Supplemental Text: **Extended methods**

Sample collection and DNA extraction

Whole blood samples were collected from dogs with iUC through either the Purdue University Veterinary Hospital, where they had biopsy-confirmed iUC, or via direct mail from owners of affected dogs. Owners of dogs not diagnosed at Purdue submitted pathology reports with diagnosis from the treating veterinarian. Blood from Shetland sheepdogs, ten years of age or older, and with no diagnosis of any type of cancer, were collected as controls. This included samples from dogs seen at the Purdue University Veterinary Hospital, those solicited from breed clubs, or dogs recruited at shows and trials. DNA was extracted from whole blood samples using cell lysis followed by phenol-chloroform, as described previously [1], with phase separation performed in 15 mL Phase Lock tubes (5-Prime, Inc. Gaithersburg, MD, USA). DNA was resuspended in a 10 mMolar (mM) tris, 0.1 mM EDTA solution and stored at -80ºC. In the case of pedigreed Shetland sheepdogs, only those that were unrelated within three generations (no common grandparents) were used in the GWAS. All dogs without pedigrees were confirmed as the reported breed using principal component clustering, distance measures and haplotype sharing with previously published pedigreed dogs as the standard [2].

Sample filtering

Samples were assessed for breed membership using principal components analysis (PCA) and distance measurements calculated in PLINK v.1.90 [3]. Significant PCs were determined using the Tracy-Widom distribution [4]. The first three PCs significantly separated the four UK herding breeds: Shetland sheepdog, Collie, Border collie, and Australian shepherd (Supplemental

Figure 1). Samples that did not cluster with their assigned breed on PCA were marked for follow-up. Samples were then added to a neighbor joining tree which included 157 previously published unrelated breeds [2]. Of the samples marked for follow-up after PCA, three putative Shetland sheepdogs did not cluster with the pedigreed Shetland sheepdogs on the multi-breed tree and were therefore excluded from the Shetland sheepdog only analysis. Samples that did not cluster with any of the UK herding breeds (two of the above Shetland sheepdogs and one Australian shepherd) were excluded from all analyses. Samples that clustered with the UK herding breeds, but not with their assigned breed, were retained for the multi-breed herding dog analyses (one Australian shepherd, two Border collies and one Shetland sheepdog). All other samples clustered with the submitted breed in both analyses (Supplementary Figure 2).

The ancestry of the four herding breed dogs, designated atypical, was determined by calculating shared haplotypes, estimated to be identical by decent (IBD) between all pairs of dogs, regardless of breed, as we described previously [2]. Breeds with average haplotype sharing above the 95% percentile of all across breed comparisons were considered possible contributors. Final relationships were visualized using Circos in R [5].

SNP genotyping and GWAS analysis

Genotypes were called in Genome Studio 2.0.4 and exported using the CanFam3.1 forward strand orientation. Data was imported into pLINK v1.9 for filtering and reformatting. Single nucleotide polymorphisms (SNPs) missing more than 10% of genotypes or with minor allele frequencies (MAF) <5% in Shetland sheepdogs were removed from the analyses.

Correction for multiple testing was applied in two ways; through Bonferroni correction based on independent loci (pair-wise $r^2<0.80$) P-Bonferroni$_{.05}=1.7x10^{-6}$, and by permuting the phenotypes

10,000 times and identifying the limit at which 95% of random associations can be excluded (P-perm 0.05=1.32e10^{-6}). Genomic inflation for the resulting association values was calculated from the p-values from uncorrected association analysis in R with λgc >1.1 indicating genome-wide inflation likely due to confounding population structure.

Additional analyses were performed on subsets of samples based on their genotype at chr13:43897196. The three sets include 29 cases and seven controls that were homozygous for the *NIPAL1*^{G256D} mutation, 11 cases and 31 controls all heterozygous for the mutation, and eight cases and 20 controls, none of which carry the mutation. The association analysis was also repeated in the full set of 101 Shetland sheepdog samples using the number of alternate alleles at chr13: 43897196, the variant within the *NIPAL1* gene, as a covariate. All analyses were corrected for underlying population structure as described above.

Whole Genome and Whole Tumor Sequencing

Paired-end data was aligned to the CanFam3.1 reference genome and variants called as described previously [6, 7]. Annotation was applied to the variant sets using snpEff v5.0 with the Ensembl gene set CanFam3.1.99 and also using VEP with the CanFam3.1plus gene annotation [8]. Polymorphic variants from the associated region were extracted using bcftools view command. Associated regions were defined as the area encompassing SNPs in LD with the most associated SNP with $r^2 \geq 0.6$ and associated with the disease with a $P \leq 1x10^{-5}$. The alignment was repeated using the reference build UU_GSD1.0 [9] and the corresponding annotation to look for missing exons.

Three tumors were sequenced at 60x coverage following the same protocols and aligned to CanFam3.1. Somatic variants were called using Mutect2 [10] with both matched normal sequence and a population-based panel of common variants to rule out germline polymorphism. In addition, RNAseq data previously published in [11] was analyzed for somatic mutation using Mutect2. Bam files were prepped following GATK best practices workflow for RNAseq short variant discovery. Variants were called using Mutect2 with matched normal RNAseq or a panel of normal RNAseq if no matched normal was available. A population-based panel of variants was used to filter out common polymorphism. All variants within the genes of interest were confirmed manually in IGV[12].

Mutation Identification

Variants identified in the 12 WGSs that followed the same pattern as the topmost associated SNPs identified in the GWAS were retained for further investigation. The position of each variant in relation to annotated genes was assessed using SNPeff and the CanFam3.1.99 gene model, VEP with the CanFam3.1plus gene model (Wucher et al, 2017), and by overlap using a bed file of exon positions from the Broad canine improved annotation v.1 [13]. Variants within coding transcripts that are predicted to alter amino acids were converted to human genome positions using lift-over (UCSC, hg19) [14, 15]. Within coding sequences, only sites for which both the DNA sequence and predicted amino acid were conserved between dogs and humans were considered for further evaluation. To confirm that no coding variants were missing due to gaps in the reference sequence we realigned the sequence reads to the UU_GSD1.0 reference sequence [16] and searched for variation within the associated gene annotation. No

new coding variants were identified. Pathogenicity of variants was predicted using CRAVAT [17, 18] and MutPred2 [19].

Putative disease-associated non-coding mutations were filtered for overlap with sites of human enhancers that had predicted local gene interactions [20], and with histone marks identifying possible enhancers and promoters determined from ChIP-Seq analysis of canine iUC tumor cell lines. The human enhancer loci were lifted to the dog assembly using UCSC lift-over [14]. Putative canine functional sites were identified by ChIP-Seq. Only those that were present in two different canine iUC cell lines and overlapped each other by at least 50% overlap and called by both SICER [21] and Macs2 [22], as described previously [23], were retained for comparisons. The allele frequency of each variant was determined using a set of 1486 dogs representing more than >200 breeds and 139 wild canids, none of which had a known diagnosis of iUC. Variants were considered potentially disease contributing if they were within a canine active enhancer, in the corresponding site of a human cis-active enhancer, and had an allele frequency of <10% in unaffected, average risk dog breeds.

Transcription factor binding sites (TFBS) were identified in regions of interest using TFBind and AIModules. Significant matches were determined using an empirically determined cut-off value for sequence similarity based on position weighted matrices from TRANSFAC R.3.4 database implemented by TFBind [24], or an odds ratio >7 and an odds ratio deficit, compared to the canonical sequence, <8, as determined by similarity calculations from position weighted matrices from the JASPAR 2022 database for the AIModule predictions [25].

Chromatin Immunoprecipitation followed by Sequencing (ChIP-seq)

ChIP-seq was performed as described in [26]. Canine iUC cell lines K9TCC_Mx and K9TCC_Ab, provided by the Knapp lab from the Werling Comparative Oncology Research Center at Purdue University, College of Veterinary Medicine, were grown to 70-80% confluency in DMEM/F12 media as previously described [27, 28]. Approximately $5x10^7$-$1x10^8$ cells were used in each assay. Cells were cross-linked using 1/10 volume of 11% formaldehyde quenched with 1/20 volume of 1M glycine after 10 minutes. Cells were washed twice with 1x PBS and harvested using a Cell Lifter (Costar #3008) in PBS. Cells were pelleted and lysed in three stages with a protease inhibitor cocktail (Roche #11873580001) added at each step. DNA in pelleted nuclei was sheared on ice using a GEX130 Ultrasonic processor (Sonics, Newtown, CT) in 3ml volumes with the following settings: power set at 30% output, 25 sec ON, then 60 sec OFF, repeated for a total of eight minutes. Magnetic beads (Invitrogen Dynabeads Protein G, Thermo Fisher Scientific, Waltham, MA), previously bound with an antibody to one of the histone modifications, were washed with phosphate-buffered saline solution (PBS), resuspended, then added to the fragmented DNA on ice and incubated at 4ºC overnight. Beads with bound DNA strands were washed five times, then rinsed with PBS. The final rinse was aspirated and 115µl of elution buffer added. DNA was eluted at 65ºC for 15 minutes, vortexing every two minutes, followed by overnight incubation on a Thermomixer R (Eppendorf, Enfield, CT). DNA was purified with phenol-chloroform after treatment with RNaseA and proteinase K.

Immunoprecipitation (IP) was repeated using 10 ug of Abcam (Cambridge, United Kingdom) antibodies to H3K4me1 (rabbit polyclonal-AbCam, #ab8895), H3K4me3 (rabbit polyclonal-AbCam, #ab8580), H3K27ac (rabbit polyclonal AbCam #ab4729) and one negative control (IgG rabbit, AbCam#46540). A 100µl input control sample was taken from shorn DNA

prior to each IP and sequenced to control for background signal in the IP. DNA was quantified using a Qubit fluorometer (ThermoFisher, Waltham, MA, USA).

The purified, IP DNA was submitted for sequencing at NISC, where paired-end libraries were made with 100bp inserts and run on an HiSeq2000 Sequencing System (Illumina). Reads were aligned to CanFam3.1 reference sequence using BWM-MEM [29] and converted to bam files using samtools 1.10 [30].

PCR conditions

PCR was carried out on a SimplyAmp thermocycler or GeneAmp PCR system 9700 (Applied Biosystems, Foster City, CA) using AmpliTaq Gold (Thermo Fisher, Waltham, MA) using standard reaction protocols with a touch-down thermocycler program starting at 65ºC annealing temperature and decreasing 0.5ºC each cycle for 20 cycles then an additional 20 cycles at 55ºC annealing temperature. Denaturing and extending temperatures were constant as per standard protocols.

Supplemental References:

1. Bell GI, Selby MJ, Rutter WI. The highly polymorphic region near the human insulin gene is composed of simple tandemly repeated sequences. Nature. 1982; 295:31-5.

2. Parker HG, Dreger DL, Rimbault M, Davis BW, Mullen AB, Carpintero-Ramirez G, et al. Genomic analyses reveal the influence of geographic origin, migration, and hybridization on modern dog breed development. Cell Rep. 2017;19(4):697-708.

3. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7.

4. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet. 2006;2(12):e190.

5. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639-45.

6. Parker HG, Whitaker DT, Harris AC, Ostrander EA. Whole genome analysis of a single scottish deerhound dog family provides independent corroboration that a SGK3 coding variant leads to hairlessness. G3 (Bethesda). 2020;10(1):293-7.

7. Plassais J, Kim J, Davis BW, Karyadi DM, Hogan AN, Harris AC, et al. Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. Nature communications. 2019;10(1):1489.

8. Wucher V, Legeai F, Hedan B, Rizk G, Lagoutte L, Leeb T, et al. FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. Nucleic Acids Res. 2017;45(8):e57.

9. Wang C, Wallerman O, Arendt M-L, Sundström E, Karlsson Å, Nordin J, et al. A new long-read dog assembly uncovers thousands of exons and functional elements missing in the previous reference. bioRxiv. 2020:2020.07.02.185108.

10. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013;31(3):213-9.

11. Parker HG, Dhawan D, Harris AC, Ramos-Vara JA, Davis BW, Knapp DW, et al. RNAseq expression patterns of canine invasive urothelial carcinoma reveal two distinct tumor clusters and shared regions of dysregulation with human bladder tumors. BMC cancer. 2020;20(1):251.

12. Robinson JT, Thorvaldsdottir H, Wenger AM, Zehir A, Mesirov JP. Variant review with the integrative genomics viewer. Cancer Res. 2017;77(21):e31-e4.

13. Hoeppner MP, Lundquist A, Pirun M, Meadows JR, Zamani N, Johnson J, et al. An improved canine genome and a comprehensive catalogue of coding genes and non-coding transcripts. PLoS One. 2014;9(3):e91172.

14. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res. 2002;12(6):996-1006.

15. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921.

16. Wang C, Wallerman O, Arendt ML, Sundstrom E, Karlsson A, Nordin J, et al. A novel canine reference genome resolves genomic architecture and uncovers transcript complexity. Commun Biol. 2021;4(1):185.

17. Douville C, Carter H, Kim R, Niknafs N, Diekhans M, Stenson PD, et al. CRAVAT: cancer-related analysis of variants toolkit. Bioinformatics. 2013;29(5):647-8.

18. Masica DL, Douville C, Tokheim C, Bhattacharya R, Kim R, Moad K, et al. CRAVAT 4: Cancer-Related Analysis of Variants Toolkit. Cancer Res. 2017;77(21):e35-e8.

19. Pejaver V, Urresti J, Lugo-Martinez J, Pagel KA, Lin GN, Nam HJ, et al. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. Nature communications. 2020;11(1):5918.

20. Nasser J, Bergman DT, Fulco CP, Guckelberger P, Doughty BR, Patwardhan TA, et al. Genome-wide enhancer maps link risk variants to disease genes. Nature. 2021;593(7858):238-43.

21. Zang C, Schones DE, Zeng C, Cui K, Zhao K, Peng W. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics. 2009;25(15):1952-8.

22. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.

23. Evans JM, Parker HG, Rutteman GR, Plassais J, Grinwis GCM, Harris AC, et al. Multi-omics approach identifies germline regulatory variants associated with hematopoietic malignancies in retriever dog breeds. PLoS Genet. 2021;17(5):e1009543.

24. Tsunoda T, Takagi T. Estimating transcription factor bindability on DNA. Bioinformatics. 1999;15(7-8):622-30.

25. Aydinli M, Liang C, Dandekar T. Motif and conserved module analysis in DNA (promoters, enhancers) and RNA (lncRNA, mRNA) using AlModules. Scientific reports. 2022;12(1):17588.

26. Stitzel ML, Sethupathy P, Pearson DS, Chines PS, Song L, Erdos MR, et al. Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. Cell Metab. 2010;12(5):443-55.

27. Dhawan D, Ramos-Vara JA, Stewart JC, Zheng R, Knapp DW. Canine invasive transitional cell carcinoma cell lines: In vitro tools to complement a relevant animal model of invasive urinary bladder cancer. Urol Oncol. 2009;27(3):284-92.

28. Shapiro SG, Knapp DW, Breen M. A cultured approach to canine urothelial carcinoma: molecular characterization of five cell lines. Canine Genet Epidemiol. 2015;2:15.

29. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589-95.

30. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2).

Supplemental Figure 1. Cladogram of dogs showing clutering by breed. Shown is the UK herding clade with the dogs included in the GWAS shaded dark green and pure-breed samples shaded light green. Affected dogs are marked with a dot near the branch. Blue stars indicated atypical examples of the reported breed from top to bottom: Shetland sheepdog 1, Border collie 1, Border collie 2, Australian shepherd 1. Grey stars represent dogs that were removed from the analysis as they are not the reported breed nor closely related.

Supplemental Figure 1

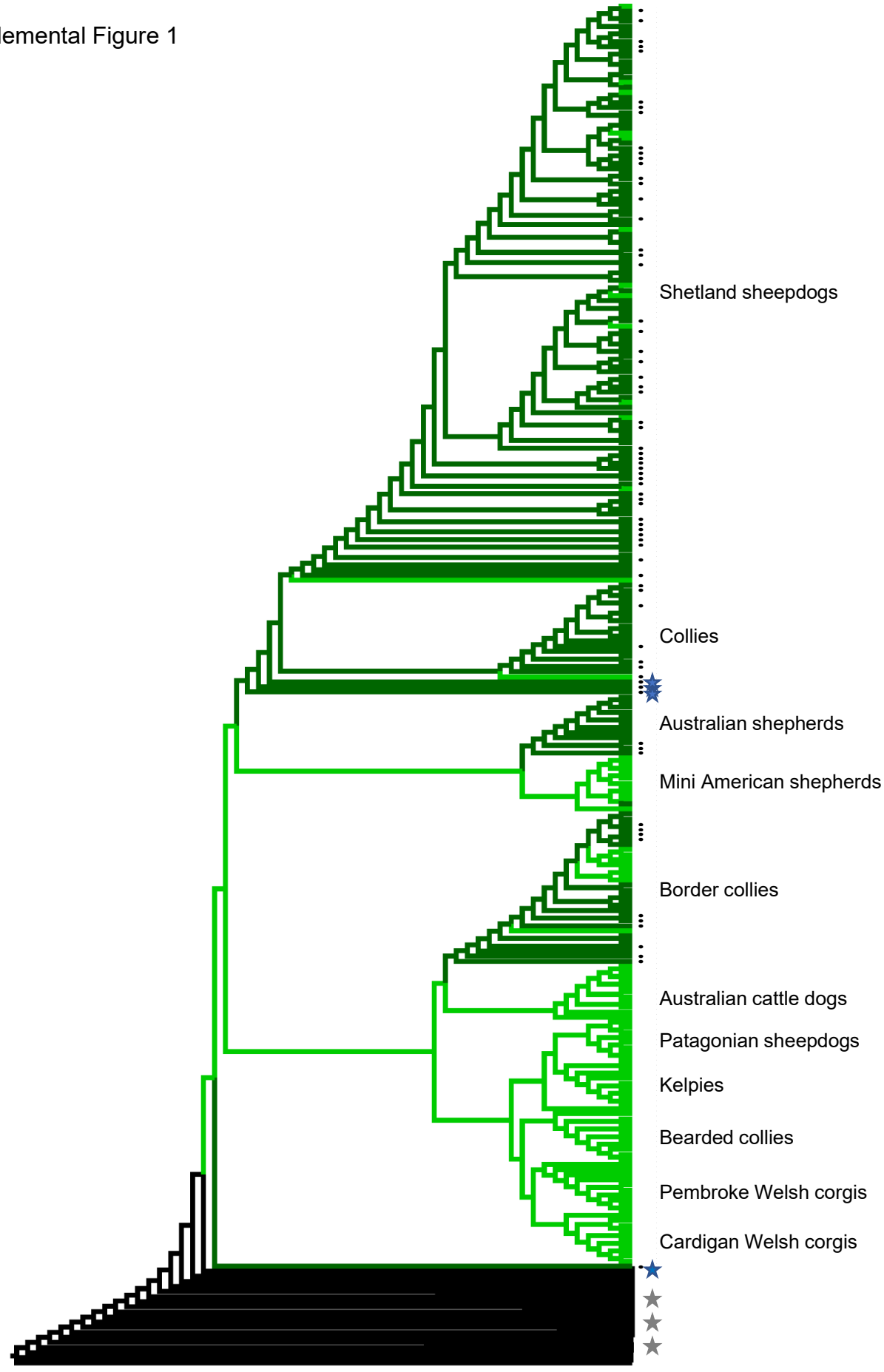
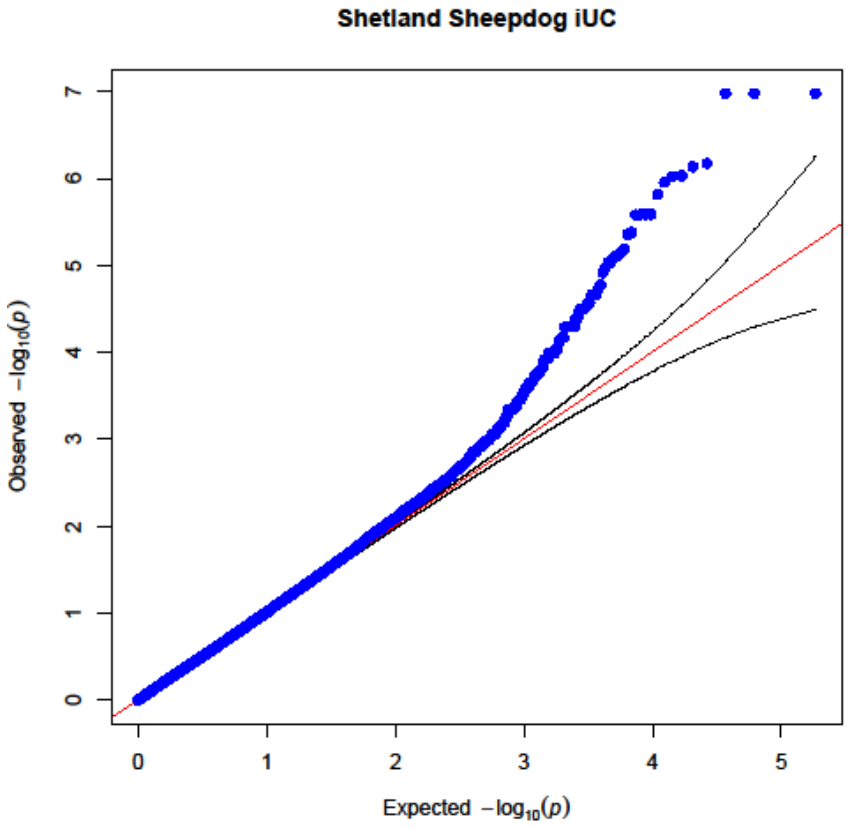


Figure S2. QQ plots corresponding to the GWAS results in Figure 1.



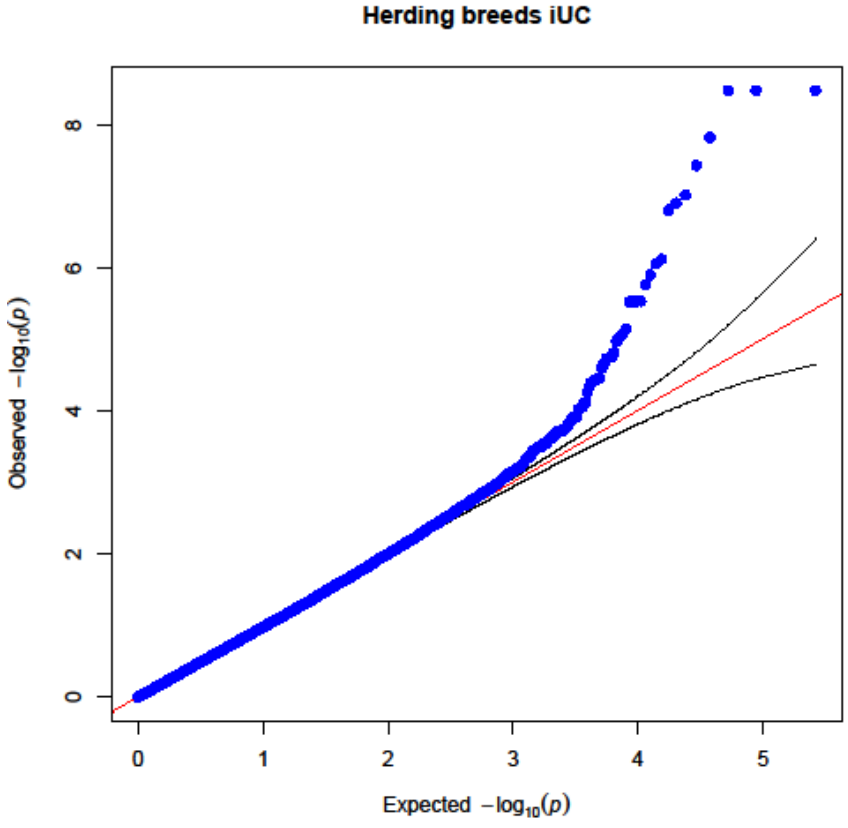


Figure S3. Manhattan plot of iUC association (black) within Shelties across chr13 overlayed with 1Mb windows of gene expression (red). The pvalues for association and excess dysregulation are plotted on the y-axis in -log scale.

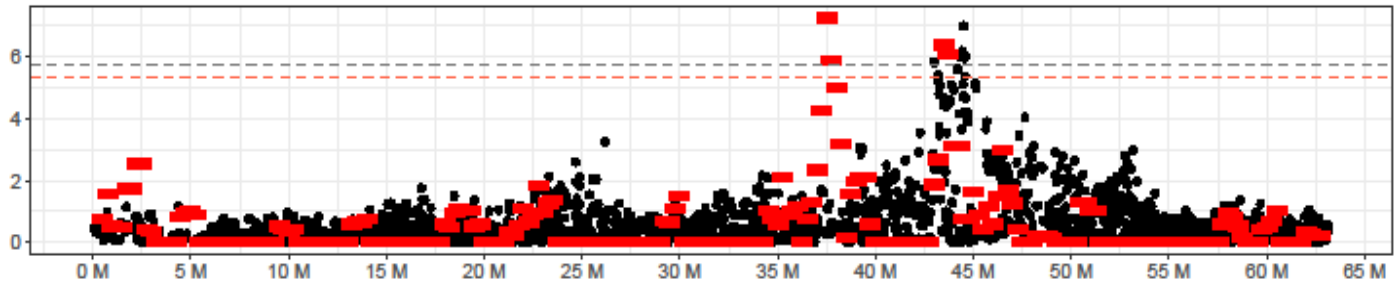
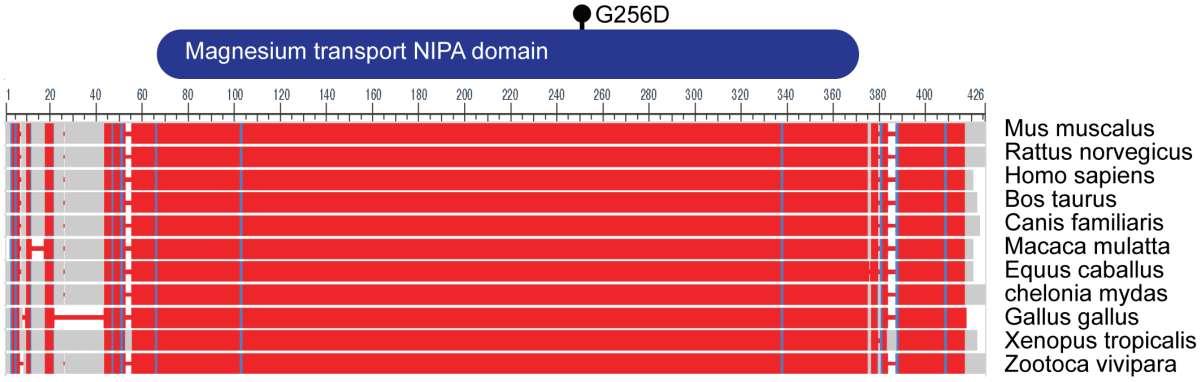
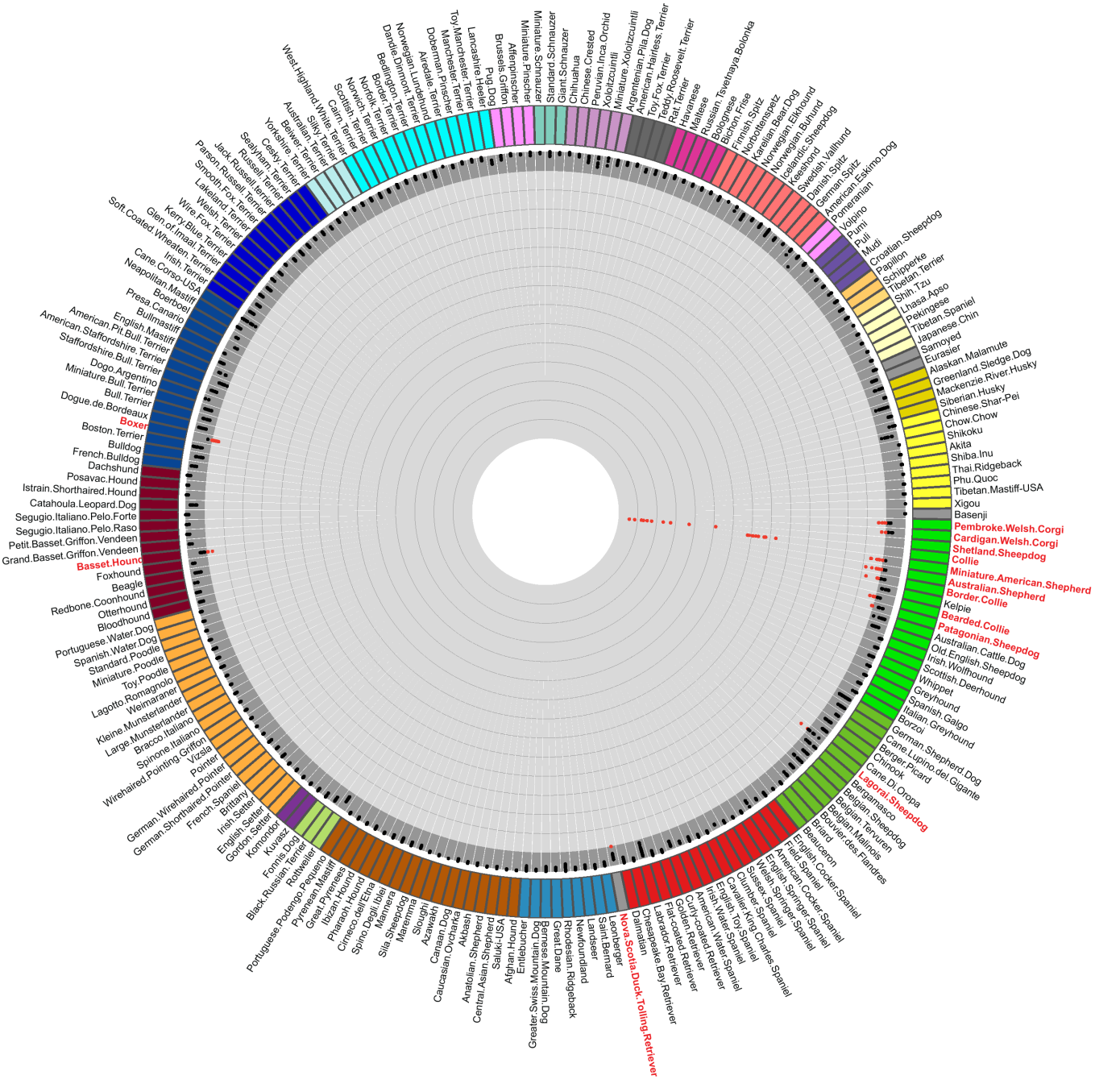


Figure S4. Multiple alignment of the NIPAL1 protein sequence in eight diverse species. Conservation of the sequence is displayed using color with red indicating highly conserved, blue is less conserved, and grey is not conserved. Alignment graphed using the NCBI Blast multiple alignment tool. The region of the gene which comprises the functional transport domain, amino acids 67 through 375, is shown above the alignment.



Supplemental Figure 5. Mixed breed analysis of atypical breed samples. Breed components are identified by excessive haplotype sharing with known pure-breeds. Haplotype sharing is calculated by pair with representatives of each breed and plotted with highest numbers toward the center of the graph. Dark grey shading indicates background sharing in which includes more than 95% of all across breed sharing. A. Atypical Border collie 1. B. Atypical Border collie 2

A. Atypical Border collie 1



B. Atypical Border collie 2

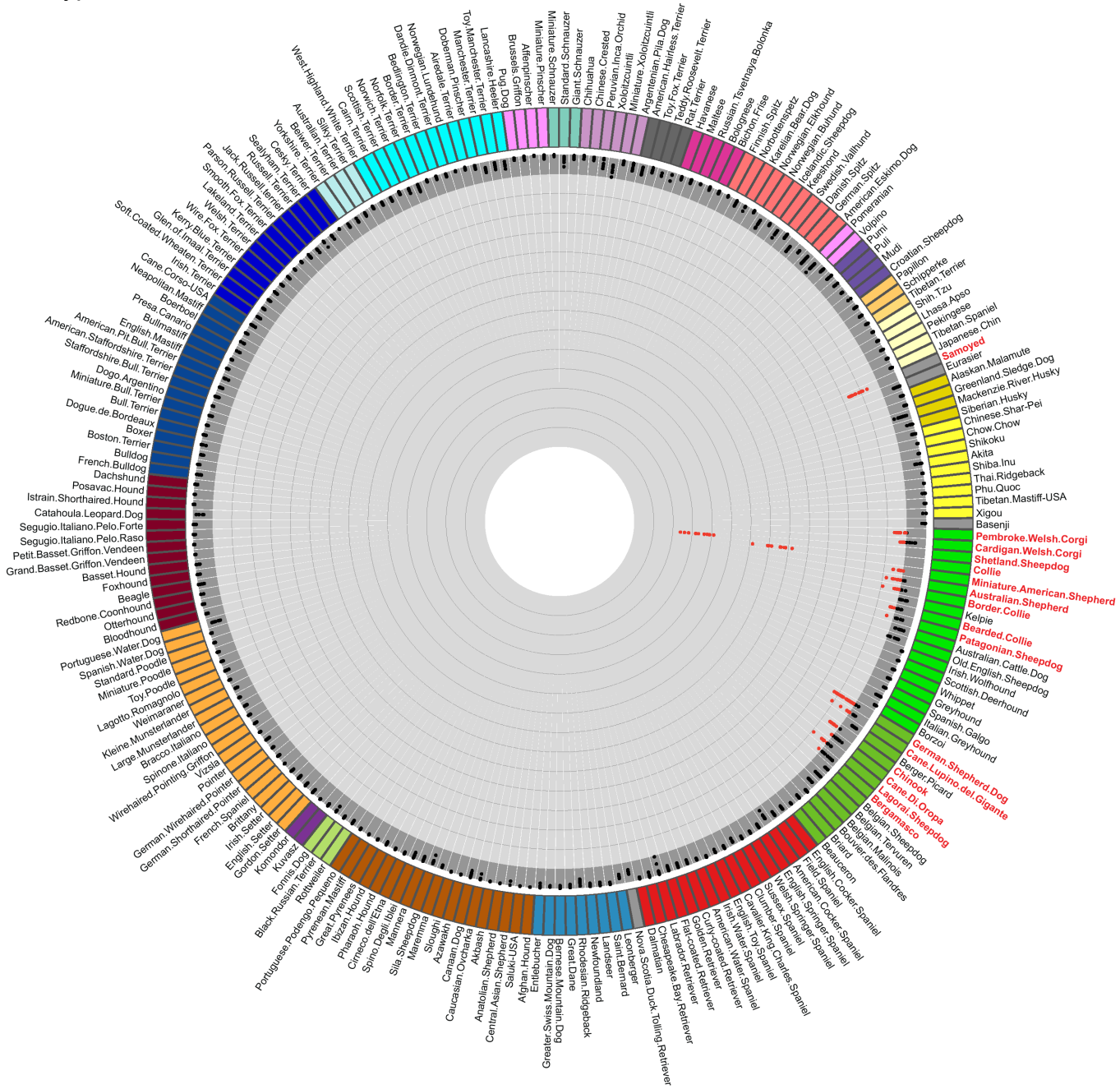


Table S3. Predicted transcription factor binding sites created by the G>A transition at chr13: 44170972 within a likely multi-gene enhancer element.

TFbind ID	JASPAR ID	TF	sense	site	avgTPM*	score
	MA1137.1	FOSL1::JUNB	+	CGATGAGAC**A**CGT	38	0.832969
	MA1137.1	FOSL1::JUNB	-	ACG**T**GTCTCATCG	38	0.838318
	MA1142.1	FOSL1::JUND	+	GATGAGAC**A**C	38	0.862818
	MA1132.1	JUN::JUNB	+	GATGAGAC**A**C	165	0.854736
	MA0099.3	FOS::JUN	-	G**T**GTCTCATC	165	0.887599
M00174		V$AP1_Q6	+	GATGAGAC**A**CG	211	0.802181
M00174		V$AP1_Q6	-	CG**T**GTCTCATC	211	0.783753
M00199		V$AP1_C	+	ATGAGAC**A**C	211	0.821283
M00199		V$AP1_C	-	G**T**GTCTCAT	211	0.880175
	MA1142.1	FOSL1::JUND	-	CG**T**GTCTCAT	38	0.858544
M00251		V$XBP1_01	-	CCTTAGTCG**T**GTCTCAT	188	0.806838
M00236		V$ARNT_01	+	TGAGAC**A**CGTCTAAGG	45	0.766628
	MA1566.1	TBX3	-	GACG**T**GTCTC	23	0.913377
M00119		V$MAX_01	+	GAGAC**A**CGTCTAAG	133	0.744371
M00119		V$MAX_01	-	CTTAGACG**T**GTCTC	133	0.744371
M00123		V$MYCMAX_02	-	TTAGACG**T**GTCT	91	0.795377
	MA0608.1	Creb3l2	+	GAC**A**CGTCT	59	0.953104
	MA0649.1	HEY2	+	GAC**A**CGTCTA	5	0.905384
MA0259.1		ARNT::HIF1A	-	AGACG**T**GT	45	0.951317
M00217		V$USF_C	+	AC**A**CGTCT	347	0.803086
M00217		V$USF_C	-	AGACG**T**GT	347	0.803086
	MA0608.1	Creb3l2	-	TAGACG**T**GT	59	0.931579

* avgTPM=average transcripts per million. For complexes, avgTPM is given for the gene with the lowest expression level. Average expression across all JUN/FOS family members is provided for AP1 binding sites (range 38-542).

Table S4. Secondary associated regions and the genes encompassed by the same.

NIPAL1 genotype	chr	position	P	region	size	Protein Coding Genes
AG	1	83750779	$1.36x10^{-6}$	chr1:83750779-84390848	640,069	ENSCAFG00000001773
AG	4	87441098	$4.52x10^{-7}$	chr4:87440198-87540261	100,063	FBXL7
AG	9	22495330	$6.23x10^{-7}$	chr9:22381987-23201516	819,529	WIPF2, RAPGEFL1, CASC3, MSL1, NR1D1, THRA, MED24, CSF3, PSMD3, GSDMA, LRRC3C, ORMDL3, GSDMB, ZPBP2, IKZF3, GRB7, **ERBB2**, MIEN1, PGAP3, PNMT, ENSCAFG00000028772, STARD3, PPP1R1B, NEUROD2, CDK12, MED1, FBXL20, ENSCAFG00000016470, STAC2, RL19_CANFA, ARL5C
GG	21	33595359	$1.67x10^{-7}$	chr21:31405230-33951574	2,546,344	OR5P2, OR10A6, ENSCAFG00000006866, ENSCAFG00000025199, NLRP10, EIF3F, TUB, RIC3, LMO1, STK33, TRIM66, RPL27A, ST5, AKIP1, C11orf16, ASCL3, TMEM9B, NRIP3, SCUBE2, TMEM41B, ENSCAFG00000007241, IPO7, ENSCAFG00000007332, ZNF143, **WEE1**, SWAP70, SBF2, ENSCAFG00000007532, ADML, AMPD3, ENSCAFG00000008138, ENSCAFG00000008290, AMPD3, RNF141, LYVE1, MRVI1, CTR9
AA	21	34085269	$5.64x10^{-7}$	chr21:33806003-34295586	489,583	MRVI1, CTR9, EIF4G2, ZBED5_CANFA
AA	28	34513379	$1.23x10^{-9}$	chr28:34513379-34747318	233,939	ENSCAFG00000032418
Positions giving in CanFam3.1. Genes extracted from the Broad improved annotation v1 Protein Coding Genes on the UCSC genome browser. Unknown proteins omitted.						

Table S5. Predicted pathogenicity of coding mutations.

	13:43897196		9:22775561	9:22775767	21:32843843
Pathogenicity					
Gene	NIPAL1		ERRB2	ERRB2	WEE1
Type	germline		somatic	somatic	somatic
Amino Acid Change	Gly256Asp		Asp277Tyr	Asp251His	Pro174Arg
CADD	**26.9**		**25.1**	**29.6**	**26.7**
DANN	**0.998**		**0.939**	**0.953**	**0.997**
ClinPred	**0.998**		**0.766**	**0.967**	**0.996**
fathmmMKL	**0.993**		**0.927**	**0.992**	**0.953**
PhD-SNP	**0.997**		**0.846**	**0.988**	**0.988**
PROVEAN	**0.932**		0.418	**0.899**	**0.924**
VEST	**0.912**		**0.688**	**0.925**	0.32
MutPred2	**0.85**		**0.66**	**0.697**	0.428
PolyPhen2	**1**		**0.988**	**0.975**	**1**
Cancer specific function					
CScape	**0.768**		**0.802**	**0.969**	**0.841**
CHASMplus	0.017		0.243	0.438	0.079
(p-value)	(0.691)		**(0.027)**	**(7.96E-04)**	(0.255)
CHASMplus bladder cancer	0.06		0.452	0.36	0.039
(p-value)	(0.234)		**(4.91E-05)**	**(0.000303)**	(0.375)
COSMIC count	317		2966	2966	377
COSMIC count urinary tract only	4		299	299	26

Significant predictions are bolded. The variant effect predictors DANN, ClinPred, fathmmMKL, PhD-SNP, PROVEAN, VEST, MutPred2, and PolyPhen2 score each variant on a scale from 0-1 with scores >0.5 indicating likely damaging or pathogenic mutations. CADD scores are provided on a phred scale with scores >20 in the top 1% of deleterious variation and >30 in the top 0.1%.

CScape uses a 0-1 scoring system to estimate oncogenic potential in mutations with 1 being most likely. The significance of cancer driver scores by CHASMplus are adjusted based on the frequency of the mutation therefore p-values are included with the score. COSMIC counts are all mutations that have been reported within the gene in the COSMIC database.