

**Pugh et al** Arginine to glutamine variant in olfactomedin like 3 (*OLFML3*) is a candidate for severe goniodysgenesis and glaucoma in the Border Collie dog breed.

#### ***Supplemental Materials and Methods***

##### *DNA samples*

The majority of the DNA samples were from buccal cells collected by owners using Isohelix DNA Buccal Swabs (Isohelix, Harrietsham, Kent, UK), Performagene (DNA Genotek, Ottawa, ON, Canada) or nylon brushes supplied by the Animal Health Trust (AHT) Genetic Services department (<http://www.aht.org.uk>). Samples were mailed to the laboratory and stored at room temperature until extraction. Samples on AHT brushes were lysed with a standard cell lysis buffer and treated with proteinase K at 50 C overnight prior to DNA extraction. DNA was extracted using the manufacturer's protocol and supplied reagents (Performagene, Isohelix), using the Maxwell16 DNA purification instrument and kits (Promega, Sydney, Australia), using Bioline Isolate II Genomic DNA kit (Bioline (Aust), Sydney, Australia) or using a phenol-chloroform extraction procedure. For the latter, 600 µL of phenol/chloroform (1:1 v:v) were added to 600 µL cheek cell lysate. The samples were mixed gently on a rotator for 15 min then centrifuged for 10 min at top speed in a microfuge. 600 µL chloroform isoamyl alcohol (24:1 v:v) were added to the aqueous layer. The sample was mixed for 5 min and then spun for 5 min as above. 600 µL of isopropanol were added to the aqueous layer. The tube was inverted several times over a period of 10 min and then centrifuged for 5 min as above. The supernatant was discarded and the pellet washed with 300 µL 70% ethanol for 2 min. The tube was then centrifuged for 3 minutes. The supernatant was removed and the pellet air dried for 30 minutes and then resuspended in 60 µL TE buffer. Some samples were

supplied from residual blood following a clinical test and DNA was extracted using a Blood and Tissue Kit (Qiagen, Manchester, UK).

#### *Genome-Wide Association Study (GWAS)*

DNA was sent to Edinburgh Genomics, University of Edinburgh, Edinburgh, UK, for genotyping using the Illumina 173k CanineHD Whole-Genome Genotyping Bead Chip (Illumina, San Diego, CA, USA), following the manufacturer's instructions. Results were filtered in PLINK v1.07 [1] initially to remove individuals that had more than 10% of missing genotypes and then SNPs that had rates of genotyping < 0.95, had minor allele frequency < 0.05 in this population or deviated from the Hardy-Weinberg equilibrium in the controls with P value of less than 0.0001. The SNV data are available via the University of Edinburgh data repository at <http://dx.doi.org/10.7488/ds/2426>.

Cases and controls were defined as described in the main text Methods section. Affected dogs that had a lower grade goniodysgenesis (grade one to three under the previous scheme), those said to be affected without a grade assigned and those that were described as 'mild' were excluded from the GWAS.

#### *Whole genome sequencing*

Whole genome sequencing (30X coverage) of DNA from three dogs with glaucoma, three with severe goniodysgenesis and three unaffected animals was performed by Edinburgh Genomics using the Illumina HiSeq X platform. One sample was extracted from blood; the remainder were from buccal cells. Reads were processed using Trimmomatic v0.35 [1] with parameters TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:100 and then mapped to the CanFam3.1 reference genome (<ftp://ftp.ensembl.org/pub/release->

85/fasta/canis\_familiaris/dna/Canis\_familiaris.CanFam3.1.dna.toplevel.fa.gz, downloaded 5th September 2016) using BWA-MEM v0.7.8 [2] with default parameters. After filtering alignments using SAMtools view v1.3 [3] with parameters -q 30 and -F 12, variants where the sequence differed from the reference sequence were identified using the GATK v3.5 best practices pipeline [4] in conjunction with canine SNVs from dbSNP build 140 (ftp://ftp.ncbi.nlm.nih.gov/snp/organisms/dog\_9615/VCF/, downloaded 5th September 2016) and the set of Ensembl canine SNVs (ftp://ftp.ensembl.org/pub/release-85/variation/vcf/canis\_familiaris/Canis\_familiaris.vcf.gz, downloaded 5th September 2016). The sequence was visualised on the Integrative Genomics Viewer ([5]; <http://software.broadinstitute.org/software/igv/>). The significance of variants was assessed using PolyPhen-2 (<http://genetics.bwh.harvard.edu/pph2/>), Mutation Taster (<http://www.mutationtaster.org>), Provean (<http://provean.jcvi.org>) and SIFT (scores obtained from Provean analysis). Structural variants were ascertained with DELLY 0.7.7 [6] and GRIDSS 2.0.1 [7] with default options. Precise unfiltered breakpoints from DELLY were cross-referenced with breakends called by GRIDSS and summarised with a python script. Gene annotations were obtained via snpEff 4.3t [8] using the canFam3.1.86 database. BAM files are available via the European Nucleotide Archive, accession numbers ERS2643230 and ERS2643240-7.

#### *Transcriptomic analysis of candidate region genes*

The FANTOM5 Consortium has produced transcriptomic analysis of the development over time of multiple organs of the mouse embryo using cap analysis of gene expression (CAGE; genome-scale 5' RACE ([9]; [9-11])). A time course of mouse eyeball development from embryonic day 12 to adult was examined using this FANTOM5 database

(<http://fantom.gsc.riken.jp/zenbu>). The samples accessed are shown in **Supplemental Table S1**. RLE normalised expression values were downloaded using the FANTOM5 Table Extraction Tool (<http://fantom.gsc.riken.jp/5/tet/>). We also used the FANTOM5 data to obtain RLE normalised expression data for human eye-related samples. RNA sequencing data for chicken tissues and cells were obtained from BioGPS ([http://biogps.org/dataset/BDS\\_00031/chicken-atlas/](http://biogps.org/dataset/BDS_00031/chicken-atlas/)). RNA sequencing data for the developing chicken eye were generated for another project (Rainger et al, in preparation). Briefly, whole eyes, optic fissure and ventral eye region were dissected at embryonic days 5, 6 and 7 and total RNA was extracted using TriZol (ThermoFisher). Poly(A)-tailed mRNA was isolated and purified from 1µg of each total RNA sample. Purified mRNA was then concentrated and fragmented before reverse-transcription and purification, and the resulting cDNA was ligated to Ion-compatible barcoded adapters and then amplified (15 cycles). Adapter-ligated libraries were then purified and size-selected to provide barcoded total RNA libraries for each sample. Libraries were quantified and equimolar pools of 3 samples were brought together for template preparation. Sequencing was performed on the Ion Proton using Ion PI Chips (ThermoFisher). Initial pseudoalignment identified 30,265 transcripts for all tissue-types and stages aligning with unique Ensembl chicken transcript (Gallus\_gallus-5.0). We then performed quantitative analyses using Kallisto [12] to reveal transcript per million (TPM) values for *OLFML3* at each embryonic stage.

#### *Analysis of canine OLFML3 gene*

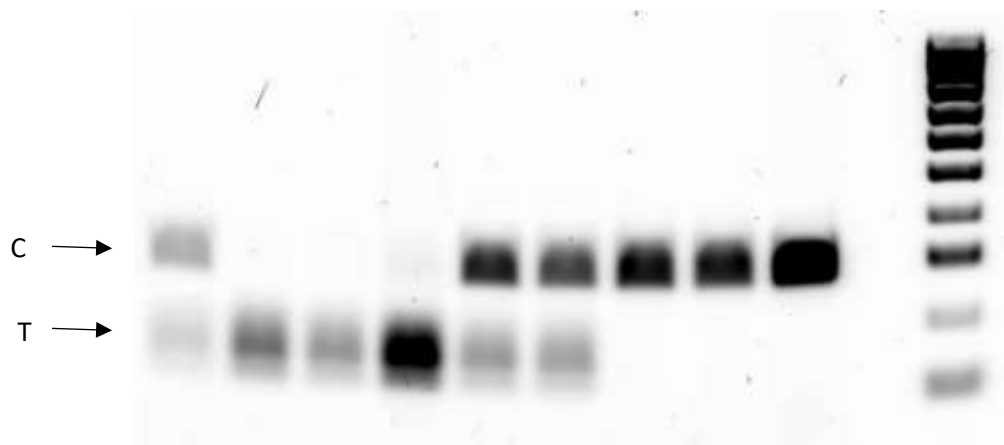
Primers for PCR to amplify the canine *OLFML3* region containing a missense variant were designed using Primer3 (<http://primer3.ut.ee/>). Two different primer sets were used: *OLFML3* set 1 (F: AATGGGGCCAACAGAGAAGA; R: GAGTTGCAACGTGTTCTCCA) and *OLFML3*

set 2 (F: TGAAGATCCTGAAGCGGTTTG; R: AGGTCAATGTAGGTGTCGGC). A number of individuals were genotyped with both sets of primers and gave the same result. The animals genotyped using the whole genome sequence also gave the same result using these primers. The PCR contained 0.5 - 1  $\mu$ M primers, 200  $\mu$ M deoxynucleotides, 1 unit of *Taq* polymerase (Roche, Mannheim, Germany) and 1 ng/ $\mu$ L template DNA in the manufacturer's *Taq* polymerase buffer. PCR conditions were 3.5 min at 94 C followed by 30 cycles of 0.5 min at 94 C, 0.5 min at 63 C and 1 min at 72 C. This was followed by 7 min at 72 C and the samples were held at 10 C. PCR products were cleaned using the Life Technologies ChargeSwitch system (Thermo Fisher Scientific, Waltham. MA. USA) and sent to Edinburgh Genomics for chain termination sequencing. Some PCR products were sent to the AGRF for clean up and sequencing. Sequences for analysis of nucleotide and amino acid conservation across species were downloaded from Ensembl (<http://ensembl.org>) and alignment was performed using Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>).

#### *Analysis of canine CDK2AP1 gene*

Primers for PCR to amplify the canine *CDK2AP1* region containing a missense variant were designed using Primer3 (<http://primer3.ut.ee/>). The primers were F: CCCACCCTCCACCAGTATG and R: AACCTTCCACGCTGATAGGC. The PCR contained 0.5  $\mu$ M primers, 1 U MyTaq DNA polymerase (Bioline, Sydney, Australia) and 1 ng/  $\mu$ L template DNA in the MyTaq Reaction Buffer. PCR conditions were 3.5 min at 94 C followed by 30 cycles of 0.5 min at 94 C, 0.5 min at 63 C and 1 min at 72 C. This was followed by 7 min at 72 C and the samples were held at 10 C. The amplicon contained the sequence GACCC in the wild type allele which was converted to GACTC in the risk allele, creating a cut site for the

restriction enzyme *Hinf*1 (GATC). Therefore the wild type allele would be uncut but the variant allele would be cut into two fragments of 135 and 165 bp. After amplification, 10 µL of the PCR product was digested in a total volume of 30 µL, containing 20 U *Hinf*1 (New England Biolabs, Massachusetts, USA) in the CutSmart buffer and incubated for 2 hours at 37 C. 10 µL of this reaction were run on a 2% agarose TAE gel and visualised with SYBR-Safe (Invitrogen). An example of the gel used for genotyping is shown in the figure.



**Figure.** Example of a gel used for genotyping the canine *CDK2AP1* gene. The two alleles are indicated at the left. The Bioline 100 bp size ladder was used.

Samples of all three genotypes were sent to the Australian Genome Research Facility (AGRF), Brisbane, Australia for clean up and chain termination sequencing.

## References for Supplemental Materials and Methods

1. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data.** *Bioinformatics* 2014, **30**(15):2114-2120.
2. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.** *Bioinformatics* 2009, **25**(14):1754-1760.
3. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics* 2009, **25**(16):2078-2079.
4. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J *et al*: **From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline.** *Curr Protoc Bioinformatics* 2013, **43**:11 10 11-33.
5. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP: **Integrative genomics viewer.** *Nat Biotechnol* 2011, **29**(1):24-26.
6. Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO: **DELLY: structural variant discovery by integrated paired-end and split-read analysis.** *Bioinformatics* 2012, **28**(18):i333-i339.
7. Cameron DL, Schroder J, Penington JS, Do H, Molania R, Dobrovic A, Speed TP, Papenfuss AT: **GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly.** *Genome Res* 2017, **27**(12):2050-2060.
8. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM: **A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w<sup>1118</sup>; iso-2; iso-3.** *Fly (Austin)* 2012, **6**(2):80-92.
9. Arner E, Daub CO, Vitting-Seerup K, Andersson R, Lilje B, Drablos F, Lennartsson A, Ronnerblad M, Hrydziuszko O, Vitezic M *et al*: **Gene regulation. Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells.** *Science* 2015, **347**(6225):1010-1014.
10. Forrest AR, Kawaji H, Rehli M, Baillie JK, de Hoon MJ, Haberle V, Lassmann T, Kulakovskiy IV, Lizio M, Itoh M *et al*: **A promoter-level mammalian expression atlas.** *Nature* 2014, **507**(7493):462-470.
11. Summers KM, Hume DA: **Identification of the macrophage-specific promoter signature in FANTOM5 mouse embryo developmental time course data.** *J Leukoc Biol* 2017, **102**(4):1081-1092.
12. Bray NL, Pimentel H, Melsted P, Pachter L: **Near-optimal probabilistic RNA-seq quantification.** *Nat Biotechnol* 2016, **34**(5):525-527.