



**Table S2. Overview of regions with poor coverage**

	GC%	Length	Complexity / length	Tm	Coverage
Avg. (low cvg. regions)	0.596501	248.8409	0.466871	87.11466	0.7
Standard dev. (low)	0.130444	235.3588	0.105623	6.578099	1.2
Avg. (other regions)	0.500257	252.8994	0.424738	82.36241	297.3
Standard dev. (others)	0.10179	548.7422	0.107763	5.65267	132.8
<b>T-test (low vs others)</b>	<b>1.49E-05</b>	0.913384	<b>0.011909</b>	<b>2.1E-05</b>	0

This table displays the GC content, length, complexity/length, and melting temperature of both the low coverage data set, and the set of all other exons, as well as their standard deviations. Complexity was measured as the difference in storage size of a compressed vs uncompressed probe (using the linux algorithm compress), and as such is an inverse measurement of the sequence complexity of the probe. Tm is the melting temperature of the probe. GC% calculates the percentage of base pairs that are either Guanine or Cytosine. A T-test was performed between these two groups, and statistically significant differences were found in the GC content, complexity/ length, and melting temperature of low coverage probes when compared to their normal counterparts.

**Table S3. List of all low coverage capture regions**

Hg 19 position	Gene	NM#	CDS Exon	GC%	Length	Complexity / length	Tm	Duplicated regions	Coverage
chr1:196658550-196658744	CFH	NM_000186	8	0.410256	195	0.461538	80.54598	Yes	0.0
chr1:196714947-196715129	CFH	NM_000186	21	0.431694	183	0.464481	81.25679	Yes	0.0

chr1:196716241-196716634	CFH	NM_000186	22	0.337563	394	0.555838	78.86064	Yes	0.5
chr2:73826528-73826648	ALMS1	NM_015120	10	0.404959	121	0.413223	78.76065	Yes	0.0
chr2:73830368-73830431	ALMS1	NM_015120	14	0.453125	64	0.328125	77.0552	Yes	0.0
chr4:47972913-47973117	CNGA1	NM_001142564	1	0.507317	205	0.482927	84.65055	Yes	0.0
chr5:178421442-178422124	GRM6	NM_000843	4	0.775988	683	0.633968	97.37303	No	1.3
chr7:33140143-33140173	RP9	NM_203288	2	0.451613	31	0.129032	68.67667	Yes	0.0
chr7:33148833-33149002	RP9	NM_203288	1	0.752941	170	0.5	94.21899	No	0.8
chr8:63998377-63998612	TTPA	NM_000370	1	0.762712	236	0.533898	95.44212	No	3.3
chrX:77359666-77359902	PGK1	NM_000291	1	0.632911	237	0.50211	90.12924	No	0.0
chrX:77365364-77365414	PGK1	NM_000291	2	0.392157	51	0.27451	72.56408	Yes	0.0
chrX:77369513-77369657	PGK1	NM_000291	4	0.531034	145	0.455172	84.61371	Yes	1.2
chrX:77373548-77373667	PGK1	NM_000291	6	0.558333	120	0.391667	85.01457	Yes	0.5
chrX:77380371-77380548	PGK1	NM_000291	9	0.52809	178	0.466292	85.13227	Yes	1.1
chrX:77380824-77380922	PGK1	NM_000291	10	0.565657	99	0.383838	84.43099	Yes	0.7
chrX:77381287-77382324	PGK1	NM_000291	11	0.365125	1038	0.630058	80.77801	Yes	0.0
chrX:153409725-153409869	OPN1LW	NM_020061	1	0.627586	145	0.448276	88.57233	Yes	0.7
chrX:153416128-153416424	OPN1LW	NM_020061	2	0.579125	297	0.525253	88.35018	Yes	0.0
chrX:153421769-153422008	OPN1LW	NM_020061	5	0.541667	240	0.5125	86.41457	Yes	4.9
chrX:153424291-153424507	OPN1LW	NM_020061	6	0.520737	217	0.483871	85.33566	Yes	0.0
chrX:153448085-153448278*	OPN1MW	NM_000513	1	0.634021	194	0.494845	89.7071	Yes	0.5
chrX:153453259-153453555	OPN1MW	NM_000513	2	0.579125	297	0.525253	88.35018	Yes	0.2
chrX:153461421-153462351	OPN1MW	NM_000513	6	0.568206	931	0.61869	89.04897	Yes	0.2
chrX:153485203-153485396*	OPN1MW2	NM_001048181	1	0.634021	194	0.494845	89.7071	Yes	0.7
chrX:153490377-153490673	OPN1MW2	NM_001048181	2	0.579125	297	0.525253	88.35018	Yes	0.4
chrX:153498539-153499469	OPN1MW2	NM_001048181	6	0.568206	931	0.621912	89.04897	Yes	0.0
chr10:85954412-85954571	CDHR1	NM_001171971	1	0.76875	160	0.5	94.68332	No	1.3
chr10:95360928-95360993	RBP4	NM_006744	UTR	0.80303	66	0.363636	91.63806	No	0.0
chr10:124221041-124221640	HTRA1	NM_002775	1	0.795	600	0.64	98.05124	No	0.7
chr11:12695969-12696381	TEAD1	NM_021961	UTR	0.784504	413	0.588378	97.24357	No	1.0
chr11:68080108-68080273	LRP5	NM_002335	1	0.825301	166	0.53012	97.11488	No	0.0

chr14:68168603-68168652	RDH12	NM_152443	UTR	0.58	50	0.26	80.06957	Yes	0.0
chr16:16297267-16297470	ABCC6	NM_001171	7	0.563725	204	0.485294	86.95134	Yes	3.1
chr16:16302585-16302716	ABCC6	NM_001171	6	0.560606	132	0.439394	85.48654	Yes	0.0
chr16:16306042-16306103	ABCC6	NM_001171	5	0.612903	62	0.306452	83.35409	Yes	0.1
chr16:16308181-16308306	ABCC6	NM_001171	4	0.619048	126	0.452381	87.70227	Yes	1.3
chr16:16313411-16313539	ABCC6	NM_001171	3	0.55814	129	0.395349	85.29733	Yes	0.0
chr16:16313679-16313804	ABCC6	NM_001171	2	0.547619	126	0.396825	84.7737	Yes	0.0
chr16:16315506-16315688	ABCC6	NM_001171	1	0.606557	183	0.486339	88.42619	Yes	0.3
chr16:16317256-16317328	ABCC6	NM_001171	UTR	0.739726	73	0.356164	89.76903	Yes	0.0
chr16:53656131-53656268	RPGRIP1L	NM_015272.2	23	0.608696	138	0.427536	87.62291	Yes	0.5
chr17:6459705-6459877	PITPNM3	NM_001165966	1	0.849711	173	0.543353	98.23755	No	0.0
chr17:79503901-79504155	FSCN2	NM_012418	5	0.729412	255	0.513725	94.23467	No	4.5

Table S3 lists all regions of poor coverage, along with their GC content, length, complexity per unit length, predicted melting temperature and coverage. Duplicated regions are defined as having one or more similar genomic loci that have more than 90% identity in sequence and in homologous length, supported by UCSC genome browser BLAT results (<http://genome.ucsc.edu/cgi-bin/hgBlat?command=start>). \*: The two regions have the same sequence.

**Table S4. *In silico* prediction of novel missense mutations**

Gene	Patient ID	Mutations	PhyloP		SIFT		Polyphen2		LRT	MutationTaster
			Score	PhyloP	Score	SIFT	Score	Polyphen2		
<i>ALMS1</i>	3779	c.9764C>G, p.S3255C	0.98	C	1.00	D	0.48	P	N	N
<i>BBS7</i>	1327	c.728G>A, p.C243Y	1.00	C	1.00	D	1.00	D	D	D
<i>CRB1</i>	3319	c.1439G>C, p.C480S	1.00	C	1.00	D	1.00	D	na	D
<i>GUCY2D</i>	1272	c.1933T>C, p.S645P	1.00	C	1.00	D	1.00	D	D	D
<i>GUCY2D</i>	1272	c.2207T>G, p.M736R	1.00	C	0.99	D	1.00	D	N	D
<i>GUCY2D</i>	3611	c.2132C>T, p.P711L	1.00	C	1.00	D	1.00	D	D	D
<i>GUCY2D</i>	3725	c.2678C>T, p.S893F	1.00	C	1.00	D	1.00	D	D	D
<i>GUCY2D</i>	3799	c.743C>G, p.S248W	1.00	C	1.00	D	1.00	D	D	D
<i>INPP5E</i>	3773	c.1861C>T, p.R621W	1.00	C	1.00	D	1.00	D	D	D

<i>PDE6A</i>	1313	c.2333A>T, p.D778V	1.00	C	1.00	D	1.00	D	D	D
<i>RDH12</i>	3740	c.692G>A, p.G231D	0.99	C	1.00	D	1.00	D	D	D
<i>SNRNP200</i>	3795	c.3133C>A, p.P1045T	1.00	C	1.00	D	1.00	D	D	D
<i>TULP1</i>	1268	c.1277C>T, p.P426L	1.00	C	1.00	D	1.00	D	D	D
<i>TULP1</i>	1268	c.1518C>A, p.F506L	0.75	N	1.00	D	1.00	D	D	D
<i>TULP1</i>	3681	c.1064A>T, p.D355V	1.00	C	0.99	D	0.99	D	D	D
<i>TULP1</i>	3771	c.961T>G, p.Y321D	1.00	C	1.00	D	1.00	D	D	D

#The prediction score and results are generated from PhyloP, SIFT, Polyphen2, LRT, and MutationTaster, and compiled by dbNSFP.

C: conserved; D: damaging; N: neutral; P, possibly damaging; na: not available. When dbNSFP don't have pre-computed scores for specific mutation, we searched the online server provided by each program to get the original prediction score, then convert the score to dbNSFP score according to dbNSFP standard.

**Table S5. Prescreening information for subset of our patient cohort**

Patient ID	LCA		arRP		RPGRIP1	AIPL1	RPE65	CRX	CRB1	GUCY2D	Solved by NGS: known LCA gene	Why prescreening failed
	APEX array	DATE	APEX array	DATE								
54	yes	Oct-03			yes	yes	yes	yes	yes	yes	yes	A
208	yes	Oct-03			yes	yes	yes	yes	yes	yes		
251					yes				yes	yes		
285	yes	Oct-03			yes	yes	yes	yes	yes			
291	yes	Oct-03			yes	yes	yes	yes	yes			
361						yes	yes	yes	yes			
393						yes	yes	yes	yes	yes	yes	B
398					yes	yes	yes	yes	yes	yes	yes	B
408	yes	Oct-03	Apr-08		yes	yes		yes	yes	yes		
418	yes	Oct-03			yes	yes		yes	yes	yes	yes	A, B
432	yes	Oct-03			yes	yes	yes	yes	yes			
486	yes	Oct-03			yes	yes	yes	yes	yes			
489												
512	yes	Oct-03	Aug-07		yes	yes	yes	yes	yes			

518	yes	Oct-03								
617	yes	Oct-03			yes	yes	yes	yes	yes	
622	yes	Oct-03			yes	yes	yes	yes	yes	yes
635					yes	yes	yes	yes	yes	
647						yes	yes	yes	yes	
690	yes	Oct-03	yes	Jun-05	yes	yes	yes	yes	yes	
704	yes	Oct-03				yes	yes	yes	yes	
787					yes	yes	yes	yes	yes	
789						yes	yes	yes	yes	
1171						yes	yes	yes	yes	
1174					yes	yes	yes	yes	yes	
1248			yes	Apr-05	yes	yes	yes	yes	yes	
1250					yes	yes	yes	yes	yes	
1251					yes	yes	yes	yes	yes	yes
1253					yes	yes	yes	yes	yes	
1255					yes	yes	yes	yes	yes	
1257					yes	yes	yes	yes	yes	
1259					yes	yes	yes	yes	yes	yes
1262					yes	yes	yes	yes	yes	
1268					yes	yes	yes	yes	yes	yes
1269					yes	yes	yes	yes	yes	
1271					yes	yes	yes	yes	yes	yes
1272					yes	yes	yes	yes	yes	yes
1273					yes	yes	yes	yes	yes	
1275					yes	yes	yes	yes	yes	
1278					yes	yes	yes	yes	yes	yes
1290						yes	yes	yes	yes	
1303						yes	yes	yes	yes	yes
1311					yes	yes	yes	yes	yes	
1312					yes	yes	yes	yes	yes	
1313					yes	yes	yes	yes	yes	

1314			yes	yes	yes	yes	yes		
1315			yes	yes	yes	yes	yes	yes	C
1318			yes	yes	yes	yes	yes		
1327			yes	yes	yes	yes	yes		
1331			yes	yes	yes	yes	yes		
1332			yes	yes	yes	yes	yes		
1379				yes	yes	yes	yes		
1380				yes	yes	yes	yes		
1381				yes	yes	yes	yes		
1413					yes	yes	yes	yes	C
1472					yes	yes	yes		
1473					yes	yes	yes	yes	C
1842									
1896									
3062									
3075									
3256									
3294	yes	Apr-08							
3311	yes	Apr-08							
3319	yes	Apr-08						yes	A
3425									
3443	yes	Apr-09							
3494									
3550	yes	Jun-10							
3551									
3557								yes	C
3561	yes	Apr-09						yes	A
3577								yes	C
3582									
3596	yes	Apr-09							
3611								yes	C

3628							
3698	yes	Jun-10					
3719							
3722					yes		C
3725					yes		C
3916					yes		C
3973							
3985	yes	Jun-10					
4012							
4013							
4016							
4019					yes		C

Columns in table S5: LCA APEX array: this indicates whether or not the sample had been screened by LCA APEX array; arRP APEX array: this indicates whether or not the sample had been screened by autosomal recessive RP APEX array; Date: the date when the sample was screened by the array; Six columns with gene names: this indicates whether or not the sample had been Sanger sequenced for the frequently mutated exons in the corresponding gene; Solved by NGS: known LCA gene: this indicates whether or not mutations in known LCA genes were identified by our targeted NGS method in the sample; Why prescreening failed: this is to explain why the mutations in known LCA genes had not been captured by the prescreening (A: the mutations had not been covered by the array; B: the mutant exon had not been covered by Sanger sequencing, only the frequently mutated exons in the corresponding gene were Sanger sequenced; C: the sample had neither been screened by LCA APEX array nor been Sanger sequenced for the corresponding genes identified in the sample).