



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e. g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

The genetic basis of previously unexplained aniridia

Hildegard Nikki Hall

Thesis submitted for the degree of Doctor of Philosophy

University of Edinburgh

2025

I declare that this thesis is my own work. Where the contribution of others is included, this is clearly indicated and acknowledged.

The work has not been submitted for any other degree or professional qualification.

Hildegard Nikki Hall, December 2024

Abstract

Classical aniridia [MIM 106210] is a rare, panocular malformation characterised by a spectrum of iris and foveal hypoplasia, later complicated by cataracts, lens subluxation, glaucoma and keratopathy. It is overwhelmingly associated with *PAX6* [MIM 607108] haploinsufficiency. *PAX6* is a highly-conserved transcription factor that orchestrates eye development. Whilst the molecular genetic basis of aniridia and *PAX6* disruption is broadly known, there remain unexplained areas that have guided the work in this thesis.

PAX6 missense variants can partially or fully phenocopy haploinsufficiency, causing classical aniridia. Why they also cause rare “worse than null” phenotypes is unclear. In order to study the effect of variant class on phenotype, I performed systematic, detailed ocular phenotyping followed by cluster analysis of *PAX6* eye malformations. This compared all available missense variant cases (n=161) with a size-matched control of likely gene-disruptive (LGD) cases. Whilst LGD variants were almost universally associated with classical aniridia, the phenotypic spectrum of missense cases was wider and appeared to cluster separately, ranging in severity from isolated foveal hypoplasia (hypomorphic) to bilateral severe microphthalmia (worse than null). Recurrent substitutions of key residues in the DNA-binding paired domain of *PAX6* accounted for the most severe cases, notably p.Ser54Arg and p.Asn124Lys. The impact of these variants was investigated in vitro, studying the binding of purified *PAX6* paired domain to well-characterised *PAX6* DNA targets, the LE9 and SIMO enhancers. Electromobility shift assays showed that binding was altered by p.Ser54Arg and p.Asn124Lys compared to wild type, though (unlike classical aniridia variants) not totally abrogated. Their impact differed between the two DNA targets studied. Whilst not explaining all of the observed genotype-phenotype correlation, these data suggest that the variants do alter protein-DNA interactions - consistent with results of protein modelling - and could affect the stoichiometry of *PAX6* target gene activation.

I noted a striking similarity between the rare, severe *PAX6* missense cases above and those seen in six families with heterozygous variants in *MAB21L1* [MIM 601280], all

resulting in Arg51 and Phe52 substitutions. All had aniridia and/or microphthalmia, 3/6 had both, and 2/6 had anomalous optic discs. *MAB21L1* is involved in eye development downstream of *PAX6*, but was not known to cause aniridia. To characterise this further, ocular phenotyping of a CRISPR-Cas9 gene-edited mouse strain carrying *Mab21l1* p.Arg51Leu was performed. Adult mice heterozygous for this allele had a highly-penetrant excavated optic nerve anomaly. The phenotype overlaps with but is milder than the human counterpart. The homozygous animals showed severe microphthalmia. There are still questions to be answered regarding the mechanisms in both human and mouse, which may differ, but the human genetic data strongly support this being a new aniridia-associated gene with a gain of function effect.

A small proportion of aniridia is still genetically uncharacterised, indicating that there could be novel ways of inactivating *PAX6* or additional disease loci. To address this, I undertook analysis of 37 unrelated mutation-negative individuals with classical aniridia (n=51 including relatives) using paired-end, short-read whole genome sequencing (WGS) technology. Following an in-house analysis pipeline, causative variants were identified in 22/37 cases (60%). 15/22 were structural variants, including: chromosome 11 inversions, one separating *PAX6* from its critical downstream enhancers; deletions of this downstream enhancer region; whole/partial gene deletions of *PAX6*, and lastly three *FOXC1* gene deletions. The remainder were intragenic *PAX6* single nucleotide variants, including two deep intronic splice variants. RT-PCR analysis of RNA from patient-derived lymphoblastoid cell lines was performed in a subset, confirming missplicing in vitro. Notwithstanding historical differences in *PAX6* screening, it is remarkable that over half the cohort had a disease mechanism involving either *PAX6* or its regulatory region, highlighting the unusually strong disease-gene link.

In summary, this thesis fills gaps in our understanding of aniridia genetics through whole genome diagnosis of unsolved cases, underlining the importance of the wider *PAX6* locus and adding to the known variants which inactivate it, as well as through characterisation of severe aniridia-microphthalmia phenotypes caused by rare missense variants in *PAX6* and in the novel aniridia-associated gene *MAB21L1*.

Lay abstract

Classical aniridia is a rare disease, present at birth, in which the coloured iris and other structures within the eye have not developed properly. Individuals with this condition tend to develop complications such as a cloudy lens (cataract), instability of the lens, glaucoma (due to high eye pressure), and progressive clouding of the clear cornea at the front of the eye. These issues typically result in sight impairment. Whilst aniridia itself means ‘lack of iris’, the term is also used to refer to the condition overall.

The cause of aniridia is genetic, whether it is inherited from a parent or a new gene change in the baby’s DNA. Genetic variants that cause disease (also known as mutations) can simplistically be considered spelling errors in our genetic code. Variants in a specific area of the DNA, a gene called *PAX6*, account for the overwhelming majority of aniridia cases. Although much is known on a molecular level regarding aniridia and *PAX6*, there remain unexplained areas that have guided the work in this thesis.

First, I explored the link between different types of variant in *PAX6* (genotype) and the variety of eye malformations that it can cause (phenotype) - which is not limited to just aniridia. I performed a detailed analysis of phenotypic features (e.g. lack of iris, cataract, cloudy cornea) of 322 eye malformations caused by *PAX6* variants, helping to define the spectrum of eye malformations that can be caused by this gene. I studied the effect of types of variant, and found that missense variants (which result in a switch to a different amino acid, rather than simply inactivating the gene) tend to cause more diverse eye problems, including some of the severest cases featuring small malformed eyes (microphthalmia). Experimental work suggests this is in part due to altered protein-DNA binding.

Next, I described six families who have been found to have variants in a new aniridia-associated gene, *MAB21L1*. Changes in this gene appear to cause aniridia and/or microphthalmia (in all six families): in half of the families both of these were present, and one third also had an abnormal optic nerve appearance. Mice, bred using CRISPR technology to have the same genetic variant seen in human families, also have an eye

malformation, though it does not exactly mimic the human version. Unlike *PAX6*, this gene is not well understood yet, and this is an active area of research.

Lastly, in order to try to find the genetic cause of the previously-unexplained aniridia cases, I undertook analysis of the DNA of 37 mutation-negative individuals with classical aniridia (51 including relatives) using whole genome sequencing technology, which sequences an individual's entire genetic code. I was able to find a genetic cause for the aniridia in 60% of cases (22/37). These variants included a variety of ways (some novel) of disrupting either the *PAX6* gene or the nearby area of the genetic code which are essential for normal function of the gene in the eye. Three of the cases had variants affecting a different gene, *FOXC1*, known to cause a congenital eye condition that can appear very similar to aniridia.

In summary, this thesis fills in gaps in our understanding of aniridia, which may lead to improved genetic diagnosis rates. This in turn could open treatment avenues, given that there is the potential for future genetic therapies in aniridia.

Acknowledgements

My wholehearted thanks to the colleagues, collaborators, patients, friends and family who helped with this work. I am grateful to the Wellcome Trust for funding this research and to ECAT, especially Jo Ness, Andrew Jackson and Sarah Walmsley.

Thank you to my Principal Supervisors, David FitzPatrick and then Wendy Bickmore, for superb guidance and inspiration. Many thanks to Ian Jackson and Joe Rainger, second supervisors. Thank you to Veronica van Heyningen, without whom this would not have happened. Thanks to Kathy Williamson, walking *PAX6* encyclopaedia, for teaching me human genetics over working coffees. Thanks to Hemant Bengani, Shipra Bhatia, Morad Ansari and Jacqui Rainger for their generous help and *PAX6* knowledge. Thank you to Liusaidh Owen, Brianda Hernandez-Moran, Mark Handley, Alex McDonnell, Angela Sandilands and other FitzPatrick lab members.

Thank you to Lisa McKie, mouse eye guru, Amy Findlay, Sally Cross, Zuhair Mustafa and Roly Megaw. I appreciated the IJJ lab mouse eye meetings and Christmas dinners, the latter nicely spaced from FitzPatrick Christmas dinners in a swanky restaurant in July. Thanks to the E2/C1 lab members, for scientific help and friendship: MJ Kempen, Kiko Sanchez-Luque, Thomas Williams, Martin Reijns, Carole-Ann Martin, Paula Carroll, Olga Murina, Žy Tarnauskaite, Patty Heyn, Clare Logan, David Parry, Maggie Macdonald, Maarten van den Ancker, Nick Younger, Kirsty Uttley, Emma Clarke and Andrea Robertson. Thanks to my PhD class and crossword compatriots including Kathryn Jackson-Jones, Lloyd Bibby, Juliet Luft, Lauren Kane and Becka Hughes.

Thanks to the IGC bioinformatics team, particularly Alison Meynert, Mike Halachev, Lana Talmane and Graeme Grimes. Thanks to Joe Marsh and Rolando Hernandez Trapero for protein structure expertise. Thank you to the BRF staff and IGC core teams, particularly Lorraine Rose, Anna Thornburn, Stephen Brown, Jeffrey Joseph, Allyson Ross, Craig Nicol, Stewart McKay, Maggie Arbuckle and Katie Browne.

Thank you to my family for all their support, especially to Lindsey and Nick who have aniridia. My father revealed himself to be a generous and eagle-eyed proof reader. Lastly, thank you for everything to my son Fergus and husband David.

Abbreviations

aa	amino acid
AAK	aniridia-associated keratopathy
aCGH	array comparative genomic hybridisation
ACMG	American College of Medical Genetics and Genomics
ACTA2	actin alpha 2
AGCS	Association for Clinical Genomic Science
ARA / ARS	Axenfeld-Rieger Anomaly/Syndrome
AS / ASD	anterior segment (dysgenesis)
BL	bilateral
BMP	bone morphogenetic protein
bp	base pairs
BRF	Biomedical Research Facility
CA	classical aniridia
CAGE	cap analysis of gene expression
CCO	congenital corneal opacification
cGAS	cyclic guanosine monophosphate
ChIP	chromatin immunoprecipitation
CNV	copy number variation/variant
CRE	<i>cis</i> -regulatory elements
CRISPR	clustered regularly interspaced short palindromic repeats
DDD	Deciphering Developmental Disorders study
DNA / RNA	(deoxy)ribonucleic acid (ss-, single-stranded; ds-, double-stranded)
DRR	downstream regulatory region (of <i>PAX6</i>)
ECAT	Edinburgh Clinical Academic Track, University of Edinburgh
EMSA	electrophoretic mobility shift assay (also, gel-shift assay)
ExAC	Exome Aggregation Consortium (predecessor to gnomAD)
FAM	carboxyfluorescein label
FID	family identification number
FISH	fluorescence <i>in situ</i> hybridisation
FLAG	FLAG epitope tag (octapeptide DYKDDDDK)
FOX	forkhead box (e.g. FOXC1)

gnomAD	genome Aggregation Database
GRCh38	Genome Reference Consortium human build 38
HD	homeodomain
HGUeye	MRC Human Genetics Unit eye malformations cohort
IGC	Institute of Genetics and Cancer (formerly, Institute of Genetics and Molecular Medicine), University of Edinburgh
IGV	Integrative Genome Viewer
IOP	intraocular pressure
iPSC	induced pluripotent stem cell
ITPR1	inositol 1,4,5-triphosphate (IP3) receptor type 1
Kd	equilibrium dissociation constant
LCL	lymphoblastoid cell line
LE	left eye
LE9	LE9 lens enhancer upstream of <i>PAX6</i>
LGD	likely gene-disrupting (-disruptive, -disruption)
logMAR	logarithm of the minimal angle of resolution (visual acuity)
LOVD	Leiden open variation database
LRS	long read sequencing
MAB21L1	male abnormal 21-like protein 1
MAC / MC	microphthalmia (anophthalmia) coloboma
MAF	minor allele frequency
MANE	matched annotation from NCBI and EMBL-EBI
MI	mild (in context of mild aniridia spectrum phenotypic category)
MLPA	multiplex ligation-dependent probe amplification
MRC	Medical Research Council
MRI	magnetic resonance imaging
N3	N3 (also, N-3) <i>SOX2</i> enhancer
NA	not available (missing data)
NCBI	National Center for Biotechnology Information
NGS	next generation sequencing
NHS	national health service (UK)
NMD	nonsense-mediated decay
NTase	nucleotidyltransferase
OCT	optical coherence tomography

PAGE	polyacrylamide gel electrophoresis
PAX6	paired-box 6
PBS	phosphate-buffered saline
PCR	polymerase chain reaction (dd-, droplet digital; RT-, reverse transcription)
PD	paired domain (of PAX6 unless otherwise specified)
PDB	Protein DataBase
PFV	persistent foetal vasculature
PID	personal identification number
PITX2	paired-like homeodomain transcription factor 2 (pituitary homeobox 2)
poly(dI-dC)	poly(deoxyinosinic-deoxycytidylic) acid
PTC	premature termination codon
QC	quality control
RE	right eye
RPE	retinal pigment epithelium
RPID	research participant ID (equivalent to PID, above)
SDM	site-directed mutagenesis
SDS	sodium dodecyl-sulphate
SIMO	SIMO (E+153) <i>PAX6</i> autoenhancer
SNV	single nucleotide variant
SOX2	SRY-box 2
SV	structural variant (bSV, balanced)
TBE	Tris borate EDTA buffer
TGFβ	transforming growth factor beta
UTR	untranslated region
VEP	variant effect predictor
VUS	variant of unknown significance
WAGR	Wilms tumour, aniridia, genital anomalies and intellectual disability (historically ‘retardation’)
WES	whole exome sequencing
WGS	whole genome sequencing
WT	wild type
WT1	Wilms Tumour 1

Contents

Abstract.....	iii
Lay abstract	v
Acknowledgements	vii
Abbreviations	viii
Chapter 1 <i>Introduction</i>.....	1
1.1 Introductory note	1
1.2 Structure and development of the human iris.....	2
1.2.1 Iris anatomy.....	2
1.2.2 Iris development and its genetic control.....	2
1.2.3 Iris colour	4
1.2.4 Etymology of the iris.....	5
1.3 Eye malformations: a genetic and developmental overview	7
1.4 Clinical features of classical aniridia.....	9
1.4.1 Iris phenotype	9
1.4.2 Foveal hypoplasia, nystagmus and reduced vision	11
1.4.3 Cataract	12
1.4.4 Glaucoma	12
1.4.5 Keratopathy.....	12
1.4.6 Strabismus.....	13
1.4.7 Mild microphthalmia.....	13
1.4.8 Neuroendocrine abnormalities	13
1.5 Identification of <i>PAX6</i> as the principal aniridia gene	14
1.6 <i>PAX6</i> biology and structure	15
1.7 <i>PAX6</i> mutation spectrum.....	18
1.7.1 Monoallelic variants	18
1.7.1.1 Likely gene-disruptive variants	18
1.7.1.2 Missense variants	19
1.7.1.3 C-terminal extensions	19
1.7.1.4 Structural and copy number variants	20
1.7.1.5 Variants disrupting the <i>cis</i> -regulation of <i>PAX6</i>	20

1.7.2	Biallelic variants.....	21
1.8	Other forms of <i>PAX6</i>-related eye disease	22
1.9	Genetic differential diagnosis of classical aniridia	22
1.10	Syndromic forms of aniridia	24
1.10.1	WAGR syndrome (<i>WT1</i> and <i>PAX6</i> contiguous deletion).....	24
1.10.2	Gillespie syndrome (<i>ITPR1</i>).....	24
1.10.3	Multisystemic smooth muscle dysfunction (<i>ACTA2</i>)	24
1.11	Investigation and clinical management of aniridia	25
1.11.1	Genetic investigation strategy and clinical management	25
1.11.2	Targeted therapeutic avenues in aniridia.....	29
1.12	Conclusion and aims	30
Chapter 2	<i>The phenotypic spectrum of <i>PAX6</i> missense variants and their effects on DNA-protein binding.....</i>	33
2.1	Introduction.....	33
2.2	Methods	34
2.2.1	Phenotypic analysis of <i>PAX6</i> eye malformations	34
2.2.1.1	Ethical approval	34
2.2.1.2	Case selection	34
2.2.1.3	Structured phenotypic data recording	35
2.2.1.4	Phenotypic analyses and statistics	37
2.2.1.5	Accession numbers and protein domains.....	37
2.2.1.6	Computational structural modelling of the impact of missense variants.....	38
2.2.2	DNA-protein binding studies.....	39
2.2.2.1	Site-directed mutagenesis of plasmid DNA	39
2.2.2.2	Protein expression and purification.....	42
2.2.2.3	Oligonucleotide probe design.....	44
2.2.2.4	Electrophoretic mobility shift assay (EMSA).....	44
2.2.2.5	Imaging and quantification of EMSA	45
2.2.2.6	Analyses and statistics for DNA-protein binding study.....	47
2.3	Results.....	48
2.3.1	Defining the distinctive phenotypic spectrum of <i>PAX6</i> missense mutations.....	48
2.3.2	Distribution of variants amongst the domains of <i>PAX6</i>	51

2.3.3	Missense variants in key residues in the PAX6 paired domain cause structurally very severe eye phenotypes.....	54
2.3.4	Protein structural analysis per residue informed the selection of PAX6 variants for DNA-protein binding studies	59
2.3.5	Purification of PAX6 paired domain protein variants of interest	61
2.3.6	Comparison of DNA-protein binding of different PAX6 paired domain (PD) variants .	63
2.3.6.1	Choice of DNA target sequences	63
2.3.6.2	Optimisation of electrophoretic mobility shift assay (EMSA)	65
2.3.6.3	PAX6 DNA-protein binding curves and calculation of Kd values	67
2.3.6.4	Comparison of DNA-protein binding of different PAX6 paired domain variants by EMSA	69
2.4	Discussion	72
 Chapter 3 <i>Missense variants in MAB21L1 cause eye malformations</i>		
	<i>in human and mouse</i>	79
3.1	Introduction	79
3.2	Methods.....	81
3.2.1	Ethical approval	81
3.2.2	Human participant case selection and clinical phenotype data collection	81
3.2.3	Accession numbers.....	82
3.2.4	Mab21l1 mouse line with a targeted mutation	82
3.2.4.1	Generation of the CRISPR knock-in Mab21l1R51L mouse line	82
3.2.4.2	Breeding and husbandry.....	83
3.2.4.3	Genotyping of the mice	83
3.2.4.4	Phenotyping of the mouse line.....	83
3.2.4.4.1	Adult mice	84
	Ocular examination and imaging.....	84
	Histology.....	85
3.3	Results	86
3.3.1	Phenotypic analysis, human: monoallelic missense variants in <i>MAB21L1</i> cause aniridia and microphthalmia	86
3.3.1.1	Clinical case descriptions	86
3.3.1.1.1	Aniridia	86
3.3.1.1.2	Microphthalmia/MAC spectrum	86
3.3.1.1.3	Other ocular features.....	87

3.3.1.1.4	Non-ocular features	87
3.3.2	Protein structural context of MAB21L1 residue substitutions	92
3.3.3	Description of a CRISPR mouse <i>Mab21l1</i> line	96
3.3.4	Phenotypic analysis, mouse: <i>Mab21l1</i> ^{R51L/+} heterozygous mice have bilateral anomalous optic discs	98
3.3.5	Phenotypic analysis, mouse: <i>Mab21l1</i> ^{R51L/R51L} homozygous mice have a severe complex panocular eye malformation	102
3.4	Discussion	105
Chapter 4	<i>Whole genome sequencing study of “PAX6-negative” aniridia....</i>	111
4.1	Introduction.....	111
4.2	Methods	112
4.2.1	Clinical research participants and phenotypes	112
4.2.2	Ethical approval	112
4.2.3	Preparation of genomic DNA & quality control (QC)	112
4.2.4	Droplet digital PCR	112
4.2.5	Whole genome sequencing (WGS).....	113
4.2.6	WGS mapping, alignment, QC and SNV/INDEL variant calling	113
4.2.7	Structural variant (SV) calling	114
4.2.7.1	IGV visualization	114
4.2.7.2	Bioinformatic SV calling	115
4.2.8	<i>De novo</i> analysis	115
4.2.9	Variant filtering	116
4.2.10	Detection of intronic splice variants.....	116
4.2.11	<i>Cis</i> -regulatory variant analysis.....	116
4.2.12	Variant prioritisation and pathogenicity criteria	116
4.2.13	Experimental analysis of splice variants	117
4.2.13.1	RT-PCR and nested PCR of lymphoblastoid cell line-derived RNA.....	117
4.3	Results.....	118
4.3.1	Assembling the cohort and prescreening.....	118
4.3.1.1	Cohort description, phenotypes and prior screening.....	118
4.3.1.2	Assessment of quality and quantity of genomic DNA	122
4.3.1.3	Additional screening of <i>PAX6</i> and <i>MAB21L1</i>	122
4.3.1.4	Droplet digital PCR as a screen for <i>PAX6</i> locus copy number variants	123
4.3.1.5	Final cohort for whole genome sequencing	125

4.3.2	Structural variants identified in WGS aniridia cohort	127
4.3.2.1	Genomic copy number variants.....	130
4.3.2.1.1	FOXC1 whole gene deletions	130
4.3.2.1.2	PAX6 partial or whole gene deletions.....	132
4.3.2.1.3	Deletions of PAX6 cis-regulatory regions.....	137
4.3.2.2	Balanced structural rearrangements	142
4.3.2.3	Complex structural rearrangements.....	144
4.3.2.3.1	Likely benign reciprocal translocation t(1,9).....	144
4.3.2.3.2	Identification of a second and causative complex structural variant (inversion/deletion) affecting PAX6	146
4.3.2.3.3	A tandem duplication and deletion affecting PAX6	147
4.3.2.4	Comparison with bioinformatic structural variant callers	149
4.3.3	Sequence variants identified in the WGS aniridia cohort	152
4.3.3.1	Sequence variants affecting the <i>PAX6</i> coding exons	154
4.3.3.2	Splicing analysis of the <i>PAX6</i> locus	157
4.3.3.2.1	Essential splice site variants	157
4.3.3.2.2	Deep intronic splice variants.....	160
4.3.3.2.3	In vitro validation of selected splicing variants	163
4.3.3.2.4	RNAseq analysis from LCLs showed insufficient coverage of splice junctions to detect splice variants	167
4.3.3.3	No plausibly causative variants identified in <i>PAX6</i> cis-regulatory elements	168
4.3.3.4	<i>De novo</i> analysis of WGS trios	168
4.3.3.4.1	<i>De novo</i> missense variant of uncertain pathogenicity in <i>AHCYL2</i> , which interacts with a syndromic aniridia gene	169
4.4	Discussion	171
Chapter 5	<i>Final discussion and future challenges</i>	175
5.1	Advances in whole genome sequencing in relation to genetic eye disease	175
5.2	Are any disorders purely monogenic?.....	177
5.3	Towards a better understanding of cis-regulatory variants in <i>PAX6</i>	179
5.4	How much can you predict a <i>PAX6</i> phenotype from genotype?	180
5.5	Determining the impact of missense variants	181
5.6	Concluding remarks.....	182
	<i>References.....</i>	183

Appendices	213
Appendix A Supplemental material for PAX6 chapter.....	213
Appendix B Supplemental material for <i>MAB21L1</i> chapter.....	245
Appendix C Supplemental material for WGS chapter	274

Chapter 1 Introduction

1.1 Introductory note

Much of this chapter has been published as a review article (1). The parts of the publication that formed the basis of this introduction were solely written by me, and have been updated and/or expanded. The section here entitled ‘Targeted therapeutic avenues in aniridia’ includes material that I contributed towards a further review publication (2).

Classical aniridia is a dominantly-inherited, bilateral, panocular malformation characterised by a spectrum of iris and foveal hypoplasia with reduced vision from birth. It typically also features cataracts, secondary glaucoma and progressive keratopathy [MIM 106210] (1,3), and affects 1 in 40,000-72,000 live births (4–6). Haploinsufficiency for *PAX6* [MIM 607108] is the principal cause. It is a striking phenotype with a remarkably strong disease-gene link.

The transcription factor *PAX6* is highly conserved and critical to the control of eye and brain development (7). In addition to exonic variants and intragenic deletions, several disease-causing alleles have been identified which disrupt the *cis*-regulation of *PAX6*, either through structural variation (8–10) or, described in one individual, through a single nucleotide variant of the SIMO element (11). Whilst aniridia is typically caused by gene-disruptive variants (e.g. deletions, nonsense, essential splice variants), missense variants in *PAX6* are associated with both aniridia and more diverse phenotypes such as optic nerve anomalies (12,13), congenital corneal opacification (14) and microphthalmia (15,16). The microphthalmia, anophthalmia and coloboma (MAC) spectrum is more commonly due to other genes including *SOX2* [MIM 184429] and *OTX2* [MIM 600037] (16). The phenotypic spectrum associated with different classes of *PAX6* variants was historically a poorly studied area, especially given how much else is known about the locus; in parallel to this work other efforts have been made to understand it (17). It is unclear why some recurrent *PAX6* variants cause hypomorphic phenotypes such as foveal hypoplasia whilst others cause severe “worse than null” phenotypes, and whether this is related to the DNA-binding properties of the variant.

Finally, there remain a proportion of aniridia patients with no genetic diagnosis, indicating that there could be undiscovered ways of inactivating *PAX6* or additional disease loci. Thus, whilst much is known about the molecular genetic basis of aniridia and mechanism of *PAX6* disruption, these unexplained areas have guided the work described in this thesis.

1.2 Structure and development of the human iris

1.2.1 Iris anatomy

Sitting between the cornea and lens of the eye, the adult human iris is a 12 mm vivid, multilayered and photoprotective disc. Through control of pupil size in response to light or accommodation, it regulates light passage to the retina and is an optical necessity for sharp vision. It lies anterior to the ciliary body and itself plays a structurally important role in aqueous dynamics and the iridocorneal angle configuration, both of which affect intraocular pressure and glaucoma risk.

A macroscopic view of the iris shows a central pupillary region demarcated from the ciliary region by the iris collarette (**Figure 1 A,B**). The sphincter muscle, responsible for miosis (pupil constriction) under parasympathetic nervous control, lies in this inner pupillary region, whereas the dilator muscle, which produces sympathetically-mediated mydriasis (pupil dilatation), lies in the more peripheral ciliary region.

1.2.2 Iris development and its genetic control

Of the principal components of the iris, the bilayered epithelium and smooth muscle have a different embryological origin to the overlying stroma:

The iris pigmented epithelium is neuroectodermal, growing out from the anterior margins of the optic cup from the third month of gestation; relatively late compared to the rest of the eye (**Figure 1 C-E**). The retinal pigment epithelium derives from the same neuroectoderm. Specification of iris progenitor cells to a non-neuronal fate is accompanied by the expression of transcription factors including *PAX6* and *OTX1*, and of members of the Wnt signalling pathway (18,19). Notch genes are also implicated in this patterning (20).

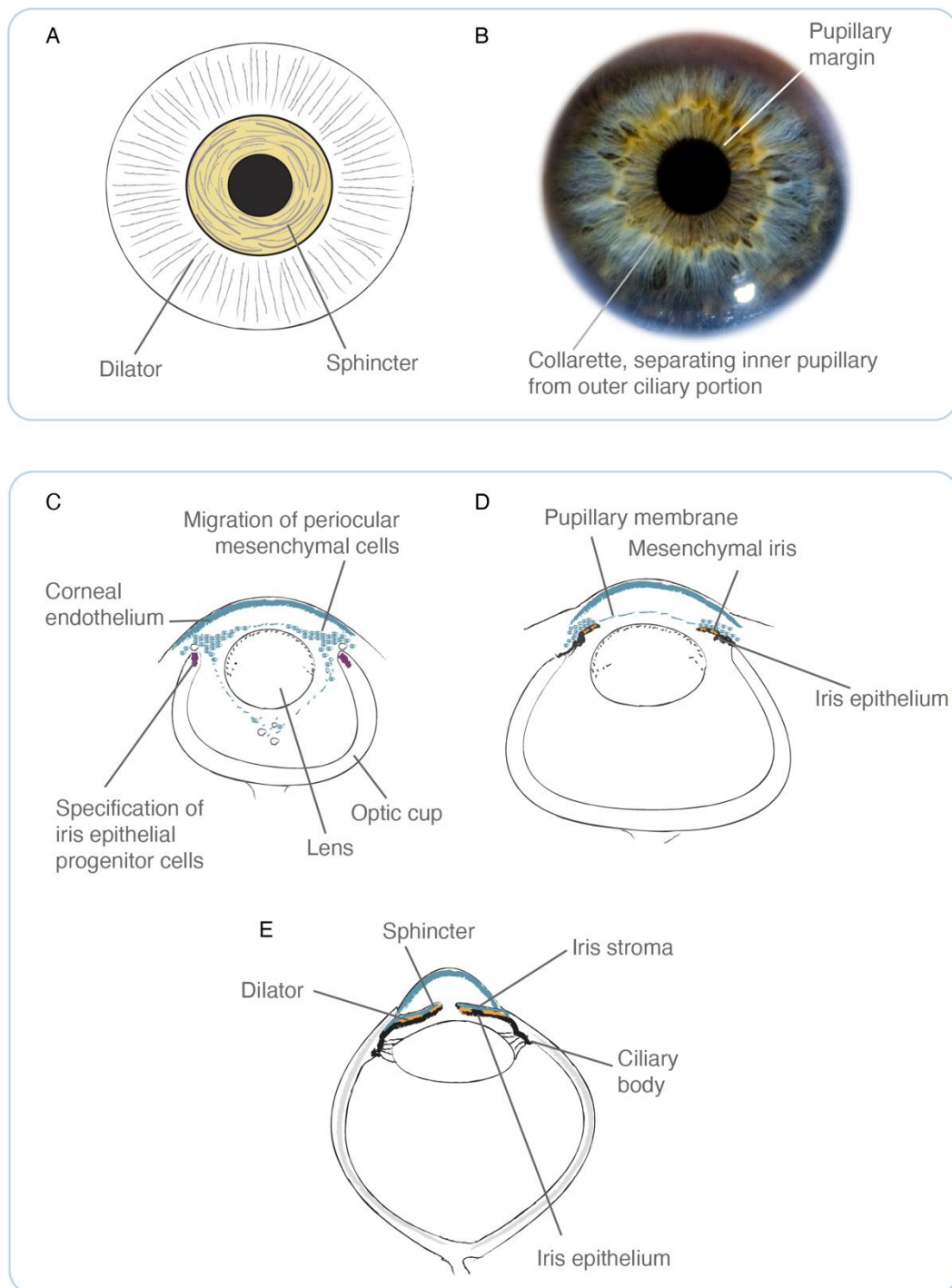


Figure 1 The structure and development of the human iris. (A) Cartoon of the iris musculature, showing the positions of the sphincter (shaded yellow) and dilator muscles. (B) Macroscopic photo of normal adult iris (image credit Chris Moody), showing the inner pupillary portion and outer ciliary portion. (C) Cartoon of the developing eye in a 7-8 week human embryo, showing the migration of periocular mesenchymal cells (shaded in teal) into the developing anterior chamber, which will go on to form the mesenchymal iris, pupillary membrane, and corneal endothelium and stroma. The tips of the optic cup, which is of neuroectodermal origin, are shaded in purple showing the specification of iris epithelial progenitor cells. (D) Cartoon of developing human eye at 4-5 months gestation, showing the iris pigmented epithelium (black) growing out from the optic cup margins. The iris musculature (orange) is just starting to develop. The mesenchymal iris (teal) forms a layer above these, and the pupillary membrane spans the pupil aperture. (E) Cartoon of adult eye, showing a well-formed anterior chamber and iridocorneal

angle; the iris stroma (teal), musculature (orange) and epithelium (black); lens attached by zonules to the ciliary muscle. The cartoons derived from multiple sources, primarily Ida Mann (21).

Two members of the TGF β superfamily (*Bmp4* and *Tgfb2*) and *Foxc1* have been identified as direct downstream targets of Pax6 in murine iris development (22). In addition to evidence that they are expressed in mouse iris development (23–25), *BMP4* and *BMP7* are highly expressed in developing chick iris smooth muscle (26).

The iris sphincter and dilator muscles are rare examples of ectodermally-derived muscle (26,27). They are not fully formed histologically in human fetuses until the sixth and eighth months, respectively (21).

By contrast, the pupillary membrane and iris stroma are mesenchymal in origin, arising from the third wave of migrating neural crest cells (28). During the second month of gestation, after separation of the lens but before the appearance of iris epithelium, a thin layer of mesodermally-derived tissue is visible anterior to the lens (21). The central part of this will become the pupillary membrane, which regresses late in gestation. Mesenchymal cells migrate along the anterior border of the iris epithelium, along with stromal melanocytes deriving from neural crest (29). *PAX6* is also expressed during morphogenesis of this part of the iris, though in a more transitory manner and at lower levels than in the ectodermal tissues (30). The *Small eye* rat (*rSey*) shows impaired migration of neural crest cells (31).

1.2.3 Iris colour

The striking colour of the iris is directly related to both the concentration of pigment within iris stromal melanocytes and the type of melanin (32), with brown eyes being more pigmented with a higher eumelanin:pheomelanin ratio. Final adult eye colour is not reached at birth, with iris stromal development continuing postnatally. In infants of white European ancestry, the iris appears bluer for the initial months of life due to lower stromal melanin content (33). Genetically, the *OCA2-HERC2* region of chromosome 15q explains the large majority of variation in human eye colour (34).

1.2.4 Etymology of the iris

The name of this anatomical structure is borrowed from ancient Greek Ἥρις meaning rainbow. Iris was the messenger goddess of the rainbow in ancient Greek mythology, and also gives her name to the flower genus, both pictured in **Figure 2** (35).



Figure 2 Iris etymology. Both the ocular structure and the flower pictured in (A) derive from the goddess in Greek Mythology, depicted in (B).

(A) Z-stack of 17 photographs of an iris 'Verschiedenfarbige Schwertlilie (Iris versicolor)' created by Reinhold Möller (36);



(B) Oil painting by Guy Head (c.1793), 'Iris Carrying the Water of the River Styx to Olympus for the Gods to Swear By' (37)

Table 1 Definition of terms used in the description of aniridia and related eye malformations

Term	Working definition
<i>Iris</i>	
Aniridia	1. (<i>Clinical sign</i>) Absence of the iris, which can be congenital or traumatic 2. (<i>Disease</i>) Panocular eye malformation described in section 1.3
Complete aniridia	Absence of the iris, with no visible iris tissue. (Distinction is not made depending on whether any remnant is detectable gonioscopically.)
Partial aniridia	Incomplete aniridia with visible iris rim (e.g. in classical aniridia) or, less commonly, absence of part of the iris (e.g. Gillespie syndrome, see below)
Iris hypoplasia	Incomplete iris formation. Often describes the milder aniridia spectrum iris phenotype, but is a neutral term not implying a particular condition/gene. Mildest manifestation of it is iris transillumination due to stromal hypoplasia
Congenital mydriasis	Iris malformation giving the appearance of fixed dilated pupils. In Gillespie syndrome this is due to aplasia of the iris sphincter muscle
Iris coloboma	Inferior iris defect resulting from failure of optic fissure closure
Iridolenticular strands	Strands of iris tissue between the iris and the lens, often seen in Gillespie syndrome or anterior segment dysgeneses
Corectopia	Displacement of the pupil, usually congenital, which may be misshapen
<i>Anterior segment</i>	
Posterior embryotoxon	Iridocorneal angle malformation resulting in an anteriorly displaced Schwalbe's line. Seen in Axenfeld-Rieger syndrome and a small percentage of normal eyes
Keratopathy	General term for corneal pathology. In context of aniridia, this is a progressive corneal opacification (aniridia associated keratopathy, AAK)
Peters anomaly	Congenital corneal opacification (leukoma) related to a posterior corneal defect, often with iris adhesions to the lens or cornea: kerato-irido-lenticular dysgenesis
Sclerocornea	Congenital corneal opacification, often associated with flattening (cornea plana). However, sclerocornea is arguably a redundant and unhelpful term (38)
Cataract	Congenital or acquired lens opacity, with prognosis and need for surgery dependent on age of onset, symptoms, morphological type and laterality
Congenital aphakia	Failure of lens formation, often with drainage angle abnormalities and glaucoma
<i>Posterior segment and external eye</i>	
Foveal hypoplasia	Failure of specialisation of the central macula for fine detailed central vision, visible on OCT as a shallow/absent foveal dip; associated with nystagmus. Core feature of classical aniridia (not Gillespie syndrome) and albinism
Optic nerve hypoplasia	Small optic disc, classically with peripapillary halo of pigmentation; level of vision correlating with papillomacular nerve fibre integrity rather than disc size
Optic nerve coloboma	Excavated optic disc anomaly due to defective optic fissure closure and thus deeper inferiorly, which may involve the whole disc and/or coexist with MAC
Chorioretinal coloboma	Defect in inferior choroid and retina resulting from failure of optic fissure closure (resulting in bare sclera being visible on fundoscopy), may be small or extend all the way to the optic disc; part of MAC spectrum
Microphthalmia (<i>microphthalmos</i>), MAC	Small eye(s) (axial length <2 SD from age mean), usually with some degree of malformation of the internal ocular structures (cf nanophthalmos, small but with normal anatomy). On spectrum (MAC) with anophthalmia, where the eyes but not adnexa are absent (either 'true' or 'clinical' anophthalmia) and coloboma
Ptosis	Drooping of the upper eyelid

1.3 Eye malformations: a genetic and developmental overview

Structural eye malformations are rare diseases associated with significant visual impairment (39). As well as the microphthalmia-anophthalmia-coloboma (MAC) and aniridia spectrum disorders, which are the focus of much of the work presented here, this overlapping group of primarily genetic conditions also includes other anterior segment dysgeneses, congenital cataracts and primary congenital glaucoma.

Two comprehensive Scandinavian studies have shown an aniridia prevalence of 1 in 40,000-72,000 live births (4,5), in keeping with older US epidemiological data (6). MAC spectrum disorders are overall more prevalent than this, with coloboma affecting 1 in 5000, microphthalmia 1 in 7000-11,000 and anophthalmia 1 in 30,000-66,000 live births (16,40,41).

Eye malformations result from perturbation of the tightly-regulated genetic control of ocular development. An overview of the key terms and clinical signs used in describing eye malformations is given in **Table 1**, above. The severity and number of structures affected depend on where the developmental process is disrupted. In brief:

MAC spectrum

1. Variants in the genes controlling the specification of the eye field in the forebrain (3 weeks' human gestation) and early eye organogenesis can result in anophthalmia or microphthalmia. Nearly 100 causative genes are now implicated (16,42). In 2003, the FitzPatrick laboratory identified *SOX2* as the major causative gene in bilateral anophthalmia or severe microphthalmia [MIM 206900], discovered initially via a translocation breakpoint (43), and subsequently also *OTX2* (44) [MIM 610125]. These two genes, which together coregulate *RAX* in specification of the eye field (45), between them account for 60% of bilateral severe cases, many of which have associated systemic or syndromic manifestations (16,46). Other eye field transcription factors (*RAX*, *VSX2*, *PAX6*), and members of the retinoic acid signalling pathway (*STRA6*, *ALDH1A3*, *RARB*) are also important causative genes (42).

2. Failure of optic fissure closure (5th-7th week of gestation) results in iris, chorioretinal and/or optic disc coloboma. Uveal coloboma is often associated with microphthalmia, and occurs when the various processes required for successful optic fissure closure are disrupted (apposition of the two epithelial margins, basement membrane breakdown, regulation of cell adhesion factors, fusion of the epithelial layer) (42,47). As with the rest of the MAC spectrum, there is considerable locus heterogeneity. Most coloboma genes are also microphthalmia genes (discussed above), but some cause predominantly isolated coloboma including *FZD5* involved in WNT-signalling (48,49) [MIM 620731] and *ACTG1* (50).

Aniridia

3. *PAX6* gene disruption results not just in aniridia, through perturbation of the iris development processes described above, but to a panocular eye malformation. This is due to the major role of *PAX6* in regulating organogenesis and differentiation of multiple ocular structures, from the end of the 3rd week of gestation and continuing postnatally, hence affecting the development of the retina (fovea) and optic nerve as well as the anterior segment (51). This is discussed in depth in the following sections.

Anterior segment dysgeneses

4. Failure of the lens to form or to separate causes various forms of kerato-iridolenticular dysgenesis (38), including congenital aphakia, Peter's anomaly and cataract. Key lens development genes include *PAX6*, *SIX3*, *SOX2*, *MAB21L1* and *FOXE3* (52). At the more severe end, failure of lens formation is a mechanism that can lead to microphthalmia (53,54) hence the overlap in the causative genes. Congenital aphakia is caused by biallelic *FOXE3* variants and accompanied phenotypically by a thin silvery cornea [MIM 610256]. *FOXE3*, *PAX6* and others (54,55) can also cause a sign called Peter's anomaly (type 2) where the lens fails to separate from the overlying cornea, resulting in congenital corneal opacification (CCO) (38).
5. Disruption to the anterior segment development can also occur predominantly as an iridotrabecular dysgenesis. Two important examples of this are primary

congenital glaucoma, presenting with buphthalmos due to raised intraocular pressure and CCO, of which the major genetic cause is *CYP11B1* variants [MIM 231300]; and Axenfeld-Rieger anomaly/syndrome, due to *PITX2* or *FOXC1* variants (56) [MIM 180500 and 602482, respectively], which often feature posterior embryotoxon and CCO. Axenfeld-Rieger features in detail later (**section 1.9**) as it is in the genetic differential diagnosis for aniridia (the ocular sign, rather than the condition).

6. Failure of regression of the foetal intraocular (hyaloid) vessels during the 8th month of gestation results in a spectrum of structural abnormalities now termed persistent foetal vasculature (PFV). In a large study of 246 children under 2 years old undergoing cataract surgery, 24% had PFV (57). Both PFV and especially congenital cataract have diverse aetiologies, both genetic and non-genetic, and are significant contributors to childhood visual impairment (58,59).

We will next look at classical aniridia in more detail.

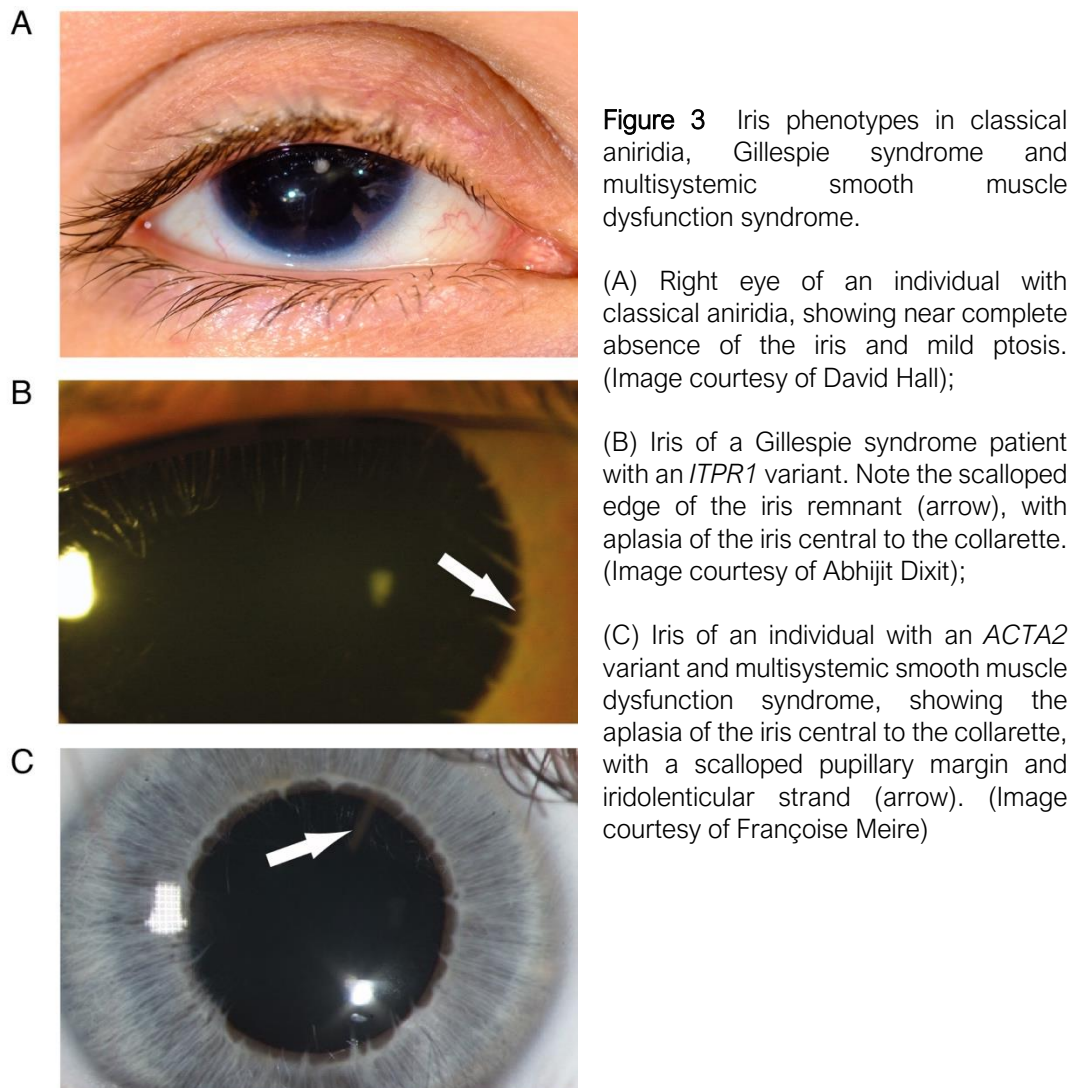
1.4 Clinical features of classical aniridia

First described in 1818 (60), classical aniridia is a bilateral, panocular malformation [MIM 106210]. The term aniridia alone can refer to both the clinical sign (iris aplasia, or partial/moderate to severe hypoplasia) or to the panocular condition (2). The latter is a spectrum comprising aniridia with foveal hypoplasia, with or without optic nerve hypoplasia, with the later development of cataracts, glaucoma and/or keratopathy. Mild ptosis is also sometimes seen as a feature, as in **Figure 3 A**. Arguably, ‘classic’ aniridia could be a more accurate term than ‘classical’, though I have not adopted it.

1.4.1 Iris phenotype

The typical iris phenotype is of complete (**Figure 3 A**) or partial aniridia, often with a thin crescent of iris visible. The distinction between partial and total aniridia is of limited use as, armed with a gonioscopy lens to view the iridocorneal angle, even those with apparently complete aniridia may have a detectable iris stump (61). Lesser degrees of iris hypoplasia are also seen, sometimes as subtle as a complete iris but with abnormal architecture or transillumination defects. A grading of the iris hypoplasia-aniridia spectrum has been proposed by Grønskov (62), **Figure 4**.

Cases of apparent partial aniridia that have aplasia of the sphincter and a scalloped margin (described as congenital mydriasis) are important to differentiate from other aniridia, due to the genetic differential diagnosis (*ACTA2* and *ITPR1*) and systemic implications for this distinct phenotype (**Figure 3 B,C**).



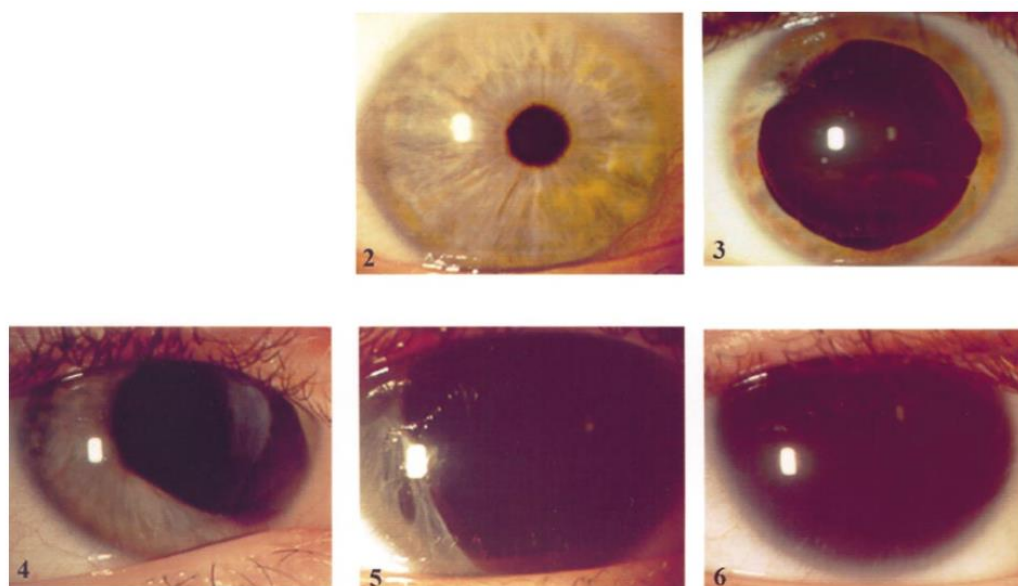


Figure 4 Grading of iris hypoplasia proposed by Grønskov and adapted from their paper (62). The degree of hypoplasia was categorised as follows:

1. Stromal hypoplasia with central pupil (no image)
2. Stromal hypoplasia with eccentric pupil
3. Circumpupillary aplasia
4. Atypical sector coloboma (< half circumference absent)
5. Subtotal aniridia (> half circumference absent)
6. Complete aniridia

1.4.2 Foveal hypoplasia, nystagmus and reduced vision

Interestingly, foveal hypoplasia in aniridia occurs without the chiasmal misrouting seen in albinism (41). Foveal hypoplasia is an almost universal finding in classical aniridia, though with varying grade (63), and is a more consistent finding than the iris phenotype (64). It typically results in reduced vision around the level of 6/30 (0.7 logMAR) or worse (with wide variation), such that the majority of patients are eligible for sight impairment registration. Normal vision is extremely rare in aniridia.

Foveal hypoplasia is typically associated with nystagmus (61). The nystagmus seen in aniridia is a form of infantile nystagmus syndrome (INS) with the onset within the first 6 months of life (typically around 2 months rather than at birth); INS can occur secondary to ocular conditions such as aniridia but can also occur in isolation (65,66). INS secondary to ocular disorders such as aniridia is sometimes alternatively referred to as sensory nystagmus.

In addition to foveal hypoplasia, the optical consequences of lack of iris tissue also contribute to reduced vision and glare, and in some patients optic nerve hypoplasia and amblyopia may also feature. These factors act from childhood and, as complications develop over time, glaucoma and especially keratopathy can further impair vision.

1.4.3 Cataract

Whilst the whole spectrum of cataract morphology is seen in aniridia, there are some types which are more common, including polar and subcapsular (both anterior and posterior). A progressive “cartwheel” form of posterior subcapsular cataract has been described by Bremond-Gignac as pathognomonic of *PAX6* variants (64).

Cataracts in aniridia are associated with capsular fragility that render surgery more difficult (67) and often become symptomatic in young adulthood (mean age of cataract surgery 30 (68)). Lens subluxation also occurs.

1.4.4 Glaucoma

Whilst congenital glaucoma presenting with buphthalmos is rare in aniridia, glaucoma occurs in approximately half of aniridia patients, in later childhood or in adulthood (69). Mechanisms of glaucoma described in aniridia include open angle glaucoma, rotation of the iris stump causing synechial chronic angle closure, and goniodysgenesis (70). A minority of patients undergo childhood cataract surgery which itself carries an independent risk of secondary glaucoma (71).

1.4.5 Keratopathy

Abnormal corneal epithelial homeostasis and limbal stem cell deficiency, reflecting impairment of the corneal maintenance role of *PAX6*, lead to progressive corneal opacification (72). Signs of aniridia-associated keratopathy (AAK, synonymous with ARK, “aniridia-related”) are detectable in childhood but it typically becomes symptomatic in adulthood, the cornea becoming slowly cloudier and more vascularised with a gradual conjunctivalisation over decades. Early signs of AAK include an increased central corneal thickness, reduced corneal sensitivity and degradation of the limbal palisade morphology. The published grading of AAK includes grade 1 (peripheral vascularisation) to grade 4 (opaque cornea) and grade 5

(end stage irregular cornea with no discernible layers) (73). AAK constitutes the principal cause of visual loss in later life (74).

1.4.6 Strabismus

In common with many ocular conditions which reduce vision and disrupt fusion, strabismus is very common in aniridia. A review of 379 aniridia records by the Bremond-Gignac group showed 150 (40%) had a strabismus diagnosis (75). The squint can turn in either direction: of the 73 able to maintain fixation and complete an orthoptic examination in this study, there was an equal split between eso- and exodeviations. Previous literature suggested a preponderance of esotropia (61).

1.4.7 Mild microphthalmia

Whilst other ocular phenotypes such as moderate to severe microphthalmia are seen in conjunction with *PAX6* missense variants (discussed below), this is not considered part of the classical aniridia phenotype. However, some degree of mild microphthalmia is well recognised in classical aniridia, in one study affecting 11% of individuals (76).

1.4.8 Neuroendocrine abnormalities

Non-ocular phenotypes, in particular neuroanatomical and endocrine abnormalities have been identified in studies of aniridia patients. These are often asymptomatic or subtle and as such may be underreported, especially as aniridia is traditionally considered a solely ocular condition. MRI brain studies have shown structural abnormalities in the pineal gland, anterior commissure, corpus callosum and frontoparietal cortical areas in aniridia (77–79). These are primarily radiological studies with no, or limited (77), evidence of potential functional consequences. However, subtle deficits in olfaction and hearing have been described, and specifically there is evidence of auditory interhemispheric transfer deficits in children with aniridia (80,81). Impaired glucose tolerance has been reported in aniridia, and the prevalence of diabetes is higher than the general population (68,82,83). This is supported by functional and animal work on the role of Pax6 in the pancreas (84).

Below I will discuss other clinical phenotypes of *PAX6*-related eye disease (**section 1.8**), a genetic differential diagnosis for aniridia (**1.9**) and the notable syndromic forms

of aniridia (1.10). Prior to that, we shall explore the story of the disease-gene link between aniridia and *PAX6*.

1.5 Identification of *PAX6* as the principal aniridia-associated gene

Classical aniridia is strongly associated with heterozygous loss of function of *PAX6* (Paired box 6, GRCh38 chr11:31784792-31811353) [MIM 607108], on chromosome 11p13 reverse strand. Animal models and human patients were key to the identification of this gene in the early 1990s (51). The 25th anniversary was commemorated in an excellent review (85).

Comparative mapping data had suggested that human aniridia was homologous to the murine *Sey* phenotype (86,87), a strain first reported in 1967 as having microphthalmia in heterozygotes (88) - and later shown to also feature iris hypoplasia, cataracts and corneal opacification (89). *Pax6* was isolated along with other *Pax* genes from mouse expression libraries (90). In parallel, the gene was independently identified from positional cloning and shown to be deleted in human aniridia (91). Concurrent analysis of *Sey* alleles confirmed the causal link between *Pax6* and semidominant *Sey* (92).

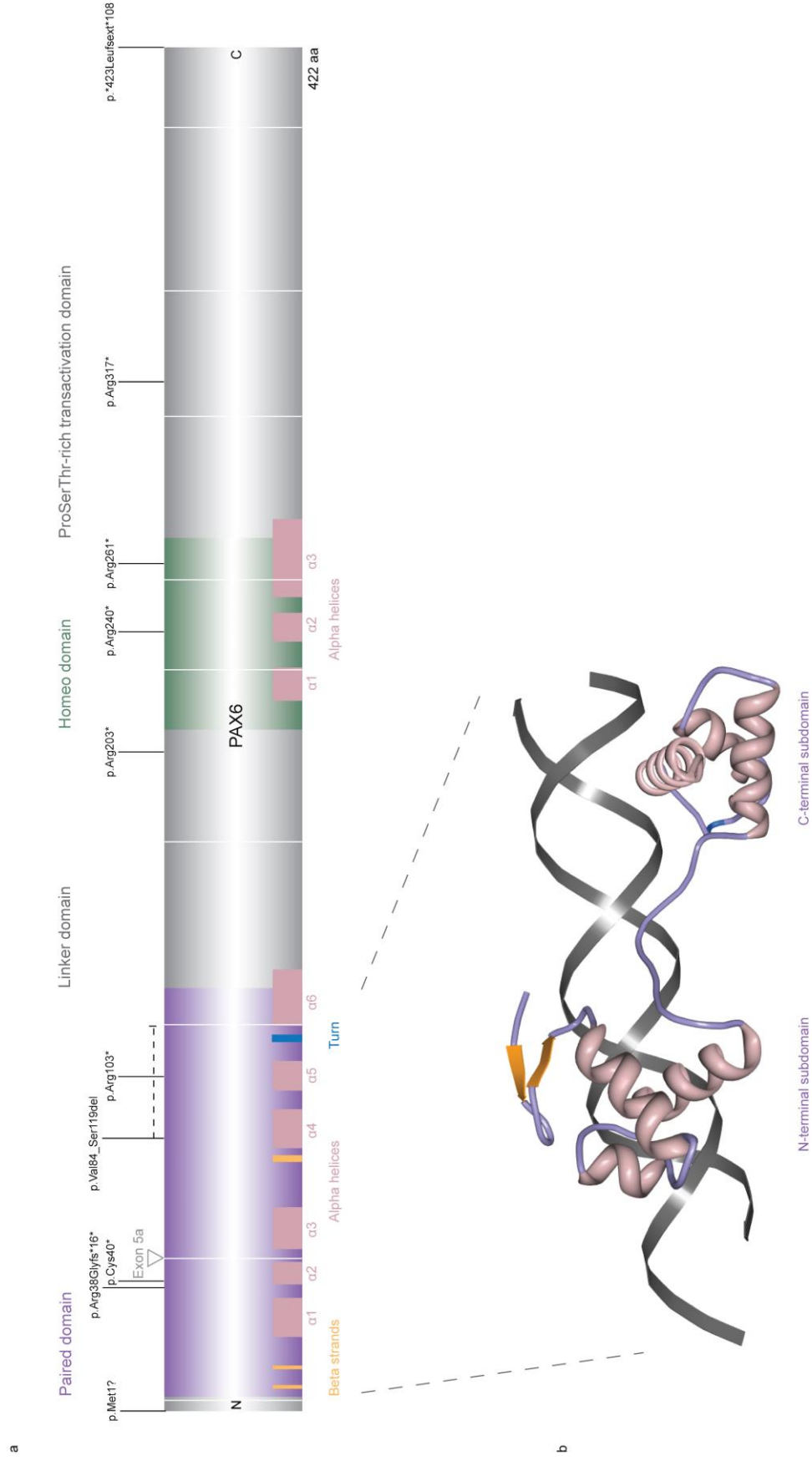
Establishing the link between human *PAX6* variants and the semidominant *Sey*, the authors (92) first observed the importance of dosage sensitivity. *PAX6* is entirely haploinsufficient as quantified by its impressive pLI score of 1.00, with any score ≥ 0.9 indicating extreme intolerance to truncating variation (93,94). Whilst *PAX6* haploinsufficiency leads to aniridia, whole-gene duplication causes non-aniridic eye malformations, and homozygous loss-of-function is perinatal lethal in humans and mice (88,95–99).

1.6 PAX6 biology and structure

PAX6 is a highly-conserved transcription factor critical to the control of both ocular and neural development (7). In addition to orchestrating the developing eye tissue, it is expressed in the central nervous system and pancreas, as well as in adult tissues such as the cornea, brain and pancreas where it is involved in homeostasis (77,100–104).

The most biologically relevant human *PAX6* isoform, isoform a, encodes a 422 amino acid, 46.7 kDa protein (UniProt P26367-1 (105), NM_000280, closest related Ensembl transcript ENST00000643871, NP_000271). It is labelled as the canonical isoform on UniProt but is not consistently recognised as such elsewhere. Most of the *PAX6* variant literature uses this isoform, including this thesis. The less abundant, alternatively spliced PAX6(5a), isoform b, (436 amino acid, 48.2 kDa, UniProt P26367-2, NM_001368894, ENST00000640368, NP_001355823) contains an extra 14 residues between Gln47 and Val48, encoded by exon 5a (**Figure 5a**); at the time of submission (2024) this has been chosen as the MANE Select transcript and, as it is longer, often described as canonical. PAX6 contains two DNA binding domains, the bipartite paired domain (**Figure 5b**) and the homeodomain, as well as a proline-serine-threonine-rich C-terminal transactivation domain. The paired- and homeodomains can bind DNA both cooperatively and independently, widening the repertoire of targets (106). A third isoform has been identified arising from an additional internal promoter (107,108), though the role of such isoforms in humans, potential transcripts of which have been identified *in silico*, remains an interesting question (109).

The subcellular localisation of PAX6 is predominantly nuclear. Nuclear localisation signals have been identified in the paired domain and near the homeodomain (110,111). However, non-nuclear localisation has also been observed. In chick retina, PAX6 was nuclear in the ganglion cell and inner nuclear layers, and cytoplasmic in the photoreceptor and outer nuclear layer (112), which may reflect its roles in diverse cellular processes (113). In quail and nematode, paired-less isoforms are found in the cytoplasm (108,114). In the mouse brain, SPARC and Pax6 have been shown to co-localise and physically interact, and the authors suggest a model in which SPARC facilitates nuclear-cytoplasmic shuttling of Pax6 (115), a role that has also been shown for Karyopherin 13 (111).



In PAX6(5a), the additional 14 amino acids between alpha helices 2 and 3 of the paired domain modulate its binding and transactivation potential, preferentially binding DNA with the C-terminal subdomain, acting as a molecular toggle (118,119). The ratio of PAX6(5a) to canonical PAX6 is higher in adult than embryonic tissue (120), with the PAX6(5a) isoform having a potential role in postnatal eye development (121). Other *PAX6* transcripts whose function is currently uncertain may also have a regulatory role.

PAX6 is thus able to increase its target diversity by alternative splicing as well as through the use of different DNA binding domains (113). Further fine-tuning is achieved through cooperative binding with other transcription factors, such as SOX2 (122,123): together the two transcription factors bind the *DC5* enhancer to drive δ -crystallin expression in the initiation of lens development (122,124). Targets of Pax6 include transcription factors (e.g. Six3 (125) and Foxc1 (22)), signalling molecules (e.g. Wnt) and other crystallins (e.g. alpha A crystallin (126)); for targets in iris development see above (**section 1.2.2**). These PAX6 target genes give rise to diverse ocular phenotypes when mutated in humans: missense variants in *SIX3* cause microphthalmia, coloboma and holoprosencephaly [MIM 157170]; *FOXC1* variants cause anterior segment dysgenesis [MIM 601631], and alpha-A-crystallin (*CRYAA*) variants cause cataract [MIM 604219].

Some of the *PAX6* gene regulatory networks have been elucidated. An elegant and phenotypically-relevant example is that reciprocal inhibition between Pax6 and Pax2 that establishes the retinal-optic nerve boundary in mammals (127); *PAX2* variants cause papillorenal syndrome in humans [MIM 120330]. Little is published regarding potential regulatory networks involving *PAX6* at the fovea. Returning to the lens, Pax6 was shown by ChIP to directly regulate *Prox1* and *Mab21l1* as part of a lens differentiation network (128).

Transcriptional activation of *PAX6* in the different tissues of the eye requires the coordinated activity of *cis*-regulatory elements, which are mostly conserved across mammals and zebrafish (129,130). These were initially identified via human translocation cases, detailed below.

1.7 PAX6 mutation spectrum

Reflecting the many possible ways of disrupting PAX6 function, 472 unique sequence variants are reported (LOVD PAX6 database, version 170616, previously hosted by the MRC HGU at lsdb.hgu.mrc.ac.uk/home.php?select_db=PAX6; current version (131)). All classes of sequence variation are present and no domain of the protein is spared. The 10 commonest aniridia variants are marked on **Figure 5**.

Whilst the majority of *PAX6* variants are heterozygous loss of function associated with classical aniridia, the wider range of phenotypes spans subtle iris hypoplasia to perinatal lethality. *PAX6* eye malformations are dominantly inherited and fully penetrant, though can show variable expressivity. In keeping with historical observations that one third of aniridia is sporadic and two thirds familial (61,132), in a large cohort from Wessex Regional Genetics Laboratory, approximately 40% diagnosed *PAX6* variants causing aniridia were *de novo* and 60% inherited from a parent (133).

1.7.1 Monoallelic variants

1.7.1.1 Likely gene-disruptive variants

The commonest intragenic variants are encoded by single nucleotide substitutions resulting in a premature termination codon (PTC), i.e. nonsense variants, with p.Arg240*, p.Arg317* and p.Arg203* being the three most frequently reported (129 reports in the LOVD database); these mutational hotspots correspond to CpG dinucleotides (134). Frameshifting insertions, small deletions and duplications, as well as many splice site variants, also fall into this category of likely gene-disruptive variants (presumed loss of function). PTCs >50-55 nucleotides upstream of the last coding exon-exon junction are predicted to be subject to nonsense-mediated decay (NMD). Though there could be exceptions to the NMD rules (135,136), the known truncating variants in humans phenocopy null alleles and are notably absent from the 3' end of exon 12 and the coding portion of exon 13 (134), consistent with NMD. In addition to the classical aniridia phenotype described above, some likely gene-disruptive variants appear to be hypomorphic, though this is rare and milder phenotypes appear to be more common with missense variants (68,137). The

overwhelming majority of LGD variants have a classical aniridia phenotype, with higher grades of foveal and iris hypoplasia than missense variants (68).

1.7.1.2 Missense variants

Missense variants in *PAX6*, which may lead to production of a full length protein with altered structure and function (138), have been published associated with classical aniridia, mimicking haploinsufficiency, as well as with diverse non-classical phenotypes, many but not all of which are milder phenotypes (68,134). These include optic nerve anomalies (12), hypomorphic phenotypes such as foveal hypoplasia or *ectopia pupillae* (139) and, less commonly, microphthalmia (15,16).

In addition to canonical *PAX6*, two missense variants in the *PAX6*(5a) isoform (affecting the alternatively spliced exon 5a itself) have been described in five Japanese families, at least three unrelated. They demonstrate congenital corneal opacification with or without foveal hypoplasia and cataract (140), with functional work suggesting subtle changes in binding by *PAX6*(5a) and transactivation (118).

The structural and functional protein consequences of many missense variants (in addition to many splice site and in-frame deletions) have not been determined experimentally. The functional consequences of missense variants, unlike most PTCs, are dependent on their location within the protein. For example, missense variants at the C terminal amino acid of *PAX6* impair transactivation mediated through the homeodomain (141), and clearly those in the DNA-binding domains have the potential to disrupt this critical function of the transcription factor. This will be explored further in Chapter 2.

1.7.1.3 C-terminal extensions

Frameshift duplications and substitutions predicted to result in C-terminal extensions are notably common (34 LOVD database reports; **Figure 5**), some with variant phenotypes including exudative retinopathy (137), and many with nystagmus and foveal hypoplasia (133).

1.7.1.4 Structural and copy number variants

Structural variants (SV) such as chromosomal rearrangements and whole/partial gene deletions (copy number variants, CNV) disrupting *PAX6* account for approximately one third of sporadic aniridia cases, and are also found in some familial cases (142), such that overall these account for approximately 10% all aniridia cases (143,144). Balanced SV are harder to detect with modern array- and NGS-based screening, which we will explore further in Chapter 4. The expected phenotype is classical aniridia, except for contiguous gene deletions causing WAGR syndrome which are discussed below. Lastly, whole gene duplication of *PAX6*, in the form of partial 11p duplication, is rare but reported (97), with variable ocular and systemic phenotypes. The authors summarise a total of 11 cases and describe eyes with various mild anterior or posterior segment abnormalities, nearly all with nystagmus.

1.7.1.5 Variants disrupting the *cis*-regulation of *PAX6*

Via analysis of translocation and inversion breakpoints in aniridia patients, several disease-causing alleles were identified which disrupt the long-range regulation of *PAX6* (**Figure 6**). These were located 11 kb, 22 kb (8), 125 kb and 150 kb (SIMO element) (145,146) downstream (telomeric) to *PAX6*. Analysis of the SIMO element showed that this is an autoregulatory enhancer binding site whose disruption impairs *PAX6* maintenance (11). The phenotypes strongly resemble, and were initially thought to be identical to, heterozygous deletion or PTC-causing variants of *PAX6*. However, whilst they do have classical aniridia, recent evidence from a large cohort suggests individuals with 3' regulatory region deletions have better vision and less severe grades of foveal hypoplasia and keratopathy (147).

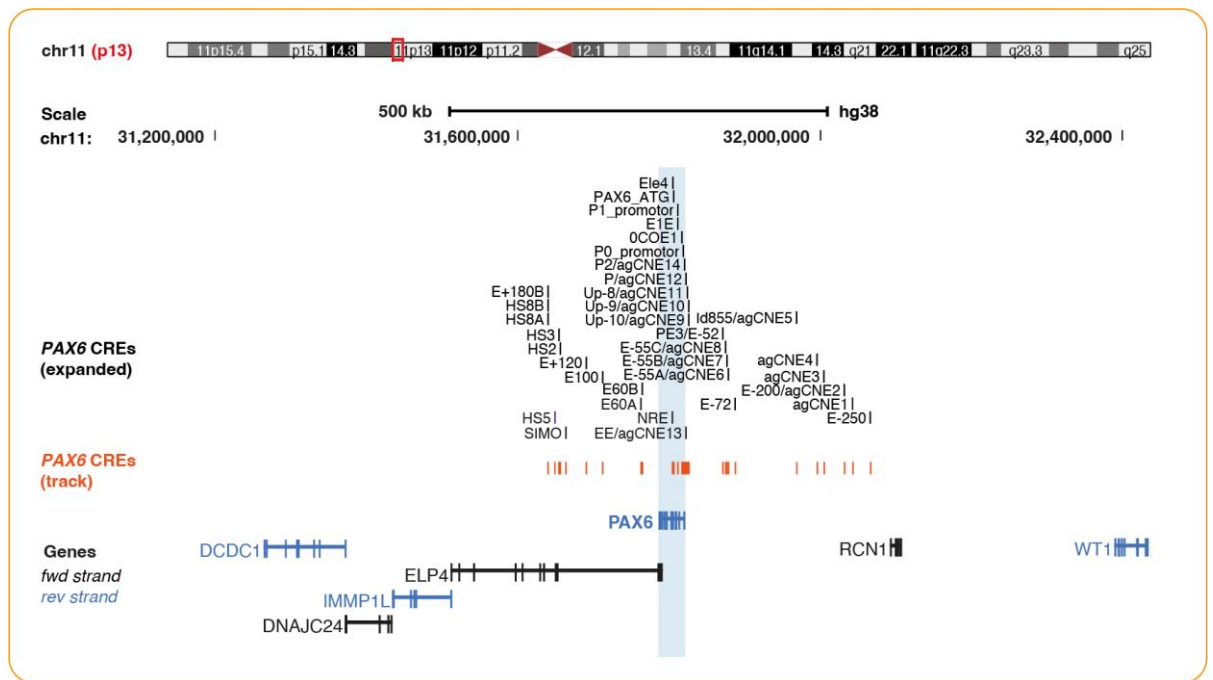


Figure 6 Location of *PAX6* regulatory elements in relation to *PAX6* and neighbouring genes. The chromosomal location is indicated by a red box on the ideogram of chromosome 11 at the top. Below this are genomic coordinates (hg38) for scale. At the bottom are the genes: *PAX6* is indicated in blue, as are the other genes on the reverse strand, whereas the forward strand genes are in black. The area encompassing the *PAX6* gene coordinates is shaded in blue. *PAX6* cis-regulatory elements are shown as an orange track above *PAX6*, with the individual elements annotated as an expanded track above (this includes most but all CREs, but includes all known CREs active in the eye).

Variants in *PAX6* regulatory regions are now more easily detectable by array-based comparative genomic hybridisation (CGH). Ansari *et al.* identified 5 aniridia patients with deletions involving plausible 3' regulatory regions and suggest a 'critical region' for transcriptional activation spanning 244 kb (148). Plaisancié *et al.* identified a further 9 such patients (10) in their large study of non-coding regions of *PAX6* in aniridia, in addition to several interesting variants of unknown significance (VUS) in non-coding regions.

1.7.2 Biallelic variants

Biallelic variants in *PAX6* with two likely gene-disruptive alleles cause severe CNS defects with anophthalmia and perinatal lethality (95,99). A similarly severe ocular phenotype was reported in a viable compound heterozygous child with trisomy 21 (149). However, two viable compound heterozygote families have been published with

aniridia or coloboma, and no severe systemic effects, through having a combination of a likely gene-disruptive and presumed mildly hypomorphic missense allele (62,150).

1.8 Other forms of *PAX6*-related eye disease

It is well recognised that *PAX6*-related eye malformations form a spectrum that extends beyond classical aniridia. Some of these phenotypes above can be considered variable expressivity of classical aniridia or hypomorphic phenotypes, for example milder iris abnormalities including *ectopia pupillae* (139), or the non-iris features of classical aniridia, either in isolation or combination, namely cataract, foveal hypoplasia, nystagmus, glaucoma and keratopathy (151–153). However, there are also more striking and somewhat distinct phenotypes described. Severe anterior segment dysgenesis including congenital corneal opacification (CCO) is described. Often this is labelled as Peters anomaly (14,38) but this can be an ambiguous term historically (along with sclerocornea in particular) and a more modern approach to naming CCO taking into account the genetics and developmental mechanisms is helpful (38). Isolated optic nerve anomalies are also well described in the *PAX6* eye malformation literature, including hypoplasia, aplasia, and excavated optic disc anomalies (12,13).

Lastly, *PAX6* coloboma and microphthalmia cases are described (15,16,154) [MIM 120200]. The mechanism of the more severe MAC cases, which were previously not described, will be explored in Chapter 2.

1.9 Genetic differential diagnosis of classical aniridia

The classical aniridia phenotype is almost exclusively associated with heterozygous loss-of-function of *PAX6* to the extent that it is almost considered a monogenic disorder. The other aniridia-associated genes that follow in this section are, broadly speaking, phenotypically distinct but included here for completeness. In addition, two syndromic forms of aniridia discussed in the next section are an important genetic differential diagnosis (*ITPRI*, *ACTA*), again with their own distinct phenotypes.

The most well-established aniridia-like phenotypes associated with other genes occur with dominantly-inherited *FOXC1* and *PITX2* variants (155,156), associated with Axenfeld-Rieger syndrome (ARS). A large study of 128 ARS patients found aniridia

in 8% *PITX2* and 13% *FOXC1* cases. Both genes are associated with systemic features: hearing and congenital cardiac defects (*FOXC1*), and dental, craniofacial and periumbilical features (*PITX2*). The ocular features of ARS (or Axenfeld-Rieger anomaly, ARA, as syndromic features are not always present), include corectopia, polycoria, posterior embryotoxon, glaucoma, and iris hypoplasia or aniridia. Whilst congenital corneal opacification is common, a progressive later-onset keratopathy similar to aniridia does not feature. Foveal hypoplasia is also not part of the condition, though there is a reported co-occurrence in one family with a *PITX2* variant (157). Reported aniridia cases with *PITX2* variants include both intragenic nonsense and missense variants, deletions (158–160) and one with a telomeric regulatory variant (148). *FOXC1* cases featuring aniridia include missense and whole gene deletions (148,155,156,160). The vast majority of these have features of ARA in addition to aniridia. These studies also highlight that congenital glaucoma presenting with buphthalmos is a common feature of *FOXC1*-associated aniridia.

Similarly, aniridia has been reported as a feature of other anterior segment dysgeneses: *CYP11B1* (primary congenital glaucoma), *PXDN* (variable phenotypes, but 2 cases including aniridia and microspherophakia) and *FOXE3* (primary congenital aphakia) (161,162). These are all recessively inherited.

An aniridia family with a segregating missense variant in *TRIM44*, ~4 Mb away from *PAX6* on 11p13, was described (163), but we advise caution in ascribing pathogenicity as this is a single pedigree (albeit supported by *in vitro* functional data suggesting that the two *PAX6* 3'UTR SNVs cosegregating were unlikely to be pathogenic).

There remain a small but significant proportion of aniridia patients with no genetic diagnosis. Following an analysis of 42 *PAX6*-negative aniridia cases with array CGH (Comparative Genomic Hybridisation), the authors were left with 27/42 unexplained cases (148). Since this publication, the Gillespie syndrome cases in this cohort have been shown to have *ITPR1* variants (164,165) as described below. A similar study of *PAX6*-negative cases left 17/47 with no genetic diagnosis (10): the cohort overall this equated 10% aniridia cases remaining unexplained.

Based on the above data of others and our own, we estimated the proportion of unexplained classical aniridia cases to be 5-10% at the start of this work (**Figure 7c**, below).

Lastly, we can now add *MAB21L1* to the list of aniridia (and microphthalmia) differential diagnosis genes (166,167), but this is the subject of Chapter 3.

1.10 Syndromic forms of aniridia

1.10.1 WAGR syndrome (*WT1* and *PAX6* contiguous deletion)

Contiguous deletion of *WT1* and *PAX6*, which are less than a megabase apart on chromosome 11p, results in a syndrome known as WAGR [MIM 194072]: tumour, aniridia, genitourinary anomalies and intellectual disability (historically, R from ‘mental retardation’). Obesity is also sufficiently recognised as part of the syndrome for WAGRO to be in common usage [MIM 612469], with *BDNF* causally implicated in this and the intellectual disability (168,169). The ocular phenotype is classical aniridia and these individuals are at risk of and require screening for nephroblastomatosis and end-stage renal failure (170,171).

1.10.2 Gillespie syndrome (*ITPR1*)

In 2016, the causative gene in Gillespie syndrome [MIM 206700] was identified via whole exome sequencing projects as *ITPR1*, with *de novo* monoallelic variants acting in a dominant-negative mechanism related to multimeric interactions (164) and biallelic cases also described (165). Gillespie syndrome is a triad of aniridia, non-progressive cerebellar ataxia and intellectual disability. The iris phenotype is pictured in **Figure 3**.

1.10.3 Multisystemic smooth muscle dysfunction (*ACTA2*)

As shown in **Figure 3**, the *ACTA2* iris phenotype closely resembles that of Gillespie syndrome: so-called congenital mydriasis (but really a form of partial aniridia) resulting from aplasia of the sphincter muscle and overlying tissues. Interestingly the products of both *ACTA2* and *ITPR1* are both components of cGMP kinase signalling complexes and studies in tracheal smooth muscle suggest an interaction (172,173).

This is an important genetic diagnosis as the syndrome is associated with life-threatening thoracic and cerebral aneurysms which can be screened for radiologically. The syndrome also includes congenital mydriasis, pulmonary hypertension, bladder hypotonia and hypoperistalsis of the gut (174) [MIM 613834].

1.11 Investigation and clinical management of aniridia

1.11.1 Genetic investigation strategy and clinical management

The investigation and management of aniridia cases is multidisciplinary, usually involving genetics, paediatric and ophthalmology services, and even more so in syndromic cases. I published a suggested approach to working up new aniridia and congenital mydriasis cases, shown in **Figure 7**, along with a general outline of clinical management recommendations (1).

For the majority of aniridia cases, genetic work up comprises two tests: an NGS gene panel (that ideally includes *PAX6*, *FOXC1*, *PITX2* and *ITPR1*) and an assessment of copy number such as array CGH. The array is particularly important to detect deletions encompassing *WT1*, which determine the need for Wilms tumour screening (4) and monitoring of renal function. In familial aniridia cases, the chance of a *WT1* deletion is very low, with only two WAGR syndrome familial cases reported (144,175). However, the array retains excellent diagnostic value, as the high-resolution assessment of copy number can detect whole gene or regulatory region deletions in *PAX6* (or rarely, *FOXC1/PITX2*) (148,176,177). We therefore have not distinguished between the investigation of sporadic and familial aniridia cases, though clearly in familial cases where the genetic cause is already known, this should be tested directly.

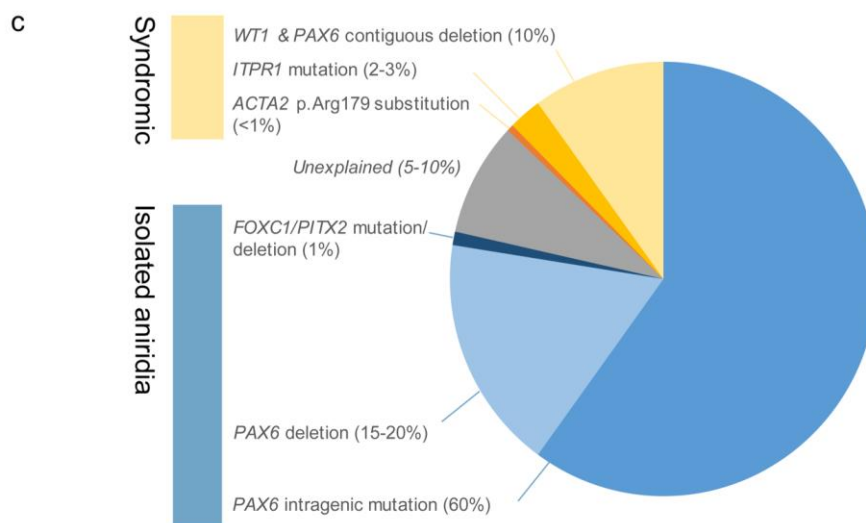
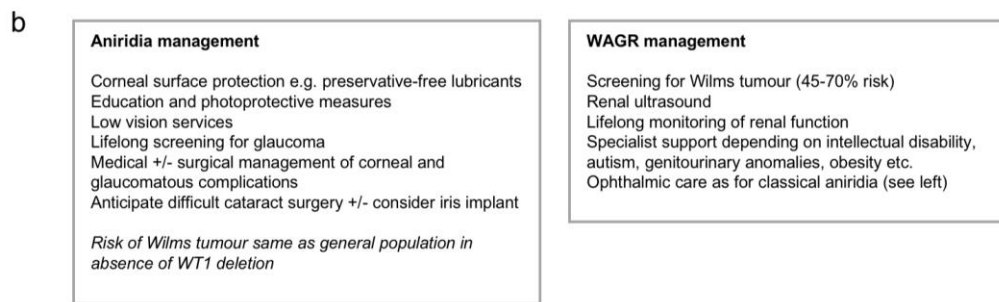
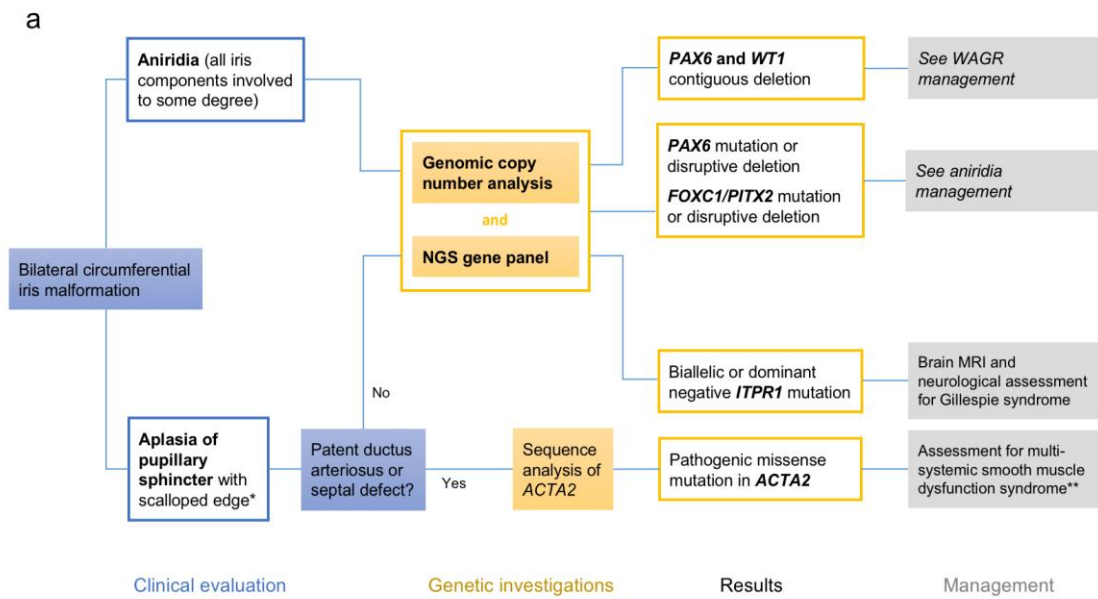


Figure 7 Genetic investigation, management and causes of isolated and syndromic aniridia, adapted from my publication (1). (a) Flowchart showing an example approach to the genetic investigation of new congenital aniridia/congenital mydriasis cases) (1). This will usually consist of an aniridia NGS gene panel – in the UK this includes *PAX6*, *FOXC1*, *PITX2* and *ITPR1* – and copy number analysis (e.g. chromosomal array or array-based comparative genomic hybridisation). The latter is important to look for contiguous deletion of *PAX6* and *WT1*. We have suggested the combination of iris and cardiac features which should prompt consideration of *ACTA2* sequencing – this could be tested with other genes such as *ITPR1* depending on clinical judgement. * Gillespie syndrome-like iris (see **Figure 3 B,C**); ** A suggested surveillance regimen for the serious complications of *ACTA2* multisystemic smooth muscle dysfunction has recently been outlined (178). (b) Brief outline of the management considerations in aniridia and WAGR syndrome, from an Ophthalmology perspective. (c) Pie chart showing the genetic causes of isolated and syndromic aniridia highlighting the proportion of unexplained cases. Note that these frequencies are calculated as a percentage of all aniridia cases (using the term broadly and including Gillespie syndrome). For example, whilst *ITPR1* accounts for 2-3% aniridia cases, it is the only known cause of Gillespie syndrome. They are estimated and approximate frequencies derived from published sources (4,142–144,148) or, where no published data is available, estimated from the HGU eye cohort of >400 aniridia patients

There are two causes of aniridia associated with life-limiting complications. In WAGR syndrome, Wilms tumour survival rates approach 95% at 4 years, but WAGR carries a late mortality risk from end-stage renal failure (179,180). The International WAGR Syndrome Association produce a comprehensive management summary for physicians including screening protocols (wagr.org). Individuals with *ACTA2* multisystemic smooth muscle dysfunction are at risk of acute, life-threatening vascular events including aortic dissections and ischaemic strokes in childhood (181); it is important to identify these rare cases early to instigate appropriate vascular surveillance (178).

For patients with apparent congenital mydriasis with pupillary sphincter aplasia (**Figure 3 B,C**), the two known genetic causes are *ITPR1* and *ACTA2* p.Arg179 substitutions. In most cases, the iris defect is detected before the other features of these conditions. However, there is one clinical feature which can aid the decision of whether to test *ACTA2* in addition to *ITPR1* (or opt to include a panel containing the former). In the largest series of *ACTA2* p.Arg179 cases (33 individuals), the only feature reliably present at birth in all cases, other than the iris defect, was a cardiac defect: either a large patent ductus arteriosus (PDA, 91%) or aortopulmonary septal defect (9%). Based on this and given the serious prognostic implications, it seems reasonable based on current evidence to only test *ACTA2* if both the iris and cardiac features are present. This would be the decision of the clinical team, in this situation

typically a Clinical Geneticist and Ophthalmologist with paediatric experience working together.

One of the principal benefits of NGS gene panels for patients with unexplained ocular disease is the number of eye genes that can be tested simultaneously and cost effectively. For example, known *PAX6* target genes, such as *BMP4*, *FOXC1* and *SIX3* discussed above – which are not aniridia-associated genes but cause diverse eye malformations – would be included in such a panel, which is now standard practice in the UK. However, active consideration of the individual genes being tested is still clearly required: for example, *ACTA2* does not feature on the Genomics England Structural Eye Disease Panel (v4.0, 7 Aug 2024) (182), but does feature on several other panels involving thoracic aneurysms and cerebral vascular disease, so would need to be actively tested in this scenario if there was a clinical suspicion.

If all of the suggested investigations are negative, options for pursuing further investigation include looking for balanced chromosome translocations and whole exome or whole genome sequencing (WGS), depending on availability. The latter may be more often pursued in a research rather than diagnostic laboratory setting but is increasingly used in routine public healthcare. WGS analysis in aniridia is the subject of the research presented in Chapter 4.

In addition to the genetic tests outlined in **Figure 7**, optical coherence tomography (OCT) is valuable for evaluating foveal hypoplasia, though this is postponed until the child can cooperate unless handheld OCT is available. MRI brain may be helpful in classical aniridia if there are relevant neurological symptoms (e.g. olfactory, auditory), as well as in Gillespie syndrome to assess cerebellar structures.

Lastly, it is also helpful to alert families to the existence and benefits of patient organisations, such as Aniridia Network UK (aniridia.org.uk), Aniridia Foundation International (www.aniridia.net) and IWSA (wagr.org).

1.11.2 Targeted therapeutic avenues in aniridia

As discussed earlier, the development of the eye, orchestrated by *PAX6*, is not complete at birth and continues postnatally (including iris and fovea), and *PAX6* also plays a role in tissue maintenance of the adult cornea. Therefore, genetic therapies for aniridia may potentially be of benefit in the early postnatal period, and throughout life for preserving corneal clarity.

Knowledge of the specific variant in aniridia will become important for determining eligibility for some future genetic therapies. Ataluren targets in-frame nonsense variants and showed promise in a *Sey* mouse model with a nonsense variant in the linker region (183,184). A phase II clinical trials as an oral preparation vs placebo for aniridia patients with a nonsense variant (STAR trial, ClinicalTrials.gov ID NCT02647359) was negative, with a primary endpoint of a change in maximum reading speed using MNREAD acuity charts (185).

Ataluren previously failed to meet the primary endpoint in phase 3 trials for cystic fibrosis (186) and initially Duchenne muscular dystrophy (DMD) (187), albeit showing benefit in a subgroup of the latter. A further phase III trial was conducted for DMD, and subsequent meta-analysis of the two trials showed evidence that it delayed disease progression (188); it has now been conditionally approved from use in DMD by the European Medicines Agency. This raises the question of whether the most useful primary endpoint was chosen in the aniridia trial, in a condition where visual function is already very poor.

More recently, a further agent with nonsense suppression properties, amlexanox (used in allergic respiratory tract conditions), was able to rescue the disease phenotype in an iPSC-derived optic cups and limbal epithelial stem cell models from aniridia patients, in work by the Moosajee laboratory (189).

Employing an alternative strategy that is not conditional on a nonsense variant, it was recently shown that the antidepressant duloxetine enhances the expression of *PAX6* protein via the ERK/MEK pathway. In *PAX6* gene-edited human limbal stem cells, duloxetine enhanced *PAX6* expression and normalised the expression of the target genes normally repressed by *PAX6* in mutant cells (190). In contrast to nonsense

suppression, which risks producing some truncated or abnormal protein, in this study only wild type PAX6 was enhanced. Duloxetine has now been given topically to wild type mice and upregulates Pax6 *in vivo* without corneal toxicity including in LPS-induced corneal injury (191). Using a similar strategy and target pathway *in vivo* in a mouse aniridia model, another MEK inhibitor (PD0325901) given topically or systemically in the early postnatal period enhanced PAX6 expression in the cornea, improving corneal clarity, visual function and intraocular pressure (192,193). These suggest exciting avenues for clinical translation that could apply to other diseases involving genes that, like PAX6, are highly dosage sensitive.

Lastly, phenotypic rescue of the *Pax6* small eye (*Sey*) mouse, which carries an exon 8 truncating variant, has been demonstrated using *in vivo* CRISPR germline correction of the *Sey* allele (194). The authors were able to insert a non-disruptive endogenous 3xFLAG tag immediately upstream of the *Pax6* start codon, in both the *Sey* mouse (calling this new strain *Fey*) and a wild type control strain (named *Fax*), in order to be able to monitor expression of the alleles. They then corrected the pathogenic variant in the *Fey* mouse using CRISPR/Cas9 microinjection of zygotes, achieving a 25% allelic correction rate, with restoration of Pax6 expression levels and a normal eye examination. In the future, CRISPR editing of limbal stem cells of young *Fey* mice would potentially be a model for treating children with aniridia.

1.12 Conclusion and aims

The striking disease-gene link between aniridia and *PAX6* has been a paradigm of haploinsufficiency and gene regulation studied over the past three decades. Building on this, the research presented here approaches specific gaps in our understanding from both the gene side and the disease side. The aims are as follows:

1. To explore the phenotypic spectrum of *PAX6* missense variants, and to address the hypothesis that the severe “*worse than null*” phenotypes are due to changes in the DNA-binding properties of the variant.

2. To phenotype and explore a new aniridia and microphthalmia gene, *MAB21L1*, missense variants of which phenotypically resemble *PAX6* missense variants.
3. To identify the genetic cause of classical aniridia in an unexplained (*PAX6*-negative) cohort through whole genome sequencing, with the potential to find additional disease loci or previously undiscovered ways of inactivating *PAX6*.

Chapter 2 The phenotypic spectrum of PAX6 missense variants and their effects on DNA-protein binding

2.1 Introduction

Please note: the majority of the work presented in this chapter was published in 2019 (195) and this paper, on which I am joint first author, is available open access and included in the thesis as Appendix A.1. The parts of the published paper which overlap with sections below were written by me, and have been expanded upon to provide further detail for the thesis.

The remarkably strong relationship between *PAX6* haploinsufficiency and classical aniridia has contributed to *PAX6* becoming one of the best studied transcription factors in human disease. However, there remain gaps in our understanding of missense variation in *PAX6* and in particular how it leads to certain “worse than null” phenotypes.

Missense variants are frequently associated with hypomorphic phenotypes. *PAX6* missense variants can additionally partially or fully phenocopy haploinsufficiency to cause classical aniridia (196). Then there are rarer reports at the most severe end of the spectrum, such as severe microphthalmia and coloboma, which are poorly understood in terms of variant type and genetic mechanism.

In order to define the spectrum of phenotypes associated with *PAX6* variants, I performed systematic, detailed ocular phenotyping followed by cluster analysis of *PAX6* eye malformations. The distinct clusters emerging from this unbiased phenotypic characterisation then informed future analyses and collaborative work with Dr Joe Marsh performing protein modelling to predict the structural effects of causative amino acid substitutions.

Focussing on the severe end of the phenotypic spectrum, I explored the phenotype of previously-unpublished (now published (195)) severe microphthalmia cases caused by

recurrent variation of residues in the paired domain. Lastly, I investigated the effect of these variants on protein-DNA binding *in vitro*.

2.2 Methods

2.2.1 Phenotypic analysis of PAX6 eye malformations

2.2.1.1 Ethical approval

The studies in this thesis complied with the principles of the Declaration of Helsinki (197). The MRC Human Genetics Unit eye malformations (HGUeye) cohort was assembled and maintained with ethical approval from the Scotland A UK Multicentre Research Ethics Committee, using the following protocols: “Linkage and mutation analysis of developmental eye genes in cases of severe structural eye malformations” (06/MRE00/76), and “An investigation of the genetic causes of human eye malformations” (16/SS/0201).

A small number of cases are from the Deciphering Developmental Disorders (DDD) Study, which was collected under a protocol approved by Cambridge UK Multicentre Research Ethics Committee: “The Deciphering Developmental Disorders Study” (10/H0305/83).

2.2.1.2 Case selection

All available eye malformation cases with disease-associated missense variants in *PAX6* were included in the phenotypic analysis. These were sourced as follows:

1. Published cases, mainly collated via the LOVD Human *PAX6* Mutation Database (134,198). The database also includes unpublished cases submitted to the database, and was for many years the most comprehensive repository of pathogenic *PAX6* variants. It included all published cases up until at least the start of this study in 2017. Of note for future analyses: the database is no longer curated by the MRC Human Genetics Unit but the current version remains hosted by LOVD <https://databases.lovd.nl/shared/genes/PAX6>.
2. Previously-unpublished cases from the MRC Human Genetics Unit eye malformations (HGUeye) cohort, which was maintained by the FitzPatrick

laboratory; principally severe microphthalmia anophthalmia coloboma (MAC) spectrum cases

3. Previously-unpublished cases from our collaborator Dr Jay Self (University of Southampton; sequenced by Salisbury NHS Foundation Trust genetic testing program); these cases were foveal hypoplasia cases at the milder end of the spectrum
4. Previously-unpublished cases from the DDD (Deciphering Developmental Disorders) study (original DDD publication (199))

Cases were excluded if there was doubt as to pathogenicity, such as an allele inherited from a normal parent, or an allele appearing in the Exome Aggregation Consortium (ExAC) database (94) or, for more recent cases, the Genome Aggregation Database, gnomAD (200). Well-phenotyped, mutation positive, affected relatives were included to capture the phenotypic variability at each allele. In the case of very large pedigrees where not all family members could be included, care was taken to include a representative of each phenotype within a family.

In order to obtain a “true haploinsufficiency” comparison group, a size-matched sample of cases with likely gene-disrupting (LGD) variants was chosen. These were proband-only cases. Other than three previously-unpublished HGUeye/collaborator cases which constituted <2% of the group, this “control” cohort was selected at random from the much greater number of LGD cases available on LOVD. Randomisation was performed using the =RAND() function in Excel. LGD cases were distinguished from others in LOVD on the basis of the protein consequence (the ‘ProteinInfo’ column), typically worded ‘Protein synthesis unlikely’ or ‘NMD predicted’. The LGD group included predominantly stop-gain (nonsense) and frameshift variants, all predicted to result in a premature stop.

2.2.1.3 Structured phenotypic data recording

Cases were phenotyped systematically, consulting clinical information from the referral to the HGUeye cohort and original publications if relevant, with a focus on structured, detailed ocular phenotyping. Access to original ophthalmology case notes was not possible.

The data were collected reflecting the features present at a single point in time (the point of referral or publication); this was not a longitudinal study. The presence or absence of the following features were noted, or recorded as data not available (NA). Aniridia (unspecified), complete aniridia, partial aniridia, iris hypoplasia, foveal hypoplasia, nystagmus, cataract, glaucoma, keratopathy, optic nerve hypoplasia, optic nerve aplasia, excavated optic disc anomaly, coloboma (iris/choroid), microphthalmia, microcornea, congenital corneal opacity, posterior embryotoxon, corectopia/pseudopolycoria, and iridocorneal adhesions. In addition, important ocular and extraocular comorbidities, and age at assessment, were recorded (but not included in the final clustering analysis).

Insufficient examination detail was available for meaningful grading of foveal hypoplasia (201), keratopathy (202) or iris hypoplasia/aniridia (62). Data were recorded per individual rather than per eye, as laterality of signs was only available for a small proportion of patients.

For several key phenotypic features, a logical spectrum was present (e.g. iris hypoplasia < partial aniridia < complete aniridia; or optic nerve hypoplasia < aplasia) and so these were grouped to give an overall score, to avoid redundancy. The side effect of this was that the severe end of the spectrum of these features (aniridia, optic nerve aplasia, anophthalmia, severe congenital corneal opacification/anterior segment dysgenesis, ASD) were scored as a 2 in what was otherwise a 0 or 1 scoring for the absence or presence of a phenotypic feature, giving more weight to the severe end of these. However, as they were also the biologically most severe, this was considered an acceptable compromise.

Optic nerve malformations were not reliably distinguishable from the (often limited) clinical description, for example optic nerve coloboma and Morning Glory disc anomaly, and so the presence of either of these or a similar malformation was recorded as an excavated optic disc anomaly. Similarly, subclassification was not attempted for congenital corneal opacification (CCO), where the clinical label often seemed unreliable or variable (e.g. Peters anomaly vs sclerocornea, as discussed by Nischal (38)) with no anterior segment ultrasound data available for standardisation of these terms. Therefore, these were recorded simply as CCO.

Whilst detailed phenotype information regarding the presence of the above features was fed into the clustering methodology, a simpler ocular phenotype classification was required for the protein structural analysis performed in collaboration by Joe Marsh, using this data. For this, all cases were sorted into one of four broad categories of phenotype. The four categories were: AS, anterior segment; CA, classical aniridia; MC, microphthalmia/coloboma (MAC spectrum); MI, mild (mild aniridia spectrum or mild mixed phenotype). Cases were classified by the phenotype that was predominant.

2.2.1.4 Phenotypic analyses and statistics

All analyses for the phenotypic data were performed using R (203). Alpha was set at 0.05.

The phenotypic clustering and heat map were produced using the R packages cluster and gplots, and the function heatmap.2 (R script in **Appendix A.2.1**). Within heatmap.2, the functions dist and daisy were combined to compute the distance (dissimilarity) matrix; combining both these functions allowed missing data, NA, to be included in clustering. Euclidean distance (root sum-of-squares of differences) was chosen as the metric argument within daisy; clustering with Manhattan (sum of absolute differences) and Gower was also performed in order to see the effect on clustering results. Dr Liusaidh Owen (FitzPatrick laboratory) collected part of the phenotypic data and helped to write the R scripts.

2.2.1.5 Accession numbers and protein domains

PAX6 protein domains and regions were numbered using isoform NP_000271.1; P26367-1 (-5a isoform, 422 aa) with the following amino acid residue ranges for each domain: paired domain (PD), residues 4 - 130; linker region, residues 131 - 209; homeodomain (HD), residues 210 - 269; PST-rich transregulatory region, residues 279 - 422 (204); for cartoon depiction of domains please see **Figure 5** in the Introduction chapter.

PAX6 variants were described using the biologically relevant NM_000280.4 transcript (6969 nt, -5a isoform, 13 exons, 422 aa).

An update, NM_000280.5 (2500 nt, -5a isoform, 13 exons, 422 aa), was released in May 2020 based on CAGE and polyA data, as part of the MANE project (205). I noticed that this was drastically shorter than NM_000280.4 (**Figure 8**, below) and wrote to NCBI. This query triggered an update to NM_000280.6 in Dec 2021, due to the following reasons outlined by NCBI:

“I have reviewed the UTRs of NM_000280.5. The 5' end was (incorrectly) trimmed to the CAGE peak at chr 11:31,811,121 in version 5. This resulted in a 7-nt 5' terminal exon. ...I have updated to NM_000280.6 to extend the 5' terminal exon to the CAGE peak at chr 11:31,811,322. The length of the updated transcript is 2701 nt.”

(NCBI Staff Scientist, personal communication).

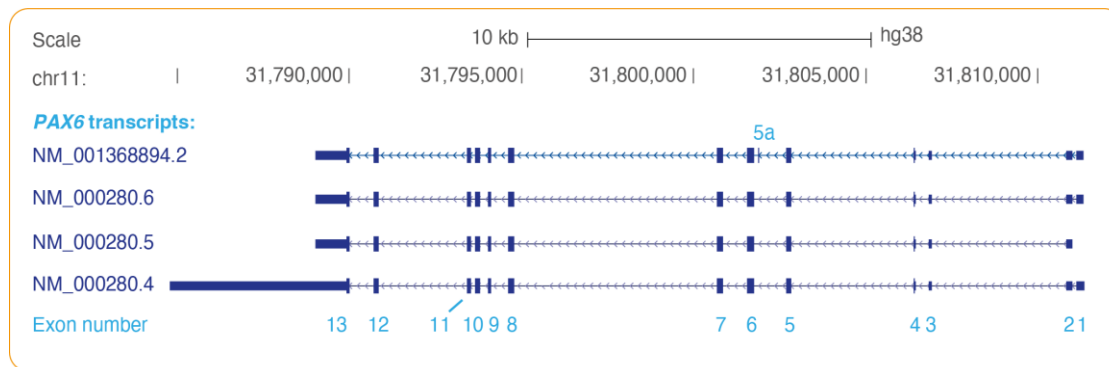


Figure 8 Selected human *PAX6* transcripts. This cartoon shows the *PAX6*(+5a) isoform b, NM_001368894.2, which is the MANE Select transcript, and the latest three versions of the more abundant and biologically relevant NM_000280 transcript, isoform a, which does not have the alternatively spliced exon 5a. The hg38 coordinates on chromosome 11 are indicated at the top, and exon numbers at the bottom (numbered from left to right as *PAX6* is transcribed on the reverse strand). The ‘erroneous’ update (see text), NM_000280.5, can be seen to almost abolish exon 1 in the 5'UTR.

For further commentary on *PAX6* transcripts, including the +5a isoform, please see Introduction **section 0**.

2.2.1.6 Computational structural modelling of the impact of missense variants

This work was performed by our collaborator Dr Joe Marsh. Some of the results are included in brief here as, firstly, my work on phenotypic characterisation of missense

variants fed directly into it, and secondly, the modelling results informed the choice of variants for DNA-protein binding studies. The structural impact of missense variants was modelled using FoldX using both the PAX6 protein in monomeric form and the DNA-bound complex. For the paired domain, the models used the X-ray crystal structure of the DNA-bound PAX6 PD (Protein DataBase, PDB ID: 6PAX), which is based on an “optimal” consensus binding sequence (116). For details of the homeodomain modelling, and further methodological details including the FoldX thresholds for “destabilising folding” and “perturb protein:DNA interactions”, please see the Methods of our publication (**Appendix A.1** (195)).

2.2.2 DNA-protein binding studies

2.2.2.1 Site-directed mutagenesis of plasmid DNA

In order to purify PAX6 paired domain protein variants of interest, site-directed mutagenesis was performed on plasmid DNA to be used as bacterial expression vector, as follows:

The pGEX-6P-1-N-FLAG plasmid vector was used for protein purification. The basic pGEX-6P1 plasmid had previously been modified in-house by Martin Reijns (Senior Research Fellow, Andrew Jackson lab) and by Hemant Bengani (postdoctoral scientist, FitzPatrick lab), to add the FLAG and *PAX6* paired domain sequences, respectively. The *PAX6* gene fragment encoded the first 136 amino acids of PAX6, of which the paired domain is amino acids 4-131 (105).

The complete sequence of the pGEX-6P-1 plasmid vector is available from Addgene (<https://www.addgene.org/vector-database/2887/>) and the construct and relevant sites of note are shown here in **Figure 9**.

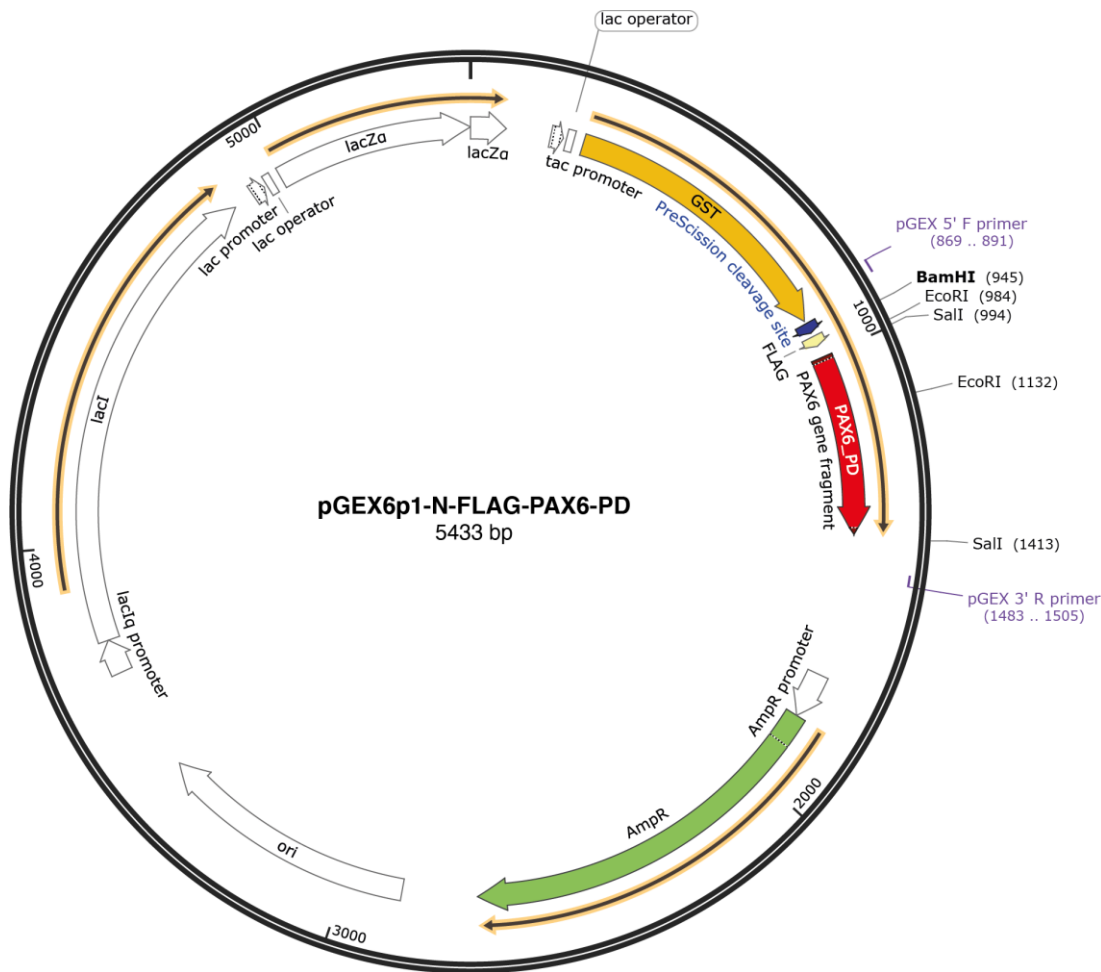


Figure 9 Plasmid map created using SnapGene software (GSL Biotech), showing the pGEX6P1-N-FLAG plasmid after cloning, resulting in a 5433 bp plasmid, pGEX6P1-N-FLAG-PAX6-PD. The *PAX6* gene fragment encoding the first 136 amino acids of the protein (of which the paired domain is amino acids 4-131, shown in bright red) was inserted such that it is under the control of the *tac* promoter, induced with the lactose analogue IPTG. This vector enabled bacterial expression of the *PAX6* paired domain as a GST fusion protein with an N-terminal FLAG tag. The GST tag (yellow) was cleaved with PreScission protease at the site marked, and Glutathione Sepharose beads were used for protein purification. An advantage of this cleavage enzyme is its relatively high activity at cold temperatures, allowing the reaction to be performed at 4°C, protecting the protein product. Other features of note in the plasmid are the pBR322 origin (*ori*), the ampicillin resistance gene (encoding β -lactamase; *AmpR*, green) and the *lac* repressor (*lacI*). The open reading frames are shown as thin yellow and grey arrows.

I further modified the plasmid using site-directed mutagenesis (SDM) to generate clones encoding *PAX6* paired domain protein variants of interest. I used the QuikChange II site-directed mutagenesis kit (Agilent), including their primer design tool. The target variant and SDM primers used are listed in **Table 2**.

Variant	Target base change (NM_000280.4)	Codon change	Oligo name	Oligo sequence 5' to 3'	Tm
p.Cys52Arg	c.154T>C	TGT → CGT	C52R_F	CAGAATTTTACTCACACGTCC GTTGGACACCTGCA	80.3
			C52R_R	TGCAGGTGTCCAACGGACGTG TGAGTAAAATTCTG	80.3
p.Ser54Arg	c.162T>G	AGT → AGG	S54R_F	TACCTGCCCAGAATTTTCCTCA CACATCCGTTGGA	81.4
			S54R_R	TCCAACGGATGTGTGAGGAAA ATTCTGGGCAGGTA	81.4
p.Arg92Gln	g275a	CGG → CAG	R92Q_F	GGACGGGCACTCCTGCTTATA CTGGGCT	77.1
			R92Q_R	AGCCCAGTATAAGCAGGAGTG CCCGTCC	77.1
p.Ser121Leu	c362t	TCA → TTA	S121L_F	CGAAGAAGTCTGTTTATTGATA ACACGCTTGGTATGTTATCGT	76.0
			S121L_R	ACGATAACATACCAAGCGTGTT ATCAATAAACAGAGTTCTTCG	76.0
p.As124Lys	c372g	AAC → AAG	N124K_F	TGCGAAGAAGTCTCTTTATTGA TGACACGCTTGGTATGT	77.9
			N124K_R	ACATACCAAGCGTGTCATCAAT AAAGAGAGTTCTTCGCA	77.9

Table 2 Design of oligonucleotide primers for site-directed mutagenesis

SDM was performed as 40 µl PCR reactions each using Phusion High-Fidelity PCR Master Mix with HF Buffer (New England Biolabs), 1 µl 10 µM SDM primers, 1 µl of 20 ng plasmid DNA template or water control, and 18 µl water. The following PCR conditions were used, due to the length of the amplification required: 98°C 3:00, (98°C 0:30, 65°C 0:30, 72°C 3:00) x26 cycles, 72°C 5:00 min. The PCR product was run and visualised on a 1% agarose gel. 0.5 µl (10 units) of DpnI restriction endonuclease (New England Biolabs) was added to each 35 µl sample of PCR product, mixed by pipette and incubated at 37°C for 30 min.

2 µl (40 ng) of each PCR product was transformed into 35 µl DH5α competent cells (Invitrogen, thawed on ice). Following 30 min incubation on ice, the cells were heat shocked at 42°C for 45 sec in a water bath, terminated on ice (5 min). The cells were incubated at 37°C, 500 rpm for 75 min with 0.5 ml each of SOC medium (Invitrogen), then centrifuged for 30 sec at 10,000 rpm. 100 µl of the supernatant was retained and

mixed with the pellet, and this was spread over L-AMP agar plates at room temperature. The plates were incubated overnight at 37°C.

Clones from each plate were selected and inoculated into 3.8 ml LB media containing 50 µg/ml ampicillin and incubated overnight at 37°C with shaking.

QIAprep Miniprep Kit (Qiagen) was used to isolate plasmid DNA as follows. Cells were pelleted by centrifugation at 4000 rpm for 5 min, resuspended in Buffer P1 (containing LyseBlue reagent and RNase A) 250 µl and transferred to 1.5 ml microcentrifuge tubes. 350 µl each of first lysis buffer, Buffer P2, and then Buffer N3, were added and mixed by inverting each time. The mixture was centrifuged at 13,000 rpm (same speed for the remainder of this paragraph) for 10 min; the supernatant was spun through the spin column for 1 min, discarding the flow through; washed with 750 µl Buffer PE (containing ethanol as directed), re-spun for 1 min, discarding the flow through; re-spun for 1 min. The column was transferred to a fresh 1.5 ml microcentrifuge tube. The DNA was eluted in 30 µl EB Buffer (10 mM Tris-HCl, pH 8.5), standing for 1 min prior to centrifugation for 1 min.

The concentration of plasmid DNA was ascertained by Nanodrop 1000 (Thermo Fisher Scientific) spectrophotometry. Sanger sequencing of the DNA from each clone was performed by the HGU technical services team.

2.2.2.2 Protein expression and purification

Using the plasmid DNA as a bacterial expression vector, the protein variants were expressed as GST fusion proteins and then purified using glutathione sepharose beads as follows:

2 µl (40 ng) of each plasmid DNA was transformed into 30 µl *E. Coli* BL21-Gold(DE3) Competent Cells (Agilent) using a 30 min incubation on ice and 45-sec 42°C heat shock, terminated on ice. After growing out for 1 hour in SOC medium at 37°C, the cells were plated on L-AMP agar plates (ampicillin 100 µg/ml) and incubated overnight at 37°C.

1-3 colonies of each plasmid clone were inoculated in 10 ml LB media (containing 50 µg/ml ampicillin) and incubated overnight at 37°C and 250 rpm. 10 ml of this culture

were added to 400 ml fresh media and further incubated under the same conditions. When the optical density (OD₅₉₅) reached 0.5-0.8, IPTG was added (1 mM) to induce protein expression via the tac promoter. 3 h later, the incubation was terminated and all of the subsequent steps were carried out at 4°C or on ice. Following centrifugation at 6000 rpm for 4 min (5400 g with Beckman JLA-16.250 rotor), the pellet was resuspended in 10 ml NETN buffer with protease inhibitor (NETN: 20 mM Tris-HCl pH 8, 100 mM NaCl, 1 mM EDTA, 0.5% NP40; cOmplete protease inhibitor cocktail, Roche: 1 tablet dissolved per 50 ml buffer). Following the addition of 1 mM DTT, the cells were lysed (3x 20 sec bursts of sonication) and centrifuged at 20,000 rpm for 30 min (48,400 g with Beckman JA-20 rotor).

The protein was purified using Glutathione Sepharose 4B beads (Cytiva, prepared as 50% slurry with cold PBS), 30 µl beads/ml supernatant, incubated overnight at 4°C. Following centrifugation (2000 rpm, 3 min) with three NETN buffer washes then two PreScission cleavage (PSC) buffer washes, the protein was cleaved from the beads using PreScission cleavage enzyme (10 µl/sample, rolling overnight incubation). Following centrifugation (3000 rpm for 2 min details), the first protein eluate was removed, followed by a 5 min rolling incubation with PSC buffer and re-spin to obtain the 2nd eluate.

Protein samples were aliquoted and stored at -80°.

Protein samples were prepared for SDS-PAGE as follows: the sample was mixed with SDS sample loading buffer and reducing agent, and boiled at 85°C for 10 min. The samples were loaded onto a NuPAGE 4-12% Bis-Tris gel (Thermo Fisher) and run in MES-SDS buffer with NuPAGE antioxidant at 150 V for 45-60 min, with BenchMark Pre-stained Protein Ladder (Thermo Fisher). The gel was incubated overnight with InstantBlue Coomassie stain (Expedeon) at 4°C. Gels were imaged on a standard flatbed desktop scanner and then dried between sheets of acetate. Proteins were quantified using Qubit 3.0 Fluorometer (Life Technologies, Thermo Fisher) following the manufacturer's instructions.

2.2.2.3 Oligonucleotide probe design

Fluorescein-labelled, double-stranded DNA oligonucleotides were designed (Integrated DNA Technologies) with the sequence of the LE9, N3 and SIMO enhancers, HPLC purified, along with identical unlabelled ‘cold’ probes. For the hot probes, the FAM label was present as a 5’ modification (5’ 6-FAM) on one strand of the duplex oligonucleotide (**Table 3**).

Enhancer/oligo name	Strand	Sequence (* 56-FAM in bold only present in hot probe)
LE9: LE9_Hot*/LE9_Cold	5’ to 3’	/56-FAM /ATG AGA GAT CTT TCC GCT CAT TGC CCA TTC AAA TAC AAT TGT AGA TC
	3’ to 5’	GAT CTA CAA TTG TAT TTG AAT GGG CAA TGA GCG GAA AGA TCT CTC AT
SIMO: SIMO_Hot*/SIMO_Cold	5’ to 3’	/56-FAM /TGT CAA TGC CTG AAG TGA TAC GCT CTG A
	3’ to 5’	TCA GAG CGT ATC ACT TCA GGC ATT GAC A
N3: N3_Hot*/N3_Cold	5’ to 3’	/56-FAM /TTT GTT ATA AAC TAA GCT TTC AGT AAT CCC AAA CAA AAG AGT CAG TT
	3’ to 5’	AAC TGA CTC TTT TGT TTG GGA TTA CTG AAA GCT TAG TTT ATA ACA AA

Table 3 Oligonucleotide DNA enhancer sequences for EMSA experiments

2.2.2.4 Electrophoretic mobility shift assay (EMSA)

The starting point for my EMSA experiments including the binding buffer composition was the method outlined in Bhatia *et al.* (11). Starting initially with 1% agarose gels, I rapidly switched to using native PAGE gels, 6% DNA Retardation Gel (Life Technologies) to improve resolution and separation. At the start of each EMSA experiment, the gel was assembled in the tank in cold 0.5x TBE buffer, and each well was flushed thrice using a 25G needle and syringe. The gel was prepared and run in a cold room at 4°C.

Protein samples for EMSA were thawed on ice from -80°C. A master mix was prepared containing binding buffer (25 mM HEPES-NaOH pH 7.6, 1 mM KCl, 1 mM DTT, 0.1 mM EDTA, 0.1 mM PMSF, 100 mM NaCl, 10% glycerol), FAM-labelled DNA oligos 1 µM (covered in foil to preserve the fluorescent label), and 1 µg/µl

poly(deoxyinosinic-deoxycytidylic), poly(dI-dC). 15 µl reaction mixtures were prepared on ice containing nuclease-free water, the master mix and protein eluate, mixing by pipette.

The DNA-protein reaction mixture above was incubated for 30 min at room temperature, covered. During this period, the pre-prepared empty gel was run at 100 V at 4°C until the samples were ready to load.

The samples were carefully loaded onto the gel with the first well containing Orange G dye in order to visualise progress. The gel was run, covered or in the dark at 4°C, for approximately 75 min (15-well gel) or 90 min (10-12 well gels). Replicates were performed as separate DNA:protein binding reactions on different days and with two batches of purified protein.

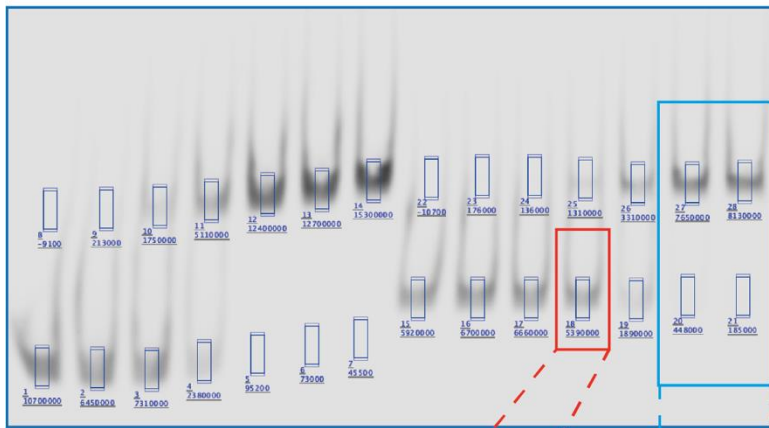
2.2.2.5 Imaging and quantification of EMSA

EMSA gels were imaged on the FLA-5100 Fluorescent Image Analyser (Fujifilm) with 473 nm laser and DBR1 filter (530DF20). Densitometry of the raw TIFF image files of the gels was performed using Image Studio Lite software (LI-COR Biosciences) using a protocol I developed, following principles outlined by Gassman (206) and Miller (207) and is depicted and described in **Figure 10**. This software allows consistent, reliable densitometry including an unbiased background subtraction performed for each band, and stores the raw data and image analysis separately. Densitometry values for the bound and unbound bands were exported, allowing calculation of the proportion of binding for each lane.

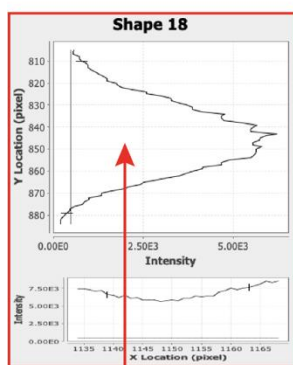
A



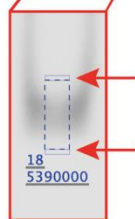
B



C



D



E

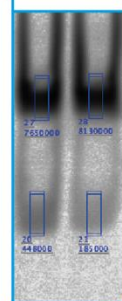


Figure 10 Densitometry of EMSA gel images. (A) Raw TIFF image of EMSA gel. (B-E) Screenshots from the densitometry process developed for EMSA using Image Studio Lite. Raw TIFF images of the gel were imported into the software. (B) The bound and unbound bands were quantified as follows. Long rectangles of a constant size were placed vertically across each band (covering the middle third of the width of the band (206); longer than the height of the band). An orange box highlights one of these bands, the one shown in C and D. (C) Pixel intensity profile of the highlighted band, obtained by clicking on the rectangle over the band ('shape 18' here). The symmetry of this intensity profile was used to ensure correct manual placement of the rectangle over the band in the vertical plane. The peak area (orange arrow) was used for densitometry. (D) Close-up view of the highlighted band, with orange arrows showing the strips at the top and

bottom of the rectangle: the mean pixel density of these strips (which fall above and below each band) was used by the software to perform the background subtraction. (E) shows a useful feature of this software, which allows the display of the gel to be changed without altering the raw data or pixel density being measured. In B, the lower bands in the blue highlighted area are virtually invisible, whereas changing the display reveals them allowing correct placement of the rectangle for quantification

2.2.2.6 Analyses and statistics for DNA-protein binding study

The analysis of quantified EMSA binding data was performed using GraphPad Prism (versions 8-10, GraphPad Software). For binding curves, the analysis method selected was a 'Specific binding with Hill slope', a non-linear fit using least squares regression with no constraints placed on Bmax, h (Hill slope) or Kd. This allowed calculation of the protein concentration at which 50% of the DNA would be bound (corresponding to the Kd, the equilibrium dissociation constant), and this was calculated for each batch of wild type protein and DNA combination used.

The Kd of the wild type was used as the concentration to compare to the binding of the missense variant proteins. Data on the binding of different PAX6 variants was normalised that of WT and expressed as a proportion of this (both normalised and non-normalised are presented in the Results). Binding of different variants was compared using repeated measures one-way ANOVA with Dunnett's multiple comparison test. Alpha was set at 0.05.

2.3 Results

2.3.1 Defining the distinctive phenotypic spectrum of PAX6 missense variants

Collation of available published and unpublished *PAX6* missense variant cases for phenotyping resulted in 161 missense cases from 110 families, and a size-matched control group of nonsense and frameshift variants predicted to result in nonsense-mediated decay (NMD); the latter group we will refer to as likely gene-disrupting (LGD) variants (208); they are equivalent to a loss of function group. As I shall describe, analysis of these cohorts showed that the missense group had a distinct and far broader phenotypic spectrum as compared with the LGD group.

I performed clustering analysis of the detailed phenotypic data which is described below and in the heat map, **Figure 11**. A simpler categorisation was required for some of the additional analyses, so, informed by the clustering, I categorised all the cases into four phenotypic categories based on predominant phenotype: classical aniridia, microphthalmia-coloboma, mild phenotypes (aniridia spectrum or mixed), and anterior segment dysgeneses. The latter comprised cases where the predominant features included congenital corneal opacification, iridocorneal adhesions, corectopia and/or posterior embryotoxon. In addition to the analyses presented here, a detailed phenotyping table of the MAC cases is presented in Table 1 of the publication (195) in **Appendix A.1**, and the full phenotypic data collection spreadsheet in Supplementary Table S1 of the same paper.

When comparing *PAX6* missense cases to a size-matched control of n=161 likely gene-disruptive LGD cases, the effect of variant class on phenotype was striking (**Figure 11** and **Figure 12**, $p < 0.001$). The phenotypic spectrum associated with missense variants was notably wider than with LGD, with distinct phenotypic clusters emerging. In particular, MAC cases were almost exclusively missense and clustered together.

The proportion of non-classical aniridia cases was higher in the missense group (**Figure 12**). The proportions of individuals with aniridia ($p < 2.2 \times 10^{-16}$) and with microphthalmia ($p = 2.811 \times 10^{-14}$, Fishers' exact test) are different between the two variant classes.

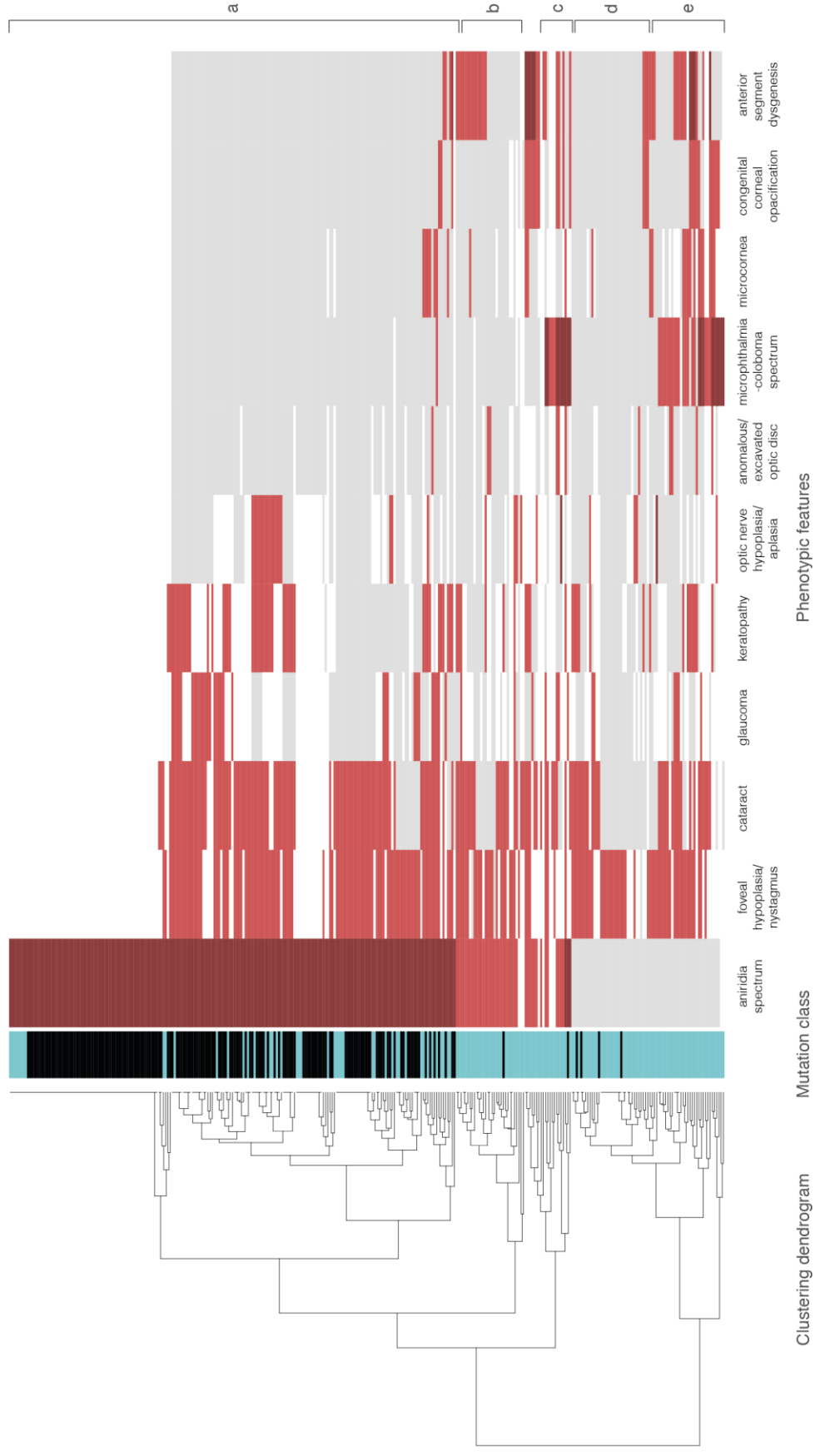


Figure 11 Phenotype-driven clustering analysis of PAX6 eye malformation cases. The heat map displays phenotypic information for all PAX6 missense variant cases (n=161 individuals, turquoise) and a random size-matched sample of likely gene-disruptive (LGD) cases (n=161, black; nonsense and frameshift variants with predicted nonsense-mediated decay). The phenotypic clustering and heatmap were produced in R (using "hclust" and "heatmap.2" functions, and within the latter using Euclidean distance as the metric argument within the function "daisy"); the development of the R script for clustering was performed with Dr Liusaidh Owen (FitzPatrick lab). The R script and alternative heat maps showing the effect of altering the distance metric, which is an important methodological consideration with a potentially big influence on clustering results, are available in **Appendix A.2**. The dendrogram on the left shows the output of the clustering algorithm, and the ocular phenotypic features are scored on the x-axis (see heatmap key). Patterns and clusters can be discerned and are annotated on the right of the figure. The phenotypes tend to differ by variant class: the LGD cases are predominantly associated with the classical aniridia spectrum (a), whilst the missense variants lead to a wider phenotypic spectrum with distinct clusters. The latter include mild aniridia spectrum disease (b) and predominantly-isolated foveal hypoplasia (d). The severe microphthalmia/MAC cases tend to cluster together, into those with aniridia (c) and without (e). The aniridia-associated features of foveal hypoplasia, cataract, glaucoma and keratopathy are more evenly distributed across both variant classes.

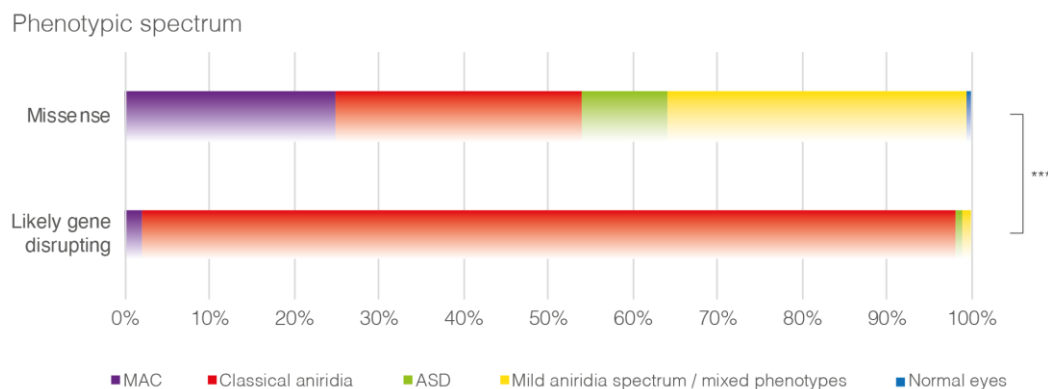


Figure 12 Percentage component bar chart illustrating the phenotypic spectrum, in broad categories, of *PAX6* eye malformations caused by missense vs LGD variants. Each line represents $n=161$ individuals with *PAX6* variants, classified into one of 4 phenotypic categories which, in order of appearance from left to right, are: MAC, microphthalmia-coloboma spectrum (purple); classical aniridia (red); anterior segment dysgeneses (green); mild aniridia spectrum/mixed phenotypes (yellow). The phenotypic spectra of LGD vs missense variants were different ($p<0.001$, Pearson's Chi-squared, with $X^2 = 158.41$ and $df = 4$). The overwhelming majority of LGD variants resulted in classical aniridia phenotypes. Missense variants resulted in 25% MAC phenotypes, 29% classical aniridia, 16% anterior segment dysgeneses (where features such as congenital corneal opacification or posterior embryotoxon predominated), and 35% hypomorphic phenotypes. 1 individual in the missense group, with p.(Arg242Thr), had no apparent ocular phenotype (blue, far right), and was included as the relative of an individual carrying the same variant with a mild iris phenotype.

2.3.2 Distribution of variants amongst the domains of PAX6

The distribution of variants amongst the domains of the PAX6 protein differed between variant class ($p < 0.001$, Pearson's Chi-square test, **Figure 13**), with disease-associated missense variants disproportionately likely (but not exclusively) to be found in the paired domain. Specific, recurrent missense variants in the paired domain were associated with severe microphthalmia phenotypes and are described in the following section (2.3.3).

Distribution of variants between PAX6 domains

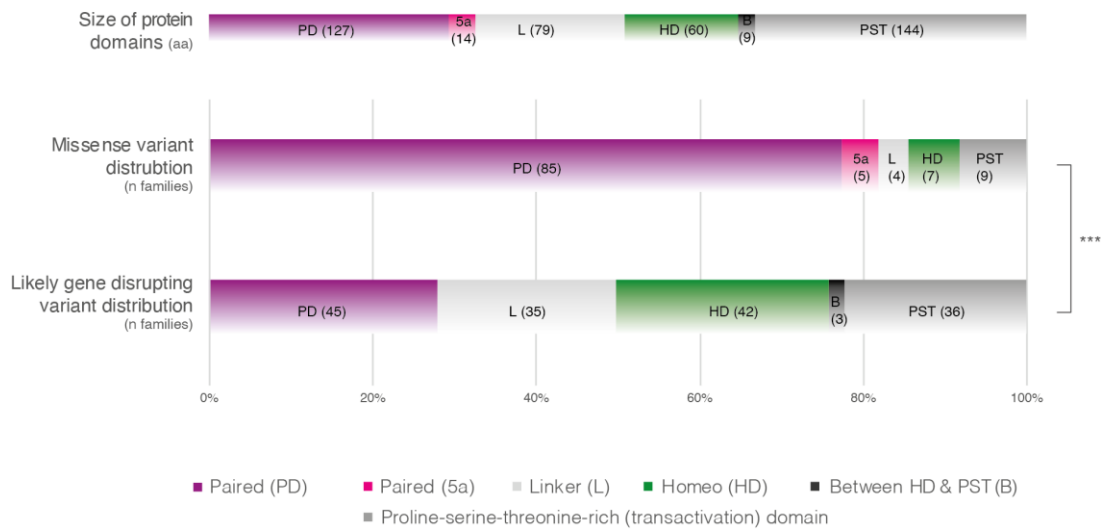


Figure 13 Percentage component bar chart showing the distribution of variants amongst the different protein domains of PAX6, for missense cases (n=110 families total) vs likely gene-disrupting cases (LGD, n=161 families). The relative sizes of each domain (in amino acids) is shown above; note the alternatively-spliced 5a domain data is shown after rather than in its 'true' location within the paired domain (PD) here. The distribution of variants amongst the different domains of the protein differed between the two variant classes ($p < 0.001$, Pearson's Chi-square test, with $X^2 = 75.848$ and $df = 5$). Missense variants are disproportionately likely to affect the paired domain, whereas there is a more equal distribution of LGD variants amongst the different domains, closer to what would be expected by chance based on the size of the domains.

Within the missense variants, the different categories of phenotypes exhibited were distributed as shown in **Figure 14**. As expected, the presumed benign variants present in gnomAD (dark blue) are much sparser in the DNA binding domains of PAX6. The disease-causing variants are clustered in the DNA binding domains, predominantly paired domain. All the variants in the alpha helix of the homeodomain are associated with MAC.

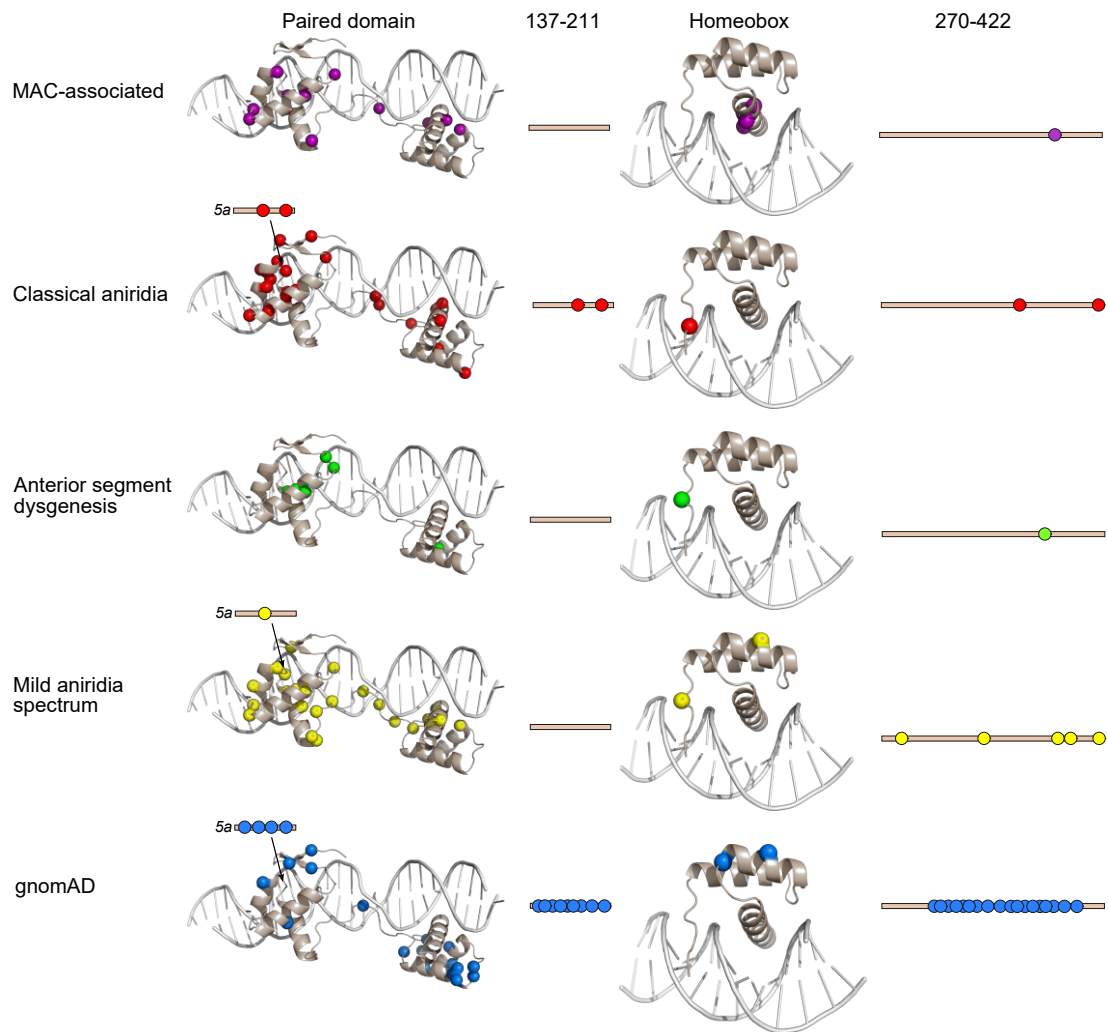


Figure 14 Location of missense variants within PAX6 stratified by broad phenotype. Model of PAX6 (isoform NP_000271.1, 422 amino acids) DNA-binding domains: paired domain (PDB ID: 6PAX) and homeodomain (homology model of the complex based on PDB ID: 3INQ) bound to dsDNA with residues coloured according to the broad phenotype(s) associated with their substitution. Top line, substitutions associated with a predominantly MAC phenotype (magenta/purple); second line, classical aniridia phenotype; third line, anterior segment dysgenesis; fourth line, milder aniridia-spectrum phenotypes. Missense variants present in gnomAD (therefore putatively benign) are shown on the bottom line, notably much sparser in the DNA binding regions. Figure from the Supplemental Information of our paper (195) created by Joe Marsh based on my phenotypic classification.

2.3.3 Missense variants in key residues in the PAX6 paired domain cause structurally very severe eye phenotypes

The FitzPatrick and, prior to that, van Heyningen labs had identified recurrent missense variants in the paired domain of *PAX6* amongst individuals in the HGU eye cohort responsible for microphthalmia-coloboma spectrum cases: out of a total of 372 individuals with bilateral MAC, 17 individuals from 15 families had been identified as having eight different, ultrarare heterozygous *PAX6* missense variants. As expected, these cluster together in the phenotypic analysis presented above.

To complement the full phenotypic spectrum outlined in the sections above, encompassing MAC, classical aniridia, other anterior segment dysgenesis and milder phenotypes, the pivot table in **Appendix A.3** gives a per-residue overview of the broad phenotypes associated with amino acid substitutions in *PAX6*, highlighting recurrently substituted residues. However, here we focus on the more severe end of this spectrum.

Focussing on the MAC cases, seven of these cases altered residues in the *PAX6* paired domain: p.(Arg26Gln); p.(Gly36Val); p.(Arg38Trp), 2 cases; p.(Arg38Gln); p.(Gly51Arg), 2 cases; p.(Ser54Arg), 2 cases; and p.(Asn124Lys), 5 cases; and one the homeodomain, p.(Asn260Tyr). (The structural aspects are discussed in the next section but for visualisation of residue locations please skip down to **Figure 16 D** below.) At the start of this project these cases were unpublished, but we have now published full phenotypic information for these variants (Table 1 in **Appendix A.1** (195)).

Some of these phenotypes were more severe than those few MAC cases that had been previously described. As part of this project, I collated and examined the phenotypic information for these most severe cases in detail, including photographs where available, as it is central to the experimental work to follow.

The phenotypes of the p.Ser54Arg and p.Asn124Lys were of particular interest and are outlined along with the variant characteristics in **Table 4**. In brief, they all had microphthalmia and in the 6/7 with detailed phenotyping information including level of vision, the individuals were blind with a bilateral severe eye malformation. 4/7 had congenital corneal opacification/sclerocornea; 2/7 had coloboma (chorioretinal/optic

disc); 3/7 had lens subluxation; 2/7 had retinal detachment; 1/7 had aniridia (though of note, in a very severe microphthalmia it is not always possible/meaningful to assess the iris); 1 had aphakia; and 2/7 had mild intellectual disability and other systemic phenotypic features.

Photos of the eyes of several of these individuals are seen in **Figure 15**.

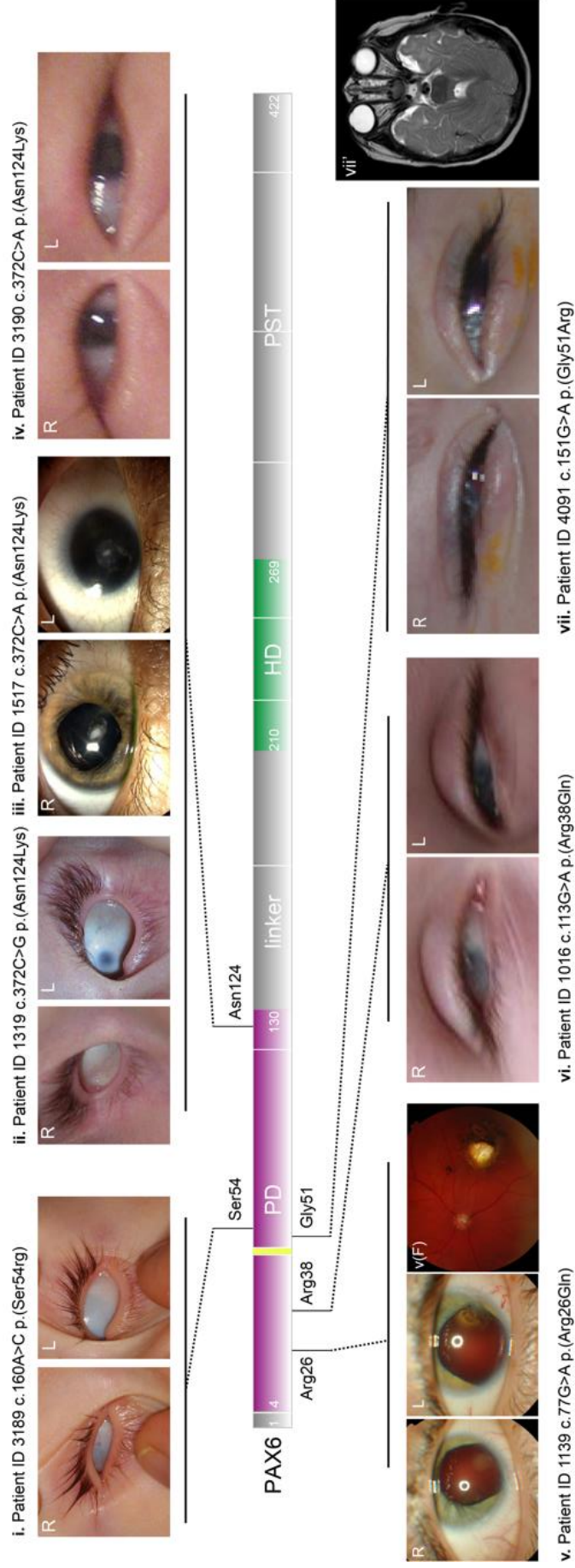


Figure 15 Clinical photographs from patients with severe eye malformations secondary to *PAX6* missense variants, adapted from (195). The ideogram of *PAX6* (isoform NP_000271.1, 422 amino acids) shows the position of the variants with respect to the paired domain (PD, purple), homeodomain (HD, green), the alternatively-spliced exon 5a (14 amino acids, yellow box) that are present in the *PAX6* isoform NP_001595.2 (436 amino acids), and the *PAX6* coding exon boundaries (white vertical bars). None of these severe malformations were seen in variants in the PD-HD linker region or the C-terminal Proline-Serine-Threonine-rich transregulatory region, which are labelled. Above the ideogram are the clinical photos (images i-iv) of individuals with *PAX6* p.(Ser54Arg) or p.(Asn124Lys) PD variants, which are recurrent and uniquely associated with severe, SOX2-like bilateral eye malformations. Images v-vii, below, show the eye phenotypes of individuals with *PAX6* PD variants; these are recurrent and are associated with more variable MAC and aniridia phenotypes. v(F), fundus image of patient ID 1139 with bilateral choroidal coloboma; vii, magnetic resonance image (MRI) showing bilateral mild microphthalmia in patient ID 4091. L left eye, R right eye.

PAX6 residue, helix, domain		Ser54, αIII PD			Asn124, αVI PD			
Family ID	3189	494	3190	1517	1319	339	5	
Patient ID	3189	494	3190	1517 ^a	1319	339	5	
chr11 GRCh38	g.31801758T>G		g.31801756A>C		g.31800842G>C			
ENST00000643871.1, NM_000280.4	c.160A>C		c.162T>G		c.372C>G			
ENSP00000492315.1	p.(Ser54Arg)		p.(Asn124Lys)		p.(Asn124Lys)			
Previously reported	No	No	No		No			
Residue conserved in paralogs	Yes				No			
Causative substitution of residue in paralog	Yes				No			
Inheritance	de novo	de novo	de novo	de novo	de novo	ND	ND	
Sex	Male	Male	Male	Female	Female	Male	Female	
Age at last assessment y	1	15	1.5	13	6	36	60	
Ophthalmology								
MAC spectrum	BL severe M, sclerocornea	BL M	BL severe M, choroidal C, sclerocornea	L M, BL retinal and optic disc C	BL extreme M, UL microcornea, sclerocornea	M,C	BL M, microcornea, sclerocornea	
Iris phenotype	Normal	NA	None	L aniridia, R abnormal architecture	None	NA	None	
Other ocular features	R congenital aphakia, L small lens, mild ONH	ND	Lens subluxation	BL cat, L lens subluxation, nystagmus, keratopathy	L posterior cyst & PFV, R disc abnormality: possible Morning Glory	Retinal detachment, UL phthisis	High myopia, nystagmus, L eye cat & lens subluxation, retinal detachment & phthisis	
Visual acuity (logMAR)	Blind	Blind	Nearly blind	Blind: 1.3 R, PL R	Blind: NPL R, PL L	NA	Blind: BL NPL	
Neurology, development & additional features								
Neuroimaging			Slightly enlarged lat ventricles		Normal MRI			
Intellectual disability		Mild	Mild	No	No	NA	No	
Additional clinical features	Patent foramen ovale	Short stature, microcephaly, protruding ears, downward slanting palpebral fissures	Mild microcephaly, hearing loss, UL cryptorchidism, congenital hyperinsulinism, epilepsy, autism	None	BL epicanthus	NA	None	

Abbreviations: BL, bilateral; Cat, cataract; C, coloboma; CF, counting fingers; FH, foveal hypoplasia; lat, lateral; M, microphthalmia; HM, hand movements; L, left; NA, not available; ND, not determined; NPL, no perception of light; ONH, optic nerve hypoplasia; Pat, paternal; PD, paired domain; PFV, persistent foetal vasculature; PL, perception of light; R, right; UL, unilateral; y, years. ^a, Subject 19 in Yogarajah *et al.* (77) with no individual clinical details given and analysed as an embedded component of a brain MRI grouped dataset

Table 4 Table adapted from (195), with thanks to Kathy Williamson, detailing the phenotypes of the individuals with novel p.Ser54Arg and p.Asn124Lys variants

2.3.4 Protein structural analysis per residue informed the selection of PAX6 variants for DNA-protein binding studies

Although it is well established that missense variants can lead to loss of function or hypomorphic phenotypes, why they also cause these rare “worse than null” phenotypes is unclear. We investigated their structure first, via protein molecular modelling performed by collaboration with Dr Joe Marsh.

This is his team’s work on my phenotypic categorised data (detailed in our paper (195)). To summarise briefly, it showed that MAC-associated variants are associated with a predicted effect to perturb DNA binding. The 86 missense variants from 161 individuals were classified on the basis of the FoldX results in to 3 categories

1. Those predicted to disrupt protein folding
2. Those predicted to disrupt the interaction with DNA
3. Unknown molecular effect

Over half of the MAC-associated variants (12/22) were predicted to perturb DNA binding. Only 10/64 of the non-MAC variants fell into this category (**Figure 16**). Our hypothesis was therefore that the missense variants may affect interactions of the paired domain with DNA, and so I selected variants for DNA-protein binding studies.

The two variants responsible for severe microphthalmia, p.Ser54Arg and p.Asn124Lys, were of primary interest. For comparison, I selected two variants whose phenotype was entirely classical aniridia, p.Cys52Arg and p.Ser121Leu, and one presumed neutral variant (present in gnomAD), p.Arg92Gln (rs769095184).

The classical aniridia variants were selected not only on the basis of published human phenotype but also predicted effect on protein structure and location, in discussion with Joe Marsh. The p.Cys52Arg variant associated with classical aniridia was predicted to disrupt DNA binding but not destabilise PAX6 folding. The p.Ser121Leu variant was surprisingly predicted to have only a mild impact structurally by FoldX, but was a residue in contact with DNA so it was hypothesised that it might disrupt or modulate the specificity of DNA binding.

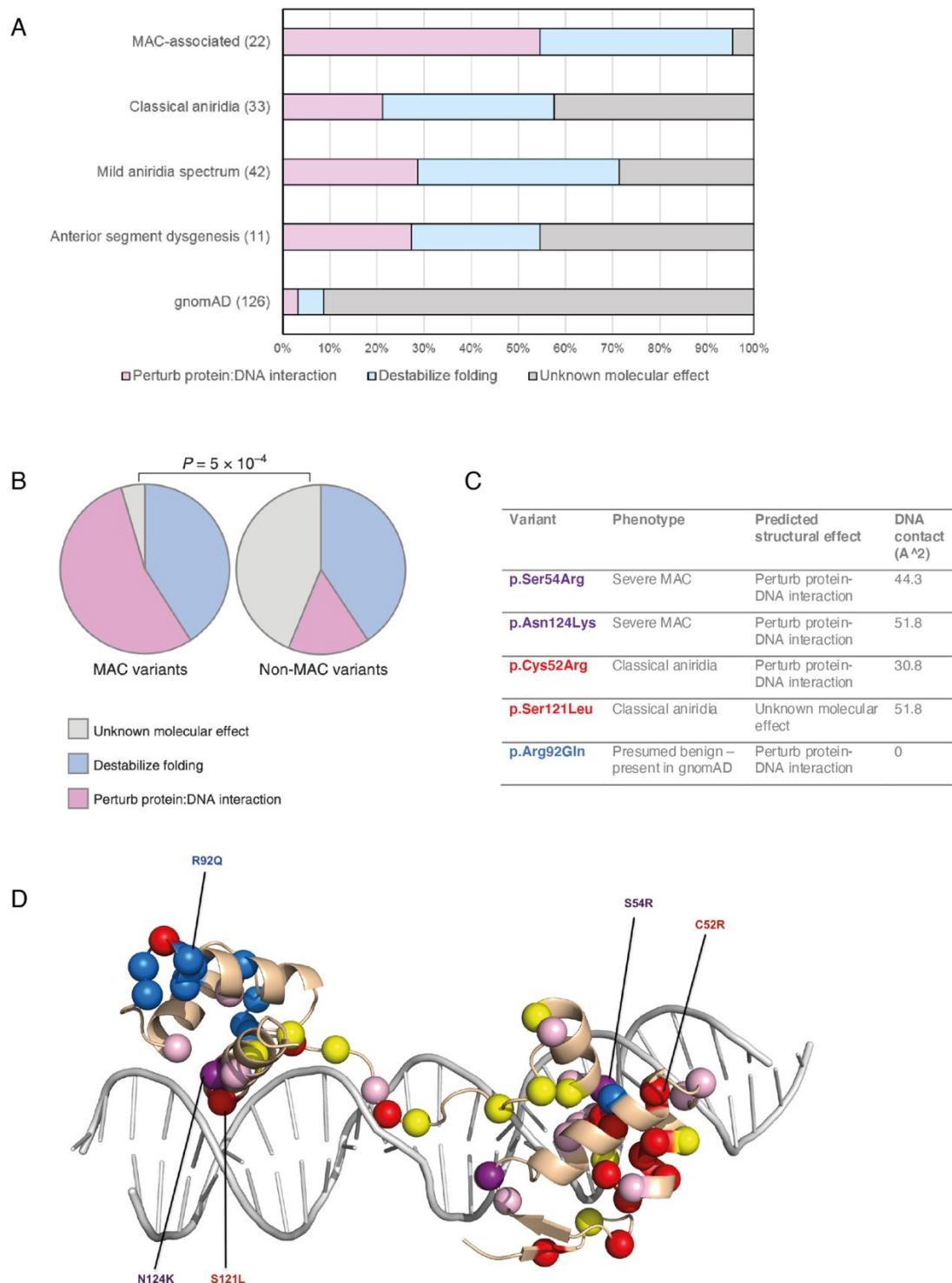
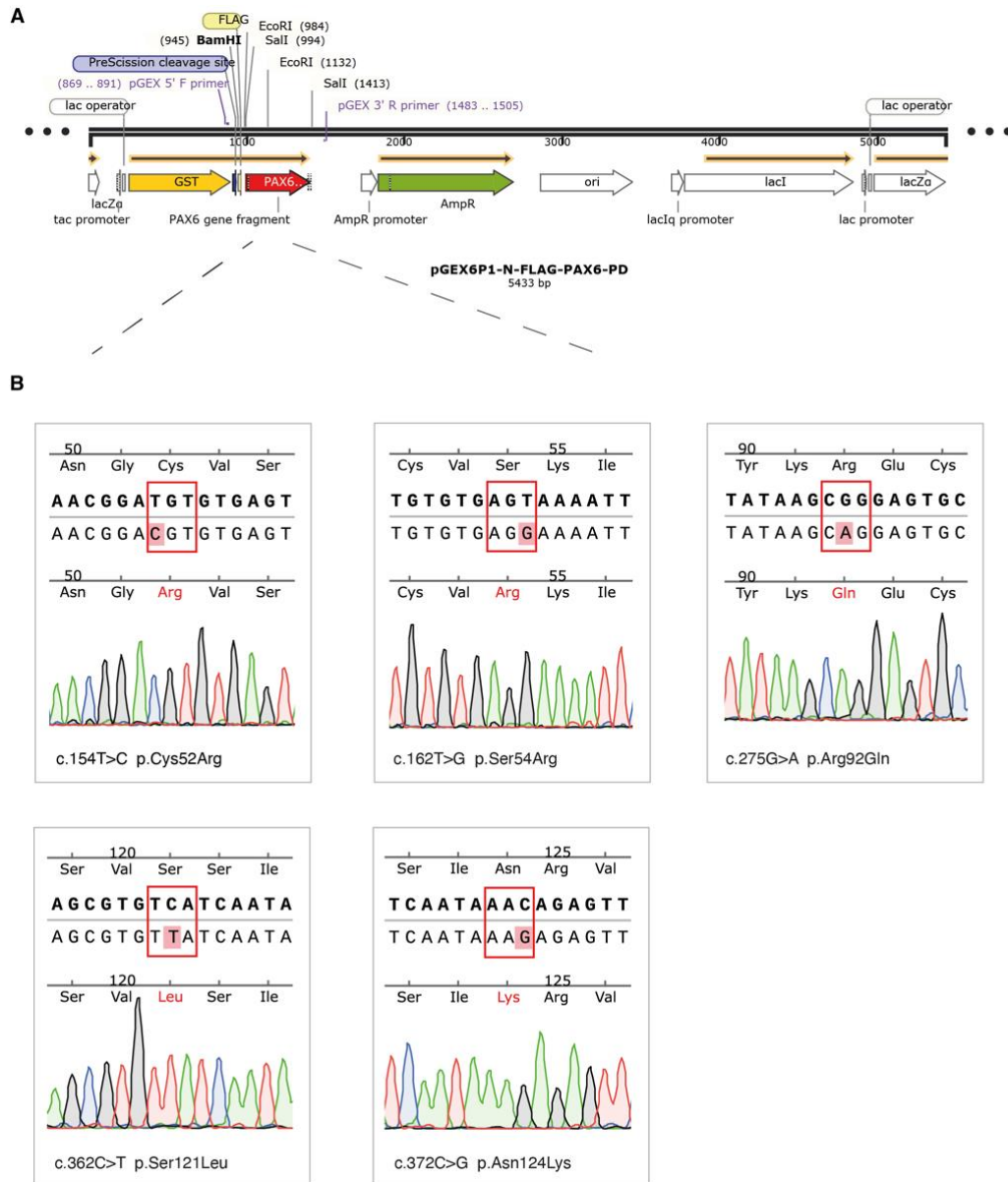


Figure 16 Protein structural analysis informing the choice of missense variants for further study. (A) Percentage component bar chart showing the predicted molecular/structural effects (using FoldX) of different missense variants stratified by their predominant associated phenotype. Based

on data from n=108 families, and n=126 missense variants which were listed in gnomAD (v2.1) and therefore putatively benign. (B) Data from the same analysis as (A), but comparing predominantly MAC phenotype associated variants with non-MAC. MAC variants are more likely to be associated with a predicted effect on DNA-protein binding. (C) Table showing the phenotype summary and predicted structural impact (FoldX) and DNA contact of variants selected for further study. The colour of the variants listed in the first column corresponds to the phenotype category shown in (D). (D) Model of the PAX6 (isoform NP_000271.1, 422 amino acids) paired domain (PDB ID: 6PAX) bound to dsDNA (based on a consensus binding sequence (116)) with residues coloured by the broad phenotype category associated with substitution. Purple, MAC; red, classical aniridia; yellow, mild aniridia spectrum; pink, mixed phenotype categories; blue, gnomAD (putatively benign). The missense variants selected for further study, listed in the table in C, are marked out here using their abbreviated names.

2.3.5 Purification of PAX6 paired domain protein variants of interest

In order to purify proteins for DNA-protein binding studies *in vitro*, I introduced substitutions encoding the missense variants of interest (p.Cys52Arg, p.Ser54Arg, p.Ser121Leu, p.Asn124Lys, p.Arg92Gln) into a plasmid containing the sequence encoding the PAX6 paired domain. The pGEX-6P1-N-FLAG-PAX6-PD plasmid (with the N-FLAG PAX6 paired domain sequence having been previously inserted by my postdoctoral colleague Hemant Bengani) was used and the variants were introduced by site-directed mutagenesis (SDM); key features of this plasmid and its structure after cloning are shown in **Figure 17**, along with Sanger sequencing results.



28

Figure 17 (A) Plasmid map created using SnapGene software (GSL Biotech), for the 5433 bp plasmid, pGEX6P1-N-FLAG-PAX6-PD. A fragment encoding the first 136 amino acids of PAX6 (of which the paired domain is amino acids 4-131, shown in bright red) was inserted such that its expression is under the control of the tac promoter, induced with the lactose analogue IPTG. This vector enabled bacterial expression of the PAX6 paired domain as a GST fusion protein with an N-terminal FLAG tag. The GST tag (yellow) was cleaved with PreScission protease at the site marked, and Glutathione Sepharose beads were used for protein purification. (B) Results of site-directed mutagenesis (SDM). Chromatograms generated from ab1 file from Sanger sequencing of plasmid clone following SDM to introduce variants of interest using pGEX primers (marked in A; F, forward and R, reverse). Chromatogram key: A, green; C, blue; G, black; T, red. SDM allowed the generation of the following PAX6 paired domain variants seen in the human population and human disease: p.Cys52Arg, p.Ser54Arg, p.Arg92Gln (presumed neutral variant), p.Ser121Leu and p.Asn124Lys.

Using this bacterial expression vector, the resultant purified protein eluates of the PAX6 paired domain are shown in **Figure 18**.

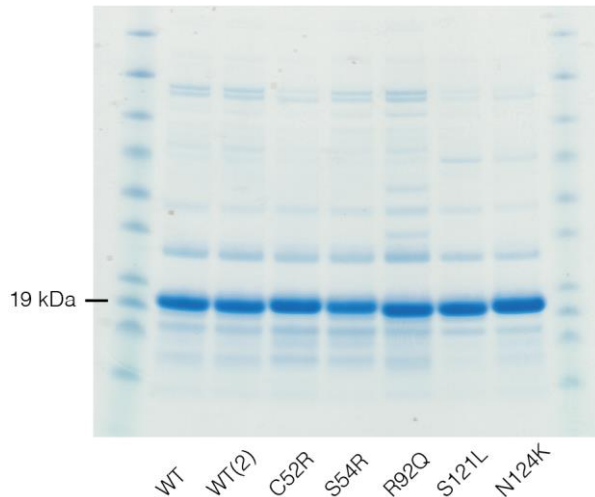


Figure 18 SDS-PAGE gel showing the purified PAX6 paired domain (PD) protein with 3.5 µg protein loaded per lane. 10 µl of BenchMark Pre-stained Protein Ladder are loaded each side, indicating molecular weights of 180 down to 6 kDa. The predicted molecular weight of the 136-aa portion of PAX6 plus FLAG is approximately 16 kDa. Two preparations of wild type PAX6 PD protein, WT and WT(2), were expressed using different colonies from transformation of the same plasmid DNA to increase yield. The variants created by site-direct mutagenesis of WT plasmid DNA were p.(Cys52Arg), p.(Ser54Arg), p.(Arg92Gln), p.(Ser121Leu) and p.(Asn124Lys).

2.3.6 Comparison of DNA-protein binding of different PAX6 paired domain (PD) variants

2.3.6.1 Choice of DNA target sequences

FAM-labelled, duplex oligonucleotide probes were created from the sequences of the LE9, SIMO and N3 enhancers (as described above in 2.2.2.3). LE9 and SIMO are both *PAX6* autoregulatory enhancers (i.e. they both bind *PAX6* and influence *PAX6* expression), whereas N3 is a *PAX6*-responsive *SOX2* enhancer. LE9 and N3 also bind *SOX2*.

LE9 was first identified in human lens epithelial cells, within the head surface ectoderm-specific enhancer region upstream of *PAX6* (209). The distal half of LE9 was shown to bind Pax6 in a paired domain-dependent manner (but interestingly not the alternatively spliced Pax6(+5a) isoform), and the proximal half of LE9 contains plausible high-mobility group (HMG) domain recognition sites, with Sox2 and Sox3 synergistically activating LE9 with Pax6 (209).

SIMO, previously introduced in **section 1.7.1.5**, is in the *PAX6* downstream regulatory region and also known as E+153 due to its position 153 kb away from *PAX6* (210). Structural variants (146) and a single nucleotide variant (11) disrupting SIMO have been described with the phenotype mimicking *PAX6* haploinsufficiency in the eye. The family with the translocation which led to the discovery of the SIMO enhancer (146) was originally published by Simola *et al.* (145), hence the name SIMO.

The *SOX2* regulator N3 (also N-3) is an autoregulatory enhancer, shown to be upstream of *Sox2* and active in the diencephalon, optic vesicle and lens placode: it is cooperatively activated by SOX2 and PAX6 (211).

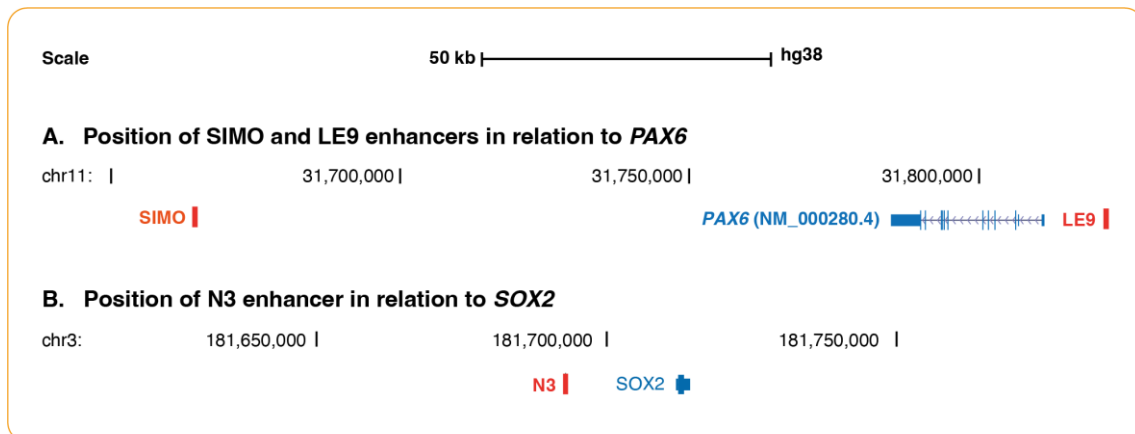


Figure 19 Genomic locations of enhancer sequences chosen as DNA targets for EMSA experiments. A 50 kb scale bar is shown at the top and all genomic coordinates are hg38. (A) The genomic positions of the SIMO (E+153) and LE9 enhancers, shown in red, in relation to *PAX6* (blue, coding exons shown by taller vertical bars and non-coding exons by shorter ones) on chromosome 11. SIMO is downstream and LE9 upstream of *PAX6*, which is transcribed on the reverse strand. (B) The genomic position of the N3 enhancer in relation to *SOX2* on chromosome 3. N3 is upstream of *SOX2*.

2.3.6.2 Optimisation of electrophoretic mobility shift assay (EMSA)

Binding of the purified PAX6 PD protein to the LE9, SIMO and N3 enhancer sequences in a dose-dependent manner was demonstrated on EMSA both using 1% agarose gels and then native PAGE, based on the methods outlined in Bhatia *et al.* (11) (**Figure 20**). Differences in the DNA-binding of different protein variants of interest was noted on pilot experiments (**Figure 21**) and then subsequently confirmed (**section 2.3.6.4** below).

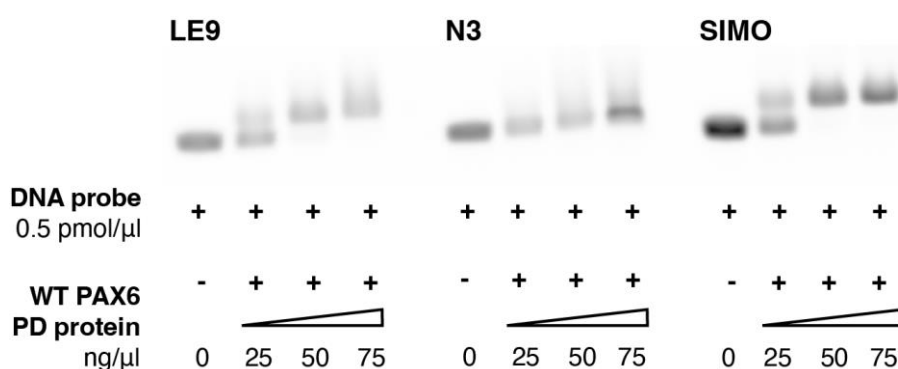


Figure 20 Initial EMSA experiment using purified PAX6 PD protein and FAM-labelled LE9, N3 and SIMO enhancer DNA sequences, with each 20 μl DNA:protein binding reaction run on a 1% agarose gel. Each DNA probe is present in all 4 lanes, respectively, at 0.5 pmol/μl, with the first lane being DNA probe alone. The subsequent 3 lanes also contain increasing concentration of wild type purified PAX6 PD protein at 25, 50 and 75 ng/μl. The N3 enhancer required higher concentrations of PAX6 protein to elicit binding (in keeping with observations of others (211)), and so due to constraints of the amount of protein available, the other two enhancer sequences, LE9 and SIMO, were chosen for the majority of the further experiments.

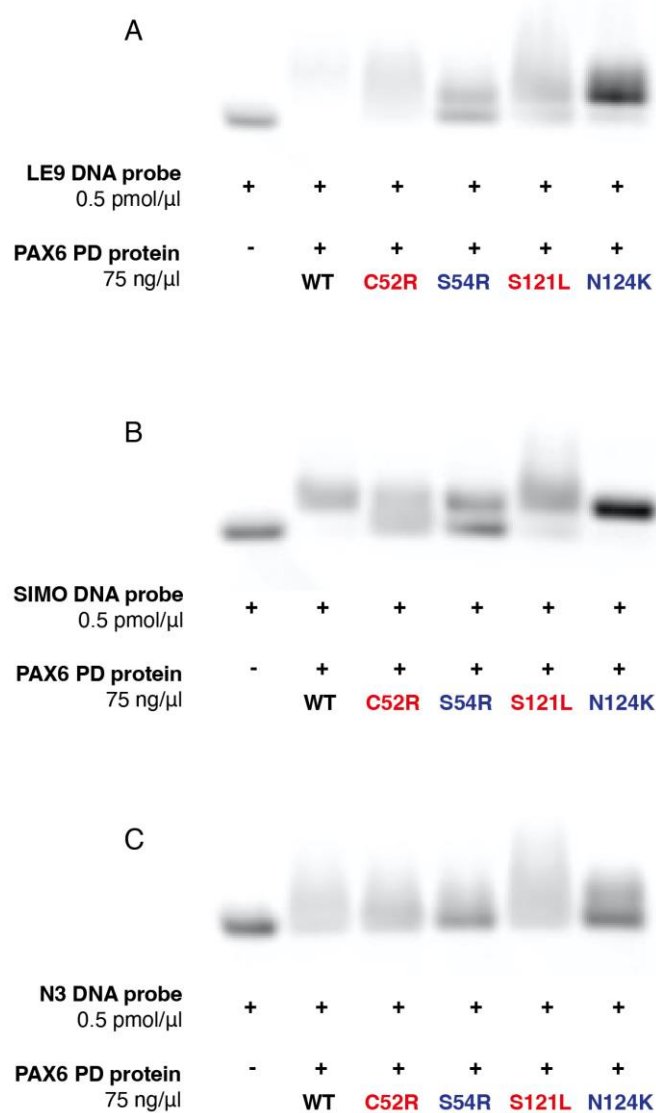


Figure 21 Initial EMSA comparison of the DNA binding properties of wild type (WT) purified PAX6 paired domain (PD) protein versus protein carrying different missense variants of interest. Binding was compared against 3 enhancer sequences: (A) LE9, (B) SIMO, (C) N3. The FAM-labelled DNA probe is present in all lanes, the first lane DNA probe alone and subsequent lanes containing purified protein 75 ng/μl. C52R, p.(Cys52Arg), and S121L, p.(Ser121Leu) are variants associated with classical aniridia and are labelled in red (for consistency with **Figure 16**), and S54R, p.(Ser54Arg), and N124L, p.(Asn124Lys), are severe MAC associated variants, shown in purple.

However, various modifications were taken to optimise the protocol, allowing better separation of the bands which in turn enabled me to develop a reliable densitometry quantification method and calculate DNA:protein binding curves and a K_d value for the wild type protein. This provided a more objective comparison between the variants.

Firstly, poly(deoxyinosinic-deoxycytidylic), poly(dI-dC) 50 ng/μl was added to reduce non-specific interactions (212). Secondly, moving from agarose to native polyacrylamide (PAGE) gels achieved better resolution and separation of the band shifts (contrast **Figure 20** **Figure 21** above to **Figure 22** below). Finally, I was able to develop a protocol for reliable quantification of the EMSA gels. This is described in the Methods (**2.2.2.5**), uses Image Studio Lite software (LI-COR), and adheres to principles of reliable densitometry outlined by Gassman (206), in particular: performing an unbiased background subtraction per individual band, performing the densitometry using a constant width approximating to the middle third of the band, and using the peak area rather than peak height (206) (illustrated in **Figure 10 C**).

In addition to the LE9 and SIMO enhancer DNA sequences, pilot data was also obtained for the N3 enhancer (such as in **Figure 21** above). Cold competition experiments were performed with all 3 enhancers, showing expected results for LE9 and N3 but not SIMO (**Appendix A.5.1, Supplementary Figure 3**). However, due to larger protein volumes required for each N3 binding curve experiment (and limited yields from protein expression), I decided to proceed with LE9 and SIMO only for further experiments.

2.3.6.3 PAX6 DNA-protein binding curves and calculation of K_d values

Binding curve experiments were performed with PAX6 wild type PD protein. A non-linear fit with the Hill slope was performed in Prism (GraphPad Software), with no constraints on any parameters (though constraining B_{max} to 1 did not significantly alter the results). This allowed calculation of the K_d for each wild type protein-DNA combination for variant comparison (**Figure 22**).

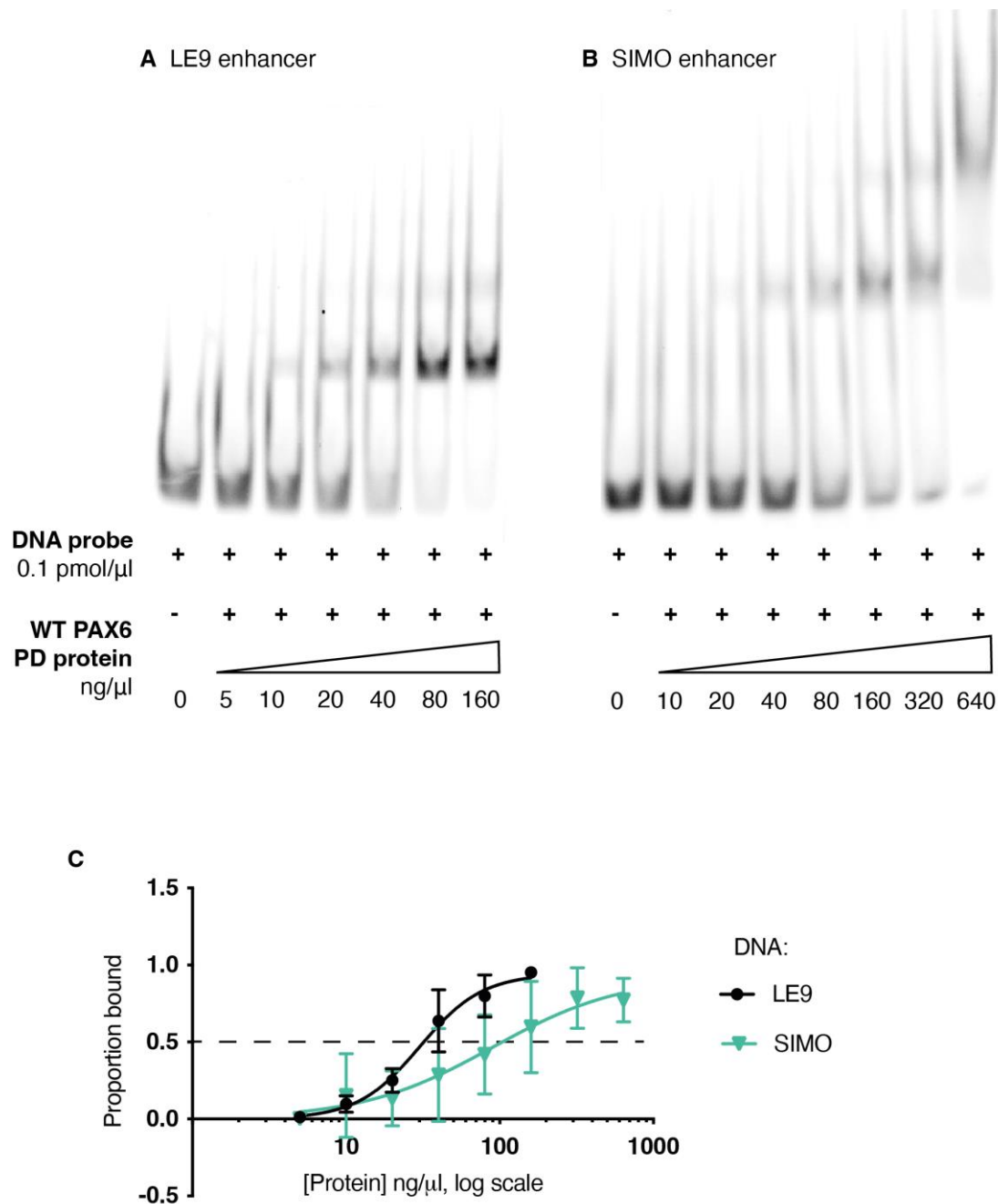


Figure 22 EMSA results showing binding curves of wild type purified PAX6 paired domain to LE9 and SIMO oligonucleotide probes. Representative EMSA gels are shown for (A) LE9 DNA, and (B) SIMO DNA. In each gel, FAM-labelled DNA probe is present in all lanes at 0.1 pmol/μl. The first lane is DNA probe alone, and subsequent lanes have increasing concentrations of purified PAX6 paired domain. As the protein concentration increases, so does the proportion of bound protein-DNA. (C) Quantification of EMSA binding curve experiments (n=5 for each DNA sequence), plotting the proportion of bound labelled DNA-protein complex (vs unbound DNA probe) on the y-axis against the increasing protein concentration. A non-linear fit (specific binding with Hill slope) allowing calculation of the K_d values for each protein-DNA combination: for this batch of PAX6 wild type protein and DNA probes, K_d for LE9 = 33 ng/μl and K_d for SIMO = 89

ng/μl. The gels from all the EMSA experiments that contributed to this graph are in **Appendix A.5.2**, along with the binding curve for the other batch of protein.

2.3.6.4 Comparison of DNA-protein binding of different PAX6 paired domain variants by EMSA

I used EMSA to test this subset of purified PAX6 paired domain missense variants in order to compare their binding to the LE9 and SIMO enhancers. I used the K_d concentration of the wild type protein (the concentration at which 50% was bound), as calculated in the experiments above, in order to standardise the comparison.

The results of these experiments showed significant differences in the DNA-binding of the variants, and the EMSA are presented in **Figure 23**. Full statistical details of the repeated-measures one-way ANOVA and multiple comparisons are in **Appendix A.6**.

One of the severe MAC-associated variants, p.Ser54Arg, caused an 85% reduction in binding to LE9 (95% CI 81 to 88%) but had no significant effect on SIMO binding, whilst the other, p.Asn124Lys, showed reduced binding to both (62% to LE9, 95% CI 44 to 80%; and 42% to SIMO, CI 28 to 57%). Of note, the N3 enhancer in the pilot (**Figure 21**) showed a very similar binding pattern to LE9. The presumed neutral variant p.Arg92Gln unexpectedly showed a similar binding profile to the phenotypically severe variant p.Asn124Lys.

In contrast, the classical aniridia variants, p.Cys52Arg26 and p.Ser121Leu, largely disrupted binding to both targets as expected. However, the binding of the latter to SIMO was variable and skewed by an outlying result as evident in **Figure 23**: this outlier though was binding in a similar pattern to the pilot in **Figure 21**, and as noted at the end of **Section 2.3.4**, this was a variant that had a more ambiguous structural prediction.

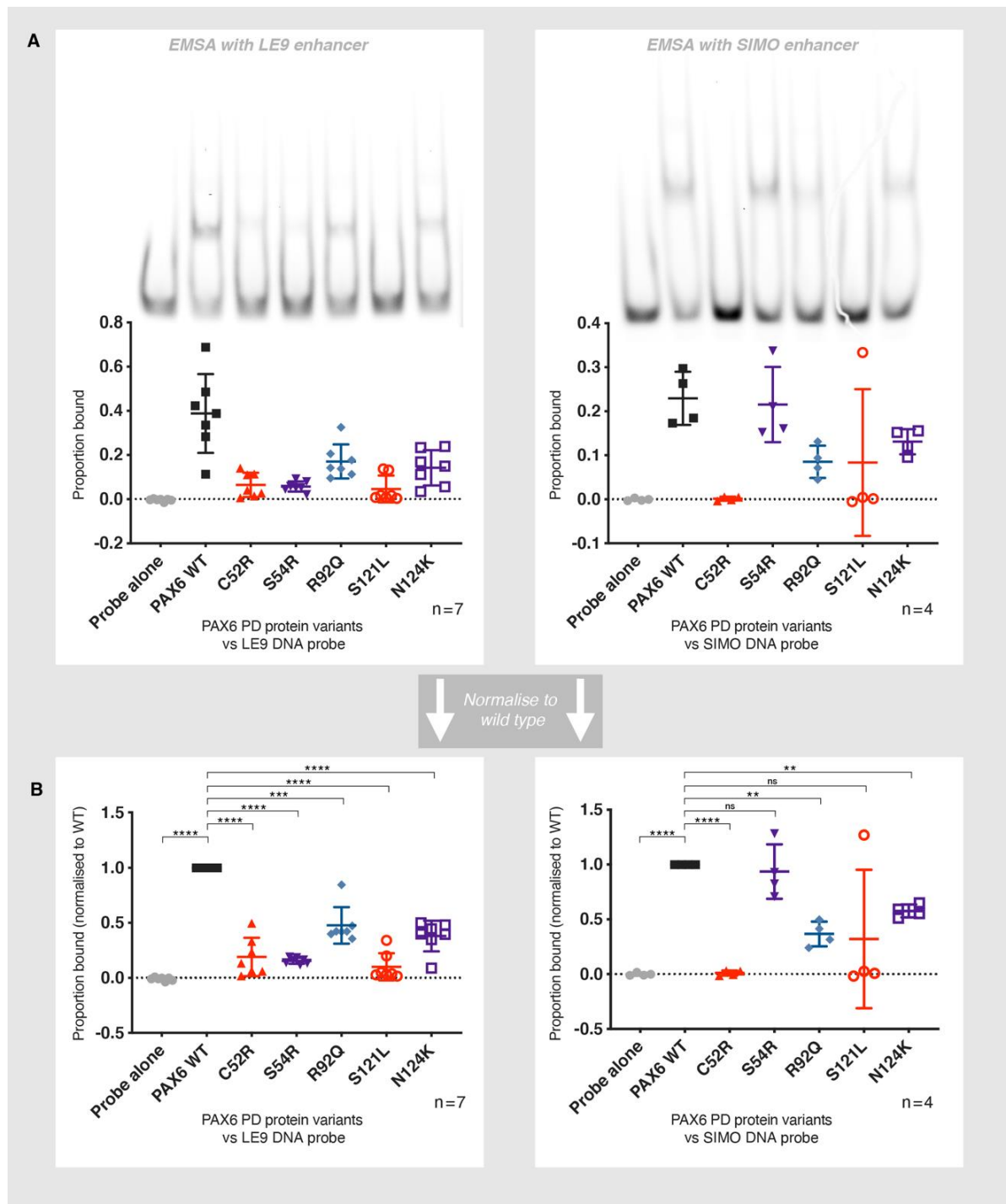


Figure 23 Effect of PAX6 variants on protein-DNA binding. Electrophoretic mobility shift assay (EMSA) comparing the binding of wild type and missense variants of the PAX6 paired domain (PD) protein to two known DNA targets of PAX6, the LE9 enhancer (left hand panels) and the SIMO auto-enhancer (right hand panels). The DNA probes were FAM-labelled duplex oligonucleotides at 0.1 pmol/μl. The purified paired domain variants were used at the K_d concentration of the wild type (WT) proteins, as determined by binding curves for each batch. (A) Representative EMSA gel image with the corresponding densitometry quantification of multiple

experiments below it. Error bars on the scatter dot plots show mean $\pm$ SD. The remainder of the EMSA gels which contributed to the graphs in this Figure are shown in **Appendix A.5.3**. (B) The same data shown normalised to the WT protein (lane 2 in all panels, black) within each experiment. Overall, the different PAX6 proteins showed differences in binding to LE9 ($p<0.0001$) and to SIMO ($p=0.0393$, one-way ANOVA). The severe MAC-associated variants (lanes 4 and 7, purple) showed different binding profiles to the two PAX6 DNA targets: p.Ser54Arg showed severely reduced binding to LE9, but not to SIMO; p.Asn124Lys showed a moderate reduction in binding to both targets. As anticipated, the classical aniridia-associated variants, p.Cys52Arg and p.Ser121Leu (lanes 3 and 6, respectively, red), mostly showed severely reduced binding. The presumed neutral variant present in gnomAD, p.Arg92Gln (lane 5, blue), showed moderate reduction in binding to both targets in a similar pattern to p.Asn124Lys. The statistical significance indicators above the graphs show the results of Dunnett's multiple comparison test.

These *in vitro* results support the premise that the difference in ocular phenotypes between individuals carrying *PAX6* missense variants may in part be related to alteration of the DNA-binding properties of their PAX6 proteins.

2.4 Discussion

I have shown via an unbiased clustering methodology that the human ocular phenotypic spectrum of PAX6 missense variants is greater than that of likely gene disruptive variants. After describing the cluster of severe MAC phenotypes which are due to recurrent missense variants in the paired domain, I provided some potential mechanistic insight via DNA:protein binding studies which demonstrate perturbation of DNA binding to different PAX6 targets *in vitro*.

Central to this work was describing and publishing the recurrent *PAX6* missense variant cases in the paired domain which cause very severe microphthalmia, which seem to be a “worse than null” effect – where the null phenotype is of course classical aniridia. In some respects the phenotypes of these variants are reminiscent of *SOX2*-like phenotypes (213), though we identified no cases of anophthalmia (and indeed there are no known monoallelic *PAX6* anophthalmia cases). In my opinion, these cases also bear a striking phenotypic similarity to the group of patients discussed in the next chapter with *MAB21L1* missense variants, an observation that has not been explored yet mechanistically. Whilst *PAX6* variants have for some time been recognised as a possible cause of microphthalmia, accounting for 1-2% cases (16,214), and indeed mild microphthalmia is common (and perhaps often unrecognised) in classical aniridia, this study defines the severe end of the microphthalmia spectrum and highlights the recurrently hit residues in the paired domain.

The most phenotypically severe variants were p.Ser54Arg and p.Asn124Lys. In EMSA experiments, the effect of the former was to severely reduce, but not abrogate, DNA binding to LE9 compared to wild type PAX6 paired domain, with no effect on SIMO binding. By contrast, p.Asn124Lys moderately reduced binding to both targets. The effect of both variants on perturbing DNA interaction is consistent with protein structural analysis and these residues both have significant contact with DNA. The relative positioning of the residues within their subdomains of the paired domain and their conservation are also of interest. (These aspects are explored further in the Discussion of our paper (195) **Appendix A.1.**)

Further work using saturation mutagenesis in a yeast one-hybrid assay, by my HGU colleagues Alex McDonnell (PhD) and Grzegorz Kudla (Principal Investigator), confirmed this effect of reduced affinity of Ser54Arg to LE9, and also found increased binding of this variant to BLX, a SELEX-derived PAX6 and SOX2 binding site (215). These results provide further support that these variants may alter the target activation of PAX6.

In future work I would like to explore the effect of these *PAX6* missense variants on binding to various enhancers in the presence of SOX2 (such as purified SOX2 HMG domain protein) to explore their effect on cooperativity. In addition to LE9, N3 (a SOX2 enhancer, discussed in 2.3.6.1 due to the pilot work here) and DC5 (δ -crystallin gene lens enhancer) would be useful targets to explore this with. Of note, the levels of transcription factor required to activate N3 and DC5 are different (211), with a several fold higher concentration required to produced cooperative binding with N3 *in vitro*.

With notable exceptions (discussed below), MAC phenotypes related to *PAX6* are overwhelming associated with variants in the paired domain (16,17), which is known to interact with SOX2 (216), the commonest gene associated with MAC (16). The LE9 enhancer has plausible HMG domain binding sites and is activated synergistically by *Pax6* and *Sox2* *in vitro* (209). The range of PAX6 binding sites *in vivo* is notably diverse (128,217).

Considering the results of the phenotypic and protein structural analysis, and DNA-protein binding experiments, we hypothesise that these “worse than null” MAC-associated variants alter, but do not abrogate, DNA binding, affecting the repertoire and/or degree of target activation of PAX6. This may involve LE9 itself, affecting *PAX6* autoregulation in the lens or the response to SOX2 (209), or other forms of SOX2 cooperativity such as PAX6-SOX2 co-binding to the DC5 enhancer to initiate lens development (216). Although failure of lens development can lead to microphthalmia (53), the phenotypes involved in severe MAC could also be consistent with disruption earlier in oculogenesis.

Two previous PAX6 variants, p.Gly36Arg and p.Arg44Gln, have been studied by Narasimhan *et al.* and found to have reduced DNA binding affinity but could not be shown to alter cooperativity with SOX2 (218). However, these substitutions cause classical aniridia rather than more complex phenotypes.

It is clear that the molecular properties of a variant are not sufficient to explain everything. The variant p.Arg92Gln (rs769095184) was chosen as a presumed benign variant because it was present in gnomAD (in eight individuals, East Asian, giving an allele frequency of 0.0004349). In my experiments, p.Arg92Gln in fact showed moderately reduced binding to both LE9 and SIMO (a similar pattern to p.Asn124Lys) which was at first surprising given the gnomAD data, but was in fact in keeping with the structural predictions which were to moderately perturb DNA interaction. After the above experiments, this variant was reported, albeit with limited segregation information, as being associated with classical aniridia in one Chinese family with four affected members (219). Overall these data would support this being a mild but vulnerable allele rather than an entirely neutral variant.

Since this work was performed and published (195), there have been significant advances in the field of computational structural and pathogenicity predictors. At the very start of this project a combination of CADD, SIFT and PolyPhen2 would have been a typical combination of variant effect predictors (VEPs) to consult, but now the advent of AlphaFold (220), AlphaMissense (221) and many others (222) has dramatically changed this landscape. The predictions of high performing modern VEPs (223–225), with some older predictors for comparison, are listed in **Table 5**, with the modern VEPs predicting the Arg92Gln allele to be much milder.

Table 5 Ten variant effect predictions for the PAX6 missense variants in this study. The order and colours indicate the year of preprint/publication of the VEP: AlphaMissense, CPT and SNPred 2023, green (221,222); EVE, MetaRNN 2021, turquoise (226,227); BayesDel 2017, blue (228); REVEL, SIFT4G 2016, violet (229,230); CADD 2014, pink (231); Polyphen2 2010, red (232). All the 5 most modern VEPs, in green or turquoise, accurately predict the phenotype and experimental data presented here for every variant, whereas the older VEPs are less consistently accurate. The bottom two rows show the default and PAX6-specific optimised thresholds, respectively, presented by Andhika *et al.* (224). Without the gene specific thresholds, only 1/5 of the older VEPs, SIFT4G (an update to the original SIFT), accurately predicts all the variants. With the gene-specific threshold, 2/5 do (SIFT4G, REVEL). For further sources and information regarding the nature of the VEPs listed, for example whether they are supervised or based on human population data, please see Table 2 in Livesey *et al.* (225), and (233).

variant	Alpha Mis- sense	CPT	SNPr ed	EVE	Meta RNN	Bayes Del	REVEL	SIFT 4G	CADD	Polyphen2 / \	
										_Hum Div	_Hum Var
Cys52 Arg	1.000	0.843	0.999	0.959	0.994	0.575	0.977	0.001	33.0	1.000	1.000
Ser54 Arg	1.000	0.840	0.999	0.962	0.988	0.552	0.969	0.001	30.3	1.000	0.993
Arg92 Glu	0.111	0.304	0.304	0.258	0.616	-0.012	0.717	0.176	28.2	0.992	0.473
Ser121 Leu	1.000	0.858	0.999	0.882	0.983	0.535	0.842	0.001	25.2	0.899	0.464
Asn124 Lys	1.000	0.781	0.996	0.932	0.938	0.245	0.929	0.003	28.1	0.997	0.923
thres- hold	≥ 0.560					> 0.069	> 0.500	≤ 0.050	> 20.0	≥ 0.447	
optim- ised	> 0.967					> 0.380	> 0.770	≤ 0.030	> 25.3	≥ 0.900	

Cohort studies have found missense variants almost exclusively in the paired domain (68) whereas here we show that disease-causing variants are principally, but certainly *not* exclusively, located in the paired domain (**Figure 14**), with homeodomain variants existing across all phenotypic categories. All the variants in the alpha helix of the PAX6 homeodomain, known as the recognition helix, were associated with MAC. One of these cases, first reported by our publication (195), is an individual with isolated

severe microphthalmia and p.Asn260Tyr. This residue is critical for DNA binding and interaction with the C-terminal subdomain of the paired domain in zebrafish (234), and is invariant in PAX homeodomains as demonstrated by Kathy Williamson in Figure S6 of our publication (195); namely, in PAX3, PAX4, PAX6 and PAX7. Variants in the recognition helix of the homeodomain in *PAX3* cause Waardenburg syndrome type 1 (235), and in *PAX7* a case of cleft palate is reported (236).

Limitations

Concerning the range of phenotypes analysed, we gave careful consideration as to how representative our cohort is. At one end of the spectrum, the HGU eye malformations cohort is notably enriched for severe eye malformations and MAC. Conversely, it and the PAX6 LOVD database (which formed the initial backbone of this cohort, and includes the HGU cohort and more) may be underrepresentative of the milder end of the phenotypic spectrum. These cases, such as isolated foveal hypoplasia (237), can also be due to LGD variants and may historically have been less likely to be screened for *PAX6* variants. A recent analysis of 86 patients with *PAX6* aniridia found the missense cases to have overall milder phenotypes (68). One caveat is that, whilst the groups in our phenotypic analysis were size matched, the LGD group were proband only, resulting in the missense group coming from fewer families than the LGD (110 vs 161). However, if this has introduced any bias the effect of this would logically be to diversify the LGD group, and this is the opposite to what was observed. On balance, we feel that the phenotypic analysis here accurately reflects the range of moderate and severe phenotypes associated with *PAX6* missense variants, but less so the range of milder cases.

Lastly, a further brief technical note. The wet lab work presented here involved purified PAX6 paired domain proteins. As evident from the SDS-PAGE gels in **Figure 18**, the purified proteins used in these experiments were not pure. We did commission the University's protein production facility to challenge them to improve on this, but they were unable to achieve significantly better purity without unacceptable compromise on yield. The small volume of protein produced was also a limit to these experiments as a new batch required new binding curve experiments.

Conclusion

In conclusion, the research presented in this chapter has defined the ocular phenotypes associated with missense variants in PAX6, with particular focus on recurrent paired domain variants responsible for the interesting MAC phenotypes at the severe end of this spectrum. These variants are likely to perturb DNA binding *in vivo*, as demonstrated *in vitro* here, and affect the repertoire of target activation of PAX6.

Chapter 3 Missense variants in *MAB21L1* cause eye malformations in human and mouse

3.1 Introduction

Please note: the majority of the work presented in this chapter was published in 2022 (166) and this paper, on which I am joint first author, is available open access and included as Appendix B.1. The parts of the published paper which overlap with sections below were written by me, and have been expanded upon to provide further detail for the thesis.

Monoallelic missense variants in the *MAB21L1* gene [MIM 601280] have recently been identified in seven unrelated families with aniridia and/or microphthalmia by the FitzPatrick lab and others (166,167). In this chapter, I characterise the phenotype of *MAB21L1* variants in both human and mouse (166).

MAB21L1, full name Mab-21 Like 1, is one of the paralogs of the *C. elegans* gene *mab21* involved in cell fate decisions and sensory organ development (238–240). The highly conserved MAB21-like proteins are a subfamily of nucleotidyltransferase (NTase) fold proteins. The best known members of these are cyclic GMP-AMP synthase (cGAS) and oligo adenylate synthase 1, which sense cytosolic DNA and double-stranded RNA, respectively, to trigger the innate immune response (241,242). Their enzymatic activity is dependent on this activation by nucleic acid binding and results in impressive signal amplification (243,244). Despite high sequence and structural homology with cGAS, much remains unknown about the molecular function of MAB21L1 including whether it retains NTase activity *in vivo*. The crystal structure of MAB21L1 is known, and it does demonstrate mild affinity for nucleic acids *in vitro* (245). If it does retain NTase activity, neither the activating ligand, substrate or effector molecule have been conclusively identified (244,245). Though the mechanism is not understood, the downstream response of MAB21L1 activity is to influence

transcriptional control, and there are multiple lines of evidence including strong human and animal genetic evidence for its role in ocular development (166,240,246–248).

Human *MAB21L1* is a single exon gene on chromosome 13q13, reverse strand, and nestled within an intron of the neurobeachin gene *NBEA* (249), which is transcribed on the forward strand. In humans, biallelic loss of function of *MAB21L1* is reported in at least 5 families causing corneal dystrophy, along with cerebellar, craniofacial and genital anomalies (250,251), with unaffected carrier parents. *Mab21l1* null mice have microphthalmia, only a rudimentary lens, no iris or ciliary body, a thin retina, and are infertile with preputial gland hypoplasia (the mouse preputial gland and human scrotum having the same ectodermal abdominal origin) (248).

Previous work from the FitzPatrick lab by Joe Rainger and colleagues (252), and by others (253), has demonstrated that human mono- and biallelic variants in the closely related *MAB21L2* cause severe MAC and limb malformations. Similar to *MAB21L1-NBEA*, *MAB21L2* and *LRBA*, which are on chromosome 4, are also a conserved nested pair within the metazoan genome (249). *Mab21l2* null mice have major developmental defects in not just the eye but also body wall and multiple solid organs, resulting in embryonic lethality (254). Heterozygote null mice for both *Mab21l1* or *Mab21l2* show apparently normal eyes.

Whilst some of their developmental expression overlaps, *Mab21l1* is expressed in the lens (247) whilst *Mab21l2* is expressed in the retina (240). *Mab21l1* expression is reduced in *Pax6* Small eye homozygotes, suggesting that its function is dependent on *Pax6* (248). In the same study, *Mab21l1* null mice showed normal *Pax6* expression, implying that the effects of *Mab21l1* are downstream of *Pax6*. ChIP-seq studies in mouse forebrain and lens chromatin suggest that *Mab21l1* expression is directly regulated by *Pax6* (128).

In this chapter, I characterise the ocular phenotype associated with *MAB21L1/Mab21l1* variants in humans and mouse, the latter in a CRISPR mouse line carrying one of the human Arg51 variants.

3.2 Methods

3.2.1 Ethical approval

The human participants in this research were recruited to the MRC Human Genetics Eye Malformations (HGUeye) cohort with ethical approval from the Scotland A UK Multicentre Research Ethics Committee as described in **section 2.2.1.1**.

All murine studies complied with UK Home Office regulations, with study protocols approved under the relevant Home Office project licences (60/4424, P1914806F). Approval for the mouse experiments was given by the Wellcome Trust as part of the fellowship application. In planning and managing these experiments, I was guided by the principles of the 3Rs (255) to both minimise the number and improve the welfare of mice involved in this project.

3.2.2 Human participant case selection and clinical phenotype data collection

MAB21L1 human eye malformation cases described are families or individuals referred to the MRC Human Genetics Unit eye malformations (HGUeye) cohort and from collaborating clinicians (166), with genomic DNA, phenotypic information and consent supplied by the referring unit. I sought and collated structured phenotypic information, seeking further detail and photography from the referring clinicians where needed.

I did not personally identify these variants; this was done prior to this study. Causative variants were identified using Sanger sequencing, whole exome sequencing, or candidate gene panel sequencing by: Professor David FitzPatrick and his team, in particular Kathy Williamson (Postdoctoral Scientist); the NHS South East Scotland Regional Genetics Service, in particular Morad Ansari (Clinical Scientist); the UK10K project at the Wellcome Sanger Institute (256).

3.2.3 Accession numbers

The protein reference and RefSeq accession numbers (257) for this project are as follows:

Human *MAB21L1* (chromosome 13):

GRCh38 NC_000013.11 (genomic sequence)

GenBank NM_005584.5 (transcript, 1080 nucleotides open reading frame)

GenPept NP_005575.1 (359 amino acids)

UniProt Q13394 (MB211_HUMAN)

PDB: 5EOG (human MAB21L1)

Mouse *Mab21l1* (chromosome 3):

GenBank NM_010750.3 (transcript 2793 bp, 1080 nt open reading frame)

GenPept NP_034880.1 (359 aa)

UniProt O70299 (MB211_MOUSE)

3.2.4 Mab21l1 mouse line with a targeted mutation

3.2.4.1 Generation of the CRISPR knock-in Mab21l1R51L mouse line

The CRISPR design and generation of this mouse line was performed by Joe Rainger (then Postdoctoral Scientist, now Principal Investigator), with the technical help of the local transgenic facility including Jacek Mendrychowski for the zygote microinjections. I subsequently took over the management of the colony and performed ocular phenotyping of the mice. I have summarised the generation of the line here as it is key to the project.

CRISPR-Cas9 genome editing was targeted to the arginine 51 codon of *Mab21l1* in C57Bl6/J Cr1 zygotes. Guide DNA oligonucleotides (Mab21l1_guide_C and Mab21l1_guide_D, **Appendix B.2.1** table) targeting the single coding exon of *Mab21l1* were designed and used for sgRNA synthesis (HiScribe™ T7 High Yield

RNA Synthesis Kit, New England Biolabs) (258). The repair template DNA, synthesised commercially (IDT Technologies), was designed to include the Arg51 codon change as well as additional silent nucleotide substitutions to minimise subsequent homology directed repair events at edited loci. Microinjections of zygotes were performed, injecting sgRNA, Cas9 mRNA and repair template (Tebu-bio).

3.2.4.2 Breeding and husbandry

Founders carrying the targeted variant were backcrossed onto the parental strain and tested to ensure germline transmission of the targeted variant, then used to establish a colony. The colony was principally maintained in the heterozygous (*Mab21l1*^{R51L/+}) state, backcrossing a male heterozygote with a female wild type animal from the parental C57Bl/6JCrI strain to minimise genetic drift. At least every 10 generations, a female heterozygous animal was crossed with a male wild type to refresh the sex chromosome. Heterozygous intercrosses (siblings or cousins) were set up when offspring of all 3 genotypes were needed experimentally. Occasionally heterozygous mice from these crosses were used in breeding for colony maintenance. Mice were maintained under normal husbandry conditions with a standard diet.

Mendelian ratios of the offspring were assessed statistically by comparing observed versus expected frequencies using chi square (with 2 degrees of freedom, df, for heterozygous crosses, and 1 df for heterozygous x wild type crosses).

3.2.4.3 Genotyping of the mice

Genotyping of adult mice was performed on ear clip tissue lysate, using PCR of genomic DNA and Sanger sequencing using the primers listed (**Appendix Table B.2.1**). Due to the close homology between *Mab21l1* and *Mab21l2*, Sanger sequencing of the *Mab21l2* locus was also performed on a subset of animals.

3.2.4.4 Phenotyping of the mouse line

Age-matched wild type mice from the colony were used as controls. Littermates were used preferentially, then half-sibling or cousin controls. Very occasionally, wild type animals from the parental C57BL/6JCrI strain were used. As this was a coisogenic

strain, all of these options are listed as reasonable control animals by the Jackson Laboratory (Jackson Laboratory Handbook, ISBN 978-0-578-04182-7). Founders were not used for phenotyping experiments (despite no overt developmental phenotype being observed).

3.2.4.4.1 Adult mice

Ocular examination and imaging

Ocular phenotyping was performed with the aid and training of Lisa McKie (research technician, Ian Jackson's group). Assistance with fundus imaging was from Craig Nicol (graphic designer/photographer).

Adult and control mice were examined at 2-3 months old unless otherwise stated, using: slit lamp biomicroscopy of the anterior segment, indirect ophthalmoscopy (retinal examination), topical endoscopic fundus photography and histological examination. In a smaller subset of mice, intraocular pressure (IOP) and optical coherence tomography (OCT) were performed under brief general anaesthesia; the OCT was performed by Lisa McKie.

More detailed methods for these examination procedures, which are broadly as previously published (259) are as follows:

Anterior segment examination: The cornea and iris were examined undilated using a normal slit lamp biomicroscope (with the scruffing hand resting on the chin rest). Dilating drops (1% tropicamide) were instilled 10 min prior to examination in order to examine the lens.

Fundus (retinal) examination and photography: 1% tropicamide and 2.5% phenylephrine were instilled to both eyes at least 10 min prior to examination to dilate the pupils. Fundi were examined using an indirect ophthalmoscope and 90D lens. For fundus photography, an additional topical anaesthetic eye drop was instilled (0.4% oxybuprocaine). A Nikon camera (D40X, AF 85mm, F1.8, AF Nikkor lens) was set up with the lens attached to a topical endoscope (BERCI Tele-Otoscope; HOPKINS straight forward 0°, 3mm diameter, with halogen cold light fountain). Carbomer gel 0.2% was applied as a coupling agent to the mouse cornea, carefully positioned at the

tip of the endoscope. The image was captured by a second person when the retina was in focus. Mice were kept in a darkened room whilst the dilating drops wore off.

Fundus image analysis: The optic discs were delineated semi-automatically using the Vampire Annotation Tool (260,261). The resultant two-dimensional optic disc sizes (area of the ellipse) were compared using a 2-tailed *T* test. For IOP and optic disc measurement, the means of left and right eye measurements were used for statistical analysis. Alpha was set at $p=0.05$.

Intraocular pressure measurement: IOP was measured using the Icare Tonolab tonometer, two minutes after general anaesthetic given via intraperitoneal injection of ketamine/medetomidine. Each IOP result from the Tonolab is a mean of six readings, with an alert displayed and prompt to repeat if the standard deviation is high.

Optical coherence tomography: Optical coherence tomography (OCT) using Spectralis (Heidelberg Engineering) was performed largely as described (262). Mice were anaesthetised by inhalation of 4% isoflurane, and then 2% isoflurane for maintenance. Pupils were dilated as above. Carbomer gel 0.2% was applied to prevent drying of the cornea during the scan. OCT scans were obtained, positioning the probe of the OCT scan 5 mm away from the eye and adjusting the focus until clear images could be captured. Mice were placed in a heat box and monitored until recovered from the anaesthetic.

Histology

Mice were culled and enucleated eyes were preserved in Davidson's fixative overnight at 4°C. The eyes were then dehydrated in an ethanol series, wax embedded, sectioned (7-8 μm), floated in a warm water bath onto glass slides, dried in a 60°C oven, and stained with Haematoxylin and Eosin as previously described (263). Images were captured with the Nanozoomer XR scanner (Hamamatsu, Japan) and viewed with NDP.view2 software (Hamamatsu).

3.3 Results

3.3.1 Phenotypic analysis, human: monoallelic missense variants in *MAB21L1* cause aniridia and microphthalmia

Heterozygous missense variants in *MAB21L1* affecting Arg51 were identified in five families, and affecting Phe52 in a further family, giving a total of six affected families (**Table 6**). As outlined above (**section 3.2.2**) the variants were identified by my colleagues and collaborators. In this section and in our paper (166), I phenotype the cases of this new aniridia and microphthalmia-associated gene:

3.3.1.1 Clinical case descriptions

Pedigrees of the six families and clinical images of two families are depicted in **Figure 24** below this section. The phenotypic features are described briefly here and summarised **Table 6** with the molecular variant details. For more detailed clinical descriptions for each individual or family, see **Appendix B.4**.

In summary, all six families had aniridia and/or microphthalmia. Aniridia was present in 4/6 families, microphthalmia in 4/6; and both in 3/6:

3.3.1.1.1 Aniridia

The aniridia was partial (moderate loss of iris tissue, for example corresponding to iris examples 3-5 in Grønskov's classification (62)) in 2/6, unspecified in 1/6, and profound in 1/6, as depicted in **Figure 24**. None of the families had a documented normal iris: 1/6 had a milder iris phenotype consisting of an irregular pupillary margin, and the remaining 1/6 had severe microphthalmia.

3.3.1.1.2 Microphthalmia/MAC spectrum

Microphthalmia plus or minus other MAC spectrum features were seen in 4/6. 1/6 had a family member with bilateral severe microphthalmia, no view of the internal ocular structures and no perception of light. 1/6 had a microphthalmia with a chorioretinal colobomata and microcornea (in all three family members).

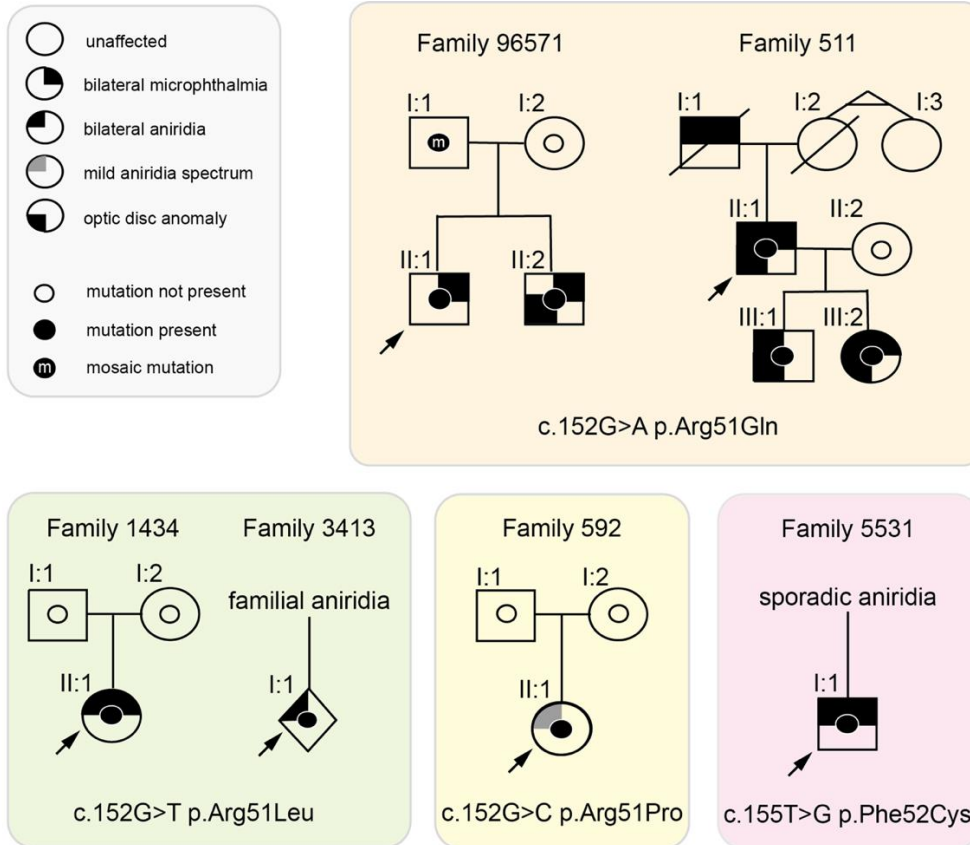
3.3.1.1.3 Other ocular features

4/6 had nystagmus, 3/6 with confirmed foveal hypoplasia; in the remaining 2/6 only limited phenotypic data was available. 3/6 families had cataract, 2/6 with lens instability or subluxation. None had aphakia. 1/6 had glaucoma. 2/6 were described as having keratopathy or an opaque cornea but, as both individuals had microphthalmia and one had phthisis, this is difficult to interpret and could be multifactorial. 2/6 had optic disc anomalies, including congenitally excavated and hypoplastic nerves. 2/6 were myopic, known to be high myopia in one of these. Whilst visual acuity data was only available for three individuals, they were all severely sight impaired, reflecting the gravity of these eye malformations.

3.3.1.1.4 Non-ocular features

There were no non-ocular phenotypic features of note. In particular, none of the affected individuals were reported to have genital anomalies.

A



B

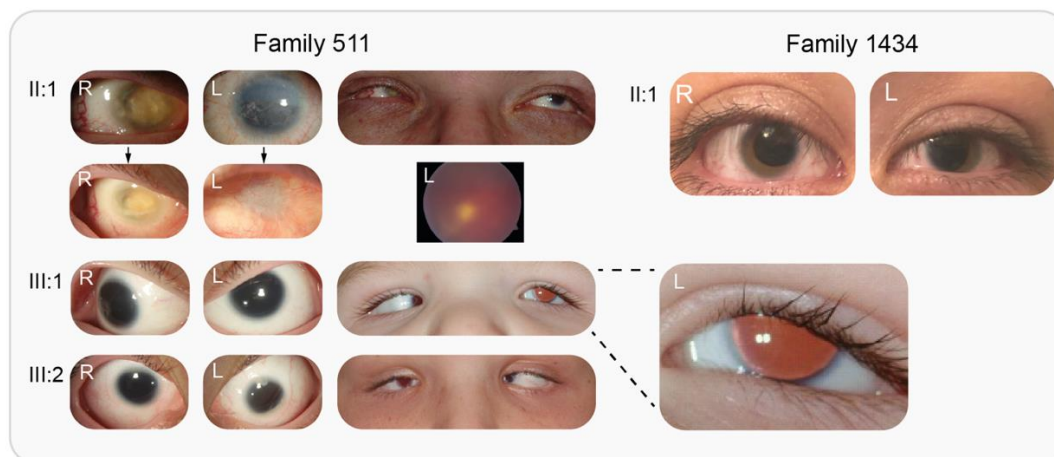


Figure 24 Pedigrees and clinical images showing MAB21L1 Arg51 and Phe52 substitution causes microphthalmia and aniridia. A. Pedigrees for the six families with MAB21L1 variants and bilateral microphthalmia and/or aniridia, ordered by variant: *c.152G>A p.(Arg51Gln)* (orange shaded box), *c.152G>T p.(Arg51Leu)* (green), *c.152G>C p.(Arg51Pro)* (yellow) and *c.155T>G p.(Arg52Cys)* (pink). B. Clinical images of individuals with MAB21L1 Arg51-related eye malformations. Family 511 all have profound aniridia, microcornea, choroidal coloboma (just visible in II:1's L fundus photo) and optic disc anomalies. The progression of disease in II:1 over one decade is shown between the upper and lower photos (indicated by arrowheads), with worsening of phthisis in the

right and pannus in the left. An enlarged retroilluminated image of III:1's L eye is shown highlighting complete aniridia. Individual 1434 II:1, showing bilateral partial aniridia and microphthalmia, worse on the L. Chromatograms are shown in **Appendix B.3**. Abbreviations: dn, *de novo*; R, right eye; L, left eye. Nucleotide and amino acid numbering are based on GenBank NM_005584.5 and GenPept NP_005575.1, respectively.

Table 6 Clinical and molecular features of individuals with *MAB21L1* heterozygous variants

Family ID	96571	511	1434	3413	592	5531
Case	II:1	II:1	II:1	I:1	II:1	I:1
Sex	male	male	female	ND	female	male
GRCh38: NC_000013.11	g.35475987C>T	g.35475987C>T	g.35475987C>A	g.35475987C>A	g.35475987C>G	g.35475984A>C
GenBank: NM_005584.5	c.152G>A	c.152G>A	c.152G>T	c.152G>T	c.152G>C	c.155T>G
GenPept: NP_005575.1	p.(Arg51Gln)	p.(Arg51Gln)	p.(Arg51Leu)	p.(Arg51Leu)	p.(Arg51Pro)	p.(Phe52Cys)
Inheritance	paternal ^a	paternal ^a	paternal ^b	ND	de novo	ND ^c
Growth						
Age at last assessment, years	7.5	2	48	11	23	11
Z scores: Birthweight	2.9	3.05				
Height	2.08	1.2				
Weight	1.31	0.76				
OFC, cm	51	49.5				
Ocular features						
Microphthalmia	BL severe	BL	BL	BL		BL
Coloboma			LE small, inferior to disc	choroidal: RE infero- temporal; LE small, temporal to disc	LE nasal choroidal	
Aniridia			BL	BL	BL	BL

Microcornea, mm			BL	BL: 6								
Keratopathy	LE opaque vascularised cornea	RE opaque cornea	BL progressive									
Glaucoma							BL with surgery LE					
Cataract	no view		LE previous lensectomy		BL with RE lenticulus, LE lens instability		BL with progressive subluxation		BL			
Nystagmus		LE	BL	BL			BL		BL			
Foveal hypoplasia		LE	insufficient view				BL		BL			
Optic disc anomaly		LE	RE no view; LE congenitally excavated	BL grey hypoplastic	BL grey hypoplastic							
Myopia							BL high		BL			limited phenotypic data
Additional details	BL no view of internal structures	RE no view internally	RE phthisis of unknown cause						limited phenotypic data			
Visual acuity	BL NPL	BL PL					BL HM					
Non-ocular features												
	hyperthyroidism (also in mother) normal MRI brain scan, echo and renal US						BL mild sensorineural hearing loss, normal MRI brain scan					

^a, the unaffected father was gonosomal mosaic (at a level of approximately 27 percent) for the variant; ^b, his deceased father had microphthalmia and aniridia; ^c, sporadic case of aniridia. Abbreviations: BL, bilateral; echo, echocardiogram; HM, hand movements; LE, left eye; MRI, magnetic resonance imaging; ND, not determined; NPL, no perception of light; OFC, occipitofrontal circumference; PL, perception of light; RE, right eye; US, ultrasound.

3.3.2 Protein structural context of MAB21L1 residue substitutions

To assess the protein structural characteristics of MAB21L1 variants, we took together the monoallelic (heterozygous) missense MAB21L1 variants described above and published (166,167), the published biallelic missense variant (250), and published missense variants in the paralog MAB21L2 (252,253).

Firstly, this shows distinct clustering of monoallelic variants around Arg51. Arg51 itself is recurrently substituted in both genes. In MAB21L1, the adjacent Phe52 is substituted, whilst in MAB21L2 the other reported residue is Glu49 (**Figure 25**).

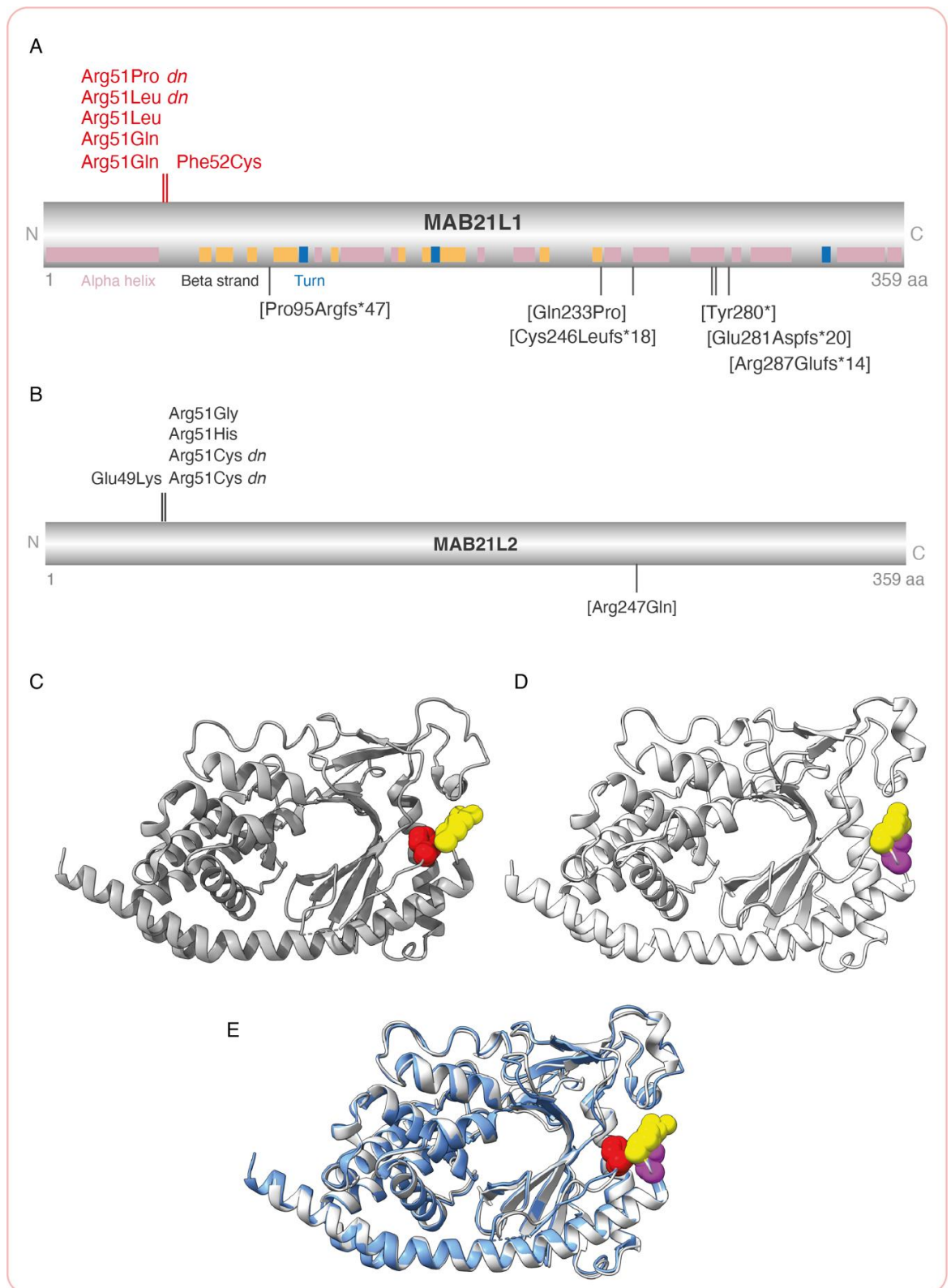


Figure 25 Structure of MAB21L1 and MAB21L2 and clustering of substituted residues around Arg51. (A, B) Protein bar schematics showing (A) MAB21L1, and (B) MAB21L2, and the linear positions of causative variants. The first and last amino acid (aa) residues are numbered and, for

MAB21L1, the position of secondary structural features such as alpha helices (pink) and beta strands (yellow) are indicated. Heterozygous causative variants (autosomal dominant inheritance) are shown above the protein bars, with the variants in this study shown in red, and published homozygous variants, both missense and likely gene disrupting, are shown below in brackets (autosomal recessive). (C) Cartoon of the crystal structure of MAB21L1 (PDB: 5EOG) with the Arg51 residue in yellow and Phe52 residue in red. (D) AlphaFold prediction of the MAB21L2 structure, with the Glu49 residue in magenta and the Arg51 residue in yellow. (E) Superimposition of (C) and (D) structures highlighting the clustering of the substituted residues in an unstructured loop in both paralogs. With thanks to Dr Rolando Hernandez Trapero (Postdoctoral Research Fellow, Marsh group) for his help with (C-E).

Collaboration with Joe Marsh's group used FoldX (264) to model the effects of MAB21L1 and MAB21L2 substitution on protein thermodynamic stability (166). This showed that all the missense variants examined are predicted to be destabilising to protein structure, as indicated by values of mean $\Delta\Delta G$ of $>+1.5$ kcal/mol (265), except the biallelic MAB21L2 Arg247Gln substitution. (Of note, previous experimental work has shown that this variant is also destabilising (245).)

Table 7 FoldX values of MAB21L1 and MAB21L2 substitutions

Variant	Gene	Mean $\Delta\Delta G$ (kcal/mol)	Standard deviation
Q233P	MAB21L1	7.65775	0.0604902
R51Q		5.51197	0.387943
F52C		5.34457	0.0648261
R51G	MAB21L2	5.13351	0.176716
R51C		4.59367	0.380072
R51H		4.52398	0.413263
R51L	MAB21L1	3.78305	0.248232
R51P		3.57492	0.309426
E49K	MAB21L2	1.99549	0.51072
R247Q		-0.647993	0.171486
<i>Mean $\Delta\Delta G$ of $>+1.5$ kcal/mol is predicted to be destabilising to protein structure</i>			

These results highlight the structural relevance of this region. Of the heterozygous MAB21L1 substitutions, Arg51Pro had both the mildest destabilising effect and also the mildest phenotype.

The monoallelic MAB21L1 variants were also predicted by multiple variant effect predictors (VEPs) to be likely pathogenic (**Supplementary Table 1 in Appendix B.5**). (The term likely pathogenic as used by VEPs in this context should not be confused for the specific clinical genetics meaning of this term, such as the ACMG pathogenicity classification (266,267).)

3.3.3 Description of a CRISPR mouse *Mab21l1* line

The murine *Mab21l1* gene is a single exon gene on chromosome 3 which, as with the human ortholog on chromosome 13, is nested within an intron of *Nbea*. Using CRISPR-Cas9 microinjection of zygotes, the *Mab21l1* arginine 51 residue was targeted to introduce one of the substitutions seen in the human families described above, p.Arg51Leu (**Figure 26**). The repair template in **Figure 26** was used to knock in the substitution.

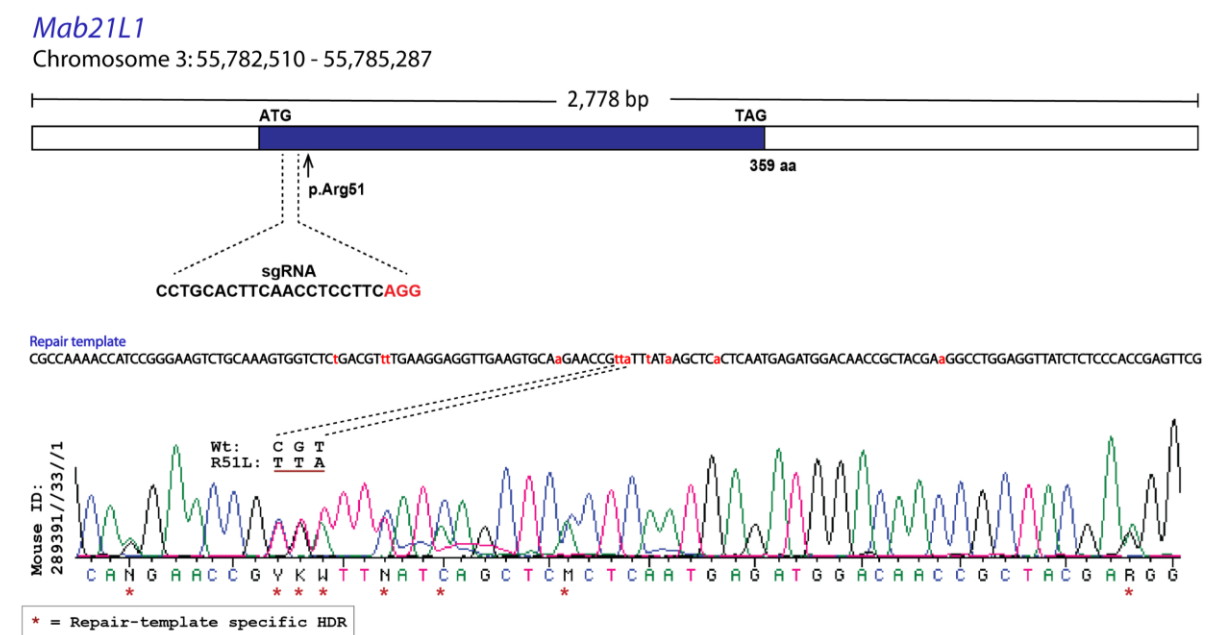


Figure 26 Schematic showing CRISPR-Cas9 sgRNA guide (guide C in the oligonucleotide table in **Appendix B.2.1**) and repair template sequences used in the generation of the *Mab21l1* mouse line, and their locations relative to the region encoding Arg51. *Mab21l1* is a single exon gene on mouse chromosome 3 (GRCm38:CM000996.2 coordinates in figure), nested within the neurobeachin gene. The *Mab21l1* p.Arg51Leu variant was introduced, along with the silent substitutions in the flanking regions (red asterisks) specific to the repair template, as shown. The Sanger sequencing chromatogram shows a heterozygous *Mab21l1*^{R51L/+} animal from the N1 generation (offspring of the backcross of a founder mouse onto the parental strain). Figure adapted from Joe Rainger (166).

From the littered-down animals, a founder female animal was identified to be carrying the correctly gene-edited variant (**Figure 26** lower panel) and was backcrossed to a wild type male from the parental strain C57Bl/6JCrI. Sanger sequencing showed 3 out of 8 of this litter were heterozygous, *Mab21l1*^{R51L/+}, confirming germline transmission

of the targeted variant and the silent substitutions flanking it. Heterozygous mice were crossed to produce viable and fertile *Mab21l1*^{R51L/R51L} homozygotes. As the resultant strain differed from the established inbred C57Bl/6JCr1 strain at one locus, we considered this a coisogenic strain (268).

I calculated the mean litter size for the colony to be 6.7 mice per litter, with no significant difference in litter size between heterozygous crosses and heterozygous x wild type crosses (**Figure 27**). This compares to a published mean litter size of 4.9 mice/litter for C57Bl/6J overall in the JAX breeding manual.

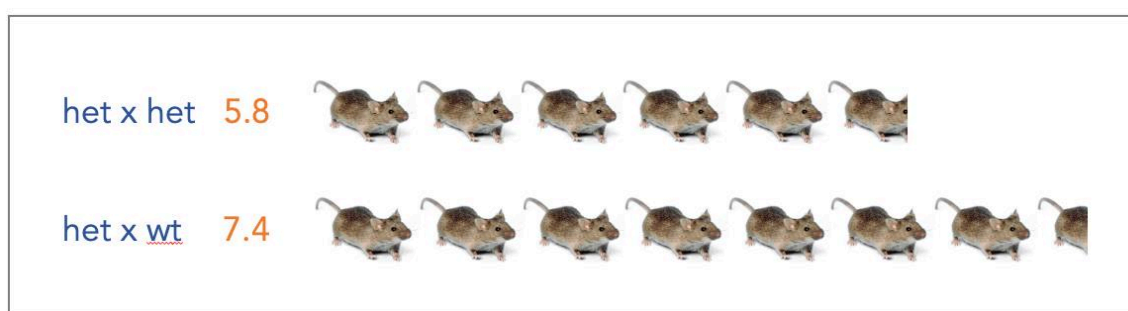


Figure 27 The average litter size for a cross of heterozygous (het) mice (usually cousin matings) was 5.8 mice/litter, which was not significantly different to heterozygous by wild type (WT) crosses, 7.4 mice/litter ($p=0.015$, unpaired 2-tailed *T* test, comparing the size of $n=14$ litters of each type). The mean litter size for the whole colony, including all types of mating, was 6.7 mice/litter.

There was no evidence of perinatal lethality and, at weaning, the ratios of the offspring inheriting the edited allele were Mendelian (**Table 8** below).

The mice appeared systemically healthy. Weights recorded in female mice on the day they had plugged were within the normal range for the mouse strain ($24.5\text{g} \pm 2.7$ SD at a mean of 22.7 weeks of age, $n=7$). The only visible external abnormality was the ocular phenotype of the homozygote mice, who had small eyes. Both this and, first, the more subtle ocular phenotype of the heterozygote animals, are described in the following sections.

Table 8 Mendelian ratios. Comparison of the observed versus expected ratios of genotypes from intercrosses of the *Mab21l1*^{R51L} line mice (n=14 litters of each type), establishing that the observed ratios were consistent with Mendelian genetics.

	<i>Mab21l1</i> ^{+/+}	<i>Mab21l1</i> ^{R51L/+}	<i>Mab21l1</i> ^{R51L/R51L}
Het x het cross			
Observed	16	44	14
Expected (1:2:1)	18.50	37	18.50
<i>p</i> =0.252, chi square with 2 degrees of freedom, pooled numbers from n=14 litters			
Het x wt cross			
Observed	48	36	
Expected (1:1)	42	42	
<i>p</i> =0.190, chi square with 1 degree of freedom, pooled numbers from n=14 litters			

3.3.4 Phenotypic analysis, mouse: *Mab21l1*^{R51L/+} heterozygous mice have bilateral anomalous optic discs

Mab21l1^{R51L/+} heterozygote mice all had anomalous, excavated optic discs: these were present on examination in 13/13 animals and confirmed as bilateral in 9/13 (**Figure 28**, remainder of images in **Appendix B.6.1**).

The phenotype was striking enough that a genotyping sample error resulting in 2 heterozygous mice being initially mislabelled as wild type was detected through recognition of the specific optic disc phenotype.

Quantitative analysis of fundal images confirmed that the optic discs were larger in heterozygous mice than in wild type littermate/cousin controls (*p*=0.00004, n=8 heterozygous and n=4 wild type, **Figure 28**). The excavated optic disc anomaly is also evident on histological H&E stained sections (**Figure 29**). Intraocular pressure tested

in a smaller subset of *Mab21l1*^{R51L/+} mice was within the normal range, and not different to wild type or homozygous littermates (n=5/2/3 heterozygous/WT/homozygous, respectively, p=0.5734, Kruskal-Wallis test) consistent with the optic nerve phenotype being a developmental rather than glaucomatous phenomenon; the other factor in support of this was the absence of buphthalmos. The mice had no other apparent abnormalities, and in particular slit lamp examination of the iris and anterior segment was normal.

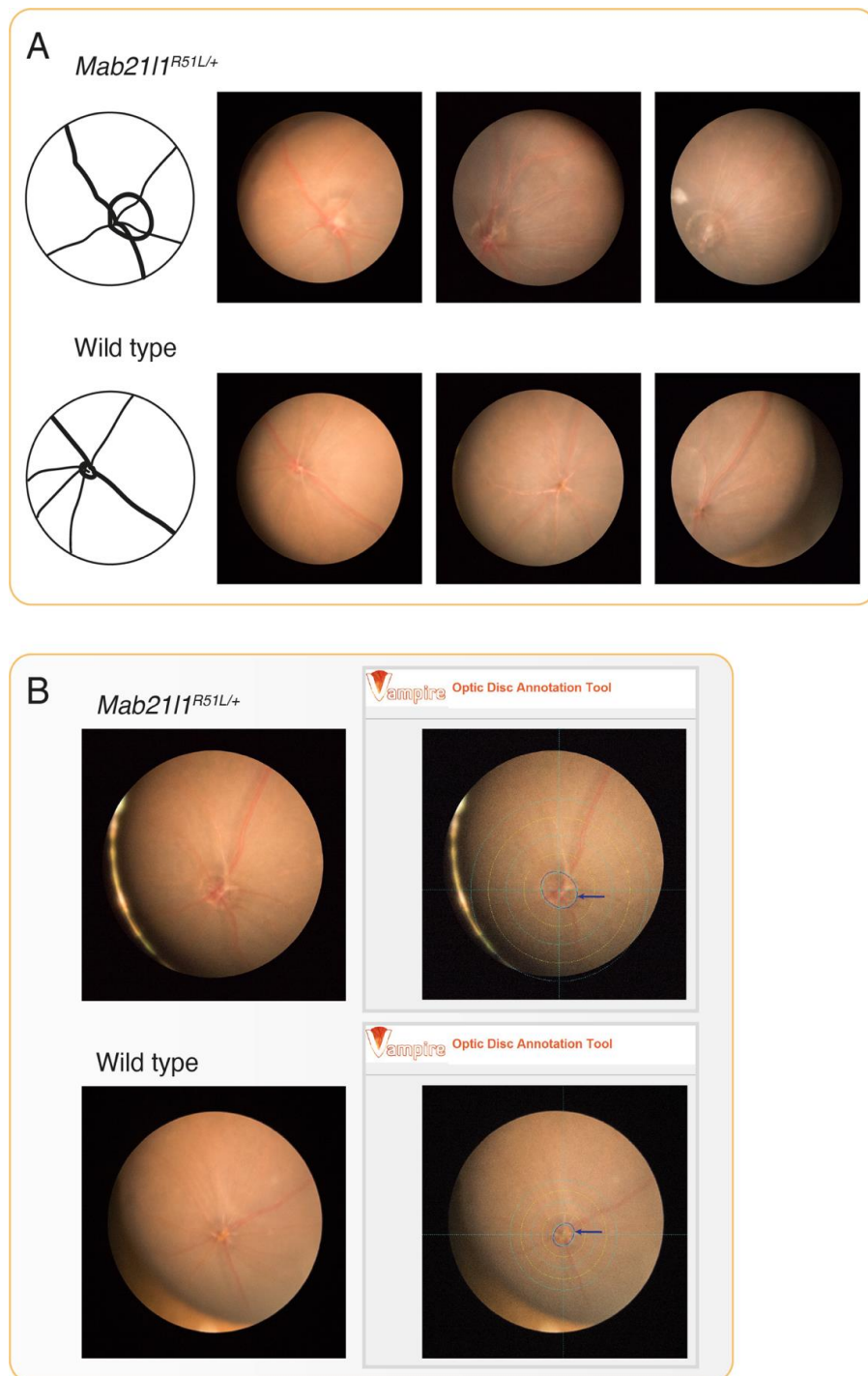


Figure 28 (A) Retinal photographs showing the anomalous optic disc phenotype of *Mab21l1*^{R51L/+} mice, upper panel (images from one eye of three different mice, representative of the phenotype which was present in all 13 heterozygous mice examined and no controls; all available images are shown in **Appendix B.6.1**), versus age-matched wild type (WT) littermate/cousin controls, lower panel. The cartoon to the left outlines the larger optic disc. (B) Example of annotation using the Vampire Annotation Tool to measure optic disc size. Retinal images are shown with semi-automated disc annotation (navy blue ellipse, arrow). Upper panel, *Mab21l1*^{R51L} heterozygote; lower panel, WT littermate control

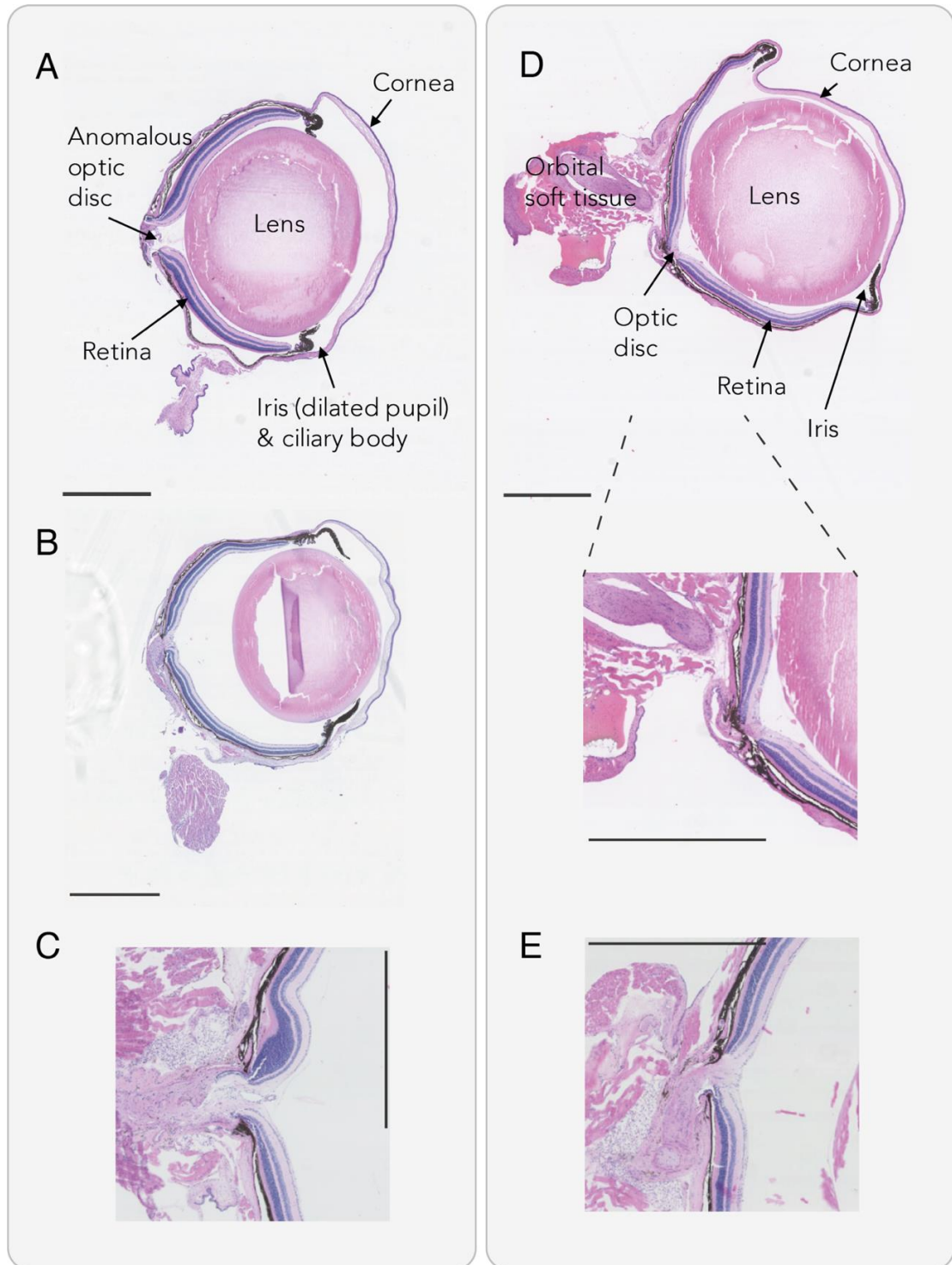


Figure 29 H&E stained wax sections of A-C, one eye from different adult *Mab21l1*^{R51L/+} mice; D-E, one eye of different wild type littermate or cousin controls. The anomalous excavated optic nerve appearance is visible in the heterozygous mice. The iris had been pharmacologically dilated prior to culling in order to permit fundus photography. All scale bars 1mm.

3.3.5 Phenotypic analysis, mouse: *Mab21l1*^{R51L/R51L} homozygous mice have a severe complex panocular eye malformation

Homozygous *Mab21l1*^{R51L} mice examined (n=7 examination and fundus photography, n= 3 for histology) showed microphthalmic eyes with a severe, bilateral, panocular eye malformation (**Figure 30**). These were microphthalmic eyes with disorganised anterior and posterior segments, and histological sectioning showed abnormalities of the cornea, iris, ciliary body, lens, retina and optic nerve (**Figure 31**). There was hyperplasia of pigmented uveal tissue, further obscuring any possibility of a fundal view on examination. The most severely affected eyes had only a rudimentary lens and retina, and optic nerve aplasia.



A External photographs



B Macroscopic images of enucleated eyes

Figure 30 Photographs of homozygote vs heterozygous *Mab21l1* mice and enucleated eyes. Unlike the homozygous animals, which were clearly recognisable from the external appearance, heterozygous and wild type mice appeared the same, with differences only seen on fundus examination and histology, as described above.

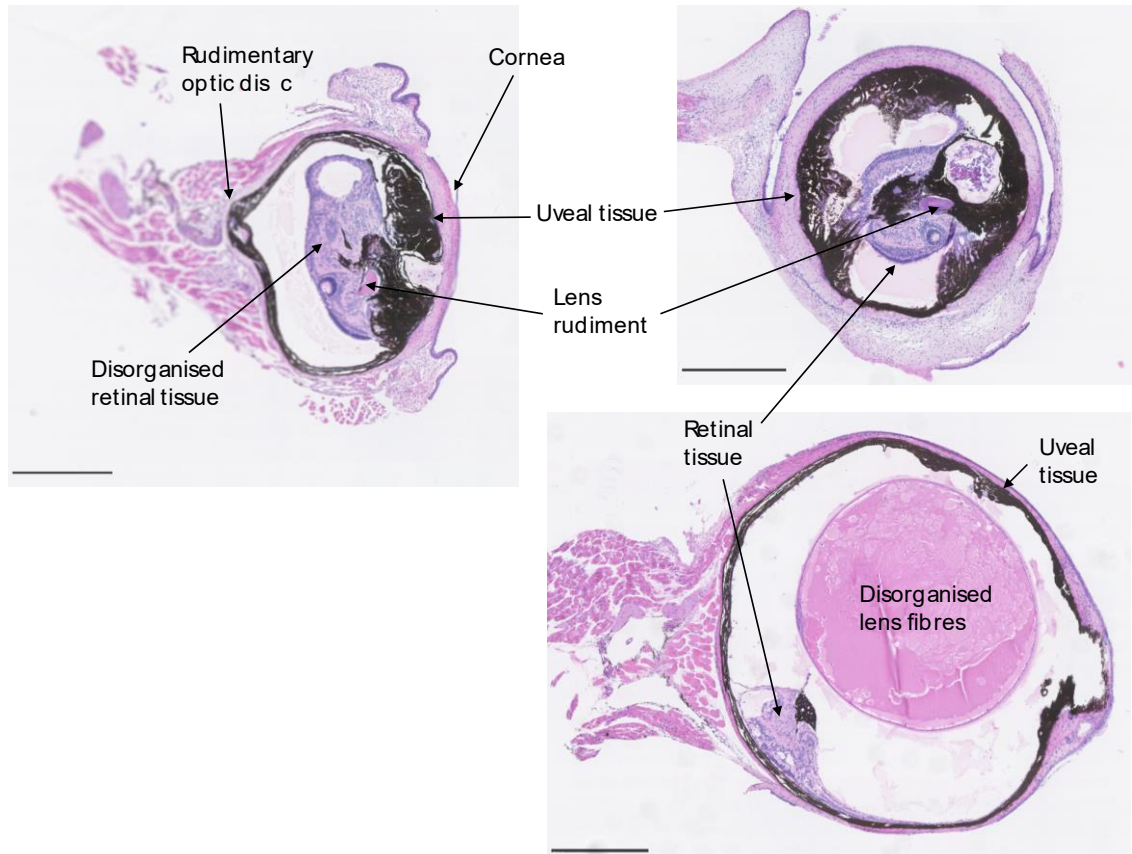


Figure 31 H&E-stained adult eye sections from two *Mab21l1*^{R51L/R51L} mice showing a severe eye malformation with disorganised anterior and posterior segments, and optic nerve aplasia. This phenotype was present in all 7 animals examined, the images here illustrating the range of size and disorganisation of the eyes. Scale bars 1mm.

A summary of the mouse phenotypes from the different genotypes is shown in **Figure 32**.

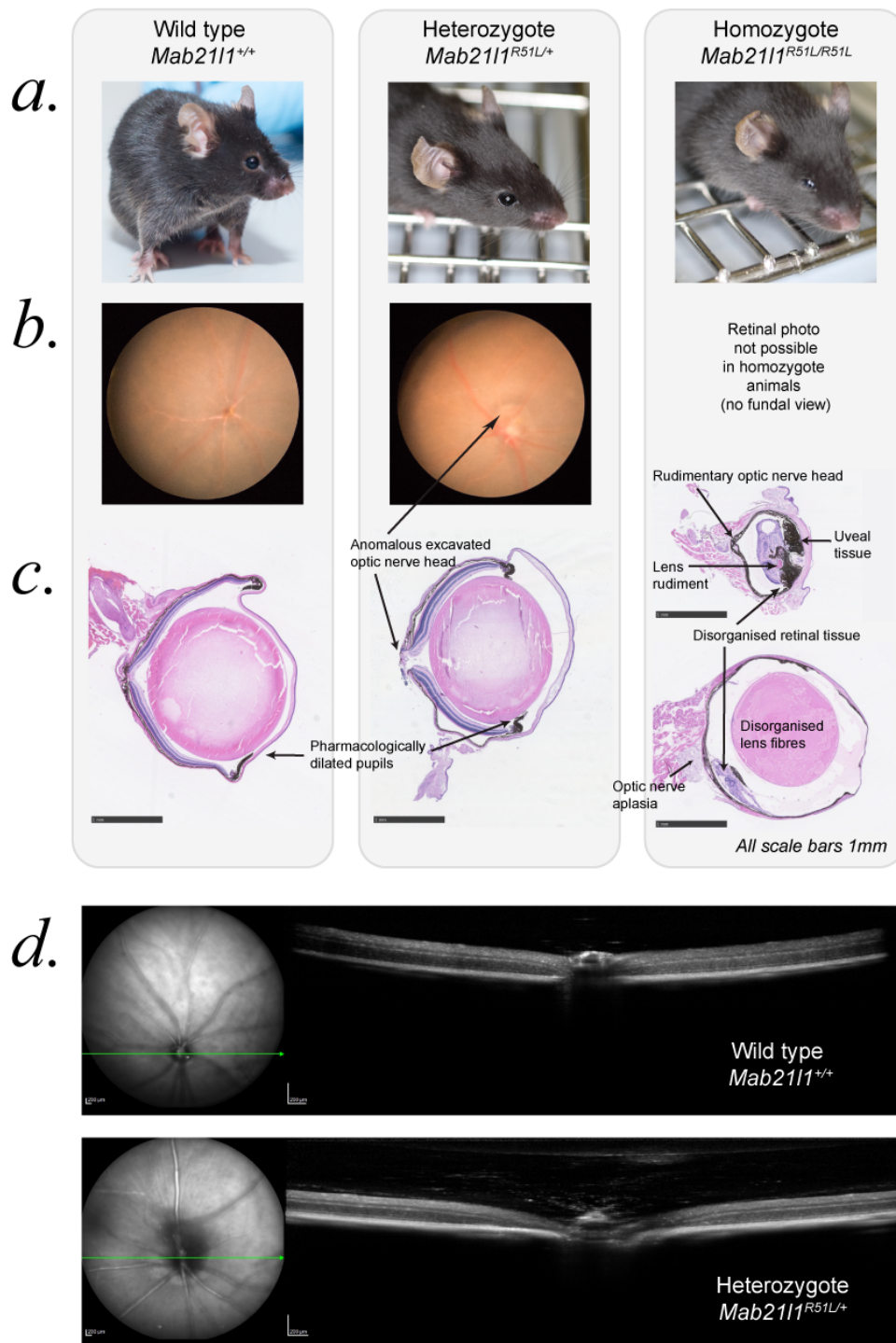


Figure 32 Summary of *Mab21l1*^{R51L} ocular phenotypes. Left; wild type (WT), middle; *Mab21l1*^{R51L/+}, right; *Mab21l1*^{R51L/R51L}. Mice were examined at 2-3 months of age. (A) Photographs, showing the normal external appearance of heterozygous and severe microphthalmia in homozygous mice. (B) Representative retinal photographs showing the anomalous, excavated optic disc phenotype (arrow) present in all the 13 heterozygous mice examined (further images in **Appendix B.6.1**). No retinal view was possible in the microphthalmic eyes of homozygotes, often with uveal tissue obscuring the cornea. (C) H&E-stained wax sections of eyes, with the abnormal features for each genotype labelled. The excavated optic nerve anomaly of the heterozygous mice is seen. These

mice had recently-administered dilating drops for fundal examination, but the iris appeared normal on slit lamp examination. Homozygotes had a severe panocular eye malformation including microphthalmia, severely disorganised retina and uveal tissue, along with hypoplasia or aplasia of both the optic nerve and lens. Eyes from two different age-matched homozygotes are shown to illustrate the spectrum of microphthalmia. Scale bars = 1 mm. (D) Optical coherence tomography (OCT) of WT and heterozygous mice, showing the enlarged optic nerve phenotype.

3.4 Discussion

As we have seen in the preceding chapters, classical aniridia is both a highly distinctive and strongly *PAX6*-related phenotype, with missense variation resulting in some cases in severe MAC spectrum phenotypes. The six human families described here with heterozygous *MAB21L1* variants have a combination of microphthalmia and aniridia, the majority with features typically associated with classical aniridia (nystagmus, foveal hypoplasia, cataract and lens instability, or anomalous optic discs).

Phenotypic evidence

In many respects, the *MAB21L1* cases described here phenocopy a subset of severe aniridia/MAC associated with recurrent *PAX6* missense variants described in Chapter 2. Notably, however, none of *MAB21L1* families described had intellectual disability or developmental delay, which does feature in several severe *PAX6* missense cases.

Whilst the published biallelic loss of function *MAB21L1* human phenotype is corneal dystrophy with cerebellar, craniofacial and genital anomalies (250,251), the monoallelic phenotype described here is an isolated and more severe eye malformation, significantly compromising vision, with no remarkable systemic features. The human genetic evidence would be consistent with a gain of function effect in the eye of these missense variants, and the highly localised clustering of the heterozygous variants is also consistent with this.

During the course of this work, Seese *et al* reported the identification of the Arg51Leu variant in a family with microphthalmia and aniridia (167). In addition to our publication of the six families here (166), there is one unpublished Arg51Gln family

within the HGU cohort. A further series of five unrelated Chinese families has been published by Wang *et al* (269) with Arg51Leu, Arg51Gln and Phe52Cys variants. They describe the same pattern of eye malformation but also identified blepharophimosis in all cases. Whilst we did not collect this data systematically, and often microphthalmia is associated with developmental abnormalities of the orbit due to socket growth, one of the families here (FID 511) could be described as having a phenotype similar to blepharophimosis. Equally, the occurrence or penetrance of this feature as part of the phenotype may differ between populations. Even more recently, Meng *et al* report on a further *MAB21L1* pedigree with Phe52Leu causing microphthalmia and lens subluxation (270), and Zucco *et al* report two aniridia microphthalmia cases, Arg51Gln and Phe52Cys, but the latter being the same case as in our cohort (159). This brings the total reported to date to 14 unrelated families with dominantly-inherited, monoallelic *MAB21L1* eye malformations.

Examination of the CRISPR-targeted *Mab21l1*^{R51L} mouse model has revealed that monoallelic/heterozygous Arg51Leu substitution causes an optic nerve malformation, which is overall a less severe manifestation than the human counterpart. One of the human individuals described here with Arg51Gln does have a similar optic nerve malformation, as does the unpublished but well-phenotyped unrelated family (also Arg51Gln) in which this also features. Biallelic/homozygous mutant mice have a severe panocular malformation, with microphthalmia, disorganisation of anterior and posterior segments and optic nerve aplasia. In contrast to the human phenotypic evidence, where the “null” phenotype is quite distinct to the missense, homozygote Arg51Leu mice have a similar phenotype to the published null (248) animals, which also have small eyes with a severe panocular malformation. The *Mab21l1*^{R51L/+} heterozygote phenotype is worse than the (apparently normal) heterozygous knock-out (248), though it is not clear to what extent the optic discs of adult animals were studied as retinal photographs or examination results were not explicitly described. The human and mouse phenotypic evidence is summarised in **Table 9** and taken together suggests that there may be important differences in dosage sensitivity and/or disease mechanism between humans and mice for this gene.

Table 9 Phenotype-genotype correlations according to variant category in human and mouse *MAB21L1/Mab21l1* cases

<i>MAB21L1/Mab21l1</i> variant type		Human phenotype	Mouse phenotype
Missense Arg51 variants	<i>Monoallelic</i>	Aniridia, microphthalmia +/-blepharophimosis (159,166,167,269,270)	Excavated optic disc anomaly (166)
	<i>Biallelic</i>	None reported	Severe microphthalmia and panocular malformation (166)
Likely gene disruptive/null variants	<i>Monoallelic</i>	None reported	Apparently normal (see note in text) (271)
	<i>Biallelic</i>	Cerebellar hypoplasia +/- microcephaly, intellectual disability, corneal dystrophy, craniofacial and genital anomalies (250,251)	Severe microphthalmia, impaired notochord/neural tube differentiation, preputial gland atrophy (248)

Structural considerations and relevance of the Arg51 region

The clustering of *MAB21L1* monoallelic missense variants around Arg51 is striking. Furthermore, monoallelic variants in *MAB21L2* also affect this region, namely Arg51 and Glu49. The MAB21L1 and MAB21L2 subunits have 94% amino acid sequence similarity, with a RMSD (root mean square deviation) of 1.08 Å, indicating that the structures are very similar. The clustering of the residues in both paralogs can be seen in **Figure 33**. These are residues in an unstructured loop that are involved in fold stabilisation within the protein (245), which is in keeping with the FoldX modelling reported above predicting all the monoallelic variants in both genes being destabilising to protein structure.

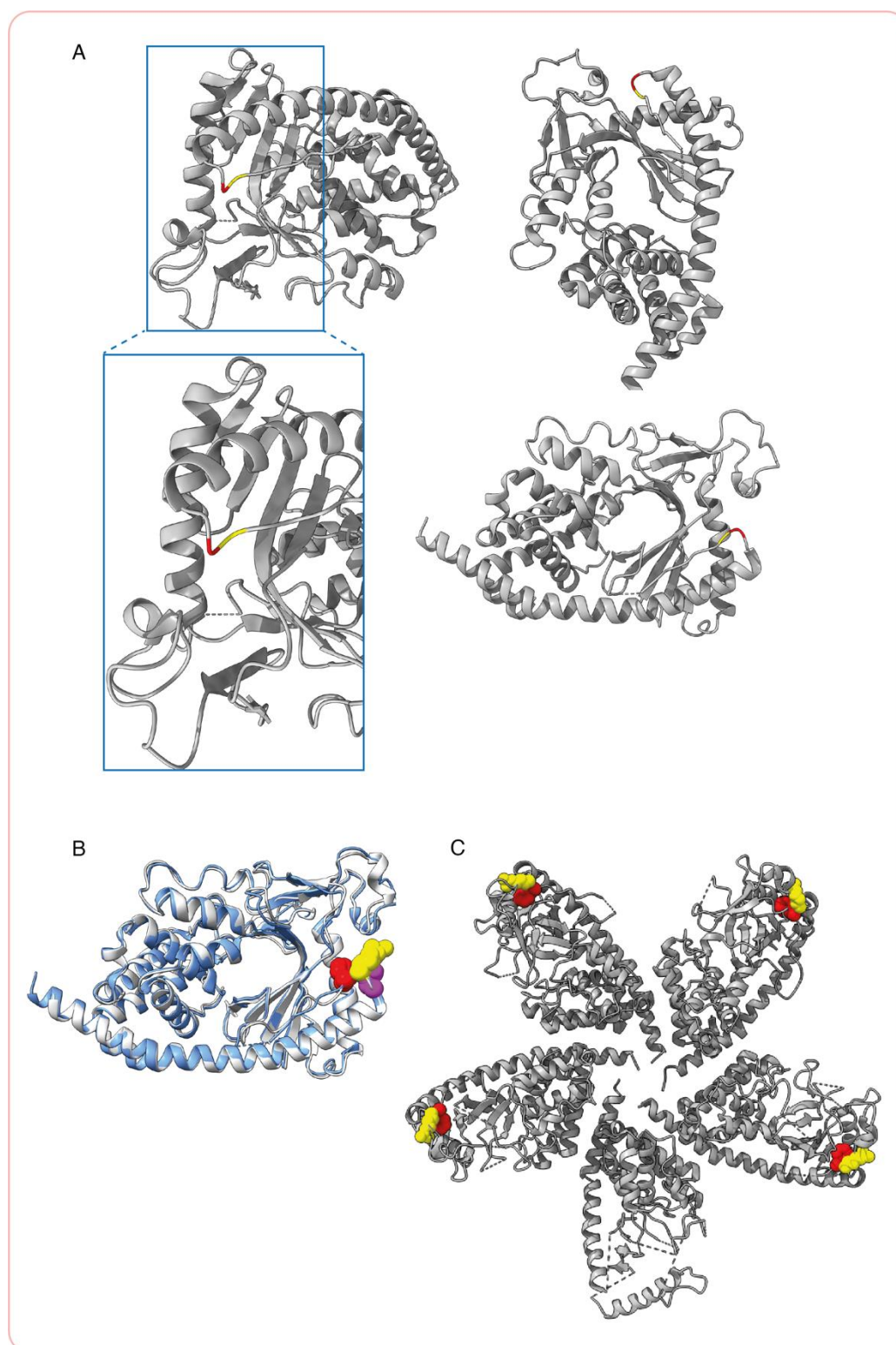


Figure 33 Structure of MAB21L1 highlighting the Arg51 region. (A) Cartoon of the crystal structure of the MAB21L1 subunit (PDB: 5EOG) with the Arg51 residue in yellow and Phe52 residue in red. The structure is shown in several different orientations. A higher magnification image is shown (blue box) of the area surrounding the unstructured loop occupied by the residues. In the rest of the figure, the coloured residues are shown as spheres for emphasis; here

they are not, so as not to obscure the loop. (B) Superimposition of the crystal structure of MAB21L1 and AlphaFold prediction of MAB21L2, with Glu49 in magenta, Arg51 in yellow and Phe52 in red. This is the same image as **Figure 25E** (C) Crystal structure of MAB21L1 as a pentamer (PDB: 5EOG). Chain D does not contain amino acids 42-64 so the Arg51 and Phe52 are not represented there. With thanks to Dr Rolando Hernandez Trapero (Postdoctoral Research Fellow, Marsh group) for his help with this figure.

Whilst the crystal structure of human MAB21L1 has been determined (245), its molecular function, activating ligand and substrate remain elusive. Some clues have been gained by Hemant Bengani and other colleagues in the FitzPatrick group. These are described in more detail in our paper (166), **Appendix B.1**. Analysis of protein-protein interactions from IP-MS (using GFP-tagged peptides from HEK293 cells) shows three apparent mutant-specific interactors: GALNT2, LRRFIP1 and MSI2/Musashi-2. The latter is perhaps of most interest; it is an RNA-binding protein that binds 3'UTR and regulates gene products with roles in development (272,273), including photoreceptor development (274), and SOX2 interaction (275). This is a potential candidate for a gain of function interaction, though this could not be confirmed by reciprocal IP-MS. We hypothesise that RNA-bound Musashi-2 is the *in vivo* activator of MAB21L1, and a competitive antagonist effect of Musashi2 binding in the mutant protein is the gain-of-function interaction. However, further work is required to investigate this hypothesis and confirm wild type and mutant interactions of MAB21L1 (and MAB21L2) in ocular tissues, and their relationship to PAX6 interactors. The phenotypic differences between the two structurally-similar genes could be explained by the relative spatiotemporal expression patterns during development.

Chapter 4 Whole genome sequencing study of “*PAX6*-negative” aniridia

4.1 Introduction

Please note: the work presented in this chapter has been published in 2023 (276) and this paper, on which I am first author, is available open access and included in the thesis as Appendix C.1. The large parts of the published paper which overlap with sections below were written by me, and some have been expanded upon to provide further detail for the thesis.

A proportion of classical aniridia remains genetically unexplained. The MRC Human Genetics Unit (HGU) Eye Malformations cohort contains approximately 443 individuals with aniridia from 347 families. Of these, 10-15% remain unexplained, this cohort being by nature enriched for unsolved cases (148).

In order to investigate unexplained cases, I assembled a classical aniridia cohort for whole genome sequencing (WGS).

As well as its potential for the discovery of new genetic mechanisms, WGS is increasingly widespread as a clinical diagnostic and not just research tool (277). Aniridia is an attractive candidate for a WGS study, due to its specific congenital phenotype, strong but not exclusive disease-gene link, and the granular knowledge of *PAX6* available following 30 years of diagnostic analysis of this locus.

4.2 Methods

4.2.1 Clinical research participants and phenotypes

This project used clinical information and biological samples from individuals referred to the MRC HGU Eye Malformation Study (HGUeye). All affected individuals had classical aniridia, or where limited clinical information was available, ‘aniridia’. I excluded those with known significant additional ocular phenotypes, such as moderate or severe microphthalmia, coloboma or severe congenital corneal opacification. Pseudonymised family IDs and personal (research participant) ID and details of their relevant clinical features are provided in the Results section in **Table 11**.

4.2.2 Ethical approval

This research participant cohort was collected and maintained using consent processes and experimental protocols approved by the Scotland A UK Multicentre Research Ethics Committee, references 06/MRE00/76 and 16/SS/0201, as described in **2.2.1.1**.

4.2.3 Preparation of genomic DNA & quality control (QC)

I assessed the quality and concentration of patient genomic DNA (gDNA) samples by agarose gel electrophoresis, Nanodrop 1000 spectrophotometry (Thermo Fisher Scientific, Inchinnan, UK), +/- Qubit 3 fluorometer HS assay (Invitrogen, Thermo Fisher Scientific). In the case of one family trio, I extracted DNA for the parents from patient-derived lymphoblastoid cell lines (LCL) as no gDNA samples of sufficient quality were available.

All probands had had prior Sanger sequencing of the coding regions of *PAX6* and *MAB21L1*, performed by a combination of referring centres, previous HGU staff, Kathy Williamson and myself.

4.2.4 Droplet digital PCR

I performed droplet digital PCR with four amplicons spanning the wider *PAX6* locus as a pilot screen in 13 individuals prior to finalising the WGS cohort, to identify copy number variants (CNVs) which had not been adequately assessed prior to this study.

Primers (Universal Probe Library assay design centre, Roche) and UPL probe pairs were targeted to three intragenic *PAX6* regions (exon 2, intron 7, 3'UTR) and the critical SIMO enhancer in the downstream regulatory region (primer table in **Appendix C.2.4**). PCR reaction mixtures were prepared (20 µl reaction containing 2 µl gDNA at 20 ng/µl, 10 µl of 2x ddPCR Supermix (Bio-Rad, Watford, UK), 2 µl of 2.5 µM UPL probe, 2 µl each forward and reverse primer (10 µM), 1 µl 20x RNaseP and 1 µl dH₂O). The QX200 Droplet Digital PCR system (Bio-Rad) was used following manufacturer's protocols (droplet generation, PCR, and plate reader). Droplets were generated in the QX200 Droplet Generator; PCR was then performed under the following conditions: 95° C 10 min, (94° C 30 sec, 55° C 1 min) x40, 95° C 10 min. The plate was then analysed by the droplet reader and CNV count was analysed using Bio-Rad QuantaSoft software.

4.2.5 Whole genome sequencing (WGS)

WGS was performed by BGI (New Territories, Hong Kong) for 9 samples, and Edinburgh Genomics (Edinburgh, UK) for the remaining 42 samples.

In brief, gDNA was sheared using a Covaris ultrasonicator, and the fragments were A-tailed, size selected and adaptor ligated prior to PCR amplification. Libraries were clustered onto a flow cell for sequencing using HiSeqX (Illumina). Further detail regarding the WGS and library preparation are given in **Appendix C.2**.

4.2.6 WGS mapping, alignment, QC and SNV/INDEL variant calling

WGS samples were processed with Bcbio v0.9.7, which uses bwa v0.7.13 (278) to align reads to the human reference genome assembly GRCh38, samblaster v0.1.22 (279) to mark duplicates, and GATK v3.4.0 (280) to realign indels and recalibrate base quality scores. Families and singletons were genotyped following GATK best practices (281) with version 4.0.2.1 of the toolkit. HaplotypeCaller was used to generate GVCFs, which were imported into a database via GenomicsDBImport and genotyped with GenotypeGVCFs. Variant quality score recalibration was carried out with GATK VariantRecalibrator and ApplyVQSR separately for SNVs and INDELs. Low quality (GQ < 20) genotypes were filtered. I acknowledge the enormous help of

Alison Meynert whose core bioinformatics team performed this pipeline processing and supplied the methods in this paragraph.

4.2.7 Structural variant (SV) calling

4.2.7.1 IGV visualization

I inspected the aligned WGS data to detect SV at known aniridia loci, using IGV (Integrative Genome Viewer, Broad Institute, CA, USA) (282), via visualisation of breakpoints and coverage. The gene and regulatory regions of *PAX6* were examined in all cases. Where *PAX6* was negative, *FOXC1* and *PITX2* were included with their regulatory regions, and *MAB21L1*; in some cases, this was expanded to additional loci (*FOXE3*, *RARB*, *ADAMTSL1*, *CYP1B1*); for a list of coordinates see **Table 10**. In IGV, reads were coloured both by insert size (to detect breakpoints of deletions/insertions) and by pair orientation (to detect breakpoints of chromosomal rearrangements such as inversions).

All genomic coordinates are GRCh38 (hg38).

Chromosome	Genomic coordinates (GRCh38)		Gene / Region	Notes
	Start	End		
chr9	18472082	18912949	ADAMTSL1	
chr9	93847400	93977000	BARX1 and regions translocation breakpoint/del	Not an aniridia gene. Included as breakpoint of the t(1,9) translocation in case RPID 356
chr7	129223015	129432211	AHCYL2	Not an aniridia gene. Included as <i>de novo</i> variant in this gene identified in RPID 2469.
chr2	38065603	38078181	CYP1B1	
chr22	45500839	45603134	FBLN1	
chr6	1395001	1615000	FOXC1 + 5' gene desert	

chr1	47414072	47420052	FOXE3
chr3	4491348	4849840	ITPR1
chr3	25426263	25599931	RARB
chr13	35471789	35478695	MAB21L1
chr11	31311354	32111353	PAX6 locus + 5'/3' regulatory regions
chr11	35660805	35811380	TRIM44
chr4	110615801	111715800	PITX2 + 5' gene desert

Table 10 Genomic regions examined on IGV for direct visualisation of SV breakpoints

4.2.7.2 Bioinformatic SV calling

Copy number variants were called for each family with Canvas v1.38 (283) using the “SmallPedigree-WGS” workflow (284). Structural variants were called with Manta v1.3.2 (285,286). CNV/SV overlapping all genes in the EyeG2P dataset (described in Variant Filtering, below) were examined. This was performed by Dr David Parry (Postdoctoral Researcher) and Dr Jose Campos-Parada (Bioinformatician). I examined the results in the loci of interest and compared them to direct visual inspection on IGV.

4.2.8 *De novo* analysis

For parent-child trios, short variants arising *de novo* in the child were identified by David Parry using VASE (287), which he developed, with the following criteria: read depth ≥ 10 in parents and child, genotype quality score ≥ 30 in the child and ≥ 20 in both parents, variant allele frequency ≥ 0.3 in the child and < 0.05 in both parents, maximum of one variant allele called at site and variant allele absent from gnomAD v3.0. Potential *de novo* variants were subsequently filtered to exclude low complexity, telomeric and centromeric regions and to select only variants within coding regions, splice regions (exonic positions within 3 bp of an intron/exon junction or intronic positions within 8 bp of an intron/exon junction) or intronic variants with a SpliceAI (288) delta score ≥ 0.5 . The variant output was analysed by myself.

4.2.9 Variant filtering

From all the variants identified in an individual, we selected only those that are rare, predicted to be functional, and potentially relevant to eye disorders by using the G2P plugin (289,290) in VEP [release 90.1] and the Eye Gene Panel (291). In short, we extracted only variants satisfying the inheritance requirements of the genes in the Eye Gene Panel, with MAF in public databases < 0.0001 for monoallelic and X-linked genes and $MAF < 0.005$ for biallelic genes and annotated by VEP to have one of the following consequences: stop gained, stop lost, start lost, frameshift variant, inframe insertion/deletion, missense variant, coding sequence variant, initiator codon variant, transcript ablation, transcript amplification, protein altering variant, splice donor/acceptor variant (i.e., canonical splice site) or splice region variant (i.e., either within 1-3 bases of the exon or 3-8 bases of the intron). I acknowledge the help of Dr Mihail Halachev and the core bioinformatics team.

4.2.10 Detection of intronic splice variants

Variants within the *PAX6* locus (chr11:31784779-31817961, GRCh38) were annotated with SpliceAI delta scores using SpliceAI v1.3 (288), with the help of David Parry, using transcript coordinates from the “GENCODE basic” transcript set from Ensembl release 95.

4.2.11 *Cis*-regulatory variant analysis

The GRCh38 coordinates of 35 different *cis*-regulatory elements chosen for analysis are given in **Supplementary Table 2** in **Appendix C.3**. Many thanks to Dr Shipra Bhatia for her CRE expertise in compiling this list. A BED file was created from this table, and bedtools (292) was then used to extract the *PAX6* CRE variants from the cohort VCF file, the latter by David FitzPatrick. Subsequent filtering used the gnomAD (200) allele frequency data in the VCF file.

4.2.12 Variant prioritisation and pathogenicity criteria

I assessed all of the bioinformatics WGS pipeline output and results of my SV analyses and assessed the likely pathogenicity of variants, in the context of the aniridia variant literature where applicable. I examined all plausibly causative variants on IGV. For

sequence variants I applied ACMG/AGVS pathogenicity criteria (293,294), using the Bayesian tool built into DECIPHER. I also acknowledge the help of Morad Ansari (Clinical Scientist, NHS South East Scotland Clinical Genetic Service) for advice on correct application of the pathogenicity criteria. I verified variant nomenclature and adherence to HGVS recommendations using Alamut software (Sophia Genetics) and Variant Validator (295).

4.2.13 Experimental analysis of splice variants

4.2.13.1 RT-PCR and nested PCR of lymphoblastoid cell line-derived RNA

Patient-derived lymphoblastoid cell lines (LCLs) were recovered from liquid nitrogen storage. I grew the cells in suspension in RPMI media containing 15% FCS and penicillin/streptomycin, and incubated at 37°/5% CO₂. I extracted RNA from the LCLs following the principles of the phenol-chloroform method (296) using TRIzol reagent (Invitrogen) following the manufacturer's instructions. DNase treatment was performed using TURBO DNase kit (Invitrogen). cDNA was obtained from total RNA using the SuperScript First Strand Synthesis System for RT-PCR kit (Invitrogen). I performed a nested PCR across the *PAX6* locus using four overlapping RT-PCR primer pairs (courtesy of Kathy Williamson) spanning exons 1 to 5, exons 3 to 8, exons 7 to 12 and exons 9 to 13 (primer table **Appendix C.2.6**). The products were run on a 2% low melting point agarose gel with ethidium bromide. I excised bands of interest, performed a gel DNA extraction (Zymoclean Gel DNA Recovery Kit, Zymo Research, Freiburg im Breisgau), and sent this for Sanger sequencing to look for missplicing.

4.3 Results

4.3.1 Assembling the cohort and prescreening

4.3.1.1 Cohort description, phenotypes and prior screening

In this section I shall summarise the phenotypic information and screening performed by referring laboratories and by us prior to finalising the cohort for WGS.

Those with classical aniridia phenotypes were prioritised for this study. Of 443 individuals from 347 families with aniridia in HGU eye malformations cohort, 45 ‘*PAX6* negative’ families were considered in total after review of phenotype and prior screening results (**Table 11**). These are divided in the table into those that made the final cohort, 37 families, and then 8 families excluded from the WGS cohort in the final round of pre-WGS screening described in the following subsections.

Note that due to the historic nature of the cohort, ‘*PAX6* negative’ status was subject to geographical and historical variation in screening practices. The majority but not all (approximately 60%), had had some form of deletion analysis.

Clinical phenotype information and prior screening results for the aniridia WGS cohort (1/3)

FID	PID	Proband/ relative	Pheno inherit- ance	PAX6 method*	11p13 del analysis	MAB21L1	Other genes** screened	ddPCR CNV PAX6	Phenotype***	Comments
75	75	p (trio)	sporadic	direct	FISH, aCGH	WT	no	normal	aniridia	
182	182	p	sporadic	direct	aCGH	WT	no	normal	partial aniridia, atrophic iris, congenital cataracts, nystagmus, corneal endothelial degeneration	
224	224	p	unk	direct	aCGH	WT	no	normal	aniridia	
356	356	p (trio)	sporadic	DHPLC	FISH	WT	no	normal	bilateral aniridia, congenital cataracts	46,XX t(1;9)(p36.1;q22) found at amniocentesis
535	535	p	sporadic	direct	aCGH	WT	no	no	classical bilateral aniridia	
660	660	p	unk	direct	aCGH	WT	no	normal	aniridia, possible foveal hypoplasia	
723	724	p	familial	DHPLC	likely FISH	WT	no	no	aniridia, type II diabetes	
774	774	p	unk	DHPLC	no	WT	no	normal	bilateral aniridia, cataract, drusen at macula and nasal to optic disc	
877	877	p	unk	DHPLC	FISH	WT	no	normal	bilateral aniridia, glaucoma, bilateral cataracts in young adulthood	
999	999	p	unk	DHPLC	no	WT	no	normal	partial aniridia	
1019	1019	p	sporadic	DHPLC	FISH	WT	no	no	aniridia	
1142	1142	p	sporadic	DHPLC	unk	WT	no	no	aniridia	
1190	1191	p	familial	DHPLC	unk	WT	no	no	partial aniridia, early onset cataracts with surgery in late 20s	
1271	1271	p	sporadic (+ aff son)	DHPLC	aCGH	WT	no	no	bilateral aniridia, cataract extraction both eyes as a teenager	
1304	1304	p	familial	DHPLC	unk	WT	no	no	aniridia	
1326	1329	p (duo)	sporadic (+ aff sibs)	DHPLC	unk	WT	no	normal	aniridia, nodular corneal dystrophy	other family members have classical phenotype
1326	1328	aff rela- tive (duo)	familial	<i>not tested (relative tested)</i>			no	normal	aniridia	nephew of 1326_1328
1358	1358	p	sporadic	DHPLC	"chromosome s normal"	WT	no	no	aniridia, congenital glaucoma, spherophakia, high myopia from birth	consanguinity, aff sib, likely recessive or gonadal mosaic
1361	1361	p	familial	DHPLC	FISH	WT	no	no	aniridia	
1451	1451	p	unk	DHPLC	FISH	WT	no	no	aniridia, congenital glaucoma, bicuspid aortic valve, mild aortic stenosis	
1468	1468	p	unk	DHPLC	likely FISH	WT	no	no	bilateral aniridia	

Clinical phenotype information and prior screening results for the aniridia WGS cohort (continued; 2/3)

FID	PID	Proband/ relative	Pheno inherit- ance	PAX6 method*	11p13 del analysis	MAB21LI	Other genes** screened	ddPCR CNV PAX6	Phenotype***	Comments	
1496	1496	p	unk	DHPLC	FISH	WT	no	no	aniridia		
1500	1500	p	sporadic	DHPLC	aCGH	WT	no	no	isolated bilateral aniridia, cataracts (posterior subcapsular), absent foveal reflexes, no nystagmus, vision R 6/12 N5, L 6/36 N5; corneas clear with fine limbal vessels peripherally		
1524	1524	p	familial	DHPLC	unk	WT	CHX10	no	aniridia spectrum: iris & foveal hypoplasia, congenital corneal opacification with small lenses apposed to posterior surface of cornea		
1607	1607	p	sporadic	DHPLC	FISH	WT	no	no	sporadic aniridia, congenital glaucoma		
1635	1635	p (trio)	sporadic	DHPLC, plus direct for PD	aCGH	WT	FOXC1 SOX2 OTX2	no	variant aniridia: inferior iris defects, peripheral corneal vascularisation, inferior lens opacities, foveal hypoplasia, nystagmus, poor vision		
1646	1647	p	unk	DHPLC	unk	WT	no	no	aniridia with glaucoma		
1648	1648	p	unk	direct	aCGH, FISH	WT	no	no	aniridia		
1732	1732	p	unk	DHPLC	unk	WT	no	no	bilateral aniridia, glaucoma, opaque R cornea, aortic stenosis requiring neonatal surgery; not dysmorphic		
1878	1879	p	unk	DHPLC	FISH	WT	no	no	aniridia spectrum anterior chamber abnormality, abnormal iris; multiple operations in early childhood for bilateral glaucoma		
1943	1943	p	familial	DHPLC	"11p normal"	WT	no	no	bilateral aniridia		
2134	2134	p (trio)	sporadic	DHPLC	likely FISH	WT	no	no	bilateral aniridia, nystagmus, glaucoma		
2197	2197	p	unk	unk; rescree- ned direct	"chromosome s normal"	WT	no	no	bilateral aniridia, diaphragmatic eventration, undescended testicle & hydrocele; small kidneys		
2464	2466	p (duo)	sporadic	DHPLC	unk	WT	no	no	aniridia	half sibs both affected	
2464	2464	aff relative (duo)		not tested (relative tested)							half sib of 2464_2466
2469	2469	p (trio)	sporadic	DHPLC	unk	WT (uni- lateral reads)	no	normal	aniridia	only LCL DNA available for unaffected parents	
3612	3612	p	sporadic	direct	aCGH, FISH	WT	no	normal	bilateral aniridia, relatively well-preserved foveas, vision 0.4 logMAR, mild keratopathy as teenage, mild cataracts		
4340	4340	p	sporadic (+ aff son)	direct	unk	WT	SOX2 OTX2	normal	bilateral aniridia, reasonably good vision		
5645	5645	p (trio)	sporadic	direct (by referring centre)	MLPA, microarray	WT	FOXC1PI TX2 ITPR1	no	isolated bilateral partial aniridia		

Clinical phenotype information and prior screening results for the aniridia WGS cohort (continued; 3/3)

FID	PID	Proband/ relative	Pheno inherit- ance	PAX6 method*	11p13 del analysis	MAB21L1	Other genes** screened	ddPCR CNV PAX6	Phenotype***	Comments
EXCLUDED prior to WGS (reason in comments):										
851	851	p (duo)	familial	DHPLC	unk	WT	no	no	aniridia	insufficient gDNA stocks
851	852	aff rela- tive (duo)	familial	<i>not tested (relative tested)</i>						
1144	1144	p	familial	DHPLC	no	failed	no	no	aniridia	gDNA degraded
1201	1201	p (trio)	sporadic	DHPLC	unk	WT	no	SIMO probe CN=1	bilateral aniridia (L. partial), no cataract/glaucoma	deletion in PAX6 DRR (SIMO) identified on ddPCR
1233	1233	p	unk (adopted)	DHPLC	unk	failed	no	no	aniridia, strabismus	insufficient gDNA stocks
1606	1606	p	familial	DHPLC; rescreened by direct	FISH	WT	no	no	aniridia	PAX6 rescreened: NM_000280.4:c.1261del p.(Leu421Tyrfs*104)
2024	2024	p	sporadic	unk; rescreened by direct	unk, possibly FISH	WT partial	no	no	bilateral aniridia, congenital cataracts, nystagmus	PAX6 rescreened: NM_000280.4:c.781C>T, p.(Arg261*)
4674	4674	p	unk	direct (PD only); rescree- ned by direct	unk	WT	SOX2 OTX2	no	bilateral aniridia	PAX6 rescreened: NM_000280.4:c.949C>7, p.(Arg317*)
5656	5656	p (trio)	sporadic	unk; rescreened by direct	array	WT partial	no	no	bilateral aniridia	PAX6 rescreened: NM_000280.4:c.112_116del p.(Arg38Vfs*16)

Abbreviations: FID/PID, family/personal ID number; p, proband; aff, affected; sibs, siblings; unk, unknown; DHPLC, denaturing high performance liquid chromatography; direct, direct Sanger sequencing; PD, paired domain; FISH, fluorescence in-situ hybridisation; aCGH, array comparative genomic hybridisation; MLPA, multiplex ligation-dependent probe amplification; WT, wild type; ddPCR, droplet digital PCR (using 4 probes spanning PAX6/ SIMO).

Notes: * Prior PAX6 screening mainly performed in the HGU but in some cases by the referring centre. ** Other genes screened: results of any other genes screened other than PAX6 and MAB21L1. *** Phenotype information: where simply 'aniridia' is listed, this limited information was all that was received with the referral.

Table 11 Clinical phenotype information and prior screening results for the aniridia WGS cohort

4.3.1.2 Assessment of quality and quantity of genomic DNA

3/45 families were excluded on the basis of insufficient quality or quantity of proband DNA held for WGS: family IDs (FID) 851, 1144 and 1233.

In the case of FID 2134 and 2024, gDNA for the proband was extracted afresh from peripheral blood leucocyte (PBL) stocks which had been preserved in liquid nitrogen. In the case of FID 2469, whilst the proband gDNA was satisfactory, no gDNA or PBLs were available for the unaffected parents. However, lymphoblastoid cell lines (LCLs) were available for both parents. In this trio only, therefore, DNA was extracted from LCLs for the parental samples and we were cognisant of this when analysing the results.

In the remaining cases, gDNA of adequate quality and quantity was available for probands and relatives.

4.3.1.3 Additional screening of *PAX6* and *MAB21L1*

For some cases where the previous *PAX6* sequencing result was ambiguous, incomplete or missing, I resequenced all the exons and splice sites of *PAX6*. This resulted in the detection of 4 unrelated individuals with 4 different likely gene-disruptive variants in *PAX6* (**Table 11**, page 3/3), which are all causative variants for aniridia and 3 of which are in the top 10 commonest alleles seen in aniridia (Introduction, **Figure 5**). These 4/45 cases were therefore excluded from the WGS cohort.

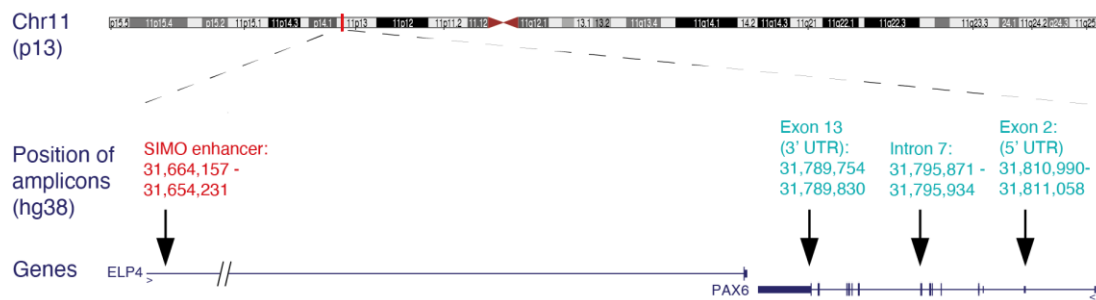
Additionally, based on our now-published (166) *MAB21L1* human genetic data (**Chapter 3**), we chose to screen *MAB21L1* prior to WGS in order to look for Arg51 variants. Eight cases were screened by me and the remainder by my colleague Kathy Williamson. No new plausibly causative *MAB21L1* variants were detected, though much of the larger HGUeye cohort had already had this gene screened.

4.3.1.4 Droplet digital PCR as a screen for *PAX6* locus copy number variants

As 40% of the potential cohort had not had any sort of deletion analysis, I performed a screen for copy number variants (CNVs) on a smaller subset of 13 randomly chosen cases as a pilot screen using droplet digital PCR (ddPCR).

This revealed one case (1/45) with a deletion of the probe (75 bp) covering the SIMO auto-enhancer, a critical *cis*-regulatory region of *PAX6* (**Figure 34**). This was a child with sporadic bilateral aniridia, complete in the right eye and partial in the left. A deletion of SIMO would be predicted to cause loss of function of the gene and thus is consistent with the aniridia phenotype (11). This family trio, FID 1201, was excluded from further analysis/WGS and the result was reported back to the referring clinicians.

A ddPCR amplicons



B Copy number determined by ddPCR, FID 1201

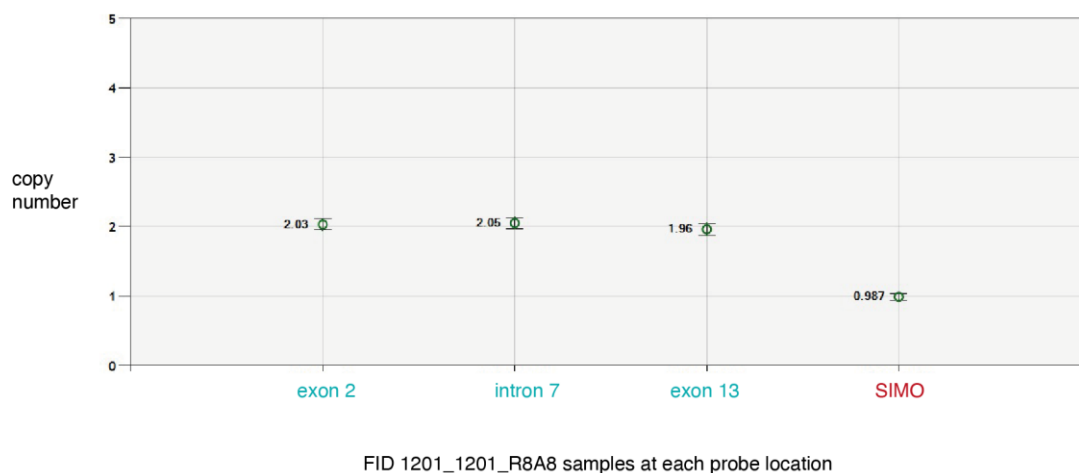


Figure 34 Copy number variation (CNV) at the *PAX6* locus assessed by droplet digital PCR, revealing a heterozygous deletion involving the cis-regulatory element, SIMO. (A) The CNV of 4 probes spanning the *PAX6* locus were tested, 3 amplicons covering the gene itself coloured in turquoise (5' UTR in exon 2, the middle of the gene in intron 7, and 3' UTR in exon 13) and the final one, shown in red, covering SIMO, a critical *cis*-regulatory enhancer region which is approximately 150 kb 3' to the start of *PAX6*. (B) In this case, FID 1201, the intragenic amplicons showed a normal copy number, but the 75 bp amplicon targeting the SIMO region showed a copy number of 1, indicating a heterozygous deletion. The exact breakpoints were not mapped. This would be expected to cause loss of function of one allele of *PAX6* and fully explain the aniridia phenotype

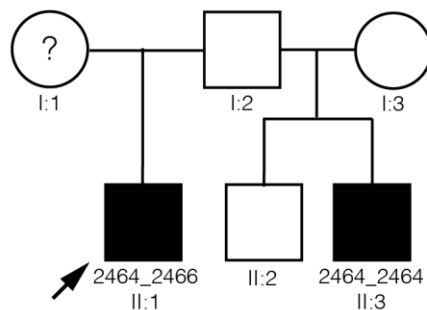
All the other cases screened as part of the pilot cohort returned a normal copy number of 2 across the *PAX6* locus. Although, based on this pilot experiment, we predicted that some of the remaining cases in the WGS cohort could be solved by ddPCR, we elected not to expand this to the remainder of the cohort. Whilst this left potential

PAX6 deletions within the WGS cohort, this was at an acceptable level and the ease of detection of these deletions on WGS was also of interest.

4.3.1.5 Final cohort for whole genome sequencing

Thus based on the above, a total of 8/45 families were excluded. The final cohort of 51 individuals (39 affected) from 37 families proceeded to WGS. Of these, 29 were single probands, 6 were trios (proband plus unaffected parents) and 2 were duos (proband plus affected relative). Pedigrees for the two duos are shown in **Figure 11**.

A Family 2464, duo (affected half siblings)



B Family 1326, duo (affected aunt and nephew)

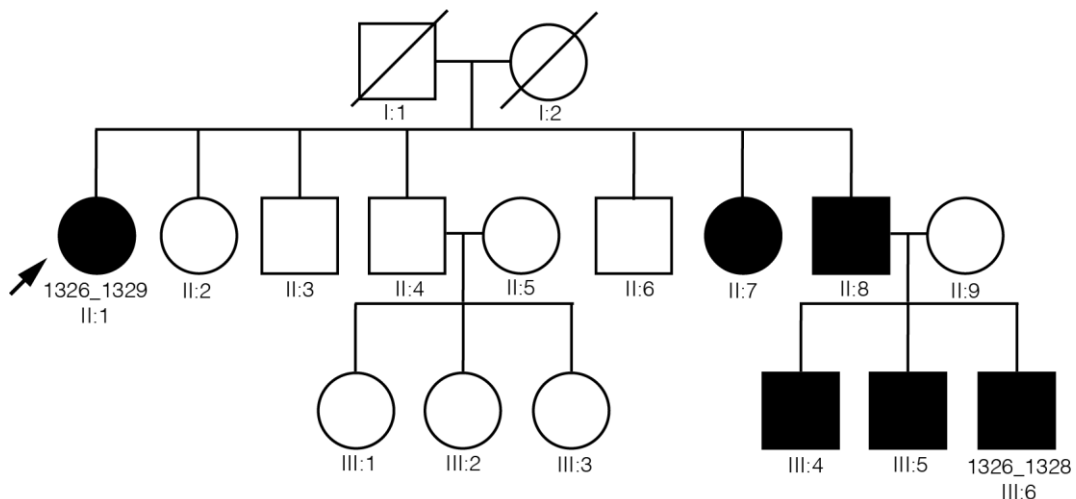


Figure 35 Pedigrees for the two families included in the WGS cohort as duos (proband plus affected relative). Family members with aniridia are shaded in black. Those with an ID number underneath (FID_PID, family_personal ID respectively) were sent for WGS. (A) Family 2464, affected half siblings with aniridia sharing an unaffected father (I:2), who is presumed to have gonadosomatic mosaicism. The mother of 2464_2466 (I:1) is recorded as “?” as it is not known for certain that she was unaffected. (B) Family 1326, a large family with several members affected by aniridia. The aunt (proband; 1326_1329, II:1) and nephew (1326_1328, III:6) duo was included in the WGS cohort

WGS was performed on Illumina HiSeq X by BGI, Hong Kong, under project number F16FTSEUHT0678 (9 samples), and by Edinburgh Genomics, UK, under project number X18041 (42 samples).

The following analyses of WGS data to identify structural and sequence variants were performed in parallel.

Please note, throughout the results section, I have used principally the Family ID (FID) rather than personal ID (PID) for simplicity. The majority (29/37 families) are single probands (where FID=PID) and, in those that are not, I discuss one causative variant affecting either both members (duo) or the proband of a nuclear trio, so it tends to be clearer discussing per family. All FID and PIDs are listed in **Table 11**. (In the publication (276) a slightly different approach was taken and the default currency for discussion is the individual research participant ID.)

4.3.2 Structural variants identified in WGS aniridia cohort

The variants identified in the aniridia WGS project are divided into two sections: this section, covering large scale Structural Variants (SVs), and then the next (4.3.3) covering sequence variants (comprising SNVs and short insertions/deletions). The plausibly pathogenic variants detected are summarised in two tables which lead their respective sections: **Table 12**, here, covering SVs, and **Table 13** in the sequence variant section.

The expertise of the HGU core bioinformatics service is acknowledged here, particularly Alison Meynert, Graeme Grimes and Mihail Halachev. All of my WGS data interpretation and analysis that is described in the following sections rests on their performing joint genotyping, annotation and running the variant filtering pipeline.

Seventeen unique, ultra-rare, heterozygous structural variants (SVs) were detected in 15 families (**Table 12**), using a combination of direct inspection of candidate loci using the Integrative Genomics Viewer (IGV) (297) and genome-wide bioinformatic tools (Canvas (298) and Manta (285)).

Affected individuals from 3/17 families had overlapping SVs all resulting in a whole gene deletion of *FOXC1*. The remainder of the SVs detected (14/17) all affected the wider *PAX6* locus, demonstrating a variety of ways of inactivating the gene. Affected individuals from 2 families had complex SVs, one (FID 356) with a *de novo* inversion adjacent to a deletion around the N-terminus of *PAX6*, and the other (FID 1271) with a tandem duplication adjacent to a deletion involving the C-terminus.

FID	PID	Proband/ duo/trio	Heterozygous variant type	De novo	Region affected*	Size (kb)	Chr	Coordinates (GRCh38)
Simple deletions affecting PAX6 transcription unit:								
75	75	p (trio)	deletion	yes	PAX6 intron 7/exon 8	0.126	11	31,794,763-31,794,889
723	724	p	deletion	unk	PAX6 whole gene deletion	191	11	31,714,074-31,905,175
1496	1496	p	deletion	unk	C-termini of both <i>ELP4</i> and <i>PAX6</i> (intron 8 onwards)	19	11	31,777,022-31,796,016
1524	1524	p	deletion**	unk	PAX6 exons 6 and 7	1.4	11	31,800,812-31,802,170
2464	2464, 2466	duo	deletion	unk	C-termini of both <i>ELP4</i> and <i>PAX6</i> (intron 4 onwards)	41	11	31,762,967-31,803,741
Deletions affecting cis-regulation of PAX6:								
1190	1191	p	deletion	unk	PAX6 DRR	674	11	31,009,535-31,683,449
1361	1361	p	deletion	unk	PAX6 DRR	598	11	31,118,961-31,716,799
1646	1647	p	deletion	unk	PAX6 DRR	842	11	30,837,881-31,680,038
Balanced SVs affecting PAX6 locus:								
535	535	p	inversion	unk	11p13-11p14.3 with intragenic <i>PAX6</i> breakpoint	7300	11	24,479,030-31,792,704
774	774	p	inversion	unk	inversion separating <i>PAX6</i> from DRR; breakpoints <i>ELP4&RAG2</i>	4900	11	31,701,464-36,593,500
Complex SVs affecting PAX6 locus:								
356	356	p (trio)	inversion (complex)	yes	N-terminus of <i>PAX6</i> up to intron 4	6	11	31,805,500-31,811,582
1271	1271	p	deletion (complex)	yes	upstream of <i>PAX6</i> (P0, EE enhancer)	30	11	31,811,583-31,842,057
			tandem duplication (complex)	unk	final intron of <i>ELP4</i>	13	11	31,765,420-31,778,063
			deletion (complex)	unk	C-termini of both <i>ELP4</i> and <i>PAX6</i> (intron 8 onwards)	16	11	31,780,738-31,796,994
Simple deletions affecting FOXC1:								
1142	1142	p	deletion	unk	<i>FOXC1</i> whole gene deletion	82.6	6	1,544,064-1,626,702
1451	1451	p	deletion	unk	<i>FOXC1</i> whole gene deletion	33	6	1,581,517-1,615,082
1732	1732	p	deletion	unk	<i>FOXC1</i> whole gene deletion	81.6	6	1,604,452-1,686,063
Abbreviations: FID/PID, family/personal ID; p, proband; DRR, downstream (3') regulatory region. Notes: *Exon numbering for <i>PAX6</i> is as per NM_000280.4 (-5a, 13 exons) **This individual's deletion was also detected independently to this study elsewhere and is patient 11 in the cited publication (299)								

Table 12 Structural variants considered likely pathogenic detected in WGS aniridia cohort

4.3.2.1 Genomic copy number variants

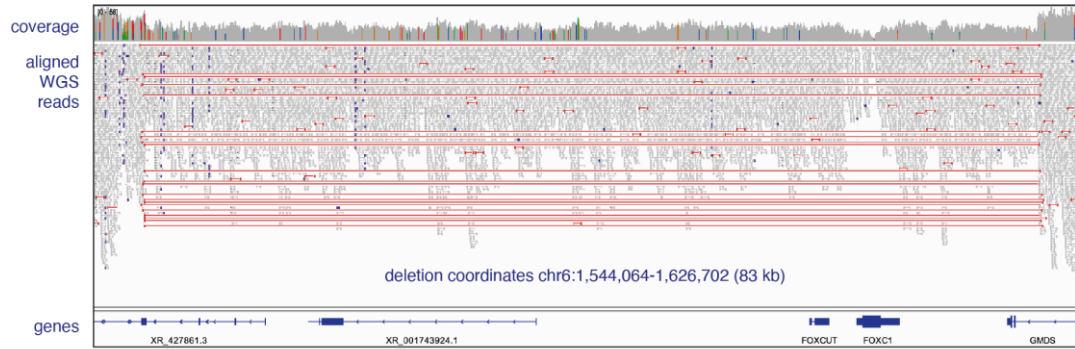
4.3.2.1.1 *FOXC1* whole gene deletions

Three unique chromosome 6p deletions encompassing *FOXC1* were identified (**Figure 36**) in 3 probands, FIDs 1142, 1451 and 1732. These CNV ranged from 33 to 83 kb in size and in each case resulted in a heterozygous whole gene deletion of *FOXC1*. All three individuals had aniridia, 2/3 had glaucoma (one confirmed as congenital), and 2/3 had congenital aortic or aortic valve anomalies. As discussed in the introduction **Section 1.9**, *FOXC1* deletions are a well-described cause of an anterior segment dysgenesis involving aniridia, particularly in conjunction with congenital glaucoma (148,155). The combination of this ocular phenotype with congenital aortic valvular disease is very characteristic (155,300,301)

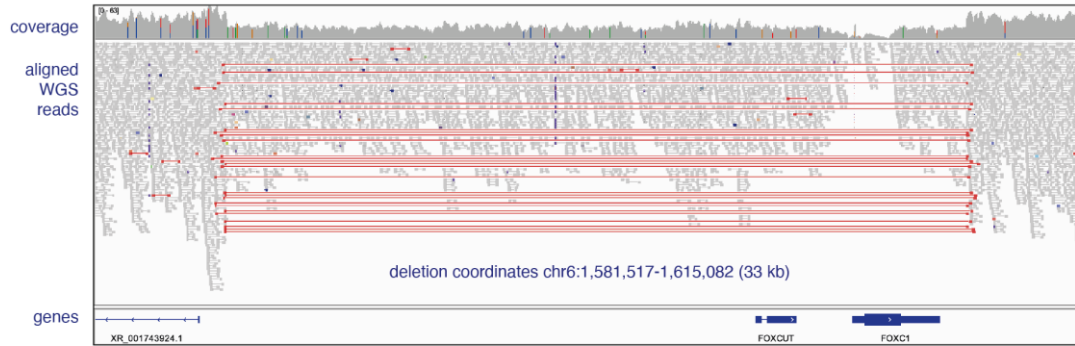
Deletions encompassing *FOXC1* on chromosome 6 detected on WGS

6p25.3 

A FID 1142



B FID 1451



C FID 1732

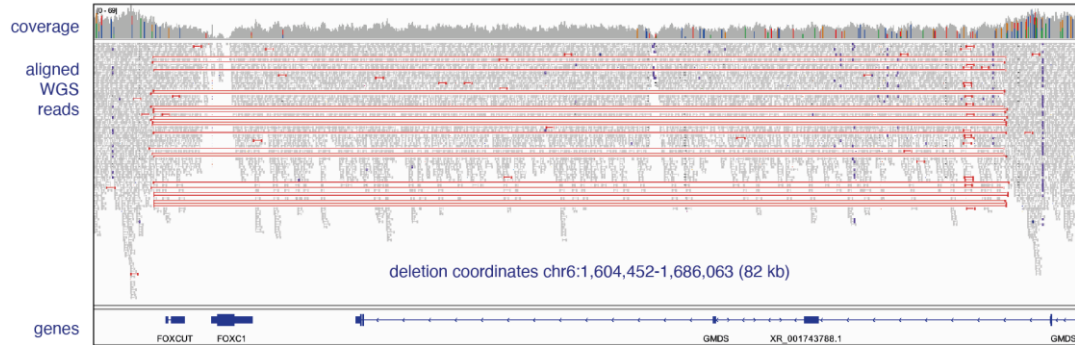


Figure 36 Three unique CNVs on chromosome 6 resulting in heterozygous whole gene deletions of *FOXC1*. For each individual (A-C) identified by family ID number, an IGV snapshot of the aligned WGS data is shown. Corresponding drop in coverage is seen on the coverage track above the WGS reads. Estimated coordinates from WGS are given (GRCh38). Individual 150 bp reads are paired (viewed with a connecting line) and coloured by insert size, with a red colour indicating that the inferred insert size between paired reads is larger in the reference genome than in the sample, suggesting a deletion. The coverage over the *FOXC1* gene itself was poor, as expected for an area of high GC content. (A) FID 1142, a young child with sporadic aniridia and an 83 kb deletion encompassing *FOXC1*, chr6:1,544,064-1,626,702. (B) FID 1451, a child with aniridia presenting as congenital glaucoma, bicuspid aortic valve and mild aortic stenosis. 33 kb deletion, chr6:1,581,517-1,615,082. (C) FID 1732, a child with bilateral aniridia and glaucoma, and aortic stenosis requiring neonatal surgery. 82 kb deletion, chr6:1,604,452-1,686,063.

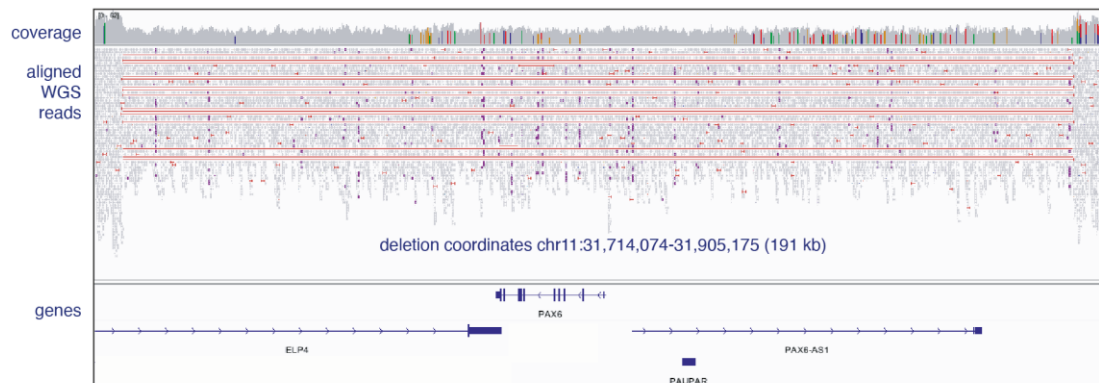
4.3.2.1.2 *PAX6* partial or whole gene deletions

Analysis of the single proband WGS data by direct inspection on IGV revealed three individuals with heterozygous copy number variants (CNVs) affecting the *PAX6* gene itself (**Figure 37**). Of these, one was a 191 kb deletion resulting in a whole gene deletion of *PAX6* as well as some of its up- and down-stream regulatory regions. The other two, with 19 kb and 1.4 kb deletions, had partial gene deletions of *PAX6*. All three individuals had aniridia. In FID 1524, the proband had an aniridia spectrum phenotype comprising iris hypoplasia, foveal hypoplasia, congenital corneal opacification with slightly small lenses apposed to the posterior surface of the cornea; this individual's variant was subsequently found to have been detected independently by others (299).

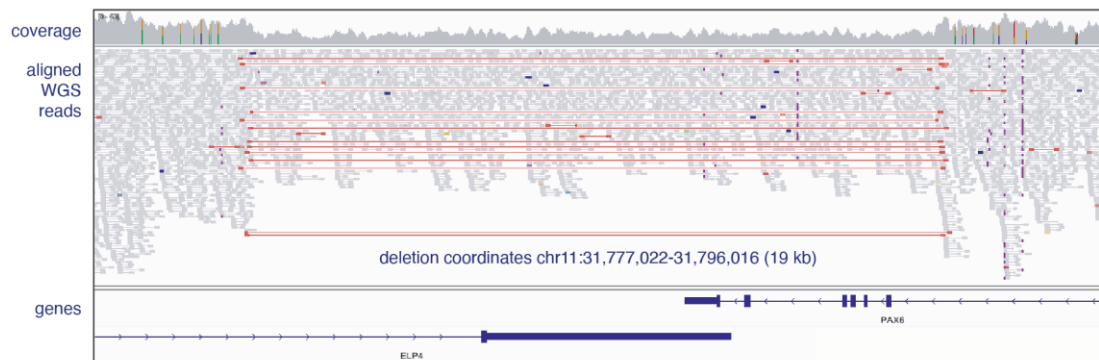
Deletions encompassing *PAX6* on chromosome 11 detected on WGS: single probands



A FID 723



B FID 1496



C FID 1524

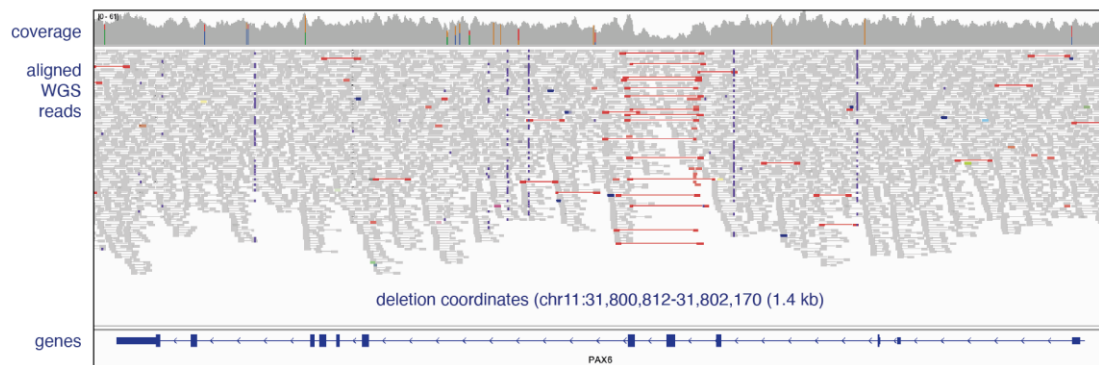


Figure 37 CNVs on chromosome 11 resulting in heterozygous whole or partial gene deletions of *PAX6* in three single probands with aniridia. For each individual (A-C), identified by their family ID number (FID), an IGV snapshot of the aligned WGS data is shown. Corresponding drop in coverage can be seen on the coverage track above the WGS reads. Estimated coordinates of the deletion from WGS are given (all GRCh38). The individual 150 bp reads are paired (viewed here with a line connecting each pair) and coloured by insert size, with a red colour suggesting a deletion. The equivalent colour coding for an insertion is dark blue for paired reads, or purple for a single base insertion within a read. Other colours denoting SNVs are visible on the coverage track. (A) FID 723, 191 kb chromosome 11 deletion resulting in whole gene deletion of *PAX6* and

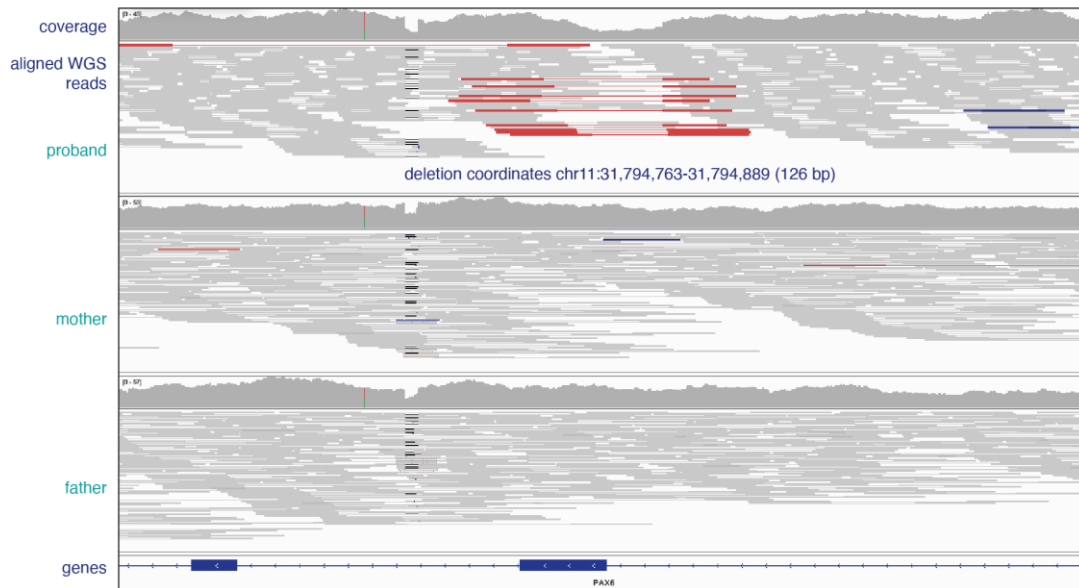
much of its wider regulatory regions, chr11:31,714,074-31,905,175. (B) FID 1496, 19 kb deletion involving C-terminal portions of both *PAX6* and *ELP4* (which are transcribed in opposite directions, towards each other); *PAX6* is missing from the end of intron 7 onwards; chr11:31,777,022-31,796,016. (C) FID 1524, proband is a young boy with familial aniridia whose phenotype comprised partial aniridia, foveal hypoplasia and congenital corneal opacification. WGS showed a 1358 bp deletion involving exons 6 and 7 of *PAX6*, chr11:31,800,812-31,802,170.

Analysis of the trio and duo WGS data on IGV showed two families with heterozygous partial *PAX6* gene deletions (**Figure 38**). The proband of FID 75, who has sporadic aniridia, has a small *de novo* deletion involving the intron 7-exon 8 boundary of *PAX6*. The half siblings with aniridia of FID 2464, who share an unaffected father, have a 41 kb deletion encompassing the majority of the *PAX6* gene (intron 4 onwards) and the C-terminus of *ELP4*, similar to FID 1496 in **Figure 37** above.

Deletions involving *PAX6* detected on WGS in family trio and duo data

11p13 

A FID 75



B FID 2464

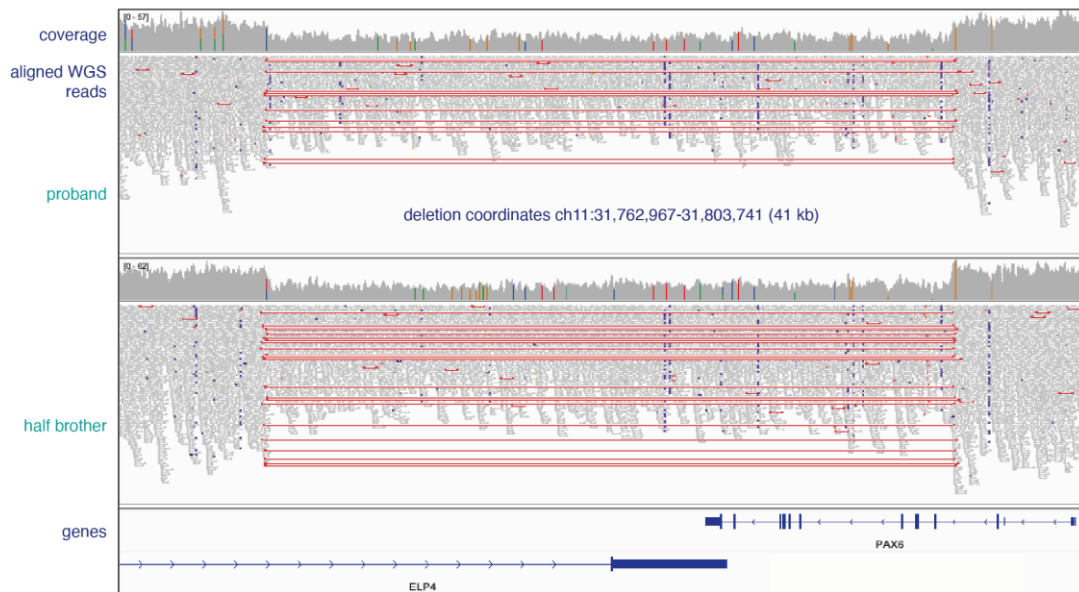


Figure 38 chromosome 11 CNVs resulting in partial gene deletion of *PAX6*: aligned WGS data (IGV) from a family trio and duo. As previously, the data is displayed with paired reads connected by lines, red colouring suggesting a deletion based on the insert size versus the reference genome. There is a corresponding drop in coverage depth over the area of the heterozygous deletion. Deletion coordinates are estimated (GRCh38). Each family is identified by the FID number. (A) FID 75 (trio): the proband with sporadic aniridia has a *de novo* 126 bp deletion involving the intron 7 and exon 8 of *PAX6*, chr11:31,794,763-31,794,889. (B) FID 2464 (duo): d two affected half siblings with aniridia, showing a 41 kb deletion involving a large proportion of the *PAX6* gene, from intron 4 onwards, along with the C-terminus of *ELP4*; chr11:31,762,967-31,803,741. Their father was unaffected and presumed to have gonadal mosaicism

4.3.2.1.3 Deletions of *PAX6* cis-regulatory regions

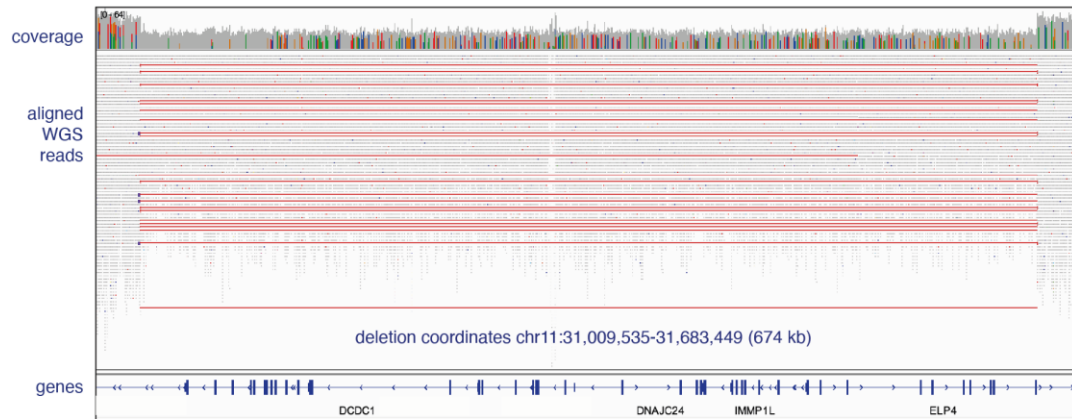
Whilst the previous sections describe whole or partial aniridia-associated gene deletions, I identified three individuals in whom *PAX6* itself was intact but large areas of the *PAX6* downstream regulatory region were deleted. All three had fairly classical aniridia phenotypes. In all three, key *cis*-regulatory elements driving *PAX6* expression in the developing eye were deleted, most notably SIMO and HS5 which have both been implicated in classical aniridia (11,302). Aligned WGS data for these deletions is shown in **Figure 39**.

The two subsequent figures put these deletions into a wider perspective. **Figure 40** shows them in the context of the full extent of all the SVs detected on chromosome 11 (including balanced SVs described in the next section) and their relationship to the *PAX6* topologically-associated domain. **Figure 41** looks in on a tenfold smaller area to illustrate the relationship between the SVs and the wider *PAX6* locus including individual *cis*-regulatory elements (CREs) and neighbouring genes.

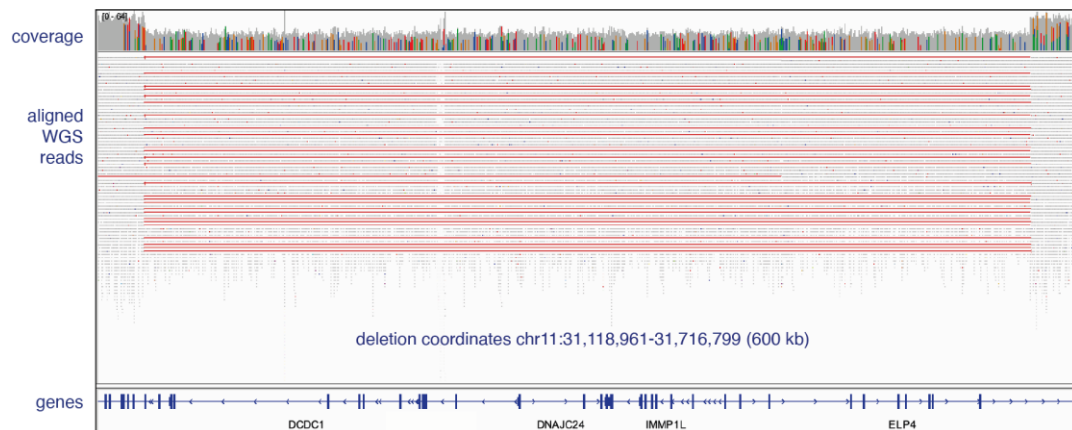
Deletions on chromosome 11 involving the downstream regulatory region of *PAX6*

11p13 - 11p14.1 

A FID 1190



B FID 1361



C FID 1646

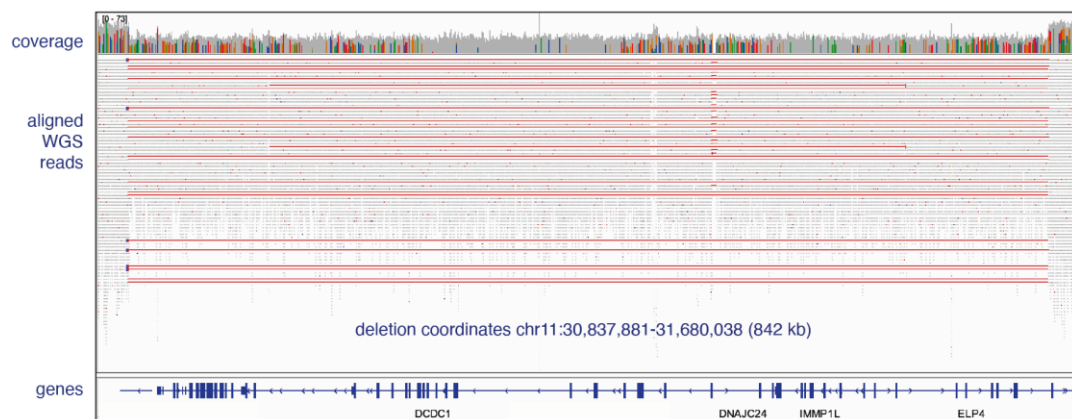


Figure 39 Deletions detected on WGS involving the *PAX6* downstream regulatory region on chromosome 11. All of these were detected in single probands, identified by the FID number. Aligned WGS data is shown for each individual with, as previously, the data displayed with paired reads connected by lines; red suggests a deletion based on the insert size compared to the

reference genome. The coverage track above shows a corresponding decrease over the area of the heterozygous deletion. Estimates of the deletion coordinates are in GRCh38. (A) FID 1190: an individual with partial aniridia and early onset cataracts, requiring surgery in late twenties; chr11:31,009,535-31,683,449. (B) FID 1361: an individual with aniridia from a large family with many affected members (the latter not in this cohort); chr11:31,118,961-31,716,799. (C) FID 1646: an individual with aniridia and glaucoma; chr11:30,837,881-31,680,038.

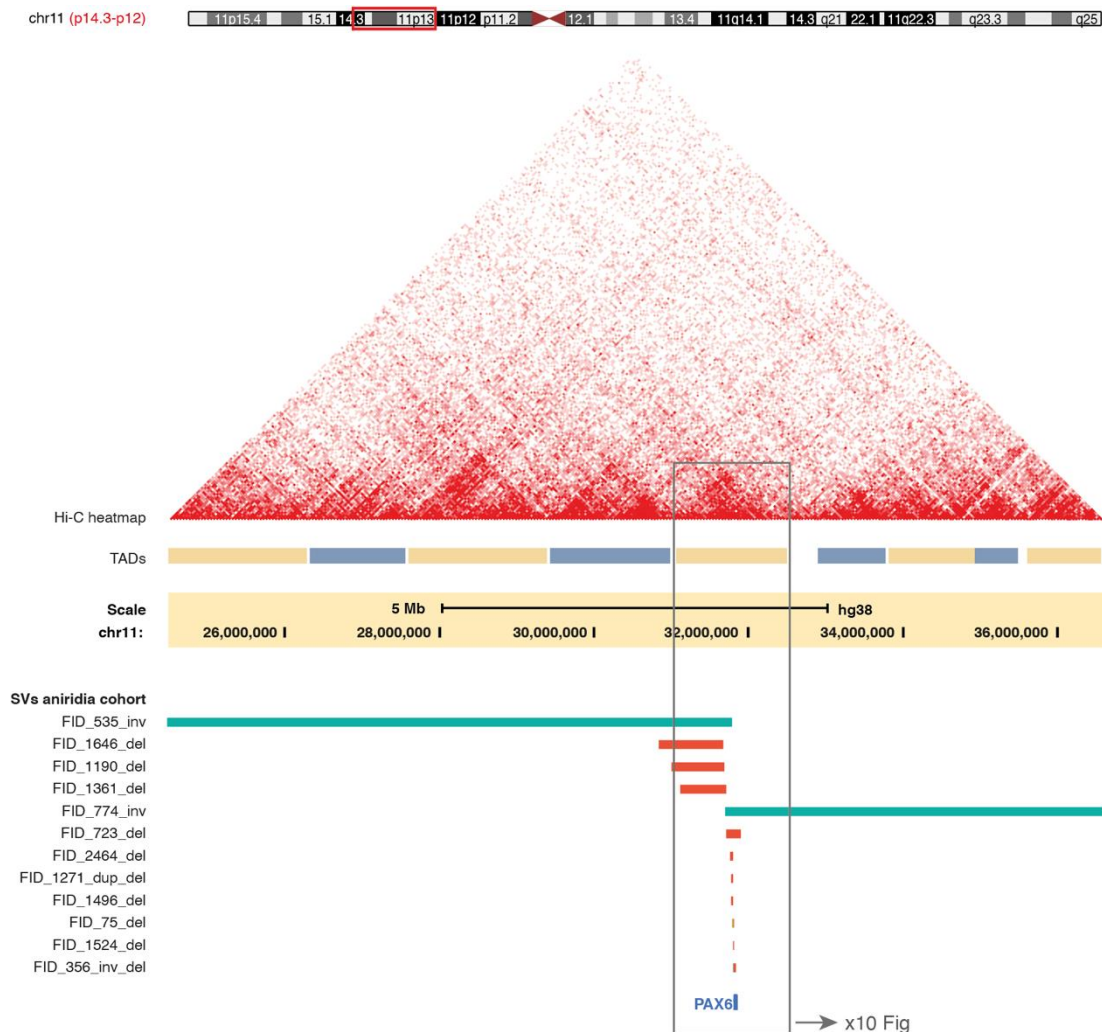


Figure 40 Structural variants (SV) identified on chromosome 11 and their relationship to the PAX6 locus. This figure shows approximately 12 Mb (just under one tenth) of chromosome 11, marked on the ideogram at the top and then expanded below. For each SV, shown as horizontal bars, the family ID (FID) is given, along with the type of SV (inv, inversion, teal; del, deletion, red; dup, duplication, green). The two largest SVs are FID 535, 7.3 Mb, and FID 774, 4.9 Mb, both inversions. The topologically associated domains (TADs), illustrated by the Hi-C heatmap (data from cortex (303,304)), are marked in yellow and blue blocks. The area of the PAX6 TAD is marked (grey rectangle): this allows easy comparison with **Figure 41** which shows this area in 10x higher resolution and allows us to see the relationship of the SVs with the *PAX6* cis-regulatory elements and neighbouring genes.

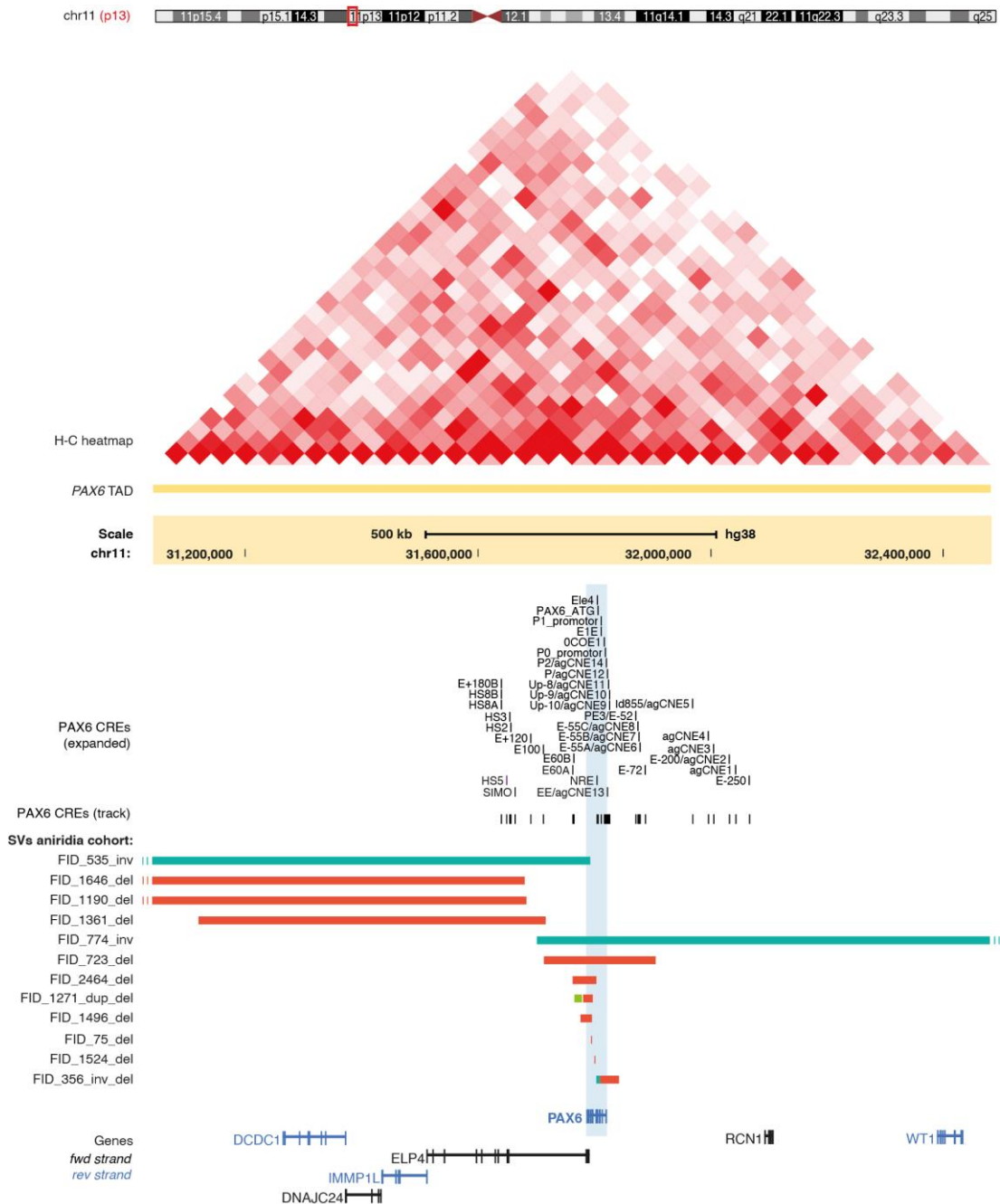


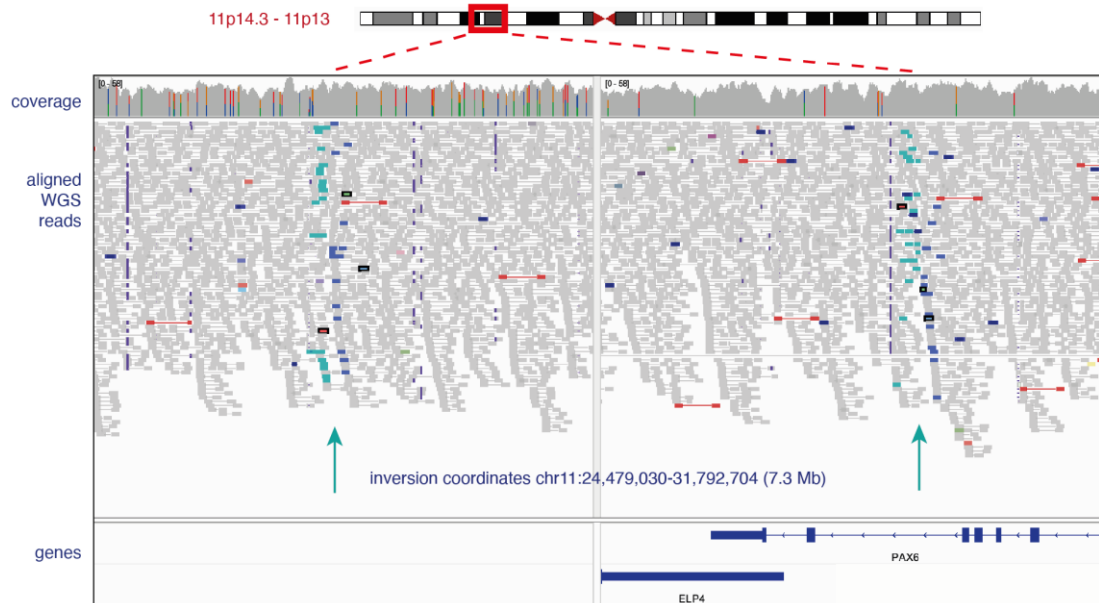
Figure 41 Structural variants (SV) identified on WGS affecting the wider *PAX6* locus. Each SV is shown as horizontal bars identified by the family ID (inv, inversion, teal; del, deletion, red; dup, duplication, green). This figure covers the exact area of the *PAX6* topologically associated domain (TAD), as indicated by the Hi-C heatmap (data from cortex (303,304)). **Figure 40** shows the full span of the SVs, whereas this is a 10x smaller area allowing us to see the relationship of the SVs with the *PAX6* cis-regulatory elements (CREs), above, and with neighbouring genes, below. A track of all the CREs is shown and then expanded above, with the locations of key CREs expressed in the eye (particularly SIMO, HS5 E60A, NRE, EE) easily visible towards the bottom of this group. The coordinates of the *PAX6* itself are shaded blue, making it clear that 7 of the SVs have intragenic breakpoints and FID 723 results in a whole gene deletion. Of the remaining SVs, 3 are deletions of the downstream regulatory region, taking out some CREs implicated in aniridia, notably SIMO and HS5. Similarly, FID 774 inverts the *PAX6* gene itself, disrupting its relationship with these enhancers.

4.3.2.2 Balanced structural rearrangements

Two inversions of chromosome 11 were detected by visual inspection of WGS data around the *PAX6* locus. The full extent of these SVs can be seen above in **Figure 40**. Whilst one, FID 535, has an intragenic *PAX6* breakpoint, the other leaves *PAX6* intact but has a breakpoint between the gene itself and CREs including SIMO and HS5 (**Figure 41**), thus disrupting the spatial relationship between the gene and enhancers critical for its expression in the eye. The aligned WGS data for these variants is shown in **Figure 42**.

Chromosome 11 inversions detected on WGS
split screen showing each breakpoint (arrows)

A FID 535



B FID 774

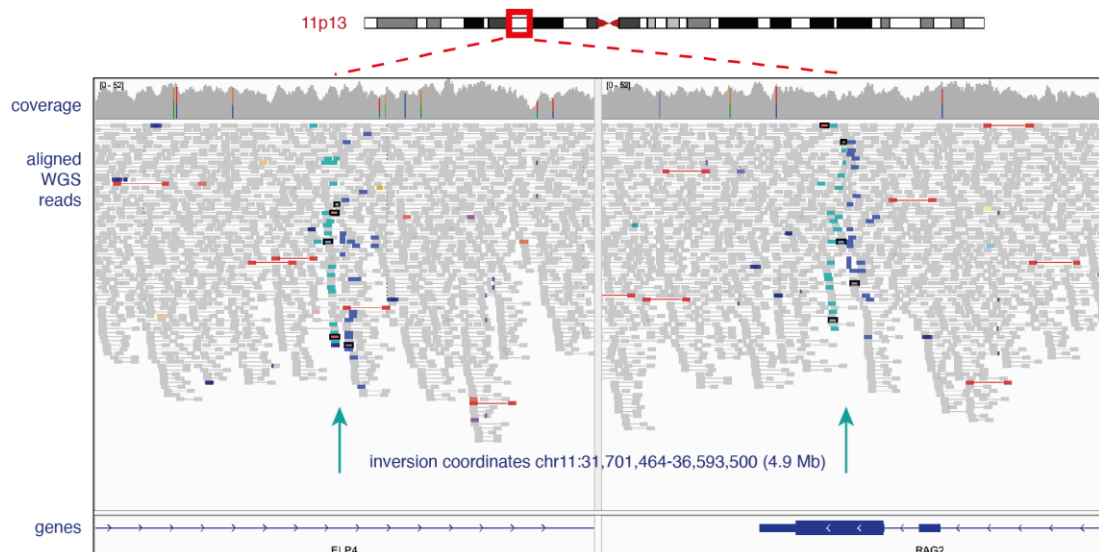


Figure 42 Balanced structural rearrangements of chromosome 11 detected on WGS. Two inversions were detected and aligned WGS data is shown for each, identified by FID. As both inversions were too large to capture on single IGV screen, a split-screen view of each breakpoint (arrows) is shown. The red box on the chromosome 11 ideogram shows the inverted region. At each breakpoint, coloured reads affecting approximately 50% of the reads can be seen: the reads are coloured by pair orientation, with the blue and teal colours indicate that paired read has been found in an unexpected orientation to its mate (with respect to the reference genome), implying an inversion. Consistent with the balanced nature of the rearrangement, no drop in coverage is seen on the coverage track.

(A) FID 535: an individual with classical bilateral aniridia, a 7.3 Mb inversion of chromosome 11 and an intragenic *PAX6* breakpoint in intron 11. Approximately 9 kb of aligned WGS data is shown around each breakpoint; chr11:24,479,030-31792704.

(B) FID 774: an individual with bilateral aniridia, central lens opacification and drusen at the macula and nasal to the optic disc. An inversion of 4.9 Mb was detected which leaves *PAX6* intact but disrupts its relationship with critical enhancers; chr11:31,701,464-36,593,500.

4.3.2.3 Complex structural rearrangements

Proband FID 356 was described by the referring clinicians at age 16 months as having sporadic bilateral aniridia and congenital cataract. WGS data for FID 356 revealed a causative *de novo* complex structural variant, but not the one initially expected from the referral:

4.3.2.3.1 Likely benign reciprocal translocation *t*(1,9)

Proband FID 356 had a known *de novo* *t*(1,9) reciprocal chromosomal translocation, detected on routine cytogenetic analysis following the aniridia diagnosis. This had been mapped by the referring laboratory to *t*(1,9)(p36.1; q22). *PAX6* Sanger sequencing and FISH had been normal. I hoped that analysis of this *de novo* complex structural variant in further detail using WGS might reveal new aniridia loci, in a similar way to *SOX2* and anophthalmia (43).

Through inspection of the known approximate breakpoints on IGV, sites consistent with the breakpoints were identified on WGS, with the paired-end reads mapping to reciprocal sites on chromosomes 1 and to two sites on chromosome 9 (**Figure 43**), indicating a more complex SV. Analysis both visually and bioinformatically of the discrepant paired-end reads crossing the three breakpoints enabled precise mapping of the breakpoints. This was consistent with a paracentric inversion of chromosome 9, followed by the known balanced translocation *t*(1,9) as depicted in the ideograms (**Figure 43 A and C**).

Reciprocal translocation t(1,9) in FID 356 (likely benign)

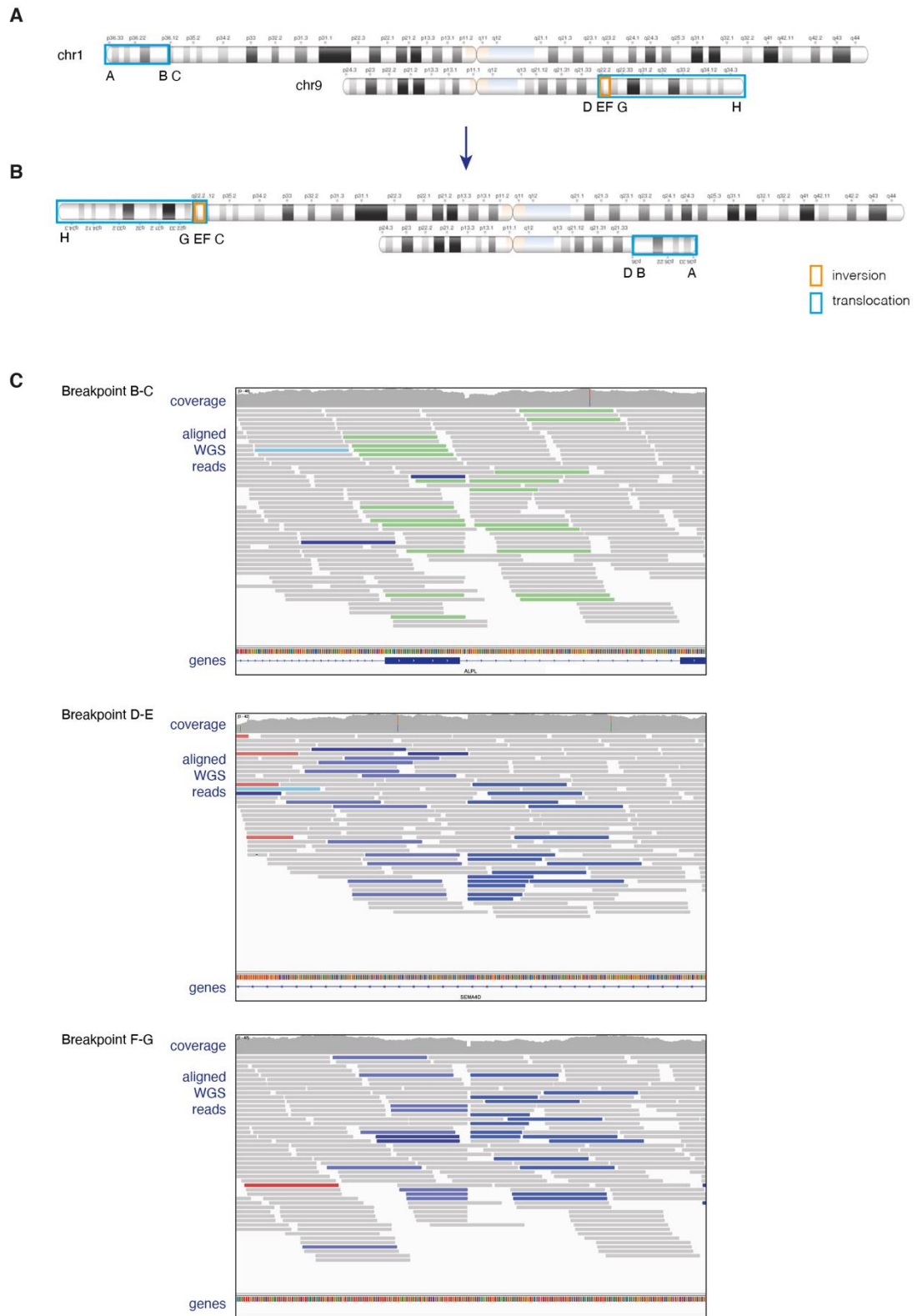


Figure 43 *De novo* reciprocal chromosomal translocation t(1,9) in the proband of trio FID 356. (A) Ideogram before translocation showing intact chromosomes 1 and 9 with the areas of subsequent structural rearrangement marked out: 4.5 Mb paracentric inversion of chromosome 9 (orange) and t(1,9) translocation (blue). Each end of the rearranged segments are marked with

a letter A-H for orientation purposes below. (B) Ideogram after structural rearrangement, showing the 4.5 Mb paracentric inversion of chromosome 9 in orange, and balanced reciprocal chromosome translocation between chromosomes 1 and 9 in blue. The letters at the end of each segment correspond to those above. (C) WGS data viewed in IGV (alignments coloured by insert size and pair orientation) from individual 356_356, the proband from FID 356. This shows aligned reads over each of the 3 breakpoints of the SV, the letters corresponding to those in the ideogram above. Data from the parents is not shown but confirmed that the translocation was *de novo*. The genomic sites of the breakpoint lie in the centre of the panel for each of the 3 breakpoints (blue arrow of top panel). This is a heterozygous SV with approximately half the reads either side of the breakpoint coloured to indicate that they have paired mate mapping to a different chromosomal location. Breakpoint B-C: chromosome 1 breakpoint, at chr1:21,560,758 (p36.12), in the *ALPL* gene, with the green-coloured reads either side of the breakpoint denoting paired mates mapping to two different areas of chromosome 9. The chromosome 9 breakpoints are labelled D-E, at chr9:89,372,728 (q22.1), and F-G, at chr9:93,847,511 (q22.31). The light purple reads on the left of each of these two panels indicates that the paired mate maps to chromosome 1; the blue reads on the right indicate a difference in pair orientation consistent with the chromosome 9 inversion.

Potential effects of this translocation were predicted to include:

- A gene fusion between *ALPL* and *SEMA4D*
- Disruption of both of these genes
- Disruption of a possible critical enhancer region of *BARX1*

BARX1 was the most immediately relevant candidate for further attention as it is known to be expressed in the developing eye and specifically iris (305). It encodes a homeodomain-containing transcription factor initially described as implicated in the *Drosophila* bar eye phenotype (306). The translocation breakpoint here falls 100 kb proximal (upstream) to the 5' end of *BARX1*.

Despite initial hope that this translocation might lead to the discovery of a novel aniridia-associated gene, closer inspection of WGS data revealed a small, previously undetected complex structural variant that in fact directly involves *PAX6*, described in the next section. The reciprocal t(1,9) translocation was therefore considered likely benign with no clinical impact suspected for any of the three breakpoints.

4.3.2.3.2 Identification of a second and causative complex structural variant (inversion/deletion) affecting *PAX6*

Examination of trio WGS data for FID 356 indicated a *de novo* 6 kb heterozygous inversion in the proband encompassing the start of *PAX6* up to intron 4, with the adjacent 30 kb of DNA upstream of *PAX6* deleted (**Figure 44 A**). This would be

predicted to abrogate *PAX6* function in this allele and therefore trump previous speculation regarding the t(1,9) translocation possibly being causative of the aniridia.

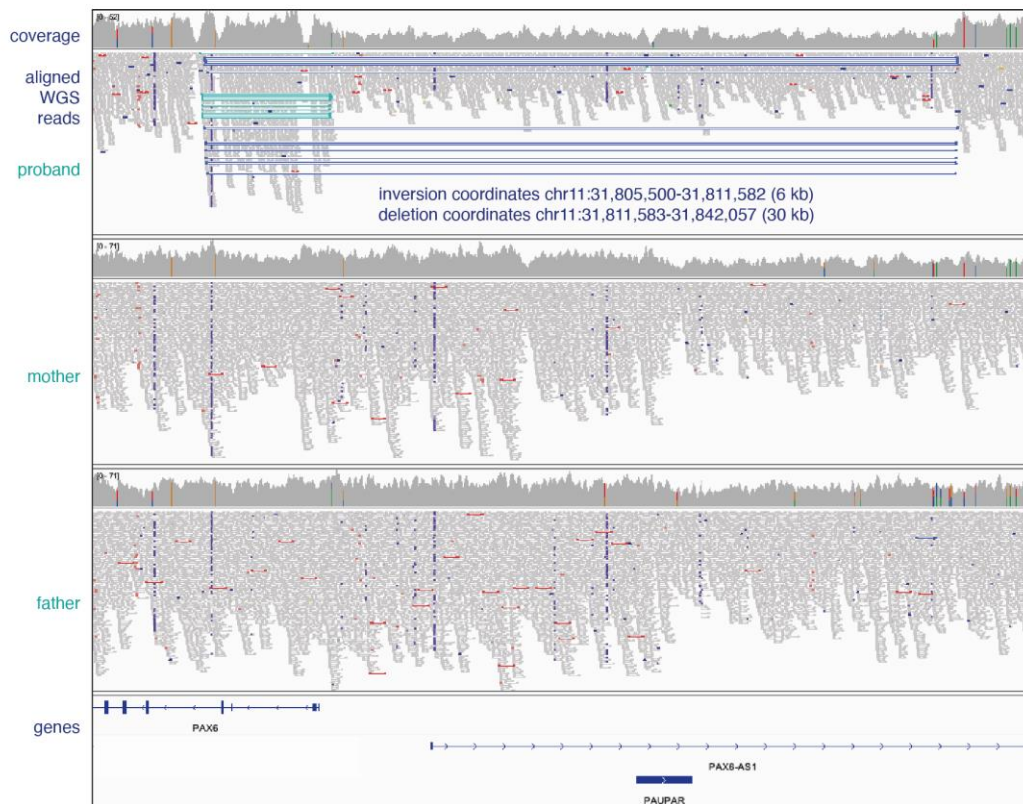
4.3.2.3.3 A tandem duplication and deletion affecting PAX6

A complex structural variant was identified in one further proband (FID 1271), an individual with bilateral aniridia and cataracts, this time involving the 3' end of the *PAX6* gene. WGS data indicates a 13 kb tandem duplication and nearby 16 kb deletion (**Figure 44 B**) encompassing the final 6 exons of *PAX6*.

Complex structural rearrangements of chromosome 11 detected on WGS

11p13

A FID 356: inversion and deletion



B FID 1271: tandem duplication and deletion

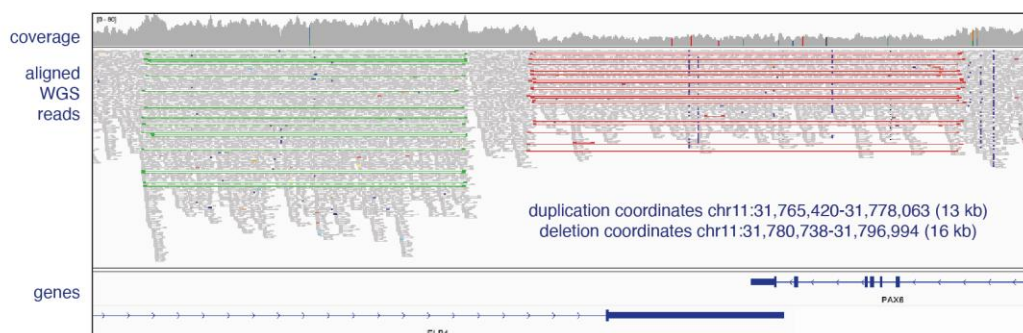


Figure 44 Complex structural rearrangements of chromosome 11 in two individuals with bilateral aniridia and cataracts. Aligned WGS data viewed in IGV, with reads viewed as pairs and coloured both by insert size and by pair orientation. Breakpoint coordinates (GRCh38) are visualised/estimated from WGS data. (A) FID 356: trio comprising child with sporadic bilateral aniridia and congenital cataracts, and unaffected parents. The child's WGS data indicates a 6 kb inversion (chr11:31,805,500-31,811,582, teal) involving the P1 promotor, all of the 3'UTR and the first coding exon (exon 4) of *PAX6*; next to this a 30 kb deletion (chr11:31,811,583-31,842,057, navy), which is immediately upstream of *PAX6* and deletes the P0 promotor and several enhancers including EE. The blue and teal colours both flag up paired reads with abnormal pair orientation. In IGV, pair orientation is determined first and only if this is as expected will abnormal insert size then be flagged. Therefore the reads across the deletion are not flagged in

red (as they would be in a simple deletion, due to the abnormal insert size between paired reads) as they also span the inversion, and so are coloured in teal due to the abnormal pair orientation. However, the drop in coverage depth clearly shows the deleted area. (B) FID 1271: a 16 kb deletion involving the 6 last exons of *PAX6* (chr11:31,780,738-31,796,994, red) and a 13 kb tandem duplication (chr11:31,765,420-31,778,063, green) affecting the final intron of *ELP4*. Reads with abnormal pair orientation in a pattern that suggests a tandem duplication flag up in green on IGV, whilst those spanning a deletion are flagged in red to show the abnormal insert size between pairs. In further support of this, the coverage depth is increased over the putative duplication and decreased over the putative deletion. The duplication coordinates do not directly overlap with critical *PAX6* enhancers.

4.3.2.4 Comparison with bioinformatic structural variant callers

I compared direct visual SV identification with the results of genome-wide bioinformatic SV calling tools. Canvas (284) detected, as expected, all the deletions >10 kb in size affecting the *PAX6* and *FOXC1* loci (**Figure 45**). It did not detect the two smallest deletions, affecting *PAX6*, 126 bp and 1.36 kb in size (FID 75 and 1524, respectively). It also called the more complex deletion-duplication case (FID 1271). As a CNV caller based on read depth data, it was unable to detect the inversions.

By contrast, Manta (285) was able to detect all the SV in the cohort including balanced SVs, but with more noise such that clean calls are difficult even when focussed to loci of interest. It detected many overlapping SVs, with each small-medium sized SV overlapping with several much larger potentially-spurious inversions/duplications (example of FID 723 in **Figure 46 A and B**). Conversely, larger SVs would overlap with a great number of shorter calls, as can be seen on the whole chromosome 11-level view in **Figure 46 C**.

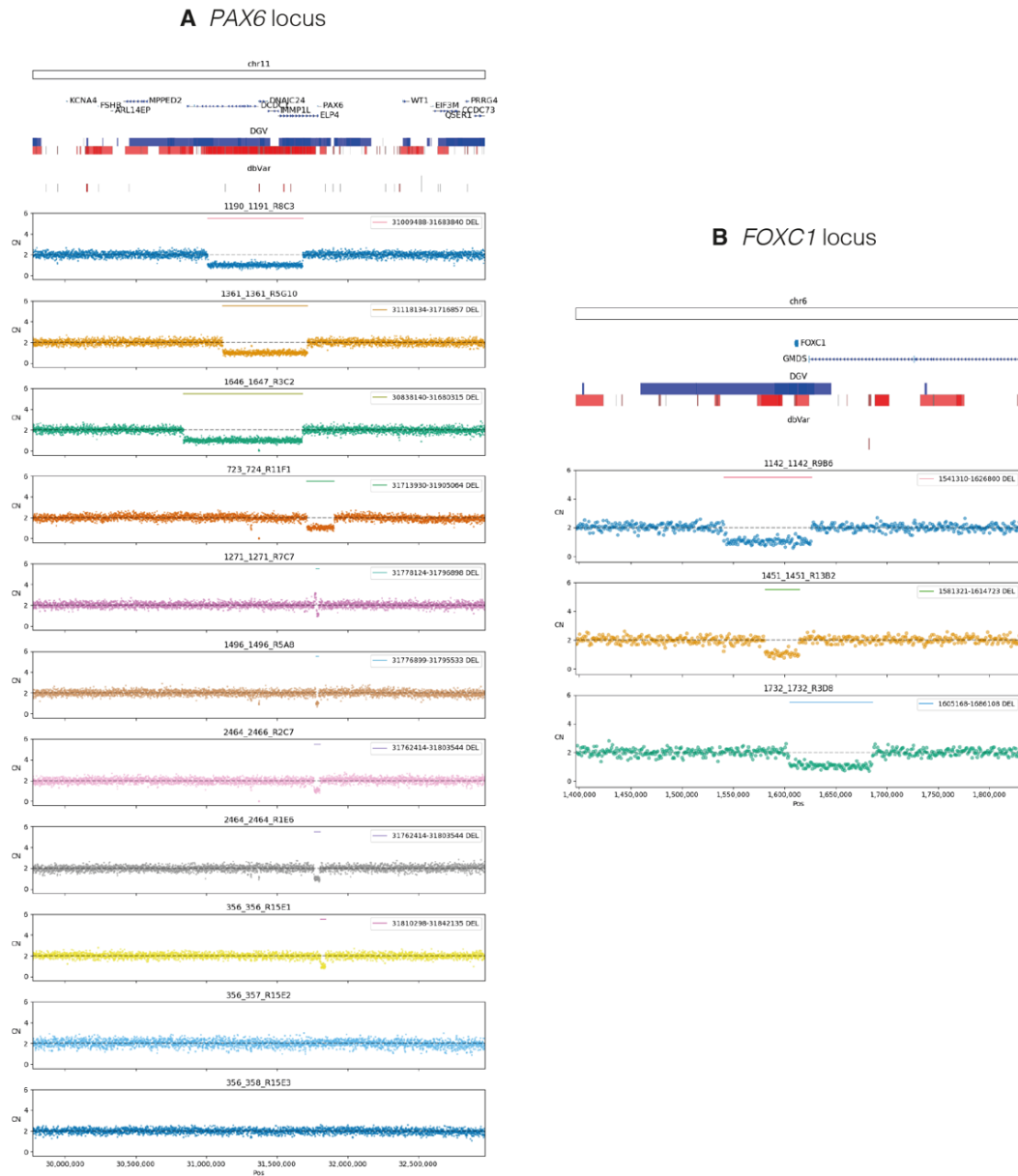
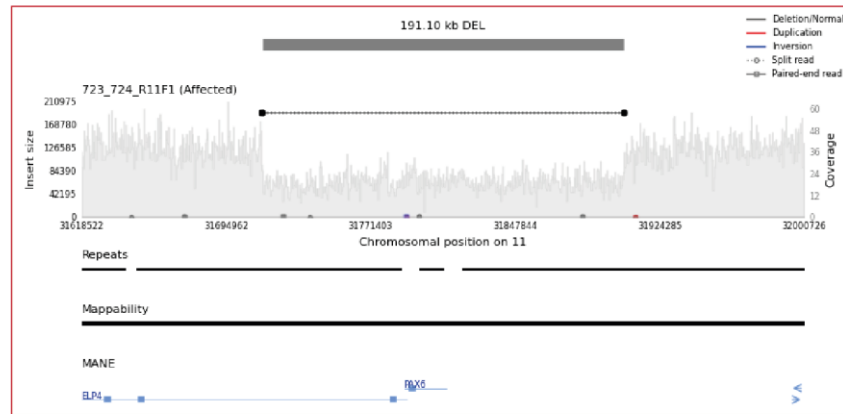


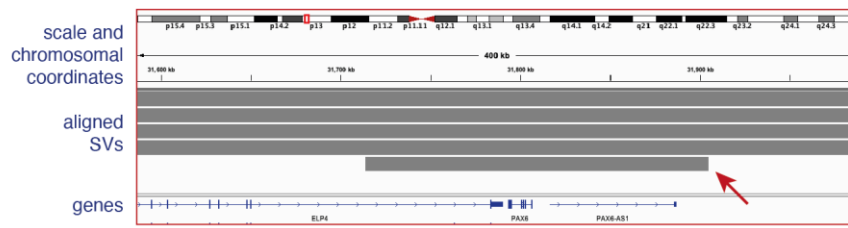
Figure 45 Bioinformatic CNV calling with Canvas across the (A) *PAX6* chromosome 11 and (B) *FOXC1* (chromosome 6) loci. These samplots figures kindly provided by David Parry show intersection of Canvas CNV calls for these two loci. Chromosomal coordinates (GRCh38) are at the bottom, with individual variants shown above. Annotations are shown from two databases of structural variation (307): the Database of Genomic Variants (DGV), containing variants seen in healthy individuals, are shown, with blue (upper DGV track) showing copy number gain and red (lower track) showing CN loss; and dbVar (specifically, the nstd186, or NCBI Curated Common SVs study), which includes data from gnomAD Structural Variants, 1000 Genomes SV, and DECIPHER Common CNVs. (A) The trio FID 356, duo FID 2464, and above that the single probands (FIDs 1496, 1271, 723, 1646, 1361 and 1190), all show reduced copy number (CN) consistent with a deletion, and 1271 additionally has an adjacent area of increased CN, consistent with duplication. (B) Three individuals with reduced CN over the *FOXC1* locus (FIDs 1142, 1451, 1732).

Structural variant calling with Manta:
example of FID 723, *PAX6* whole gene deletion (191 kb)

A



B



C

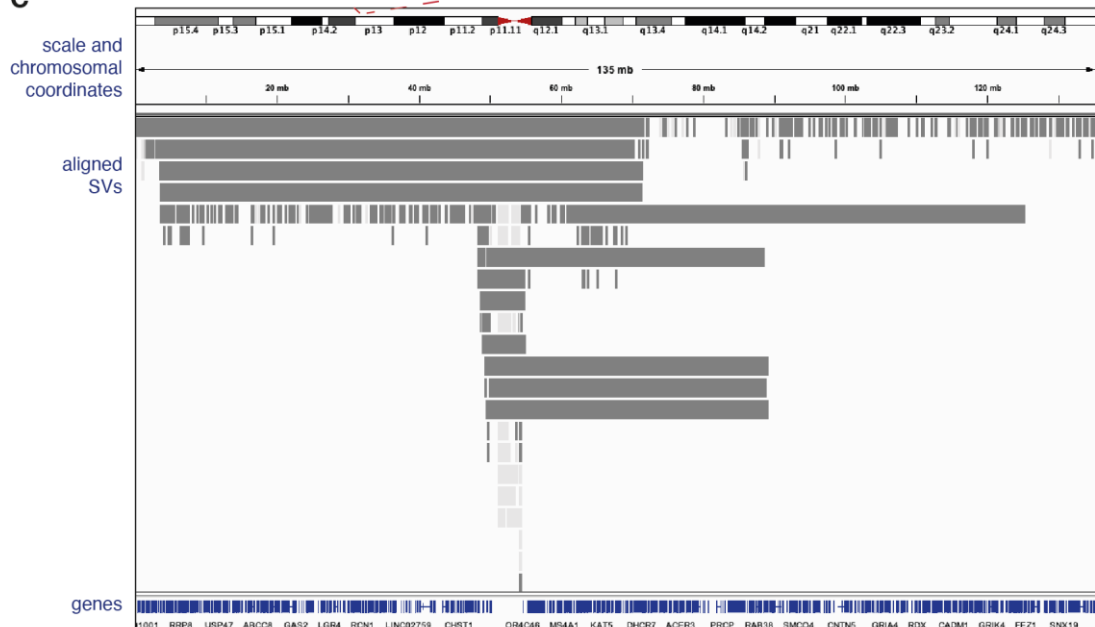


Figure 46 Bioinformatic SV calling with Manta: example of FID 723, who has a 191 kb deletion encompassing the *PAX6* gene. (A) Figure generated with samplots showing the clean call of the deletion, achieved by limiting the results to a narrow region of the genome. A plot of insert size aligned to the chromosomal position (GRCh38) is shown, with the gene track below showing the MANE select

transcripts. (B) The Manta call of the same deletion (red arrow) seen with all the SV calls overlapping that region, shown on IGV over a 400 kb segment of chromosome 11. (C) A 135 mb, whole chromosome view in IGV of all the Manta calls detected on chromosome 11 for this patient, including those filtered out (lighter grey). This illustrates the level of noise that would affect an 11p variant.

4.3.3 Sequence variants identified in the WGS aniridia cohort

WGS VCF files for all 51 individuals were filtered using the Eye gene2phenotype dataset (289) with the VEP-G2P plugin (290) to identify high and moderate impact changes in genes known to cause genetic eye disease. The likely causative sequence variants detected, all in *PAX6*, are summarised in **Table 13** along with American College of Medical Genetics and Genomics pathogenicity classifications. This is the second of the two results tables for the variants detected on WGS, the first covering structural variants, above (**Table 2**). I will then discuss these variants and the experimental analysis of those affecting splicing. Lastly, I will discuss the *de novo* trio analysis results.

FID	PID	Inheritance	Intron	GRCh38	CDS variant	Predicted consequence	ACMG/ACGS	Novel
Coding likely gene disruptive variants								
2134 (trio)	2134	unknown*		chr11:31,793,725_31,793,726insAC	c.842_843ins GT **	p.(Pro282 Tyrfs*84)	P (0.999) ; PM2, PVS1, PP4 mod	yes
Essential splice site variants affecting 5' non-coding exons								
877	877	unknown	IVS3+1	chr11:31,806,848C>G	c.-52+1G>C	p.(?) donor loss	LP (0.9) ; PM2, PS1 supp, PP3, PP4 mod – see note (1)	yes; see note (2)
1019	1019	unknown	IVS3+1	chr11:31,806,848C>A	c.-52+1G>T	p.(?) donor loss	LP (0.949) ; PM2, PS4 supp, PP3, PS1 supp, PP4 mod (1)	no (308)
1500	1500	unknown	IVS2-2	chr11:31,806,927del	c.-128-2del	p.(?) acceptor loss	LP (0.949) ; PM2, PS4 mod., PP3, PP4 mod	no (10,144,309,310) VCV000430969.2
5645 (trio)	5645	de novo	IVS2-2	chr11:31,806,927del	c.-128-2del	p.(?) acceptor loss	LP (0.997) ; PM2, PS4 mod, PP3, PS2, PP4 mod	as above
Deep intronic variants affecting splicing								
1635 (trio)	1635	de novo	IVS8+68	chr11:31,794,562G>C	c.682+68C>G ***	p.(?) donor gain	LP (0.988) ; PM2, PP3, PS2, PP4 mod	yes
3612	3612	unknown	IVS6+334	chr11:31,801,227C>T	c.357+334G>A ****	p.(?) donor gain	LP (0.9) ; PM2, PS4 supp., PP3, PP4 mod.	no (uncertain pathogenicity (10)) VCV000559621.2

Abbreviations: FID/PID, family/personal ID; CDS, coding sequence; ACMG/ACGS, American College of Medical Genetics/Association for Clinical Genomic Science pathogenicity classification (posterior probability); LP, likely pathogenic; P, pathogenic; for criteria (PM2 etc) see (294,311); mod, moderate; supp, supporting; IVS, intervening sequence.

Notes: *paternal WGS sample invalid therefore trio incomplete;

The c. coordinates are the same for the MANE transcript NM_001368894.2 (+5a, 14 exons) as for the more biologically relevant NM_000280.4 (~5a, 13 exons) used here, apart from: **2134: NM_001368894.2:c.884_885insGT; ***1635, NM_001368894.2:c.724+68G>C; ****3612, NM_001368894.2:c.399+334G>A.

(1) Arguable whether PVS1 could be applied to this canonical splice site in a non-coding exon (this would only further increase the pathogenicity prediction).

(2) These c.-52+1 variants appear to be novel; a different substitution of the same nucleotide, G>A, has been published (312), is in ClinVar RCV000984349.1, and appears 3 times in the PAX6 LOVD database. For non-novel variants, the ClinVar reference is given

Table 13 Likely pathogenic PAX6 sequence variants detected in the aniridia cohort (NM_000280.4)

4.3.3.1 Sequence variants affecting the *PAX6* coding exons

I identified one individual with a likely gene disruptive variant affecting a coding exon of *PAX6* (**Figure 47, Table 13**), in an individual with bilateral aniridia, nystagmus and glaucoma, and unaffected parents (trio FID 2134). The variant identified was NM_000280.4(*PAX6*):c.842_843insGT, a 2 bp insertion causing a frame shift and premature stop, p.(Pro282Tyrfs*84). This constitutes the only “false negative” of the *PAX6* screening, which was performed many years ago by denaturing HPLC analysis.

The family was included in the WGS cohort as a trio, but WGS results led to the discovery of historic error in the sample ID of the father (2134-2137):

WGS analysis showed a kinship coefficient of 0.4991 for the WGS data from parental samples 2134-2137 and 2134-2138, indicating that the samples must be duplicates. No Y chromosome data was present in either sample suggested a duplicate of the maternal sample. All three stock DNA samples were sent for STR profiling using the GenePrint 10 system (Promega), which was run by Jeffrey Joseph of the HGU core services team. This confirmed that the stock DNA held for both parents was the same at each of the 10 STR loci (**Table 14**), and excluded a plating error at the time of WGS panel preparation. It was not possible to obtain a fresh DNA sample.

We were thus unable to confirm whether the variant detected was *de novo*. The joint genotyping for the whole cohort was rerun without the duplicate sample.

Sequence variant detected on WGS: 2 bp insertion *PAX6* exon 10 causing frameshift stop

11p13 

FID 2134

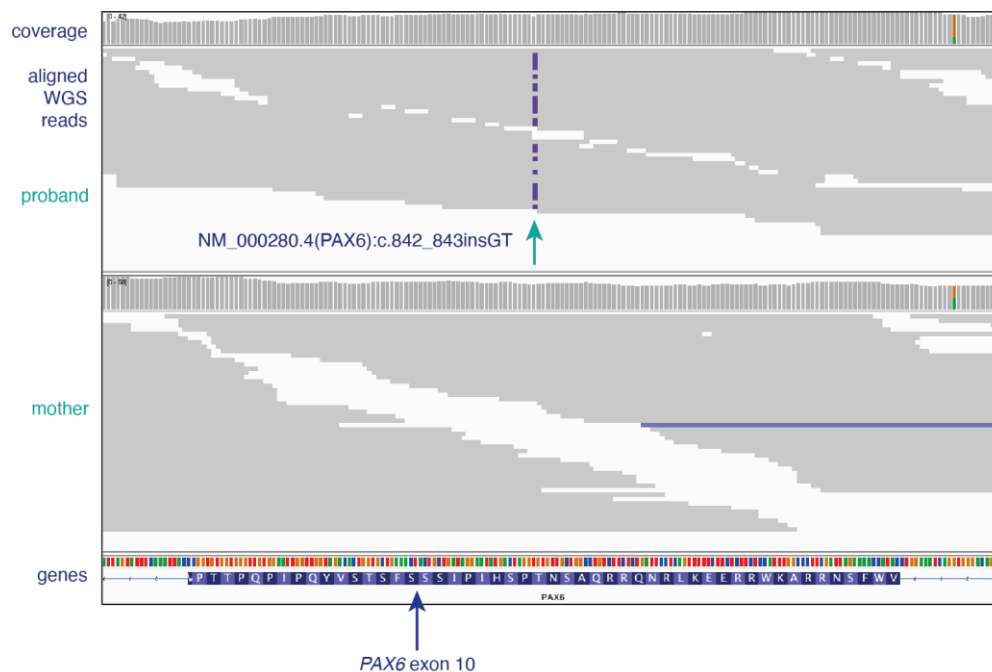


Figure 47 Aligned WGS data for family 2134 showing a likely gene-disruptive sequence variant affecting exon 10 of *PAX6* in the proband. The proband has sporadic bilateral aniridia, nystagmus and glaucoma. The IGV snapshot shows aligned WGS data from proband (top panel) demonstrating a 2 bp insertion (arrow) in exon 10 of *PAX6* (at chr11:31,793,725, NM_000280.4:c.842_843insGT), and below that of the unaffected mother not carrying the variant.

Sample Name	Locus	Allele 1	Allele 2	Size 1	Size 2
2134_2134_R45E2 Proband sample	TH01	6	9	161.32	173.26
	D21S11	29	32	218.90	232.94
	D5S818	11	12	130.16	134.21
	D13S317	11		187.35	
	D7S820	11		232.03	
	D16S539	9		278.70	
	CSF1PO	11	12	337.24	341.33
	AMEL	X	Y	104.10	109.92
	vWA	14	18	138.56	154.36
	TPOX	8	11	268.76	280.75
2134_2137_R2A10 Parental sample	TH01	9	10	173.30	177.30
	D21S11	29	32	218.90	230.99
	D5S818	11	12	130.20	134.23
	D13S317	11		187.41	
	D7S820	10	11	228.09	232.08
	D16S539	9		278.61	
	CSF1PO	12		341.15	
	AMEL	X		104.20	
	vWA	15	18	142.39	154.36
	TPOX	11		280.75	
2134_2138_R45F3 Parental sample	TH01	9	10	154.43	173.27
	D21S11	29	32	218.95	231.03
	D5S818	11	12	130.16	134.21
	D13S317	11		187.34	
	D7S820	10	11	228.02	232.14
	D16S539	9		278.64	
	CSF1PO	12		341.18	
	AMEL	X		104.13	
	vWA	15	18	142.40	154.43
	TPOX	11		280.80	

Note: Short tandem repeat (STR) profile for 10 loci run using the GenePrint10 System (Promega). Allele binning based on size performed in GeneMapper 4.0. Where a single allele is listed, this is in the homozygous state. Each allele number is coloured for visualisation purposes. The sample name is in the format FID_PID_RackPosition (internal DNA sample location).

Table 14 Short tandem repeat profiling of family 2134. This confirms that the two parental samples were the same at the 10 STR loci tested, and indicating (as suspected on WGS) that the “paternal” sample was in fact a duplicate of the maternal sample

4.3.3.2 Splicing analysis of the *PAX6* locus

4.3.3.2.1 Essential splice site variants

I identified heterozygous variants in four families affecting the essential splice sites (ESS) in the *PAX6* 5' UTR; three single probands and one trio (**Figure 48, Table 13**). These variants all flank exon 3 of *PAX6*, the last fully non-coding exon as the *PAX6* translation start site is in exon 4. The 5' UTR was not always screened on Sanger sequencing historically, hence why such “*PAX6* negative” cases have been found to persist in the cohort.

The two heterozygous SNVs, FID 877 and 1019, affect the first nucleotide at the start of intron 3 (IVS3+1, or c.-52+1), G>C and G>T respectively (**Figure 48 A**). The latter (G>T) has been published (308), as has IVS3+1 G>A (312,313). Changing this nucleotide was shown on reverse transcription (RT)-PCR to result in skipping of exons 3, 4, 5 and 5a (312). This is in line with bioinformatic splicing predictions (**Table 15**). Although the published functional work above is for the G>A allele, the splice prediction (abolishment of the splice donor site) is the same as for changing the nucleotide G>C and G>T.

The third variant, NM_000280.4(*PAX6*):c.-128-2del (IVS2-2), a heterozygous single base pair deletion, was found in both the single proband FID 1500 (**Figure 48 A**) and *de novo* in the proband of the FID 5645 trio (affected proband and unaffected parents; **Figure 48 B**). These families are not known to be related, and it is a variant that has been previously published (10,144,309,310). SpliceAI predicts acceptor loss at -1 bp which is in agreement with all 4 of the Alamut predictors which also predict abolishment of the natural splice acceptor site (**Table 15**).

These cases are therefore predicted to have straightforward heterozygous loss of *PAX6*, in keeping with the classical aniridia phenotype.

CDS variant (PAX6 NM_000280.4)	FID(s)	Region affected	Analysis range (Alamut)	Splice site	Alamut SSF [0-100]	Alamut MaxEnt [0-16]	Alamut NNSPLICE [0-1]	Alamut GeneSplicer [0-21]	Splicing prediction	SpliceAI delta score [0-1]
c.-52+1G>C	877	essential splice site (IVS3+1)	c.-128-84 (intron 2) to c.-52+85 (intron 3)	c.-52 N	100.00 → -				donor loss	0.58
c.-52+1G>T	1019	essential splice site (IVS3+1)	c.-128-45 (intron 2) to c.-52+124 (intron 3)	c.-52 N	100.00 → -				donor loss	0.58
c.-128-2del	1500 + 5645	essential splice site (IVS2-2)	c.-128-124 (intron 2) to c.52+45 (intron 3)	c.-128 N	86.66 → -	10.74 → -	0.99 → -	11.42 → -	acceptor loss	0.65
c.682+68C>G	1635	deep intronic (IVS8+68)	c.644 (exon 8) to c.682+175 (intron 8)	c.682+67	- → 84.79	- → 8.39	- → 0.93	- → 2.09	donor gain	0.97
c.357+334G>A	3612	deep intronic (IVS6+334)	c.357+212 to c.358-248 (intron 6)	c.357+331	- → 71.80	1.32 → 7.33 (+456.7%)	- → 0.80		donor gain	0.46

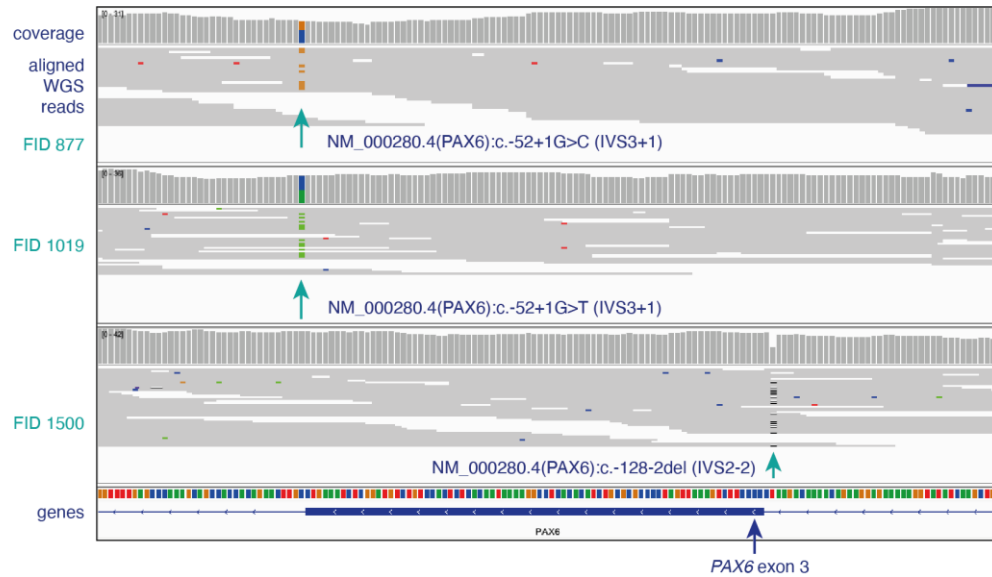
CDS, coding sequence; FID, family IDr; N, natural splice site; SSF, splice site finder
Thresholds: SSF ≥ 70, MaxEnt ≥ 0, NNSPLICE ≥ 0.4, GeneSplicer ≥ 0

Table 15 Predicted effects on splicing of PAX6 sequence variants identified in the WGS cohort

Single nucleotide variants detected on WGS involving *PAX6* 5' UTR essential splice sites

11p13 

A Single probands FIDs 877, 1019, 1500



B FID 5645 trio

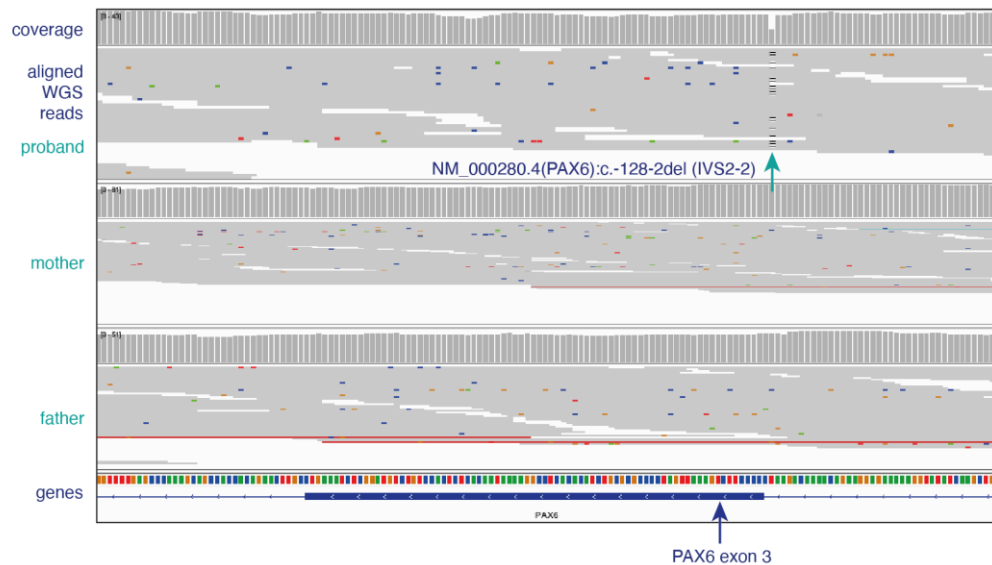


Figure 48 Single nucleotide variants affecting essential splice sites of *PAX6*. Aligned WGS data (shown in IGV) showing 3 heterozygous variants in 4 families disrupting the essential splice sites flanking exon 3. Mismatched bases are coloured (A, green; C, blue; G, orange; T, red) and, when greater than 20% reads for a nucleotide are affected, the coverage track bars are coloured in proportion to the read count. Thus the heterozygous SNVs here in panel A are easily visible on the coverage track. Equally, the heterozygous single base pair deletions in lower panel A and panel B are visible as a black dash in the alignments, and a corresponding ~50% drop in coverage depth above it. (A) Single probands: FID 877, who has bilateral aniridia, glaucoma and bilateral cataracts in young adulthood, has the heterozygous substitution Chr11(GRCh38):g.31806848C>G, which as *PAX6* is on the reverse strand is

NM_000280.4(PAX6):c.-52+1G>C (IVS3+1). This is visible at the marked arrow as a blue and orange bar on the coverage track. FID 1019, an individual with sporadic aniridia, has a different substitution of the same nucleotide Chr11(GRCh38):g.31806848C>A, which is NM_000280.4(PAX6):c.-52+1G>T (IVS3+1), visible at the marked arrow as a blue and green bar on the coverage track. FID 1500 has isolated bilateral aniridia, posterior subcapsular cataracts, absent foveal reflexes, subtle peripheral corneal changes and relatively good vision (6/12 N5 right, 6/36 N5); they were found to have Chr11(GRCh38):g.31806927del, which is NM_000280.4(PAX6):c.-128-2del (IVS2-2), a single base pair deletion affected the -2 position of intron 2. (B) Family 5645, a trio of child affected by bilateral aniridia and unaffected parents, carries exactly the same variant as FID 1500. WGS data for the proband and parents are shown together and the deletion is arrowed.

4.3.3.2.2 Deep intronic splice variants

Two unrelated affected individuals with aniridia were identified as having deep intronic variants in *PAX6*, highlighting one of the strengths of WGS.

Firstly, an individual with sporadic bilateral aniridia, with relatively well-preserved foveal regions and vision (FID 3612) was found to have the variant NM_000280.4(PAX6):c.357+334G>A. This is deep at the +334 position in intron 6 (**Figure 49 A, Table 13**).

This variant was detected both by the variant filtering pipeline and by the dedicated splice tool SpliceAI, which was run on the cohort by David Parry. Although the variant filtering pipeline only reports essential splice and splice region variants, it detected this variant because there exist three alternative transcripts of *PAX6* (variants 10, 23 and 26 on NCBI, all encoding the paired-less isoform d) whose preceding exon extends into that intronic space (which is otherwise maintained as intron across the other 50 *PAX6* transcripts listed in RefSeq). For these 3 transcripts, this variant is in the +3 position and therefore within the splice region. The exon 6 in these 3 variants is non-coding and one of them is included for illustrative purposes in **Figure 49 A**. Without this quirk of filtering, the variant would only have been detected by SpliceAI; this predicted a donor gain caused by the variant 3 bp away from this site (for all the common transcripts including -5a and +5a). The predictors within Alamut, SSF, MaxEnt and NNSPLICE all agree with this and strongly predicted a donor gain (**Table 15**). Similar to the group who first reported this variant as a VUS (10), we were unable to do functional work to confirm missplicing, but a more recent paper has demonstrated missplicing *in vitro* in a minigene assay (314)

For the second case, SpliceAI predicted with 0.97 confidence a donor gain caused by a *de novo* deep intronic variant in the proband with variant aniridia of trio FID 1635. The phenotype is of a congenital anterior chamber abnormality very much on the classical aniridia spectrum, with inferior iris defects, vascularisation of peripheral cornea, inferior lens opacities and foveal hypoplasia, nystagmus, and poor vision. The variant, chr11(NM_000280.4(PAX6):c.682+68C>G), confirmed on WGS as *de novo* (**Figure 49 B**), was at the +68 position in intron 8, with the donor gain predicted to be 1 bp away from it. All four splice prediction tools in the Alamut visual splicing predictions agreed with the splicing prediction (**Table 15**).

Experimental functional work was performed for FID 1635, for whom a lymphoblastoid cell line was available, described in the following section.

Figure 49 (right) Deep intronic splice variants in *PAX6* identified on WGS. This shows aligned WGS data from 2 families. In the coverage track, the heterozygous variants show up as a bicoloured vertical bar: mismatched bases are coloured (A, green; C, blue; G, orange; T, red) and the coverage track bars are coloured in proportion to the read count. The mismatched bases can also be seen on the aligned reads below.

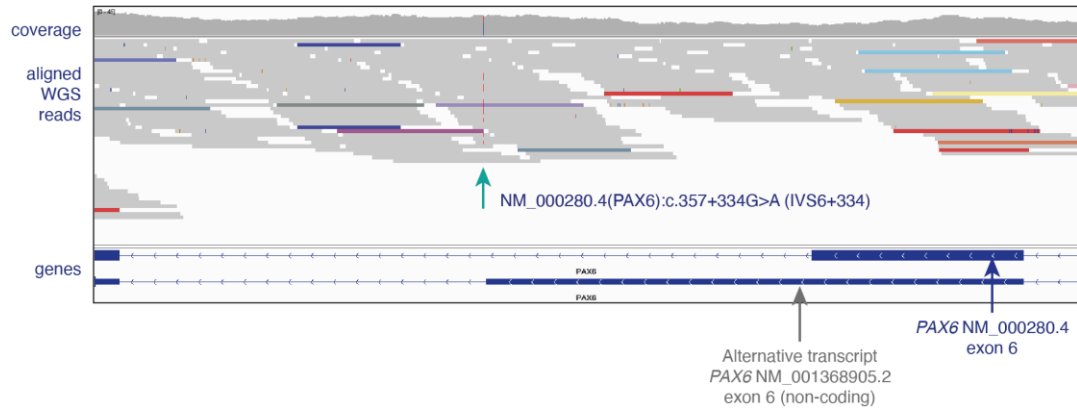
(A) FID 3612, an individual with sporadic aniridia and relatively preserved foveas, was found to have Chr11(GRCh38):g.31801227C>T, which is NM_000280.4(*PAX6*):c.357+334G>A, a deep intronic variant in intron 6. A 1 kb span of aligned WGS data is shown, with an arrow indicating the heterozygous SNV within intron 6, shown as a blue-red vertical bar on the coverage track. Its position relative to a rarer alternative *PAX6* transcript (encoding a paired less isoform d) is also shown, which has a smaller intron 6; the variant is in a +3 position here which meant that the variant was detected both on SpliceAI and on the standard filtering pipeline which includes only splice regions. The strong donor gain prediction seen in 3/4 Alamut predictors (**Table 15**) is consistent with the donor gain prediction (0.46 confidence) of SpliceAI at the same position.

(B) FID 1635 (trio): the proband, a child with a sporadic, classical aniridia spectrum eye malformation, was found to carry Chr11(GRCh38):g.31794562G>C, which is NM_000280.4(*PAX6*):c.682+68C>G. A 330 bp span of aligned WGS data is shown with the *de novo* heterozygous SNV within intron 8 is indicated with an arrow and is indicated by a blue-orange vertical bar on the coverage track. The unanimous prediction of a donor gain from the Alamut splice predictors (**Table 15**) was consistent with the 0.97 confidence donor gain prediction of SpliceAI. Experimental confirmation of missplicing for this variant is described in the next section.

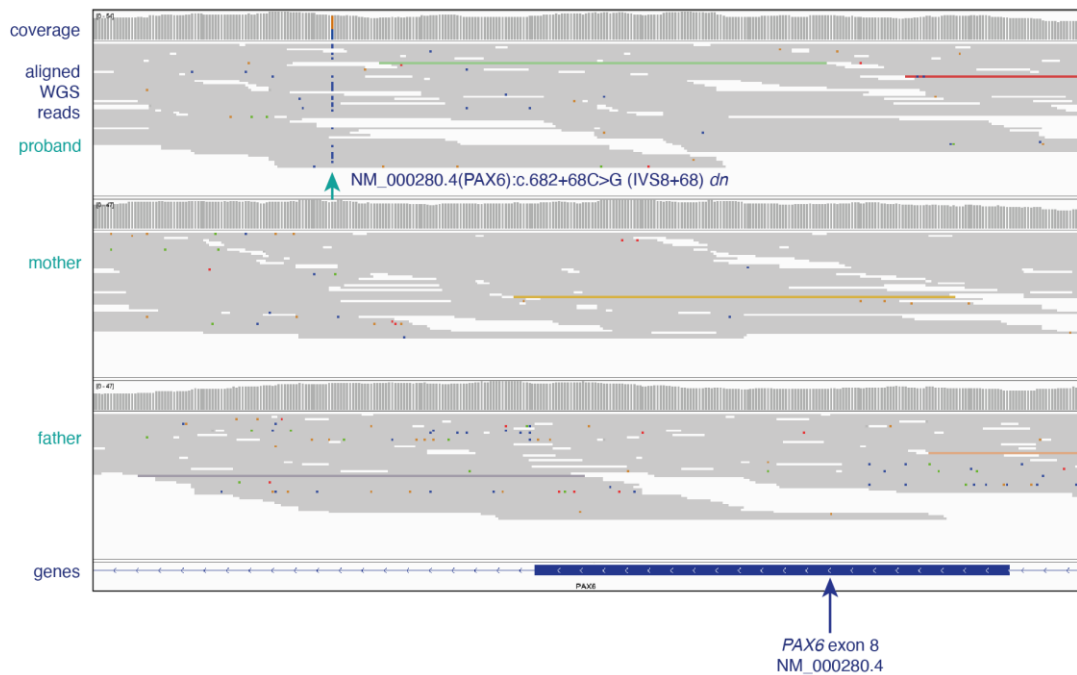
PAX6 deep intronic splice variants detected on WGS



A Single proband FID 3612



B FID 1635 trio



4.3.3.2.3 In vitro validation of selected splicing variants

Patient-derived lymphoblastoid cell lines (LCL) were available for a small proportion of the WGS cohort (the more historical rather than recent additions to the cohort), including one of the probands with a deep intronic splicing variant, FID 1635. In order to test the hypothesis that these variants disrupted splicing *in vitro*, LCLs were recovered from liquid nitrogen for further investigation.

In addition to the deep intronic case, FID 1635, three further cases were selected for further investigation: FID 1500, carrying an essential splice site variant (NM_000280.4(PAX6):c.-128-2del (IVS2-2)); a historical case carrying a different *PAX6* deep intron 8 variant previously shown to affect splicing (FID 1292, NM_000280.4:c.683-113G>T); and an unaffected (unrelated) individual as a control.

LCLs were grown up and RNA was successfully extracted from all 4 cell lines. cDNA was obtained from total RNA and a nested PCR was performed using four primer pairs spanning different overlapping sections of the *PAX6* gene (**Figure 50 A**). *PAX6* plasmid cDNA (courtesy of Hemant Bengani) containing the coding exons only was used as an additional control.

The nested PCR confirmed missplicing of the *PAX6* gene *in vitro* in a pattern consistent with variants (**Figure 50 B**):

FID 1500, with an essential splice site variant at the end of intron 2, showed evidence of abnormal splicing between exons 1 and 5.

We had wondered whether the deep intronic intron 8 control case, FID 1292, might show a similar pattern of splicing to FID 1635. However, whilst this case showed evidence of abnormal splicing *in vitro*, this was not in the same pattern as our proband of interest, FID 1635.

The case of interest, FID 1635, showed evidence of missplicing using the primers spanning both exons 3-8 and 7-12. The former (exon 3-8) was not necessarily predicted; Sanger sequencing of the gel-purified product appeared to show skipping of exon 5 and part of exon 6. The missplicing between exons 7 and 12 was fully expected and Sanger sequencing of the gel-purified product confirmed skipping of exons 9, 10 and 11 (**Figure 50 C**).

Figure 50 *In vitro* confirmation of missplicing in deep intronic splice variant case FID 1635. (A) Locations of the four nested primer pairs used to assess overlapping sections of the *PAX6* locus; exonic locations within *PAX6* depicted by double arrowheads: exons 1 to 5 ('green' pair), exons 3 to 8 ('orange' pair), exons 7 to 12 ('yellow' pair) and exons 9 to 13 ('pink' pair; products not shown as no evidence of missplicing in any of the cases). (B) Agarose gel images showing the products of the nested PCR performed with the LCL-derived cDNA. Samples are labelled as FID_PID (family_personal ID). Gel B1, the green primer pair covering exons 1 to 5, showed evidence of missplicing only in the sample from FID 1500, who has an essential splice site variant in the 5'UTR, NM_000280.4:c.-128-2del. The plasmid *PAX6* cDNA only contained cDNA for the coding exons, which start at exon 4, so did not yield a product for the green or orange primers. Gel B2, the orange pair spanning exons 3 to 8, showed evidence of missplicing compared to controls in the deep intronic variant case, FID 1635. As the variant is in intron 8, this was not necessarily anticipated. Gel B3, the yellow primers covering exons 7 to 12, confirmed *in vitro* evidence of missplicing in both FID 1635 and in the deep intronic 'positive control' case FID 1292, with the smaller splicing products suggesting exon skipping. In each gel image, the bands outlined were gel purified and sent for Sanger sequencing. (C) Sanger sequencing of the gel-purified misspliced band from FID 1635 (gel B3, panel B) over the exon 8-9 junction confirmed skipping of exons 9, 10 and 11.

4.3.3.2.4 RNAseq analysis from LCLs showed insufficient coverage of splice junctions to detect splice variants

In collaboration with Dr Jamie Ellingford's group (University of Manchester), we wanted to see whether RNAseq was useful to detect splice variants. The RNA that I extracted in the section above from the 4 patient-derived LCLs was sent for mRNAseq, performed by the Wellcome Trust Clinical Research Facility, Edinburgh (project E07446), on Illumina NextSeq 550 with poly-A enrichment and paired end 75 bp reads.

I supplied RNAseq data and variant information for each sample and the data were analysed by Charlie Rowlands (PhD student, Jamie Ellingford's group).

The known variants were not observed in the individual LCL RNAseq data, which were sequenced with a depth of 30M reads per sample. This was due to poor coverage of splice junctions. Charlie Rowlands examined how well *PAX6* splice junctions were covered in control individuals using RNAseq data from different tissues (blood, n=150; LCL, n=91; muscle, n=184). Even though LCLs had the best coverage of the three tissues, across the *PAX6* splice junctions in question, the coverage was in the order of 0.1 reads per million, which at 30M read depth was at most three reads mapping to each splice junction.

However, my LCL sequencing and variant data was a useful contribution to the publication testing the predictions of the Ellingford group's minimum required sequencing depth (MRSD) tool (315), which estimates the sequencing depth required for RNAseq to detect missplicing of a gene of interest. The tool, available at <https://mcgm-mrsd.github.io>, predicts that *PAX6* splice junctions would not be sufficiently covered in any tissue (blood/fibroblasts/LCL/muscle) even at >100M read depth.

This emphasises the need for performing targeted PCRs such as those described above to detect missplicing in *PAX6*.

4.3.3.3 No plausibly causative variants identified in *PAX6* cis-regulatory elements

Intersection of a bed file with 35 known *PAX6* CREs and the annotated VCF files from the WGS did not reveal any plausibly causative variants. Further details of this analysis, of the 39 variants identified and why we do not consider them clinically significant are given in **Appendix C.3**.

4.3.3.4 *De novo* analysis of WGS trios

De novo analysis of the trio whole genomes (affected proband with unaffected parents) was performed, with Mike Halachev and David Parry using the bioinformatic tool VASE and then review of the variants by me for plausible pathogenicity.

One of the 6 trios, FID 2134, was excluded due to paternal WGS data being essentially identical to the mother's, due to a presumed historic sample ID error as discussed above in **4.3.3.1**. However, as a *PAX6* SNV was identified in the proband of this trio, the only impact of the absence of the father's data was being unable to confirm that the variant was *de novo*.

Table 16 shows the *de novo* variants discovered in the 5 complete trios (low impact variants excluded, leaving modifier, moderate and high impact variants). Between 1 and 5 such *de novo* variants were detected in each family, which is in line with the expected number of candidate variants of 3 (316).

In keeping with the Eye G2P variant filtering results described above and listed in **Table 13**, FID 1635 and 5645 returned *PAX6 de novo* variants predicted to disrupt splicing.

FID 356 and 75 trios did return a single *de novo* variant each: a frameshift variant in *CRYBG2* and a missense variant in *PRDM5*, respectively. However, from a disease causation perspective these were trumped by the structural variants detected in both of these cases affecting the *PAX6* locus (Table 12 above).

This leaves one trio.

4.3.3.4.1 *De novo missense variant of uncertain pathogenicity in AHCYL2, which interacts with a syndromic aniridia-associated gene*

FID 2469 trio returned more *de novo* results (1 high and 4 moderate impact variants); perhaps of relevance is that the parents' WGS data was from DNA extracted from a lymphoblastoid cell line, whereas the proband was not.

Amongst the 5 *de novo* variants detected, all listed in **Table 16**, is a missense variant in *AHCYL2* (*adenosylhomocysteinase-like 2*), p.(Leu352Met).

This allele, NM_015328.4(*AHCYL2*):c.1054C>A, is not present in gnomAD, and is predicted to be pathogenic/disruptive to normal gene function by multiple variant effect predictors. including: AlphaMissense 0.922; CPT 0.669; popEVE -3.881; SNPred 0.977; ESM-1b -11.189; CADD 24.3; REVEL 0.73. Based on ACMG/ACGS criteria, the pathogenicity prior probability is calculated as 0.5 (Uncertain), though I acknowledge this should not strictly be applied in this case as it is not a known disease locus. Therefore, based on this case alone, no claim can be made for the clinical significance of this variant. However, there is a reported interaction of *AHCYL2* with *ITPR1* (317), the cause of Gillespie syndrome, which merits some further discussion below.

FID	Gene	Region	GRCh38 coordinates	CDS variant	Predicted consequence	CADD/REVEL	ACMG/ACGS	Novel	Comment
75	<i>PRDM5</i>	exon 16/16	chr4:120695125C>T	NM_018699.4: c.1879G>A	p.(Val627Ile)	22.6		yes	known eye gene (1) but <i>de novo</i> <i>PAX6</i> SV identified
356	<i>CRYBG2</i>	exon 3/22	chr1:26354223_26354233del	ENST00000475866.3: c.729_739del	p.(Ala244Leufs*4)	n/a		yes	<i>de novo</i> <i>PAX6</i> SV identified
1635	<i>PAX6</i>	IVS8 +68	chr11:31794562G>C	NM_000280.4: c.682+68C>G	p.(?) / donor gain	19.83	P (0.994)	yes	disrupts splicing <i>in silico</i> & <i>in vitro</i> (2)
1635	<i>NYNRIN</i>	exon 9/9	chr14:24416718_24416719dup	NM_025081.3: c.4969_4970dup	p.(Ala1658Argfs*19)	34		no	trumped by <i>PAX6</i> variant above
2469	<i>AHCYL2</i>	exon 8/17	chr7:129405125 C>A	NM_015328.4:c.1054C>A	p.(Leu352Met)	25 / 0.73	[U (0.5)]*	yes	interacts with ITPR1
2469	<i>IGLV3-1</i>	exon 2/2	chr22:22881377 G>A	ENST00000390319.2:c.327G>A	p.(Trp109*)	35		no	encodes a Bence-Jones protein
2469	<i>ZBTB11</i>	exon 2/11	chr3:101671997 G>T	NM_014415.4: c.527C>A	p.(Ser176 Tyr)	25.8		yes	AR intellectual disability gene
2469	<i>H3C4</i>	exon 1/1	chr6:26197043G>T	NM_001376937.1:c.208C>A	p.(Arg70Ser)	23		yes	
2469	<i>ZNF74</i>	exon 5/5	chr22:20406879_20406880ins s GGGATCCCA	NM_003426.4:c.1846_1847ins GGGATCCCA	p.(Pro615_Ile616 insArgAspPro)	n/a		yes	
5645	<i>PAX6</i>	IVS2-2	chr11:31,806,927del	NM_015328.4:c.-128-2del	p.(?) / acceptor loss	24.4	LP (0.9)	no (3)	canonical splice variant

Notes: Variants considered not to be even possibly causative are greyed out. All variants relate to the proband of the trio identified by the family ID (FID). ACMG/ACGS classification calculated where the variant is considered possibly causative; *as *AHCYL2* is not a known disease locus this scoring should not strictly be applied, hence the brackets.

(1) Homozygous *PRDM5* variants cause Brittle Cornea Syndrome. I published AD pedigree (+consanguinity), with *ARS* phenotype including iris hypoplasia (318). However, for this case FID75 and the next, FID 356, *PAX6* structural variants are a clear genetic explanation for the aniridia

(2) See functional splicing work presented above for FID 1635 (3) ClinVar VCI000430969.2

Abbreviations: subs, substitution; del, deletion; dup, duplication; ins, insertion; IVS, intervening sequence/intron; CDS, coding sequence; P, pathogenic; U, uncertain; LP, likely pathogenic (ACMG/ACGS criteria/references see **Table 13**)

Table 16 De novo variants detected in the trio WGS of the aniridia cohort

4.4 Discussion

In an aniridia cohort already enriched for unexplained cases (148), a genetic explanation for a majority of previously unsolved families, 22/37 (60%), is arguably a high diagnostic rate.

Although 3 families were explained by *FOXC1* deletions, the remainder of the solved cases (19/22) were attributable to effects on the *PAX6* locus, highlighting the very strong disease-gene link with classical aniridia. As I shall examine in the final discussion, when strictly limiting the phenotype to classical aniridia, it essentially remains a monogenic disorder, despite our greater understanding of the genetics of more varied aniridia-involving phenotypes. This study did not identify likely pathogenic variants involving loci other than *PAX6* or *FOXC1*.

The SVs identified in this study demonstrate the diverse ways of inactivating *PAX6*, namely simple partial or whole deletion of the gene, deletion of a key downstream *cis*-regulatory region (DRR), inversion with an intragenic breakpoint, inversion separating *PAX6* from its DRR, and lastly more complex SVs involving deletion with an inversion or duplication.

The 3 heterozygous *FOXC1* deletions are mostly in keeping with previously published phenotypes (56). In one of the research participants in this work, though, the only phenotypic information available was ‘aniridia’. The other two individuals displayed typical features of *FOXC1* deletions, both with glaucoma (known to be congenital in one) and congenital aortic stenosis. Due to the ocular phenotypic overlap with classical aniridia, *FOXC1* variants are reliably identified in larger aniridia series, both from the FitzPatrick lab (148) and others (64,319). Incidentally, at the Aniridia Europe conferences, where families with aniridia are well represented, there is a small minority of *FOXC1* variant families amongst the delegates (personal experience of 2018 and 2021 meetings) and they are the only ‘minority subgroup’ big enough to be represented after *PAX6*. On the molecular aspects, *FOXC1* is a direct downstream target of *PAX6* in the developing eye (22) although not much detail is published on their interaction.

Whilst I cannot make any serious claim for pathogenicity based on n=1, the *de novo* missense variant in *AHCYL2* does merit reporting as a VUS. *AHCYL2* (*adenosylhomocysteinase-like* 2) somewhat confusingly encodes adenosylhomocysteinase 3 (UniProt Q96HN2), also known as: “IP(3)Rs binding protein released with IP(3) 2”, IRBIT2 or long-IRBIT (a homologue of IRBIT). There is no known Mendelian disease-gene link with aniridia or any other eye disease. However, it is expressed in the eye (GTEx Portal data, accessed Feb 2022) and demonstrates a low affinity interaction with ITPR1 (317), the causative gene in Gillespie syndrome, a syndromic aniridia, so if further cases emerge it is not an implausible link. As well as reporting the finding briefly in our publication (276), I have entered the variant on GeneMatcher (submission “Aniridia2469”), but the only replies to date do not relate to ocular phenotypes.

Novelty

This was the first dedicated WGS study of a classical aniridia cohort, now published (276) and, prior to that, results presented at the Aniridia Europe and Aniridia North American conferences in 2021 (320). Recently, a very large clinical diagnostic cohort (BaMaRa Necker, France) of iris malformation cases has been reported/summarised briefly amongst an excellent review, including many classical aniridia cases who have undergone WGS as well as other more diverse anterior segment dysgeneses (64). Their cohort includes individuals undergoing primary investigation for aniridia and also non-classical phenotypes (though coloboma was excluded); full results have not yet been published, but is a powerful advert for the combination of targeted NGS and WGS in the increasingly diverse genetic landscape of these conditions.

Limitations

Clearly many of these CNVs could have been diagnosed without WGS, using array-based screening. This was where the cohort’s prior screening was deficient, largely due to historical and geographical variation in screening for gene deletions (both in method and completeness). However, the pilot screening of 13 cases that I performed using ddPCR only diagnosed one case, and we considered this an acceptable rate for

proceeding to WGS. Furthermore, it was helpful to observe structural variation on WGS, especially given the changing landscape of bioinformatic callers, and to correlate the results of a visual exploration of this locus with Canvas and Manta. The direct inspection of WGS on IGV was very high yield. However, the main limitation is that it is very labour intensive, made more so by examining loci other than *PAX6* where this was normal. *PAX6* was the only diagnostically rewarding locus examined.

There are several limitations of short-read WGS in respect of SV detection, which will be explored further in the Final Discussion chapter. In addition, there are further disadvantages inherent in methods involving PCR-based amplification, including biased coverage due to high or low GC content in certain regions of the genome (321). The best example of this here is the coverage of the GC-rich *FOXC1* gene, as seen in **Figure 36**, where the depth of coverage can be seen to drop dramatically over the gene itself.

We did not detect any novel disease genes. The inclusion of more trios would have given us greater power to do this. It may also be that classical aniridia is genuinely and exclusively a *PAX6*-associated phenotype.

Advantages of WGS

Gene regulation. Playing to the several strengths of WGS, I detected a large number of structural variants disrupting the relationship between *PAX6* and its critical downstream regulatory region. However, we did not identify any single nucleotide changes in CREs, with the only known such case remaining the SIMO variant published a decade ago (11).

Balanced SVs. The two balanced SVs detected in this cohort realistically would not have been diagnosed without the careful visual study of the *PAX6* locus, given the noise we observed with Manta. For deletions >10kb, there was 100% concordance between those detected on visual inspection and bioinformatically by Canvas. Compared to traditional cytogenetic methods, WGS is able to offer much higher (nucleotide-level) resolution of SV breakpoints.

Splicing variants. With its better per-based coverage of 5'UTR than WES, WGS had strengths in detecting the non-coding exon splicing variants. However, its real advantage was the ability to detect the two deep intronic splicing variants. The combination of approaches (bioinformatic, visual exploration of the data and experimental) was used to thoroughly explore the observed genetic variation and to confirm missplicing for one of them *in vitro*.

Identification of novel disease genes. Although we did not identify novel aniridia-associated genes, as discussed in the section above, clearly this possibility remains an advantage of WGS as an investigation strategy. In addition to greater trio numbers, it would also theoretically be possible to look for common areas of the genome that were deleted in those cases that remain unexplained in the hopes of identifying causative loci.

In summary, this study identified causative variants, some novel, in approximately 60% of those with genetically unsolved aniridia. The results support the use of WGS as both a primary clinical investigation in classical aniridia and demonstrate its power in solving previously unexplained cases, as well as highlighting the limitations of current WGS analysis capabilities. Both long-read sequencing (322) and larger cohorts with greater numbers of trios may help to explain the remaining cases.

Chapter 5 Final discussion and future challenges

In this thesis, I have diagnosed 60% of an unexplained aniridia cohort using whole genome sequencing, and explored the phenotype and impact of severe missense variants in *PAX6* and *MAB21L1*. Several discussion points raised by these themes, particularly in the realms of variant detection and interpretation, merit further consideration.

5.1 Advances in whole genome sequencing in relation to genetic eye disease

The success in diagnosing this aniridia cohort is thanks to the astonishing gains that have been made in the availability, affordability and interpretation of whole genome sequencing (WGS) over the past 15 years (**Figure 51**, below) (324,325). WGS has now entered national healthcare systems and is having a strong impact on genetic diagnosis and gene discovery for rare diseases: this is especially strong for monogenic disorders and when nuclear trios (proband plus unaffected parents) are sequenced (326,327). The results presented here for aniridia support the use of WGS as a primary investigation in rare diseases, demonstrating its value even when (and for some analyses, especially when) a small number of loci are likely to be causative.

At the end of the previous chapter, I discussed the many advantages of WGS over other methods such as whole exome sequencing (WES). This project, though, highlights areas that remain a challenge in WGS interpretation.

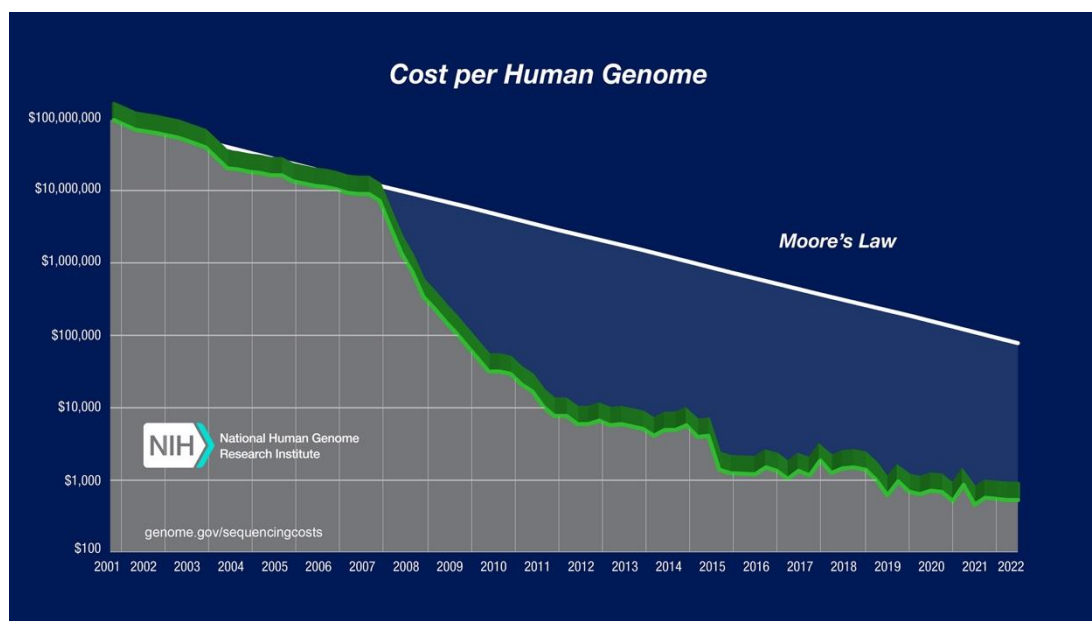


Figure 51 The cost of human whole genome sequencing 2001-2022: figure produced by the National Human Genome Research Institute (325), National Institute of Health. The x-axis shows years 2001 to 2022, and the y-axis shows the cost in USD per human genome, on a logarithmic scale. The green line shows the price per genome, whilst the white line shows the hypothetical price if the technology had continued to follow Moore's Law, which it outpaced. The aniridia WGS project presented here became possible following the major cost decrease in 2015-2016 and was conceived soon after this.

Due to the nature of short-read WGS, with small fragments aligned to a reference genome, balanced structural variants (bSV) remain hard to detect (328): traditional karyotyping or fluorescence *in situ* hybridisation remain a gold standard for their detection, despite their lower resolution, as microarrays can only detect unbalanced SVs. In short-read sequencing, there are limited numbers of sequencing reads whose mapping will be split by the breakpoints, and so callers such as Manta rely heavily on paired reads spanning a bSV. This fails in areas where reads cannot be confidently mapped, for example due to highly repetitive sequences (which are enriched for SVs (329)). These technical limitations can be mitigated by third-generation or long-read sequencing, and by combining existing orthogonal approaches to detect bSVs (328). As well as having read lengths of potentially tens of kilobases (c. 100x that of short-read technology), long-read sequencing (LRS) is also PCR-free (330), which can minimise GC bias (which affects genes encoding several key transcription factors in ocular development including *FOXC1*).

Damián *et al.* used LRS to detect causative bSVs involving *PAX6* or its regulatory region in two unsolved aniridia cases (322): in one of these cases, they reclassified a t(6;11) balanced translocation as pathogenic years after it was first identified and dismissed, by mapping a breakpoint to the DNase I hypersensitive site 2 enhancer within the downstream regulatory region of *PAX6*. LRS has been used to identify a causative mobile element insertion in *EYS* in a family with retinitis pigmentosa (331) and deep intronic splice variants in Usher syndrome (332). These are just a few ocular examples of the ability of LRS to improve diagnostic yield in Mendelian disorders.

In addition to the advantages of LRS in SV detection and repetitive sequences, it offers a high sensitivity for mosaic alterations, detection of DNA methylation (333) and the possibility of haplotype phasing (334). Determining whether a second variant affecting a causative gene is in *cis* or *trans* to the first variant is crucial to diagnosing biallelic compound heterozygous cases; currently this requires sequencing parental DNA which may not be available. Despite its exciting potential, wider adoption of LRS is at present limited by cost, accuracy of SNV detection and throughput, and the lack of widespread bioinformatic and clinical diagnostic expertise for interpretation of variants (333).

5.2 Are any disorders purely monogenic?

Although many other genes are now known to cause an eye malformation which features aniridia (the sign rather than the condition), still no gene has been identified which truly phenocopies *PAX6* classical aniridia with its combination of marked iris and foveal hypoplasia at birth and distinct keratopathy in later life. The eye malformation seen with *MAB21L1* monoallelic Arg51 substitutions does overlap with this, but has a more severe MAC component. In many ways this *PAX6* predominance is entirely unsurprising given the unique niche of *PAX6* as the master regulator of eye development. It is even less surprising when you also factor in the other roles of *PAX6* in the forebrain and pancreas, and consequently the systemic neuroendocrine and neuroanatomical changes that often accompany classical aniridia.

It is my opinion, therefore, that the remaining unsolved “true” classical aniridia cases are entirely due to disruption of *PAX6*, likely via *cis*-regulatory element (CRE) variants, deep intronic splicing variants and structural variants. These were all

categories of variant identified in this research, and ones identified as likely culprits for the diagnostic gap in Mendelian rare diseases (335). So whilst the answer to the question in the title of this section is usually ‘no’, classical aniridia seems to come very close to being monogenic.

Despite the purity of the classical aniridia-*PAX6* relationship, there is of course a ‘but’. In the real world, deep careful phenotyping is time consuming and not always achievable. Aniridia may be the first clue to a new, undifferentiated eye condition, as it can be detected on routine newborn examination. Our increasing knowledge of the genetic differential diagnosis for this rare sign is therefore highly valuable in the initial investigation: informing who requires, for example, *ACTA2* screening or an array CGH to assess *WT1*, in order to identify life-altering syndromic forms of aniridia (1).

Many genetic diagnoses will of course be made (in the UK and elsewhere) by NGS gene panel testing. This research (166) together with the recent work of others (159,167,269,270) cements *MAB21L1* as a causative gene for aniridia and microphthalmia. At the time of writing, 2024, *MAB21L1* is not on the current NHS Structural Eye Disease panel, version 4.0 (182), which covers both aniridia and microphthalmia as indications (or indeed on any other panel with a monoallelic requirement); a situation for which I am actively seeking remedy.

The similarities between the *MAB21L1* monoallelic eye phenotype and the severe end of the *PAX6* missense spectrum are notable. It is not simply that they both cause bilateral eye malformations with aniridia and/or microphthalmia: the constellation of accompanying features is also very similar across both groups, with many demonstrating one or more of corneal opacification, nystagmus, choroidal coloboma, lens subluxation, anomalous optic discs and severe sight impairment. Neither the *PAX6* or *MAB21L1* missense cases are, as a group, structurally as severe as *SOX2*, even *SOX2* missense cases (many of which are anophthalmic) (16), but there is nevertheless overlap. Certainly, *SOX2* would be an important genetic differential diagnosis for such cases, along with *OTX2* and *VSX2* (42). As well as assessing the impact of the *PAX6* variants on *SOX2* cooperative binding (for example using the N3 enhancer) initially in EMSA, future experiments with both *PAX6* and *MAB21L1* variants could look more genome-wide at patterns of binding using ChIP-seq in cell

lines, for example the human lens epithelial cell line CD5a. Lastly, the Moosajee laboratory have successfully made a fibroblast-derived human induced pluripotent stem cell line from the individual described in Chapter 2 with aniridia and microphthalmia and the p.(Asn124Lys) variant (336), which will act as a valuable tool for disease modelling and potential identification of therapeutic targets.

5.3 Towards a better understanding of *cis*-regulatory variants in *PAX6*

Many of the remaining unsolved aniridia cases are likely to be disrupting the *cis*-regulation of *PAX6* in ways we are not currently well able to detect or interpret. Deletions of the downstream regulatory region (DRR) are well reported in the literature, with a 244 kb “critical region” required for transcriptional activation previously proposed by the FitzPatrick laboratory (148). These patients have a mild classical aniridia phenotype, interestingly with mild or no aniridia keratopathy (reviewed in (64)). One Chinese family is reported with a 681 kb DRR deletion and described as having iris coloboma (337): my review of the clinical images presented though suggests that they in fact have partial aniridia, as well as having typical features consistent with classical aniridia of nystagmus and posterior polar cataracts.

It is remarkable that the SIMO point mutation case (11) is still the only pathogenic SNV detected in a *cis*-regulatory element (CRE). Plaisancié *et al.* report 6 VUS in 4 patients involving SNVs in *PAX6* DRR CRE but very sensibly do not claim pathogenicity for these given that they are present in gnomAD, albeit one at 0.003% (10). A better understanding of the *PAX6* regulatory landscape will feed into our ability to detect and interpret genetic variants in noncoding regions, especially single nucleotide variants.

PAX6 has the advantage of several decades of discovery and parsing of its CREs. Whilst regulatory elements upstream and around the first few introns were well recognised at the time, in 2001 Veronica van Heyningen and Dirk-Jan Kleinjan were guided by chromosomal breakpoints in individuals with classical aniridia to look for regulatory elements far downstream of *PAX6*, which they identified, coining the concept of the downstream regulatory region, DRR (338). With the Price laboratory,

they went on to demonstrate that the DRR was essential for *Pax6* for correct expression in the eye and brain, and identified tissue-specific enhancers (339). A detailed, interactive browser of all the *PAX6* CREs is hosted by the Bickmore group at the MRC HGU (210) drawing together the wealth of evidence we have about *PAX6* CREs today (Figure 52).

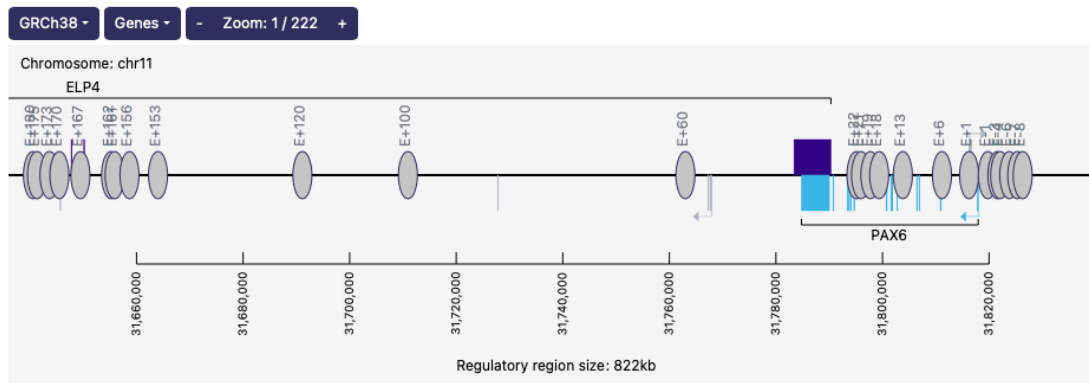


Figure 52 The PAX6 Regulatory Browser. Screenshot of the interactive web application detailing all known *PAX6* cis-regulatory elements (CRE), their tissue expression and the publications identifying them (210). The *PAX6* exons and transcription start site are shown in blue, and CREs in grey ovals. E+153 indicated towards the left of the image is also known as SIMO.

Using a dual enhancer-reporter assay in zebrafish embryos, Uttley *et al.* recently demonstrated the unique expression patterns of the retinal enhancers HS5 and NRE *in vivo*. Combined with single-cell RNAseq, they showed the precise cell types and spatiotemporal activity of the two enhancers in the developing eye (340).

5.4 How much can you predict a *PAX6* phenotype from genotype?

Certain genotype-phenotype correlations have recently become established for *PAX6* eye malformations, all with the anchor or reference point of the distinctive “null” classical aniridia phenotype. Missense phenotypes, though, remain difficult to predict. Confirming the wide severity spectrum I outlined in Chapter 2, missense variants show striking phenotypic variability and include a large proportion of non-aniridia phenotypes – as high as two thirds non-aniridia in a recent study of 125 Russian aniridia families (341) – as well as being capable of mimicking the null phenotype (64). Despite the notable severe MAC cases discussed in this work affecting specific

residues of the paired and homeodomain (195), overall missense variants are statistically milder than nonsense with better visual acuity and lower grades of foveal and iris hypoplasia (68). *Cis*-regulatory deletions in the DRR also show milder phenotypes rather than being the entirely classical aniridia cases we originally thought, as discussed above, with little to no keratopathy (64).

Careful reporting of genotype-phenotype correlations remains an important part of our ability to diagnose developmental disorders safely (342). However, missense variants continue to present a challenge and *in silico* predictors are, as we shall see, increasingly valuable both in the clinical prediction sphere and for answering wider structure-function questions about proteins.

5.5 Determining the impact of missense variants

Pertinent to this issue is the impressive recent advances in the field of computational prediction, both the prediction of structural effects (especially with FoldX and then AlphaFold) and clinical variant effect (especially AlphaMissense).

Variant effect predictors (VEPs) are traditionally algorithms which predict the likely fitness effect or pathogenicity of a variant, by harnessing diverse biological data (structural differences, conservation, functional annotation etc). In the age of machine learning, newer VEPs vastly outperform their predecessors from several years ago (223), as illustrated by the comparison in Chapter 2, **Table 5**, with modern VEPs all predicting the missense pathogenicity accurately in terms of correlation with our phenotyping and functional work.

A concern regarding benchmarking of VEPs is the potential for bias from data circularity, either at the variant or gene level, where there are overlap or similarities in the training set and performance evaluation set (225). One way of feeding reliable data to VEPs that limits these concerns is from high-throughput experimental strategies known as multiplexed assays of variant effect (MAVE) (343). My work on *PAX6* missense variants led to a Deep Mutational Scanning experiment, a subclass of MAVE, performed by the Kudla and Marsh laboratories: McDonnell *et al.* used a yeast-1-hybrid assay (including assessing binding to LE9) to perform saturation mutagenesis, assessing substitutions in all of the amino acids in the *PAX6* paired domain (344),

quantifying the effect of >2500 variants. In agreement with my EMSA results, they observed that Ser54Arg results in a substantial loss of affinity with LE9, and noted increased binding with another target sequence, BLX. This supports the hypothesis that missense variants alter the DNA binding specificity to differing amounts depending on the target, in a variant-specific fashion, accounting for the diversity of phenotypes (344). Experimentally-derived variant effect maps derived from work such as this have the potential to significantly improve our interpretation of variants in these genes (often key genes); it is of course worth remembering though that data from MAVEs only cover a minute proportion of the human genome (266).

A further strategy to improve variant effect prediction is the concept of gene-specific thresholds for computational predictors (345), which has recently been demonstrated for *PAX6* by Andhika *et al.* (224), showing the best-performing predictor in their study was AlphaMissense with a gene-specific threshold. Enhancing the power of variant interpretation is important both on an individual level when confronted with VUS in a patient's sequencing report, and at a wider level to provide scalable methods of interpreting the overwhelming amount of sequencing data we are now faced with.

5.6 Concluding remarks

The work presented here highlights the importance of the wider *PAX6* locus in genetic eye disease, and builds on our already rich understanding of the gene. I have touched on this in work cited above, but in concluding this thesis allow me to draw attention again to the spatiotemporal context of this work. It follows a long and inspiring history of *PAX6* and aniridia work both afar and locally in the MRC Human Genetics Unit in Edinburgh, in particular but not limited to that of Professor Veronica van Heyningen (*PAX6* and the disease-gene link), Professor Nick Hastie (*WT1*), Professor David FitzPatrick (gene discovery in eye malformations, especially MAC), and lastly Dr DJ Kleinjan, Dr Shipra Bhatia and Professor Wendy Bickmore (*cis*-regulation). My thanks to these scientists and to the many individuals with eye malformations who consented to contribute their DNA towards genetic research.

References

1. Hall HN, Williamson KA, FitzPatrick DR. The genetic architecture of aniridia and Gillespie syndrome. *Hum Genet.* 2019 Sep 1;138(8–9):881–98.
2. Burgess FR, Hall HN, Megaw R. Emerging Gene Manipulation Strategies for the Treatment of Monogenic Eye Disease. *The Asia-Pacific Journal of Ophthalmology.* 2022 Aug;11(4):380.
3. Hingorani M, Hanson I, van Heyningen V. Aniridia. *Eur J Hum Genet.* 2012 Oct 1;20(10):1011–7.
4. Grønskov K, Olsen JH, Sand A, Pedersen W, Carlsen N, Bak Jylling AM, et al. Population-based risk estimates of Wilms tumor in sporadic aniridia. A comprehensive mutation screening procedure of PAX6 identifies 80% of mutations in aniridia. *Hum Genet.* 2001 Jul 1;109(1):11–8.
5. Eden U, Iggman D, Riise R, Tornqvist K. Epidemiology of aniridia in Sweden and Norway. *Acta Ophthalmol.* 2008 Nov 1;86(7):727–9.
6. Shaw MW, Falls HF, Neel JV. Congenital Aniridia. *Am J Hum Genet.* 1960 Dec 1;12(4 Pt 1):389–415.
7. Prosser J, van Heyningen V. PAX6 mutations reviewed. *Hum Mutat.* 1998;11(2):93–108.
8. Lauderdale JD, Wilensky JS, Oliver ER, Walton DS, Glaser T. 3' deletions cause aniridia by preventing PAX6 gene expression. *Proc Natl Acad Sci USA.* 2000 Dec 5;97(25):13755–9.
9. McBride DJ, Buckle A, van Heyningen V, Kleinjan DA. DNase hypersensitivity and ultraconservation reveal novel, interdependent long-range enhancers at the complex Pax6 cis-regulatory region. Hwang SPL, editor. *PLoS ONE.* 2011 Jan 1;6(12):e28616.
10. Plaisancié J, Tarilonte M, Ramos P, Jeanton-Scaramouche C, Gaston V, Dollfus H, et al. Implication of non-coding PAX6 mutations in aniridia. *Hum Genet.* 2018 Oct;137(10):831–46.
11. Bhatia S, Bengani H, Fish M, Brown A, Divizia MT, de Marco R, et al. Disruption of autoregulatory feedback by a mutation in a remote, ultraconserved PAX6 enhancer causes aniridia. *Am J Hum Genet.* 2013 Dec 5;93(6):1126–34.
12. Nallathambi J, Neethirajan G, Shashikant S, Vijayalakshmi P, Sundaresan P. PAX6 missense mutations associated in patients with optic nerve malformation. *Mol Vis.* 2006 Mar 30;12:236–42.

13. Azuma N, Yamaguchi Y, Handa H, Tadokoro K, Asaka A, Kawase E, et al. Mutations of the PAX6 gene detected in patients with a variety of optic-nerve malformations. *Am J Hum Genet.* 2003 Jun 1;72(6):1565–70.
14. Prokudin I, Simons C, Grigg JR, Storen R, Kumar V, Phua ZY, et al. Exome sequencing in developmental eye disease leads to identification of causal variants in GJA8, CRYGC, PAX6 and CYP1B1. *Eur J Hum Genet.* 2014 Jul 1;22(7):907–15.
15. Chassaing N, Causse A, Vigouroux A, Delahaye A, Alessandri JL, Boespflug-Tanguy O, et al. Molecular findings and clinical data in a cohort of 150 patients with anophthalmia/microphthalmia. *Clin Genet.* 2014 Oct 1;86(4):326–34.
16. Williamson KA, FitzPatrick DR. The genetic architecture of microphthalmia, anophthalmia and coloboma. *Eur J Med Genet.* 2014 Aug 1;57(8):369–80.
17. Lima Cunha D, Arno G, Corton M, Moosajee M. The Spectrum of PAX6 Mutations and Genotype-Phenotype Correlations in the Eye. *Genes.* 2019 Dec 1;10(12):1050–22.
18. Liu H, Xu S, Wang Y, Mazerolle C, Thurig S, Coles BLK, et al. Ciliary margin transdifferentiation from neural retina is controlled by canonical Wnt signaling. *Dev Biol.* 2007 Aug 1;308(1):54–67.
19. Davis-Silberman N, Ashery-Padan R. Iris development in vertebrates; genetic and molecular considerations. *Brain Res.* 2008 Feb 4;1192:17–28.
20. Bao ZZ, Cepko CL. The expression and function of Notch pathway genes in the developing rat eye. *J Neurosci.* 1997 Feb 15;17(4):1425–34.
21. Mann IC. The development of the human iris. *Br J Ophthalmol.* 1925 Oct 1;9(10):495–512.
22. Wang X, Shan X, Gregory-Evans CY. A mouse model of aniridia reveals the in vivo downstream targets of Pax6 driving iris and ciliary body development in the eye. *Biochim Biophys Acta.* 2016 Oct 20;1863(1):60–7.
23. Chang B, Smith RS, Peters M, Savinova OV, Hawes NL, Zabaleta A, et al. Haploinsufficient Bmp4 ocular phenotypes include anterior segment dysgenesis with elevated intraocular pressure. *BMC Genet.* 2001 Jan 1;2(1):18.
24. Adams D, Karolak M, Robertson E, Oxburgh L. Control of kidney, eye and limb expression of Bmp7 by an enhancer element highly conserved between species. *Dev Biol.* 2007 Nov 15;311(2):679–90.

25. Hägglund AC, Jones I, Carlsson L. A novel mouse model of anterior segment dysgenesis (ASD): conditional deletion of Tsc1 disrupts ciliary body and iris development. *Dis Model Mech*. 2017 Mar 1;10(3):245–57.
26. Jensen AM. Potential roles for BMP and Pax genes in the development of iris smooth muscle. *Dev Dyn*. 2005 Feb 1;232(2):385–92.
27. Imaizumi M, Kuwabara T. Development of the rat iris. *Invest Ophthalmol*. 1971 Oct 1;10(10):733–44.
28. Williams AL, Bohnsack BL. Neural crest derivatives in ocular development: Discerning the eye of the storm. *Birth Defect Res C*. 2015 Jun 4;105(2):87–95.
29. Sturm RA, Larsson M. Genetics of human iris colour and patterns. *Pigment Cell Melanoma Res*. 2009 Oct 1;22(5):544–62.
30. Cvekl A, Tamm ER. Anterior eye development and ocular mesenchyme: new insights from mouse models and human diseases. *Bioessays*. 2004 Apr 1;26(4):374–86.
31. Matsuo T, Osumi-Yamashita N, Noji S, Ohuchi H, Koyama E, Myokai F, et al. A mutation in the Pax-6 gene in rat small eye is associated with impaired migration of midbrain crest cells. *Nat Genet*. 1993 Apr;3(4):299–304.
32. Prota G, Hu DN, Vincensi MR, McCormick SA, Napolitano A. Characterization of melanins in human irides and cultured uveal melanocytes from eyes of different colors. *Exp Eye Res*. 1998 Sep 1;67(3):293–9.
33. Rennie IG. Don't it make my blue eyes brown: heterochromia and other abnormalities of the iris. *Eye (Lond)*. 2012 Jan 1;26(1):29–50.
34. White D, Rabago-Smith M. Genotype-phenotype associations and human eye color. *J Hum Genet*. 2011 Jan 1;56(1):5–7.
35. Roguz K, Gallagher MK, Senden E, Bar-Lev Y, Lebel M, Helicz R, et al. All the Colors of the Rainbow: Diversification of Flower Color and Intraspecific Color Variation in the Genus *Iris*. *Front Plant Sci*. 2020;11:569811.
36. Möller R. Purple Iris (*Iris versicolor*) in a garden in Bamberg, Germany [Internet]. [cited 2024 Nov 16]. Available from: [https://commons.wikimedia.org/wiki/File:Verschiedenfarbige_Schwertlilie_\(Iris_versicolor\)-20200603-RM-100257.jpg](https://commons.wikimedia.org/wiki/File:Verschiedenfarbige_Schwertlilie_(Iris_versicolor)-20200603-RM-100257.jpg)
37. Head G. *Iris Carrying the Water of the River Styx to Olympus for the Gods to Swear By* (oil on canvas, Nelson-Atkins Museum of Art; photograph of cited artwork created by Jonathunder, uploaded to Wikimedia Commons

as public domain) [Internet]. 1793 [cited 2024 Nov 16]. Available from:
<https://commons.wikimedia.org/wiki/File:IrisGuyHead.jpg>

38. Nischal KK. Genetics of Congenital Corneal Opacification--Impact on Diagnosis and Treatment. *Cornea*. 2015 Oct 1;34 Suppl 10:S24-34.
39. Rahi JS. Childhood blindness: a UK epidemiological perspective. *Eye*. 2007 Oct;21(10):1249–53.
40. Dharmasena A, Keenan T, Goldacre R, Hall N, Goldacre MJ. Trends over time in the incidence of congenital anophthalmia, microphthalmia and orbital malformation in England: database study. *British Journal of Ophthalmology*. 2017 Jun 1;101(6):735–9.
41. Morrison D, FitzPatrick D, Hanson I, Williamson K, van Heyningen V, Fleck B, et al. National study of microphthalmia, anophthalmia, and coloboma (MAC) in Scotland: investigation of genetic aetiology. *J Med Genet*. 2002 Jan 1;39(1):16–22.
42. Harding P, Moosajee M. The Molecular Basis of Human Anophthalmia and Microphthalmia. *Journal of Developmental Biology*. 2019 Sep;7(3):16.
43. Fantes J, Ragge NK, Lynch SA, McGill NI, Collin JRO, Howard-Peebles PN, et al. Mutations in SOX2 cause anophthalmia. *Nat Genet*. 2003 Apr 1;33(4):461–3.
44. Ragge NK, Brown AG, Poloschek CM, Lorenz B, Henderson RA, Clarke MP, et al. Heterozygous mutations of OTX2 cause severe ocular malformations. *Am J Hum Genet*. 2005 Jun 1;76(6):1008–22.
45. Danno H, Michiue T, Hitachi K, Yukita A, Ishiura S, Asashima M. Molecular links among the causative genes for ocular malformation: Otx2 and Sox2 coregulate Rax expression. *Proceedings of the National Academy of Sciences*. 2008 Apr 8;105(14):5408–13.
46. Gerth-Kahlert C, Williamson K, Ansari M, Rainger JK, Hingst V, Zimmermann T, et al. Clinical and mutation analysis of 51 probands with anophthalmia and/or severe microphthalmia from a single center. *Mol Genet Genomic Med*. 2013 May 1;1(1):15–31.
47. Hardy H, Rainger J. Cell adhesion marker expression dynamics during fusion of the optic fissure. *Gene Expr Patterns*. 2023 Dec;50:119344.
48. Holt R, Goudie D, Verde AD, Gardham A, Ramond F, Putoux A, et al. Individuals with heterozygous variants in the Wnt-signalling pathway gene FZD5 delineate a phenotype characterized by isolated coloboma and variable expressivity. *Ophthalmic Genet*. 2022 Dec;43(6):809–16.

49. Liu C, Widen SA, Williamson KA, Ratnapriya R, Gerth-Kahlert C, Rainger J, et al. A secreted WNT-ligand-binding domain of FZD5 generated by a frameshift mutation causes autosomal dominant coloboma. *Hum Mol Genet*. 2016 Apr 1;25(7):1382–91.
50. Rainger J, Williamson KA, Soares DC, Truch J, Kurian D, Gillessen-Kaesbach G, et al. A recurrent de novo mutation in ACTG1 causes isolated ocular coloboma. *Human Mutation*. 2017;38(8):942–6.
51. Hanson I, Jordan T, van Heyningen V. Aniridia. In: *Molecular Genetics of Inherited Eye Disorders* [Internet]. CRC Press; 1994. Available from: <http://www.worldcat.org/title/molecular-genetics-of-inherited-eye-disorders/oclc/479896370>
52. Ogino H, Ochi H, Reza HM, Yasuda K. Transcription factors involved in lens development from the preplacodal ectoderm. *Developmental Biology*. 2012 Mar 15;363(2):333–47.
53. Verma AS, FitzPatrick DR. Anophthalmia and microphthalmia. *Orphanet J Rare Dis*. 2007 Jan 1;2(1):47.
54. Plaisancié J, Ragge N k., Dollfus H, Kaplan J, Lehalle D, Francannet C, et al. FOXE3 mutations: genotype-phenotype correlations. *Clinical Genetics*. 2018;93(4):837–45.
55. Weh E, Reis LM, Happ HC, Levin AV, Wheeler PG, David KL, et al. Whole exome sequence analysis of Peters anomaly. *Hum Genet*. 2014 Dec 1;133(12):1497–511.
56. Reis LM, Tyler RC, Volkmann Kloss BA, Schilter KF, Levin AV, Lowry RB, et al. PITX2 and FOXC1 spectrum of mutations in ocular syndromes. *Eur J Hum Genet*. 2012 Dec 1;20(12):1224–33.
57. Solebo AL, Russell-Eggitt I, Cumberland P, Rahi JS. Congenital cataract associated with persistent fetal vasculature: findings from IoLunder2. *Eye*. 2016 Sep;30(9):1204–9.
58. Chen C, Xiao H, Ding X. Persistent Fetal Vasculature. *Asia Pac J Ophthalmol (Phila)*. 2019;8(1):86–95.
59. Bell SJ, Oluonye N, Harding P, Moosajee M. Congenital cataract: a guide to genetic and clinical management. *Therapeutic Advances in Rare Disease*. 2020 Jan 1;1:2633004020938061.
60. Barratta G. Osservazioni pratiche sulle principali malattie degli occhi. Vol. 2. Milan; 1818. 349 p.
61. Nelson LB, Spaeth GL, Nowinski TS, Margo CE, Jackson L. Aniridia. A review. *Surv Ophthalmol*. 1984 May 1;28(6):621–42.

62. Grønskov K, Rosenberg T, Sand A, Brøndum-Nielsen K. Mutational analysis of PAX6: 16 novel mutations including 5 missense mutations with a mild aniridia phenotype. *Eur J Hum Genet.* 1999 Apr 1;7(3):274–86.
63. Thomas MG, Kumar A, Mohammad S, Proudlock FA, Engle EC, Andrews C, et al. Structural grading of foveal hypoplasia using spectral-domain optical coherence tomography a predictor of visual acuity? *Ophthalmology.* 2011 Aug;118(8):1653–60.
64. Daruich A, Duncan M, Robert MP, Lagali N, Semina EV, Aberdam D, et al. Congenital aniridia beyond black eyes: From phenotype and novel genetic mechanisms to innovative therapeutic approaches. *Progress in Retinal and Eye Research.* 2023 Jul;95:101133.
65. Self JE, Dunn MJ, Erichsen JT, Gottlob I, Griffiths HJ, Harris C, et al. Management of nystagmus in children: a review of the literature and current practice in UK specialist services. *Eye.* 2020 Sep;34(9):1515–34.
66. Bertsch M, Floyd M, Kehoe T, Pfeifer W, Drack AV. The clinical evaluation of infantile nystagmus: What to do first and why. *Ophthalmic Genet.* 2017;38(1):22–33.
67. Schneider S, Osher RH, Burk SE, Lutz TB, Montione R. Thinning of the anterior capsule associated with congenital aniridia. *Journal of Cataract & Refractive Surgery.* 2003 Mar 1;29(3):523–5.
68. Kit V, Cunha DL, Hagag AM, Moosajee M. Longitudinal genotype-phenotype analysis in 86 patients with *PAX6*-related aniridia. *JCI Insight* [Internet]. 2021 Jul 22 [cited 2023 Nov 12];6(14). Available from: <https://insight.jci.org/articles/view/148406>
69. Grant WM, Walton DS. Progressive changes in the angle in congenital aniridia, with development of glaucoma. *Am J Ophthalmol.* 1974 Nov 1;78(5):842–7.
70. Netland PA. Management of Glaucoma in Congenital Aniridia. In: *Aniridia* [Internet]. Cham: Springer International Publishing; 2015. p. 27–37. Available from: http://link.springer.com/10.1007/978-3-319-19779-1_4
71. Mataftsi A. Incidence of and Risk Factors for Postoperative Glaucoma and Its Treatment in Paediatric Cataract Surgery. *Dev Ophthalmol.* 2016 Jan 1;57:40–8.
72. Douvaras P, Mort RL, Edwards D, Ramaesh K, Dhillon B, Morley SD, et al. Increased corneal epithelial turnover contributes to abnormal homeostasis in the Pax6(+/-) mouse model of aniridia. Cannon CJ, editor. *PLoS ONE.* 2013 Jan 1;8(8):e71117.

73. Edén U, Fagerholm P, Danyali R, Lagali N. Pathologic epithelial and anterior corneal nerve morphology in early-stage congenital aniridic keratopathy. *Ophthalmology*. 2012 Sep;119(9):1803–10.
74. Mayer KL, Nordlund ML, Schwartz GS, Holland EJ. Keratopathy in congenital aniridia. *Ocul Surf*. 2003 Apr 1;1(2):74–9.
75. Alafaleq M, Sordello L, Bremond-Gignac D. Congenital Aniridia and Ocular motility. *American Journal of Ophthalmology*. 2023 Mar;247:145–51.
76. Voskresenskaya A, Pozdeyeva N, Batkov Y, Vasilyeva T, Marakhonov A, West RA, et al. Morphometric analysis of the lens in human aniridia and mouse *Small eye*. *Experimental Eye Research*. 2021 Feb 1;203:108371.
77. Yogarajah M, Matarin M, Vollmar C, Thompson PJ, Duncan JS, Symms M, et al. PAX6, brain structure and function in human adults: advanced MRI in aniridia. *Ann Clin Transl Neurol*. 2016 May 1;3(5):314–30.
78. Free SL, Mitchell TN, Williamson KA, Churchill AJ, Shorvon SD, Moore AT, et al. Quantitative MR image analysis in subjects with defects in the PAX6 gene. *Neuroimage*. 2003 Dec 1;20(4):2281–90.
79. Mitchell TN, Free SL, Williamson KA, Stevens JM, Churchill AJ, Hanson IM, et al. Polymicrogyria and absence of pineal gland due to PAX6 mutation. *Ann Neurol*. 2003 May 1;53(5):658–63.
80. Sisodiya SM, Free SL, Williamson KA, Mitchell TN, Willis C, Stevens JM, et al. PAX6 haploinsufficiency causes cerebral malformation and olfactory dysfunction in humans. *Nat Genet*. 2001 Jul 1;28(3):214–6.
81. Bamiou DE, Free SL, Sisodiya SM, Chong WK, Musiek F, Williamson KA, et al. Auditory interhemispheric transfer deficits, hearing difficulties, and brain magnetic resonance imaging abnormalities in children with congenital aniridia due to PAX6 mutations. *Arch Pediatr Adolesc Med*. 2007 May 1;161(5):463–9.
82. Netland PA, Scott ML, Boyle JW, Lauderdale JD. Ocular and systemic findings in a survey of aniridia subjects. *J AAPOS*. 2011 Dec 1;15(6):562–6.
83. Menke A, Casagrande S, Geiss L, Cowie CC. Prevalence of and Trends in Diabetes Among Adults in the United States, 1988-2012. *JAMA*. 2015 Sep 8;314(10):1021–9.
84. Swisa A, Avrahami D, Eden N, Zhang J, Feleke E, Dahan T, et al. PAX6 maintains β cell identity by repressing genes of alternative islet cell types. *J Clin Invest*. 2017 Jan 3;127(1):230–43.
85. Cvekl A, Callaerts P. PAX6: 25th anniversary and more to learn. *Exp Eye Res*. 2017 Mar 1;156:10–21.

86. van der Meer-de Jong R, Dickinson ME, Woychik RP, Stubbs L, Hetherington C, Hogan BL. Location of the gene involving the small eye mutation on mouse chromosome 2 suggests homology with human aniridia 2 (AN2). *Genomics*. 1990 Jun 1;7(2):270–5.
87. Glaser T, Lane J, Housman D. A mouse model of the aniridia-Wilms tumor deletion syndrome. *Science*. 1990 Nov 9;250(4982):823–7.
88. Roberts RC. Small eyes - a new dominant eye mutant in the mouse. *Genetics Research*. 1967 Feb 1;9(1):121–2.
89. Jordan T, Hanson I, Zaletayev D, Hodgson S, Prosser J, Seawright A, et al. The human PAX6 gene is mutated in two patients with aniridia. *Nat Genet*. 1992 Aug 1;1(5):328–32.
90. Walther C, Guenet JL, Simon D, Deutsch U, Jostes B, Goulding MD, et al. Pax: a murine multigene family of paired box-containing genes. *Genomics*. 1991 Oct 1;11(2):424–34.
91. Ton CC, Hirvonen H, Miwa H, Weil MM, Monaghan P, Jordan T, et al. Positional cloning and characterization of a paired box- and homeobox-containing gene from the aniridia region. *Cell*. 1991 Dec 20;67(6):1059–74.
92. Hill RE, Favor J, Hogan BL, Ton CC, Saunders GF, Hanson IM, et al. Mouse small eye results from mutations in a paired-like homeobox-containing gene. *Nature*. 1991 Dec 1;354(6354):522–5.
93. PAX6 | gnomAD v4.1.0 | gnomAD [Internet]. [cited 2024 Dec 1]. Available from: https://gnomad.broadinstitute.org/gene/ENSG00000007372?dataset=gnomad_r4
94. Karczewski KJ, Weisburd B, Thomas B, Solomonson M, Ruderfer DM, Kavanagh D, et al. The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic Acids Res*. 2017 Jan 4;45(D1):D840–5.
95. Schmidt-Sidor B, Szymańska K, Williamson K, van Heyningen V, Roszkowski T, Wierzba-Bobrowicz T, et al. Malformations of the brain in two fetuses with a compound heterozygosity for two PAX6 mutations. *Folia Neuropathol*. 2009 Jan 1;47(4):372–82.
96. Schilter KF, Reis LM, Schneider A, Bardakjian TM, Abdul-Rahman O, Kozel BA, et al. Whole-genome copy number variation analysis in anophthalmia and microphthalmia. *Clin Genet*. 2013 Nov 1;84(5):473–81.
97. Aalfs CM, Fantes JA, Wenniger-Prick LJJM, Sluijter S, Hennekam RCM, van Heyningen V, et al. Tandem duplication of 11p12-p13 in a child with borderline development delay and eye abnormalities: Dose effect of the PAX6 gene product? *Am J Med Genet*. 1997 Dec 19;73(3):267–71.

98. Schedl A, Ross A, Lee M, Engelkamp D, Rashbass P, van Heyningen V, et al. Influence of PAX6 gene dosage on development: overexpression causes severe eye abnormalities. *Cell*. 1996 Jul 12;86(1):71–82.
99. Glaser T, Jepeal L, Edwards JG, Young SR, Favor J, Maas RL. PAX6 gene dosage effect in a family with congenital cataracts, aniridia, anophthalmia and central nervous system defects. *Nat Genet*. 1994 Aug 1;7(4):463–71.
100. Sivak JM, West-Mays JA, Yee A, Williams T, Fini ME. Transcription Factors Pax6 and AP-2 Interact To Coordinate Corneal Epithelial Repair by Controlling Expression of Matrix Metalloproteinase Gelatinase B. *Mol Cell Biol*. 2003 Dec 12;24(1):245–57.
101. Huettl RE, Eckstein S, Stahl T, Petricca S, Ninkovic J, Götz M, et al. Functional dissection of the Pax6 paired domain: Roles in neural tube patterning and peripheral nervous system development. *Dev Biol*. 2016 May 1;413(1):86–103.
102. Maurya SK, Mishra R. Pax6 Binds to Promoter Sequence Elements Associated with Immunological Surveillance and Energy Homeostasis in Brain of Aging Mice. *Ann Neurosci*. 2017 May 1;24(1):20–5.
103. Hart AW, Mella S, Mendrychowski J, van Heyningen V, Kleinjan DA. The developmental regulator Pax6 is essential for maintenance of islet cell function in the adult mouse pancreas. Zhou R, editor. *PLoS ONE*. 2013 Jan 1;8(1):e54173.
104. Georgala PA, Carr CB, Price DJ. The role of Pax6 in forebrain development. *Dev Neurobiol*. 2011 Aug 1;71(8):690–709.
105. The UniProt Consortium. UniProt: the universal protein knowledgebase. 2017 Jan 4;45(D1):D158–69.
106. Jun S, Desplan C. Cooperative interactions between paired domain and homeodomain. *Development*. 1996 Sep 1;122(9):2639–50.
107. Kim J, Kim J, Lauderdale JD. Analysis of Pax6 expression using a BAC transgene reveals the presence of a paired-less isoform of Pax6 in the eye and olfactory bulb. *Dev Biol*. 2006 Apr 15;292(2):486–505.
108. Carrière C, Plaza S, Caboche J, Dozier C, Bailly M, Martin P, et al. Nuclear localization signals, DNA binding, and transactivation properties of quail Pax-6 (Pax-QNR) isoforms. *Cell Growth Differ*. 1995 Dec 1;6(12):1531–40.
109. Bandah D, Rosenmann A, Blumenfeld A, Averbukh E, Banin E, Sharon D. A novel de novo PAX6 mutation in an Ashkenazi-Jewish family with aniridia. *Mol Vis*. 2008 Jan 28;14:142–5.

110. Tabata H, Koinui A, Ogura A, Nishihara D, Yamamoto H. A novel nuclear localization signal spans the linker of the two DNA-binding subdomains in the conserved paired domain of Pax6. *Genes Genet Syst.* 2018 Apr 2;17–00057.
111. Ploski JE, Shamsheer MK, Radu A. Paired-Type Homeodomain Transcription Factors Are Imported into the Nucleus by Karyopherin 13. *Mol Cell Biol.* 2004 May 13;24(11):4824–34.
112. Shin DH, Kwon BS, Chang YP, Bae S ryeon, Kim J, Kim JW. Ultramicroscopical immunolocalization of PAX6 in the adult chicken retina. 2003 Jan 1;105(3):267–72.
113. Simpson TI, Price DJ. Pax6; a pleiotropic player in development. *Bioessays.* 2002 Nov 1;24(11):1041–51.
114. Zhang Y, Ferreira HB, Greenstein D, Chisholm A, Emmons SW. Regulated nuclear entry of the *C. elegans* Pax-6 transcription factor. *Mech Dev.* 1998 Nov 1;78(1–2):179–87.
115. Tripathi R, Mishra R. Interaction of Pax6 with SPARC and p53 in brain of mice indicates Smad3 dependent auto-regulation. *J Mol Neurosci.* 2010 Jul 1;41(3):397–403.
116. Xu HE, Rould MA, Xu W, Epstein JA, Maas RL, Pabo CO. Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. *Genes Dev.* 1999 May 15;13(10):1263–75.
117. Moreland JL, Gramada A, Buzko OV, Zhang Q, Bourne PE. The Molecular Biology Toolkit (MBT): a modular platform for developing molecular visualization applications. *BMC Bioinformatics.* 2005 Feb 6;6(1):21.
118. Azuma N, Yamaguchi Y, Handa H, Hayakawa M, Kanai A, Yamada M. Missense mutation in the alternative splice region of the PAX6 gene in eye anomalies. *Am J Hum Genet.* 1999 Sep 1;65(3):656–63.
119. Epstein JA, Glaser T, Cai J, Jepeal L, Walton DS, Maas RL. Two independent and interactive DNA-binding subdomains of the Pax6 paired domain are regulated by alternative splicing. *Genes Dev.* 1994 Sep 1;8(17):2022–34.
120. Kozmik Z, Czerny T, Busslinger M. Alternatively spliced insertions in the paired domain restrict the DNA sequence specificity of Pax6 and Pax8. *EMBO J.* 1997 Nov 17;16(22):6793–803.
121. Singh S, Mishra R, Arango NA, Deng JM, Behringer RR, Saunders GF. Iris hypoplasia in mice that lack the alternatively spliced Pax6(5a) isoform. *Proc Natl Acad Sci USA.* 2002 May 14;99(10):6812–5.

122. Kondoh H, Uchikawa M, Kamachi Y. Interplay of Pax6 and SOX2 in lens development as a paradigm of genetic switch mechanisms for cell differentiation. *Int J Dev Biol*. 2004 Jan 1;48(8–9):819–27.
123. Hu C, Malik V, Chang YK, Veerapandian V, Srivastava Y, Huang YH, et al. Coop-Seq Analysis Demonstrates that Sox2 Evokes Latent Specificities in the DNA Recognition by Pax6. *J Mol Biol*. 2017 Nov 24;429(23):3626–34.
124. Kamachi Y, Uchikawa M, Tanouchi A, Sekido R, Kondoh H. Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development. *Genes Dev*. 2001 May 15;15(10):1272–86.
125. Goudreau G, Petrou P, Reneker LW, Graw J, Löster J, Gruss P. Mutually regulated expression of Pax6 and Six3 and its implications for the Pax6 haploinsufficient lens phenotype. *Proc Natl Acad Sci USA*. 2002 Jun 25;99(13):8719–24.
126. Cvekl A, Sax CM, Bresnick EH, Piatigorsky J. A complex array of positive and negative elements regulates the chicken alpha A-crystallin gene: involvement of Pax-6, USF, CREB and/or CREM, and AP-1 proteins. *Mol Cell Biol*. 1994 Nov;14(11):7363–76.
127. Schwarz M, Cecconi F, Bernier G, Andrejewski N, Kammandel B, Wagner M, et al. Spatial specification of mammalian eye territories by reciprocal transcriptional repression of Pax2 and Pax6. *Development*. 2000 Oct 15;127(20):4325–34.
128. Sun J, Rockowitz S, Xie Q, Ashery-Padan R, Zheng D, Cvekl A. Identification of in vivo DNA-binding mechanisms of Pax6 and reconstruction of Pax6-dependent gene regulatory networks during forebrain and lens development. 2015 Aug 18;43(14):6827–46.
129. Lakowski J, Majumder A, Lauderdale JD. Mechanisms controlling Pax6 isoform expression in the retina have been conserved between teleosts and mammals. *Dev Biol*. 2007 Jul 1;307(2):498–520.
130. Kleinjan DA, Seawright A, Childs AJ, van Heyningen V. Conserved elements in Pax6 intron 7 involved in (auto)regulation and alternative transcription. *Dev Biol*. 2004 Jan 15;265(2):462–77.
131. The PAX6 gene homepage - Global Variome shared LOVD [Internet]. [cited 2024 Dec 8]. Available from: <https://databases.lovd.nl/shared/genes/PAX6>
132. Vincent MC, Pujo AL, Olivier D, Calvas P. Screening for PAX6 gene mutations is consistent with haploinsufficiency as the main mechanism leading to various ocular defects. *Eur J Hum Genet*. 2003 Feb 1;11(2):163–9.

133. Cross E, Duncan-Flavell PJ, Howarth RJ, Crooks RO, Thomas NS, Bunyan DJ. Screening of a large *PAX6* cohort identified many novel variants and emphasises the importance of the paired and homeobox domains. *European Journal of Medical Genetics*. 2020 Jul 1;63(7):103940.
134. Tzoulaki I, White IMS, Hanson IM. *PAX6* mutations: genotype-phenotype correlations. *BMC Genet*. 2005 May 26;6:27.
135. Isken O, Maquat LE. Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes Dev*. 2007 Aug 1;21(15):1833–56.
136. Kervestin S, Jacobson A. NMD: a multifaceted response to premature translational termination. *Nat Rev Mol Cell Biol*. 2012 Nov 1;13(11):700–12.
137. Hingorani M, Williamson KA, Moore AT, van Heyningen V. Detailed ophthalmologic evaluation of 43 individuals with *PAX6* mutations. *Invest Ophthalmol Vis Sci*. 2009 Jun 1;50(6):2581–90.
138. Mishra R, Gorlov IP, Chao LY, Singh S, Saunders GF. *PAX6*, paired domain influences sequence recognition by the homeodomain. *J Biol Chem*. 2002 Dec 20;277(51):49488–94.
139. Hanson I, Churchill A, Love J, Axton R, Moore T, Clarke M, et al. Missense mutations in the most ancient residues of the *PAX6* paired domain underlie a spectrum of human congenital eye malformations. *Hum Mol Genet*. 1999 Feb 1;8(2):165–72.
140. Nanjo Y, Kawasaki S, Mori K, Sotozono C, Inatomi T, Kinoshita S. A novel mutation in the alternative splice region of the *PAX6* gene in a patient with Peters' anomaly. *Br J Ophthalmol*. 2004 May 1;88(5):720–1.
141. Singh S, Chao LY, Mishra R, Davies J, Saunders GF. Missense mutation at the C-terminus of *PAX6* negatively modulates homeodomain function. *Hum Mol Genet*. 2001 Apr 15;10(9):911–8.
142. Crolla JA, van Heyningen V. Frequent chromosome aberrations revealed by molecular cytogenetic studies in patients with aniridia. *Am J Hum Genet*. 2002 Nov 1;71(5):1138–49.
143. Bobilev AM, McDougal ME, Taylor WL, Geisert EE, Netland PA, Lauderdale JD. Assessment of *PAX6* alleles in 66 families with aniridia. *Clin Genet*. 2015 Dec 10;89(6):n/a-n/a.
144. Robinson DO, Howarth RJ, Williamson KA, van Heyningen V, Beal SJ, Crolla JA. Genetic analysis of chromosome 11p13 and the *PAX6* gene in a series of 125 cases referred with aniridia. *Am J Med Genet A*. 2008 Mar 1;146A(5):558–69.

145. Simola KO, Knuutila S, Kaitila I, Pirkola A, Pohja P. Familial aniridia and translocation t(4;11)(q22;p13) without Wilms' tumor. *Hum Genet.* 1983 Jan 1;63(2):158–61.
146. Fantes J, Redeker B, Breen M, Boyle S, Brown J, Fletcher J, et al. Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype. *Hum Mol Genet.* 1995 Mar;4(3):415–22.
147. Daruich A, Robert MP, Leroy C, De vergnes N, Beugnet C, Malan V, et al. Foveal Hypoplasia Grading in 95 Cases of Congenital Aniridia: Correlation to Phenotype and PAX6 Genotype. *American Journal of Ophthalmology.* 2022 May 1;237:122–9.
148. Ansari M, Rainger J, Hanson IM, Williamson KA, Sharkey F, Harewood L, et al. Genetic Analysis of 'PAX6-Negative' Individuals with Aniridia or Gillespie Syndrome. Anderson MG, editor. *PLoS ONE.* 2016 Jan 1;11(4):e0153757.
149. Solomon BD, Pineda-Alvarez DE, Balog JZ, Hadley D, Gropman AL, Nandagopal R, et al. Compound heterozygosity for mutations in PAX6 in a patient with complex brain anomaly, neonatal diabetes mellitus, and microphthalmia. *Am J Med Genet A.* 2009 Nov 1;149A(11):2543–6.
150. Chao LY, Mishra R, Strong LC, Saunders GF. Missense mutations in the DNA-binding region and termination codon in PAX6. *Hum Mutat.* 2003 Feb 1;21(2):138–45.
151. Hever AM, Williamson KA, van Heyningen V. Developmental malformations of the eye: the role of PAX6, SOX2 and OTX2. *Clin Genet.* 2006 Jun 1;69(6):459–70.
152. Sale MM, Craig JE, Charlesworth JC, FitzGerald LM, Hanson IM, Dickinson JL, et al. Broad phenotypic variability in a single pedigree with a novel 1410delC mutation in the PST domain of the PAX6 gene. *Hum Mutat.* 2002 Oct 1;20(4):322–322.
153. Lima Cunha D, Owen N, Tailor V, Corton M, Theodorou M, Moosajee M. PAX6 missense variants in two families with isolated foveal hypoplasia and nystagmus: evidence of paternal postzygotic mosaicism. *Eur J Hum Genet.* 2021 Feb;29(2):349–55.
154. Deml B, Reis LM, Lemyre E, Clark RD, Kariminejad A, Semina EV. Novel mutations in PAX6, OTX2 and NDP in anophthalmia, microphthalmia and coloboma. *Eur J Hum Genet.* 2016 Apr 1;24(4):535–41.
155. Sadagopan KA, Liu GT, Capasso JE, Wuthisiri W, Keep RB, Levin AV. Anirdia-like phenotype caused by 6p25 dosage aberrations. *Am J Med Genet A.* 2015 Mar 1;167A(3):524–8.

156. Ito YA, Footz TK, Berry FB, Mirzayans F, Yu M, Khan AO, et al. Severe Molecular Defects of a Novel FOXC1 W152G Mutation Result in Aniridia. *Investigative Ophthalmology & Visual Science*. 2009 Aug 1;50(8):3573–9.
157. Gołaszewska K, Dub N, Saeed E, Mariak Z, Konopińska J. Axenfeld-Rieger syndrome combined with a foveal anomaly in a three-generation family: a case report. *BMC Ophthalmol*. 2021 Mar 29;21(1):154.
158. Perveen R, Lloyd IC, Clayton-Smith J, Churchill A, van Heyningen V, Hanson I, et al. Phenotypic variability and asymmetry of Rieger syndrome associated with PITX2 mutations. *Invest Ophthalmol Vis Sci*. 2000 Aug 1;41(9):2456–60.
159. Zucco J, Baldan F, Allegri L, Bregant E, Passon N, Franzoni A, et al. A bird's eye view on the use of whole exome sequencing in rare congenital ophthalmic diseases. *J Hum Genet*. 2024 Jun;69(6):271–82.
160. Reis LM, Maheshwari M, Capasso J, Atila H, Dudakova L, Thompson S, et al. Axenfeld-Rieger syndrome: more than meets the eye. *J Med Genet*. 2023 Apr;60(4):368–79.
161. Zazo-Seco C, Plaisancié J, Bitoun P, Corton M, Arteche A, Ayuso C, et al. Novel PXDN biallelic variants in patients with microphthalmia and anterior segment dysgenesis. *J Hum Genet*. 2020 May;65(5):487–91.
162. Reis LM, Seese SE, Costakos D, Semina EV. Congenital anterior segment ocular disorders: Genotype-phenotype correlations and emerging novel mechanisms. *Progress in Retinal and Eye Research*. 2024 Sep;102:101288.
163. Zhang X, Qin G, Chen G, Li T, Gao L, Huang L, et al. Variants in TRIM44 Cause Aniridia by Impairing PAX6 Expression. *Hum Mutat*. 2015 Dec 1;36(12):1164–7.
164. McEntagart M, Williamson KA, Rainger JK, Wheeler A, Seawright A, De Baere E, et al. A Restricted Repertoire of De Novo Mutations in ITPR1 Cause Gillespie Syndrome with Evidence for Dominant-Negative Effect. *Am J Hum Genet*. 2016 Apr 21;98(5):981–92.
165. Gerber S, Alzayady KJ, Burglen L, Brémond-Gignac D, Marchesin V, Roche O, et al. Recessive and Dominant De Novo ITPR1 Mutations Cause Gillespie Syndrome. *Am J Hum Genet*. 2016 May 5;98(5):971–80.
166. Hall HN, Bengani H, Hufnagel RB, Damante G, Ansari M, Marsh JA, et al. Monoallelic variants resulting in substitutions of MAB21L1 Arg51 Cause Aniridia and microphthalmia. Swaroop A, editor. *PLoS ONE*. 2022 Nov 22;17(11):e0268149.

167. Seese SE, Reis LM, Deml B, Griffith C, Reich A, Jamieson RV, et al. Identification of missense MAB21L1 variants in microphthalmia and aniridia. *Human Mutation*. 2021;42(7):877–90.
168. Rodríguez-López R, Pérez JMC, Balsera AM, Rodríguez GG, Moreno TH, García de Cáceres M, et al. The modifier effect of the BDNF gene in the phenotype of the WAGRO syndrome. *Gene*. 2013 Mar 10;516(2):285–90.
169. Marlin S, Couet D, Lacombe D, Cessans C, Bonneau D. Obesity: a new feature of WAGR (del 11p) syndrome. *Clin Dysmorphol*. 1994 Jul 1;3(3):255–7.
170. Hol JA, Jongmans MCJ, Sudour-Bonnange H, Ramírez-Villar GL, Chowdhury T, Rechnitzer C, et al. Clinical characteristics and outcomes of children with WAGR syndrome and Wilms tumor and/or nephroblastomatosis: The 30-year SIOP-RTSG experience. *Cancer*. 2021 Feb 15;127(4):628–38.
171. Fischbach BV, Trout KL, Lewis J, Luis CA, Sika M. WAGR syndrome: a clinical review of 54 cases. *Pediatrics*. 2005 Oct 1;116(4):984–8.
172. Koller A, Schlossmann J, Ashman K, Uttenweiler-Joseph S, Ruth P, Hofmann F. Association of phospholamban with a cGMP kinase signaling complex. *Biochem Biophys Res Commun*. 2003 Jan 3;300(1):155–60.
173. ITPR1 - Inositol 1,4,5-trisphosphate-gated calcium channel ITPR1 - Homo sapiens (Human) | UniProtKB | UniProt [Internet]. [cited 2024 Nov 26]. Available from: <https://www.uniprot.org/uniprotkb/Q14643/entry#interaction>
174. Milewicz DM, Ostergaard JR, Ala-Kokko LM, Khan N, Grange DK, Mendoza-Londono R, et al. De novo ACTA2 mutation causes a novel syndrome of multisystemic smooth muscle dysfunction. *Am J Med Genet A*. 2010 Oct 1;152A(10):2437–43.
175. Fantes JA, Bickmore WA, Fletcher JM, Ballesta F, Hanson IM, van Heyningen V. Submicroscopic deletions at the WAGR locus, revealed by nonradioactive in situ hybridization. *Am J Hum Genet*. 1992 Dec 1;51(6):1286–94.
176. Franzoni A, Russo PD, Baldan F, D'Elia AV, Puppin C, Penco S, et al. A CGH array procedure to detect PAX6 gene structural defects. *Mol Cell Probes*. 2017 Apr;32:65–8.
177. Blanco-Kelly F, Palomares M, Vallespín E, Villaverde C, Martín-Arenas R, Vélez-Monsalve C, et al. Improving molecular diagnosis of aniridia and WAGR syndrome using customized targeted array-based CGH. Lo AWI, editor. *PLoS ONE*. 2017 Feb 23;12(2):e0172363-13.
178. Regalado ES, Mellor-Crummey L, De Backer J, Braverman AC, Adès L, Benedict S, et al. Clinical history and management recommendations of the

smooth muscle dysfunction syndrome due to ACTA2 arginine 179 alterations. *Genet Med*. 2018 Jan 4;8:269.

179. Breslow NE, Norris R, Norkool PA, Kang T, Beckwith JB, Perlman EJ, et al. Characteristics and outcomes of children with the Wilms tumor-Aniridia syndrome: a report from the National Wilms Tumor Study Group. *J Clin Oncol*. 2003 Dec 15;21(24):4579–85.
180. Brok J, Pritchard-Jones K, Geller JI, Spreafico F. Review of phase I and II trials for Wilms' tumour - Can we optimise the search for novel agents? *Eur J Cancer*. 2017 Jul 1;79:205–13.
181. Roulez FMJ, Faes F, Delbeke P, Van Bogaert P, Rodesch G, De Zaeytijd J, et al. Congenital fixed dilated pupils due to ACTA2- multisystemic smooth muscle dysfunction syndrome. *J Neuroophthalmol*. 2014 Jun 1;34(2):137–43.
182. Genomics England Panelapp [Internet]. 2024 [cited 2024 Nov 17]. Structural eye disease (Version 4.0). Available from: <https://nhsgms-panelapp.genomicsengland.co.uk/panels/509/v4.0>
183. Gregory-Evans CY, Wang X, Wasan KM, Zhao J, Metcalfe AL, Gregory-Evans K. Postnatal manipulation of Pax6 dosage reverses congenital tissue malformation defects. *J Clin Invest*. 2014 Jan 1;124(1):111–6.
184. Wang X, Gregory-Evans K, Wasan KM, Sivak O, Shan X, Gregory-Evans CY. Efficacy of Postnatal In Vivo Nonsense Suppression Therapy in a Pax6 Mouse Model of Aniridia. *Molecular Therapy Nucleic Acids*. 2017 Jun 16;7:417–28.
185. EudraCT Number 2022-001013-39 - Clinical trial results - EU Clinical Trials Register [Internet]. 2022 [cited 2024 Nov 17]. Available from: <https://www.clinicaltrialsregister.eu/ctr-search/trial/2022-001013-39/results#moreInformationSection>
186. Kerem E, Konstan MW, De Boeck K, Accurso FJ, Sermet-Gaudelus I, Wilschanski M, et al. Ataluren for the treatment of nonsense-mutation cystic fibrosis: a randomised, double-blind, placebo-controlled phase 3 trial. 2014 Jul 1;2(7):539–47.
187. McDonald CM, Campbell C, Torricelli RE, Finkel RS, Flanigan KM, Goemans N, et al. Ataluren in patients with nonsense mutation Duchenne muscular dystrophy (ACT DMD): a multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet*. 2017 Sep 23;390(10101):1489–98.
188. Campbell C, Barohn RJ, Bertini E, Chabrol B, Comi GP, Darras BT, et al. Meta-analyses of ataluren randomized controlled trials in nonsense mutation Duchenne muscular dystrophy. *J Comp Eff Res*. 2020 Oct;9(14):973–84.

189. Lima Cunha D, Sarkar H, Eintracht J, Harding P, Zhou JH, Moosajee M. Restoration of functional PAX6 in aniridia patient iPSC-derived ocular tissue models using repurposed nonsense suppression drugs. *Mol Ther Nucleic Acids*. 2023 Sep 12;33:240–53.
190. Dorot O, Roux LN, Zennaro L, Oved K, Bremond-Gignac D, Pichinuk E, et al. The antipsychotropic drug Duloxetine rescues PAX6 haploinsufficiency of mutant limbal stem cells through inhibition of the MEK/ERK signaling pathway. *The Ocular Surface*. 2022 Jan;23:140–2.
191. Moustardas P, Abbasi M, Javidjam D, Asamoah CS, Schweitzer-Chaput A, Cisternino S, et al. Duloxetine enhances PAX6 expression and suppresses innate immune responses in murine LPS-induced corneal inflammation. *The Ocular Surface*. 2024 Oct 1;34:225–34.
192. Rabiee B, Anwar KN, Shen X, Putra I, Liu M, Jung R, et al. Gene dosage manipulation alleviates manifestations of hereditary PAX6 haploinsufficiency in mice. *Sci Transl Med*. 2020 Dec 9;12(573):eaaz4894.
193. Cole JD, McHaney KM, Rabiee B, Gao J, Rodriguez C, Miller DA, et al. Long-term retinal protection by MEK inhibition in Pax6 haploinsufficiency mice. *Exp Eye Res*. 2022 May;218:109012.
194. Mirjalili Mohanna SZ, Hickmott JW, Lam SL, Chiu NY, Lengyel TC, Tam BM, et al. Germline CRISPR/Cas9-Mediated Gene Editing Prevents Vision Loss in a Novel Mouse Model of Aniridia. *Mol Ther Methods Clin Dev*. 2020 Jun 12;17:478–90.
195. Williamson KA, Hall HN, Owen LJ, Livesey BJ, Hanson IM, Adams GGW, et al. Recurrent heterozygous PAX6 missense variants cause severe bilateral microphthalmia via predictable effects on DNA-protein interaction. *Genet Med*. 2019 Nov 8;20(11):1011–2.
196. Azuma N, Hotta Y, Tanaka H, Yamada M. Missense mutations in the PAX6 gene in aniridia. *Invest Ophthalmol Vis Sci*. 1998 Dec 1;39(13):2524–8.
197. Association WM. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. *JAMA*. 2013 Nov 27;310(20):2191–4.
198. Brown A, McKie M, van Heyningen V, Prosser J. The human PAX6 mutation database. *Nucleic Acids Research*. 1998 Jan 1;26(1):259–64.
199. Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. *Nature*. 2015 Mar 12;519(7542):223–8.

200. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020 May;581(7809):434–43.
201. Thomas MG, Kumar A, Mohammad S, Proudlock FA, Engle EC, Andrews C, et al. Structural grading of foveal hypoplasia using spectral-domain optical coherence tomography a predictor of visual acuity? *Ophthalmology*. 2011 Aug 1;118(8):1653–60.
202. Lagali N, Edén U, Utheim TP, Chen X, Riise R, Dellby A, et al. In Vivo Morphology of the Limbal Palisades of Vogt Correlates With Progressive Stem Cell Deficiency in Aniridia-Related Keratopathy. *Investigative Ophthalmology & Visual Science*. 2013 Aug 7;54(8):5333–42.
203. Core Team R. www.r-project.org. Vienna, Austria; 2016. R: A language and environment for statistical computing. Available from: www.r-project.org
204. PAX6 - Paired box protein Pax-6 - Homo sapiens (Human) | UniProtKB | UniProt [Internet]. [cited 2024 Dec 15]. Available from: <https://www.uniprot.org/uniprotkb/P26367/entry>
205. Morales J, Pujar S, Loveland JE, Astashyn A, Bennett R, Berry A, et al. A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. *Nature*. 2022 Apr 14;604(7905):310–5.
206. Gassmann M, Grenacher B, Rohde B, Vogel J. Quantifying Western blots: Pitfalls of densitometry. *Electrophoresis*. 2009 Jun 1;30(11):1845–55.
207. Miller LP. Analyzing western blots with Image Studio Lite [Internet]. lukemiller.org. 2013 [cited 2023 Oct 27]. Available from: <https://lukemiller.org/index.php/2013/02/analyzing-western-blots-with-image-studio-lite/>
208. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, et al. De novo gene disruptions in children on the autistic spectrum. *Neuron*. 2012 Apr 26;74(2):285–99.
209. Aota S ichi, Nakajima N, Sakamoto R, Watanabe S, Ibaraki N, Okazaki K. Pax6 autoregulation mediated by direct interaction of Pax6 protein with the head surface ectoderm-specific enhancer of the mouse Pax6 gene. *Developmental Biology*. 2003 May 1;257(1):1–13.
210. MRC Human Genetics Unit. Institute of Genetics and Cancer PAX6 Regulatory Browser [Internet]. [cited 2024 Dec 20]. Available from: <https://regulatorybrowser.igc.ed.ac.uk/>
211. Inoue M, Kamachi Y, Matsunami H, Imada K, Uchikawa M, Kondoh H. PAX6 and SOX2-dependent regulation of the Sox2 enhancer N-3 involved in

- embryonic visual system development. *Genes Cells*. 2007 Sep 1;12(9):1049–61.
212. Larouche K, Bergeron MJ, Leclerc S, Guerin SL. Optimization of competitor poly(dI-dC) poly(dI-dC) levels is advised in DNA-protein interaction studies involving enriched nuclear proteins. *BioTechniques*. 1996 Mar 1;20(3):439–.
213. Ragge NK, Lorenz B, Schneider A, Bushby K, de Sanctis L, de Sanctis U, et al. SOX2 anophthalmia syndrome. *Am J Med Genet A*. 2005 May 15;135(1):1-7-discussion 8.
214. Chassaing N, Causse A, Vigouroux A, Delahaye A, Alessandri JL, Boespflug-Tanguy O, et al. Molecular findings and clinical data in a cohort of 150 patients with anophthalmia/microphthalmia. *Clin Genet*. 2014 Oct;86(4):326–34.
215. McDonnell AF, Plech M, Livesey BJ, Gerasimavicius L, Owen LJ, Hall HN, et al. Deep mutational scanning quantifies DNA binding and predicts clinical outcomes of PAX6 variants. *Mol Syst Biol*. 2024 Jun 7;20(7):825–44.
216. Kamachi Y, Uchikawa M, Tanouchi A, Sekido R, Kondoh H. Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development. *Genes Dev*. 2001 May 15;15(10):1272–86.
217. Coutinho P, Pavlou S, Bhatia S, Chalmers KJ, Kleinjan DA, van Heyningen V. Discovery and assessment of conserved Pax6 target genes and enhancers. *Genome Res*. 2011 Aug 1;21(8):1349–59.
218. Narasimhan K, Pillay S, Huang YH, Jayabal S, Udayasuryan B, Veerapandian V, et al. DNA-mediated cooperativity facilitates the co-selection of cryptic enhancer sequences by SOX2 and PAX6 transcription factors. 2015 Feb 18;43(3):1513–28.
219. Xiao Y, Liu X, Yang C, Liu L, Guo X, Wang Q, et al. A Novel PAX6 Heterozygous Mutation Found in a Chinese Family with Congenital Aniridia. *Genet Test Mol Biomarkers*. 2019 Jul;23(7):495–500.
220. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021 Aug;596(7873):583–9.
221. Cheng J, Novati G, Pan J, Bycroft C, Žemgulytė A, Applebaum T, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science*. 2023 Sep 19;381(6664):eadg7492.
222. Jagota M, Ye C, Albors C, Rastogi R, Koehl A, Ioannidis N, et al. Cross-protein transfer learning substantially improves disease variant prediction. *Genome Biol*. 2023 Aug 7;24(1):182.

223. Livesey BJ, Marsh JA. Updated benchmarking of variant effect predictors using deep mutational scanning. *Molecular Systems Biology*. 2023 Aug 8;19(8):e11474.
224. Andhika NS, Biswas S, Hardcastle C, Green DJ, Ramsden SC, Birney E, et al. Using computational approaches to enhance the interpretation of missense variants in the PAX6 gene. *Eur J Hum Genet*. 2024 Aug;32(8):1005–13.
225. Livesey BJ, Marsh JA. Variant effect predictor correlation with functional assays is reflective of clinical classification performance [Internet]. *bioRxiv*; 2024 [cited 2024 Nov 27]. p. 2024.05.12.593741. Available from: <https://www.biorxiv.org/content/10.1101/2024.05.12.593741v1>
226. Frazer J, Notin P, Dias M, Gomez A, Min JK, Brock K, et al. Disease variant prediction with deep generative models of evolutionary data. *Nature*. 2021 Nov;599(7883):91–5.
227. Li C, Zhi D, Wang K, Liu X. MetaRNN: differentiating rare pathogenic and rare benign missense SNVs and InDels using deep learning. *Genome Medicine*. 2022 Oct 8;14(1):115.
228. Feng BJ. PERCH: A Unified Framework for Disease Gene Prioritization. *Hum Mutat*. 2017 Mar;38(3):243–51.
229. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *The American Journal of Human Genetics*. 2016 Oct 6;99(4):877–85.
230. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. *Nat Protoc*. 2016 Jan;11(1):1–9.
231. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014 Mar;46(3):310–5.
232. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010 Apr;7(4):248–9.
233. Livesey BJ, Marsh JA. Interpreting protein variant effects with computational predictors and deep mutational scanning. *Disease Models & Mechanisms*. 2022 Jun 1;15(6):dmm049510.
234. Bruun JA, Thomassen EIS, Kristiansen K, Tylden G, Holm T, Mikkola I, et al. The third helix of the homeodomain of paired class homeodomain proteins acts as a recognition helix both for DNA and protein interactions. *Nucleic Acids Res*. 2005;33(8):2661–75.

235. Pingault V, Ente D, Dastot-Le Moal F, Goossens M, Marlin S, Bondurand N. Review and update of mutations causing Waardenburg syndrome. *Hum Mutat.* 2010 Apr;31(4):391–406.
236. Leslie EJ, Taub MA, Liu H, Steinberg KM, Koboldt DC, Zhang Q, et al. Identification of functional variants for cleft lip with or without cleft palate in or near PAX7, FGFR2, and NOG by targeted sequencing of GWAS loci. *Am J Hum Genet.* 2015 Mar 5;96(3):397–411.
237. Azuma N, Nishina S, Yanagisawa H, Okuyama T, Yamada M. PAX6 missense mutation in isolated foveal hypoplasia. *Nat Genet.* 1996 Jun 1;13(2):141–2.
238. Baird SE, Fitch DHA, Kassem IAA, Emmons SW. Pattern formation in the nematode epidermis: determination of the arrangement of peripheral sense organs in the *C. elegans* male tail. *Development.* 1991 Oct 1;113(2):515–26.
239. Mariani M, Corradi A, Baldessari D, Margaretti N, Pozzoli O, Fesce R, et al. Mab21, the mouse homolog of a *C. elegans* cell-fate specification gene, participates in cerebellar, midbrain and eye development. *Mech Dev.* 1998 Dec 1;79(1–2):131–5.
240. Wong RLY, Chow KL. Depletion of Mab21l1 and Mab21l2 messages in mouse embryo arrests axial turning, and impairs notochord and neural tube differentiation. *Teratology.* 2002 Feb 1;65(2):70–7.
241. Chen Q, Sun L, Chen ZJ. Regulation and function of the cGAS–STING pathway of cytosolic DNA sensing. *Nat Immunol.* 2016 Oct;17(10):1142–9.
242. Kranzusch PJ, Lee ASY, Berger JM, Doudna JA. Structure of Human cGAS Reveals a Conserved Family of Second-Messenger Enzymes in Innate Immunity. *Cell Rep.* 2013 May 30;3(5):1362–8.
243. Hartmann R, Norby PL, Martensen PM, Jorgensen P, James MC, Jacobsen C, et al. Activation of 2'-5' oligoadenylate synthetase by single-stranded and double-stranded RNA aptamers. *J Biol Chem.* 1998 Feb 6;273(6):3236–46.
244. Kranzusch PJ. cGAS and CD-NTase enzymes: structure, mechanism, and evolution. *Curr Opin Struct Biol.* 2019 Dec 1;59:178–87.
245. de Oliveira Mann CC, Kiefersauer R, Witte G, Hopfner KP. Structural and biochemical characterization of the cell fate determining nucleotidyltransferase fold protein MAB21L1. *Sci Rep.* 2016 Jan 1;6:27498.
246. Huang Z, Xiang JW, Zhou L, Nie Q, Wang L, Chen Z, et al. The Male Abnormal Gene Family 21 (Mab21) Members Regulate Eye Development. *Curr Mol Med [Internet].* 2016 Aug 24; Available from:

<http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=27558071&retmode=ref&cmd=prlinks>

247. Yamada R, Oguri A, Fujiki K, Shirahige K, Hirate Y, Kanai-Azuma M, et al. MAB21L1 modulates gene expression and DNA metabolic processes in the lens placode. *Dis Model Mech*. 2021 Dec 23;14(12):dmm049251.
248. Yamada R, Mizutani-Koseki Y, Hasegawa T, Osumi N, Koseki H, Takahashi N. Cell-autonomous involvement of Mab21l1 is essential for lens placode development. *Development*. 2003 May 1;130(9):1759–70.
249. Tsang WH, Shek KF, Lee TY, Chow KL. An evolutionarily conserved nested gene pair - Mab21 and Lrba/Nbea in metazoan. *Genomics*. 2009 Sep 1;94(3):177–87.
250. Rad A, Altunoglu U, Miller R, Maroofian R, James KN, Çağlayan AO, et al. MAB21L1 loss of function causes a syndromic neurodevelopmental disorder with distinctive cerebellar, ocular, craniofacial and genital features (COFG syndrome). *J Med Genet*. 2019 May 1;56(5):332–9.
251. Bruel AL, Masurel-Paulet A, Rivière JB, Duffourd Y, Lehalle D, Bensignor C, et al. Autosomal recessive truncating MAB21L1 mutation associated with a syndromic scrotal agenesis. *Clin Genet*. 2017 Feb 1;91(2):333–8.
252. Rainger J, Pehlivan D, Johansson S, Bengani H, Sanchez-Pulido L, Williamson KA, et al. Monoallelic and biallelic mutations in MAB21L2 cause a spectrum of major eye malformations. *Am J Hum Genet*. 2014 Jun 5;94(6):915–23.
253. Deml B, Kariminejad A, Borujerdi RHR, Muheisen S, Reis LM, Semina EV. Mutations in MAB21L2 result in ocular Coloboma, microcornea and cataracts. Barsh GS, editor. *PLoS Genet*. 2015 Jan 1;11(2):e1005002.
254. Yamada R, Mizutani-Koseki Y, Koseki H, Takahashi N. Requirement for Mab21l2 during development of murine retina and ventral body wall. *Dev Biol*. 2004 Oct 15;274(2):295–307.
255. NC3Rs [Internet]. [cited 2023 Dec 14]. Available from: <https://www.nc3rs.org.uk/>
256. UK10K Consortium, Walter K, Min JL, Huang J, Crooks L, Memari Y, et al. The UK10K project identifies rare variants in health and disease. *Nature*. 2015 Oct 1;526(7571):82–90.
257. Sayers EW, Cavanaugh M, Clark K, Ostell J, Pruitt KD, Karsch-Mizrachi I. GenBank. *Nucleic Acids Res*. 2019 Jan 8;47(D1):D94–9.

258. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc.* 2013 Nov 1;8(11):2281–308.
259. Jadeja S, Barnard AR, McKie L, Cross SH, White JK, Sanger Mouse Genetics Project, et al. Mouse *slc9a8* mutants exhibit retinal defects due to retinal pigmented epithelium dysfunction. *Invest Ophthalmol Vis Sci.* 2015 May 1;56(5):3015–26.
260. Perez-Rovira A, MacGillivray T, Trucco E, Chin KS, Zutis K, Lupascu C, et al. VAMPIRE: Vessel assessment and measurement platform for images of the REtina. *Annu Int Conf IEEE Eng Med Biol Soc.* 2011 Jan 1;2011:3391–4.
261. Trucco E, Ballerini L, for DR and R, 2013. Novel VAMPIRE algorithms for quantitative analysis of the retinal vasculature. *ieeexplore.ieee.org.* 2013;
262. Cross SH, McKie L, Hurd TW, Riley S, Wills J, Barnard AR, et al. The nanophthalmos protein TMEM98 inhibits MYRF self-cleavage and is required for eye size specification. *PLOS Genetics.* 2020 Apr 1;16(4):e1008583.
263. Findlay AS, McKie L, Keighren M, Clementson-Mobbs S, Sanchez-Pulido L, Wells S, et al. *Fam151b*, the mouse homologue of *C.elegans* *menorin* gene, is essential for retinal function. *Sci Rep.* 2020 Jan 16;10(1):437.
264. Delgado J, Radusky LG, Cianferoni D, Serrano L. FoldX 5.0: working with RNA, small molecules and a new graphical interface. *Bioinformatics.* 2019 Oct 15;35(20):4168–9.
265. Gerasimavicius L, Livesey BJ, Marsh JA. Loss-of-function, gain-of-function and dominant-negative mutations have profoundly different effects on protein structure. *Nat Commun.* 2022 Jul 6;13(1):3895.
266. Marsh JA, Teichmann SA. Predicting pathogenic protein variants. *Science.* 2023 Sep 22;381(6664):1284–5.
267. Brandt T, Sack LM, Arjona D, Tan D, Mei H, Cui H, et al. Adapting ACMG/AMP sequence variant classification guidelines for single-gene copy number variants. *GENETICS in MEDICINE.* 2020 Jan 28;22(2):1–9.
268. Guidelines for Nomenclature of Mouse and Rat Strains (section 5.1) [Internet]. 2022 [cited 2023 Jun 1]. Available from: <https://www.informatics.jax.org/mgihome/nomen/strains.shtml#coisogenic>
269. Wang P, Wu P, Wang J, Zeng Y, Jiang Y, Wang Y, et al. Missense Mutations in *MAB21L1*: Causation of Novel Autosomal Dominant Ocular AMD Syndrome. *Investigative Ophthalmology & Visual Science.* 2023 Mar 9;64(3):19.

270. Meng F, Li X, Zhang J, Gao Z, Yang X, Liu Z, et al. Single amino acid variation in MAB21L1 is dominantly associated with congenital eye defects. *Journal of Medical Genetics*. 2023 Jul 1;60(7):655–61.
271. Yamada R, Mizutani-Koseki Y, Hasegawa T, Osumi N, Koseki H, Takahashi N. Cell-autonomous involvement of *Mab21l1* is essential for lens placode development. *Development*. 2003 May 1;130(9):1759–70.
272. Park SM, Deering RP, Lu Y, Tivnan P, Lianoglou S, Al-Shahrour F, et al. Musashi-2 controls cell fate, lineage bias, and TGF- β signaling in HSCs. *Journal of Experimental Medicine*. 2014 Jan 6;211(1):71–87.
273. Sutherland JM, McLaughlin EA, Hime GR, Siddall NA. The Musashi Family of RNA Binding Proteins: Master Regulators of Multiple Stem Cell Populations. In: Hime G, Abud H, editors. *Transcriptional and Translational Regulation of Stem Cells* [Internet]. Dordrecht: Springer Netherlands; 2013 [cited 2024 Jul 25]. p. 233–45. Available from: https://doi.org/10.1007/978-94-007-6621-1_13
274. Sundar J, Matalkah F, Jeong B, Stoilov P, Ramamurthy V. The Musashi proteins MSI1 and MSI2 are required for photoreceptor morphogenesis and vision in mice. *Journal of Biological Chemistry*. 2021 Jan 1;296:100048.
275. Cox JL, Wilder PJ, Gilmore JM, Wuebben EL, Washburn MP, Rizzino A. The SOX2-Interactome in Brain Cancer Cells Identifies the Requirement of MSI2 and USP9X for the Growth of Brain Tumor Cells. *PLOS ONE*. 2013 May 7;8(5):e62857.
276. Hall HN, Parry D, Halachev M, Williamson KA, Donnelly K, Campos Parada J, et al. Short-read whole genome sequencing identifies causative variants in most individuals with previously unexplained aniridia. *J Med Genet*. 2023 Nov 30;jmg-2023-109181.
277. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care — Preliminary Report. *New England Journal of Medicine*. 2021 Nov 11;385(20):1868–80.
278. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM [Internet]. arXiv; 2013 [cited 2022 Sep 7]. Available from: <http://arxiv.org/abs/1303.3997>
279. Faust GG, Hall IM. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics*. 2014 Sep 1;30(17):2503–5.
280. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics*. 2013;43(1):11.10.1-11.10.33.

281. Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Auwera GAV der, et al. Scaling accurate genetic variant discovery to tens of thousands of samples [Internet]. *bioRxiv*; 2018 [cited 2022 Sep 7]. p. 201178. Available from: <https://www.biorxiv.org/content/10.1101/201178v3>
282. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinformatics*. 2013 Mar 1;14(2):178–92.
283. Canvas Copy Number Variant Caller [Internet]. Illumina; 2022 [cited 2022 Nov 30]. Available from: <https://github.com/Illumina/canvas>
284. Ivakhno S, Roller E, Colombo C, Tedder P, Cox AJ. Canvas SPW: calling de novo copy number variants in pedigrees. *Bioinformatics*. 2018 Feb 1;34(3):516–8.
285. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*. 2016 Apr 15;32(8):1220–2.
286. Manta Structural Variant Caller [Internet]. Illumina; 2022 [cited 2022 Nov 30]. Available from: <https://github.com/Illumina/manta>
287. Parry DA. VASE [Internet]. 2022 [cited 2022 Nov 30]. Available from: <https://github.com/david-a-parry/vase>
288. Jaganathan K, Panagiotopoulou SK, McRae JF, Darbandi SF, Knowles D, Li YI, et al. Predicting Splicing from Primary Sequence with Deep Learning. *Cell*. 2019 Jan 24;176(3):535-548.e24.
289. Lenassi E, Carvalho A, Thormann A, Fletcher T, Hardcastle C, Hunt SE, et al. EyeG2P: an automated variant filtering approach improves efficiency of diagnostic genomic testing for inherited ophthalmic disorders [Internet]. *medRxiv*; 2021 [cited 2022 Sep 7]. p. 2021.07.23.21261017. Available from: <https://www.medrxiv.org/content/10.1101/2021.07.23.21261017v1>
290. Thormann A, Halachev M, McLaren W, Moore DJ, Svinti V, Campbell A, et al. Flexible and scalable diagnostic filtering of genomic variants using G2P with Ensembl VEP. *Nat Commun*. 2019 Dec;10(1):2373.
291. EMBL-EBI. Gene2Phenotype [Internet]. [cited 2018 Aug 29]. Available from: <https://www.ebi.ac.uk/gene2phenotype/>
292. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Current Protocols in Bioinformatics*. 2014;47(1):11.12.1-11.12.34.
293. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint

- consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*. 2015 May;17(5):405–24.
294. Taviglian SV, Greenblatt MS, Harrison SM, Nussbaum RL, Prabhu SA, Boucher KM, et al. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genetics in Medicine*. 2018 Sep;20(9):1054–60.
 295. Freeman PJ, Hart RK, Gretton LJ, Brookes AJ, Dalglish R. VariantValidator: Accurate validation, mapping, and formatting of sequence variation descriptions. *Human Mutation*. 2018;39(1):61–8.
 296. Chomczynski P, Sacchi N. The single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction: twenty-something years on. *Nat Protoc*. 2006 Aug;1(2):581–5.
 297. Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. Variant Review with the Integrative Genomics Viewer. *Cancer Research*. 2017 Oct 31;77(21):e31–4.
 298. Roller E, Ivakhno S, Lee S, Royce T, Tanner S. Canvas: versatile and scalable detection of copy number variants. *Bioinformatics*. 2016 Aug 1;32(15):2375–7.
 299. Ma A, Yousoof S, Grigg JR, Flaherty M, Minoche AE, Cowley MJ, et al. Revealing hidden genetic diagnoses in the ocular anterior segment disorders. *Genet Med*. 2020;22(10):1623–32.
 300. Sanchez-Castro M, Eldjouzi H, Charpentier E, Busson PF, Hauet Q, Lindenbaum P, et al. Search for Rare Copy-Number Variants in Congenital Heart Defects Identifies Novel Candidate Genes and a Potential Role for FOXC1 in Patients With Coarctation of the Aorta. *Circ Cardiovasc Genet*. 2016 Feb;9(1):86–94.
 301. Ovaert C, Busa T, Faure E, Missirian C, Philip N, Paoli F, et al. FOXC1 haploinsufficiency due to 6p25 deletion in a patient with rapidly progressing aortic valve disease. *Am J Med Genet A*. 2017 Sep;173(9):2489–93.
 302. Bayrakli F, Guney I, Bayri Y, Ercan-Sencicek AG, Ceyhan D, Cankaya T, et al. A novel heterozygous deletion within the 3' region of the PAX6 gene causing isolated aniridia in a large family group. *Journal of Clinical Neuroscience*. 2009 Dec 1;16(12):1610–4.
 303. Schmitt AD, Hu M, Jung I, Xu Z, Qiu Y, Tan CL, et al. A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Reports*. 2016 Nov;17(8):2042–59.

304. Wang X, Wei X, Thijssen B, Das J, Lipkin SM, Yu H. Three-dimensional reconstruction of protein networks provides insight into human genetic disease. *Nat Biotechnol.* 2012 Jan 15;30(2):159–64.
305. Gould DB, Walter MA. Cloning, characterization, localization, and mutational screening of the human BARX1 gene. *Genomics.* 2000 Sep 15;68(3):336–42.
306. Reig G, Cabrejos ME, Concha ML. Functions of BarH transcription factors during embryonic development. *Dev Biol.* 2007 Feb 15;302(2):367–75.
307. Lappalainen I, Lopez J, Skipper L, Hefferon T, Spalding JD, Garner J, et al. dbVar and DGVa: public archives for genomic structural variation. *Nucleic Acids Research.* 2013 Jan 1;41(D1):D936–41.
308. You B, Zhang X, Xu K, Xie Y, Ye H, Li Y. Mutation spectrum of PAX6 and clinical findings in 95 Chinese patients with aniridia. *Mol Vis.* 2020;26:226–34.
309. Axton RA, Hanson IM, Love J, Seawright A, Prosser J, van Heyningen V. Combined SSCP/heteroduplex analysis in the screening for PAX6 mutations. *Mol Cell Probes.* 1997 Aug 1;11(4):287–92.
310. Vasilyeva T a., Voskresenskaya A a., Käsmann-Kellner B, Khlebnikova O v., Pozdeyeva N a., Bayazutdinova G m., et al. Molecular analysis of patients with aniridia in Russian Federation broadens the spectrum of PAX6 mutations. *Clinical Genetics.* 2017;92(6):639–44.
311. Update of the Best Practice Guidelines for Variant Classification Guidelines in Rare Disease 2020, v4 - The Association for Clinical Genomic Science. 2020 Feb 4; Available from: <https://www.acgs.uk.com/news/update-of-the-best-practice-guidelines-for-variant-classification-guidelines-in-rare-disease-2020-v4/>
312. Grønskov K, Rosenberg T, Sand A, Brøndum-Nielsen K. Mutational analysis of PAX6: 16 novel mutations including 5 missense mutations with a mild aniridia phenotype. *Eur J Hum Genet.* 1999 Apr;7(3):274–86.
313. Pedersen HR, Baraas RC, Landsend ECS, Utheim ØA, Utheim TP, Gilson SJ, et al. PAX6 Genotypic and Retinal Phenotypic Characterization in Congenital Aniridia. *Investigative Ophthalmology & Visual Science.* 2020 May 12;61(5):14.
314. Tamayo A, Núñez-Moreno G, Ruiz C, Plaisancie J, Damian A, Moya J, et al. Minigene Splicing Assays and Long-Read Sequencing to Unravel Pathogenic Deep-Intronic Variants in PAX6 in Congenital Aniridia. *Int J Mol Sci.* 2023 Jan 13;24(2):1562.

315. Rowlands CF, Taylor A, Rice G, Whiffin N, Hall HN, Newman WG, et al. MRSD: A quantitative approach for assessing suitability of RNA-seq in the investigation of mis-splicing in Mendelian disease. *The American Journal of Human Genetics*. 2022 Feb;109(2):210–22.
316. Pedersen BS, Brown JM, Dashnow H, Wallace AD, Velinder M, Tristani-Firouzi M, et al. Effective variant filtering and expected candidate variant yield in studies of rare human disease. *npj Genom Med*. 2021 Jul 15;6(1):1–8.
317. Ando H, Mizutani A, Mikoshiba K. An IRBIT homologue lacks binding activity to inositol 1,4,5-trisphosphate receptor due to the unique N-terminal appendage. *Journal of Neurochemistry*. 2009;109(2):539–50.
318. Micheal S, Siddiqui SN, Zafar SN, Venselaar H, Qamar R, Khan MI, et al. Whole exome sequencing identifies a heterozygous missense variant in the PRDM5 gene in a family with Axenfeld-Rieger syndrome. *Neurogenetics*. 2016 Jan 1;17(1):17–23.
319. Khan AO, Aldahmesh MA, Alkuraya FS. Genetic and genomic analysis of classic aniridia in Saudi Arabia. *Mol Vis*. 2011 Jan 1;17:708–14.
320. Grainger RM, Lauderdale JD, Collins JL, Trout KL, McCullen Krantz S, Wolfe SS, et al. Report on the 2021 Aniridia North America symposium on PAX6, aniridia, and beyond. *The Ocular Surface*. 2023 Jul 1;29:423–31.
321. Benjamini Y, Speed TP. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res*. 2012 May;40(10):e72.
322. Damián A, Núñez-Moreno G, Jubin C, Tamayo A, de Alba MR, Villaverde C, et al. Long-read genome sequencing identifies cryptic structural variants in congenital aniridia cases. *Hum Genomics*. 2023 Jun 2;17(1):45.
323. Bhatia S, Monahan J, Ravi V, Gautier P, Murdoch E, Brenner S, et al. A survey of ancient conserved non-coding elements in the PAX6 locus reveals a landscape of interdigitated cis-regulatory archipelagos. *Dev Biol*. 2014 Mar 15;387(2):214–28.
324. Jobanputra V, Schroeder B, Rehm HL, Shen W, Spiteri E, Nakouzi G, et al. Advancing access to genome sequencing for rare genetic disorders: recent progress and call to action. *npj Genom Med*. 2024 Mar 27;9(1):1–3.
325. Wetterstrand KA. DNA Sequencing Costs: Data [Internet]. [cited 2024 Dec 12]. Available from: <https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>
326. 100,000 Genomes Project Pilot Investigators, Smedley D, Smith KR, Martin A, Thomas EA, McDonagh EM, et al. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care - Preliminary Report. *N Engl J Med*. 2021 Nov 11;385(20):1868–80.

327. Turro E, Astle WJ, Megy K, Gräf S, Greene D, Shamardina O, et al. Whole-genome sequencing of patients with rare diseases in a national health system. *Nature*. 2020 Jul;583(7814):96–102.
328. Chaisson MJP, Sanders AD, Zhao X, Malhotra A, Porubsky D, Rausch T, et al. Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nat Commun*. 2019 Apr 16;10(1):1784.
329. Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, et al. Segmental Duplications and Copy-Number Variation in the Human Genome. *The American Journal of Human Genetics*. 2005 Jul 1;77(1):78–88.
330. Oehler JB, Wright H, Stark Z, Mallett AJ, Schmitz U. The application of long-read sequencing in clinical settings. *Human Genomics*. 2023 Aug 8;17(1):73.
331. Fernández-Suárez E, González-Del Pozo M, Méndez-Vidal C, Martín-Sánchez M, Mena M, de la Morena-Barrio B, et al. Long-read sequencing improves the genetic diagnosis of retinitis pigmentosa by identifying an Alu retrotransposon insertion in the EYS gene. *Mob DNA*. 2024 May 4;15(1):9.
332. Chandrasekhar S, Lin S, Jurkute N, Oprych K, Estramiana Elorrieta L, Schiff E, et al. Investigating Splice Defects in USH2A Using Targeted Long-Read Sequencing. *Cells*. 2024 Jul 26;13(15):1261.
333. Conlin LK, Aref-Eshghi E, McEldrew DA, Luo M, Rajagopalan R. Long-read sequencing for molecular diagnostics in constitutional genetic disorders. *Human mutation*. 2022 Sep 18;43(11):1531.
334. Olivucci G, Iovino E, Innella G, Turchetti D, Pippucci T, Magini P. Long read sequencing on its way to the routine diagnostics of genetic diseases. *Front Genet*. 2024 Mar 6;15:1374860.
335. Maroille T, Tarailo-Graovac M. Uncovering Missing Heritability in Rare Diseases. *Genes*. 2019 Apr 4;10(4):275.
336. Harding P, Lima Cunha D, Méjécase C, Eintracht J, Tualbi L, Sarkar H, et al. Generation of human iPSC line (UCLi013-A) from a patient with microphthalmia and aniridia, carrying a heterozygous missense mutation c.372C>A p.(Asn124Lys) in *PAX6*. *Stem Cell Research*. 2021 Mar 1;51:102184.
337. Guo H, Dai L, Huang Y, Liao Q, Bai Y. A Large Novel Deletion Downstream of *PAX6* Gene in a Chinese Family with Ocular Coloboma. *PLOS ONE*. 2013 Dec 11;8(12):e83073.
338. Kleinjan DA, Seawright A, Schedl A, Quinlan RA, Danes S, van Heyningen V. Aniridia-associated translocations, DNase hypersensitivity,

- sequence comparison and transgenic analysis redefine the functional domain of PAX6. *Hum Mol Genet*. 2001 Sep 15;10(19):2049–59.
339. Kleinjan DA, Seawright A, Mella S, Carr CB, Tyas DA, Simpson TI, et al. Long-range downstream enhancers are essential for Pax6 expression. *Dev Biol*. 2006 Nov 15;299(2):563–81.
340. Uttley K, Papanastasiou AS, Lahne M, Brisbane JM, MacDonald RB, Bickmore WA, et al. Unique activities of two overlapping PAX6 retinal enhancers. *Life Sci Alliance*. 2023 Nov;6(11):e202302126.
341. Vasilyeva TA, Marakhonov AV, Voskresenskaya AA, Kadyshev VV, Käsmann-Kellner B, Sukhanova NV, et al. Analysis of genotype–phenotype correlations in PAX6-associated aniridia. *Journal of Medical Genetics*. 2021 Apr 1;58(4):270–4.
342. Wright CF, FitzPatrick DR, Firth HV. Paediatric genomics: diagnosing rare disease in children. *Nat Rev Genet*. 2018 May;19(5):253–68.
343. Weile J. Multiplexed assays of variant effects contribute to a growing genotype–phenotype atlas. *Hum Genet*. 2018 Sep 11;137(9):665–78.
344. McDonnell AF, Plech M, Livesey BJ, Gerasimavicius L, Owen LJ, Hall HN, et al. Deep mutational scanning quantifies DNA binding and predicts clinical outcomes of PAX6 variants [Internet]. *bioRxiv*; 2023 [cited 2023 Nov 12]. p. 2023.07.25.550478. Available from: <https://www.biorxiv.org/content/10.1101/2023.07.25.550478v1>
345. Consortium TCA of GI. CAGI, the Critical Assessment of Genome Interpretation, establishes progress and prospects for computational genetic variant interpretation methods. *Genome Biol*. 2024 Feb 22;25(1):53.

Appendices

Appendix A Supplemental material for PAX6 chapter

- A.1 Publication 'Recurrent heterozygous PAX6 missense variants cause severe bilateral microphthalmia via predictable effects on DNA-protein interaction'



Recurrent heterozygous *PAX6* missense variants cause severe bilateral microphthalmia via predictable effects on DNA–protein interaction

Kathleen A. Williamson, PhD¹, H. Nikki Hall, MBChB, FRCOphth¹, Liusaidh J. Owen, BSc¹, Benjamin J. Livesey, MSc¹, Isabel M. Hanson, PhD¹, G. G. W. Adams, MBChB, FRCOphth², Simon Bodek, MBChB³, Patrick Calvas, MD⁴, Bruce Castle, MD⁵, Michael Clarke, MBChB, FRCOphth⁶, Alexander T. Deng, BMBCh⁷, Patrick Edery, MD⁸, Richard Fisher, MD⁹, Gabriele Gillissen-Kaesbach, Prof. Dr. med.¹⁰, Elise Heon, MD¹¹, Jane Hurst, MD¹², Dragana Josifova, MD⁷, Birgit Lorenz, MD, PhD¹³, Shane McKee, MD¹⁴, Francoise Meire, Prof. PhD¹⁵, Anthony T. Moore, BMBCh FRCOphth², Michael Parker, MD¹⁶, Charlotte M. Reiff, PD Dr. med.¹⁷, Jay Self, FRCOphth PhD^{18,19}, Edward S. Tobias, PhD, FRCP²⁰, Joke B. G. M. Verheij, MD²¹, Marjolaine Willems, MD²², Denise Williams, MD²³, Veronica van Heyningen, DPhil¹, Joseph A. Marsh, PhD¹ and David R. FitzPatrick, MD¹ 

Purpose: Most classical aniridia is caused by *PAX6* haploinsufficiency. *PAX6* missense variants can be hypomorphic or mimic haploinsufficiency. We hypothesized that missense variants also cause previously undescribed disease by altering the affinity and/or specificity of *PAX6* genomic interactions.

Methods: We screened *PAX6* in 372 individuals with bilateral microphthalmia, anophthalmia, or coloboma (MAC) from the Medical Research Council Human Genetics Unit eye malformation cohort (HGU^{eye}) and reviewed data from the Deciphering Developmental Disorders study. We performed cluster analysis on *PAX6*-associated ocular phenotypes by variant type and molecular modeling of the structural impact of 86 different *PAX6* causative missense variants.

Results: Eight different *PAX6* missense variants were identified in 17 individuals (15 families) with MAC, accounting for 4% (15/372) of our cohort. Seven altered the paired domain (p.[Arg26Gln]x1, p.[Gly36Val]x1, p.[Arg38Trp]x2, p.[Arg38Gln]x1, p.[Gly51Arg]x2,

p.[Ser54Arg]x2, p.[Asn124Lys]x5) and one the homeodomain (p.[Asn260Tyr]x1). p.Ser54Arg and p.Asn124Lys were exclusively associated with severe bilateral microphthalmia. MAC-associated variants were predicted to alter but not ablate DNA interaction, consistent with the electrophoretic mobility shifts observed using mutant paired domains with well-characterized *PAX6*-binding sites. We found no strong evidence for novel *PAX6*-associated extraocular disease.

Conclusion: Altering the affinity and specificity of *PAX6*-binding genome-wide provides a plausible mechanism for the worse-than-null effects of MAC-associated missense variants.

Genetics in Medicine (2020) 22:598–609; <https://doi.org/10.1038/s41436-019-0685-9>

Keywords: *PAX6*; missense variant; paired domain; microphthalmia; protein destabilization

INTRODUCTION

Classical aniridia (MIM 106210) is typified by absence or severe hypoplasia of the iris associated with cataracts, keratopathy, glaucoma, foveal hypoplasia, and optic nerve malformations, and absence or hypoplasia of the anterior commissure, olfactory bulb, and pineal gland.¹ It is almost exclusively caused by genomic variants that result in haploinsufficiency of *PAX6*, by gene deletion, *cis*-regulatory variant, or likely/presumed gene disruptive (LGD) variants. *PAX6* missense variants may mimic haploinsufficiency but are more commonly associated with milder phenotypes, such as

partial aniridia, ectopia pupillae (MIM 129750), isolated foveal hypoplasia (MIM 136520), Peters anomaly (MIM 604229), or optic disc malformations.^{2–4} Compound heterozygous LGD variants of *PAX6* are a rare cause of syndromic bilateral anophthalmia (absent eyes) associated with severe brain and craniofacial defects.^{5,6}

PAX6 (MIM 607108) encodes a transcription factor essential for the correct development of the eye, central nervous system, and pancreas, and for maintenance of these tissues in adulthood.^{7–10} *PAX6* binds DNA via two highly conserved DNA-binding domains: a homeodomain (HD) and a

Correspondence: David R. FitzPatrick (david.fitzpatrick@ed.ac.uk). Affiliations are listed at the end of the paper.

These authors contributed equally: Kathleen A. Williamson, H. Nikki Hall.

These senior authors contributed equally: Joseph A. Marsh, David R. FitzPatrick.

Submitted 24 July 2019; revised 14 October 2019; accepted: 15 October 2019

Published online: 8 November 2019

structurally bipartite paired domain (PD) with N-terminal (NTS) and C-terminal (CTS) subdomains, each containing three α -helices that incorporate a helix–turn–helix motif. Each subdomain can independently or cooperatively bind DNA, and can modulate the DNA-binding properties of the HD.¹¹

The diverse range of sequences that mediate binding of PAX6 to the genome have been defined using chromatin immunoprecipitation (ChIP)¹² and, subsequently, hidden Markov modeling using experimentally derived binding sites with evolutionary conservation of the bound sequences to enrich for truly functional elements.¹³ These approaches have identified between 600 (ref. ¹³) and ~2500 (ref. ¹²) high-confidence functional binding sites controlling between 90 (ref. ¹²) and 200 (ref. ^{12,13}) different genes. The quaternary structure of the interaction of the PAX6 paired domain with a specific binding site has been determined using X-ray crystallography.¹⁴ PAX6 has been shown to bind cooperatively with SOX2 *in vitro*.¹⁵ Heterozygous LGD variants in SOX2 are the most common cause of severe bilateral eye malformations.¹⁶

Here we use an unbiased approach to define the phenotypic spectrum associated with PAX6 missense variants and determine how confidently the observed phenotype can be predicted by the nature and position of each causative amino acid substitution and their predicted impact on DNA binding. To define the ocular phenotypic spectrum we reviewed all of the published cases of PAX6 missense variants and those identified in the HGU^{eye} cohort, which is enriched for severe eye malformations, particularly microphthalmia, anophthalmia, and coloboma (MAC). To look for previously undescribed extraocular phenotypes we reviewed all of the likely causative PAX6 missense variants reported in the Deciphering Developmental Disorders (DDD) study,^{17,18} which has performed exome sequencing on 6993 individuals with previously undiagnosed, diverse, and severe developmental disorders with adequate phenotypic data.

MATERIALS AND METHODS

Ethical approval

This study complied with the tenets of the Declaration of Helsinki: informed consent was obtained from all of the families, including permission to publish patient photos. The Medical Research Council (MRC) Human Genetics Unit eye malformation (HGU^{eye}) cohort was collected and maintained using protocols approved by the Scotland UK Multicentre Research Ethics Committee: 06/MRE00/76 “Linkage and mutation analysis of developmental eye genes in cases of severe structural eye malformations” and 16/SS/0201 “An investigation of the genetic causes of human eye malformations.” The Deciphering Developmental Disorders (DDD) Study was collected under a protocol approved by Cambridge UK Multicentre Research Ethics Committee: 10/H0305/83 “The Deciphering Developmental Disorders Study.”

Variant analysis

Missense variants from the HGU^{eye} cohort were identified by a variety of different methods used to perform variant analysis

of the PAX6 coding sequence from genomic DNA, including heteroduplex analyses, Sanger sequencing, and targeted next-generation sequencing (including exome sequencing), essentially as described.¹⁹ PAX6 missense variants in the DDD study (www.ddduk.org), identified from trio based exome sequencing and diagnostic filtering of the variants, was performed as described.^{17,18} Targeted sequencing of individuals with foveal hypoplasia was performed in the Salisbury NHS Foundation Trust genetic testing program. We designated PAX6 missense variants present in the gnomAD database v.2.1 (<https://gnomad.broadinstitute.org/gene/ENSG000000007372/transcript/ENST00000379132>) as putatively benign for the purposes of comparison with the pathogenic variants. All variants were numbered according to the reference sequences NC_000011.10 (genomic, GRCh38), NM_000280.4; ENST00000379132 (transcript, 1269 nucleotides open reading frame), and NP_000271.1; P26367-1; LRG_720p1; ENSP00000492315.1 (isoform, 422 amino acids).

Phenotypic cluster analysis

Nonredundant components of the ocular phenotype were systematically coded in individuals with PAX6 missense variants and compared with those from a size-matched random sample of individuals with PAX6 LGD variants (Table S1). The phenotypic data were curated from original publications, and/or from the original clinical descriptions from the referring clinicians. The clustering analyses and generation of the heat maps were performed using the R packages cluster and gplots, and the function heatmap.2 (see Table S1 for further details of the phenotyping and scripting). Within heatmap.2, the functions dist and daisy were combined to compute the distance/dissimilarity matrix, respectively (combining the two allowed missing data, NA, to be taken into account in the clustering). Euclidean distance (root sum-of-squares of differences) was selected as the metric argument within daisy. When Manhattan distances (sum of absolute differences) were selected as an argument, the heat map did not change significantly (Fig. S1a).

Computational modeling and phenotype predictions

All structural analyses used the X-ray crystal structure of the DNA-bound PD (PDB ID: 6pax) and a homology model of the DNA-bound HD, built with SWISS-MODEL²⁰ using the crystal structure of the closely related *Drosophila* Aristaless HD (PDB ID: 3lnq) as a template. The effects of all missense variants were modeled with the program FoldX,²¹ both using the protein monomer by itself and the protein:DNA complex. Default FoldX parameters were used, with ten replicates performed per variant. Variants with $\Delta\Delta G > 1.6$ kcal/mol (using the same threshold described previously²²) for the protein monomer were classified as “destabilize folding.” Variants with monomer $\Delta\Delta G \leq 1.6$ kcal/mol, where the difference between the $\Delta\Delta G$ for the full complex and the monomer was >0.25 kcal/mol, were classified as “perturb

protein:DNA interaction.” All other variants were classified as “unknown molecular effect.”

In total, 21 different computational phenotype predictors were run for all of the pathogenic and putatively benign (gnomAD) variants in the primary PAX6 isoform (Uniprot ID: P26367). The predicted properties of all variants are provided in Table S2.

To assess evolutionary conservation of PAX6 residues among orthologous proteins, 124 sequences from Ensembl Compara were used (https://www.ensembl.org/Homo_sapiens/Gene/Compara_Ortholog?db=core;g=ENSG00000007372;r=11:31784779-31818062).

Electrophoretic mobility shift assay (EMSA)

EMSA experiments were performed using purified PAX6 PD protein variants and two known DNA target sequences. The PAX6 PD was expressed from plasmid vector in pGEX-6P-1t and site-directed mutagenesis was used to generate clones expressing the variants p.Cys52Arg, p.Ser54Arg, p.Arg92Gln, p.Ser121Leu, and p.Asn124Lys. The proteins were expressed as GST-fusion proteins in *E. coli* BL21-Gold(DE3) Competent Cells (Agilent), purified with Glutathione Sepharose 4B beads and cleaved using PreScission cleavage enzyme. DNA target sequences were double-stranded Fam-labeled oligos (Integrated DNA Technologies) incorporating the PAX6 enhancer sequences LE9 or SIMO. Reaction products were resolved on native PAGE gels and densitometry performed using Image Studio Lite (LI-COR Biosciences). Quantified binding of wild-type PAX6 PD was fitted to the Hill equation (GraphPad Prism 8 software) to calculate the K_d value, which was the concentration used for comparison of the binding properties of each of the PAX6 PD protein variants. Within each experiment, the proportion bound was normalized to the wild type. One-way repeated-measures analysis of variance (ANOVA) was used, with Dunnett's multiple comparison test to compare the protein–DNA binding of the different protein variants versus the wild-type PAX6 PD. Alpha was set at $p = 0.05$.

Identification of variants in PAX6 paralogs

Eight human paralogous members of the PAX family (excluding PAX6) were selected to identify all pathogenic missense variants in the PD of each protein. Variants were identified using publicly available data from Alamut Visual v2.11 (SOPHiA Genetics), including ClinVar (pathogenic, likely pathogenic categories), and from Leiden Open Variation Database (LOVD) (<http://www.lovd.nl/3.0/>) and literature searching. All variants reported as pathogenic and with phenotype details were collated and tabulated in text form and as a component of a multiple sequence alignment (CLUSTAL O1.2.4).

RESULTS

Defining the spectrum of ocular phenotypes associated with PAX6 missense variants

To look for novel ocular phenotypes associated with PAX6 missense variants, we selected 372 individuals with bilateral

MAC from a total of 1169 unrelated screened individuals with a wide spectrum of eye malformations in the HGU^{eye} cohort (Table S3). In this MAC group, 17 individuals from 15 families have eight different ultrarare, heterozygous PAX6 missense variants altering seven different residues (Table 1, see Supplementary Materials and Methods for detailed clinical descriptions). In six affected individuals the variant had occurred de novo. In the two multiplex families the variant segregated appropriately with the disease. In the remaining seven affected individuals we were not able to determine the inheritance of the causative allele because all of the required samples were not available. Of the seven substituted residues in the MAC group, six were within the PD (Arg26, Gly36, Arg38, Gly51, Ser54, Asn124) and one in the HD (Asn260). Recurrent substitutions were seen in 4/6 PD residues, involving either identical (p.Arg38Trp, p.Gly51Arg, p.Ser54Arg, p.Asn124Lys) or different (p.Arg38Trp/Gln) amino acids. All individuals with a p.Ser54Arg or p.Asn124Lys variant were prescreened and found to be wild type for SOX2 and OTX2 coding sequences.

To identify missense variants that either mimic haploinsufficiency or represent likely hypomorphic alleles we took two different approaches. First we selected a group of 399 unrelated individuals from the HGU^{eye} cohort who were strongly enriched for classical aniridia and other known PAX6-associated ocular malformations. From this analysis, seven non-MAC individuals from three families have two different heterozygous missense variants, Val78Ala and Arg128His, both of which are in the PD. The clinical features of these individuals are summarized in Table 2. In addition, we reviewed and tabulated phenotypic and genetic data on PAX6 missense variants from other sources, including all published cases that were accessible to us (Tables S1, S4), resulting in a total of 161 affected individuals from 110 families with 86 different missense variants. As a “true haploinsufficiency” comparison group, we extracted phenotypic information from a randomly selected matched number of individuals with PAX6 LGD variants (Table S1).

Phenotypic cluster analysis identifies worse-than-null PAX6 missense variants

We used the detailed ocular phenotypic and genetics information from the set of individuals identified in the previous section to perform a cluster analysis to determine whether missense and LGD variants in PAX6 could be distinguished on the basis of their phenotypes. This confirmed that aniridia is highly correlated with LGD variants. MAC and anterior segment dysgenesis cluster separately but are both correlated with missense variants (Figs. 1a, S1). Ocular features that are highly correlated with aniridia, such as foveal hypoplasia, cataract, and keratopathy, are more evenly distributed between LGD and missense variants. The individuals with p.Ser54Arg and p.Asn124Lys variants were distinguishable as severe bilateral microphthalmia with or without iris defects, coloboma, congenital corneal opacification, retinal detachment, and lens defects comprising

Table 1 Clinical and molecular details of HGU^{eye} microphthalmia, anophthalmia or coloboma (MAC) individuals with likely causative heterozygous PAX6 missense variants

PAX6 residue, helical region and domain	Family ID	Patient ID	Arg26, αI PD		Gly36, αI-αII linker PD		Arg38, αI-αII linker PD		Gly51, αIII PD		Ser54, αIII PD		Asn124, αVI PD		Asn260, αIII HD			
			1139	1141	1368	1273	355	1016	3343 (father)	1235	4091	3189	494	3190	1517	1319	339	5
Residue conserved in paralogs	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
	ENSP0000049231.1	p.(Arg26Gln)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
	ENSP0000049231.1	p.(Arg26Gln)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
Causative substitution of residue in paralog	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
Previously reported allele	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
Inheritance	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
Sex	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
Age at last assessment (years)	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
Ophthalmology	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
	ENSP0000049231.1	p.(Arg26Gln)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
MAC spectrum	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₁	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₂	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₃	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₄	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₅	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₆	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₇	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₈	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₉	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₁₀	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₁₁	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₁₂	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₁₃	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₁₄	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₁₅	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₁₆	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₁₇	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₁₈	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₁₉	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₂₀	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₂₁	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₂₂	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₂₃	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₂₄	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₂₅	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₂₆	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₂₇	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₂₈	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₂₉	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₃₀	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₃₁	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₃₂	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₃₃	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₃₄	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₃₅	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₃₆	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₃₇	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₃₈	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₃₉	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₄₀	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₄₁	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₄₂	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₄₃	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₄₄	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₄₅	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₄₆	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₄₇	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₄₈	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₄₉	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₅₀	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₅₁	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₅₂	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₅₃	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₅₄	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₅₅	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₅₆	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₅₇	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₅₈	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₅₉	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₆₀	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₆₁	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₆₂	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₆₃	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₆₄	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₆₅	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₆₆	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₆₇	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₆₈	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₆₉	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₇₀	ENST0000043871.1	c.77(G>A)	p.(Arg38Trp)	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
BL M, BL iris C and choroid C ₇₁	ENST0000043871.1	c.77(G&gt																

BEO both eyes open, *BL* bilateral, *C* coloboma, *Car* cataract, *CF* counting fingers, *FH* foveal hypoplasia, *HD* homeodomain, *HM* hand movements, *M* microphthalmia, *Mat* maternal, *MRI* magnetic resonance image, *NA* not available, *ND* not determined, *NPL* no perception of light, *ONH* optic nerve hypoplasia, *Pat* paternal, *PD* paired domain, *PFV* persistent fetal vasculature, *Pl* perception of light, *UL* unilateral.

³Subject 19 in Yogarajah et al.⁷ with no individual clinical details given and analyzed as an embedded component of a brain MRI grouped data set.

⁴¹The recurrent variant c.112C>T p.(Arg38Trp) had been previously reported in an HGU^{eye} cohort index.

The recurrent variant c.112C>T p.(Arg38Trp) had been previously reported in an HBB⁺ cohort index. The recurrent variant c.151G>A p.(Gly51Arg) had been previously reported in an aniridia family.⁴¹

Table 2 Clinical and molecular details of HGU^{eye} non-microphthalmia, anophthalmia, or coloboma (MAC) individuals with likely causative heterozygous PAX6 missense variants

PAX6 residue, helical region and domain		Val78, αII-αIV linker PD		Arg128, αVI PD	
Family ID		1574	1574 (grandfather)	541	3299
Patient ID		1574	1576 (father)	1578 (grandfather)	541
NC_000011.10 (chr11 GRCh38)		g.31801685A>G		g.31800831C>T	
NM_000280.4		c.233T>C		c.383G>A	
NP_000271.1		p.(Val78Ala)		p.(Arg128His)	
Residue conserved	No	Yes		Yes	
in paralogs		Yes		Yes	
Causative substitution of residue in paralog	No	Yes ^a		Yes ^a	
Previously reported allele		Yes ^a		Yes ^a	
Inheritance	Pat	Pat	ND	Pat	Pat
Sex	Female	Male	Male	Female	Male
Age at last assessment (years)	11	40	67	Adult	4
Ophthalmology					
MAC spectrum	None	None	None	None	None
Iris phenotype (excluding MAC)	BL partial aniridia	BL partial aniridia	BL partial aniridia	None	Iris hypoplasia (mild)
Other ocular features	FH, nystagmus, ONH, strabismus	FH, keratopathy, ONH, strabismus	FH, nystagmus, pseudophakia, keratopathy, ONH, strabismus	Cat, FH, nystagmus	Nystagmus, BL myopia, pale fundi with hypopigmented areas in posterior pole and no visible macular structures
Visual acuity (logMAR)	0.7 BL	0.5 BL	0.7 BL	NA	1.0 right, 0.2 left
Neurology, development					
Neuroimaging	No	No	No	No	No
Intellectual disability	No	No	No	NA	NA
Other features					
Additional clinical features	None stated	None stated	None stated	NA	NA

BL, bilateral, Cat, cataract, FH, foveal hypoplasia, NA, not available, ND, not determined, ONH, optic nerve hypoplasia, Pat, paternal, PD, paired domain. ^aThe recurrent variant c.383G>A; p.(Arg128His) had been previously reported in an affected individual with corectopia (slight), corneal neovascularization, foveal hypoplasia, and nystagmus.⁴

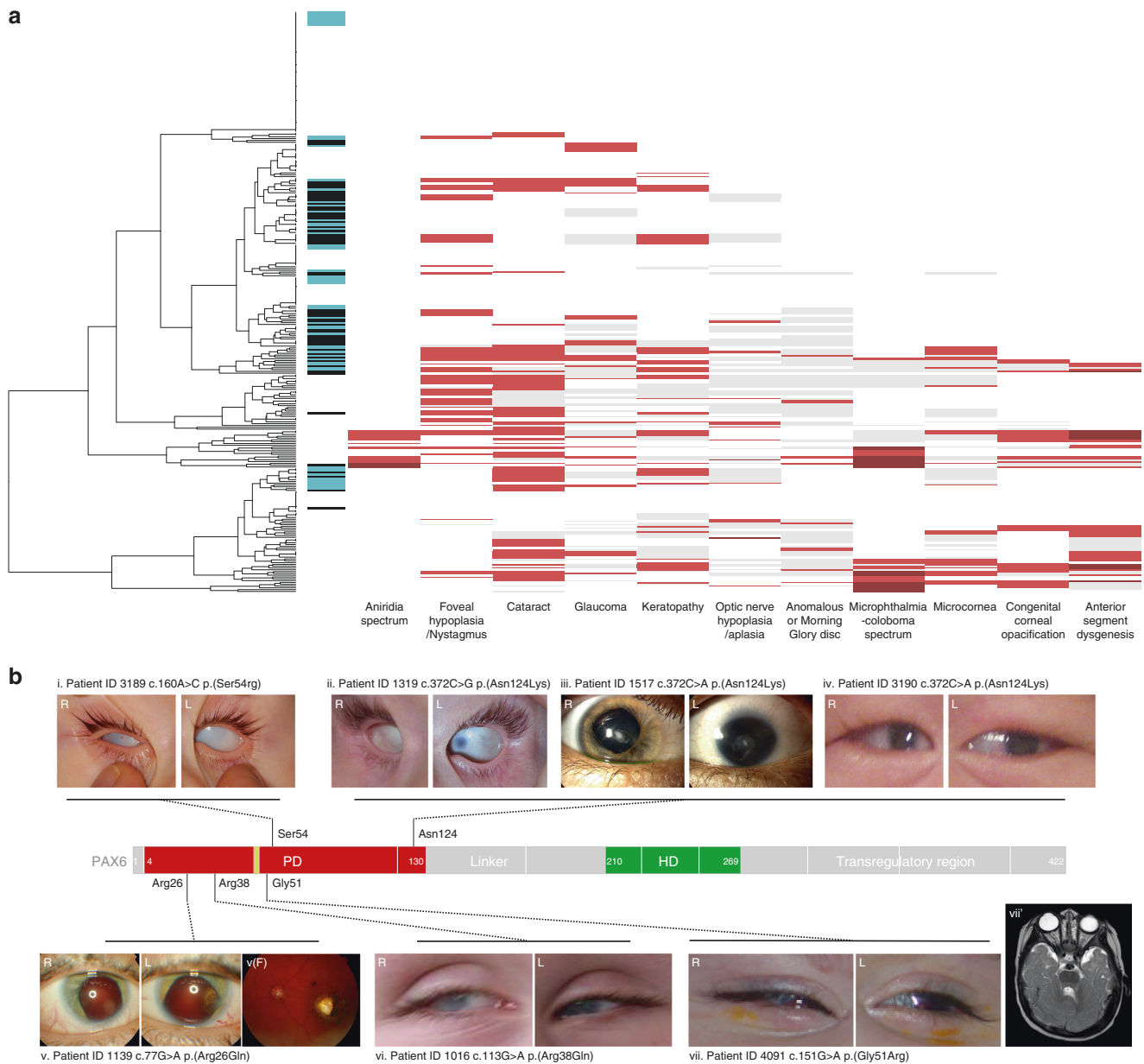


Fig. 1 Computational clustering and severity of PAX6-associated phenotypes. (a) Quantitative phenotype-driven clustering analysis. The ocular features assessed are labeled on the x-axis. Variants encoded by pathogenic PAX6 missense variants (blue; $n = 161$) and a size-matched cohort of likely/presumed gene disruptive (LGD) variants (black; nonsense and frameshift variants with predicted nonsense-mediated decay) were clustered according to phenotypic features using the R packages cluster and gplots, and the function heatmap.2 (red/dark red, feature present; gray, feature absent; white, data not available). The clustering reveals that LGD variants are predominantly associated with classical aniridia features, whereas missense variants have a wider phenotypic spectrum that includes distinct clusters of microphthalmia, anophthalmia, or coloboma (MAC) spectrum and anterior segment dysgenesis phenotypes. The aniridia-associated features foveal hypoplasia, cataract, and keratopathy, which can present in the absence of iris defects, are more evenly distributed across both classes of variants. Detailed versions of the phenotyping data set used for the clustering analysis and the heat map, with all of the missense and LGD variants labeled, are available (Table S1 and Fig. S1, respectively). (b) SOX2-like phenotype associated with recurrent PAX6 p.Ser54Arg and p.Asn124Lys variants. A schematic of PAX6 isoform NP_000271.1 (P26367-1, 422 amino acids) indicating the position of the paired domain (PD; red box), homeodomain (HD; green box), the alternatively spliced 14 amino acids (yellow box) that are present in the PAX6 + 5a isoform NP_001595.2 (P26367-2, 436 amino acids), and the PAX6 coding exon boundaries (white vertical bars). The PD-HD linker region and the C-terminal transregulatory region are labeled. Images i–iv, shown above the schematic, are the eye phenotypes of HGU^{eye} cohort MAC individuals with PAX6 p.(Ser54Arg) or p.(Asn124Lys) PD variants, which are recurrent and uniquely associated with severe bilateral defects; v–vii, shown below the schematic, are the eye phenotypes of HGU^{eye} cohort MAC individuals with PAX6 PD (NTS) variants, which are recurrent and are associated with more variable phenotypes; v(F), fundus image of patient ID 1139 with bilateral choroidal coloboma; vii', magnetic resonance image (MRI) showing bilateral mild microphthalmia in patient ID 4091. L left eye, R right eye.



Fig. 2 Protein structural analysis, DNA interaction, and paralog comparison of PAX6 variants. (a) Protein structural analysis of microphthalmia, anophthalmia, or coloboma (MAC) versus non-MAC missense variants. Locations of pathogenic missense variants highlighted on the structures of the DNA-binding domains: paired domain (PD) (PDB ID: 6pax) and homeodomain (HD) (homology model of the complex based upon PDB ID: 3lnq). Residues mutated in individuals with MAC are shown in red, while all other residues associated with pathogenic variants are shown in light blue. Figure S2 shows the variants further stratified by phenotype, and also shows other nonpathogenic variants from gnomAD observed in the human population. (b) Predicted molecular effects of MAC versus non-MAC missense variants. The *P* value is calculated with Fisher's exact test, based upon the proportion of variants in each group of unknown molecular effect. (c) Effect of PD missense variants on binding to known DNA targets of PAX6. Electrophoretic mobility shift assays (EMSA) of the PD of PAX6. Two known DNA targets of the PAX6 PD, the LE9 and SIMO enhancer elements, were selected as binding sequences. For each element, the DNA binding of wild-type PD (lane 2 of each gel image) was compared with five different mutant PDs: Cys52Arg and Ser121Leu (C52R and S121L, aniridia-associated variants, lanes 3 and 6 of each gel image), Ser54Arg and Asn124Lys (S54R and N124K, MAC-associated variants, lanes 4 and 7 of each gel image), and Arg92Gln (R92Q, gnomAD variant, lane 5 of each gel image). For the MAC-associated variants, PD-Ser54Arg showed reduced binding to the LE9 element (left panel) but not to the SIMO element (right panel), whereas PD-Asn124Lys showed a moderate reduction in binding to both the LE9 and SIMO elements (left and right panels, respectively). Quantification data of the wild-type and mutant PD binding profiles are shown above each representative EMSA gel image: black, wild-type PD normalized binding; red, C52R and S121L aniridia-associated variants; purple, S54R and N124K MAC-associated variants; blue, R92Q gnomAD variant. (d) Comparison of disease-associated PD variants in PAX6 paralogs. A multiple sequence alignment (generated by CLUSTAL O1.2.4) showing the PD of the nine members of the PAX protein family: PAX1/9, PAX2/5/8, PAX3/7, and PAX4/6. Gray shaded blocks, α -helices α I– α III of the N-terminal subdomain; green shaded blocks, α -helices α IV– α VI of the C-terminal subdomain; red bold, residues with pathogenic PAX6 missense variants identified in the HGU^{eye} MAC (blue shaded), non-MAC (yellow shaded), and Deciphering Developmental Disorders (DDD) (black underlined) cohorts analyzed in this report or conserved at these positions across the protein family (unshaded); brown shaded box, residues with previously reported pathogenic missense variants (for PAX6 these are restricted to those associated with MAC phenotypes). Residues associated with MAC cases in this report are labeled below the alignment, indicating the residue number and in brackets the number of unrelated cases with a missense variant; PAX6 Ser54 and Asn124, the residues uniquely associated with severe MAC phenotypes, are shown in bold. An extended version of this figure, detailing the reported pathogenic missense variants in the HD of PAX3/7 and PAX 4/6, is available (Fig. S6). CTS C-terminal subdomain, NTS N-terminal subdomain, WT wild type.

variants (Fig. 2b). Moreover, only 1/22 MAC variants (36 individuals with MAC) had an unknown molecular effect, compared with 28/64 non-MAC variants (125 individuals with non-MAC). Other phenotypic categories showed no significant differences in their association with variants predicted to affect protein folding or DNA binding (Fig. S3, Table S2).

We also investigated the ability of many existing sequence-based phenotype predictors to identify pathogenic PAX6 missense variants (Table S5). None of them shows any significant ability to distinguish MAC from non-MAC variants. However, many perform very well at discriminating the pathogenic PAX6 missense variants from 121 putatively benign variants present in the gnomAD database. Interestingly, FoldX, which uses structure alone, is competitive with the top-ranking predictors, which rely mostly on evolutionary conservation, demonstrating the power of protein structure for predicting both the pathogenicity and phenotype of PAX6 missense variants.

EMSA characterization of variant interactions with known PAX6-binding sites

As the structural analysis suggests that many PAX6 variants are likely to affect the interaction of the PD with DNA, we used EMSA experiments to test a subset of variants for their binding to two well-characterized PAX6-binding sites: the LE9 and SIMO elements^{24,25} (Figs. 2c, S4). One of the severe MAC-associated variants, p.Ser54Arg, caused an 85% reduction in binding to LE9 but had no effect on SIMO binding, while p.Asn124Lys showed reduced binding to both (42% to LE9, 62% to SIMO). In contrast, the classical aniridia variants p.Cys52Arg²⁶ and p.Ser121Leu²⁷ both completely disrupted binding to SIMO and caused an 80–90% disruption of

binding to LE9. We also tested p.Arg92Gln (rs769095184), which was presumed to be a neutral variant because it was present in eight individuals (East Asian, allele frequency 0.0004349) in gnomAD. It showed a LE9 and SIMO binding pattern very similar to p.Asn124Lys, which was initially surprising, but consistent with the fact that it is predicted by our structural analysis to moderately perturb the interaction with DNA. Interestingly, this variant was very recently reported, with limited segregation details, as being associated with classical aniridia in a Chinese family with four affected individuals.²⁸

PAX6 missense variants as a cause of previously undescribed nonocular phenotypes

Finally, to determine whether PAX6 missense variants may cause previously undescribed conditions with significant extraocular effects, we examined the phenotypes of individuals in the DDD cohort.^{17,18} Seven individuals had six different ultrarare, heterozygous PAX6 missense variants (Table 3). Four individuals with structural eye anomalies had three different missense variants in the PD (p.[Arg26Gly]x2, p.[Arg26Gln]x1, p.[Gly64Ala]x1). The two different Arg26 variants are recurrent (see Table 1 and ref. ²⁹). The remaining three individuals had either predominantly neurological phenotypes or multiple nonocular malformations. Two variants occur in the PD (p.Glu62Val and p.Ile117Thr) and are predicted to be mild at a structural level and not significantly perturb protein folding or the interaction with DNA. The other (p.Gln201His) is in the flexible linker between the PD and HD and is not particularly close to any known pathogenic variants. Interestingly, the three top-ranking computational phenotype predictors (FATHMM, M-CAP, and DeepSequence) are supportive of pathogenicity

Table 3 Clinical and molecular details of Deciphering Developmental Disorders (DDD) individuals with likely causative heterozygous PAX6 missense variants

DECIPHER ID	276121	277415	265016	305488	270475	307572	303218
Variant	chr11:g.31824317G>C g.31824317G>C 46XX	chr11:g.31824317G>C g.31824317G>C 46XX	chr11:g.31824316C>T g.31824316C>T 46XX	chr11:g.31823281T>A g.31823281T>A 46XX	chr11:g.31823275C>G g.31823275C>G 46XX	chr11:g.31823116A>G g.31823116A>G 46XX	chr11:g.31816257T>A g.31816257T>A 46XX
Sex	Female	Female	Female	Female	Female	Female	Female
Transcript	ENST00000241001	ENST00000241001	ENST00000241001	ENST00000241001	ENST00000241001	ENST00000241001	ENST00000241001
Exon	5/13	5/13	5/13	6/13	6/13	6/13	8/13
dNA position	76	76	77	185	191	350	603
Codons	Cgg/Ggg	Cgg/Ggg	cGg/cAg	gAg/gTg	gGc/gCc	aTg/aCa	caA/caT
Protein change	p.(Arg26Gly) ^a	p.(Arg26Gly) ^a	p.(Arg26Gln)	p.(Glu62Val)	p.(Gly64Ala)	p.(Ile117Thr)	p.(Gln201His)
Other DDD ultrarare variants	None	None	None	None	Het <i>DSPP</i> p. Asp506Gly pat; Het <i>FOXC1</i> p.Leu485Ser mat; Het <i>CREBBP</i> p. Met1046Val mat; all absent from gnomAD	None	None
Residue conserved in paralogs	Yes	Yes	Yes	No	Yes	No	No
Causative substitution of residue in paralog	Yes	Yes	Yes	No	Yes	No	Unknown
Known ClinVar allele	rs121907913 Pathogenic	rs121907913 Pathogenic	Present study, Table 1	No	No (p.Gly64Val is pathogenic rs121907920)	No	No
SIFT	Deleterious (0)	Deleterious (0)	Deleterious (0)	Deleterious (0)	Deleterious (0)	Deleterious (0)	Deleterious (0)
PolyPhen	Probably_damaging (1)	Probably_damaging (1)	Probably_damaging (1)	Probably_damaging (1)	Probably_damaging (0.999)	Probably_damaging (0.999)	Probably_damaging (0.998)
Pathogenicity/contribution (PAX6 genotype)	Pathogenic/full	Pathogenic/partial	Pathogenic/full	Unknown	Likely pathogenic/partial	Unknown	Unknown
Inheritance	Unknown	Maternally inherited	De novo constitutive	Unknown	Paternally inherited, constitutive in father	Unknown	Unknown
Family history	None	Mother has congenital cataract and optic nerve hypoplasia	None	None	Father has unilateral iris coloboma; brother has bilateral iris coloboma and horizontal nystagmus	None	None
Eye phenotype	Abnormality of the anterior segment of the globe, nystagmus	Bilateral microphthalmos, congenital cataract, iris coloboma	Cone/cone-rod dystrophy, congenital blindness, congenital cataract	None	Bilateral cataracts, horizontal nystagmus	Amblyopia	None
Other phenotype		Bilateral choanal atresia, delayed speech and language development, nephrolithiasis, secundum atrial septal defect		Cortical visual impairment, intrauterine growth retardation, ischemic stroke, seizures, severe global developmental delay	Plagiocephaly, global developmental delay with autistic features	Abnormality of the sacrum, esophageal atresia, ileal atresia, meningocele, rectal prolapse, renal duplication, short stature, stage 5 chronic kidney disease, tethered cord, tetralogy of Fallot, tracheoesophageal fistula, vesicoureteral reflux	Few cafe-au-lait spots, hemangioma, moderate global developmental delay

cDNA complementary DNA.

^aThe recurrent variant c.76C>G; p.(Arg26Gly) had been previously reported in an HGU^{ev} cohort individual with bilateral microphthalmia and iris defects.²⁹

for the two PD variants, having values similar to other pathogenic *PAX6* variants and distinct from most of the variants in gnomAD (Fig. S5). Overall, however, we found no clear evidence for major nonocular developmental defects due to *PAX6* missense variants.

DISCUSSION

The extraordinary, almost one-to-one, association of typical aniridia to heterozygous loss-of-function *PAX6* variants understandably dominates the human disease literature associated with this gene. Over many years, the MRC Human Genetics Unit has established a large research cohort of individuals with aniridia and other severe eye malformations that have been screened for causative variants in *PAX6*. This, together with the availability of phenotypic and trio exome sequencing data from the DDD study, has allowed us to use multiple independent approaches to define the phenotypic spectrum associated with missense variants in *PAX6*. This has identified a specific subset of heterozygous missense variants in the *PAX6* DNA-binding domains that are associated with severe bilateral microphthalmia, most notably the recurrent substitutions of Ser54 and Asn124 within the PD.

It has been inferred that the PD arose from gene duplication of a three-helix unit, with divergent evolution resulting in the NTS and CTS losing amino acid and DNA target sequence similarity but maintaining a twofold pseudosymmetry axis that permits similar folding and docking of each subdomain into the major groove of double-stranded DNA (dsDNA). In this regard, it is interesting to note that both Ser54 and Asn124 occur at approximately equivalent positions within their subdomains and both make direct contacts with the major groove, with similar relative positioning in the 5' and 3' regions of the DNA target binding site (Fig. 2a). Ser54 is conserved in all members of the human *PAX* family, and Asn124 in *PAX2/5/8* and *PAX4/6* (Figs. 2d, S6). In *PAX6* orthologues, Ser54 is invariant, and Asn124 is conserved in all but two avian species.

The severe bilateral microphthalmia phenotypic spectrum associated with p.Ser54Arg and p.Asn124Lys is similar to that associated with *SOX2* variants (MIM 184429).²³ There are several reasons this is an important observation. *SOX2* is a lineage-specific transcription factor, which cobinds with Pax6 to the promoter elements of many genes in presumptive lens cells.^{30,31} In chicken, *PAX6* and *SOX2* cooperatively interact at the lens-specific DC5 enhancer of delta crystallin and the N-3 enhancer of *Sox2*.^{32,33} In an optic cup progenitor cell-conditional *Sox2*-deficient mouse model, the severe ocular defects, including extreme microphthalmia, are ameliorated when crossed onto a *Pax6*-haploinsufficient background.³⁴ These data suggest that the transcriptional activation profiles of *PAX6* and *SOX2* are tightly regulated and can be interdependent, and therefore that the perturbation of the transcriptional activity of either gene can cause severe ocular defects during development. It is possible that the individuals with severe bilateral MAC and a *PAX6* variant also carry an as-yet-unidentified variant in a regulatory region or upstream regulator of *SOX2*. However, this seems unlikely, given the

number and recurrent nature of the *PAX6* variants identified in the HGU^{eye} cohort. The other important factor being that both p.Ser54Arg and all but two of the p.Asn124Lys variants occurred de novo in the affected individuals (in two adult cases inheritance could not be determined). This strongly supports monoallelic variants in *PAX6* as the cause of the severe bilateral microphthalmia.

One individual, with a phenotype similar to that associated with the recurrent p.Ser54Arg and p.Asn124Lys changes, has a variant in the HD (p.Asn260Tyr). Asn260 is invariant in all paired-type HDs and is adjacent to the C-terminal basic cluster (residues 261–267) component of the *PAX6* HD nuclear localization signal (Fig. S6). In zebrafish Pax6 this asparagine is critical for binding DNA target sequences and for CTS–HD interaction.³⁵ The two previously reported MAC-associated HD variants similarly substitute conserved residues in the recognition helix, Val256 and Phe258, and cause iris defects and severe bilateral phenotypes including microphthalmia with primary aphakia or chorioretinal and optic nerve coloboma.^{3,36} A recently reported case with p.Arg261Gly had congenital cataract and a family history of coloboma.³⁷ HD recognition helix variants are reported for *PAX3* in association with Waardenburg syndrome type 1, and for *PAX7* in a single case of cleft lip and palate (Fig. S6).^{38,39}

Taken together, the results of the protein structural analysis and EMSA experiments have allowed us to develop hypotheses regarding the worse-than-null phenotypes we observed. The most severe MAC-associated variants perturb the interaction of any one of the three DNA-binding (sub) domains. We infer that DNA binding is unlikely to be completely abrogated as the two remaining wild-type (sub) domains are sufficient for DNA interaction. However, both the affinity of these interactions and the location of bound sites are likely to be altered. The remarkable sequence diversity of *in vivo* *PAX6*-binding sites^{12,13} would predict variant-specific differential effects on both the degree and repertoire of target gene activation. In contrast, the classical aniridia variants tested experimentally retain almost no DNA-binding activity, consistent with an altered dosage haploinsufficiency phenotype similar to the LGD variants. Importantly, however, these molecular properties alone cannot explain all of the phenotypic heterogeneity observed among *PAX6* missense variants. This is highlighted by the fact that there are several examples in our data set of different patients with the same variant exhibiting very different phenotypes (Table S4).

Four other members of the *PAX* gene family are associated with robust pathogenic missense variants in the PD (Fig. 2d; see Fig. S6, Table S6 for further details). These include *PAX2* (papillorenal syndrome and focal segmental glomerulosclerosis 7; MIM 120330 and 616002), *PAX3* (Waardenburg syndrome type I and type 3 and craniofacial–deafness–hand syndrome; MIM 193500, 148820, and 122880), *PAX8* (congenital nongoitrous hypothyroidism; MIM 218700) and *PAX9* (tooth agenesis; MIM 604625). Of the eight residues that have causative substitutions detailed above in HGU^{eye}

MAC or DDD, seven are invariant across the paralogs. Asn124 is conserved only in PAX2, PAX5, and PAX8 (group II) and PAX4 (group IV); no disease-associated missense variants have been reported at this position. Six of the seven invariant residues have causative missense variants reported in PAX6 paralogs. For Ser54 a different substitution at the equivalent residue in PAX3 (p.Ser84Phe) has been reported in a single family (Table S6). Heterozygosity for this variant was not associated with a worse-than-null phenotype. Overall, of the 44 different PAX6 PD positions with pathogenic variants, 25 have pathogenic variants at equivalent positions in paralogous PAX genes. Furthermore, there are 24 PD residues with pathogenic missense variants in paralogs with no known pathogenic variants in PAX6; there is a high probability that these could be the sites of pathogenic PAX6 variants to be discovered in the future.

SUPPLEMENTARY INFORMATION

The online version of this article (<https://doi.org/10.1038/s41436-019-0685-9>) contains supplementary material, which is available to authorized users.

ACKNOWLEDGEMENTS

We thank the affected individuals and their families who contributed to this study, Alison Brown for technical assistance with the HGU^{eye} cohort (MRC HGU, Edinburgh), and Daniel Osborne for clinical data collection (Southampton Eye Unit). Dirk-Jan Kleinjan (University of Edinburgh) is thanked for helpful discussions and critical reading of the manuscript. K.A.W. and D. R.F. are supported by a MRC University Unit Award to the University of Edinburgh for the MRC Human Genetics Unit, H.N. H. is supported by a Wellcome Trust Training Fellowship (205171/Z/16/Z), B.J.L. is supported by the MRC Precision Medicine Doctoral Training Program, J.A.M. is supported by a MRC Career Development Award (MR/M02122X/1) and is a Lister Institute Research Prize Fellow, E.S.T. is supported by MRC and Chief Scientist Office of the Scottish Government Health Directorates Scottish Genomes Partnership funding (SGP/1) and by MRC and Engineering and Physical Sciences Research Council Molecular Pathology Node funding. The DDD study presents independent research commissioned by the Health Innovation Challenge Fund (grant number HICF-1009-003) and makes use of DECIPHER (<http://decipher.sanger.ac.uk>), which is funded by the Wellcome Trust (see Nature PMID: 25533962 or www.ddduk.org/access.html for full acknowledgement).

DISCLOSURE

The authors declare no conflicts of interest.

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

REFERENCES

- Hingorani M, Hanson I, van Heyningen V. Aniridia. *Eur J Hum Genet*. 2012;20:1011–1017.

- Hanson I, Churchill A, Love J, et al. Missense mutations in the most ancient residues of the PAX6 paired domain underlie a spectrum of human congenital eye malformations. *Hum Mol Genet*. 1999; 8:165–172.
- Azuma N, Yamaguchi Y, Handa H, et al. Mutations of the PAX6 gene detected in patients with a variety of optic-nerve malformations. *Am J Hum Genet*. 2003;72:1565–1570.
- Yahalom C, Blumenfeld A, Hendler K, et al. Mild aniridia phenotype: an under-recognized diagnosis of a severe inherited ocular disease. *Graefes Arch Clin Exp Ophthalmol*. 2018;256:2157–2164.
- Glaser T, Jepeal L, Edwards JG, Young SR, Favor J, Maas RL. PAX6 gene dosage effect in a family with congenital cataracts, aniridia, anophthalmia and central nervous system defects. *Nat Genet*. 1994; 7:463–471.
- Schmidt-Sidor B, Szymanska K, Williamson K, et al. Malformations of the brain in two fetuses with a compound heterozygosity for two PAX6 mutations. *Folia Neuropathol*. 2009;47:372–382.
- Yogarajah M, Matarin M, Vollmar C, et al. PAX6, brain structure and function in human adults: advanced MRI in aniridia. *Ann Clin Transl Neurol*. 2016;3:314–330.
- Osumi N, Shinohara H, Numayama-Tsuruta K, Maekawa M. Concise review: Pax6 transcription factor contributes to both embryonic and adult neurogenesis as a multifunctional regulator. *Stem Cells*. 2008; 26:1663–1672.
- Hamasaki A, Yamada Y, Kurose T, et al. Adult pancreatic islets require differential pax6 gene dosage. *Biochem Biophys Res Commun*. 2007; 353:40–46.
- Shaham O, Menuchin Y, Farhy C, Ashery-Padan R. Pax6: a multi-level regulator of ocular development. *Prog Retin Eye Res*. 2012;31:351–376.
- Singh S, Stellrecht CM, Tang HK, Saunders GF. Modulation of PAX6 homeodomain function by the paired domain. *J Biol Chem*. 2000; 275:17306–17313.
- Sun J, Rockowitz S, Xie Q, Ashery-Padan R, Zheng D, Cvekl A. Identification of in vivo DNA-binding mechanisms of Pax6 and reconstruction of Pax6-dependent gene regulatory networks during forebrain and lens development. *Nucleic Acids Res*. 2015;43:6827–6846.
- Coutinho P, Pavlou S, Bhatia S, Chalmers KJ, Kleinjan DA, van Heyningen V. Discovery and assessment of conserved Pax6 target genes and enhancers. *Genome Res*. 2011;21:1349–1359.
- Xu HE, Rould MA, Xu W, Epstein JA, Maas RL, Pabo CO. Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. *Genes Dev*. 1999;13:1263–1275.
- Narasimhan K, Pillay S, Huang YH, et al. DNA-mediated cooperativity facilitates the co-selection of cryptic enhancer sequences by SOX2 and PAX6 transcription factors. *Nucleic Acids Res*. 2015;43:1513–1528.
- Williamson KA, FitzPatrick DR. The genetic architecture of microphthalmia, anophthalmia and coloboma. *Eur J Med Genet*. 2014; 57:369–380.
- Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. *Nature*. 2015; 519:223–228.
- Deciphering Developmental Disorders Study. Prevalence and architecture of de novo mutations in developmental disorders. *Nature*. 2017; 542:433–438.
- Williamson KA, Ringer J, Floyd JA, et al. Heterozygous loss-of-function mutations in YAP1 cause both isolated and syndromic optic fissure closure defects. *Am J Hum Genet*. 2014;94:295–302.
- Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018;46:W296–W303.
- Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol*. 2002;320:369–387.
- Alibes A, Nadra AD, De Masi F, Bulky ML, Serrano L, Stricher F. Using protein design algorithms to understand the molecular basis of disease caused by protein-DNA interactions: the Pax6 example. *Nucleic Acids Res*. 2010;38:7422–7431.
- Ragge NK, Lorenz B, Schneider A, et al. SOX2 anophthalmia syndrome. *Am J Med Genet A*. 2005;135:1–7.
- Aota S, Nakajima N, Sakamoto R, Watanabe S, Ibaraki N, Okazaki K. Pax6 autoregulation mediated by direct interaction of Pax6 protein with the head surface ectoderm-specific enhancer of the mouse Pax6 gene. *Dev Biol*. 2003;257:1–13.

25. Bhatia S, Bengani H, Fish M, et al. Disruption of autoregulatory feedback by a mutation in a remote, ultraconserved PAX6 enhancer causes aniridia. *Am J Hum Genet.* 2013;93:1126–1134.
26. Chao LY, Mishra R, Strong LC, Saunders GF. Missense mutations in the DNA-binding region and termination codon in PAX6. *Hum Mutat.* 2003;21:138–145.
27. Chen P, Zang X, Sun D, et al. Mutation analysis of paired box 6 gene in inherited aniridia in northern China. *Mol Vis.* 2013;19:1169–1177.
28. Xiao Y, Liu X, Yang C, et al. A novel PAX6 heterozygous mutation found in a Chinese family with congenital aniridia. *Genet Test Mol Biomarkers.* 2019;23:495–500.
29. Hanson IM, Fletcher JM, Jordan T, et al. Mutations at the PAX6 locus are found in heterogeneous anterior segment malformations including Peters' anomaly. *Nat Genet.* 1994;6:168–173.
30. Thakurela S, Tiwari N, Schick S, et al. Mapping gene regulatory circuitry of Pax6 during neurogenesis. *Cell Discov.* 2016;2:15045.
31. Hou L, Srivastava Y, Jauch R. Molecular basis for the genome engagement by Sox proteins. *Semin Cell Dev Biol.* 2017;63:2–12.
32. Saha A, Kizaki S, De D, Endo M, Kim KK, Sugiyama H. Examining cooperative binding of Sox2 on DC5 regulatory element upon complex formation with Pax6 through excess electron transfer assay. *Nucleic Acids Res.* 2016;44:e125.
33. Inoue M, Kamachi Y, Matsunami H, Imada K, Uchikawa M, Kondoh H. PAX6 and SOX2-dependent regulation of the Sox2 enhancer N-3 involved in embryonic visual system development. *Genes Cells.* 2007;12:1049–1061.
34. Matsushima D, Heavner W, Pevny LH. Combinatorial regulation of optic cup progenitor cell fate by SOX2 and PAX6. *Development.* 2011;138:443–454.
35. Bruun JA, Thomassen EI, Kristiansen K, et al. The third helix of the homeodomain of paired class homeodomain proteins acts as a recognition helix both for DNA and protein interactions. *Nucleic Acids Res.* 2005;33:2661–2675.
36. Deml B, Reis LM, Lemyre E, Clark RD, Karimnejad A, Semina EV. Novel mutations in PAX6, OTX2 and NDP in anophthalmia, microphthalmia and coloboma. *Eur J Hum Genet.* 2016;24:535–541.
37. Chauhan BK, Medsinghe A, Baumgartner MP, et al. Case series: pyramidal cataracts, intact irides and nystagmus from three novel PAX6 mutations. *Am J Ophthalmol Case Rep.* 2018;10:172–179.
38. Pingault V, Ente D, Dastot-Le Moal F, Goossens M, Marlin S, Bondurand N. Review and update of mutations causing Waardenburg syndrome. *Hum Mutat.* 2010;31:391–406.
39. Leslie EJ, Taub MA, Liu H, et al. Identification of functional variants for cleft lip with or without cleft palate in or near PAX7, FGFR2, and NOG by targeted sequencing of GWAS loci. *Am J Hum Genet.* 2015;96:397–411.
40. Henderson RA, Williamson K, Cumming S, et al. Inherited PAX6, NF1 and OTX2 mutations in a child with microphthalmia and aniridia. *Eur J Hum Genet.* 2007;15:898–901.
41. Chen J, Zhu J. Analysis of PAX6 gene mutations in a Chinese family affected with congenital aniridia. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi.* 2016;33:523–525. [Article in Chinese].



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

© The Author(s) 2019

¹MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK; ²Moorfields Eye Hospital, London, UK; ³Department of Clinical Genetics, St Michael's Hospital, Southwell Street, Bristol, UK; ⁴CHU Toulouse, Service de Génétique Médicale, Hôpital Purpan, Toulouse, France; ⁵Peninsula Clinical Genetics, Royal Devon and Exeter Hospitals (Heavitree), Exeter, UK; ⁶Newcastle Eye Centre, Newcastle upon Tyne Hospitals NHS Foundation Trust, Royal Victoria Infirmary, Newcastle Upon Tyne, UK; ⁷Clinical Genetics, Guys and St Thomas NHS Trust, Great Maze Pond, London, UK; ⁸Hospices Civils de Lyon, Genetic Department and National HHT Reference Center, Femme-Mère-Enfants Hospital, Bron, France; ⁹Teeside Genetics Unit, The James Cook University Hospital, Middlesbrough, UK; ¹⁰Institute of Human Genetics, University of Lübeck, Lübeck, Germany; ¹¹Department of Ophthalmology and Vision Sciences, Hospital for Sick Children, Toronto, ON, Canada; ¹²Department of Clinical Genetics, Great Ormond Street Hospital for Children, London, UK; ¹³Department of Ophthalmology, Justus-Liebig-University Giessen, Giessen, Germany; ¹⁴Northern Ireland Regional Genetics Service (NIRGS), Belfast City Hospital, Belfast, UK; ¹⁵Department of Ophthalmology, Hôpital Universitaire des Enfants Reine Fabiola, Brussels, Belgium; ¹⁶Department of Clinical Genetics, Sheffield Children's NHS Foundation Trust, Sheffield, UK; ¹⁷Department of Ophthalmology, University of Freiburg, Freiburg, Germany; ¹⁸University Hospital Southampton, Southampton, UK; ¹⁹Clinical and Experimental Sciences, University of Southampton, Southampton, UK; ²⁰Academic Medical Genetics and Pathology, University of Glasgow, Queen Elizabeth University Hospital, Glasgow, UK; ²¹Department of Genetics, University of Groningen, University Medical Center, Groningen, The Netherlands; ²²Medical Genetics, CHRU de Montpellier, Montpellier, France; ²³Clinical Genetics Unit, Birmingham Women's Hospital, Birmingham, UK.

A.2 Heat maps of phenotypic data

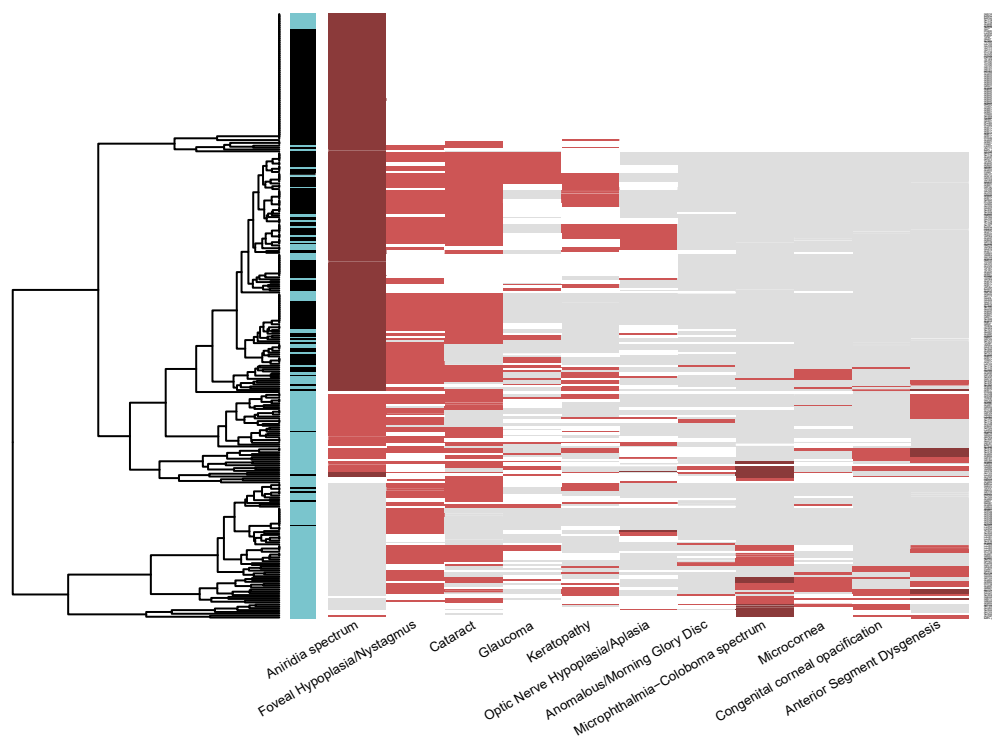
A.2.1 R script for heat map

```
setwd("/Volumes/igmm/fitzpatrick-lab/Eye genes ms and posters/PAX6  
microphthalmia ms/Draft_Nov_2018/")  
  
install.packages("gplots")  
  
library("gplots", lib.loc="~/Library/R/3.3/library")  
  
install.packages("cluster")  
  
library("cluster", lib.loc="~/Library/R/3.3/library")  
  
install.packages("readxl")  
  
library(readxl)  
  
distfunc = function(x)dist(daisy(x))  
  
colnames = c("Variant", "Aniridia spectrum", "Foveal Hypoplasia/Nystagmus",  
"Cataract", "Glaucoma", "Keratopathy", "Optic Nerve Hypoplasia/Aplasia",  
"Anomalous/Morning Glory Disc", "Microphthalmia-Coloboma spectrum",  
"Microcornea", "Congenital corneal opacification", "Anterior Segment Dysgenesis")  
  
misslgd2 <- read_excel("/Volumes/igmm/fitzpatrick-lab/Eye genes ms and  
posters/PAX6 microphthalmia ms/Draft_Nov_2018/Supplementary Table S2  
Rigorous geno_pheno classification NIKKI EDITING NOV 2018.xlsx", sheet =  
"Trimmed data for R 15 Nov 18", col_types = c("text", "numeric", "numeric",  
"numeric", "numeric", "numeric", "numeric", "numeric", "numeric",  
"numeric", "numeric"), na = "NA", col_names = TRUE)  
  
misslgdnu2 = apply(misslgd2[2:12], 2, as.numeric)
```

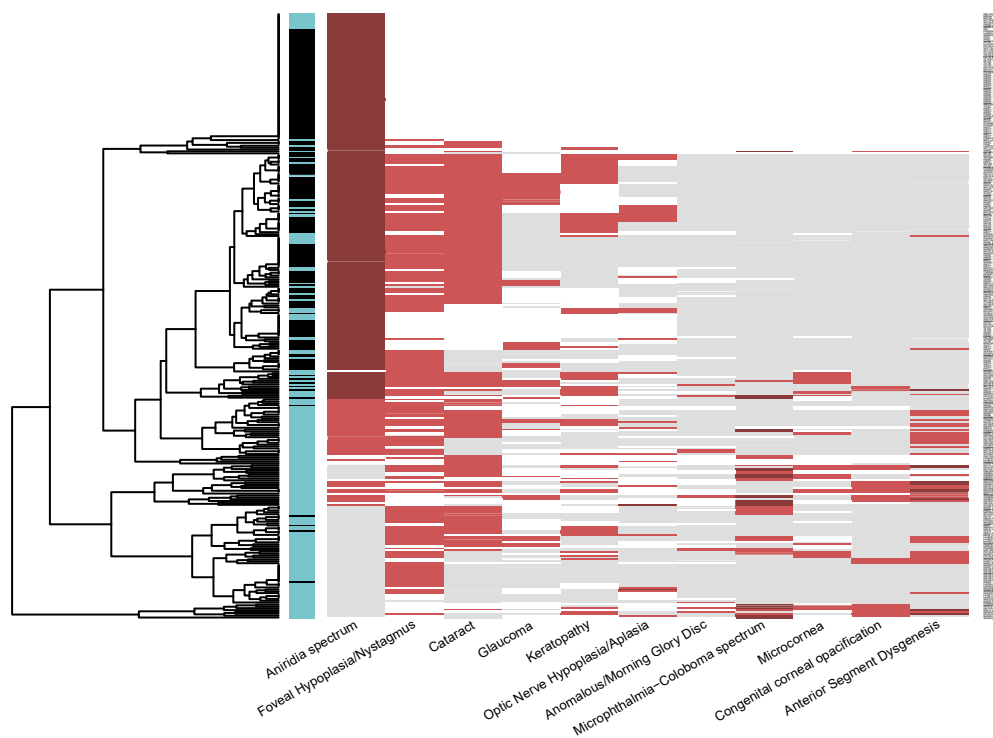
```
heatmap.2(misslgnu2, Colv = FALSE, dendrogram = "row", trace = "none", distfun = distfunc, labRow = misslgnu2$Variant, col = colorpanel(1000, "gray87", "indianred3", "indianred4"), na.color = "white", srtCol = 30, margins = c(5,5), cexRow = 0.1, cexCol = 0.6, offsetRow = 0.1, offsetCol = -0.6, labCol = colnames[2:12], RowSideColors = c(rep("cadetblue3", 161), rep("black", 161)), key = FALSE)
```

A.2.2 Alternative heat maps using different distance metrics

The choice of distance measure is an important methodological consideration with a potentially big influence on clustering results. Euclidean distance (root sum-of-squares of differences) was chosen as the distance metric, and changing this to Manhattan distance (sum of absolute differences) did not cause any major changes in the clustering (**Supplementary Figure 1** below). With Gower selected for this metric, the heat map did change slightly, with the MAC cases in more fragmented clusters (**Supplementary Figure 2**), but the rescaling involved in this method was not needed given that the data was all numeric.



Supplementary Figure 1 Heat map with Manhattan as the distance metric



Supplementary Figure 2 Heat map with Gower as the distance metric

A.3 Pivot table showing phenotypes associated with missense variants per residue

Residue	Variant (p.)	Phenotype	Family ID	No. individuals	Recurrently substituted	Additional phenotype
7	Gly7Arg	CA	1	1		
9	Asn9Ile	MI	2	1		
12	Gly12Arg	CA	3	1		
17	Asn17Lys	AS	4	1		y
	Asn17Ser	CA	5	1		
18	Gly18Arg	AS	6	1		y
	Gly18Trp	AS	7	1		retinal detachment
		MI	7	1		
19	Arg19Pro	MC	8	2		
26	Arg26Gln	MC	13	2		y
		MI	12	1		
	Arg26Gly	AS	10	2		
		MC	10	1		
			11	1		
		MI	11	1		
	Arg26Pro	MC	9	1		
		MI	9	1		
	Arg26Trp	MI	110	1		
29	Ile29Ser	CA	14	1		
33	Ala33Pro	CA	15	2		
36	Gly36Ala	MI	16	1		y
	Gly36Arg	CA	19	1		
		MI	19	1		
	Gly36Glu	MI	18	3		
	Gly36Trp	CA	17	1		
	Gly36Val	MC	20	1		
38	Arg38Gln	MC	24	1		y
	Arg38Trp	MC	22	1		
			23	1		
		MI	21	1		
39	Arg38Gln	MI	24	1		
41	Asp41Gly	MI	25	1		
42	Ile42Ser	CA	26	1		
43	Ser43Phe	CA	27	2		y
	Ser43Pro	CA	28	1		
44	Arg44Gln	CA	29	1		
46	Leu46Arg	MC	30	1		y
		MI	30	2		
	Leu46Pro	MI	31	3		
	Leu46Val	MI	32	2		
47	Gln47Arg	CA	33	1		y
			34	1		

48	Val48Glu	CA	35	1	
49	Ser49Phe	MI	36	2	
51	Gly51Ala	MC	38	1	y
	Gly51Arg	CA	41	1	
		MC	39	1	
			40	1	
		MI	42	1	
	Gly51Val	AS	37	1	
52	Cys52Arg	CA	43	1	y
		MI	43	1	
	Cys52Tyr	AS	44	1	
53	Val53Leu	CA	45	1	
54	Ser54Arg	MC	46	1	y
			47	1	
55	Lys55Thr	CA	49	1	y
	Lys55Arg	MI	48	3	
56	Ile56Thr	MI	50	1	
63	Thr63Pro	MI	51	1	
64	Gly64Ala	MC	52	2	y
	Gly64Val	MI	53	3	
66	Ile66Asn	MI	54	1	
68	Pro68Ser	MI	55	1	morning glory disc anomaly
72	Gly72Ser	MI	56	2	
73	Gly73Asp	CA	57	1	
74	Ser74Gly	CA	58	1	
		MC	58	1	
		MI	58	1	
76	Pro76Arg	MI	59	1	
78	Val78Ala	CA	60	3	
79	Ala79Glu	MI	61	1	
87	Ile87Arg	CA	63	1	y
	Ile87Lys	AS	62	1	
		CA	62	1	
93	Glu93Lys	CA	64	1	
99	Ala99Pro	MC	65	1	y
		MI	83	1	
118	Pro118Arg	MC	66	1	y
		MI	66	1	
	Pro118Gln	MI	67	2	
119	Ser119Arg	CA	68	1	y
			69	1	
			70	1	
			71	1	
		MI	68	2	

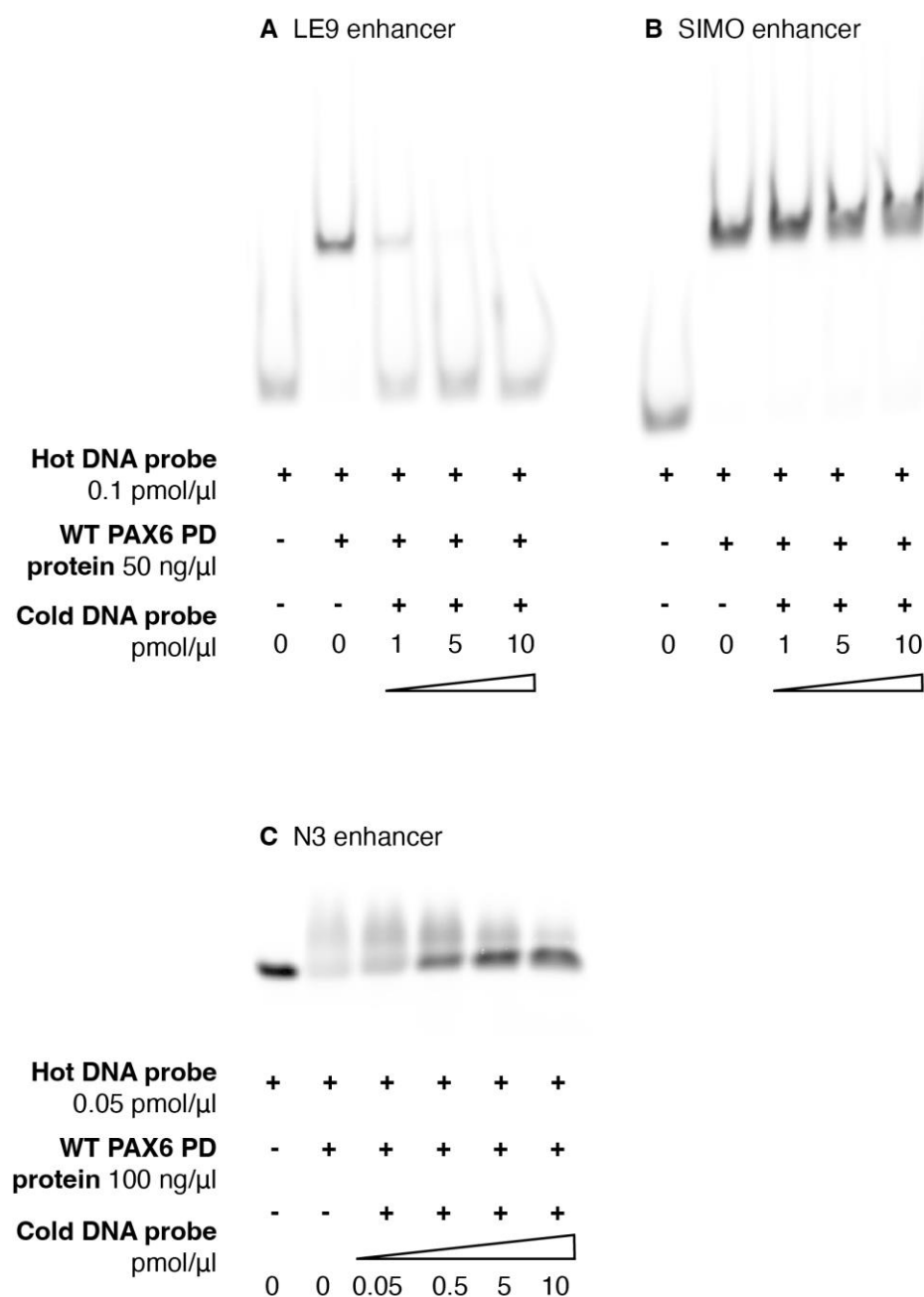
121	Ser121Leu	CA	73	2	y
	Ser121Pro	CA	72	1	
124	Asn124Lys	MC	74	1	y lens
			75	1	subluxation
			76	1	lens
			77	1	subluxation
			78	1	retinal detachment
126	Val126Asp	MI	79	1	
128	Arg128Cys	MI	80	3	y
			82	4	
	Arg128His	CA	84	1	
	Arg128Pro	MI	81	3	retinal detachment
178	Gln178His	CA	85	2	
208	Arg208Trp	CA	86	2	y
			87	2	
			88	1	
214	Arg214Gly	CA	89	1	y
	Arg214Ser	CA	90	1	
216	Ser216Pro	AS	91	3	
		MI	91	1	
242	Arg242Thr	MI	92	1	
		NO	92	1	
256	Val256Ala	MC	93	2	primary aphakia
258	Phe258Ser	MC	94	1	
260	Asn260Tyr	MC	95	1	
292	Ser292Ile	MI	96	1	
346	Pro346Ala	MI	97	1	
353	Ser353Ala	CA	98	1	
363	Ser363Pro	AS	99	1	
378	Gln378Arg	MC	100	1	optic nerve aplasia
381	Met381Val	MI	101	1	
391	Thr391Ala	MI	102	1	optic nerve aplasia
422	Gln422Arg	CA	103	1	y
		MI	104	1	
5a(13)	5aGln13Arg	AS	105	1	
5a(7)	5aVal7Asp	AS	106	1	y
			107	1	
			108	1	
		MI	109	1	

Grand Total

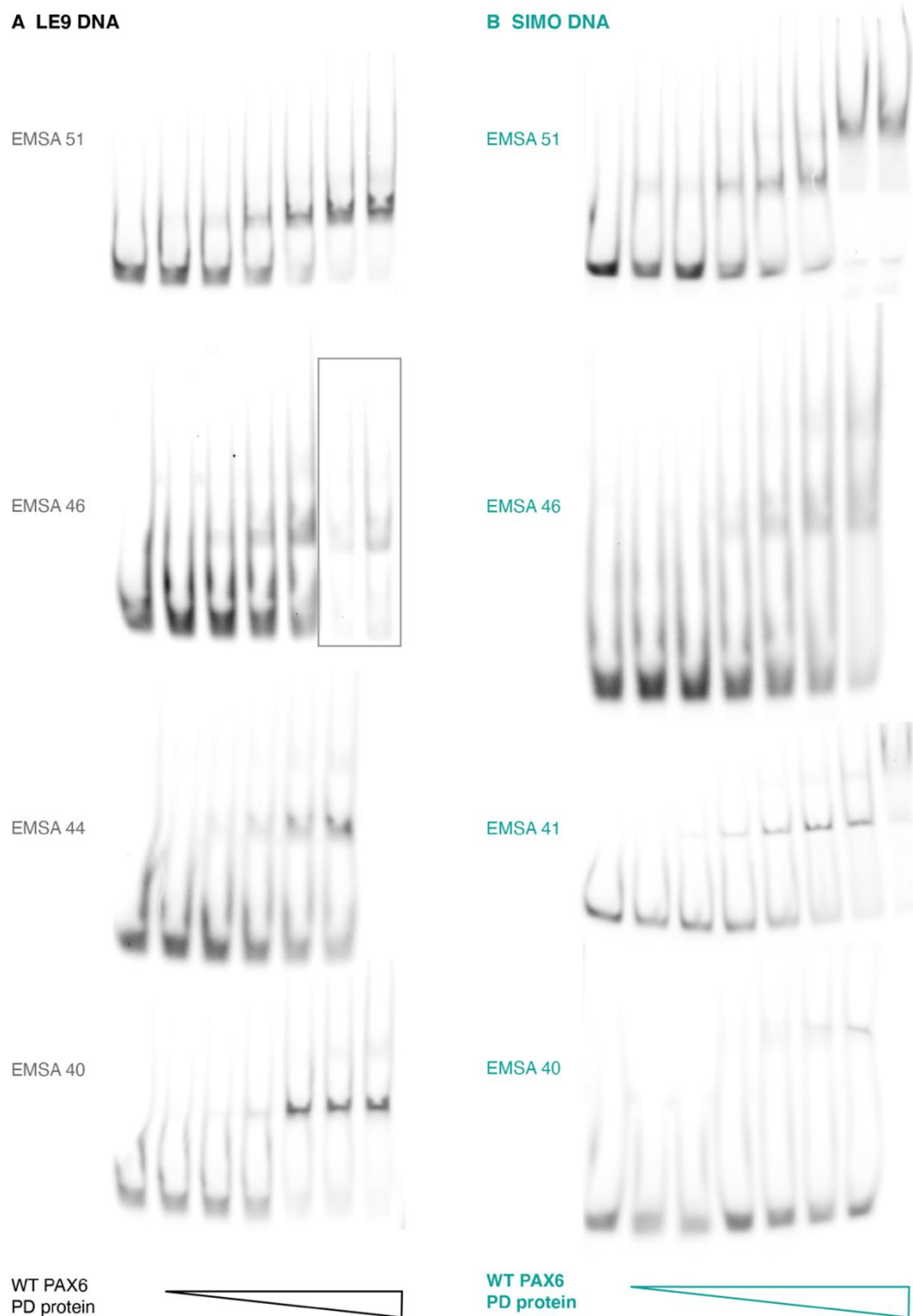
161

A.5 Supplemental EMSA data

A.5.1 Cold competition experiments

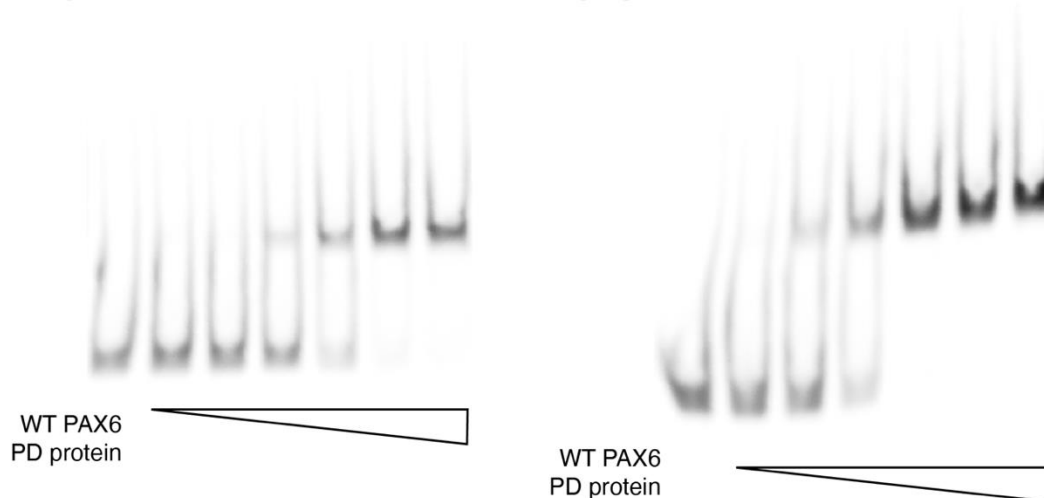
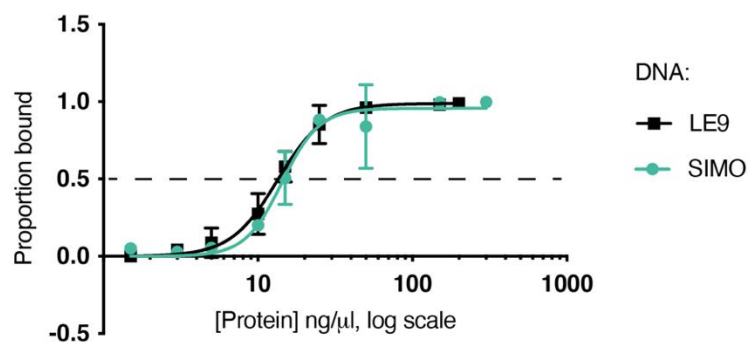


Supplementary Figure 3 Cold competition EMSA experiments in native PAGE gels with LE9 (A) and SIMO (B), and in agarose N3 (C), with wild type PAX6 paired domain (PD) purified protein. The binding of “hot” FAM-labelled DNA probe was competed out by increasing concentrations of “cold” unlabelled DNA in the case of LE9 and N3 but not SIMO. Given the results of other EMSA experiments, we could not explain the latter result.



Supplementary Figure 4 Supporting data for **Figure 22** showing the additional EMSA experiments which contributed to the binding curve calculations for this batch of PAX6 wild type protein and the LE9 and SIMO DNA sequences. In each gel, FAM-labelled DNA probe is present in all lanes at 0.1 pmol/ μ l. Concentration ranges of wild type PAX6 paired domain (PD) protein for the experiments with LE9 were: 0-80 ng/ μ l (EMSA 44 and 46) and 0-160 ng/ μ l (EMSA 40, 51, and

48, the latter being the one featured in Figure 10). In EMSA 46, 6 lanes were loaded, and the final lane spilled into the next lane (both of these are marked with a rectangular box); the bound and unbound fractions were calculated for each lane and the mean was used for the binding curve for that concentration. For the experiments with SIMO, the concentrations of PAX6 PD protein used were: 0-160 ng/μl (EMSA 40), 0-320 ng/μl (EMSA 46), and 0-640 ng/μl (EMSA 41, 51, and 48, the latter again being the one shown in Figure 10).

A LE9 DNA**B** SIMO DNA**C**

Supplementary Figure 5 Binding curves for the other PAX6 PD wild type protein batch which contributed to the EMSA experiments, binding to LE9 and SIMO DNA oligonucleotide probes. Representative EMSA gels are shown in (A) LE9, and (B) SIMO DNA. In each gel, FAM-labelled DNA probe is present in all lanes at 0.1 pmol/μl. The first lane is DNA probe alone, and subsequent lanes have increasing concentrations of purified PAX6 PD protein, indicated by the triangular wedge. As the protein concentration increases, so does the proportion of bound protein-DNA. (C) Quantification of EMSA binding curve experiments (n=4 for LE9, n=3 for SIMO), plotting proportion of bound labelled DNA-protein complex (vs unbound DNA probe) on the y-axis against increasing protein concentration (0-200 ng/μl for LE9, and 0-300 ng/μl for SIMO). A non-linear fit (specific binding with Hill slope) allowing calculation of the K_d values for each protein-DNA combination: for this batch of PAX6 wild type protein and DNA probes, K_d for LE9 = 13 ng/μl and K_d for SIMO = 14 ng/μl. The gels from all the EMSA experiments that contributed to this graph are in the next Figure.

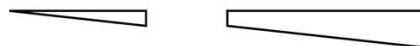
A LE9 DNA

EMSA 19

EMSA 21

WT PAX6
PD protein

EMSA 29

WT PAX6
PD protein**B SIMO DNA**

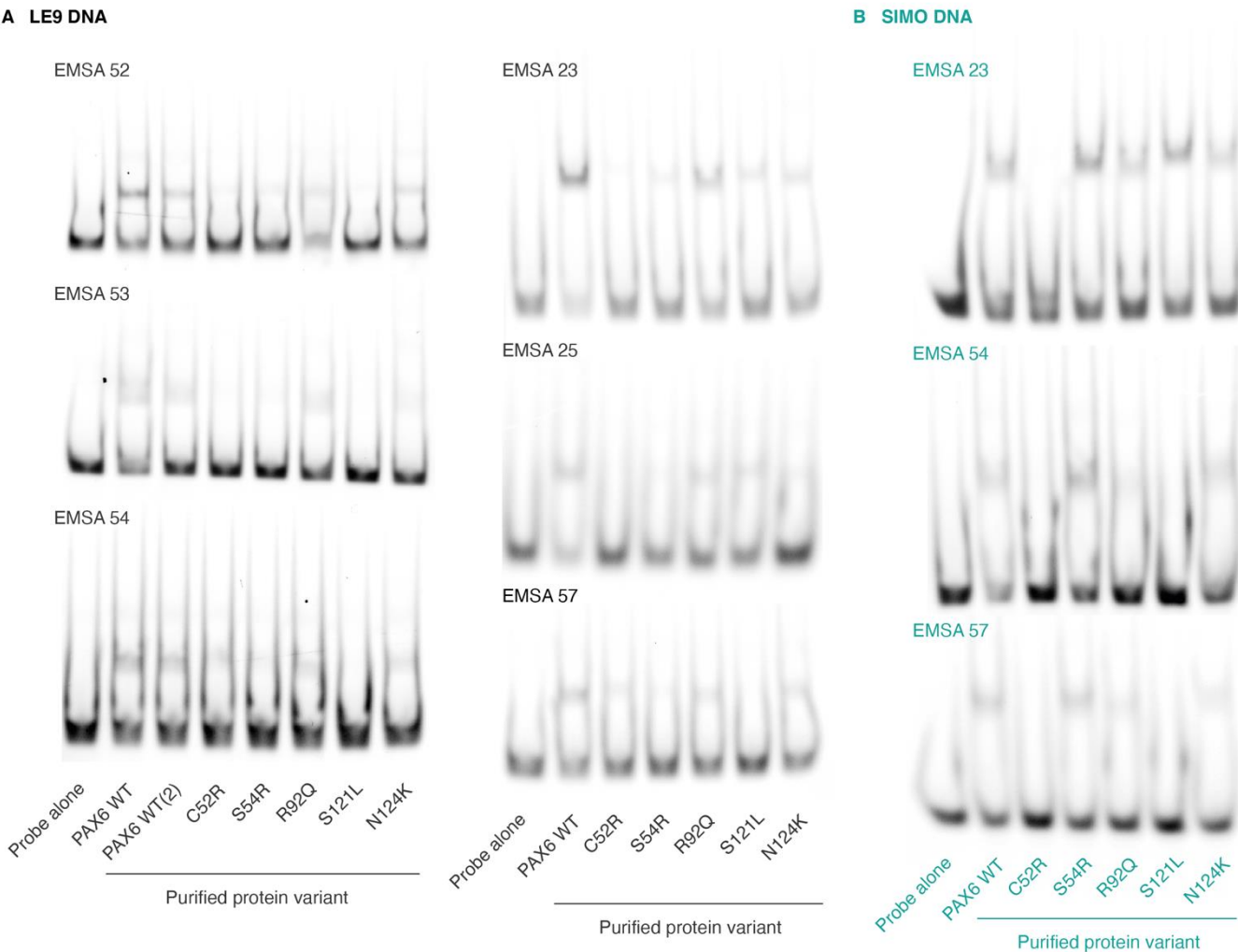
EMSA 19

EMSA 30

WT PAX6
PD protein

Supplementary Figure 6 Further data for **Supplementary Figure 5** showing the additional EMSA experiments which contributed to the binding curve calculations for this batch of PAX6 wild type protein and the LE9 and SIMO DNA sequences. In each gel, FAM-labelled DNA probe is present in all lanes at 0.1 pmol/ μ l. Concentration ranges of wild type PAX6 PD protein for the experiments with LE9 were: 0-150 ng/ μ l (EMSAs 19-21) and 0-200 ng/ μ l (EMSA 29). EMSA 29 was physically split over 2 different gels hence the repetition of the unbound, DNA probe-only lane (and break in the wedge indicating increasing protein concentration). For the experiments with SIMO, the concentrations of PAX6 PD protein used were: 0-50 ng/ μ l (EMSA 19), 0-150 ng/ μ l (EMSA 30), and 0-300 ng/ μ l (EMSA 20). EMSA 20 is not shown here, but is the representative gel shown for both DNA probes in **Supplementary Figure 5**.

A.5.3 Supplemental EMSA data for comparison of missense variant proteins



Supplementary Figure 7 Supporting data for **Figure 23** ('Effect of PAX6 variants on protein-DNA binding') showing the EMSA experiments which contributed to these results (these are in addition to EMSA 56, which is the representative gel shown in that figure). In EMSA 52-54 in (A) above, a second wild type PAX6 PD purified protein, labelled PAX6 WT(2), was included, but this lane was not part of the variant comparison experiment. These EMSA experiments shows the purified PAX6 proteins of interest (at the Kd concentration of the wild type, i.e. the concentration at which 50% is bound) binding to (A) the LE9 enhancer DNA sequences and (B) the SIMO enhancer DNA sequence. The lane labels correspond to all the gels above them, and in each experiment the first lane is the DNA probe alone, and subsequent lanes contain DNA with the purified protein variant indicated. (The numbers of the experiments are identifiers only and not significant; an LE9 and SIMO experiment were often run concurrently on the same gel, hence the overlap of some numbers.)

A.6 ANOVA results comparing EMSA binding of different PAX6 variants exported from GraphPad Prism

A.6.1 LE9

RM one-way ANOVA:

Table Analyzed	All variants vs LE9: normalised to WT				
	(EMSA #23,25,52- 54,56,57)				
Repeated measures ANOVA summary					
Assume sphericity?	No				
F	58.33				
P value	<0.0001				
P value summary	****				
Statistically significant (P < 0.05)?	Yes				
Geisser-Greenhouse's epsilon	0.3781				
R squared	0.9067				
Was the matching effective?					
F	0.7403				
P value	0.6207				
P value summary	ns				
Is there significant matching (P < 0.05)?	No				
R squared	0.01138				
ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Treatment (between columns)	4.839	6	0.8065	F (2.269, 13.61) = 58.33	P<0.0001
Individual (between rows)	0.06142	6	0.01024	F (6, 36) = 0.7403	P=0.6207
Residual (random)	0.4977	36	0.01383		
Total	5.398	48			
Data summary					
Number of treatments (columns)	7				
Number of subjects (rows)	7				
Number of missing values	0				

Multiple comparisons:

Number of families	1								
Number of comparisons per family	6								
Alpha	0.05								

Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Below threshold?	Summary	Adjusted P Value	B-?	
PAX6 WT vs. Probe alone	1.010	0.9893 to 1.031	Yes	****	<0.0001	A	Probe alone
PAX6 WT vs. C52R	0.8096	0.5803 to 1.039	Yes	****	<0.0001	C	C52R
PAX6 WT vs. S54R	0.8451	0.8093 to 0.8810	Yes	****	<0.0001	D	S54R
PAX6 WT vs. R92Q	0.5231	0.3041 to 0.7421	Yes	***	0.0007	E	R92Q
PAX6 WT vs. S121L	0.8993	0.7367 to 1.062	Yes	****	<0.0001	F	S121L
PAX6 WT vs. N124K	0.6199	0.4372 to 0.8025	Yes	****	<0.0001	G	N124K

Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	n1	n2	q	DF
PAX6 WT vs. Probe alone	1.000	-0.009987	1.010	0.005931	7	7	170.3	6
PAX6 WT vs. C52R	1.000	0.1904	0.8096	0.06573	7	7	12.32	6
PAX6 WT vs. S54R	1.000	0.1549	0.8451	0.01028	7	7	82.19	6
PAX6 WT vs. R92Q	1.000	0.4769	0.5231	0.06277	7	7	8.334	6
PAX6 WT vs. S121L	1.000	0.1007	0.8993	0.04660	7	7	19.30	6
PAX6 WT vs. N124K	1.000	0.3801	0.6199	0.05235	7	7	11.84	6

RM one-way ANOVA:

All variants vs SIMO: normalised to WT					
Table Analyzed		(EMSA #23,54,56,57)			
Repeated measures ANOVA summary					
Assume sphericity?	No				
F	11.24				
P value	0.0393				
P value summary	*				
Statistically significant (P < 0.05)?	Yes				
Geisser-Greenhouse's epsilon	0.1785				
R squared	0.7894				
Was the matching effective?					
F	2.269				
P value	0.1153				
P value summary	ns				
Is there significant matching (P < 0.05)?	No				
R squared	0.07378				
ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Treatment (between columns)	3.896	6	0.6494	F (1.071, 3.212) = 11.24	P=0.0393
Individual (between rows)	0.3932	3	0.1311	F (3, 18) = 2.269	P=0.1153
Residual (random)	1.040	18	0.05776		
Total	5.329	27			
Data summary					
Number of treatments (columns)	7				
Number of subjects (rows)	4				
Number of missing values	0				

Multiple comparisons:

Number of families	1
Number of comparisons per family	6
Alpha	0.05

Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Below threshold?	Summary	Adjusted P Value	B-?	
PAX6 WT vs. Probe alone	1.000	0.9688 to 1.032	Yes	****	<0.0001	A	Probe alone
PAX6 WT vs. C52R	0.9900	0.9353 to 1.045	Yes	****	<0.0001	C	C52R
PAX6 WT vs. S54R	0.06419	-0.5445 to 0.6729	No	ns	0.9801	D	S54R
PAX6 WT vs. R92Q	0.6329	0.3572 to 0.9086	Yes	**	0.0048	E	R92Q
PAX6 WT vs. S121L	0.6791	-0.8748 to 2.233	No	ns	0.3325	F	S121L
PAX6 WT vs. N124K	0.4234	0.2771 to 0.5698	Yes	**	0.0024	G	N124K

Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	n1	n2	q	DF
PAX6 WT vs. Probe alone	1.000	0.0004764	1.000	0.006443	4	4	155.3	3
PAX6 WT vs. C52R	1.000	0.01005	0.9900	0.01112	4	4	89.02	3
PAX6 WT vs. S54R	1.000	0.9358	0.06419	0.1238	4	4	0.5185	3
PAX6 WT vs. R92Q	1.000	0.3671	0.6329	0.05607	4	4	11.29	3
PAX6 WT vs. S121L	1.000	0.3209	0.6791	0.3160	4	4	2.149	3
PAX6 WT vs. N124K	1.000	0.5766	0.4234	0.02977	4	4	14.22	3

Appendix B Supplemental material for *MAB21L1* chapter

B.1 Publication ‘Monoallelic variants resulting in substitutions of MAB21L1 Arg51 Cause Aniridia and microphthalmia’

RESEARCH ARTICLE

Monoallelic variants resulting in substitutions of MAB21L1 Arg51 Cause Aniridia and microphthalmia

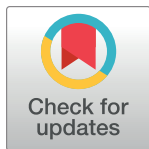
Hildegard Nikki Hall¹✉, Hemant Bengani¹✉, Robert B. Hufnagel², Giuseppe Damante³, Morad Ansari⁴, Joseph A. Marsh¹, Graeme R. Grimes¹, Alex von Kriegsheim¹, David Moore⁴, Lisa McKie¹, Jamalia Rahmat⁵, Catia Mio³, Moira Blyth⁶, Wee Teik Keng⁷, Lily Islam⁸, Meriel McEntargart⁹, Marcel M. Mannens¹⁰, Veronica Van Heyningen¹, Joe Rainger¹¹‡, Brian P. Brooks²‡, David R. FitzPatrick¹*

1 MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom, **2** National Eye Institute, National Institutes of Health, Bethesda, MD, United States of America, **3** Department of Medicine, University of Udine, Udine, Italy, **4** South East Scotland Genetic Service, Western General Hospital, Edinburgh, United Kingdom, **5** Ophthalmology Department, Hospital Kuala Lumpur, Kuala Lumpur, Malaysia, **6** University of Leeds, St. James's University Hospital, Leeds, United Kingdom, **7** Department of Genetics, Kuala Lumpur Hospital, Kuala Lumpur, Malaysia, **8** West Midlands Regional Genetics Service, Birmingham Women's and Children's NHS Foundation Trust, Birmingham, England, **9** Medical Genetics, St George's University Hospitals NHS Foundation Trust, London, United Kingdom, **10** Genome Diagnostics laboratory, Department of Clinical Genetics, Amsterdam University Medical Center, Amsterdam, The Netherlands, **11** Roslin Institute, University of Edinburgh, Edinburgh, United Kingdom

✉ These authors contributed equally to this work.

‡ These authors are joint senior authors on this work.

* david.fitzpatrick@ed.ac.uk



OPEN ACCESS

Citation: Hall HN, Bengani H, Hufnagel RB, Damante G, Ansari M, Marsh JA, et al. (2022) Monoallelic variants resulting in substitutions of MAB21L1 Arg51 Cause Aniridia and microphthalmia. PLoS ONE 17(11): e0268149. <https://doi.org/10.1371/journal.pone.0268149>

Editor: Anand Swaroop, National Eye Institute, UNITED STATES

Received: April 21, 2022

Accepted: October 6, 2022

Published: November 22, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0268149>

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: The RNA sequencing data is available under the GSE166078 series at the NCBI Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>).

Abstract

Classical aniridia is a congenital and progressive panocular disorder almost exclusively caused by heterozygous loss-of-function variants at the *PAX6* locus. We report nine individuals from five families with severe aniridia and/or microphthalmia (with no detectable *PAX6* mutation) with ultrarare monoallelic missense variants altering the Arg51 codon of *MAB21L1*. These mutations occurred *de novo* in 3/5 families, with the remaining families being compatible with autosomal dominant inheritance. Mice engineered to carry the p.Arg51Leu change showed a highly-penetrant optic disc anomaly in heterozygous animals with severe microphthalmia in homozygotes. Substitutions of the same codon (Arg51) in *MAB21L2*, a close homolog of *MAB21L1*, cause severe ocular and skeletal malformations in humans and mice. The predicted nucleotidyltransferase function of MAB21L1 could not be demonstrated using purified protein with a variety of nucleotide substrates and oligonucleotide activators. Induced expression of GFP-tagged wildtype and mutant MAB21L1 in human cells caused only modest transcriptional changes. Mass spectrometry of immunoprecipitated protein revealed that both mutant and wildtype MAB21L1 associate with transcription factors that are known regulators of *PAX6* (MEIS1, MEIS2 and PBX1) and with poly(A) RNA binding proteins. Arg51 substitutions reduce the association of wild-type MAB21L1 with TBL1XR1, a component of the NCoR complex. We found limited evidence for mutation-specific interactions with MSI2/Musashi-2, an RNA-binding proteins with

Funding: H.N.H. is funded by a Wellcome Trust fellowship (205171_Z_16_Z). D.R.F. is supported by Medical Research Council (MRC) University Unit program grant awarded to the University of Edinburgh. Funding for UK10K was provided by Wellcome under award WT091310.

Competing interests: The authors declare no conflicts of interest.

effects on many different developmental pathways. Given that biallelic loss-of-function variants in *MAB21L1* result in a milder eye phenotype we suggest that Arg51-altering monoallelic variants most plausibly perturb eye development via a gain-of-function mechanism.

Introduction

The gene *mab-21* was identified through its ability to rescue the *Caenorhabditis elegans* male abnormal 21 mutants, characterised by a homeotic transformation of the male-specific peripheral sense organs [1]. 11 human paralogs of *mab-21* have been identified each with a nucleotidyltransferase domain [2]. The best studied, *CGAS* [MIM 613973], functions in the innate immune system as a sensor of aberrant cytosolic DNA. The binding of short dsDNA induces a conformation change that activates enzymatic production of a cyclic dinucleotide which then functions as a second messenger in the interferon response cascade [3].

The *mab-21* paralog, *MAB21L1* [MIM 601280], is a single exon gene located in an intron of *NBEA* [MIM 604889] which is transcribed on the opposite strand. Biallelic loss-of-function mutations in *MAB21L1* cause a developmental disorder characterized by corneal dystrophy, microcephaly, cerebellar hypoplasia and genital anomalies [MIM 618479] [4,5]. The carrier parents of affected individuals were reported to be normal. *Mab21l1* null mice are viable but show severe bilateral microphthalmia with a small malformed lens and absence of the iris and ciliary body [PMID 12642482]. Null mice also show delayed calvarial development and male infertility with hypoplasia of the preputial glands [6,7]. Heterozygous mice apparently normal. Homozygosity for an early frameshift mutation in zebrafish *mab21l1* resulted in a late embryonic degeneration of the cornea and subsequently the lens [PMID 33570754]. The crystal structure of MAB21L1 indicates a cGAS-like capacity for catalytic activation via ligand binding although both the oligonucleotide activator and the nucleotide product are currently unknown [8].

We and others have previously reported heterozygous *de novo* missense mutations in *MAB21L2* [MIM 604357], the closest human homolog of *MAB21L1*, associated with severe bilateral eye malformations and skeletal anomalies [MIM 615877] [9–11]. These variants altered Arg51 with the most severe phenotype associated with Arg51Cys substitutions. A mouse model of this genotype resulted in a phenotype that recapitulated the human disease [12]. *Mab21l2* null mice have severe eye malformations and body wall defects with heterozygous null mice being normal [13].

Here we report monoallelic missense variants that are absent for gnomAD and result in substitution of Arg51 or, in a single case, Phe52 residues of MAB21L1 in families with severe aniridia [MIM 106210], a phenotype associated with monoallelic mutations in *PAX6* [MIM 607108], and/or microphthalmia. An apparently unrelated family has been recently reported with a heterozygous missense variant in *MAB21L1* identical to one that we have identified (c.152G>T p.(Arg51Leu)) associated with microphthalmia and aniridia [14] which provides strong support for the genotype-phenotype association. We present a mouse model of one of these mutations and study the effect of the mutant proteins on the transcriptome and protein interactome using inducible expression of tagged protein in human cells. The results are most consistent with a gain-of-function effect in Arg51-substituted MAB21L1 during embryogenesis.

Materials and methods

Recruitment, consent and mutation analysis

This project used clinical information and biological samples from individuals referred to the Medical Research Council (MRC) Human Genetics Unit Eye Malformation Study. Informed

written consent for research was obtained from all families. This cohort was collected and maintained using protocols approved by the Scotland A UK Multicentre Research Ethics Committee, references 06/MRE00/76 and 16/SS/0201. The causative variants were identified using a combination of sequencing approaches: whole exome analysis and candidate gene panel sequencing in the Wellcome Sanger Institute as part of the rare disease component of the UK10K project as described [15] and Sanger sequencing (for details see [Results](#) and [S1 Table in S7 File](#)). Samples from two families were referred following discussions with the corresponding author for clinical testing in the NHS South East Scotland Regional Genetics Services using MiSeq sequencing of a targeted gene panel which included *MAB21L1*. All variants were validated using Sanger sequencing of PCR products amplified directly from genomic DNA and were nomenclature-confirmed (<https://variantvalidator.org/>) ([S2 Table in S7 File](#)). All variant numbering is based on the human reference sequences GRCh38 NC_000013.11 (genomic, chr13). For each of the missense variants SIFT [16], PolyPhen [17], CADD [18] and REVEL [19] scores were generated using the DECIPHER web tool [20].

Structural analysis of mutations

The effects of missense mutations were modelled using the crystal structure of MAB21L1 (PDB ID: 5EOM) using FoldX 5.0 [21], which was recently shown to be the top-performing method for the identification of pathogenic missense mutations that affect protein stability [22], using all default parameters and averaging over 10 replicates.

Cloning, protein purification and enzymatic assay

Wild-type human MAB21L1 and the substitution p.(Arg51Leu) were amplified from control and patient DNA respectively and cloned in frame into the pGEX 6P1 vector (GE Life-Sciences). Purified protein was isolated from induced *E. coli* strain BL21 cultures as outlined in **Supplemental Materials and Methods**. Human OAS1 protein was used as a positive control in the enzymatic assay. A colorimetric method was used to quantitate the amount of pyrophosphate (PPi) product released upon completion of the enzymatic reaction as described [23] and detailed in **Supplemental Materials and Methods**. The resulting chromophore molybdenum blue produced was quantified by spectrophotometry at A580 nm.

Generation and RNA-based analysis of inducible human cell lines

Full-length human MAB21L1 and Arg51Leu and Arg51Gln substituted forms were amplified from the control and patient DNA and cloned downstream of green fluorescence protein (GFP) in the Gateway pcDNA-DEST53 vector according to the manufacturer's protocol, resulting in an N-terminal fusion protein. Stable cell lines were generated, selected and maintained using Human Embryonic Kidney (HEK)-293 cells with the Flp-In T-REx system (ThermoFisher) according to the manufacturer's guidelines. Details of the subcellular fractionation and Western blotting procedures are provided in **Supplemental Materials and Methods**. RNA sequencing used total RNA extracted from two biological replicates of each cell line after 12 hrs of 1 µg/ml tetracycline treatment using the RNeasy kit (QIAGEN). Random primed cDNA from poly(A) selected RNA was converted into an Illumina sequencing library using RNA Library Prep Kit from Illumina (E7420, NEB, USA) in conjunction with NEBNext® Multiplex Oligos for Illumina (E7335/E7500, NEB, USA). and single-end 50-base pair (bp) reads were generated using a NextSeq 500 (Illumina Inc, SY-415-1002). Transcript-level quantitation was performed using Salmon (v0.8.2) against the GRCh38 Ensembl reference transcriptome (release-89). Transcript-level counts were summarized to gene level using the

Bioconductor package tximport (v1.4.0). Differential expression analysis was performed with the Bioconductor package DESeq2 (v1.30.0) using the Wald significance tests.

Immunoprecipitation-mass spectrometry

Three biological replicates of HEK-293-Flp-In T-Rex cells tagged with EGFP, EGFP-MAB21L1 or mutant EGFP-MAB21L1 were seeded in T-75 flask in culturing media supplemented with 1 µg/ml tetracycline. Cells were harvested by trypsin-EDTA, washed by PBS after 12 hrs of tetracycline treatment. Cell lysis and GFP pulldown was performed using GFP Tag Immunomagnetic Beads (Sino Biologicals) according to manufacturer instructions. The pull-down beads were subjected to mass spectrometric analysis and raw data was analysed by the MaxQuant and Andromeda software package as described [24], using the pre-selected conditions for analysis (specific proteases, 2 missed cleavages, 7 amino acids minimum length). Detailed Mass Spectrometry analysis is provided in **Supplemental Materials and Methods**.

Immunoprecipitation-western blotting

HEK-293-Flp-In T-Rex cells tagged with EGFP, EGFP-MAB21L1 or mutant EGFP-MAB21L1 were seeded in T-25 flask in culturing media supplemented with 1 µg/ml tetracycline. Cells were harvested by trypsin-EDTA, washed by PBS after 12 hrs of tetracycline treatment and lysed with Nonidet P-40 lysis buffer (50mM Tris, pH 8.0, 150mM NaCl, 1.0% Nonidet P-40) in the presence of protease inhibitor (Roche Applied Science). For each immunoprecipitation, 400 µl of cell lysate were incubated with anti-TBL1XR1 antibody (ab24550, Abcam) and anti-MSI2 antibody (ab76148, Abcam) for 5 h at 4°C. Then 20 µl of Dynabeads protein A (Thermo Fischer) were added and rotated for 2 h at 4°C. Bound immune complexes were washed three times with phosphate-buffered saline. For immunoprecipitation of GFP-tagged proteins, Cell lysis and GFP pulldown was performed using GFP Tag Immunomagnetic Beads (Sino Biologicals) according to manufacturer instructions. The immune-complexes were analysed by Western blotting.

Generation and phenotyping of mouse model

All mouse work complied with United Kingdom Home Office regulations, with study protocols approved under Home Office project licences (60/4424, P1914806F). CRISPR-Cas9 gene editing methodology was used to introduce a targeted mutation of the Arg51 residue of *Mab21l1* in C57BL/6JCrI zygotes. The CRISPR design and breeding strategy for the colony are detailed further in **Supplemental Materials and Methods** and **S1 Fig**. Adult mutant and control mice were examined at 2–3 months of age unless otherwise stated, using: slit lamp bio microscopy, indirect ophthalmoscopy, Icare tonometry (intraocular pressure measurement) and endoscopic fundus imaging, all as described [25]. On fundus images, 2D optic disc size was measured semi-automatically using the Vampire Annotation Tool [26]. Optical coherence tomography (OCT) using Spectralis (Heidelberg Engineering) was performed as described [27]. For histology mice were culled and enucleated eyes were preserved in Davidson's fixative and then wax embedded, sectioned and stained with Hematoxylin and Eosin as previously described [28].

Results

Identification of MAB21L1 monoallelic missense variants altering Arg51

As part of the rare disease component of the UK10K Study [15] 384 mostly unrelated individuals with bilateral eye malformations were batch sequenced using a targeted pull-down of 1000

candidate genes, 100 of which had been chosen on the basis of their involvement in eye development. Filtering for rare variants within these 100 genes identified a heterozygous plausible deleterious variant c.152G>A p.(Arg51Gln) in MAB21L1 (ENST00000379919.6:c.152G>A, ENSP00000369251.4:p.Arg51Gln: Sift; Deleterious (0), PolyPhen; Probably damaging (0.999), CADD 30, REVEL 0.542) in a single individual (Family 511: II:1, [Fig 1A and 1B](#)) with bilateral profound aniridia and microphthalmia ([Table 1](#), [Fig 1C](#)). In this family the eye malformations were inherited as an autosomal dominant disorder and the MAB21L1 variant segregated with the phenotype ([Fig 1A and 1C](#)). The same c.152G>A p.(Arg51Gln) variant was identified in an individual referred from south-east Asian (Family 96571: II:1, [Fig 1A](#)) for clinical investigation of bilateral, severe microphthalmia ([Table 1](#)). This variant was subsequently also identified in an affected brother ([Fig 1A](#)). The affected offspring had inherited the variant from their unaffected mosaic father ([Fig 1A](#), [Table 1](#)).

Subsequence sequencing of DNA from unrelated affected individuals referred to the MRC Human Genetics Unit Eye Malformations Study identified an individual with sporadic partial aniridia and microphthalmia (Family 1434: II:1, [Fig 1A and 1C](#)) associated with *de novo* occurrence of MAB21L1 c.152G>T p.(Arg51Leu) (ENST00000379919.6:c.152G>T, ENSP00000369251.4:p.Arg51Leu: Sift; Deleterious (0), PolyPhen; Probably damaging (0.999), CADD 29.6, REVEL 0.682). The same allele was identified in an individual referred with familial aniridia (Family 3413: I:1, [Fig 1A](#)) however samples from other affected members of this family were not available for testing. A further *de novo* missense variant c.152G>C p.(Arg51Pro) (ENST00000379919.6:c.152G>C, ENSP00000369251.4:p.Arg51Pro: Sift; Deleterious (0), PolyPhen; Probably damaging (1), CADD 31, REVEL 0.694) was identified in an individual with a sporadic milder aniridia-spectrum eye malformation (Family 592). Finally, a variant affecting the adjacent codon, c.155T>G p.(Phe52Cys) (ENST00000379919.6:c.155T>G, ENSP00000369251.4:p.Phe52Cys: Sift; Deleterious (0), PolyPhen; Probably damaging (0.969), CADD 32, REVEL 0.745), was identified in an individual with sporadic microphthalmia and aniridia (Family 5531: I:1, [Fig 1A](#), [Table 1](#)). Parental samples were not available for testing in Family 5531. Sequencing chromatograms are provided ([S2 Fig](#)). None of these variants have been observed in publicly available variant databases.

Clinical phenotype. Pedigrees and clinical images are provided ([Fig 1A and 1C](#)), as well as detailed clinical descriptions ([Supplemental Clinical Descriptions](#)). The phenotypic features ([Table 1](#)) are summarised here. All 6 families had aniridia and/or microphthalmia. Aniridia was present in 4/6 families, microphthalmia in 4/6; and both in 3/6.

Aniridia. The aniridia was partial (moderate loss of iris tissue) in 2/6, unspecified in 1/6, and profound in 1/6. None of the families had a documented normal iris: 1/6 had a milder iris phenotype consisting of an irregular pupillary margin, and the remaining 1/6 had severe microphthalmia.

Microphthalmia/MAC spectrum. Of the MAC spectrum features seen in 4/6: 1/6 had a family member with bilateral severe microphthalmia, no view of the internal ocular structures and no perception of light. 1/6 had a microphthalmia with a chorioretinal colobomata and microcornea (in all 3 family members).

Other ocular features. 4/6 had nystagmus, 3/6 with confirmed foveal hypoplasia; in the remaining 2/6 only limited phenotypic data was available. 3/6 families had cataract, 2/6 with lens instability or subluxation. None had aphakia. 1/6 had glaucoma. 2/6 were described as having keratopathy or an opaque cornea, but as both individuals had microphthalmia and one had phthisis this is difficult to interpret. 2/6 had optic disc anomalies, including congenitally excavated and hypoplastic nerves.

Non-ocular features. There were no non-ocular phenotypic features of note. In particular, none of the affected individuals were reported to have genital anomalies.

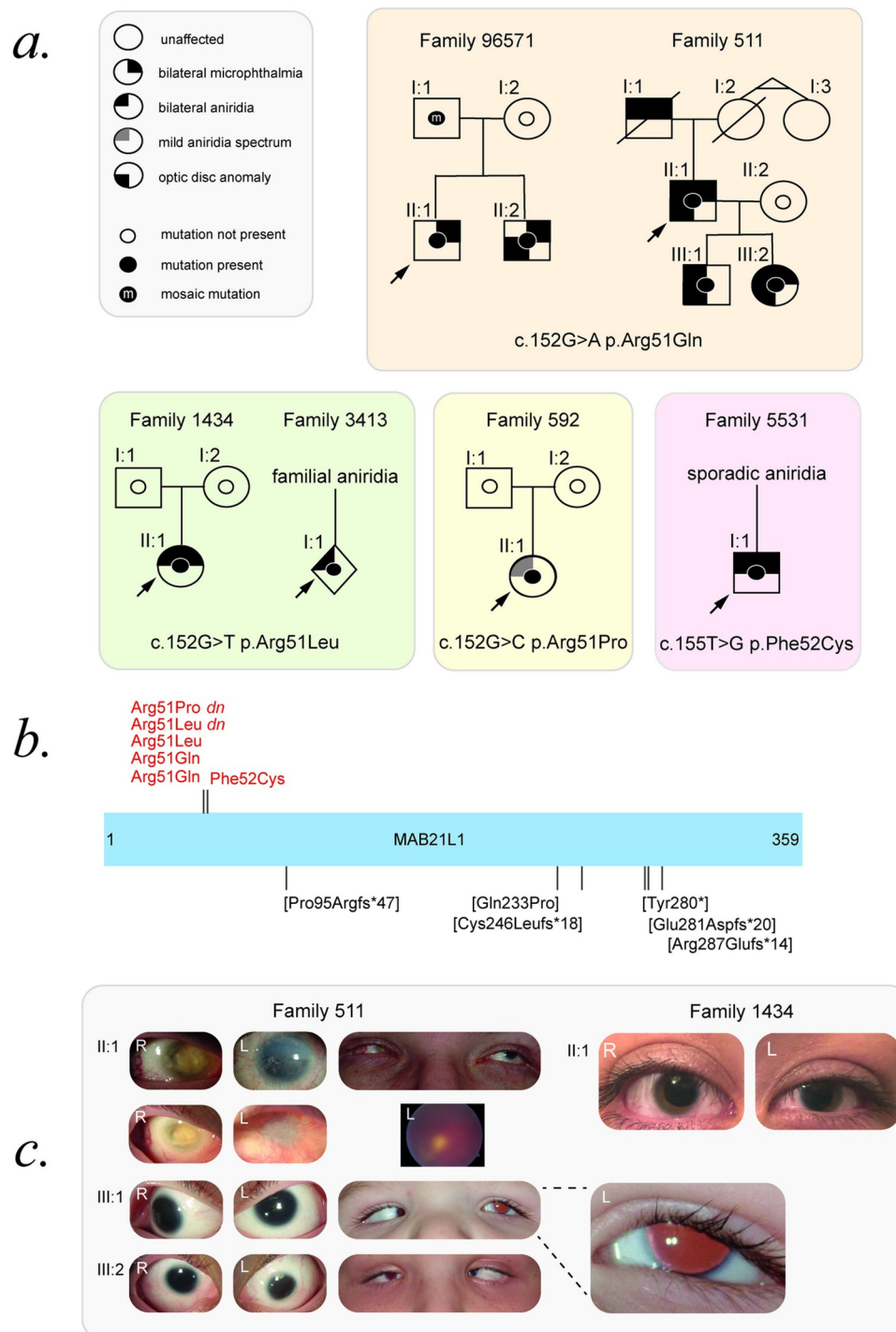


Fig 1. MAB21L1 Arg51 and Phe52 substitution causes microphthalmia and aniridia. A. Pedigrees are shown for the six families with MAB21L1 variants and bilateral microphthalmia and/or aniridia. The pedigrees are ordered by variant: c.152G>A p.(Arg51Gln) (orange shaded box), c.152G>T p.(Arg51Leu) (green shaded box), c.152G>C p.(Arg51Pro) (yellow shaded box) and c.155T>G p.(Arg52Cys) (pink shaded box). A key to the pedigree symbols is shown to the left (grey shaded box). B. A schematic of MAB21L1 represented as a linear bar and with the first and last amino acid residue numbered. The linear positions of all pathogenic variants are shown: The monoallelic variants in this study are detailed above (red text) and the published biallelic variants are detailed below (bracketed black text). C. Clinical images of individuals with MAB21L1 Arg51-related eye malformations. R, right eye; L, left eye. Family 511 all have profound aniridia, microcornea, choroidal coloboma (just visible in II:1's L fundus photo) and optic disc anomalies. The progression of disease in II:1 over one decade is shown between with the upper and lower photos, with

worsening of phthisis in the right and pannus in the left. An enlarged retroilluminated image of III:1's L eye is shown highlighting near-total aniridia. Individual 1434 II:1, showing bilateral partial aniridia and microphthalmia, worse on the L. Abbreviations: dn, *de novo*. Nucleotide and amino acid numbering are based on GenBank NM_005584.5 and GenPept NP_005575.1, respectively.

<https://doi.org/10.1371/journal.pone.0268149.g001>

Enzymatic function of MAB21L1

Wild-type and mutant (p.Arg51Leu) MAB21L1 was purified from *E. coli* in order to determine whether its predicted enzymatic function could be detected. Using an assay for colorimetric detection of pyrophosphate release [23] we could detect strong activity with 2'-5'-oligoadenylate synthase (OAS) purified by the same methods with ATP as substrate and RNA as activator molecule (Fig 2A). However, no MAB21L1-associated nucleotidyltransferase activity was detected using various nucleoside triphosphates as a substrate along with DNA or RNA as activator molecules (Fig 2A).

Structural analysis of MAB21L1 residue substitutions

We analysed the protein structural context of MAB21L1 substitutions reported above and the single reported biallelic missense variant [4] (Fig 3B) incorporating previously reported MAB21L2 residue substitutions associated with monoallelic or biallelic genotypes [10] (S3 Fig). There is a clear clustering of heterozygous variants, centred at Arg51. All the mutations are predicted to be destabilizing to protein structure (S4 Table in S7 File), except the recessive MAB21L2 substitution Arg247Gln; however, previous experimental work has conclusively demonstrated the destabilizing nature of this variant [8]. The pathogenicity of the recessive variants can almost certainly be explained by a simple loss of function caused by protein destabilization. However, while the heterozygous variants are all predicted to be somewhat disruptive, their clustering suggests a specific effect that involves this region. Thus, it seems plausible that the Arg51 substitutions are altering an interaction with another protein. It is also interesting that Arg51Pro has the mildest protein structural effect and appears to cause a milder phenotype than the other MAB21L1 heterozygous missense variants (Fig 1A, Table 1). Moreover, FoldX predicts a strong destabilizing effect of the Phe52Cys substitution, thus further indicating the structural relevance of this region.

Creation and Analysis of stable cell lines with inducible expression of wild-type and mutant MAB21L1

We created multiple independent tetracycline-inducible cell lines expressing wild-type MAB21L1 and Arg51leu and Arg51Gln variants as full-length GFP-tagged fusion proteins. Analysis of nuclear and cytoplasmic fractions revealed that MAB21L1 was present in both fraction with no evidence of mislocalisation of mutant forms (Fig 2B). RNAseq was used to assess the effect of MAB21L1 mutant variants on gene expression. Relatively few genes showed consistent differences between mutant and wild-type MAB21L1 (Fig 2C). Of these only SPARC (secreted protein acidic and rich in cysteine [MIM 182120]) had any link to eye disease [29,30]. SPARC was significantly upregulated in mutant cell lines compared to wild-type. SPARC has been reported as a PAX6 interactor [31] so we used transient transfection to over-express exogenous PAX6 in the wild-type and mutant MAB21L1 cells to see if this had any effect on the mutant-specific SPARC upregulation. Interestingly PAX6 had no effect on SPARC expression in mutant cells but induced significant downregulation in cells expressing wild-type GFP-MAB21L1 (Fig 2D). This would be consistent with wild-type MAB21L1 having a role in PAX6 mediated repression of SPARC and that function being lost with Arg51 substitution.

Table 1. Clinical and molecular features of individuals with *MAB21L1* heterozygous variants.

Family ID	96571		511			1434	3413	592	5531
Case	II:1	II:2	II:1	III:1	III:2	II:1	I:1	II:1	I:1
Sex	male	male	male	male	female	female	ND	female	male
GRCh38: NC_000013.11	g.35475987C>T		g.35475987C>T			g.35475987C>A	g.35475987C>A	g.35475987C>G	g.35475984A>C
GenBank: NM_005584.5	c.152G>A		c.152G>A			c.152G>T	c.152G>T	c.152G>C	c.155T>G
GenPept: NP_005575.1	p.(Arg51Gln)		p.(Arg51Gln)			p.(Arg51Leu)	p.(Arg51Leu)	p.(Arg51Pro)	p.(Phe52Cys)
Inheritance	paternal ^a	paternal ^a	ND ^b	paternal	paternal	de novo	ND	de novo	ND ^c
Growth									
Birth Weight z score	2.9	3.05							
Age at last assessment, years	7.5	2	48	13	11	23	ND	11	ND
Height z score	2.08	1.2							
Weight z score	1.31	0.76							
OFC, cm	51	49.5							
Ocular features									
Microphthalmia	BL severe	BL	BL		BL	BL			BL
Coloboma			LE small, inferior to disc	RE infero-temporal choroidal; LE small, temporal to disc	LE nasal choroidal				
Aniridia			BL	BL	BL	BL	BL		BL
Irregular pupil margin								BL	
Microcornea, mm			BL	BL: 6	BL: 6				
Keratopathy	LE opaque vascularised cornea	RE opaque cornea	BL progressive						
Glaucoma						BL with surgery LE			
Cataract	no view		LE previous lensectomy		BL with RE lenticonus, LE lens instability	BL with progressive subluxation		BL	
Nystagmus		LE	BL	BL	BL	BL		BL	
Foveal hypoplasia		LE	insufficient view			BL		BL	
Optic disc anomaly		LE	RE no view; LE congenitally excavated appearance, normal colour	BL gray hypoplastic	BL gray hypoplastic				
Myopia						BL high		BL	

(Continued)

Table 1. (Continued)

Family ID	96571		511			1434	3413	592	5531
Case	II:1	II:2	II:1	III:1	III:2	II:1	I:1	II:1	I:1
Sex	male	male	male	male	female	female	ND	female	male
Additional details	BL no view of internal structures	RE no view of internal structures	RE phthisis of unknown cause				limited phenotypic data		limited phenotypic data
Visual acuity	BL NPL	BL PL				BL HM			
Non-ocular features									
	hyperthyroidism (also in mother); normal MRI brain scan, echocardiogram and renal ultrasound					BL mild sensorineural hearing loss, normal MRI brain scan			

^a, the unaffected father was gonosomal mosaic (at a level of approximately 27 percent) for the variant.

^b, his deceased father had microphthalmia and aniridia.

^c, sporadic case of aniridia.

Abbreviations are BL, bilateral; HM, hand movements; LE, left eye; MRI, magnetic resonance imaging; ND, not determined; NPL, no perception of light; PL, perception of light; RE, right eye.

<https://doi.org/10.1371/journal.pone.0268149.t001>

Detection of wild-type and mutant-specific protein interactions

To identify MAB21L1 interactors that are specific either to wild-type or mutant protein we performed immunoprecipitation followed by mass spectrometry (IP-MS) on biological triplicates derived from independent clones of the inducible cell lines expressing GFP alone, wild-type MAB21L1, Arg51Leu MAB21L1 and Arg51Gln MAB21L1. More than 1000 proteins were identified using IP-MS but most were non-specific or inconsistently associated with the genotypes (Fig 3A). The levels of MAB21L1 were similar in wild-type and mutant pull-down samples suggesting uniform pulldown (Fig 3B). 72 proteins showed consistent association with all forms of MAB21L1 with no association with GFP alone. Pathway analysis of these proteins revealed a significant over-representation of RNA-binding proteins and TALE-like homeodomain containing proteins (Table 2). Indeed, four of the five most abundant proteins were transcription factors of this latter class (MEIS1, MEIS2, PBX1 and PBX3) (Fig 3D). A component of the NCor co-repressor complex, TBL1XR1, was the only wild-type MAB21L1-specific protein identified (Fig 3E). Western blot of the GFP IP using an antibody raised against TBL1XR1 showed differential, but not exclusive, binding to wild-type MAB21L1 (Fig 3F). Reciprocal immunoprecipitation using anti-TBL1XR1 antibody was not able to detect either mutant or wild-type forms of MAB21L1 using western blot (S4 Fig). MSI2/Musashi-2 was one of only three proteins showing apparently exclusively association with the mutant forms of MAB21L1 (Fig 3G, the other proteins being LRRFIP1 and GALNT2, S5 Fig). Although we were unable to confirm this interaction using reciprocal IP with a MSI2 antibody on western blot (Fig 3H), the reciprocal IP-MS using this antibody identified MAB21L1 derived peptides in each replicate of the mutant forms of MAB21L1 but only one of the wild-type replicates. This can be considered only limited evidence of a gain of function interactions since the GFP-only MS-IP showed peptides in two of the replicates, presumably derived from endogenous MAB21L1 in HEK293 cells.

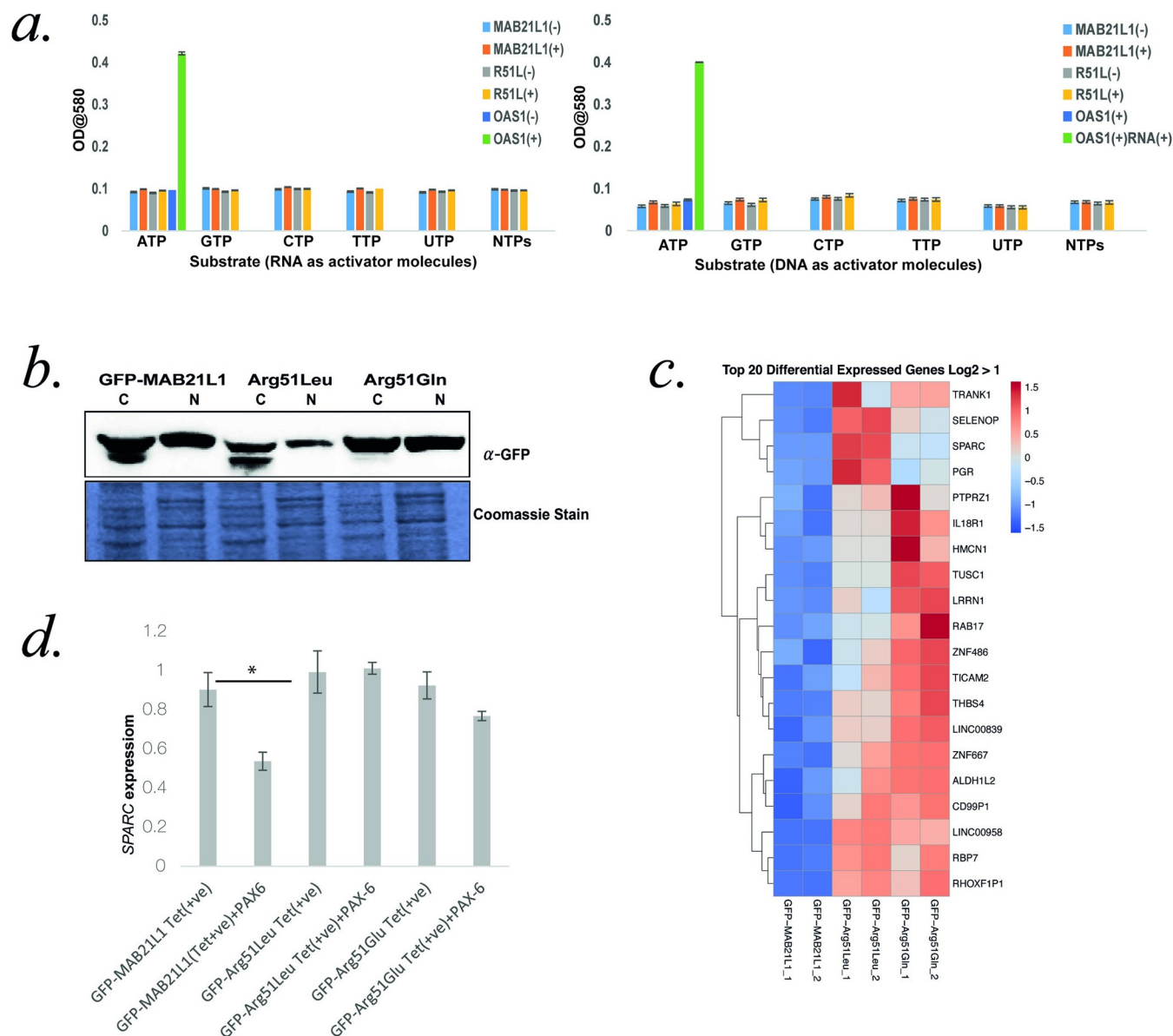


Fig 2. A. Nucleotidyltransferase activity: Graph showing the absence of nucleotidyltransferase activity in MAB21L1 and its mutant form Arg51Leu purified protein. OAS1 protein purified in the same way is a positive control and when incubated with ATP and double-stranded RNA (dsRNA), significant pyrophosphate release is detected indicating nucleotidyl transferase activity. MAB21L1 and Arg51Leu showed no activity with either ATP, CTP, GTP, UTP used as substrate separately or as an equal mixture of NTPs using DNA or RNA as an activator. The error bars represent standard errors. **B: Cellular fractionation:** Western blot analysis of cytoplasmic (C) and nuclear (N) extracts from HEK293-Flp-In cells with Tetracyclin (TET) inducible expression of GFP-tagged wild-type and mutant MAB21L1 (Arg51Leu and Arg51Gln). Wild type and mutant proteins were present in cytoplasm (C) as well as nuclear fraction (N) as detected by anti-GFP antibody. Representative Coomassie stain gel image is shown. **C: Differential Gene Expression:** Gene expression analysis by RNA Sequencing performed on GFP-tagged wild-type and mutant MAB21L1 (Arg51Leu and Arg51Gln) cells. Heatmap showing top 20 differentially expressed genes in the datasets ($p_{adj} < .05$ and $\text{Log}_2 F > 1$). The RNA sequencing data is available under the GSE166078 series at the NCBI Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>). **D: Effect of PAX6 overexpression on SPARC transcripts levels:** SPARC transcripts levels were quantified using quantitative RT-PCR using cells expressing GFP tagged Wild type and mutant MAB21L1 with or without overexpressing PAX6. GAPDH transcripts levels were used as normalization control. The levels of SPARC transcripts were significantly reduced in GFP tagged Wild type MAB21L1 cells in presence of overexpressed PAX6. There was no significant difference in the mutant cells in presence or absence of overexpressed PAX6.

<https://doi.org/10.1371/journal.pone.0268149.g002>

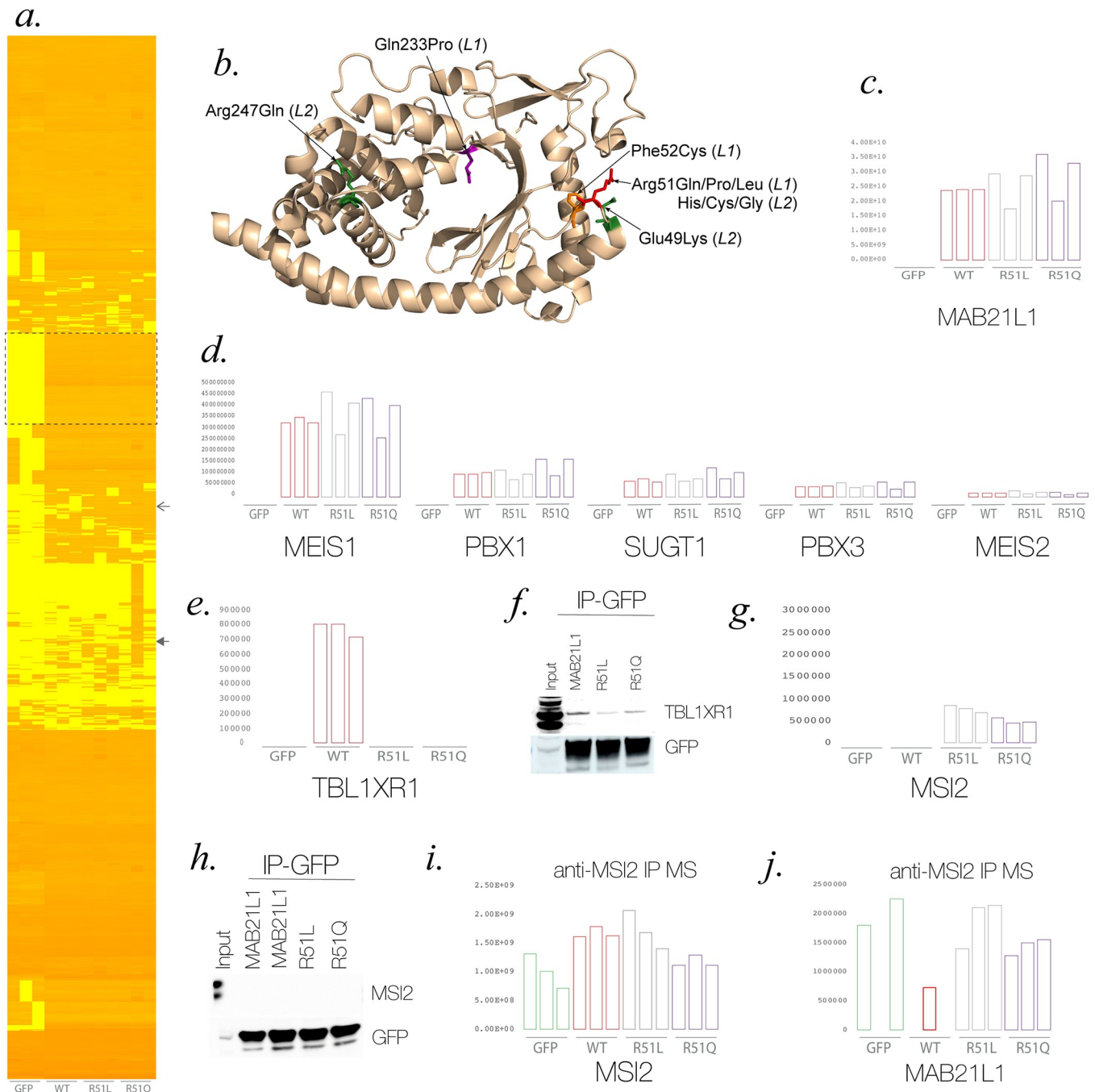


Fig 3. The Effect of disease associated missense variants on protein structure and interactions. A. A heatmap of the log-transformed quantitative mass spectrometry (MS) results of biological triplicates of anti-GFP immunoprecipitates (IP) of control (GFP), tagged wild-type (WT) and mutant MAB21L1 (R51L and R51Q) from HEK293 cells. B. Representation of the structure of MAB21L1 with the position of the amino acid substitutions seen in *MAB21L1* (L1) and *MAB21L2* (L2) annotated. C-E Graphs showing the levels of the following classes of proteins in the GFP IP-MS from biological triplicates: C. MAB21L1. D. The five most abundant proteins interacting with WT, R51L and R51Q forms of MAB21L1 (MEIS1, PBX1, SUGT1, PBX3 & MEIS2), the dotted line box indicates the position of this class of protein on the heatmap. E. Wild-type specific interactor (TBL1XR1) and other components of the NCor complex (NCOR & HDAC3), the position in the heatmap is indicated by the closed arrowhead. F. Western blot analysis of the anti-GFP IP using the anti-TBL1XR1 antibody showing differential but not exclusive binding of the wild-type MAB21L1 compared to the mutant forms. G. Mutant specific interactor MSI2/ Musashi-2, the position in the heatmap is indicated by the open arrowhead. H. Western blot analysis of the anti-GFP IP using the anti-MSI2 antibody was unable to detect interaction with wild-type or mutant forms of MAB21L1. I-J. anti-MSI2 IP-MS analysis. I. MSI2-derived peptides were present in all replicates and cell-lines in the anti-MSI2 IP. J. MAB21L1-derived peptides were detectable in all replicates of the mutant forms of MAB21L1 but in only one replicate for the wild-type. Surprisingly peptides derived from endogenous MAB21L1 were detectable in two of the three GFP-only biological replicates.

<https://doi.org/10.1371/journal.pone.0268149.g003>

Table 2. Significantly enriched terms relating to MAB21L1-interacting proteins using DAVID Functional Annotation Chart.

Category	Term	Genes	Fold Enrich	Bonferroni
GO:0044822	poly(A) RNA binding	RBM26, NCBP1, POP1, SPATS2, MRPS21, ERL1, MRPL37, NPM3, DIAPH1, SARS2, ZC3H7A, RBMX2, NUSAP1, FASTKD5, FLNB, METTL16, SNTB2	3.68	1.07E-03
UP_KEYWORDS	Phosphoprotein	RBM26, PDXDC1, POP1, FLII, GPS1, RNF219, STK4, PFAS, SMG5, TOR1AIP1, STK3, DNAJB1, ATXN3, IPO8, ZC3H7A, SALL2, RBMX2, PHKG2, NUSAP1, FLNB, METTL16, SKP2, MARK3, CEP55, EIF2A, JAGN1, NCBP1, HMGCS1, RIOK3, ZNF281, GLMN, SPATS2, RECQL, PYCR2, CDC7, ERL1, HAUS5, NPM3, GTF2F2, GPN1, DIAPH1, FASTKD5, RLIM, UBA2, CDC42EP1, GARS, SUGT1, SNTB2	1.66	1.12E-03
UP_KEYWORDS	Acetylation	RBM26, FLII, STK4, SMG5, STK3, NUSAP1, FLNB, SKP2, EIF2A, NCBP1, HMGCS1, GLMN, RECQL, PYCR2, HAUS5, NPM3, GTF2F2, GPN1, DIAPH1, PRPF4, PPP5C, SARS2, GGCX, FASTKD5, RLIM, UBA2, GARS, SUGT1	2.34	1.89E-03
UP_SEQ_FEATURE	DNA-binding region: Homeobox; TALE-type	MEIS1, PBX3, MEIS2, PBX1	56.51	1.25E-02
UP_KEYWORDS	Nucleus	POP1, FLII, GPS1, MRFAP1, STK4, SMG5, TOR1AIP1, STK3, DNAJB1, ATXN3, IPO8, ZC3H7A, SALL2, NUSAP1, SKP2, ZSCAN18, ZNF460, NCBP1, ZNF281, RECQL, PBX3, CDC7, NPM3, GTF2F2, MEIS2, GPN1, PBX1, PRPF4, PPP5C, MEIS1, RLIM, UBA2, SUGT1, INTS9	1.85	0.015

Fold Enrich, fold enrichment over homo sapiens background list; Bonferroni, p value corrected for multiple testing using the Bonferroni method.

<https://doi.org/10.1371/journal.pone.0268149.t002>

Generation and phenotyping of mice with *Mab21l1* p.Arg51Leu substitution

We used zygotic genome editing to create a mouse line harbouring *Mab21l1* p.Arg51Leu substitution (*Mab21l1*^{R51L/+}) (S1 Fig). The line was maintained as a co-isogenic strain on this C57BL/6JCrI background. Heterozygous mice were intercrossed to produce viable and fertile homozygotes (*Mab21l1*^{R51L/R51L}). The ratios of offspring were consistent with Mendelian genetics (S3 Table in S7 File). *Mab21l1*^{R51L/+} heterozygote mice showed anomalous, excavated optic discs (Fig 4A and 4B) in 13/13 heterozygotes, confirmed as bilateral in 9/13 (all fundus images are shown in S6 Fig). Quantitative analysis of fundal images show the discs were enlarged compared to wild type (n = 8 *Mab21l1*^{R51L/+}, n = 4 WT, p = 0.00004). The excavated optic disc anomaly was observed on Optical Coherence Tomography and histological sectioning (Fig 4C and 4D). Intraocular pressure tested in a subset of mice (n = 2 WT, n = 5 *Mab21l1*^{R51L/+}, n = 3 *Mab21l1*^{R51L/R51L}) was within the normal range, consistent with the optic nerve phenotype being a developmental defect rather than glaucomatous phenomenon. Slit lamp examination of the iris and anterior segment appeared normal and the mice displayed no other apparent abnormalities. All homozygous *Mab21l1*^{R51L/R51L} mice had a severe bilateral panocular eye malformation (Fig 4A and 4B). These included microphthalmia with disorganised anterior and posterior segments. There was marked hyperplasia of pigmented uveal tissue that obscured any possibility of a fundal view on examination. Histological sectioning revealed abnormalities of the cornea, iris, ciliary body, lens, retina and optic nerve (Fig 4C). The most severely affected eyes had only a rudimentary lens and retina, and optic nerve aplasia.

Discussion

Classical aniridia is a highly distinctive autosomal dominant disorder diagnosed in infancy by the combination of absence of the iris and foveal hypoplasia [32,33]. In adult life a progressive opacification of the cornea results in the relentless loss of their vision; this is currently untreatable and represented a particularly challenging aspect of the disorder for both affected

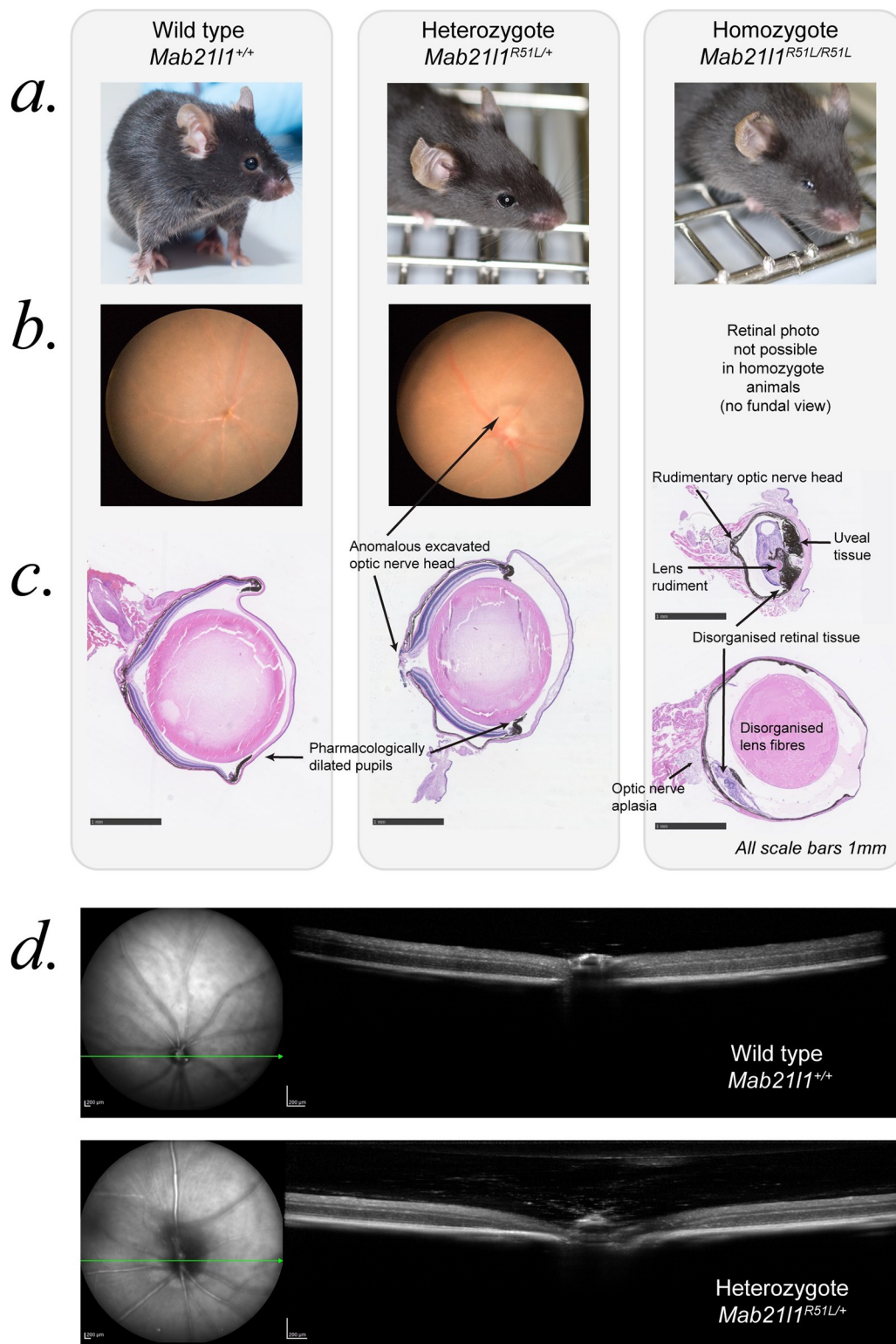


Fig 4. *Mab21l1*^{R51L} mouse phenotype. Left hand column shows wild type (WT) mice, middle column *Mab21l1*^{R51L/+} heterozygous mice and right hand column homozygous *Mab21l1*^{R51L/R51L} mice. All mice were examined as young adults, at 2–3 months of age. (A) External photographs, showing normal external appearance of heterozygous and severe microphthalmia in homozygous mice. (B) Representative retinal photographs showing the anomalous, excavated optic disc phenotype (arrow) present in all the 13 heterozygous mice examined, and none of the WT (further images in S6 Fig).

No retinal view (or OCT) was possible in homozygous animals as all eyes were microphthalmic, often with uveal tissue obscuring the cornea. (C) H&E-stained wax sections of the mouse eyes, with the abnormal features for each genotype labelled. The excavated optic nerve anomaly of the heterozygous mice is clearly seen. Note these mice had recently-administered dilating drops for fundal examination, but the iris appeared normal on slit lamp examination. Homozygous mice had a severe panocular eye malformation including microphthalmia, severely disorganised retina and uveal tissue, along with hypoplasia or aplasia of both the optic nerve and lens. Eyes from two different age-matched homozygous mice are shown to illustrate the spectrum of microphthalmia. Scale bars = 1 mm. (D) Optical coherence tomography (OCT) of WT and heterozygous mice, showing the enlarged, excavated optic nerve, with some persistent fetal vasculature seen above the optic nerve head.

<https://doi.org/10.1371/journal.pone.0268149.g004>

individuals and ophthalmologists. More than 90% of individuals with aniridia have heterozygous mutations detectable at the *PAX6* locus that appear to result in loss-of-function [34–36]. Rare monoallelic missense variants at specific residues within the PAIRED domain of *PAX6* cause a significantly more severe form of classical aniridia with microphthalmia [37] with Fig 1 of that paper demonstrating the striking similarity to the individuals with *MAB21L1* mutations reported here. The phenotypic similarity includes the nature of the iris phenotype (with a spectrum including moderate and profound absence of iris, with the loss not limited to one specific part of the iris such as in Gillespie syndrome), the severity of microphthalmia and the other associated ocular features including cataract, lens instability and foveal hypoplasia. The molecular basis of this worse-than-null phenotypic effect in the *PAX6* missense cases is unknown but is assumed to be the consequence of altered *PAX6* interaction with DNA and/or co-binding partners such as SOX2 [38–40].

Given the phenotypic similarity of monoallelic missense variants resulting in *PAX6* and Arg51 *MAB21L1* substitutions we hypothesize that the developmental function of the wild-type *PAX6* and *MAB21L1* proteins are interdependent. In this regard, it is interesting to consider similarity between the male-specific sensory rays mis-specification that characterizes the *mab-21* mutant class in *C. elegans* [1] and that seen in *mab-18* mutants caused by a mutation at the *vab-3* (*PAX6*) locus [41]. It is also striking that four of the five most abundant proteins recovered by immunoprecipitating *MAB21L1* were the transcription factors MEIS1, MEIS2, PBX1 and PBX3. These transcription factors act as both activators [42–45] and co-binding partners [46,47] of each other and *PAX6*. Although these interactions are probably relevant to the developmental role of *MAB21L1* it is difficult to link them to disease as they were not significantly altered by either of the Arg51 substitutions we studied.

Using IP-MS we could identify only one protein, TBL1XR1 that interacted with wild-type *MAB21L1* but not at all with either mutant form. TBL1XR1 mediates proteasomal degradation of NCor corepressor complex [48]. Western blotting suggested that this interaction with mutant protein was reduced rather than completely ablated (Fig 3F). Although disruption of such an interaction is a reasonable candidate for perturbing a developmental transcriptional cascade it should be noted that this interaction would be completely ablated in the individuals with homozygous loss-of-function mutations in *MAB21L1* but these individuals have significantly milder anterior segment anomalies than the individuals carrying Arg51 heterozygous missense variants. We do not therefore consider this loss of protein-protein interaction to be the likely mechanism of disease in the affected individuals we present here.

Three proteins, GALNT2, MSI2 and LRRFIP1, showed apparent mutation-specific interactions suggesting a possible gain of function effect (S5 Fig). GALNT2 is a N-acetyl-d-galactosamine-transferase 2 which localizes to the Golgi which has not previously been implicated in *PAX6* function or eye development. Biallelic loss of function mutations in *GALNT2* [MIM 602274] cause a neurodevelopmental disorder of O-linked glycosylation [MIM: 618885] [49]. LRRFIP1 [MIM: 603256] is an RNA binding protein that binds double stranded RNA [50]. LRRFIP1 has roles both as a viral sensor in innate immunity [51] and as a regulator of

canonical WNT signalling in development [50,52]. There is no direct evidence that LRRFIP1 is involved in eye development or PAX6 function.

MSI/Musashi-2 [MIM 607897] is also an RNA binding protein that regulates the translation of gene products through binding their 3'UTR regions. Its role in both cancer [53] and developmental systems [54–56] has been widely studied and it has been shown to form a complex with SOX2 [57]. This seemed a good candidate as a gain-of-function interaction but we could identify only limited evidence for this using reciprocal IP-MS and the interaction was not detectable using western blot analysis following IP. Musashi-1 was also identified as a MAB21L1 interactor but did not show any difference between mutant and wild-type proteins (S5 Fig). Musashi-1 and -2 are required for normal photoreceptor development [58].

The identification of RNA binding proteins as an overrepresented class in the list of mutation agnostic MAB21L1 interactors may be of significant functional relevance. The crystal structure of MAB21L1 suggested that activation of the nucleotidyltransferase activity required a conformational change similar to that of the mab-21 paralog cGAS (S7 Fig). The authors could demonstrate that MAB21L1 bound double stranded RNA but with significantly lower affinity than cGAS [8]. Our work would support their conclusion that any generic oligonucleotides are unlikely to function as MAB21L1 inducers. They go on to suggest that specific mRNA-RNA-binding protein complexes species may bind to MAB21L1 to induce the enzymatic activity. The fact the LRRFIP1 has cGAS like functions in sensing viral dsRNA in the cytoplasm is interesting but given that this interaction is only seen with Arg51 substitutions make this an unlikely endogenous activator. Our favoured hypothesis is that RNA-bound Musashi-2 functions as the *in vivo* activator, and a competitive antagonistic effect of Musashi-2 binding in the mutant is an important gain-of-function interaction. All the above protein-protein interaction experiments must be treated with caution given that they were performed using GFP-tagged peptides that were very highly and inducibly expressed in HEK293 cells. We suggest that future work should focus on identifying wildtype and mutant MAB21L1-specific interactions under more physiologically and developmentally relevant tissues to identify the molecular basis of the disorder.

There are several notable features regarding the phenotypes associated with monoallelic and biallelic mutation of *Mab21l1* in mice. The phenotype in *Mab21l1*^{R51L/+} mice is milder than in humans, but the optic disc anomaly is both seen in human cases and consistent with PAX6-associated disease [59]. In the process of this work Seese and colleagues [14] reported the identification of c.152G>T p.(Arg51Leu) variant in MAB21L1 in two affected members of a family which co-segregated with microphthalmia and aniridia. This family appear to be phenotypically very similar to those we have identified and this is further support for the causative nature of substitutions affecting MAB21L1 Arg51. It is interesting that the severe eye malformations in *Mab21l1*^{R51L/R51L} animals resemble those reported in *Mab21l1* null animals [6]. In contradistinction the eye phenotype in null humans is significantly milder than that seen in mice or indeed heterozygous Arg51 substitutions in humans. Together this suggests that there may be significant differences in MAB21L1/*Mab21l1* dosage sensitivity and disease mechanism between mice and humans.

Supporting information

S1 Fig. *Mab21l1* CRISPR design. (A) Schematic to illustrate the CRISPR-Cas9 sgRNA guide sequences and their relative locations to the Arginine 51 encoding region of the *Mab21l1* locus. (B) Sanger sequencing chromatogram of PCR performed using genomic DNA prepared from a gene edited mouse. The *Mab21l1* p.Arg51Leu mutation was introduced (highlighted region), along with the silent substitutions in the flanking regions (red asterisks), which were

specific to the repair template.
(DOCX)

S2 Fig. Sequences of highly specific *MAB21L1* heterozygous variants associated with microphthalmia and/or aniridia. An allelic series of *MAB21L1* heterozygous variants at position c.152G was identified in a total of five probands: two familial cases with c.152G>A (p.(Arg51Gln), chromatograms in orange shaded box), two sporadic cases with *de novo* inheritance of either the recurrent variant c.152G>T (p.(Arg51Leu), upper chromatogram in green shaded box) or the novel variant c.152G>C (p.(Arg51Pro), chromatogram in yellow shaded box), and one familial case with unknown genotypic inheritance of the recurrent variant c.152G>T (p.(Arg51Leu), lower chromatogram in green shaded box). Additionally, a sporadic case with unknown genotypic inheritance was heterozygous for the novel variant c.155T>G (p.(Phe52Cys), chromatogram in pink shaded box) in the adjacent 3' codon. The chromatogram for each proband is shown, with the Family ID and pedigree case ID detailed to the right. Sanger sequencing was used to screen for and/or validate the variant in each proband, and to test all of the available relatives (data not shown), which established segregation with the phenotype. The schematic (upper right) illustrates the highly specific positioning of the four variants identified. Nucleotide and amino acid numbering is based on GenBank: NM_005584.5 and GenPept: NP_005575.1, respectively.
(DOCX)

S3 Fig. Dominant and recessive variants of MAB21L1 and MAB21L2. Schematic representations of the linear form of MAB21L1 (blue filled bar) and MAB21L2 (purple filled bar) are shown, with the first and final amino acids numbered for each protein. For both MAB21L1 and MAB21L2 the linear positions of all published pathogenic variants are detailed on each cognate protein schematic, with the dominant heterozygous variants shown above and the recessive biallelic variants shown below. The MAB21L1 variants identified in this study are all dominantly inherited and are shown in red text. Abbreviations: *dn*, de novo. Nucleotide and amino acid numbering are based on GenBank NM_005584.5 and GenPept NP_005575.1, respectively.
(DOCX)

S4 Fig. Reciprocal IP using TBL1XR1 antibody. Western blot analysis of the anti-TBL1XR1 IP using the anti-GFP antibody was unable to detect interaction with wild-type or mutant forms of MAB21L1. TBL1XR1 was detected in all the pull down used as control for pull down experiment.
(DOCX)

S5 Fig. Mutant and WT-specific protein-protein interactions from IP-MS. A. Graphs of the log-transformed quantitative mass spectroscopy results of biological triplicates of immunoprecipitates of control (GFP), tagged wild-type (WT) and mutant MAB21L1 (R51L and R51Q) from HEK293 cells which identified as single wild-type specific interactor (TBL1XR1) which is a component of the NCor complex. Two other subunits of the NCor complex (NCOR & HDAC3) are shown for comparison. B. Graphs of the three mutant specific interactors (GALNT2, LRRFIP1, MSI2/Musashi-2) and /Musashi-1, a close homolog of MSI2, which shows interaction with all forms of MAB21L1.
(DOCX)

S6 Fig. Fundus images showing the optic nerve anomaly of Mab21l1 R51L heterozygous mice. Retinal photographs of additional Mab21l1 R51L heterozygous and wild type mice, supplementing the representative, annotated images shown in Fig 4. (A) Photographs from 8

heterozygous *Mab21l1* R51L/+ mice, left and right eyes, showing enlarged, excavated anomalous optic nerve heads, often with prominent persistent fetal vasculature. (B) Photographs from 5 wild type mice from the same colony (littermate or cousin controls), showing normal optic nerve heads, illustrating the range of normal. (C, below) On each line, a cartoon of the first fundus image is shown, illustrating the optic disc (black circle) and major blood vessels (radial lines), followed by fundus photographs from one eye of 3 different mice. (i) *Mab21l1*R51L/+ heterozygous mice (images from 3 mice representative of the optic nerve phenotype taken from images above in (A), versus (ii) age-matched wild type littermate/cousin controls. (DOCX)

S7 Fig. Human mab-21 paralogs: Peptide sequence and genomic features. A. Phylogenetic tree of the 11 human mab-21 paralogs and protein alignment (B) and genomic organisation (C) of MAB21L1, MAB21L2 and mab-21. The alignment and phylogenetic tree were generated using MUSCLE <https://www.ebi.ac.uk/Tools/msa/muscle/> (DOCX)

S1 File. Supplemental clinical descriptions.
(DOCX)

S2 File. Supplemental materials and methods.
(DOCX)

S3 File. Supplementary raw data table: Pyrophosphate assay.
(XLSX)

S4 File. Supplementary raw data table: Mass spectrometry.
(XLSX)

S5 File.
(DOCX)

S6 File.
(DOCX)

S7 File. S1 Table. Oligonucleotides used in the *MAB21L1/Mab21l1* study: Sequence and protocol details. Underlined sequence denotes universal tags with no homology to *MAB21L1*. Further details of the biological relatedness microsatellite PCR protocol are available at http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/2000s/media/200614.pdf. **S2 Table.** *MAB21L1* variant nomenclature validation. (<https://variantvalidator.org/>). **S3 Table. Mendelian ratios.** Comparison of the observed versus expected ratios of genotypes from intercrosses of the *Mab21l1* R51L line mice (n = 14 litters of each type), establishing that the observed ratios were consistent with Mendelian genetics. **S4 Table. FoldX values.** Molecular modelling performed using FoldX (Delgado et al., 2019) in order to assess the impact of MAB21L1 and MAB21L2 substitutions on protein stability. Nearly all the mutations are destabilizing to protein structure. (DOCX)

S1 Raw images.
(TIF)

Acknowledgments

We thank the patients and families for their participation. For rodent expertise we thank Lorraine Rose, Anna Thornburn, John Campbell, Jacek Mendrychowski and the staff of Central Bioresearch Services.

Author Contributions

Conceptualization: Veronica Van Heyningen, David R. FitzPatrick.

Data curation: Hemant Bengani, Morad Ansari, Joseph A. Marsh.

Formal analysis: Hildegard Nikki Hall, Hemant Bengani, Joseph A. Marsh, Graeme R. Grimes.

Funding acquisition: David R. FitzPatrick.

Investigation: Hemant Bengani, Brian P. Brooks.

Methodology: Hemant Bengani, Alex von Kriegsheim, Lisa McKie, Joe Rainger.

Project administration: David R. FitzPatrick.

Resources: Robert B. Hufnagel, Giuseppe Damante, David Moore, Jamalia Rahmat, Catia Mio, Moira Blyth, Wee Teik Keng, Lily Islam, Meriel McEntargart, Marcel M. Mannens, Brian P. Brooks.

Supervision: David R. FitzPatrick.

Validation: Hemant Bengani, Joe Rainger.

Writing – original draft: David R. FitzPatrick.

Writing – review & editing: Hildegard Nikki Hall, Hemant Bengani, Joseph A. Marsh, Joe Rainger.

References

1. Chow KL, Hall DH, Emmons SW. The mab-21 gene of *Caenorhabditis elegans* encodes a novel protein required for choice of alternate cell fates. *Development*. 1995; 121:3615–3626. <https://doi.org/10.1242/dev.121.11.3615> PMID: 8582275
2. Nikolaidis N, Chalkia D, Watkins DN et al. Ancient origin of the new developmental superfamily DANGER. *PLoS One*. 2007; 2:e204. <https://doi.org/10.1371/journal.pone.0000204> PMID: 17301879
3. Motwani M, Pesiridis S, Fitzgerald KA. DNA sensing by the cGAS-STING pathway in health and disease. *Nat Rev Genet*. 2019; 20:657–674. <https://doi.org/10.1038/s41576-019-0151-1> PMID: 31358977
4. Rad A, Altunoglu U, Miller R et al. MAB21L1 loss of function causes a syndromic neurodevelopmental disorder with distinctive cerebellar, ocular, craniofacial and genital features (COFG syndrome). *J Med Genet*. 2019; 56:332–339. <https://doi.org/10.1136/jmedgenet-2018-105623> PMID: 30487245
5. Bruel AL, Masurel-Paulet A, Rivière JB et al. Autosomal recessive truncating MAB21L1 mutation associated with a syndromic scrotal agenesis. *Clin Genet*. 2016. <https://doi.org/10.1111/cge.12794> PMID: 27103078
6. Yamada R, Mizutani-Koseki Y, Hasegawa T, Osumi N, Koseki H, Takahashi N. Cell-autonomous involvement of Mab21l1 is essential for lens placode development. *Development*. 2003; 130:1759–1770. <https://doi.org/10.1242/dev.00399> PMID: 12642482
7. Nguyen D, Yamada R, Yoshimitsu N, Oguri A, Kojima T, Takahashi N. Involvement of the Mab21l1 gene in calvarial osteogenesis. *Differentiation*. 2017; 98:70–78. <https://doi.org/10.1016/j.diff.2017.11.001> PMID: 29156428
8. de Oliveira Mann CC, Kiefersauer R, Witte G, Hopfner KP. Structural and biochemical characterization of the cell fate determining nucleotidyltransferase fold protein MAB21L1. *Sci Rep*. 2016; 6:27498. <https://doi.org/10.1038/srep27498> PMID: 27271801
9. Horn D, Prescott T, Houge G et al. A Novel Oculo-Skeletal syndrome with intellectual disability caused by a particular MAB21L2 mutation. *Eur J Med Genet*. 2015; 58:387–391. <https://doi.org/10.1016/j.ejmg.2015.06.003> PMID: 26116559
10. Rainger J, Pehlivan D, Johansson S et al. Monoallelic and biallelic mutations in MAB21L2 cause a spectrum of major eye malformations. *Am J Hum Genet*. 2014; 94:915–923. <https://doi.org/10.1016/j.ajhg.2014.05.005> PMID: 24906020

11. Deml B, Kariminejad A, Borujerdi RH, Muheisen S, Reis LM, Semina EV. Mutations in MAB21L2 result in ocular Coloboma, microcornea and cataracts. *PLoS Genet.* 2015; 11:e1005002. <https://doi.org/10.1371/journal.pgen.1005002> PMID: 25719200
12. Tsang SW, Guo Y, Chan LH, Huang Y, Chow KL. Generation and characterization of pathogenic Mab21l2(R51C) mouse model. *Genesis.* 2018; 56:e23261. <https://doi.org/10.1002/dvg.23261> PMID: 30375740
13. Yamada R, Mizutani-Koseki Y, Koseki H, Takahashi N. Requirement for Mab21l2 during development of murine retina and ventral body wall. *Dev Biol.* 2004; 274:295–307. <https://doi.org/10.1016/j.ydbio.2004.07.016> PMID: 15385160
14. Seese SE, Reis LM, Deml B et al. Identification of missense MAB21L1 variants in microphthalmia and aniridia. *Hum Mutat.* 2021; 42:877–890. <https://doi.org/10.1002/humu.24218> PMID: 33973683
15. UK10K C, Walter K, Min JL et al. The UK10K project identifies rare variants in health and disease. *Nature.* 2015; 526:82–90. <https://doi.org/10.1038/nature14962> PMID: 26367797
16. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. *Nat Protoc.* 2016; 11:1–9. <https://doi.org/10.1038/nprot.2015.123> PMID: 26633127
17. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet.* 2013;Chapter 7:Unit7.20. <https://doi.org/10.1002/0471142905.hg0720s76> PMID: 23315928
18. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* 2019; 47:D886–D894. <https://doi.org/10.1093/nar/gky1016> PMID: 30371827
19. Ioannidis NM, Rothstein JH, Pejaver V et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet.* 2016; 99:877–885. <https://doi.org/10.1016/j.ajhg.2016.08.016> PMID: 27666373
20. Foreman J, Brent S, Perrett D et al. DECIPHER: Supporting the interpretation and sharing of rare disease phenotype-linked variant data to advance diagnosis and research. *Hum Mutat.* 2022 <https://doi.org/10.1002/humu.24340> PMID: 35143074
21. Delgado J, Radusky LG, Cianferoni D, Serrano L. FoldX 5.0: working with RNA, small molecules and a new graphical interface. *Bioinformatics.* 2019; 35:4168–4169. <https://doi.org/10.1093/bioinformatics/btz184> PMID: 30874800
22. Gerasimavicius L, Liu X, Marsh JA. Identification of pathogenic missense mutations using protein stability predictors. *Sci Rep.* 2020; 10:15387. <https://doi.org/10.1038/s41598-020-72404-w> PMID: 32958805
23. Meng H, Deo S, Xiong S et al. Regulation of the interferon-inducible 2'-5'-oligoadenylate synthetases by adenovirus VA(I) RNA. *J Mol Biol.* 2012; 422:635–649. <https://doi.org/10.1016/j.jmb.2012.06.017> PMID: 22709583
24. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res.* 2011; 10:1794–1805. <https://doi.org/10.1021/pr101065j> PMID: 21254760
25. Jadeja S, Barnard AR, McKie L et al. Mouse slc9a8 mutants exhibit retinal defects due to retinal pigmented epithelium dysfunction. *Invest Ophthalmol Vis Sci.* 2015; 56:3015–3026. <https://doi.org/10.1167/iovs.14-15735> PMID: 25736793
26. Perez-Rovira A, MacGillivray T, Trucco E et al. VAMPIRE: Vessel assessment and measurement platform for images of the REtina. *Annu Int Conf IEEE Eng Med Biol Soc.* 2011; 2011:3391–3394. <https://doi.org/10.1109/IEMBS.2011.6090918> PMID: 22255067
27. Cross SH, McKie L, Hurd TW et al. The nanophthalmos protein TMEM98 inhibits MYRF self-cleavage and is required for eye size specification. *PLoS Genet.* 2020; 16:e1008583. <https://doi.org/10.1371/journal.pgen.1008583> PMID: 32236127
28. Findlay AS, McKie L, Keighren M et al. Fam151b, the mouse homologue of C.elegans menorin gene, is essential for retinal function. *Sci Rep.* 2020; 10:437. <https://doi.org/10.1038/s41598-019-57398-4> PMID: 31949211
29. Gilmour DT, Lyon GJ, Carlton MB et al. Mice deficient for the secreted glycoprotein SPARC/osteonectin/BM40 develop normally but show severe age-onset cataract formation and disruption of the lens. *EMBO J.* 1998; 17:1860–1870. <https://doi.org/10.1093/emboj/17.7.1860> PMID: 9524110
30. Norose K, Clark JI, Syed NA et al. SPARC deficiency leads to early-onset cataractogenesis. *Invest Ophthalmol Vis Sci.* 1998; 39:2674–2680. PMID: 9856777
31. Shubham K, Mishra R. Pax6 interacts with SPARC and TGF- β in murine eyes. *Mol Vis.* 2012; 18:951–956. PMID: 22539874
32. Hall HN, Williamson KA, FitzPatrick DR. The genetic architecture of aniridia and Gillespie syndrome. *Hum Genet.* 2019; 138:881–898. <https://doi.org/10.1007/s00439-018-1934-8> PMID: 30242502

33. Lima Cunha D, Arno G, Corton M, Moosajee M. The Spectrum of PAX6 Mutations and Genotype-Phenotype Correlations in the Eye. *Genes (Basel)*. 2019;10
34. Hill RE, Favor J, Hogan BL et al. Mouse small eye results from mutations in a paired-like homeobox-containing gene. *Nature*. 1991; 354:522–525. <https://doi.org/10.1038/354522a0> PMID: 1684639
35. Ton CC, Hirvonen H, Miwa H et al. Positional cloning and characterization of a paired box- and homeobox-containing gene from the aniridia region. *Cell*. 1991; 67:1059–1074. [https://doi.org/10.1016/0092-8674\(91\)90284-6](https://doi.org/10.1016/0092-8674(91)90284-6) PMID: 1684738
36. Glaser T, Walton DS, Maas RL. Genomic structure, evolutionary conservation and aniridia mutations in the human PAX6 gene. *Nat Genet*. 1992; 2:232–239. <https://doi.org/10.1038/ng1192-232> PMID: 1345175
37. Williamson KA, Hall HN, Owen LJ et al. Recurrent heterozygous PAX6 missense variants cause severe bilateral microphthalmia via predictable effects on DNA-protein interaction. *Genet Med*. 2020; 22:598–609. <https://doi.org/10.1038/s41436-019-0685-9> PMID: 31700164
38. Kamachi Y, Uchikawa M, Tanouchi A, Sekido R, Kondoh H. Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development. *Genes Dev*. 2001; 15:1272–1286. <https://doi.org/10.1101/gad.887101> PMID: 11358870
39. Inoue M, Kamachi Y, Matsunami H, Imada K, Uchikawa M, Kondoh H. PAX6 and SOX2-dependent regulation of the Sox2 enhancer N-3 involved in embryonic visual system development. *Genes Cells*. 2007; 12:1049–1061. <https://doi.org/10.1111/j.1365-2443.2007.01114.x> PMID: 17825048
40. Kondoh H, Kamachi Y. SOX-partner code for cell specification: Regulatory target selection and underlying molecular mechanisms. *Int J Biochem Cell Biol*. 2010; 42:391–399. <https://doi.org/10.1016/j.biocel.2009.09.003> PMID: 19747562
41. Zhang Y, Emmons SW. Specification of sense-organ identity by a *Caenorhabditis elegans* Pax-6 homologue. *Nature*. 1995; 377:55–59. <https://doi.org/10.1038/377055a0> PMID: 7659160
42. Zhang X, Friedman A, Heaney S, Purcell P, Maas RL. Meis homeoproteins directly regulate Pax6 during vertebrate lens morphogenesis. *Genes Dev*. 2002; 16:2097–2107. <https://doi.org/10.1101/gad.1007602> PMID: 12183364
43. Zhang X, Rowan S, Yue Y et al. Pax6 is regulated by Meis and Pbx homeoproteins during pancreatic development. *Dev Biol*. 2006; 300:748–757. <https://doi.org/10.1016/j.ydbio.2006.06.030> PMID: 17049510
44. Carbe C, Hertzler-Schaefer K, Zhang X. The functional role of the Meis/Prep-binding elements in Pax6 locus during pancreas and eye development. *Dev Biol*. 2012; 363:320–329. <https://doi.org/10.1016/j.ydbio.2011.12.038> PMID: 22240097
45. Antosova B, Smolikova J, Klimova L et al. The Gene Regulatory Network of Lens Induction Is Wired through Meis-Dependent Shadow Enhancers of Pax6. *PLoS Genet*. 2016; 12:e1006441. <https://doi.org/10.1371/journal.pgen.1006441> PMID: 27918583
46. Grebbin BM, Hau AC, Groß A et al. Pbx1 is required for adult subventricular zone neurogenesis. *Development*. 2016; 143:2281–2291. <https://doi.org/10.1242/dev.128033> PMID: 27226325
47. Agoston Z, Heine P, Brill MS et al. Meis2 is a Pax6 co-factor in neurogenesis and dopaminergic periglomerular fate specification in the adult olfactory bulb. *Development*. 2014; 141:28–38. <https://doi.org/10.1242/dev.097295> PMID: 24284204
48. Perissi V, Aggarwal A, Glass CK, Rose DW, Rosenfeld MG. A corepressor/coactivator exchange complex required for transcriptional activation by nuclear receptors and other regulated transcription factors. *Cell*. 2004; 116:511–526. [https://doi.org/10.1016/s0092-8674\(04\)00133-3](https://doi.org/10.1016/s0092-8674(04)00133-3) PMID: 14980219
49. Zilmer M, Edmondson AC, Khetarpal SA et al. Novel congenital disorder of O-linked glycosylation caused by GALNT2 loss of function. *Brain*. 2020; 143:1114–1126. <https://doi.org/10.1093/brain/awaa063> PMID: 32293671
50. Takimoto M. Multidisciplinary Roles of LRRFIP1/GCF2 in Human Biological Systems and Diseases. *Cells*. 2019;8. <https://doi.org/10.3390/cells8020108> PMID: 30709060
51. Yang P, An H, Liu X et al. The cytosolic nucleic acid sensor LRRFIP1 mediates the production of type I interferon via a beta-catenin-dependent pathway. *Nat Immunol*. 2010; 11:487–494. <https://doi.org/10.1038/ni.1876> PMID: 20453844
52. Douchi D, Ohtsuka H, Ariake K et al. Silencing of LRRFIP1 reverses the epithelial-mesenchymal transition via inhibition of the Wnt/β-catenin signaling pathway. *Cancer Lett*. 2015; 365:132–140. <https://doi.org/10.1016/j.canlet.2015.05.023> PMID: 26047573
53. Kudinov AE, Karanickolas J, Golemis EA, Bumber Y. Musashi RNA-Binding Proteins as Cancer Drivers and Novel Therapeutic Targets. *Clin Cancer Res*. 2017; 23:2143–2153. <https://doi.org/10.1158/1078-0432.CCR-16-2728> PMID: 28143872

54. Sutherland JM, McLaughlin EA, Hime GR, Siddall NA. The Musashi family of RNA binding proteins: master regulators of multiple stem cell populations. *Adv Exp Med Biol.* 2013; 786:233–245. https://doi.org/10.1007/978-94-007-6621-1_13 PMID: 23696360
55. Park SM, Deering RP, Lu Y et al. Musashi-2 controls cell fate, lineage bias, and TGF- β signaling in HSCs. *J Exp Med.* 2014; 211:71–87. <https://doi.org/10.1084/jem.20130736> PMID: 24395885
56. Bennett CG, Riemondy K, Chapnick DA et al. Genome-wide analysis of Musashi-2 targets reveals novel functions in governing epithelial cell migration. *Nucleic Acids Res.* 2016; 44:3788–3800. <https://doi.org/10.1093/nar/gkw207> PMID: 27034466
57. Cox JL, Wilder PJ, Gilmore JM, Wuebben EL, Washburn MP, Rizzino A. The SOX2-interactome in brain cancer cells identifies the requirement of MSI2 and USP9X for the growth of brain tumor cells. *PLoS One.* 2013; 8:e62857. <https://doi.org/10.1371/journal.pone.0062857> PMID: 23667531
58. Sundar J, Matakah F, Jeong B, Stoilov P, Ramamurthy V. The Musashi proteins MSI1 and MSI2 are required for photoreceptor morphogenesis and vision in mice. *J Biol Chem.* 2020; 296:100048. <https://doi.org/10.1074/jbc.RA120.015714> PMID: 33168629
59. Yokoi T, Nishina S, Fukami M et al. Genotype-phenotype correlation of PAX6 gene mutations in aniridia. *Hum Genome Var.* 2016; 3:15052. <https://doi.org/10.1038/hgv.2015.52> PMID: 27081561

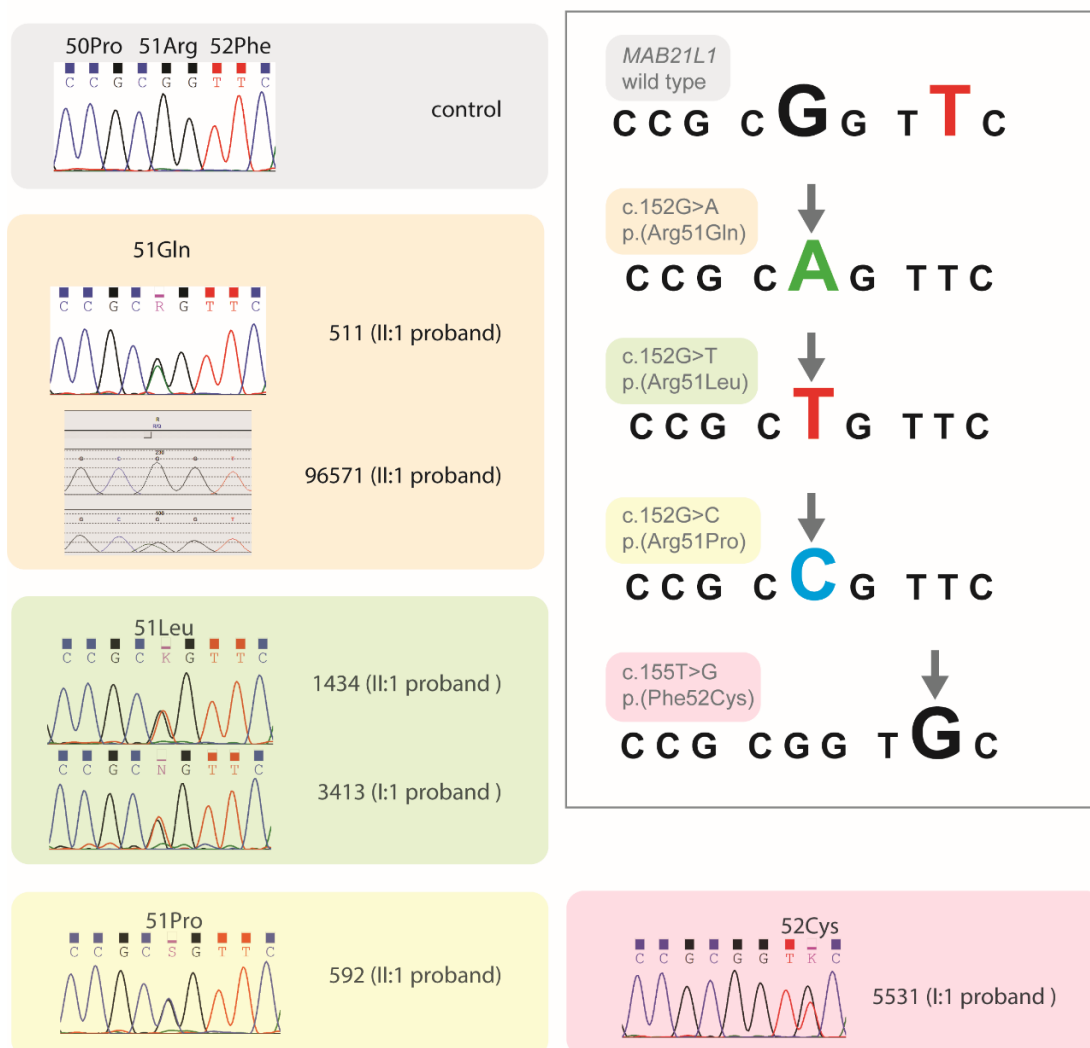
B.2 Supplemental materials and methods for *MAB21L1* project

B.2.1 Table of *MAB21L1/Mab21l1* primers and other oligonucleotides

Name	Sequence 5' to 3'	Application
MAB21L1_a1_F	<u>GTAGCGCGACGGCCAGTGAAGTTG</u> GCTCCAGCTCTAG	NM_005584 <i>MAB21L1</i> exon1 amplicon1 PCR from human total genomic DNA
MAB21L1_a1_R	<u>CAGGGCGCAGCGATGACTATCTCGG</u> ATTCTCAGTTTCAC	
MAB21L1_a2_F	<u>GTAGCGCGACGGCCAGTGCTACCTC</u> TCGGCGCGCA	NM_005584 <i>MAB21L1</i> exon1 amplicon2 PCR from human total genomic DNA
MAB21L1_a2_R	<u>CAGGGCGCAGCGATGACTTCACACT</u> CGTAGGAAACCAG	
MAB21L1_a3_F	<u>GTAGCGCGACGGCCAGTGAAAGAA</u> GTGCCTCTCCATCC	NM_005584 <i>MAB21L1</i> exon1 amplicon3 PCR from human total genomic DNA
MAB21L1_a3_Rb	<u>CAGGGCGCAGCGATGACTTGTTCG</u> ACTGCGGAGTGG	
<i>Mus_Mab21l1_F</i>	CAA GCT GGT CTA CCA CCT GAA	<i>Mab21l1</i> exon 1 PCR from mouse genomic DNA
<i>Mus_Mab21l1_R</i>	TAA ACA CCC CCA TCT GGT TC	
<i>Mab21l1_guide_C</i>	CCT GCA CTT CAA CCT CCT TC	Guide DNA sequences targeting the coding exon of <i>Mab21l1</i> used in CRISPR-Cas9 genome editing
<i>Mab21l1_guide_D</i>	GAG ATG GAC AAC CGC TAC GA	
<i>Mus_Mab21l1_F</i>	CAA GCT GGT CTA CCA CCT GAA	<i>Mab21l1</i> exon 1 PCR from mouse genomic DNA
<i>Mus_Mab21l1_R</i>	TAA ACA CCC CCA TCT GGT TC	

Underlined sequence = universal primer tags (no homology to MAB21L1)

B.3 Chromatograms



Supplementary Figure 8 Chromatogram data from Sanger sequencing showing the heterozygous *MAB21L1* variants identified in probands with aniridia and microphthalmia (case and pedigree identifiers on the right), and schematic illustrating the positions of the variants, affecting NM_005584.5:c.152 and NM_005584.5:c.155. Figure from our publication (166) and with acknowledgement of the help of Kathy Williamson.

B.4 Detailed clinical descriptions

Family 96571: c.152G>A, p.(Arg51Gln) paternal mosaicism

Two male siblings, aged 7 and 2, with non-syndromic microphthalmia/anophthalmia. The elder sibling has very severe right microphthalmia, and severe left microphthalmia with an opaque cornea, precluding a view of the internal ocular structures. Visual acuity is perception of light in the left eye only; no perception of light in the right. The younger sibling has bilateral microphthalmia and nystagmus. The right eye has an opaque cornea, again precluding visualization of internal ocular structures. The left eye has foveal hypoplasia and an anomalous optic nerve. There are no signs of cataract or glaucoma in this eye. Visual acuity is light perception in both eyes. Brain MRI, renal ultrasound and echocardiogram were normal in the older sibling, and not requested in the younger. The parents were unrelated, unaffected and had normal eyes. Testing of genomic DNA samples from both parents identified gonosomal mosaicism (at a level of ~27%) in the father.

Family 511: c.152G>A, p.(Arg51Gln) familial

A 48 year old male diagnosed with bilateral aniridia soon after birth. Both eyes are significantly microphthalmic, with a horizontal corneal diameter of 6mm in the left eye. The right eye is phthisical for unknown reasons with no view of the fundus. The left eye has progressive pannus, corneal oedema and band keratopathy. There is no discernable iris tissue, nystagmus and left post-surgical aphakia (cataract surgery aged 25). Intraocular pressure 18mmHg. View to the left posterior pole is hazy with no clear view of the fovea, but the disc appears anomalous and excavated, with a small uveal coloboma below it.

Both of his children (13 year old son and 11 year old daughter) are affected. Both children have profound bilateral aniridia, microcornea (6mm), nystagmus and gray, hypoplastic optic discs. The son has a right inferotemporal choroidal coloboma, and a small left coloboma temporal to the disc. The daughter has microphthalmia and a small left nasal choroidal coloboma; she has bilateral cataract with lenticonus in the right, and an unstable lens in the left.

His deceased parents both had ocular phenotypes: the father was similarly affected with microphthalmia and aniridia, and the mother had oculocutaneous albinism. DNA samples were not available for testing from either individual.

Individual 1434: c.152G>T, p.(Arg51Leu) *de novo*

A 23 year old woman with bilateral partial aniridia, microphthalmia and glaucoma (treated surgically in the left eye). She has bilateral cataract with progressive subluxation, in addition to nystagmus and foveal hypoplasia. She is a high myope with hand movement vision in both eyes. The only other feature of note is bilateral sensorineural hearing loss; a brain MRI scan was normal. The unaffected parents did not carry the variant.

Individual 3413: c.152G>T, p.(Arg51Leu)

An individual with bilateral aniridia. The family was lost to follow up and consequently limited phenotypic data is available.

Individual 592: c.152G>C, p.(Arg51Pro) *de novo*

An 11 year old female child with a mild aniridia-spectrum disorder and high myopia. She was born by emergency Caesarean section for fetal distress, but was subsequently well with no developmental concerns. Nystagmus, high myopia (-14 dioptries) and foveal hypoplasia were diagnosed in infancy. Electrophysiology (electroretinogram and visual evoked potentials) showed some features consistent with albinism. There was no iris transillumination but the pupillary border appeared irregular and she was described as having variant aniridia. She developed cataracts later in childhood. Her parents are unaffected and unrelated, and did not carry the variant.

Family 5531: c.155T>G, p.(Phe52Cys)

A male child with sporadic, bilateral partial aniridia and microphthalmia. The family was lost to follow-up and consequently limited phenotypic data was available.

B.5 Variant effect predictions for MAB21L1 variants

Supplementary Table 1 Ten variant effect predictions for the MAB21L1 missense variants discussed here. For consistency with **Table 5** the order and colours indicate the year of preprint/publication of the VEP: AlphaMissense, CPT and SNPred 2023, green (221,222); EVE, MetaRNN 2021, turquoise (226,227); BayesDel 2017, blue (228); REVEL, SIFT4G 2016, violet (229,230); CADD 2014, pink (231); Polyphen2 2010, red (232). However, in the case of these MAB21L1 variants there is very little disagreement, especially for the monoallelic variants (the first 4 rows) which are all predicted to be likely pathogenic.

variant	Alpha-Missense	CPT	SNPred	EVE	Meta-RNN	Bayes-Del	REVEL	SIFT4G	CADD	Polyphen2 / \ _HumDiv _HumVar	
R51L	0.999	0.268	0.987	0.774	0.969	0.269	0.682	0.000	34.0	1.000	0.999
R51P	0.999	0.494	0.993	0.890	0.972	0.272	0.694	0.000	34.0	1.000	1.000
R51Q	0.999	0.435	0.978	0.907	0.956	0.208	0.542	0.000	34.0	1.000	0.999
F52C	0.974	0.421	0.994	0.586	0.979	0.297	0.745	0.000	33.0	0.999	0.969
Q233P	0.937	0.139	0.898	0.166	0.643	0.176	0.332	0.054	25.7	0.040	0.090
thres-hold	≥0.560					>0.069	>0.500	≤0.050	>20.0	≥0.447	

B.6 Supplemental material: *Mab21/1* mouse work

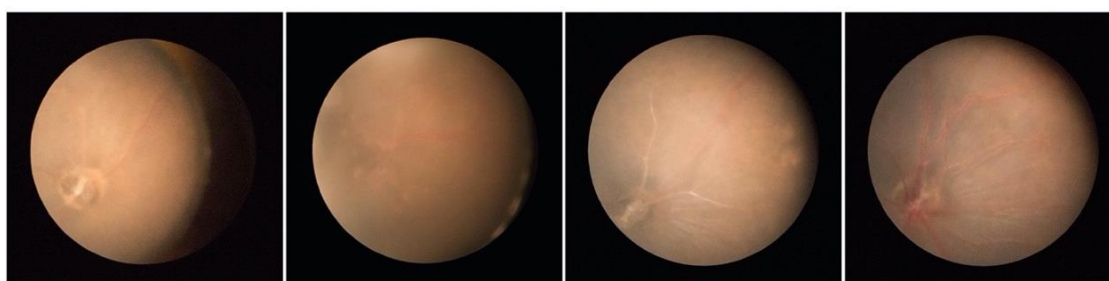
B.6.1 Fundus photographs of heterozygous *Mab21/1*^{R51L/+} and wild type mice

A.



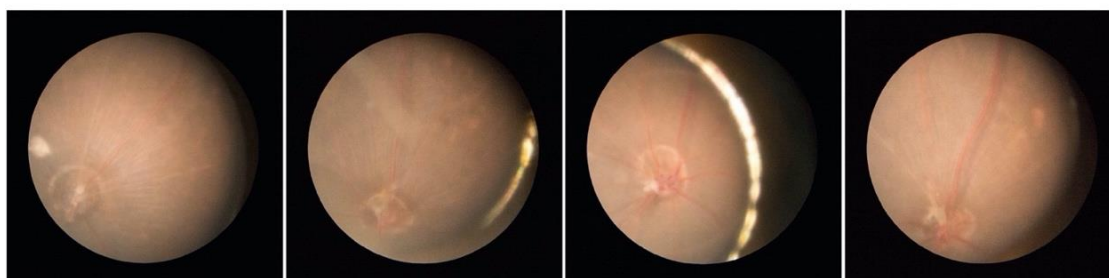
Left eye - Mouse #49 - Right eye

Left eye - Mouse #59 - Right eye



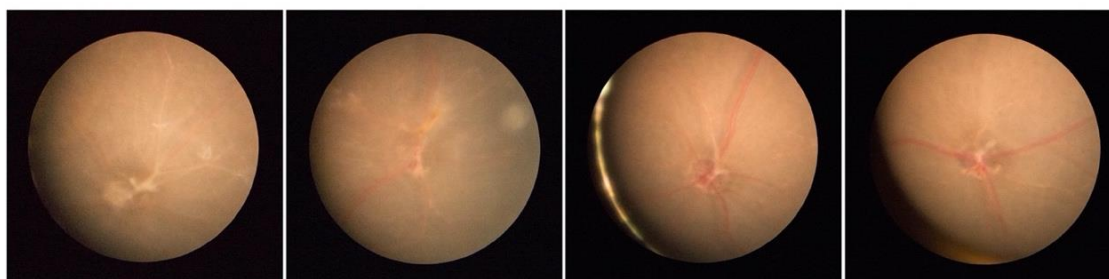
Left eye - Mouse #61 - Right eye

Left eye - Mouse #78 - Right eye



Left eye - Mouse #79 - Right eye

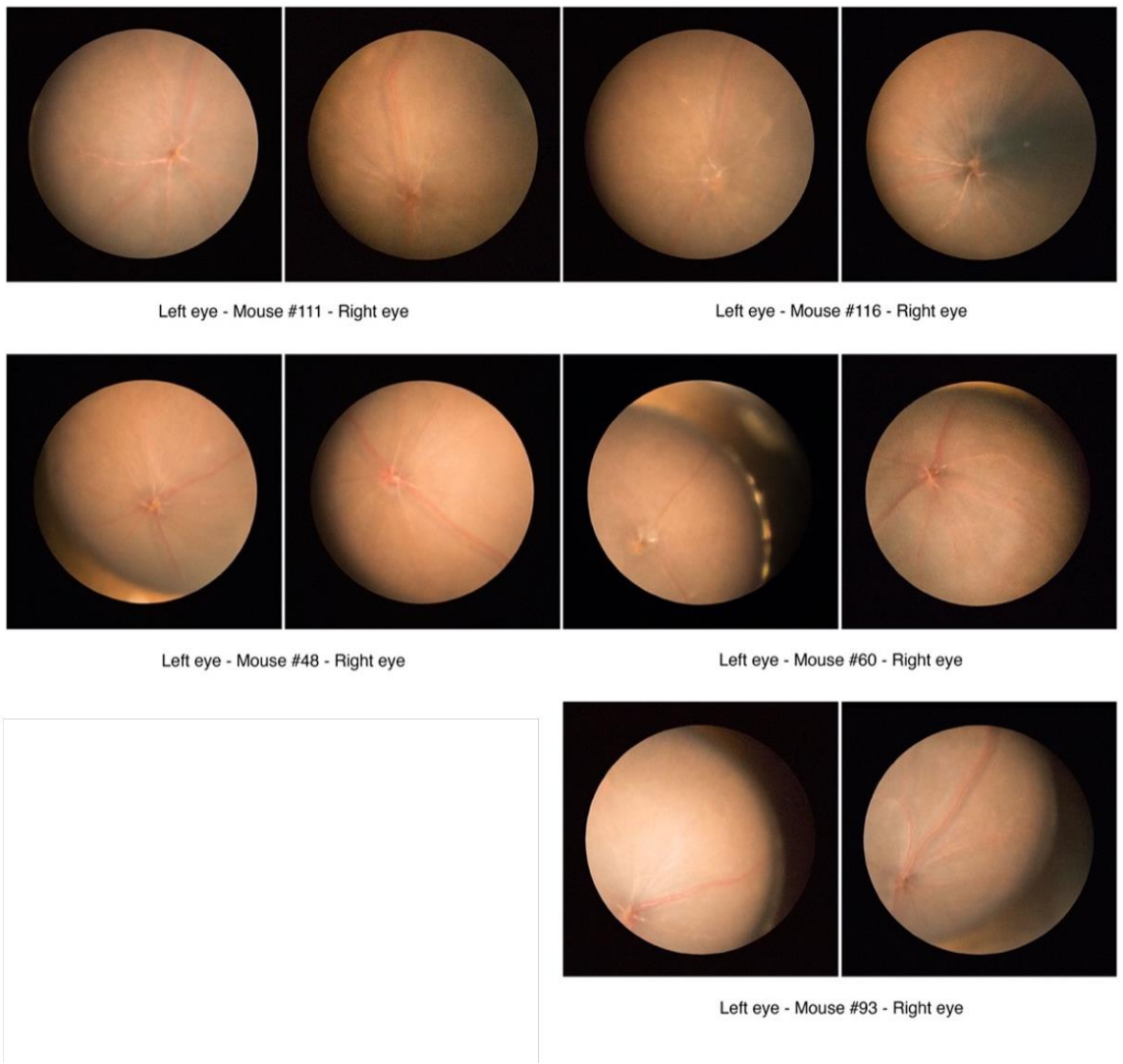
Left eye - Mouse #80 - Right eye



Left eye - Mouse #112 - Right eye

Left eye - Mouse #117 - Right eye

B.



Supplementary Figure 9 Fundus images showing the optic nerve anomaly of *Mab21l1* R51L heterozygous mice. Retinal photographs of additional *Mab21l1* R51L heterozygotes and wild type mice, supplementing the representative, annotated images shown in **Figure 28**. (A) Photographs from 8 heterozygous *Mab21l1* R51L mice, left and right eyes, showing enlarged, excavated anomalous optic nerve heads, often with prominent persistent foetal vasculature. (B) Photographs from 5 wild type mice from the same colony (littermate or cousin controls), showing normal optic nerve heads.

Appendix C Supplemental material for WGS chapter

- C.1 Publication 'Short-read whole genome sequencing identifies causative variants in most individuals with previously unexplained aniridia'



OPEN ACCESS

Original research

Short-read whole genome sequencing identifies causative variants in most individuals with previously unexplained aniridia

Hildegard Nikki Hall ,¹ David Parry,^{1,2} Mihail Halachev,¹ Kathleen A Williamson,¹ Kevin Donnelly,¹ Jose Campos Parada,¹ Shipra Bhatia,¹ Jeffrey Joseph,³ Simon Holden,⁴ Trine E Prescott,⁵ Pierre Bitoun,⁶ Edwin P Kirk,⁷ Ruth Newbury-Ecob,⁸ Katherine Lachlan,⁹ Juan Bernar,¹⁰ Veronica van Heyningen,^{3,11} David R FitzPatrick,¹ Alison Meynert¹

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/jmg-2023-109181>).

For numbered affiliations see end of article.

Correspondence to
Dr Hildegard Nikki Hall;
nikki.hall@ed.ac.uk

Received 12 March 2023
Accepted 25 September 2023

ABSTRACT

Background Classic aniridia is a highly penetrant autosomal dominant disorder characterised by congenital absence of the iris, foveal hypoplasia, optic disc anomalies and progressive opacification of the cornea. >90% of cases of classic aniridia are caused by heterozygous, loss-of-function variants affecting the *PAX6* locus.

Methods Short-read whole genome sequencing was performed on 51 (39 affected) individuals from 37 different families who had screened negative for mutations in the *PAX6* coding region.

Results Likely causative mutations were identified in 22 out of 37 (59%) families. In 19 out of 22 families, the causative genomic changes have an interpretable deleterious impact on the *PAX6* locus. Of these 19 families, 1 has a novel heterozygous *PAX6* frameshift variant missed on previous screens, 4 have single nucleotide variants (SNVs) (one novel) affecting essential splice sites of *PAX6* 5' non-coding exons and 2 have deep intronic SNV (one novel) resulting in gain of a donor splice site. In 12 out of 19, the causative variants are large-scale structural variants; 5 have partial or whole gene deletions of *PAX6*, 3 have deletions encompassing critical *PAX6* *cis*-regulatory elements, 2 have balanced inversions with disruptive breakpoints within the *PAX6* locus and 2 have complex rearrangements disrupting *PAX6*. The remaining 3 of 22 families have deletions encompassing *FOXC1* (a known cause of atypical aniridia). Seven of the causative variants occurred *de novo* and one cosegregated with familial aniridia. We were unable to establish inheritance status in the remaining probands. No plausibly causative SNVs were identified in *PAX6* *cis*-regulatory elements.

Conclusion Whole genome sequencing proves to be an effective diagnostic test in most individuals with previously unexplained aniridia.

INTRODUCTION

Historically the molecular genetic investigation of Mendelian disorders has focused on sequencing of the coding regions of causative genes often in combination with genomic copy number analysis. Depending on the phenotypic specificity of the disease under investigation, these tests could

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Aniridia is a strikingly specific phenotype, carrying a 90%–95% positive predictive value for *PAX6* haploinsufficiency.

WHAT THIS STUDY ADDS

⇒ This is the first dedicated whole genome sequencing (WGS) study of a classical aniridia cohort, highlighting the diagnostic power of WGS even in well-characterised Mendelian disorders.
⇒ It offers particular advantages in detecting deep intronic variants, *cis*-regulatory, and balanced or complex structural variants.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ This study suggests that short-read WGS merits consideration as a primary investigation for classic aniridia.

involve sequencing of a single gene, a panel of genes or whole exome analysis. One of the motivations for restricting diagnostic analysis to the coding regions of genes has been the availability of well-characterised and validated approaches to predict the consequence of each variant and to assign a confidence to its pathogenicity.^{1,2}

The wider adoption of whole genome sequencing (WGS) as a diagnostic tool,³ together with the guidelines that aim to standardise the interpretation of variants outside the coding regions of genes, provides an opportunity to increase the utility of diagnostic genetic analyses.⁴ Here we have used short-read WGS to try to identify causative variants in individuals with aniridia in whom no diagnosis was found by prior molecular genetic testing approaches. Classic aniridia has major advantages in such a study as a Mendelian disease in which the phenotype in early childhood (congenital absence of the iris with foveal hypoplasia) has a high (~0.9) positive predictive value for detecting a heterozygous loss-of-function mutation at a single locus (*PAX6*).⁵



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY. Published by BMJ.

To cite: Hall HN, Parry D, Halachev M, et al. *J Med Genet* Epub ahead of print: [please include Day Month Year]. doi:10.1136/jmg-2023-109181

PAX6 encodes a dosage critical transcription factor that is essential for vertebrate eye and brain development,⁶ and many *cis*-regulatory elements (CRE) controlling its expression during development have been functionally characterised.^{7,8} Diagnostic analysis of this locus has been in routine clinical use now for 30 years, providing a very large data set of causative variants and established disease mechanisms.⁵

As this study will show, WGS is able, with reasonable sensitivity, to identify causative variants in individuals with a clinical diagnosis of aniridia who have previously, often repeatedly, tested negative for mutations in the coding region of *PAX6*.

METHODS

Clinical research participants

This project used clinical information and biological samples from individuals referred to the Medical Research Council (MRC) Human Genetics Unit Eye Malformations Study. All affected individuals had classical aniridia; those with significant additional ocular phenotypes were excluded, such as severe microphthalmia or severe congenital corneal opacification. Pseudonymised research participant identifiers (RPIDs), relevant clinical features and molecular analyses performed prior to this study are provided in [table 1](#).

Preparation of genomic DNA and quality control

The quality and concentration of patient genomic DNA (gDNA) samples were assessed by agarose gel electrophoresis, NanoDrop 1000 spectrophotometry (Thermo Fisher Scientific, Inchinnan, UK), and/or Qubit 3 fluorometer high sensitivity (HS) assay (Invitrogen, Thermo Fisher Scientific). In the case of one family trio, the DNA for the parents was extracted from patient-derived lymphoblastoid cell lines (LCLs). All probands had had prior Sanger sequencing of the coding regions of *PAX6* and *MAB21L1*.

Whole genome sequencing

WGS was performed by BGI (New Territories, Hong Kong) for 9 samples and Edinburgh Genomics (Edinburgh, UK) for the remaining 42 samples.

Detailed methods for massively parallel sequencing library preparation are found in online supplemental information. In brief, gDNA was sheared using a Covaris ultrasonicator, and the fragments were A-tailed, size selected and adaptor ligated prior to PCR amplification. Libraries were clustered onto a flow cell for sequencing using HiSeqX (Illumina).

WGS mapping, alignment, quality control and single nucleotide variant/INDEL variant calling

WGS samples were processed with Bcbio V.0.9.7, which uses BWA V.0.7.13⁹ to align reads to the human reference genome assembly hg38, samblaster V.0.1.22¹⁰ to mark duplicates, and GATK V.3.4.0¹¹ to realign small insertions and deletions (INDELs) and recalibrate base quality scores. Families and singletons were genotyped following GATK best practices¹² using V.4.0.2.1 of the toolkit. HaplotypeCaller was used to generate GVCFs, which were imported into a database via GenomicsDBImport and genotyped with GenotypeGVCFs. Variant quality score recalibration was carried out with GATK VariantRecalibrator and ApplyVQSR separately for single nucleotide variants (SNVs) and INDELs. Low-quality (GQ <20) genotypes were filtered.

Structural variant calling

IGV visualisation

Direct inspection of the aligned WGS data was performed to detect structural variant (SV) at known aniridia loci, using IGV

(Integrative Genomics Viewer, Broad Institute, Massachusetts, USA),¹³ via visualisation of breakpoints and coverage. The gene and regulatory regions of *PAX6* were examined in all cases. Where *PAX6* was negative, *FOXC1* and *PITX2* were included with their regulatory regions, and *MAB21L1*; in some cases, this was expanded to additional loci (*FOXE3*, *RARB*, *ADAMTSL1*, *CYP11B1*); for a list of coordinates, see online supplemental table S1. In IGV, reads were coloured both by insert size (to detect breakpoints of deletions/insertions) and by pair orientation (to detect breakpoints of chromosomal rearrangements such as inversions).

Bioinformatic SV calling

CNVs were called for each family with Canvas V.1.38¹⁴ using the 'SmallPedigree-WGS' workflow.¹⁵ SVs were called with Manta V.1.3.2.^{16,17} CNV/SV overlapping all genes in the EyeG2P data set (described in the Variant filtering section) were examined.

De novo analysis

For parent-child trios, short variants arising *de novo* in the child were identified using VASE¹⁸ with the following criteria: read depth ≥ 10 in parents and the child, genotype quality score ≥ 30 in the child and ≥ 20 in both parents, variant allele frequency ≥ 0.3 in the child and < 0.05 in both parents, with a maximum of one variant allele called at site and variant allele absent from gnomAD V.3.0. Potential *de novo* variants were subsequently filtered to exclude low complexity, telomeric and centromeric regions, and to select only variants within coding regions, splice regions (exonic positions within 3 bp of an intron/exon junction or intronic positions within 8 bp of an intron/exon junction) or intronic variants with a SpliceAI¹⁹ delta score ≥ 0.5 .

Variant filtering

From all the variants identified in an individual, we selected only those that are rare, predicted to be functional and potentially relevant to eye disorders by using the G2P plugin^{20,21} in VEP (V.90.1 (16)) and the Eye Gene Panel (<https://www.ebi.ac.uk/gene2phenotype/downloads>; accessed 29 August 2018). In short, we extracted only variants satisfying the inheritance requirements of the genes in the Eye Gene Panel, with minor allele frequency (MAF) in public databases < 0.0001 for monoallelic and X-linked genes and MAF < 0.005 for biallelic genes and annotated by VEP to have one of the following consequences: stop gained, stop lost, start lost, frameshift variant, inframe insertion/deletion, missense variant, coding sequence variant, initiator codon variant, transcript ablation, transcript amplification, protein altering variant, splice donor/acceptor variant (ie, canonical splice site) or splice region variant (ie, either within 1–3 bases of the exon or 3–8 bases of the intron).

Detection of intronic splice variants

Variants within the *PAX6* locus (chr11:31784779-31817961; GRCh38) were annotated with SpliceAI delta scores using SpliceAI V.1.3¹⁹ using transcript coordinates from the 'GENCODE basic' transcript set from Ensembl V.95.

Cis-regulatory variant analysis

The GRCh38 coordinates of 35 different CREs chosen for analysis are given in online supplemental table S4. A BED file was created from this table and BEDTools²² was then used to extract the *PAX6* CRE variants from the cohort VCF file. Subsequent filtering used the gnomAD²³ allele frequency data in the VCF file.

Table 1 Clinical features and prior molecular analyses of the WGS cohort

Individual (RPID)	Family (FID)	Inheritance	PAX6 screen method	11p13 del analysis	Other genes screened*	ddPCR CNV PAX6	Phenotype
Singletons							
182	182	Sporadic	Direct	aCGH	No	Normal	Partial aniridia, atrophic iris, congenital cataracts, nystagmus, corneal endothelial degeneration.
224	224	Unknown	Direct	aCGH	No	Normal	Aniridia.
535	535	Sporadic	Direct	aCGH	No	No	Bilateral aniridia.
660	660	Unknown	Direct	aCGH	No	Normal	Aniridia, possible foveal hypoplasia.
724	723	Familial	DHPLC	Uncertain	No	No	Aniridia, type 2 diabetes.
774	774	Unknown	DHPLC	No	No	Normal	Bilateral aniridia, cataract, drusen at the macula and nasal to the optic disc.
877	877	Unknown	DHPLC	FISH	No	Normal	Bilateral aniridia, glaucoma, cataracts in young adulthood.
999	999	Unknown	DHPLC	No	No	Normal	Partial aniridia.
1019	1019	Sporadic	DHPLC	FISH	No	No	Aniridia.
1142	1142	Sporadic	DHPLC	Uncertain	No	No	Aniridia.
1191	1190	Familial	DHPLC	Uncertain	No	No	Partial aniridia, cataracts with surgery in late 20s.
1271	1271	Sporadic	DHPLC	aCGH	No	No	Bilateral aniridia, cataract extraction both eyes as teenager.
1304	1304	Familial	DHPLC	Uncertain	No	No	Aniridia.
1358	1358	Familial	DHPLC	Karyotype	No	No	Aniridia, congenital glaucoma, spherophakia, high myopia from birth, consanguinity, affected sibling, likely recessive or gonadal mosaic.
1361	1361	Familial	DHPLC	FISH	No	No	Aniridia.
1451	1451	Unknown	DHPLC	FISH	No	No	Aniridia, congenital glaucoma, bicuspid aortic valve with mild aortic stenosis.
1468	1468	Unknown	DHPLC	FISH	No	No	Bilateral aniridia.
1496	1496	Unknown	DHPLC	FISH	No	No	Aniridia.
1500	1500	Sporadic	DHPLC	aCGH	No	No	Bilateral aniridia, cataracts (posterior subcapsular), absent foveal reflexes, no nystagmus, vision right 6/12 N5 and left 6/36 N5, corneas clear with fine limbal vessels peripherally.
1524	1524	Familial	DHPLC	Unk	CHX10	No	Aniridia spectrum: iris and foveal hypoplasia, congenital corneal opacification with small lenses apposed to posterior surface of cornea.
1607	1607	Sporadic	DHPLC	FISH	No	No	Aniridia, congenital glaucoma.
1647	1646	Unknown	DHPLC	Unk	No	No	Aniridia with glaucoma.
1648	1648	Unknown	Direct	aCGH, FISH	No	No	Aniridia.
1732	1732	Unknown	DHPLC	Unk	No	No	Bilateral aniridia, glaucoma, opaque right cornea, aortic stenosis requiring neonatal surgery; not dysmorphic.

Continued

Developmental defects

Table 1 Continued

Individual (RPID)	Family (FID)	Inheritance	PAX6 screen method	11p13 del analysis	Other genes screened*	ddPCR CNV PAX6	Phenotype
1879	1878	Unknown	DHPLC	FISH	No	No	Aniridia spectrum anterior chamber abnormality; multiple operations in early childhood for bilateral glaucoma.
1943	1943	Familial	DHPLC	'11 p normal'	No	No	Bilateral aniridia.
2197	2197	Unknown	Unk; rescreened direct	'chromosomes normal'	No	No	Bilateral aniridia, diaphragmatic eventration, undescended testicle and hydrocele; small kidneys.
3612	3612	Sporadic	Direct	aCGH, FISH	No	Normal	Bilateral aniridia, relatively well-preserved foveas, vision 0.4 logMAR† mild keratopathy as teenager, mild cataracts.
4340	4340	Sporadic	Direct	Unk	SOX2, OTX2	Normal	Bilateral aniridia, reasonably good vision.
Affected relative pairs							
1329	1326	Familial	DHPLC	Unk	No	Normal	Aniridia, nodular corneal dystrophy; relatives have classic aniridia.
1328			Not tested				Aniridia, nephew of 1326_1329.
2466	2464	Familial	DHPLC	Unk	No	No	Aniridia, half-sibling both affected.
2464			Not tested				Aniridia, half-sibling of 2464_2466.
Trio (affected child, unaffected parents)							
75	75	Sporadic	Direct	FISH, aCGH	No	Normal	Aniridia.
356	356	Sporadic	DHPLC	FISH	No	Normal	Bilateral aniridia, congenital cataracts, <i>de novo</i> 46,XX t(1;9)(p36.1;q22) found at amniocentesis.
1635	1635	Sporadic	DHPLC, plus direct for PD	aCGH	FOXC1, SOX2, OTX2	No	Variant aniridia: inferior iris defects, peripheral corneal vascularisation, inferior lens opacities, foveal hypoplasia, nystagmus, poor vision.
2134	2134	Sporadic	DHPLC	Likely FISH	No	No	Bilateral aniridia, nystagmus, glaucoma.
2469	2469	Sporadic	DHPLC	Unk	No	Normal	Aniridia, DNA from unaffected parents derived from lymphoblastoid cell lines.
5645	5645	Sporadic	Direct (by referring centre)	MLPA, microarray	FOXC1, PITX2, ITPR1	No	Bilateral partial aniridia.

†Direct means direct Sanger sequencing.

*Other genes screened: results of any other relevant (non-PAX6) genes sequenced prior to inclusion in this study. Note MAB21L1 was screened in all patients as part of the study.

†LogMAR is the logarithm of the minimum angle of resolution.

aCGH, array comparative genomic hybridisation; ddPCR, droplet digital PCR (using four probes spanning PAX6/SIMO); del, deletion; DHPLC, denaturing high-performance liquid chromatography; FID, family identifier; MLPA, multiplex ligation-dependent probe amplification; PD, paired domain; RPID, research participant identifier; unk, unknown; WGS, whole genome sequencing.

Variant nomenclature (Human Genome Variation Society (HGVS)) was checked using Alamut software (Sophia Genetics) and VariantValidator.²⁴ Variant numbering is according reference sequences NC_000011.10 (GRCh38), NM_000280.4 (-5a, 13 exons) and NP_000271.1 (422 amino acids).

Experimental analysis of splice variants

RT-PCR and nested PCR of LCL-derived RNA

Patient-derived LCLs were recovered from liquid nitrogen storage. The cells were grown in suspension in Roswell Park Memorial Institute (RPMI) media containing 15% fetal calf serum (FCS) and penicillin/streptomycin, and incubated at 37°C/5% CO₂. RNA was extracted from LCLs following the principles of the phenol-chloroform method²⁵ using TRIzol reagent (Invitrogen) following the manufacturer's instructions.

DNase treatment was performed using TURBO DNase kit (Invitrogen). cDNA was obtained from total RNA using the SuperScript First-Strand Synthesis System for RT-PCR kit (Invitrogen). A nested PCR across the *PAX6* locus was performed using four overlapping RT-PCR primer pairs spanning exons 1–5, exons 3–8, exons 7–12 and exons 9–13 (online supplemental table S2). The products were run on a 2% low melting point agarose gel with ethidium bromide. Bands of interest were excised and a gel DNA extraction was performed (Zymoclean Gel DNA Recovery Kit, Zymo Research, Freiburg im Breisgau) and sent for Sanger sequencing to look for mis-splicing.

RESULTS

Assembling the cohort

DNA samples were available for 443 individuals with aniridia from 347 families recruited to the MRC Human Genetics Unit Eye Malformation Study. Forty-five families considered to be 'PAX6-negative' on previous screening were considered for inclusion in the WGS analysis. Of the 45 families, 3 were excluded on the basis of quality or quantity of the stored DNA and 4 were excluded following identification of various gene-disruptive variants via amplicon-based resequencing of all *PAX6* coding exons in all probands. Another family, RPID 1201, was excluded when a deletion of a critical *cis*-regulatory region 3' of the *PAX6* gene (online supplemental figure S1) was found using droplet

digital PCR across the *PAX6* locus following prescreening of 13 randomly chosen unrelated probands for CNVs.

DNA samples from a final cohort of 39 affected individuals from 37 families together with 12 unaffected relatives were sent for WGS. The family structures comprised 29 singleton probands, 6 trios (proband plus both unaffected parents) and 2 affected relative pairs (online supplemental figure S2). The proband phenotypes and molecular analyses of *PAX6* locus performed prior to this study (including the results from the referring centre) are detailed in table 1.

The following analyses of WGS data to identify sequence variants and SVs were performed in parallel.

Identification of sequence variants in the *PAX6* transcriptional unit and regulatory region from WGS

WGS VCF files for all 51 individuals were filtered with the VEP-G2P plugin²¹ using the EyeG2P²⁰ data set to detect high-impact and moderate-impact changes within genes known to cause genetic eye disease. Likely pathogenic *PAX6* sequence variants identified, described in the following, are listed in table 2 along with the American College of Medical Genetics and Genomics (ACMG) pathogenicity classifications.¹²

This revealed one proband (RPID 2134) with a novel frame-shift variant (NM_000280.4 *PAX6*: c.842_843insGT) affecting

Table 2 Likely pathogenic *PAX6* sequence variants (NM_000280.4)

Individual (RPID)	Family	Inheritance	Intron	GRCh38	CDS variant	Predicted consequence	ACMG/ACGS	Previously reported
Coding loss-of-function variants								
2134	2134	Unknown	n/a	chr11:31793725_31793726insAC	c.842_843insGT	p.(Pro282Tyrfs*84)	P (0.999) PM2, PVS1, PP4 mod	No
Essential splice site variants affecting 5' non-coding exons								
877	877	Unknown	IVS3+1	chr11:31806848C>G	c.-52+1G>C	p.(?) donor loss	LP (0.9) PM2, PS1 supp, PP3, PP4 mod*	No
1019	1019	Unknown	IVS3+1	chr11:31806848C>A	c.-52+1G>T	p.(?) donor loss	LP (0.949) PM2, PS4 supp, PP3, PS1 supp, PP4 mod*	Yes ²⁶
1500	1500	Unknown	IVS2-2	chr11:31806927del	c.-128-2del	p.(?) acceptor loss	LP (0.949) PM2, PS4 mod, PP3, PP4 mod	Yes ^{28–30 39} VCV000430969.2
5645	5645	<i>De novo</i>	IVS2-2	chr11:31806927del	c.-128-2del	p.(?) acceptor loss	P (0.997) PM2, PS4 mod, PP3, PS2, PP4 mod	As above
Deep intronic donor gain variants								
1635	1635	<i>De novo</i>	IVS8+68	chr11:31794562G>C	c.682+68C>G	p.(?) donor gain	LP (0.988) PM2, PP3, PS2, PP4 mod	No
3612	3612	Unknown	IVS6+334	chr11:31801227C>T	c.357+334G>A	p.(?) donor gain	LP (0.9) PM2, PS4 supp, PP3, PP4 mod	Yes, as uncertain pathogenicity ²⁹ VCV000559621.2

ACMG/ACGS pathogenicity classification¹ (and posterior probability calculated using the Bayesian framework tool in DECIPHER²).

*Arguable whether PVS1¹ could be applied to this canonical splice site in a non-coding exon.

ACGS, Association for Clinical Genomic Science; ACMG, American College of Medical Genetics and Genomics; CDS, coding sequence; del, deletion; ins, insertion; LP, likely pathogenic; mod, moderate; n/a, not available; P, pathogenic; RPID, research participant identifier; supp, supporting.

Developmental defects

a coding exon of *PAX6* constituting the sole ‘false negative’ for the prior screening of the *PAX6* coding region (WGS data viewed in IGV in online supplemental figure S3). This case had been screened many years previously using denaturing HPLC analysis. Although this was one of the ‘trios’, the ‘paternal’ sample was erroneously a duplicate of the maternal sample (short tandem repeat profiling in online supplemental table S3). We were thus unable to confirm whether the variant detected was *de novo*.

Essential splice site variants

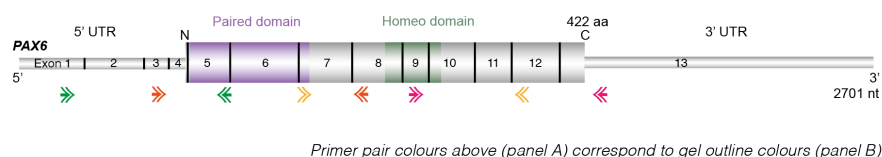
Four families (three singletons, one trio) have heterozygous essential splice site (ESS) variants flanking exon 3, part of the *PAX6* 5'UTR. In RPID 877 and RPID 1019, both variants affect the 5' base of intron 3 (IVS3+1, or c.-52+1), G>C and G>T, respectively (online supplemental figure S4A). Only the latter variant has been reported previously.²⁶ A different variant at the same position (c.-52+1G>A) has been previously reported as resulting in skipping of exons 3, 4, 5 and 5a²⁷; exon 4 contains the translation start site. RPID 1500 and RPID 5645 have identical and previously reported ESS variants (NM_000280.4 (*PAX6*): c.-128-2del (IVS2-2)).^{28–30} This variant occurred *de novo* in RPID 5645 (online supplemental figure S4B). The predicted effects of these sequence variants on splicing using the predictors in Alamut and SpliceAI are detailed in online supplemental table

S4. Nested RT-PCR was performed on LCL-derived cDNA from RPID 1500 and showed evidence of abnormal splicing between exons 1 and 5 (figure 1B(i)).

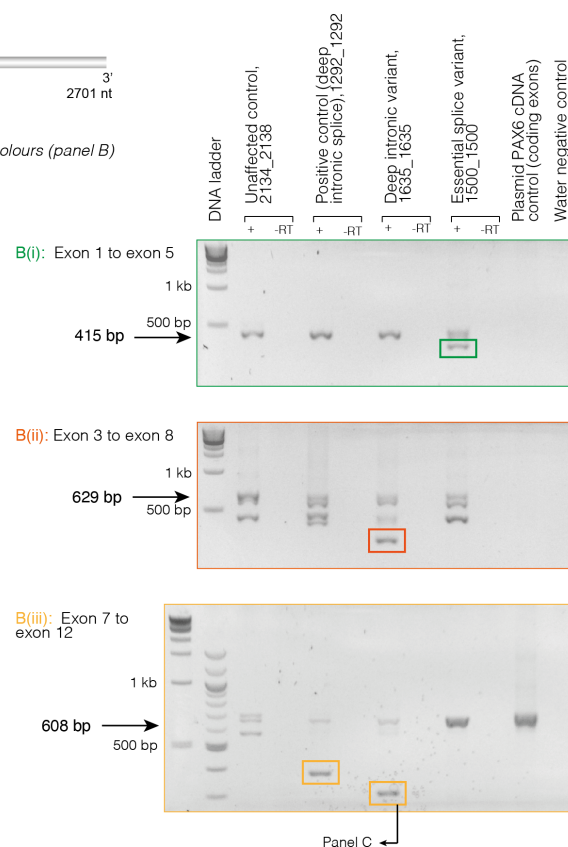
Deep intronic variants affecting splicing

In RPID 3612, a variant was detected in intron 6 (NM_000280.4(-*PAX6*):c.357+334G>A) (online supplemental figure S5A). SpliceAI, SSF, MaxEnt and NNSPLICE all predict this to result in a donor gain (online supplemental table S4). One previous occurrence of this variant has been reported in aniridia.²⁹ In RPID 1635, a novel *de novo* variant within intron 8 (NM_000280.4(-*PAX6*):c.682+68C>G) was identified (online supplemental figure S5B). SpliceAI, SSF, MaxEnt and NNSPLICE predict a donor gain consequence (online supplemental table S4). Nested RT-PCR was performed on LCL-derived cDNA from RPID 1635 and showed abnormal-sized bands using the primers spanning both exons 3–8 and 7–12. Sequencing the exons 3–8 product revealed unexpected skipping of exon 5 and part of exon 6. The exons 7–12 product-derived sequence showed skipping of exons 9, 10 and 11 (figure 1).

A Exonic location of primers for nested PCR



B RT-PCR products showing missplicing



C Sanger sequencing of missplicing product (RPID 1635)

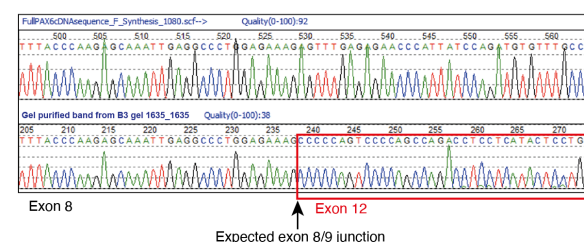


Figure 1 (A) Location of the four RT-PCR nested primer pairs spanning exons 1–5 (‘green’ pair), exons 3–8 (‘orange’ pair), exons 7–12 (‘yellow’ pair) and exons 9–13 (‘pink’ pair; products not shown as there was no evidence of mis-splicing in any of the cases). (B) Agarose gel images showing the RT-PCR products from the LCL-derived cDNA template. Samples are labelled as FID_RPID (family identifier_individual research participant identifier). (B(i)) Evidence of mis-splicing using green primer pair covering exons 1–5 in RPID 1500 (NM_000280.4:c.-128-2del). (B(ii)) The orange pair spanning exons 3–8 showed unexpected mis-splicing in RPID 1635 (intron 8 donor gain). (B(iii)) The yellow primers covering exons 7–12 showed mis-splicing in RPID 1635 and RPID 1292 (previously identified deep intronic variant in intron 8, not from this cohort and included as a positive control) suggestive of exon skipping. (C) Sanger sequence of gel-purified mis-spliced band confirming skipping of exons 9, 10 and 11 in RPID 1635. LCL, lymphoblastoid cell line.

Variants in *PAX6* CREs

To identify causative CRE mutations at the *PAX6* locus, we first created a BED file listing the GRCh38 genome coordinates of 35 previously characterised CREs^{31,32} (online supplemental table S5). Intersecting this BED file with the annotated cohort VCF file identified 39 variants that passed quality filters and were present in at least 1 out of 52 sequenced individuals (online supplemental table S6). Of 35 CREs examined, 24 encompassed one or more variant (online supplemental table S5). Of 39 variants, 20 had allele count within the study population of nine or greater. Of the remaining 19 variants, 15 had allele counts of one. Of 39 variants, 38 were present in gnomAD (online supplemental table S6), and the variant absent from gnomAD was a non-transmitted allele from an unaffected father in a trio.²³ No *de novo* CRE SNVs or INDELs were identified. It thus seems very unlikely that any of the CRE variants are of clinical significance for aniridia.

Categorisation of *de novo* variants in genes other than *PAX6*

The only individual with *de novo* SNVs or INDELs outwith *PAX6* and with no causative SVs (described in the next section) was RPID 2469, who had five such variants (online supplemental table S7). This was the aforementioned trio for which the parents' DNA was extracted from LCLs. Only the variant in the gene

encoding adenosylhomocysteinase 3 (*AHCYL2*: p.(Leu352Met)) merited further consideration. This variant (NM_015328.4(*AHCYL2*):c.1054C>A) is not present in gnomAD, and has CADD and REVEL scores of 24.3 and 0.73, respectively. There is no known Mendelian disease–gene link for *AHCYL2* and no claim can be made on the clinical significance of this variant.

Identification of large-scale SVs from WGS

A combination of direct inspection of candidate loci using the IGV³³ and genome-wide bioinformatic tools (Canvas³⁴ and Manta¹⁶) was used to identify SVs from the available WGS data. A total of 17 different ultra-rare heterozygous SVs affecting *PAX6* and *FOXC1* were detected in 15 families (table 3 for genomic coordinates).

When compared with direct visual inspection, Canvas detected all deletions > 10 kb at the *PAX6* and *FOXC1* loci (online supplemental figure S6) but not the two smallest *PAX6* deletions (126 bp, RPID 75, and 1.36 kb, RPID 1524). Canvas also called the *PAX6* deletion-duplication in RPID 1271 but was unable to detect the inversions in RPID 535 and RPID 774 as it uses read depth only. Manta detected all likely causative SVs but together with many false positive calls, so in practice these were identified solely by direct visualisation of the breakpoint regions using IGV.

Table 3 Likely pathogenic structural variants altering the *PAX6* or *FOXC1* loci

Individual (RPID)	Family	Variant type	Inheritance	Region affected†	Size (kb)	Genomic coordinates (GRCh38)
Simple deletions affecting <i>PAX6</i> transcription unit						
75	75	Deletion	<i>De novo</i>	<i>PAX6</i> intron 7/exon 8	0.126	chr11:31794763-31794889
724	723	Deletion	Unknown	<i>PAX6</i> whole gene deletion	191	chr11:31714074-31905175
1496	1496	Deletion	Unknown	C-termini of both <i>ELP4</i> and <i>PAX6</i> (intron 8 onwards)	19	chr11:31777022-31796016
1524	1524	Deletion	Unknown	<i>PAX6</i> exons 6 and 7	1.4	chr11:31800812-31802170*
2464	2464	Deletion	Segregating	C-termini of both <i>ELP4</i> and <i>PAX6</i> (intron 4 onwards)	41	chr11:31762967-31803741
Simple deletions altering <i>PAX6</i> cis-regulation						
1191	1190	Deletion	Unknown	<i>PAX6</i> DRR	674	chr11:31009535-31683449
1361	1361	Deletion	Unknown	<i>PAX6</i> DRR	598	chr11:31118961-31716799
1647	1646	Deletion	Unknown	<i>PAX6</i> DRR	842	chr11:30837881-31680038
Balanced rearrangements altering <i>PAX6</i> locus						
535	535	Inversion	Unknown	11p13-11p14.3 with intragenic <i>PAX6</i> breakpoint	7300	chr11:24479030-31792704
774	774	Inversion	Unknown	Inversion separating <i>PAX6</i> from DRR; breakpoints <i>ELP4</i> and <i>RAG2</i>	4900	chr11:31701464-36593500
Complex structural variants altering <i>PAX6</i> locus						
356	356	Inversion	<i>De novo</i>	N-terminus of <i>PAX6</i> up to intron 4	6	chr11:31805500-31811582
		Deletion		Upstream of <i>PAX6</i> (P0, EE enhancer)	30	chr11:31811583-31842057
1271	1271	Tandem duplication	Unknown	Final intron of <i>ELP4</i>	13	chr11:31765420-31778063
		Deletion		C-termini of both <i>ELP4</i> and <i>PAX6</i> (intron 8 onwards)	16	chr11:31780738-31796994
Simple deletions affecting <i>FOXC1</i>						
1142	1142	Deletion	Unknown	<i>FOXC1</i> whole gene deletion	82.6	chr6:1544064-1626702
1451	1451	Deletion	Unknown	<i>FOXC1</i> whole gene deletion	33.0	chr6:1581517-1615082
1732	1732	Deletion	Unknown	<i>FOXC1</i> whole gene deletion	81.6	chr6:1604452-1686063
*This individual's structural variant was detected independently elsewhere and is patient 11 in the cited publication. ³⁵						
†Exon numbering for <i>PAX6</i> is as per NM_000280.4 (-5a, 13 exons)						
DRR, Downstream regulatory region; RPID, research participant identifier.						

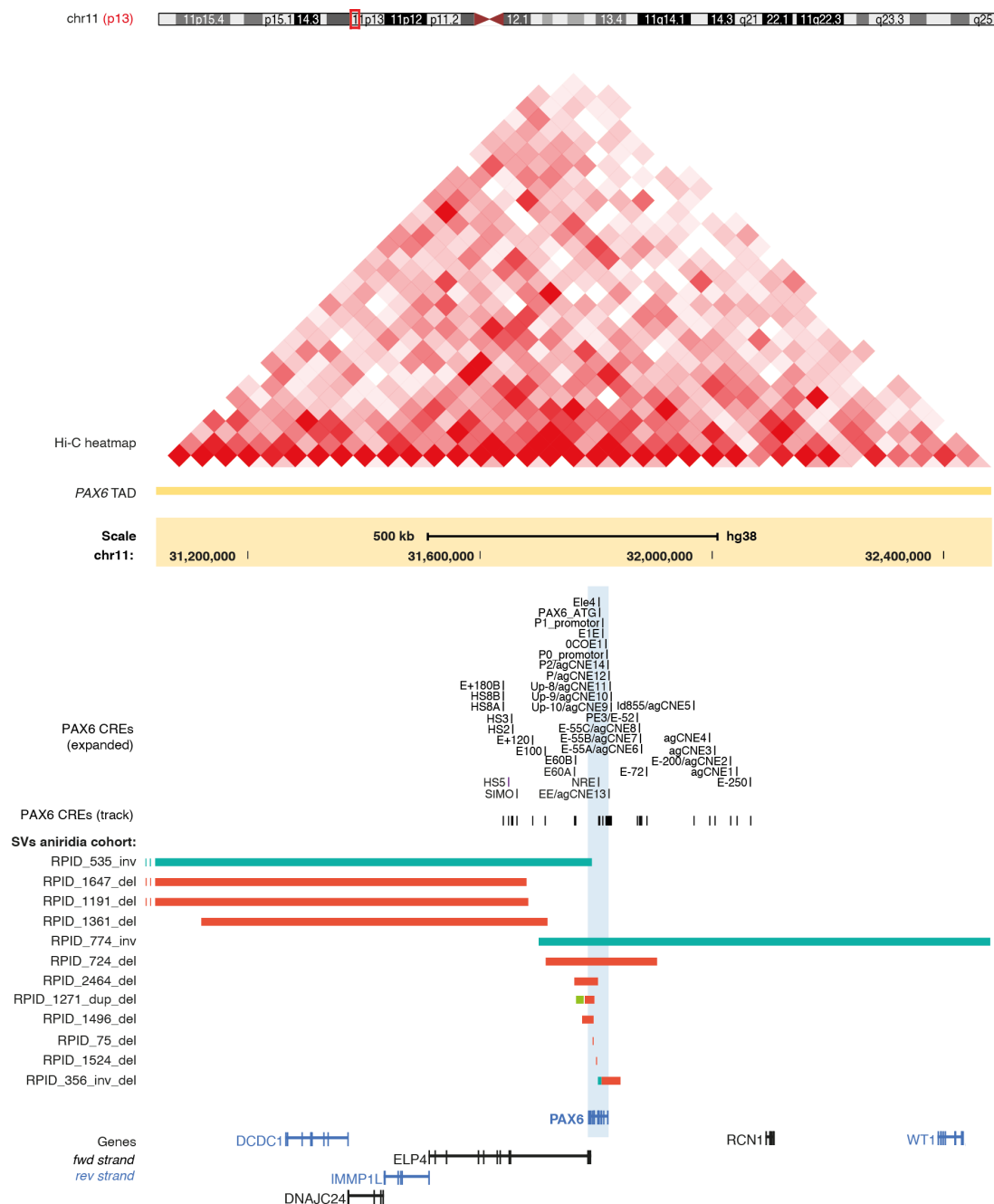


Figure 2 Structural variants (SV) identified on WGS affecting the wider *PAX6* locus. Each SV is shown as horizontal bars (inv, inversion, teal; del, deletion, red; dup, duplication, green). The *PAX6* topologically associated domain (TAD) is indicated by the Hi-C heatmap. The position of *PAX6* cis-regulatory elements (CREs) is shown as track. Gene regulatory features such as the promoters and ATG are included. The position of *PAX6* is shaded blue. Seven of the SVs have intragenic breakpoints and RPID 724 has a whole gene deletion. Three SVs are deletions of the downstream regulatory region, taking out CREs implicated in aniridia, notably SIMO and HS5. Similarly, the SV seen in RPID 774 inverts *PAX6*, disrupting its relationship with these enhancers. A smaller scale version of this figure, showing the full span of the largest SVs, is shown in online supplemental figure S9. RPID, research participant identifier; WGS, whole genome sequencing.

Whole or partial deletions of *PAX6*

Five individuals or families were found to have simple heterozygous deletions involving the *PAX6* transcription unit (figure 2, online supplemental figure S7): a whole gene deletion of 191 kb (RPID 724) and four partial deletions of 19 kb (RPID 1496), 1.4 kb (individual 1524), 0.13 kb (RPID 75) and 41 kb (RPID 2464). Each of these variants is expected to result in *PAX6* haploinsufficiency. The variant detected in individual 1524 was subsequently found to have been identified independently by others.³⁵

Deletions encompassing *PAX6* cis-regulatory domains

RPID 1191, RPID 1361 and RPID 1647 were identified with likely causative deletions encompassing well-characterised CREs that control the developmental expression of *PAX6* (figure 2, online supplemental figures S8 and S9).

Complex structural rearrangements of chromosome 11 detected on WGS

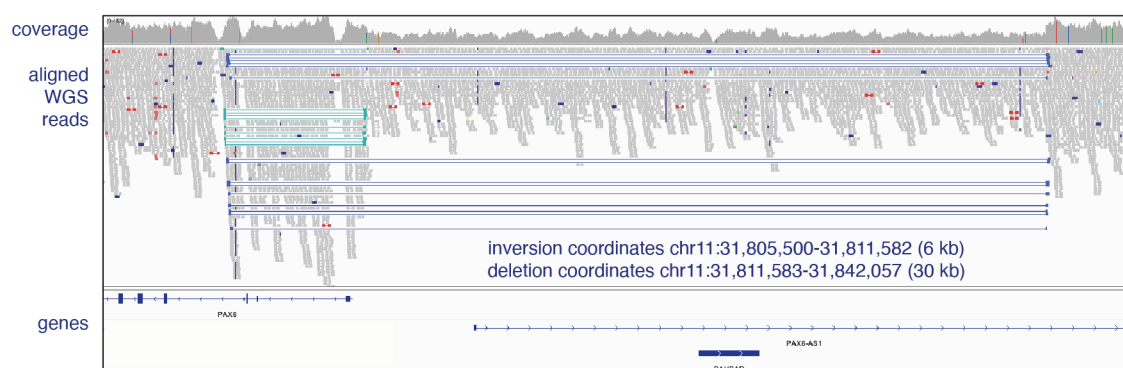
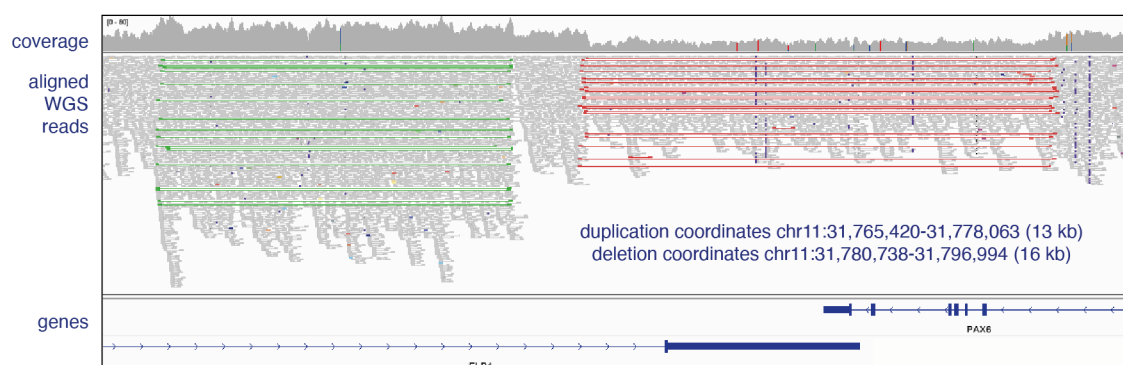
11p13 **A** RPID 356: *de novo* inversion and deletion**B** RPID 1271: tandem duplication and deletion

Figure 3 Complex structural rearrangements of chromosome 11 in two unrelated individuals with aniridia. Aligned WGS data viewed with IGV, with reads viewed as pairs and coloured both by insert size and by pair orientation. Coordinates estimated from WGS data (GRCh38). (A) RPID 356: an individual with sporadic bilateral aniridia and congenital cataracts. Trio WGS data, including the unaffected parents, are shown in online supplemental figure S11. WGS data show a *de novo* 6 kb inversion involving the P1 promoter, all of the 3'UTR and the first coding exon (exon 4) of *PAX6*; next to this is a 30 kb deletion, which deletes the P0 promoter and several enhancers, including EE. The blue and teal colours both denote paired reads with abnormal pair orientation. In IGV, pair orientation is determined first, and only if this is as expected will abnormal insert size then be flagged. Therefore, the reads across this deletion are not flagged in red (as they would be in a simple deletion) as they also span the inversion. A drop in coverage depth is seen in the deleted area. (B) RPID 1271: an individual (single proband) with bilateral aniridia and cataracts. WGS data indicate a 16 kb deletion (red) involving the six last exons of *PAX6* and a 13 kb tandem duplication (green) affecting the final intron of *ELP4*. Coverage depth is increased over the putative duplication and decreased over the putative deletion. IGV, Integrative Genomics Viewer; RPID, research participant identifier; WGS, whole genome sequencing.

Balanced structural rearrangements disrupting *PAX6*

Two inversions of chromosome 11 were detected with breakpoints within or very close to *PAX6* (figure 2, online supplemental figures S9 and S10). The *PAX6* gene is directly disrupted in RPID 535, while in RPID 774 the breakpoint is between *PAX6* and the critical CREs SIMO and HS5.

Complex structural rearrangements disrupting *PAX6*

RPID 356 was found to carry a *de novo* 6 kb heterozygous inversion with breakpoints at start of *PAX6* in intron 4, with an adjacent 30 kb region of *PAX6* deleted. In proband 1271, a 16 kb deletion encompassing the final six exons of *PAX6* was associated with a 13 kb tandem duplication immediately 3' of

PAX6 (figure 3; also figure 2 and online supplemental figures S9 and S11).

Deletions encompassing *FOXC1*

Three heterozygous chromosome 6p deletions encompassing *FOXC1* were identified in three probands: RPID 1142, RPID 1451 and RPID 1732 (table 3, online supplemental figure S12). These ranged from 33 kb to 83 kb in size. All three probands had aniridia; two out of three had glaucoma (one confirmed as congenital) and two out of three had congenital aortic or aortic valve anomalies (table 1). The combination of aniridia with congenital glaucoma and aortic valvular disease would be consistent with previously reported *FOXC1* deletions.³⁶

Breakpoint identification in a coincidental *de novo* reciprocal translocation t(1,9)

RPID 356 has a *de novo* reciprocal translocation t(1,9)(p36.1;q22), which was detected by routine cytogenetic analysis following the clinical diagnosis of aniridia. Given that no *PAX6* coding region mutation was identified on initial screening, the family was referred to our study to determine whether the breakpoint of this translocation could identify a novel locus or mechanism causing aniridia. However, as shown above, this individual has a second SV which disrupts *PAX6* and explains the phenotype. Using IGV, discrepant paired-end reads mapped a single breakpoint on chromosome 1 and two different breakpoints on chromosome 9, consistent with a paracentric inversion on chromosome 9 and a reciprocal translocation with chromosome 1 (online supplemental figure S13). No clinical impact is suspected for these three breakpoints.

DISCUSSION

In purely diagnostic terms, short-read WGS has significant advantages over short-read whole exome sequencing (WES). First, WGS allows reliable analysis of the whole transcription unit of each gene. This power is evidenced by our identification of previously cryptic causative variants in the 5'UTR and deep intronic regions of *PAX6* in 6 out of 22 (27.3%) diagnosed cases. The 5'UTR ESS variants perturb *PAX6* splicing; however, consequential changes to the length of the *PAX6* upstream ORF³⁷ and/or disruption of VAX2 binding³⁸ may also have mechanistic significance. More notably, we detected two deep intronic variants and tested the functional consequence of the novel one, predicted to result in an intron 8 donor site gain, using cDNA from an LCL derived from the proband. We could demonstrate the expected exon skipping 3' to this variant, but we also found aberrant splice events 5' to this intron, suggesting a more complex effect on splicing. While WGS may have better coverage of 5'UTRs than WES, it is particularly the deep intronic regions where it has a unique advantage.

A second advantage of WGS is more uniform per base coverage when compared with WES. This significantly improves our ability to detect disease-associated balanced structural variants (bSV) and CNVs. Our initial CNV screen was performed via direct inspection of the coverage depth change and unexpected pairing of end sequencing in proband BAM files using IGV.³³ This proved to be the most diagnostically rewarding analysis undertaken in this study, yielding 15 of the 22 new diagnoses. Of these 15, 13 were CNVs (10 at *PAX6* locus and 3 encompassing *FOXC1*) and 2 were balanced SVs (bSVs) with *PAX6*-disruptive breakpoints, the latter not easily detectable by bioinformatic SV calling with Manta due to noise. The high CNV yield^{26 39 40} reflects both prescreening of the cohort for *PAX6* coding variants and the historic nature of some samples in our cohort, as these anomalies would almost certainly be detected by modern high-resolution, array-based methods of copy number assessments now used in clinical diagnostic laboratories throughout the world. On the contrary, the two bSVs would be unlikely to be detected on standard clinical testing other than WGS. The identification of an apparently coincidental *de novo* balanced reciprocal translocation in RPID 356 is interesting but has been observed in other developmental disorders in which a second intragenic SV is subsequently determined to be causative.⁴¹

The third, and possibly most exciting, advantage of WGS in the diagnostic investigation of classic aniridia is the ability to identify causative *cis*-regulatory variants affecting the developmental expression of *PAX6*. CNVs and bSVs encompassing

CREs of *PAX6* but leaving the transcription unit intact have been recognised as resulting in functional haploinsufficiency for many years.^{42 43} Predicting the consequence of SNVs within CRE remains challenging, and currently only one *de novo* plausibly causative CRE SNV in classic aniridia has been reported.⁴⁴ We did not identify any additional CRE SNVs in this study, although we did identify four SVs affecting only the *PAX6* downstream regulatory region (three deletions and one bSV), leaving the gene itself intact. A similar *PAX6* bSV is recently reported amongst a large, more diverse clinical diagnostic cohort⁴⁵.

On the basis of the work from others^{36 45 46} and ourselves,⁴⁷ the identification of *FOXC1* deletions is not surprising from a human genetics perspective. There is remarkably little information about developmental genetic interactions between these two genes, although it has been shown that *FOXC1* is a downstream direct target of *PAX6* in the developing iris and ciliary body.⁴⁸

A fourth strength of WGS is that it permits a search for new candidate genes, and the mechanisms of inactivating known genes, in unexplained cases. We did not identify any likely causative variants at loci other than *PAX6* or *FOXC1*. A study with a larger number of trios would have greater power to detect new candidate loci.

We consider that the data presented here provide evidence that short-read WGS merits consideration as a primary investigation for classic aniridia. It certainly should be considered in cases with a normal array-based assessment of genome-wide copy number and *PAX6* coding region sequencing. We are mindful that WGS analysis is not currently capable of explaining all cases of aniridia, and there remain 15 out of 37 families in this study in whom we have still not identified a causative variant. One useful emerging diagnostic technology is long-read nanopore-based genome sequencing. This may be particularly useful in identifying bSV missed by the short-read technologies.^{49–51}

Author affiliations

¹Institute of Genetics and Cancer, The University of Edinburgh MRC Human Genetics Unit, Edinburgh, UK

²Illumina United Kingdom, Edinburgh, UK

³MRC Human Genetics Unit, The University of Edinburgh, Edinburgh, UK

⁴East Anglia Regional Genetics Service, Addenbrooke's Hospital, Cambridge, UK

⁵Department of Medical Genetics, Telemark Hospital, Skien, Norway

⁶Consultations de Génétique médicale, Service de Pédiatrie, CHU Paris-Nord, Hôpital Jean Verdier, Bondy, France

⁷Centre for Clinical Genetics, Sydney Children's Hospital Randwick, Randwick, New South Wales, Australia

⁸Department of Clinical Genetics, University Hospitals Bristol NHS Foundation Trust, Bristol, UK

⁹University Hospital Southampton, NHS Foundation Trust Wessex Clinical Genetics Service, Southampton, UK

¹⁰Department of Genetics, Hospital Ruber Internacional, Madrid, Spain

¹¹Institute of Ophthalmology, University College London, London, UK

Twitter Hildegard Nikki Hall @nikkihallyg

Acknowledgements We thank the patients, their families and the clinicians/scientists who submitted research samples to the MRC Human Genetics Unit Eye Malformation Study. These include Professor Jurgen Kohlhase, Dr Eduardo Silva, Professor Birgit Lorenz, Mr Ken Nischal, Professor Alain Verolles, Professor Lawrence Hirst, Dr Stavits Shalev, Dr Peter Turnpenny, Professor Anthony Moore, Dr Debbie Shears, Dr Katherine Bushby, Dr Ian Young, Dr Eamonn Maher, Dr William Reardon, Dr Gyorgy Fekete, Dr Sebastiana Bianca, Dr Matt Hawker, Dr Fiona Stewart, Mr David Taylor, Miss Isabelle Russell-Eggitt, Dr Miranda Splitt, Dr Victoria Murday, Dr John Tolmie, Dr Sahar Mansour and the late Dr Robin Winter. Thanks to Angela Sandilands for managing the research database over many years. Thanks to Dr Sevan Yazar, Dr Graeme Grimes, Dr Andrea Robertson and Dr Morad Ansari for their help. Thanks to the MRC Institute of Genetics and Cancer Core sequencing services, in particular Stephen Brown.

Contributors HNH and DRF conceived and planned the study. Molecular analyses were performed by HNH, KW and JJ. Analyses of WGS data were performed by HNH, DP, MH, AM, KD, JCP and DRF. SB provided CRE data and advice. SH, TEP, EPK, JB,

KL and PB and others reviewed the manuscript and contributed key samples to the HGU Eye Malformations database, which was developed by VvH, KW and DRF. The manuscript was written by HNH and DRF with contributions from the coauthors. HNH is responsible for the overall content and acts as guarantor.

Funding HNH was funded by a Wellcome ECAT PhD studentship (205171/Z/16/Z). DRF's research programme and MH, KW, SB, JJ, DP and AM's posts are or were funded by an MRC University Unit award to the MRC Human Genetics Unit, The University of Edinburgh. SB was supported by an MRC University Unit grant (MC_UU_00007/2). KD and JCP were funded by the Scottish Genomes Partnership: Chief Scientist Office of the Scottish Government Health Directorates (SGP/1) and the Medical Research Council Whole Genome Sequencing for Health and Wealth Initiative (MC/PC/15080).

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval This study involves human participants. This research participant cohort was collected and maintained using consent processes and the experimental protocols were approved by the Scotland A UK Multicentre Research Ethics Committee (references 06/MRE00/76 and 16/SS/0201). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. WGS sequence data have been deposited at the European Genome-Phenome Archive (EGA), which is hosted by the EBI and the CRG, under accession number EGAS00001006878. This was deposited in July 2023 and will soon be available publicly. GitHub links and publications for the various components are provided within the Methods section. Further codes are available from the corresponding author on reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

ORCID iD

Hildegard Nikki Hall <http://orcid.org/0000-0001-7082-1523>

REFERENCES

- Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015;17:405–24.
- Tavtigian SV, Greenblatt MS, Harrison SM, et al. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genet Med* 2018;20:1054–60.
- The 100,000 Genomes Project Pilot Investigators, Smedley D, Smith KR, et al. 100,000 Genomes pilot on rare-disease diagnosis in health care. *N Engl J Med* 2021;385:1868–80.
- Ellingford JM, Ahn JW, Bagnall RD, et al. Recommendations for clinical interpretation of variants found in non-coding regions of the genome. *Genome Med* 2022;14:73.
- Hall HN, Williamson KA, FitzPatrick DR. The genetic architecture of aniridia and Gillespie syndrome. *Hum Genet* 2019;138:881–98.
- Marquardt T, Ashery-Padan R, Andrejewski N, et al. Pax6 is required for the multipotent state of retinal progenitor cells. *Cell* 2001;105:43–55.
- Bhatia S, Gordon CT, Foster RG, et al. Functional assessment of disease-associated regulatory variants in vivo using a versatile dual colour transgenesis strategy in zebrafish. *PLoS Genet* 2015;11:e1005193.
- Rowan S, Siggers T, Lachke SA, et al. Precise temporal control of the eye regulatory gene Pax6 via enhancer-binding site affinity. *Genes Dev* 2010;24:980–5.
- Li H. Aligning sequence reads, clone sequences and assembly Contigs with BWA-MEM. 2013.
- Faust GG, Hall IM. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* 2014;30:2503–5.
- Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high-confidence variant calls: the genome analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 2013;43:11.
- Poplin R, Ruano-Rubio V, DePristo MA, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. *Genomics* [Preprint] 2018.
- Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 2013;14:178–92.
- Canvas Copy Number Variant Caller. 2022. Available: <https://github.com/illumina/canvas> [Accessed 1 Dec 2022].
- Ivakhno S, Roller E, Colombo C, et al. Canvas SPW: calling de novo copy number variants in pedigrees. *Bioinformatics* 2018;34:516–8.
- Chen X, Schulz-Trieglaff O, Shaw R, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* 2016;32:1220–2.
- Manta Structural Variant Caller. 2022. Available: <https://github.com/illumina/manta> [Accessed 1 Dec 2022].
- Parry DA. Vase. 2022. Available: <https://github.com/david-a-parry/vase> [Accessed 1 Dec 2022].
- Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, et al. Predicting splicing from primary sequence with deep learning. *Cell* 2019;176:535–48.
- Lenassi E, Carvalho A, Thormann A, et al. EyeG2P: an automated variant filtering approach improves efficiency of diagnostic genomic testing for inherited ophthalmic disorders. *Genetic and Genomic Medicine* [Preprint] 2021.
- Thormann A, Halachev M, McLaren W, et al. Flexible and Scalable diagnostic filtering of genomic variants using G2P with Ensembl VEP. *Nat Commun* 2019;10:2373.
- Quinlan AR. BEDTools: the Swiss-army tool for genome feature analysis. *Curr Protoc Bioinformatics* 2014;47:11.
- Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;581:434–43.
- Freeman PJ, Hart RK, Gretton LJ, et al. VariantValidator: accurate validation, mapping, and formatting of sequence variation descriptions. *Hum Mutat* 2018;39:61–8.
- Chomczynski P, Sacchi N. The single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction: twenty-something years on. *Nat Protoc* 2006;1:581–5.
- You B, Zhang X, Xu K, et al. Mutation spectrum of PAX6 and clinical findings in 95 Chinese patients with aniridia. *Mol Vis* 2020;26:226–34.
- Grenskov K, Rosenberg T, Sand A, et al. Mutational analysis of PAX6: 16 novel mutations including 5 missense mutations with a mild aniridia phenotype. *Eur J Hum Genet* 1999;7:274–86.
- Axton RA, Hanson IM, Love J, et al. Combined SSCP/heteroduplex analysis in the screening for PAX6 mutations. *Mol Cell Probes* 1997;11:287–92.
- Plaisancié J, Tarilonte M, Ramos P, et al. Implication of non-coding PAX6 mutations in aniridia. *Hum Genet* 2018;137:831–46.
- Robinson DO, Howarth RJ, Williamson KA, et al. Genetic analysis of chromosome 11P13 and the PAX6 gene in a series of 125 cases referred with aniridia. *Am J Med Genet A* 2008;146A:558–69.
- Bhatia S, Monahan J, Ravi V, et al. A survey of ancient conserved non-coding elements in the PAX6 locus reveals a landscape of interdigitated cis-regulatory archipelagos. *Dev Biol* 2014;387:214–28.
- Lima Cunha D, Arno G, Corton M, et al. The spectrum of PAX6 mutations and genotype-phenotype correlations in the eye. *Genes (Basel)* 2019;10:1050.
- Robinson JT, Thorvaldsdóttir H, Wenger AM, et al. Variant review with the integrative Genomics viewer. *Cancer Res* 2017;77:e31–4.
- Roller E, Ivakhno S, Lee S, et al. Canvas: versatile and Scalable detection of copy. *Bioinformatics* 2016;32:2375–7.
- Ma A, Yousoof S, Grigg JR, et al. Revealing hidden genetic diagnoses in the ocular anterior segment disorders. *Genet Med* 2020;22:1623–32.
- Sadagopan KA, Liu GT, Capasso JE, et al. Aniridia-like phenotype caused by 6P25 dosage aberrations. *Am J Med Genet A* 2015;167A:524–8.
- Filatova AY, Vasilyeva TA, Marakhonov AV, et al. Upstream ORF Frameshift variants in the Pax6 5'UTR cause congenital Aniridia. *Hum Mutat* 2021;42:1053–65.
- Lee J, Suh Y, Jeong H, et al. Aberrant expression of PAX6 gene associated with classical aniridia: identification and functional characterization of novel noncoding mutations. *J Hum Genet* 2021;66:333–8.
- Vasilyeva TA, Voskresenskaya AA, Kasmann-Kellner B, et al. Molecular analysis of patients with Aniridia in Russian Federation broadens the spectrum of Pax6 mutations. *Clin Genet* 2017;92:639–44.
- Bobilev AM, McDougal ME, Taylor WL, et al. Assessment of PAX6 alleles in 66 families with aniridia. *Clin Genet* 2016;89:669–77.
- Ireland M, English C, Cross I, et al. A de novo translocation T(3;17)(Q26.3;Q23.1) in a child with Cornelia de Lange syndrome. *J Med Genet* 1991;28:639–40.
- Fantes J, Redeker B, Breen M, et al. Aniridia-associated cytogenetic Rearrangements suggest that a position effect may cause the mutant phenotype. *Hum Mol Genet* 1995;4:415–22.
- Lauderdale JD, Wilensky JS, Oliver ER, et al. 3' deletions cause aniridia by preventing PAX6 gene expression. *Proc Natl Acad Sci U S A* 2000;97:13755–9.

- 44 Bhatia S, Bengani H, Fish M, *et al.* Disruption of autoregulatory feedback by a mutation in a remote, ultraconserved PAX6 enhancer causes aniridia. *Am J Hum Genet* 2013;93:1126–34.
- 45 Daruich A, Duncan M, Robert MP, *et al.* Congenital Aniridia beyond black eyes: from phenotype and novel genetic mechanisms to innovative therapeutic approaches. *Prog Retin Eye Res* 2023;95:S1350-9462(22)00093-3:101133:..
- 46 Khan AQ, Aldahmesh MA, Al-Amri A. Heterozygous FOXC1 mutation (M161K) associated with congenital glaucoma and aniridia in an infant and a milder phenotype in her mother. *Ophthalmic Genet* 2008;29:67–71.
- 47 Ansari M, Rainger J, Hanson IM, *et al.* Genetic analysis of 'PAX6-negative' individuals with aniridia or Gillespie syndrome. *PLoS One* 2016;11:e0153757.
- 48 Wang X, Shan X, Gregory-Evans CY. A mouse model of aniridia reveals the in vivo downstream targets of PAX6 driving Iris and ciliary body development in the eye. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 2017;1863:60–7.
- 49 Hu L, Liang F, Cheng D, *et al.* Location of balanced Chromosome-translocation Breakpoints by long-read sequencing on the Oxford nanopore platform. *Front Genet* 2019;10:1313.
- 50 Marks P, Garcia S, Barrio AM, *et al.* Resolving the full spectrum of human genome variation using linked-reads. *Genome Res* 2019;29:635–45.
- 51 Damián A, Núñez-Moreno G, Jubin C, *et al.* Long-read genome sequencing identifies cryptic structural variants in congenital Aniridia cases. *Hum Genomics* 2023;17:45:45:..

C.2 Supplemental Methods for WGS study

C.2.1 Detailed whole genome sequencing methods

Further detailed methods for the massively parallel sequencing library preparation follow here, paraphrased from the methods I received from each company:

C.2.2 For BGI Hong Kong WGS

Next generation sequencing libraries were prepared as follows: 1µg genomic DNA was sheared by a Covaris ultrasonicator. The DNA fragments were assessed by gel electrophoresis, purified using AxyPrep Mag PCR clean-up kit (Axygen), incubated with End Repair Mix at 20°C for 30 min and then purified again. The repaired DNA fragments were A-tailed at the 3' end (37°C for 30 min), adapter ligated (16°C for 16 hr), purified and then size selected (350 bp) prior to PCR amplification and purification. MGIEasy DNA library construction kits (MGI Tech) were used.

C.2.3 For Edinburgh Genomics WGS

In addition to my QC outlined above, gDNA samples were further evaluated by Edinburgh Genomics for quantity using Quant-iT Picogreen reagent (Thermo Fisher) and Spectramax XPS Gemini plate reader (Molecular Devices, Wokingham, UK), and quality using an AATI Fragment Analyzer (Agilent, Stockport, UK) (total amount >1000 ng and quality score >5 passed sample QC). gDNA samples were pre-normalised to a concentration range of 5-40 ng/µL.

Next generation sequencing libraries were prepared using an Illumina SeqLab specific TruSeq Nano High Throughput library preparation kit in conjunction with the Hamilton MicroLab STAR and Clarity LIMS X (4.2) Edition. 200 ng gDNA sample inputs were sheared to a 450 bp mean insert size using a Covaris LE220 focused-ultrasonicator. The inserts were blunt ended, A-tailed, and size selected, with TruSeq adapters ligated to each fragment before PCR amplification.

Insert size for each library was evaluated using the Caliper GX Touch with a HT DNA 1k/12K/HI SENS LabChip and HT DNA HI SENS Reagent Kit, ensuring mean fragment sizes of 530 bp to 950 bp. The Roche LightCycler 480 and a Kapa Illumina Library Quantification kit were used to ensure the concentration of each library was

between 5.5 nM and 40 nM. Libraries are normalised to 1.5 nM and denatured for clustering and sequencing at 300 pM using a Hamilton MicroLab STAR with Genologics Clarity LIMS X (4.2) Edition. Libraries were clustered onto a HiSeqX Flow cell v2.5 on cBot2s and the clustered flow cell transferred to a HiSeqX for sequencing using a HiSeqX Ten Reagent kit v2.5.

C.2.4 Primers

C.2.5 Primers used in droplet digital PCR

Oligo name	Region targeted	UPL no	Forward primer	Reverse primer
PAX6_ddPCR_ex2	5'UTR (exon 2)	12	cgctgacagccattttatt	ggtgtgtgagagcaattctcag
PAX6_ddPCR_ivs7b	Middle of <i>PAX6</i> (deep intron 7)	10	aatgacccccaaggatac	ctctggccaagttttgagga
PAX6_ddPCR_3utr	3'UTR (exon 13)	60	aaaggtccttggttcaagtcca	ttagagccgcttcagttctaca
PAX6_ddPCR_SIMO	SIMO	72	aaacagattactccttgaaacaggaa	gcgggaactactctgacaaca

Primer name	Sequence 5' to 3'
<i>Exon 1 to exon 5:</i>	
Green outer F (B1466)	GCCAGTGAGGAGCGGTG
Green outer R (B1468)	CTGCAGAATTCGGGAAATG
Green inner F (B1527)	CACATTAACACACTTGAGCC
Green inner R (B1463)	TGTGAGCTAGCTCTACAATC
<i>Exon 3 to exon 8:</i>	
Orange outer F (B1677)	GCGCAGATGTGTGAGGCC
Orange outer R (B1678)	GTTGGAAGTGTGAGTTGG
Orange inner F (B1675)	GAGAGTGGACAGACATCCG
Orange inner R (B1676)	TCCTTCCTGTTGCTGGCAG
<i>Exon 7 to exon 12:</i>	
Yellow outer F (B1523)	AGCAACAGATGGGCGCAG
Yellow outer R (B1524)	ATTCACCGAAGGGCTGGTG
Yellow inner F (B1525)	GGCATGTATGATAAACTAAGG
Yellow inner R (B1526)	GCAGGAGTATGAGGAGGTC
<i>Exon 9 to 3' UTR:</i>	
Pink outer F (B1681)	GCCCGAGAAAGACTAGCAG
Pink outer R (1682)	CTGAAGCGGCTCTAACAGC
Pink inner F (B1679)	CAAAATAGATCTACCTGAAGC
Pink inner R (B1680)	TGTTGTGTCCCATAGTCAC

C.3 Variants in *PAX6* cis-regulatory elements

No plausibly causative *cis*-regulatory element (CRE) variants were identified, a very relevant negative result when considering the overall analysis of the cohort.

To identify causative CRE variants at the *PAX6* locus I first created a BED file listing the GRCh38 genome coordinates of 35 previously characterised CREs (17,323) with the help of Shipra Bhatia (**Supplementary Table 2**, below). I acknowledge David FitzPatrick who conducted the following analysis. Intersecting this BED file with the annotated cohort VCF file identified 39 variants that passed quality filters and were present in at least 1/52 sequenced individuals (**Supplementary Table 3**). 24/35 CREs examined encompassed one or more variant. 20/39 variants had allele count within the study population of 9 or greater. 15 of the remaining 19 variants had allele counts of 1. Notably, 38/39 variants were present in gnomAD, and the variant absent from gnomAD was a non-transmitted allele from an unaffected father in a trio (200). No *de novo* CRE variants were identified. In conclusion, none of the CRE variants identified in this cohort are likely to be causative.

Supplementary Table 2 GRCh38 coordinates of *PAX6* cis-regulatory elements examined

Chromosome	Start	Finish	CRE Name	No. Variants
chr11	31640399	31641876	E+180B	1
chr11	31640562	31641230	HS8B	1
chr11	31641320	31641927	HS8A	0
chr11	31649471	31650142	HS5	1
chr11	31655216	31655799	HS3	3
chr11	31655891	31656443	HS2	1
chr11	31664033	31664626	SIMO	1
chr11	31691131	31691766	E+120	1
chr11	31712697	31713381	E100	1
chr11	31763231	31763878	E60A	2
chr11	31763748	31765309	E60B	2
chr11	31803813	31804584	NRE	1
chr11	31804143	31804358	Ele4	0
chr11	31806399	31806399	PAX6_ATG	0
chr11	31810960	31811514	P1_promotor	0
chr11	31811160	31811217	E1E	0
chr11	31816261	31816745	0COE1	3
chr11	31817808	31817962	P0_promotor	0
chr11	31819763	31820175	P2_agCNE14	0
chr11	31821668	31821983	EE_agCNE13	0
chr11	31822082	31822520	P_agCNE12	0
chr11	31823784	31824857	Up-8_agCNE11	4
chr11	31825288	31825612	Up-9_agCNE10	1
chr11	31826303	31826903	Up-10_agCNE9	1
chr11	31870906	31871516	PE3_E-52	1
chr11	31874695	31875223	E-55C_agCNE8	1
chr11	31876017	31876485	E-55B_agCNE7	2
chr11	31877182	31877419	E-55A_agCNE6	2
chr11	31886918	31887337	E-72	2
chr11	31967909	31968253	Id855_agCNE5	0
chr11	31994909	31995566	agCNE4	3
chr11	32003264	32003431	agCNE3	0
chr11	32031078	32031697	E-200_agCNE2	1
chr11	32041336	32041684	agCNE1	3
chr11	32063979	32064758	E-250	1

Supplementary Table 3 Variants called within *PAX6* CREs

Chromosome	Position	ID	Ref	Alt	AC	In gnomAD
chr11	31640570	rs116344813	C	T	1	Y
chr11	31641108	rs7480296	C	T	105	Y
chr11	31649995	rs75155674	A	G	3	Y
chr11	31655378	rs527750251	GAA	G,GA	3,55	Y
chr11	31655412	.	CT	C	1	Y
chr11	31655529	rs141010562	T	C	1	Y
chr11	31655957	rs546016853	T	C	1	Y
chr11	31664450	rs77294851	G	A	1	Y
chr11	31691356	rs2032490	C	T	72	Y
chr11	31713193	rs7926476	T	C	35	Y
chr11	31763753	.	C	CT	1	Y
chr11	31763865	rs113252042	G	A	1	Y
chr11	31765243	.	T	G	1	Y
chr11	31803970	rs694617	T	G	94	Y
chr11	31816401	rs1540318	C	T	13	Y
chr11	31816580	rs4440995	G	A	13	Y
chr11	31816704	rs5790869	C	CG	106	Y
chr11	31823785	rs1805270	G	T	13	Y
chr11	31823821	.	C	G	1	Y
chr11	31824099	rs1806172	A	G	3	Y
chr11	31824409	rs35186763	CT	C,CTT	16,4	Y
chr11	31825436	.	C	CT	2	Y
chr11	31826416	rs370398382	TA	T	1	Y
chr11	31870966	rs662155	C	T	86	Y
chr11	31875104	rs564983219	C	A	1	Y
chr11	31876300	rs72896305	G	A	20	Y
chr11	31876337	rs140066452	C	T	1	Y
chr11	31877325	rs746615173	C	T	1	Y
chr11	31877339	rs11031506	T	C	40	Y
chr11	31887106	.	C	G	1	N
chr11	31887179	rs11031520	G	T	18	Y
chr11	31995087	rs224674	T	G	12	Y
chr11	31995236	rs11031547	T	C	11	Y
chr11	31995530	rs224673	T	C	12	Y
chr11	32031461	rs1002229	A	T	9	Y
chr11	32041474	rs193052496	A	G	1	Y
chr11	32041530	rs224611	T	C	50	Y
chr11	32041683	rs224612	T	C	75	Y
chr11	32064251	rs7128319	G	A	3	Y