuous translocation events
18 that suggest contig chimerism. Following a genome-wide review of interchromosomal scaffold
19 discrepancies with the linkage map, the sequence breakpoints were manually determined and
20 the incorrect sequence linkages were separated. Also, to assess the assembly of expanded
21 heterozygous loci 'insertions', the same reference DNA sequences (Illumina short read inserts
22 300 bp) were aligned to the chromosomes to detect homozygous deletions in the read
23 alignments using Manta [82], a structural variant detection algorithm. In addition, the contigs
24 sequences were aligned to the *Felis_catus_8.0* using BLAT [83] at 99% identity and scored
25 alignment insertion length at 0.5 to 50 kb length to further refine putatively falsely assembled

1 heterozygous loci that when intersected with repeat tracks suggested either error in the
2 assembly or inability to correctly delineate the repetitive copy.

3 **Chromosome builds.** Upon correction and completion of the scaffold assembly, the genetic
4 linkage map [16] was used to first order and orient all possible scaffolds by using the
5 Chromonomer tool similarly to the previously reported default assembly parameter settings [84].
6 A final manual breakage of any remaining incorrect scaffold structure was made considering
7 various alignment discordance metrics, including to the prior reference *Felis_catus_8.0* that
8 defined unexpected interchromosomal translocations and lastly paired end size discordance
9 using alignments of BAC end sequences from the Cinnamon DNA source
10 (<http://ampliconexpress.com/bac-libraries/ite>).

11 **Gene Annotation.** The *Felis_catus_9.0* assembly was annotated using previously described
12 NCBI [85, 86] and Ensembl [87] pipelines, that included masking of repeats prior to *ab initio*
13 gene predictions and evidence-supported gene model building using RNA sequencing data [88].
14 RNA sequencing data of varied tissue types
15 (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Felis_catus/104) was used to further
16 improve gene model accuracy by alignment to nascent gene models that are necessary to
17 delineate boundaries of untranslated regions as well as to identify genes not found through
18 interspecific similarity evidence from other species.

19 **Characterizing feline sequence Variation.** Seventy-four cat WGSs from the 99 Lives Cat
20 Genome Project were downloaded from the NCBI short read archive (SRA) affiliated with NCBI
21 biosample and bioproject numbers (**S2 Data**). All sequences were produced with Illumina
22 technology, on either an Illumina HiSeq 2500 or XTen instrument using PCR-free libraries with
23 insert lengths ranging from 350 bp to 550 bp, producing 100 – 150 bp paired-end reads. WGS
24 data was processed using the Genome analysis toolkit (GATK) version 3.8 [89, 90]. BWA-MEM
25 from Burrows-Wheeler Aligner version 0.7.17 was used to map reads to *Felis_catus_9.0*

1 (GCF_000181335.3) [91]. Piccard tools version 2.1.1 (<http://broadinstitute.github.io/picard/>) was
2 used to mark duplicate reads, and samtools version 1.7 [92] was used to sort, merge and index
3 reads. Tools used from GATK 3.8 consisted of IndelRealigner and RealignerTargetCreator for
4 indel realignment, BaseRecalibrator for base quality score recalibration (BQSR) [93], and
5 HaplotypeCaller and GenotypeGVCFs for genotyping [94]. The variant database used for BQSR
6 was built by first genotyping non-recalibrated BAMs and applying a strict set of filters to isolate
7 high confidence variants. To determine the final variant call set, post BQSR, a less-strict GATK
8 recommended set of filters, was used. All filtering options are outlined in supplementary material
9 **(S6 Table)**. The set of unrelated cats was determined using vcftools' relatedness2 function on
10 SNP genotypes to estimate the kinship coefficient, ϕ , for each pair of cats [95]. First, related
11 cats were identified as sharing potential sibling and parent-child relationships if $\phi > 0.15$. Next,
12 cats with the highest number of relatives were removed in an iterative fashion until no
13 relationships with $\phi > 0.15$ remained. To detect population structure among the sequenced cats,
14 a principal component analyses was conducted using SNPRelate version 1.16.0 in the R
15 Statistical Software package. The SNV set generated after the appropriate quality control
16 measures were used was further filtered for non-biallelic SNVs and sites displaying linkage
17 disequilibrium (r^2 threshold = 0.2) as implemented in the SNPRelate package.

18 **Measuring coding variant impacts on human disease genes.** VCF summary statistics and
19 allele counts were computed using various functions from vcftools [95] and vcflib
20 (<https://github.com/vcflib/vcflib>). SNVs impacts of synonymous, missense, and LoF were
21 determined using Ensembl's variant effect predictor (VEP) with annotations from Ensembl
22 release 98 [87, 96]. Cat and human orthologs were identified using reciprocal best hits blast,
23 where Ensembl 98 protein fasta sequences were used as queries. Using pLI for human genes
24 obtained from gnomAD [56, 97], genes were assigned to constraint groups based on weak
25 constraint ($pLI < 0.1$), moderate constraint ($pLI > 0.1$ and $pLI < 0.9$), and strong constraint ($pLI >$

1 0.9). The observed per kb SNV density, Y^D , for each constraint group and SNV impact was
2 calculated as, $Y_{GI}^D = X_{GI}/C_G \times 1000bp$, where C is the total length of the coding sequence and X
3 is the number of SNVs within C . The subscript I refers to SNV impacts from the subset of all
4 SNVs, X . The subscript G refers to the subset of either C or X that are found within a particular
5 constraint group. For example, when G represents genes under weak constraint and I
6 represents LoF SNVs, X_{GI} would be all LoF SNVs within the coding sequence of genes under
7 weak constraint. The expected per kb SNV density, E^D , for each constraint group and impact
8 was calculated as, $E_{GI}^D = \frac{X_G(X_I+X)}{C_G} \times 1000bp$. The 95% confidence intervals surrounding the
9 expected per kb SNV densities were calculated from a random distribution generated by 10,000
10 permutations, where SNV impacts were shuffled randomly across variant sites. The observed
11 singleton fraction, Y^F , for each constraint group and impact was calculated as, $Y_{GI}^F = X_{GIP}/X_{GI}$,
12 where the subscript P refers to SNVs identified as singletons. The expected singleton fraction
13 was calculated as, $E_B^F = X_P/X$, where the 95% confidence intervals surrounding the expected
14 singleton fractions were also calculated from a random distribution generated by 10,000
15 permutations. For each permutation, SNV MAFs were shuffled so variants were randomly
16 assigned “singleton” status.

17 **Structural variant identification and analysis.** To discover SVs in the size range of <100 kb,
18 aligned reads from all cats to *Felis_catus_9.0* were used as input for the LUMPY [73] and Delly2
19 [72] SV callers. For LUMPY, the empirical insert size was determined using samtools and
20 `pairend_distro.py` for each BAM. Discordant and split-reads extracted from paired-end data
21 using SpeedSeq [98] were used as input along with each aligned BAM, minimum mapping
22 threshold of 20, and empirical mean and standard deviation of insert size. Samples were called
23 independently. SVTyper [98] was used to genotype each SV before merging all resulting VCFs
24 using BCFtools [92]. For Delly2 [72], all SVs were called for individual BAMs independently,
25 then merged into a single BCF. For each sample, variants were then re-called using the merged

1 results from all samples. The individual re-called VCFs were then merged into a single file using
2 BCFtools. Given the poor resolution of LUMPY for small insertions, only the Delly2 calls were
3 considered and variants were required to be found in more than 2 individuals. A convergence of
4 calls was determined by reciprocal overlap of 50% of the defined breakpoint in each caller as
5 our final set. SVs were annotated using SnpEff [99], which was used to count gene region
6 intersects. SVs were considered exonic if they were annotated as exon_region, frameshift,
7 start_lost, or stop_gained.

8 **Disease variant discovery for dwarfism.** To discover causal variants associated with
9 dwarfism, three unrelated affected cats with disproportionate dwarfism from the 99 Lives
10 genome dataset were examined for SVs. Identified SVs were considered causal candidates if
11 they were, 1) concordant with affection status and an autosomal dominant inheritance pattern,
12 2) and located within the ~5.2 Mb dwarfism critical region located on cat chromosome
13 B1:170,786,914 – 175,975,857 [27]. After initial identification, candidate variants were prioritized
14 according to their predicted impact on protein coding genes. High priority candidate SVs were
15 further characterized manually in affected individuals using the integrated genomics viewer
16 (IGV) [100]. STIX (structural variant index) was used to validate candidate SVs by searching
17 BAM files for discordant read-pairs that overlapped candidate SV regions
18 (<https://github.com/ryanlayer/stix>). After manual characterization, SV breakpoints were validated
19 with PCR amplification and Sanger sequencing. For further genotyping of candidate SVs, all
20 sample collection and cat studies were conducted in accordance with an approved University of
21 California, Davis Institutional Animal Care and Use protocols 11977, 15117, and 16691 and
22 University of Missouri protocols 7808 and 8292. DNA samples from dwarf and normal cats were
23 genotyped for the candidate SV identified in the three sequenced dwarfism cats [27]. PCR
24 primers were designed using the known SV sequence breakpoints (**S6a File**). For validation,
25 PCR amplification products were sanger sequenced and compared against Felis_catus_9.0. For

1 screening, all samples from previous linkage and GWAS studies [27] were genotyped using the
2 primers, UGDH_mid_F, UGDH_del_R, and UGDH_dn_R (**S7 Table**) in a single reaction. PCR
3 products were separated by gel electrophoresis (80V, 90 minutes) in 1.25% (w/v) agarose in 1X
4 TAE. A 622 bp amplicon was expected from the normal allele and a 481 bp amplicon from the
5 affected allele (**S6 Fig.**).

6 **Histological characterization of dwarf cat growth plates.** Cat owners voluntarily submitted
7 cadavers of stillborn dwarf and musculoskeletally normal kittens (7 and 2 respectively, dying
8 from natural causes) via overnight shipment on ice. Distal radius including physis (epiphyseal
9 plate) were collected and fixed in 10% neutral buffered formalin. These tissues were decalcified
10 in 10% EDTA solution. After complete decalcification the tissues were dehydrated with gradually
11 increasing concentrations of ethanol and embedded in paraffin. Frontal sections of distal radial
12 tissues were cut to 6 μm and mounted onto microscope slides. The samples were then
13 dewaxed, rehydrated, and stained with hematoxylin and eosin to evaluate the tissue structure
14 and cell morphology. The sections were also stained with toluidine blue to determine the
15 distribution and quantity of proteoglycans. These samples were subjectively assessed for
16 chondrocyte and tissue morphology and growth plate architecture by a pathologist (KK) who
17 was blinded to the sample information.

18 **Declarations**

19 **Ethics approval and consent to participate:** All sample collection and cat studies were
20 conducted in accordance with an approved University of California, Davis Institutional Animal
21 Care and Use protocols 11977, 15117, and 16691 and University of Missouri protocols 7808
22 and 8292.

23 **Availability of data and materials:** The datasets supporting the conclusions of this article are
24 available in NCBI's sequence read archive. The whole genome sequence data generated in this
25 study have been submitted to the NCBI BioProject database

1 (<http://www.ncbi.nlm.nih.gov/bioproject/>) under accession number PRJNA16726. Illumina WGS
2 data used in this study can be found under the NCBI BioProject accession PRJNA308208.
3 Source code used for analyses is publically available on GitHub ([https://github.com/mu-feline-](https://github.com/mu-feline-genome/Felis_catus_9.0_analysis)
4 [genome/Felis_catus_9.0_analysis](https://github.com/mu-feline-genome/Felis_catus_9.0_analysis)).

5 **Competing interests:** The authors disclose there are no conflicts of interest. The funders had
6 no role in study design, data collection, data analysis, interpretation of results, or decision to
7 publish.

8 **Funding:** Funding for this project has been provided in part by Nestlé Purina to pay for
9 personnel salaries and sequencing (W.C.W. and R.M.), Wisdom Health unit of Mars Veterinary
10 to pay for personnel salaries (L.A.L.), Zoetis to pay for RNA sequencing used in annotation
11 (L.A.L.), the University of Missouri College of Veterinary Medicine Gilbreath-McLorn endowment
12 was used for the BioNano optical map (L.A.L.), and Winn Feline Foundation W15-008 (W.J.M.),
13 Winn Feline Foundation/Miller Trust MT14-009 (W.J.M.), and Morris Animal Foundation D16FE-
14 011 (W.J.M.) were all used for PacBio sequencing.

15 **Authors' contributions:** Conceptualization: L.A.L., W.J.M., and W.C.W. Data Curation: R.M.B.,
16 B.W.D., F.H.G.F, Formal analysis: R.M.B., B.W.D., F.H.G.F., W.A.B, K.K., Funding Acquisition:
17 L.A.L., R.M., W.J.M., and W.C.W. Supervision: L.A.L. Validation: L.A.L. Writing (original draft):
18 R.M.B., L.A.L., and W.C.W. Writing (review and editing): all authors.

19 **Acknowledgements:** We appreciate the donation of funding and samples from cat breeders,
20 especially Terri Harris, who also provided the image in Fig. 4a. We appreciate Thomas R. Juba
21 for his technical assistance of the variant validation and genotyping. We thank Susan Brown at
22 Kansas State University for the generation of the BioNano map.

23

24

1

2 References

- 3 1. Mauler DA, Gandolfi B, Reiner CR, O'Brien DP, Spooner JL, Lyons LA, et al. Precision
4 Medicine in Cats: Novel Niemann-Pick Type C1 Diagnosed by Whole-Genome Sequencing. *J*
5 *Vet Intern Med.* 2017;31(2):539-44. doi: 10.1111/jvim.14599. PubMed PMID: 28233346;
6 PubMed Central PMCID: PMC5354023.
- 7 2. Fang H, Wu Y, Yang H, Yoon M, Jiménez-Barrón LT, Mittelman D, et al. Whole genome
8 sequencing of one complex pedigree illustrates challenges with genomic medicine. *BMC*
9 *medical genomics.* 2017;10(1):10.
- 10 3. Wise AL, Manolio TA, Mensah GA, Peterson JF, Roden DM, Tamburro C, et al.
11 Genomic medicine for undiagnosed diseases. *The Lancet.* 2019.
- 12 4. Shendure J, Findlay GM, Snyder MW. Genomic medicine—progress, pitfalls, and
13 promise. *Cell.* 2019;177(1):45-57.
- 14 5. Moses L, Niemi S, Karlsson E. Pet genomics medicine runs wild. Nature Publishing
15 Group; 2018.
- 16 6. Nicholas FW. Online Mendelian Inheritance in Animals (OMIA): a comparative
17 knowledgebase of genetic disorders and other familial traits in non-laboratory animals. *Nucleic*
18 *acids research.* 2003;31(1):275-7.
- 19 7. Online Mendelian Inheritance in Animals (OMIA). Sydney School of Veterinary Science,
20 03/12/2019. World Wide Web URL: <http://omia.org/>. Available from: <http://omia.org/>.
- 21 8. Lyons LA. DNA mutations of the cat: the good, the bad and the ugly. *J Feline Med Surg.*
22 2015;17(3):203-19. Epub 2015/02/24. doi: 10.1177/1098612X15571878. PubMed PMID:
23 25701860.
- 24 9. Kittleson MD, Meurs KM, Harris SP. The genetic basis of hypertrophic cardiomyopathy
25 in cats and humans. *J Vet Cardiol.* 2015;17 Suppl 1:S53-73. Epub 2016/01/19. doi:
26 10.1016/j.jvc.2015.03.001. PubMed PMID: 26776594; PubMed Central PMCID:
27 PMC5909964.
- 28 10. Menotti-Raymond M, David VA, Schäffer AA, Stephens R, Wells D, Kumar-Singh R, et al.
29 Mutation in CEP290 discovered for cat model of human retinal degeneration. *Journal of*
30 *Heredity.* 2007;98(3):211-20.
- 31 11. Lyons LA, Biller DS, Erdman CA, Lipinski MJ, Young AE, Roe BA, et al. Feline polycystic
32 kidney disease mutation identified in PKD1. *J Am Soc Nephrol.* 2004;15(10):2548-55. Epub
33 2004/10/07. doi: 10.1097/01.ASN.0000141776.38527.BB. PubMed PMID: 15466259.
- 34 12. Wang P, Mazrier H, Caverly Rae J, Raj K, Giger U. A GNPTAB nonsense variant is
35 associated with feline mucopolipidosis II (I-cell disease). *BMC Vet Res.* 2018;14(1):416. Epub
36 2018/12/29. doi: 10.1186/s12917-018-1728-1. PubMed PMID: 30591066; PubMed Central
37 PMCID: PMC6307278.
- 38 13. Spycher M, Bauer A, Jagannathan V, Frizzi M, De Lucia M, Leeb T. A frameshift variant
39 in the COL5A1 gene in a cat with Ehlers-Danlos syndrome. *Anim Genet.* 2018;49(6):641-4.
40 Epub 2018/09/25. doi: 10.1111/age.12727. PubMed PMID: 30246406.
- 41 14. Jaffey JA, Reading NS, Giger U, Abdulmalik O, Buckley RM, Johnstone S, et al. Clinical,
42 metabolic, and genetic characterization of hereditary methemoglobinemia caused by
43 cytochrome b5 reductase deficiency in cats. *Journal of veterinary internal medicine.* 2019.
- 44 15. Hug P, Kern P, Jagannathan V, Leeb T. A TAC3 Missense Variant in a Domestic
45 Shorthair Cat with Testicular Hypoplasia and Persistent Primary Dentition. *Genes.*
46 2019;10(10):806.
- 47 16. Li G, Hillier LW, Grahn RA, Zimin AV, David VA, Menotti-Raymond M, et al. A High-
48 Resolution SNP Array-Based Linkage Map Anchors a New Domestic Cat Draft Genome

1 Assembly and Provides Detailed Patterns of Recombination. *G3* (Bethesda). 2016;6(6):1607-16.
2 doi: 10.1534/g3.116.028746. PubMed PMID: 27172201; PubMed Central PMCID:
3 PMCPMC4889657.

4 17. Low WY, Tearle R, Bickhart DM, Rosen BD, Kingan SB, Swale T, et al. Chromosome-
5 level assembly of the water buffalo genome surpasses human and goat genomes in sequence
6 contiguity. *Nat Commun*. 2019;10(1):260. Epub 2019/01/18. doi: 10.1038/s41467-018-08260-0.
7 PubMed PMID: 30651564; PubMed Central PMCID: PMCPMC6335429.

8 18. Ananthasayanam S, Kothandaraman, H., Nayee, N., Saha, S., Baghel, D.S.,
9 Gopalakrishnan, K., Peddamma, S., Singh, R.B., Schatz, M. First near complete haplotype
10 phased genome assembly of River buffalo (*Bubalus bubalis*). *bioRxiv*. 2019;(April 26).

11 19. Gordon D, Huddleston J, Chaisson MJ, Hill CM, Kronenberg ZN, Munson KM, et al.
12 Long-read sequence assembly of the gorilla genome. *Science*. 2016;352(6281):aae0344. Epub
13 2016/04/02. doi: 10.1126/science.aae0344. PubMed PMID: 27034376; PubMed Central PMCID:
14 PMCPMC4920363.

15 20. Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, et al. Single-molecule
16 sequencing and chromatin conformation capture enable de novo reference assembly of the
17 domestic goat genome. *Nat Genet*. 2017;49(4):643-50. Epub 2017/03/07. doi: 10.1038/ng.3802.
18 PubMed PMID: 28263316; PubMed Central PMCID: PMCPMC5909822.

19 21. Ontiveros ES, Ueda Y, Harris SP, Stern JA, 99 Lives Consortium. Precision medicine
20 validation: identifying the MYBPC 3 A31P variant with whole-genome sequencing in two Maine
21 Coon cats with hypertrophic cardiomyopathy. *Journal of feline medicine and surgery*.
22 2018;1098612X18816460.

23 22. Oh A, Pearce JW, Gandolfi B, Creighton EK, Suedmeyer WK, Selig M, et al. Early-onset
24 progressive retinal atrophy associated with an IQCB1 variant in African black-footed cats (*Felis*
25 *nigripes*). *Scientific reports*. 2017;7:43918.

26 23. Aberdein D, Munday JS, Gandolfi B, Dittmer KE, Malik R, Garrick DJ, et al. A FAS-ligand
27 variant associated with autoimmune lymphoproliferative syndrome in cats. *Mamm Genome*.
28 2017;28(1-2):47-55. Epub 2016/10/23. doi: 10.1007/s00335-016-9668-1. PubMed PMID:
29 27770190.

30 24. Genova F, Longeri M, Lyons LA, Bagnato A, 99 Lives Consortium, Strillacci MG. First
31 genome-wide CNV mapping in FELIS CATUS using next generation sequencing data. *BMC*
32 *Genomics*. 2018;19(1):895. Epub 2018/12/12. doi: 10.1186/s12864-018-5297-2. PubMed PMID:
33 30526495; PubMed Central PMCID: PMCPMC6288940.

34 25. Smit A, Hubley R, Green P. 2013–2015. RepeatMasker Open-4.0. 2013.

35 26. Gandolfi B, Alhaddad H, Abdi M, Bach LH, Creighton EK, Davis BW, et al. Applications
36 and efficiencies of the first cat 63K DNA array. *Sci Rep*. 2018;8(1):7024. Epub 2018/05/08. doi:
37 10.1038/s41598-018-25438-0. PubMed PMID: 29728693; PubMed Central PMCID:
38 PMCPMC5935720.

39 27. Lyons LA, Fox DB, Chesney KL, Britt LG, Buckley RM, Coates JR, et al. Localization of
40 a feline autosomal dominant dwarfism locus: a novel model of chondrodysplasia. *bioRxiv*.
41 2019:687210.

42 28. Lipinski MJ, Froenicke L, Baysac KC, Billings NC, Leutenegger CM, Levy AM, et al. The
43 ascent of cat breeds: genetic evaluations of breeds and worldwide random-bred populations.
44 *Genomics*. 2008;91(1):12-21. Epub 2007/12/07. doi: 10.1016/j.ygeno.2007.10.009. PubMed
45 PMID: 18060738; PubMed Central PMCID: PMCPMC2267438.

46 29. Yeh CH, Bellon M, Nicot C. FBXW7: a critical tumor suppressor of human cancers. *Mol*
47 *Cancer*. 2018;17(1):115. Epub 2018/08/09. doi: 10.1186/s12943-018-0857-2. PubMed PMID:
48 30086763; PubMed Central PMCID: PMCPMC6081812.

49 30. TICA. Munchkin 2015 [October 26, 2015].

50 31. Wen Y, Li J, Wang L, Tie K, Magdalou J, Chen L, et al. UDP-glucose dehydrogenase
51 modulates proteoglycan synthesis in articular chondrocytes: its possible involvement and

1 regulation in osteoarthritis. *Arthritis Res Ther.* 2014;16(6):484. Epub 2014/12/04. doi:
2 10.1186/s13075-014-0484-2. PubMed PMID: 25465897; PubMed Central PMCID:
3 PMCPMC4298080.

4 32. Clarkin CE, Allen S, Kuiper NJ, Wheeler BT, Wheeler-Jones CP, Pitsillides AA.
5 Regulation of UDP-glucose dehydrogenase is sufficient to modulate hyaluronan production and
6 release, control sulfated GAG synthesis, and promote chondrogenesis. *J Cell Physiol.*
7 2011;226(3):749-61. Epub 2010/08/19. doi: 10.1002/jcp.22393. PubMed PMID: 20717929.

8 33. Montague MJ, Li G, Gandolfi B, Khan R, Aken BL, Searle SM, et al. Comparative
9 analysis of the domestic cat genome reveals genetic signatures underlying feline biology and
10 domestication. *Proc Natl Acad Sci U S A.* 2014;111(48):17230-5. doi:
11 10.1073/pnas.1410083111. PubMed PMID: 25385592; PubMed Central PMCID:
12 PMCPMC4260561.

13 34. Yu Y, Grahn RA, Lyons LA. Mocha tyrosinase variant: a new flavour of cat coat
14 coloration. *Anim Genet.* 2019;50(2):182-6. Epub 2019/02/05. doi: 10.1111/age.12765. PubMed
15 PMID: 30716167.

16 35. Yu Y, Shumway KL, Matheson JS, Edwards ME, Kline TL, Lyons LA. Kidney and cystic
17 volume imaging for disease presentation and progression in the cat autosomal dominant
18 polycystic kidney disease large animal model. *BMC Nephrology.* 2019;20(1):259.

19 36. Daetwyler HD, Capitan A, Pausch H, Stothard P, van Binsbergen R, Brondum RF, et al.
20 Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in
21 cattle. *Nat Genet.* 2014;46(8):858-65. Epub 2014/07/16. doi: 10.1038/ng.3034. PubMed PMID:
22 25017103.

23 37. Bai B, Zhao WM, Tang BX, Wang YQ, Wang L, Zhang Z, et al. DoGSD: the dog and wolf
24 genome SNP database. *Nucleic Acids Res.* 2015;43(Database issue):D777-83. Epub
25 2014/11/19. doi: 10.1093/nar/gku1174. PubMed PMID: 25404132; PubMed Central PMCID:
26 PMCPMC4383968.

27 38. Atanur SS, Diaz AG, Maratou K, Sarkis A, Rotival M, Game L, et al. Genome
28 sequencing reveals loci under artificial selection that underlie disease phenotypes in the
29 laboratory rat. *Cell.* 2013;154(3):691-703. Epub 2013/07/31. doi: 10.1016/j.cell.2013.06.040.
30 PubMed PMID: 23890820; PubMed Central PMCID: PMCPMC3732391.

31 39. Hermsen R, de Ligt J, Spee W, Blokzijl F, Schafer S, Adami E, et al. Genomic landscape
32 of rat strain and substrain variation. *BMC Genomics.* 2015;16:357. Epub 2015/05/07. doi:
33 10.1186/s12864-015-1594-1. PubMed PMID: 25943489; PubMed Central PMCID:
34 PMCPMC4422378.

35 40. Teng H, Zhang Y, Shi C, Mao F, Cai W, Lu L, et al. Population Genomics Reveals
36 Speciation and Introgression between Brown Norway Rats and Their Sibling Species. *Mol Biol*
37 *Evol.* 2017;34(9):2214-28. Epub 2017/05/10. doi: 10.1093/molbev/msx157. PubMed PMID:
38 28482038; PubMed Central PMCID: PMCPMC5850741.

39 41. Yang J, Li WR, Lv FH, He SG, Tian SL, Peng WF, et al. Whole-Genome Sequencing of
40 Native Sheep Provides Insights into Rapid Adaptations to Extreme Environments. *Mol Biol Evol.*
41 2016;33(10):2576-92. Epub 2016/07/13. doi: 10.1093/molbev/msw129. PubMed PMID:
42 27401233; PubMed Central PMCID: PMCPMC5026255.

43 42. Chen ZH, Zhang M, Lv FH, Ren X, Li WR, Liu MJ, et al. Contrasting Patterns of
44 Genomic Diversity Reveal Accelerated Genetic Drift but Reduced Directional Selection on X-
45 Chromosome in Wild and Domestic Sheep Species. *Genome Biol Evol.* 2018;10(5):1282-97.
46 Epub 2018/05/24. doi: 10.1093/gbe/evy085. PubMed PMID: 29790980; PubMed Central
47 PMCID: PMCPMC5963296.

48 43. Choi JW, Chung WH, Lee KT, Cho ES, Lee SW, Choi BH, et al. Whole-genome
49 resequencing analyses of five pig breeds, including Korean wild and native, and three European
50 origin breeds. *DNA Res.* 2015;22(4):259-67. Epub 2015/06/29. doi: 10.1093/dnares/dsv011.
51 PubMed PMID: 26117497; PubMed Central PMCID: PMCPMC4535618.

- 1 44. Jagannathan V, Gerber V, Rieder S, Tetens J, Thaller G, Drogemuller C, et al.
2 Comprehensive characterization of horse genome variation by whole-genome sequencing of 88
3 horses. *Anim Genet.* 2019;50(1):74-7. Epub 2018/12/14. doi: 10.1111/age.12753. PubMed
4 PMID: 30525216.
- 5 45. Bimber BN, Ramakrishnan R, Cervera-Juanes R, Madhira R, Peterson SM, Norgren Jr
6 RB, et al. Whole genome sequencing predicts novel human disease models in rhesus
7 macaques. *Genomics.* 2017;109(3-4):214-20.
- 8 46. Xue C, Raveendran M, Harris RA, Fawcett GL, Liu X, White S, et al. The population
9 genomics of rhesus macaques (*Macaca mulatta*) based on whole-genome sequences. *Genome*
10 *Res.* 2016;26(12):1651-62. Epub 2016/12/10. doi: 10.1101/gr.204255.116. PubMed PMID:
11 27934697; PubMed Central PMCID: PMC5131817.
- 12 47. The 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP,
13 Kang HM, et al. A global reference for human genetic variation. *Nature.* 2015;526(7571):68-74.
14 Epub 2015/10/04. doi: 10.1038/nature15393. PubMed PMID: 26432245; PubMed Central
15 PMCID: PMC4750478.
- 16 48. Ottoni C, Van Neer W, De Cupere B, Daligault J, Guimaraes S, Peters J, et al. The
17 palaeogenetics of cat dispersal in the ancient world. *Nature Ecology & Evolution.*
18 2017;1(7):0139.
- 19 49. Vigne J-D, Guilaine J, Debue K, Haye L, Gérard P. Early taming of the cat in Cyprus.
20 *Science.* 2004;304(5668):259-.
- 21 50. Van Neer W, Linseele V, Friedman R, De Cupere B. More evidence for cat taming at the
22 Predynastic elite cemetery of Hierakonpolis (Upper Egypt). *Journal of Archaeological Science.*
23 2014;45:103-11.
- 24 51. Driscoll CA, Menotti-Raymond M, Roca AL, Hupe K, Johnson WE, Geffen E, et al. The
25 Near Eastern origin of cat domestication. *Science.* 2007;317(5837):519-23. Epub 2007/06/30.
26 doi: 10.1126/science.1139518. PubMed PMID: 17600185; PubMed Central PMCID:
27 PMC5612713.
- 28 52. Kurushima JD, Lipinski MJ, Gandolfi B, Froenicke L, Grahn JC, Grahn RA, et al.
29 Variation of cats under domestication: genetic assignment of domestic cats to breeds and
30 worldwide random-bred populations. *Anim Genet.* 2013;44(3):311-24. Epub 2012/11/23. doi:
31 10.1111/age.12008. PubMed PMID: 23171373; PubMed Central PMCID: PMC3594446.
- 32 53. Festing MF. Inbred strains should replace outbred stocks in toxicology, safety testing,
33 and drug development. *Toxicologic pathology.* 2010;38(5):681-90.
- 34 54. Sundaram L, Gao H, Padigepati SR, McRae JF, Li Y, Kosmicki JA, et al. Predicting the
35 clinical impact of human mutation with deep neural networks. *Nat Genet.* 2018;50(8):1161-70.
36 Epub 2018/07/25. doi: 10.1038/s41588-018-0167-z. PubMed PMID: 30038395; PubMed Central
37 PMCID: PMC6237276.
- 38 55. Frazer KA, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its
39 contribution to complex traits. *Nature Reviews Genetics.* 2009;10(4):241.
- 40 56. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of
41 protein-coding genetic variation in 60,706 humans. *Nature.* 2016;536(7616):285-91. Epub
42 2016/08/19. doi: 10.1038/nature19057. PubMed PMID: 27535533; PubMed Central PMCID:
43 PMC5018207.
- 44 57. Gudbjartsson DF, Helgason H, Gudjonsson SA, Zink F, Oddson A, Gylfason A, et al.
45 Large-scale whole-genome sequencing of the Icelandic population. *Nat Genet.* 2015;47(5):435-
46 44. Epub 2015/03/26. doi: 10.1038/ng.3247. PubMed PMID: 25807286.
- 47 58. The 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD,
48 Durbin RM, et al. A map of human genome variation from population-scale sequencing. *Nature.*
49 2010;467(7319):1061-73. Epub 2010/10/29. doi: 10.1038/nature09534. PubMed PMID:
50 20981092; PubMed Central PMCID: PMC3042601.

- 1 59. Gabor L, Malik R, Canfield P. Clinical and anatomical features of lymphosarcoma in 118
2 cats. *Australian Veterinary Journal*. 1998;76(11):725-32.
- 3 60. Louwerens M, London CA, Pedersen NC, Lyons LA. Feline lymphoma in the post-feline
4 leukemia virus era. *J Vet Intern Med*. 2005;19(3):329-35. Epub 2005/06/16. doi: 10.1892/0891-
5 6640(2005)19[329:flitpl]2.0.co;2. PubMed PMID: 15954547.
- 6 61. Fabrizio F, Calam AE, Dobson JM, Middleton SA, Murphy S, Taylor SS, et al. Feline
7 mediastinal lymphoma: a retrospective study of signalment, retroviral status, response to
8 chemotherapy and prognostic indicators. *J Feline Med Surg*. 2014;16(8):637-44. Epub
9 2013/12/25. doi: 10.1177/1098612X13516621. PubMed PMID: 24366846.
- 10 62. Tetzlaff MT, Yu W, Li M, Zhang P, Finegold M, Mahon K, et al. Defective cardiovascular
11 development and elevated cyclin E and Notch proteins in mice lacking the Fbw7 F-box protein.
12 *Proc Natl Acad Sci U S A*. 2004;101(10):3338-45. Epub 2004/02/10. doi:
13 10.1073/pnas.0307875101. PubMed PMID: 14766969; PubMed Central PMCID:
14 PMCPMC373463.
- 15 63. Tsunematsu R, Nakayama K, Oike Y, Nishiyama M, Ishida N, Hatakeyama S, et al.
16 Mouse Fbw7/Sel-10/Cdc4 is required for notch degradation during vascular development. *J Biol*
17 *Chem*. 2004;279(10):9417-23. Epub 2003/12/16. doi: 10.1074/jbc.M312337200. PubMed PMID:
18 14672936.
- 19 64. Mao JH, Perez-Losada J, Wu D, Delrosario R, Tsunematsu R, Nakayama KI, et al.
20 Fbxw7/Cdc4 is a p53-dependent, haploinsufficient tumour suppressor gene. *Nature*.
21 2004;432(7018):775-9. Epub 2004/12/14. doi: 10.1038/nature03155. PubMed PMID: 15592418.
- 22 65. Perez-Losada J, Wu D, DelRosario R, Balmain A, Mao JH. Allele-specific deletions in
23 mouse tumors identify Fbxw7 as germline modifier of tumor susceptibility. *PLoS One*.
24 2012;7(2):e31301. Epub 2012/02/22. doi: 10.1371/journal.pone.0031301. PubMed PMID:
25 22348067; PubMed Central PMCID: PMCPMC3278431.
- 26 66. Maser RS, Choudhury B, Campbell PJ, Feng B, Wong KK, Protopopov A, et al.
27 Chromosomally unstable mouse tumours have genomic alterations similar to diverse human
28 cancers. *Nature*. 2007;447(7147):966-71. Epub 2007/05/23. doi: 10.1038/nature05886. PubMed
29 PMID: 17515920; PubMed Central PMCID: PMCPMC2714968.
- 30 67. Roversi G, Picinelli C, Bestetti I, Crippa M, Perotti D, Ciceri S, et al. Constitutional de
31 novo deletion of the FBXW7 gene in a patient with focal segmental glomerulosclerosis and
32 multiple primitive tumors. *Sci Rep*. 2015;5:15454. Epub 2015/10/21. doi: 10.1038/srep15454.
33 PubMed PMID: 26482194; PubMed Central PMCID: PMCPMC4612309.
- 34 68. Mahamdallie S, Yost S, Poyastro-Pearson E, Holt E, Zachariou A, Seal S, et al.
35 Identification of new Wilms tumour predisposition genes: an exome sequencing study. *Lancet*
36 *Child Adolesc Health*. 2019;3(5):322-31. Epub 2019/03/20. doi: 10.1016/S2352-4642(19)30018-
37 5. PubMed PMID: 30885698; PubMed Central PMCID: PMCPMC6472290.
- 38 69. Havrilla JM, Pedersen BS, Layer RM, Quinlan AR. A map of constrained coding regions
39 in the human genome.
- 40 70. Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: an approach to discover,
41 genotype, and characterize typical and atypical CNVs from family and population genome
42 sequencing. *Genome Res*. 2011;21(6):974-84. Epub 2011/02/18. doi: 10.1101/gr.114876.110.
43 PubMed PMID: 21324876; PubMed Central PMCID: PMCPMC3106330.
- 44 71. Klambauer G, Schwarzbauer K, Mayr A, Clevert DA, Mitterecker A, Bodenhofer U, et al.
45 cn.MOPS: mixture of Poissons for discovering copy number variations in next-generation
46 sequencing data with a low false discovery rate. *Nucleic Acids Res*. 2012;40(9):e69. Epub
47 2012/02/04. doi: 10.1093/nar/gks003. PubMed PMID: 22302147; PubMed Central PMCID:
48 PMCPMC3351174.
- 49 72. Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. DELLY: structural
50 variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*.

1 2012;28(18):i333-i9. doi: 10.1093/bioinformatics/bts378. PubMed PMID: 22962449; PubMed
2 Central PMCID: PMCPMC3436805.

3 73. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for
4 structural variant discovery. *Genome Biol.* 2014;15(6):R84. doi: 10.1186/gb-2014-15-6-r84.
5 PubMed PMID: 24970577; PubMed Central PMCID: PMCPMC4197822.

6 74. Kosugi S, Momozawa Y, Liu X, Terao C, Kubo M, Kamatani Y. Comprehensive
7 evaluation of structural variation detection algorithms for whole genome sequencing. *Genome*
8 *Biol.* 2019;20(1):117. Epub 2019/06/05. doi: 10.1186/s13059-019-1720-5. PubMed PMID:
9 31159850; PubMed Central PMCID: PMCPMC6547561.

10 75. Chaisson MJP, Sanders AD, Zhao X, Malhotra A, Porubsky D, Rausch T, et al. Multi-
11 platform discovery of haplotype-resolved structural variation in human genomes. *Nat Commun.*
12 2019;10(1):1784. Epub 2019/04/18. doi: 10.1038/s41467-018-08148-z. PubMed PMID:
13 30992455; PubMed Central PMCID: PMCPMC6467913.

14 76. Horton WA, Hall JG, Hecht JT. Achondroplasia. *Lancet.* 2007;370(9582):162-72. Epub
15 2007/07/17. doi: 10.1016/S0140-6736(07)61090-3. PubMed PMID: 17630040.

16 77. Foldynova-Trantirkova S, Wilcox WR, Krejci P. Sixteen years and counting: the current
17 understanding of fibroblast growth factor receptor 3 (FGFR3) signaling in skeletal dysplasias.
18 *Hum Mutat.* 2012;33(1):29-41. Epub 2011/11/03. doi: 10.1002/humu.21636. PubMed PMID:
19 22045636; PubMed Central PMCID: PMCPMC3240715.

20 78. Pontius JU, Mullikin JC, Smith DR, Agencourt Sequencing T, Lindblad-Toh K, Gnerre S,
21 et al. Initial sequence and comparative analysis of the cat genome. *Genome Res.*
22 2007;17(11):1675-89. Epub 2007/11/03. doi: 10.1101/gr.6380007. PubMed PMID: 17975172;
23 PubMed Central PMCID: PMCPMC2045150.

24 79. Li H. Minimap and miniasm: fast mapping and de novo assembly for noisy long
25 sequences. *Bioinformatics.* 2016;32(14):2103-10. Epub 2016/05/07. doi:
26 10.1093/bioinformatics/btw152. PubMed PMID: 27153593; PubMed Central PMCID:
27 PMCPMC4937194.

28 80. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an
29 integrated tool for comprehensive microbial variant detection and genome assembly
30 improvement. *PLoS One.* 2014;9(11):e112963. doi: 10.1371/journal.pone.0112963. PubMed
31 PMID: 25409509; PubMed Central PMCID: PMCPMC4237348.

32 81. Lam ET, Hastie A, Lin C, Ehrlich D, Das SK, Austin MD, et al. Genome mapping on
33 nanochannel arrays for structural variation analysis and sequence assembly. *Nature*
34 *biotechnology.* 2012;30(8):771.

35 82. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Kallberg M, et al. Manta:
36 rapid detection of structural variants and indels for germline and cancer sequencing applications.
37 *Bioinformatics.* 2016;32(8):1220-2. doi: 10.1093/bioinformatics/btv710. PubMed PMID:
38 26647377.

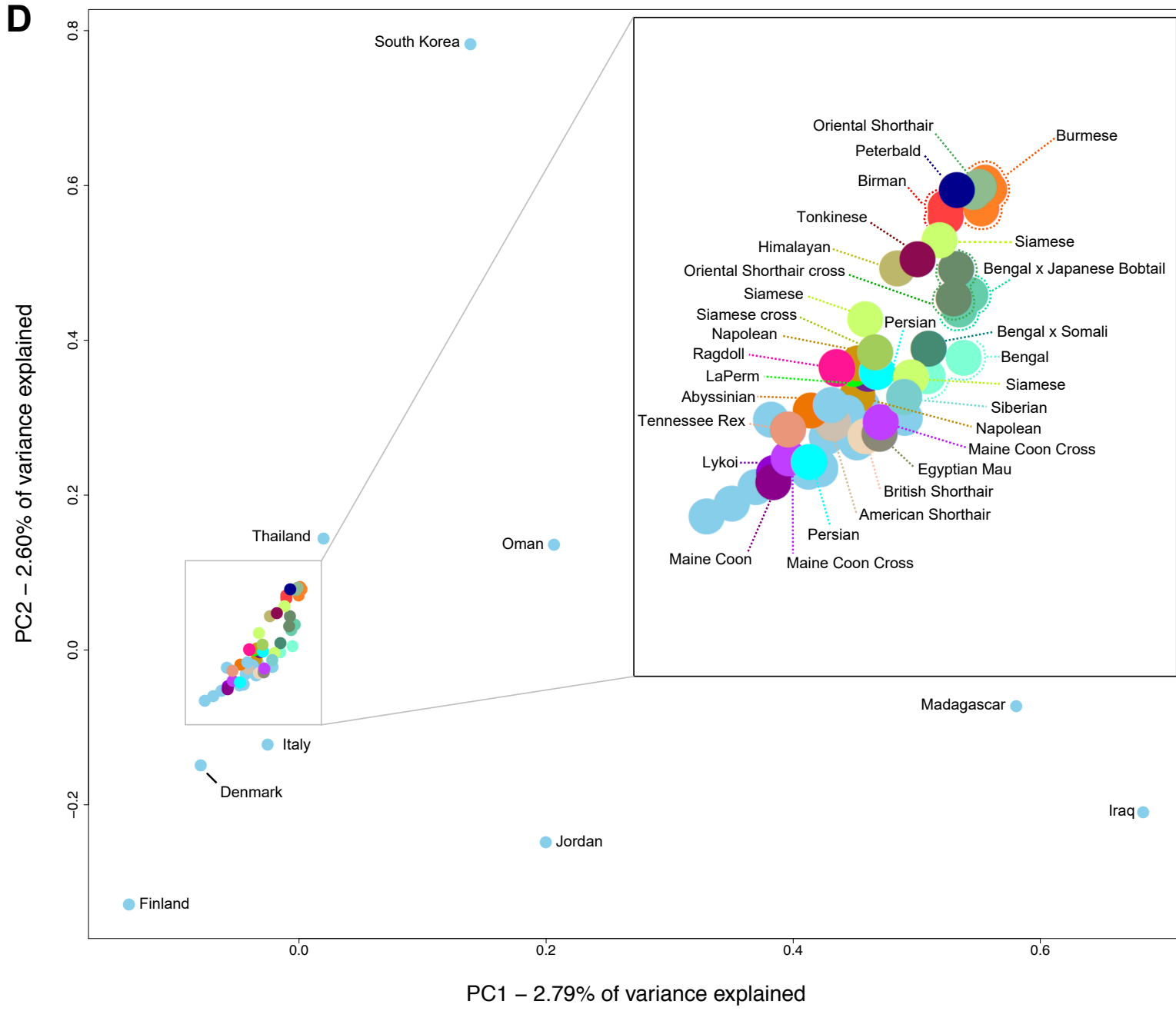
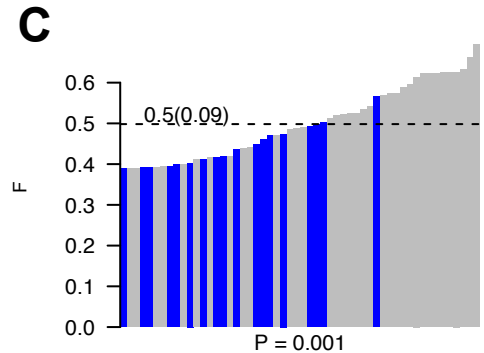
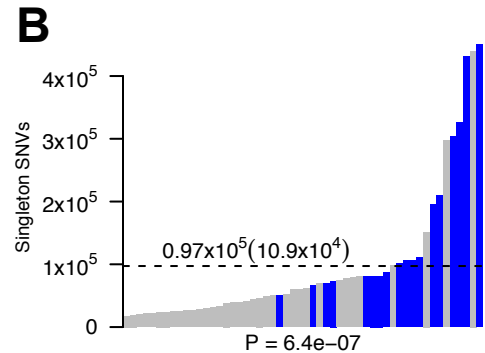
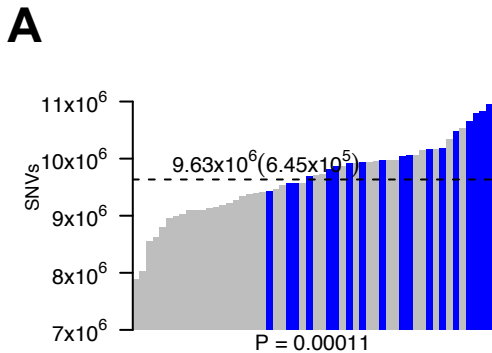
39 83. Kent WJ. BLAT--the BLAST-like alignment tool. *Genome Res.* 2002;12(4):656-64. Epub
40 2002/04/05. doi: 10.1101/gr.229202. PubMed PMID: 11932250; PubMed Central PMCID:
41 PMCPMC187518.

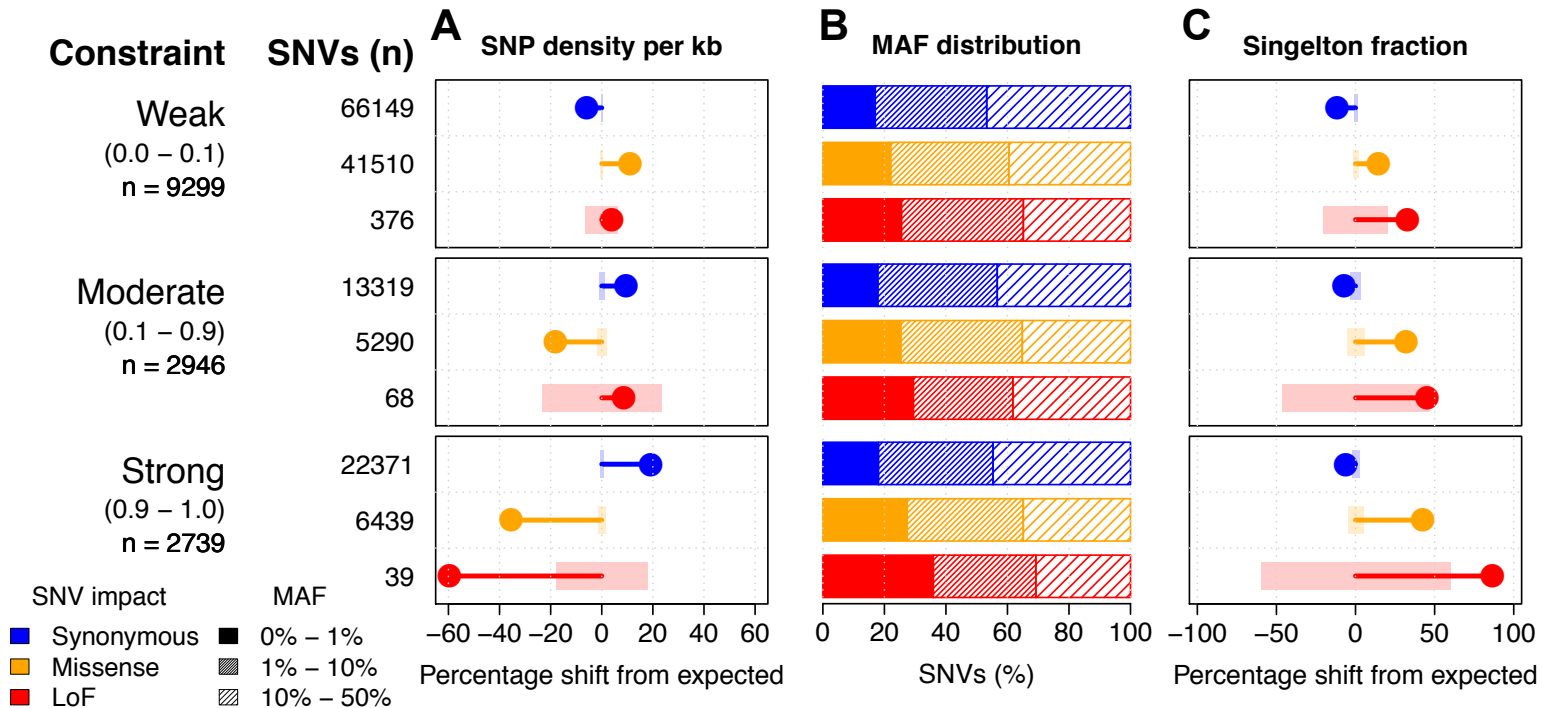
42 84. Small C, Bassham S, Catchen J, Amores A, Fuiten A, Brown R, et al. The genome of the
43 Gulf pipefish enables understanding of evolutionary innovations. *Genome biology.*
44 2016;17(1):258.

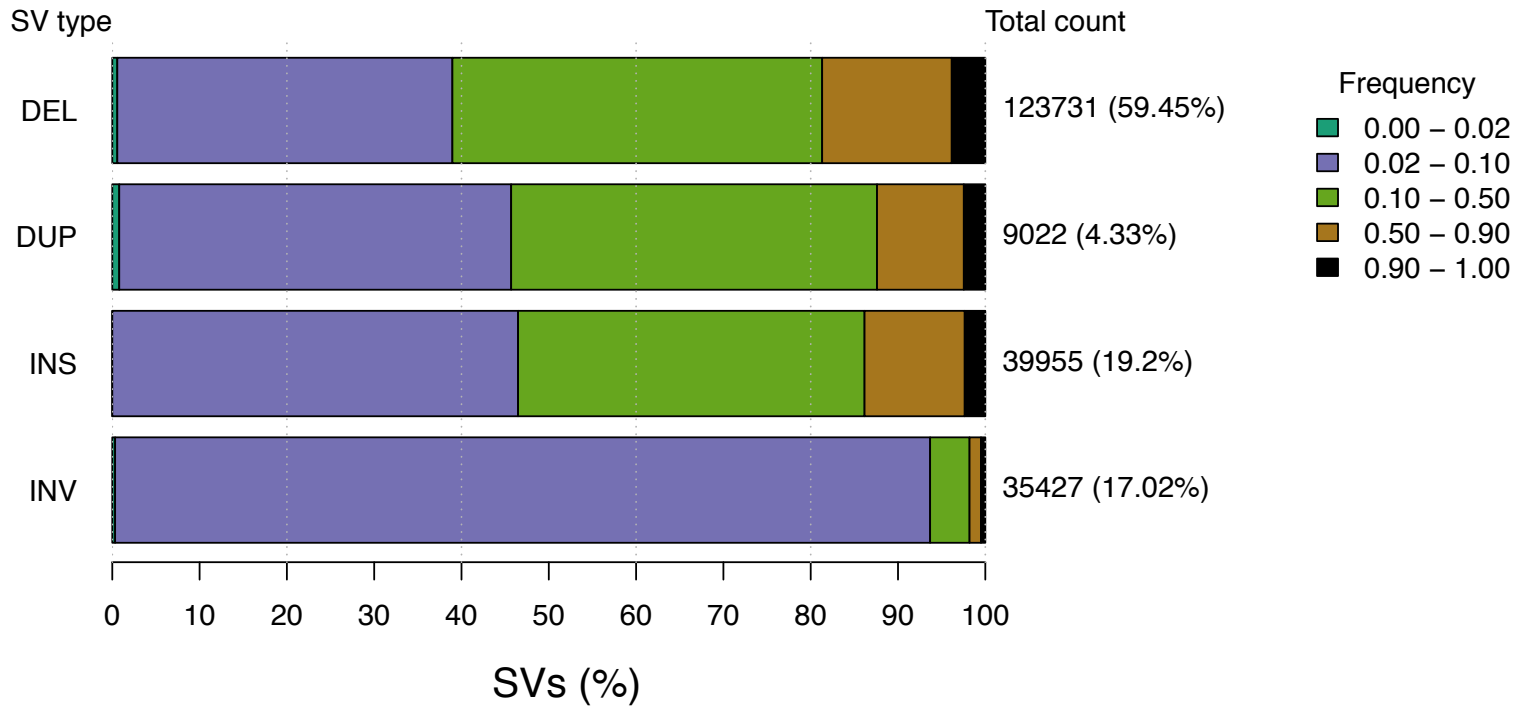
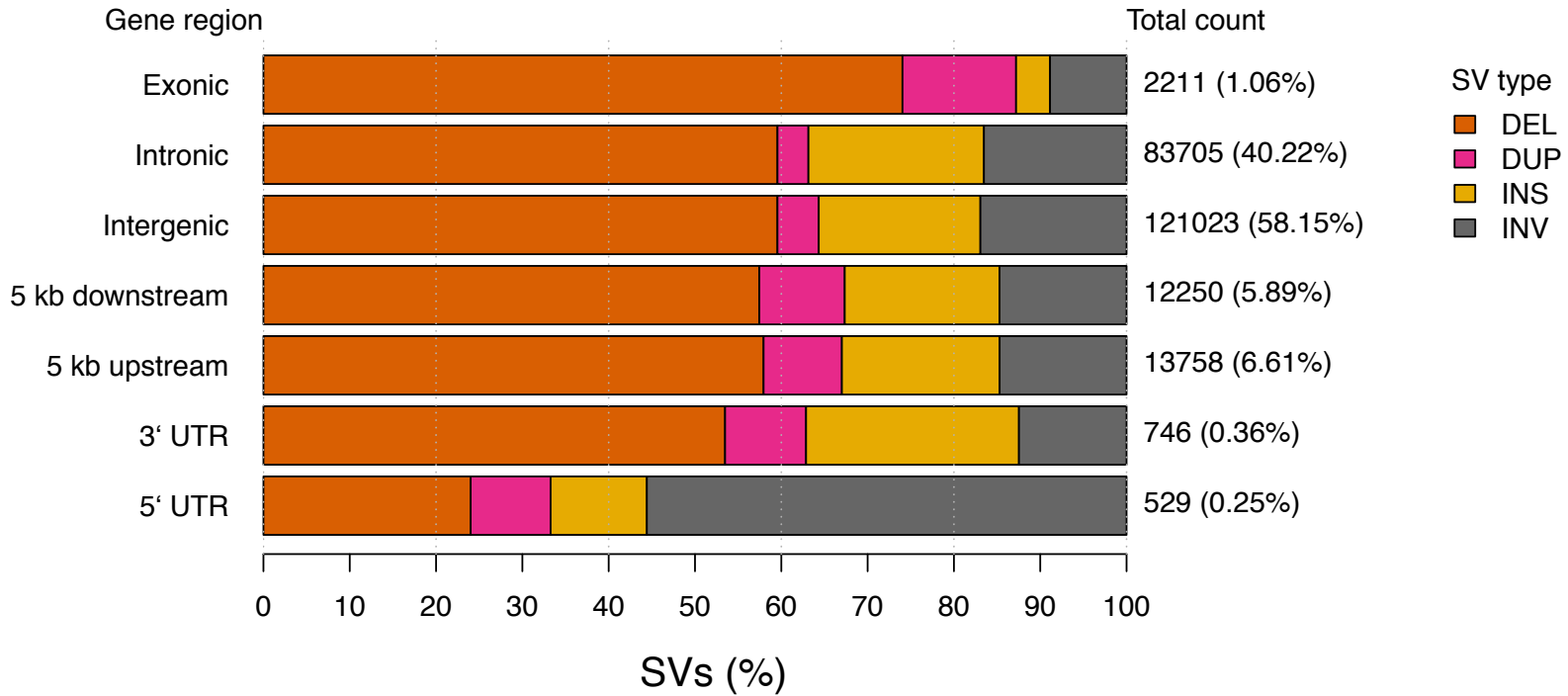
45 85. Pruitt KD, Tatusova T, Brown GR, Maglott DR. NCBI Reference Sequences (RefSeq):
46 current status, new features and genome annotation policy. *Nucleic Acids Res.*
47 2012;40(Database issue):D130-5. Epub 2011/11/29. doi: 10.1093/nar/gkr1079. PubMed PMID:
48 22121212; PubMed Central PMCID: PMCPMC3245008.

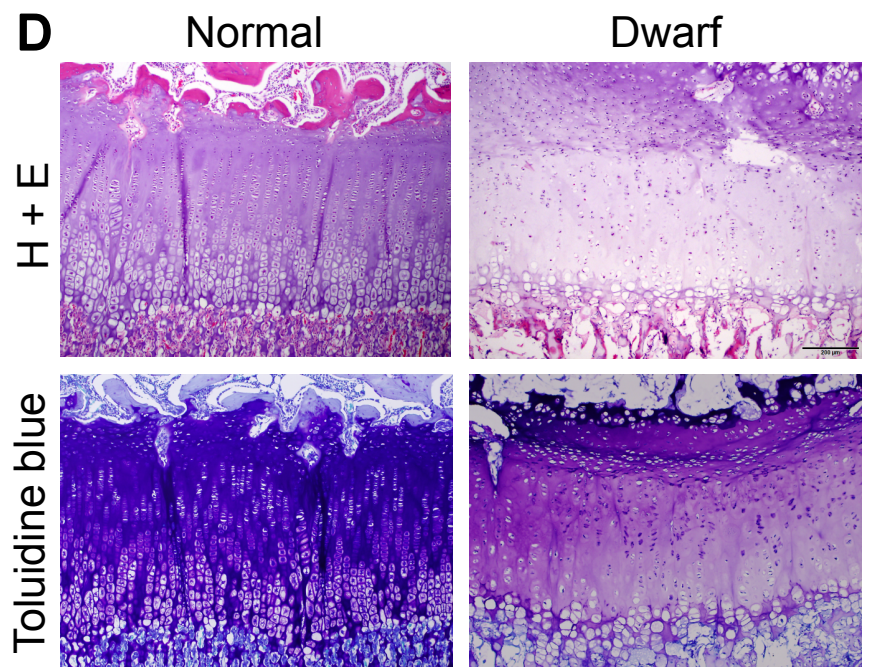
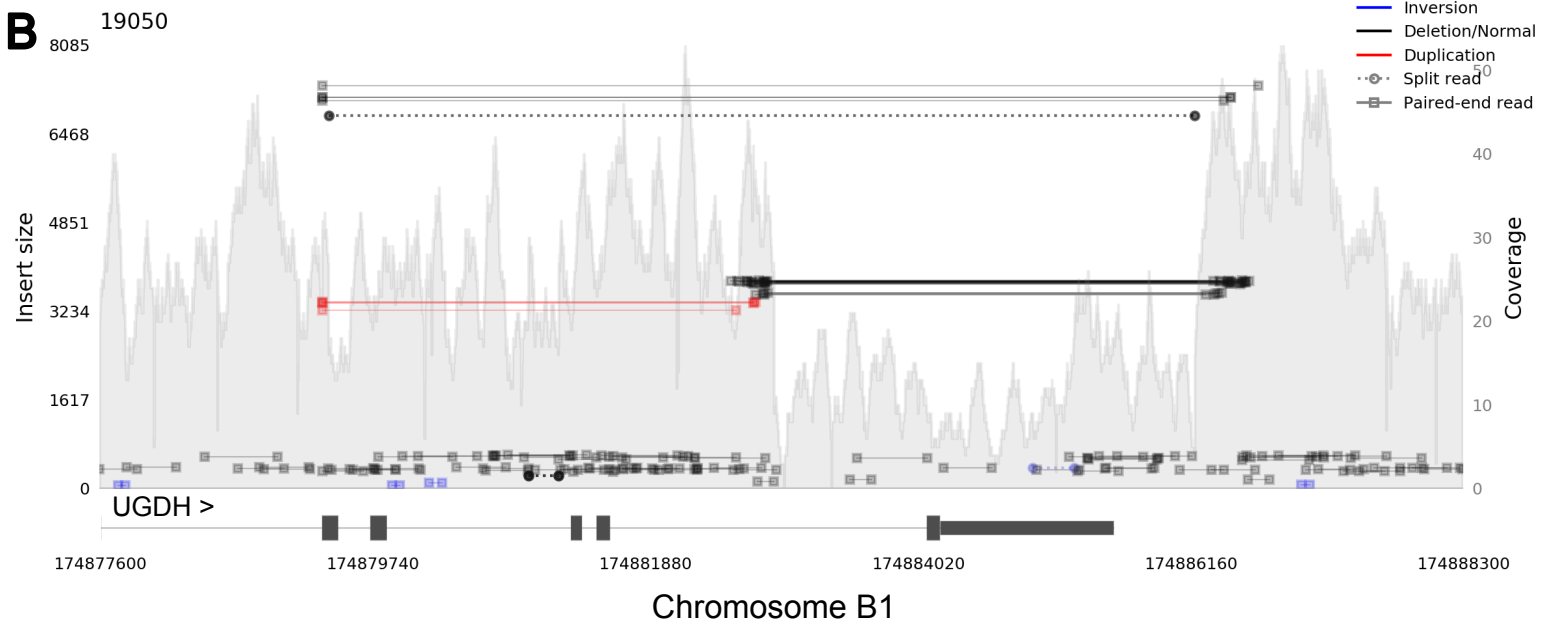
49 86. Thibaud-Nissen F, Souvorov A, Murphy T, DiCuccio M, Kitts P. Eukaryotic genome
50 annotation pipeline. *The NCBI Handbook [Internet] 2nd edition: National Center for*
51 *Biotechnology Information (US); 2013.*

- 1 87. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018.
2 Nucleic Acids Res. 2018;46(D1):D754-D61. Epub 2017/11/21. doi: 10.1093/nar/gkx1098.
3 PubMed PMID: 29155950; PubMed Central PMCID: PMC5753206.
- 4 88. Visser M, Weber KL, Lyons LA, Rincon G, Boothe DM, Merritt DA. Identification and
5 quantification of domestic feline cytochrome P450 transcriptome across multiple tissues. J Vet
6 Pharmacol Ther. 2019;42(1):7-15. Epub 2018/09/02. doi: 10.1111/jvp.12708. PubMed PMID:
7 30171610; PubMed Central PMCID: PMC6322962.
- 8 89. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The
9 Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA
10 sequencing data. Genome Res. 2010;20(9):1297-303. Epub 2010/07/21. doi:
11 10.1101/gr.107524.110. PubMed PMID: 20644199; PubMed Central PMCID:
12 PMC2928508.
- 13 90. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A,
14 et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best
15 practices pipeline. Curr Protoc Bioinformatics. 2013;43:11 0 1-33. Epub 2014/11/29. doi:
16 10.1002/0471250953.bi1110s43. PubMed PMID: 25431634; PubMed Central PMCID:
17 PMC4243306.
- 18 91. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
19 arXiv preprint arXiv:13033997. 2013.
- 20 92. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
21 Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9. Epub 2009/06/10.
22 doi: 10.1093/bioinformatics/btp352. PubMed PMID: 19505943; PubMed Central PMCID:
23 PMC2723002.
- 24 93. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework
25 for variation discovery and genotyping using next-generation DNA sequencing data. Nature
26 genetics. 2011;43(5):491.
- 27 94. Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA,
28 et al. Scaling accurate genetic variant discovery to tens of thousands of samples. BioRxiv.
29 2018:201178.
- 30 95. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant
31 call format and VCFtools. Bioinformatics. 2011;27(15):2156-8. doi:
32 10.1093/bioinformatics/btr330. PubMed PMID: 21653522; PubMed Central PMCID:
33 PMC3137218.
- 34 96. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl
35 Variant Effect Predictor. Genome Biol. 2016;17(1):122. Epub 2016/06/09. doi: 10.1186/s13059-
36 016-0974-4. PubMed PMID: 27268795; PubMed Central PMCID: PMC4893825.
- 37 97. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. Variation
38 across 141,456 human exomes and genomes reveals the spectrum of loss-of-function
39 intolerance across human protein-coding genes. BioRxiv. 2019:531210.
- 40 98. Chiang C, Layer RM, Faust GG, Lindberg MR, Rose DB, Garrison EP, et al. SpeedSeq:
41 ultra-fast personal genome analysis and interpretation. Nat Methods. 2015;12(10):966-8. doi:
42 10.1038/nmeth.3505. PubMed PMID: 26258291; PubMed Central PMCID: PMC4589466.
- 43 99. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for
44 annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the
45 genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly (Austin). 2012;6(2):80-92.
46 Epub 2012/06/26. doi: 10.4161/fly.19695. PubMed PMID: 22728672; PubMed Central PMCID:
47 PMC3679285.
- 48 100. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al.
49 Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-6. Epub 2011/01/12. doi:
50 10.1038/nbt.1754. PubMed PMID: 21221095; PubMed Central PMCID: PMC3346182.
- 51





A**B**

A**D****B****C**