Supplementary Tables

for Hall *et al*, Short-read whole genome sequencing identifies causative variants in most individuals with previously unexplained aniridia

Supplementary Tables:

S1: Genomic regions examined on IGV for direct visualisation of SV breakpoints

S2: Nested primers used in RT-PCR splicing analysis

S3: STR profiling

S4: Splicing predictions

S5: CREs examined

S6: CREs examined with frequencies

S7: *De novo* variants

Table S1: Genomic regions examined on IGV for direct visualisation of SV breakpoints

Chromosome	Genomic coordinates (hg38)		Gene / Region	Notes
	Start	End		
chr9	18472082	18912949	ADAMTSL1	
chr9	93847400	93977000	BARX1 and regions translocation breakpoint/del	Not an aniridia gene. Included as was a breakpoint of the t(1,9) translocation in case RPID 356
chr7	129223015	129432211	AHCYL2	Not an aniridia gene. Included as *de novo* 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 S2: Nested primers used in RT-PCR splicing analysis: oligonucleotide sequences

Primer name	Sequence 5' to 3'
Exon 1 to exon 5:	
Green outer F (B1466)	GCCAGTGAGGAGCGGTG
Green outer R (B1468)	CTGCAGAATTCGGGAAATG
Green inner F (B1527)	CACATTAACACACTTGAGCC
Green inner R (B1463)	TGTGAGCTAGCTCTACAATC
Exon 3 to exon 8:	
Orange outer F (B1677)	GCGCAGATGTGTGAGGCC
Orange outer R (B1678)	GTTGGAACTGATGGAGTTGG
Orange inner F (B1675)	GAGAGTGGACAGACATCCG
Orange inner R (B1676)	TCCTTCCTGTTGCTGGCAG
Exon 7 to exon 12:	
Yellow outer F (B1523)	AGCAACAGATGGGCGCAG
Yellow outer R (B1524)	ATTCACCGAAGGGCTGGTG
Yellow inner F (B1525)	GGCATGTATGATAAACTAAGG
Yellow inner R (B1526)	GCAGGAGTATGAGGAGGTC
Exon 9 to 3' UTR:	
Pink outer F (B1681)	GCCCGAGAAAGACTAGCAG
Pink outer R (1682)	CTGAAGCGGCTCTAACAGC
Pink inner F (B1679)	CAAAATAGATCTACCTGAAGC
Pink inner R (B1680)	TGTTGTGTCCCCATAGTCAC

Table S3: Short tandem repeat (STR) genotype profiling of family 2134 confirming 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

Sample Name	Locus	Allele 1	Allele 2	Size 1	Size 2
2134_2134_R45E2	TH01	6	9	161.32	173.26
Proband sample	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	TH01	9	10	173.30	177.30
Parental sample	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	TH01	9	10	154.43	173.27
Parental sample	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: 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

Table S4: Predicted effects on splicing of *PAX6* sequence variants identified in the WGS cohort

CDS variant (PAX6 NM_000280 .4)	**RPID(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 predic-tion**	**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; RPID, research participant identifier; N, natural splice site; SSF, splice site finder
Thresholds: SSF ≥ 70, MaxEnt ≥ 0, NNSPLICE ≥ 0.4, GeneSplicer ≥ 0

Table S5: GRCh38 coordinates of *PAX6 cis*-regulatory elements

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

Table S6: 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

Table S7: *De novo* analysis results of trio whole genomes in the aniridia cohort

Individual (RPID)	Gene	Region	Chr	Genomic coordinates (hg38)	CDS variant	Protein / splicing consequence	CADD/ REVEL score	ACMG/ACGS	Novel	Comment
75	*PRDM5*	Exon 16/16	4	g.120695125C>T	NM_018699.4: c.1879G>A	p.(Val627Ile)	22.6		yes	***De novoPAX6* SV identified**
356	*CRYBG2*	Exon 3/22	1	g.26354223_26354233del	ENST00000475866.3:c .729_739del	p.(Ala244Leufs*4)	n/a		yes	***De novoPAX6* SV identified**
1635	***PAX6***	**IVS8+68**	**11**	**g.31794562G>C**	**NM_000280.4: c.682+68C>G**	**p.(?) / donor gain**	**19.83**	**P (0.994)**	**yes**	**disrupts splicing *in vitro(2)***
	NYNRIN	Exon 9/9	14	g.24416718_24416719dup	NM_025081.3: c.4969_4970dup	p.(Ala1658Argfs*19)	34		no	trumped by *PAX6* variant above
2469	*AHCYL2*	Exon 8/17	7	g.129405125 C>A	NM_015328.4: c.1054C>A	p.(Leu352Met)	25 / 0.73	U (0.5)	yes	interacts with ITPR1
	IGLV3-1	Exon 2/2	22	g.22881377 G>A	ENST00000390319.2:c .327G>A	p.(Trp109*)	35		no	encodes a Bence-Jones protein
	ZBTB11	Exon 2/11	3	g.101671997 G>T	NM_014415.4: c.527C>A	p.(Ser176 Tyr)	25.8		yes	AR intellectual disability gene
	H3C4	Exon 1/1	6	g.26197043 G>T	NM_001376937.1:c.20 8C>A	p.(Arg70Ser)	23		yes	
	ZNF74	Exon 5/5	22	g.20406879_ 20406880ins GGGATCCCA	NM_003426.4: c.1846_1847ins GGGATCCCA	p.(Pro615_Ile616ins ArgAspPro)	n/a		yes	
5645	***PAX6***	**IVS2-2**	**11**	**g.31,806,927del**	**NM_015328.4: c.-128-2del**	**p.(?) / acceptor loss**	**24.4**	**LP (0.9)**	**no (3)**	**canonical splice variant**

Abbreviations: subs, substitution; del, deletion; dup, duplication; ins, insertion; IVS, intervening sequence/intron; CDS, coding sequence; P, pathogenic; U, uncertain; LP, likely pathogenic (all ACMG/ACGS criteria as previously)