Supplemental Information.

Supplemental Text:

Postulating a molecular structural basis for *ABCA4* allele severity

We used a topological model of the ABCR molecule and three dimensional homology models of its two nucleotide binding domains (Suppl. Fig.2) to better understand possible relationships between disease severity and estimates of biochemical abnormalities or molecular structural aberrations. Missense mutations are located throughout the protein including two exocytoplasmic domains (ECD-1 and ECD-2), two nucleotide binding domains (NBD-1 and NBD-2), and transmembrane regions.

Four severe mutations (C54Y, R220C, L244P, and R602W) were located in the ECD-1 domain (Supp. Fig.2B). Replacement of a Cys residue at position 54 with a bulky Tyr may result in structural rearrangements due to sterical clashes with other amino acids. Moreover, it has been demonstrated that ECD-1 and ECD-2 are linked through one or more disulfide bridges (1) and a mutation at Cys54 may carry a destabilizing effect. Likewise, the R220C mutation might induce cross link products with native Cys residues producing misfolded protein that might even fail to insert properly into membranes. A Pro residue instead of a Leu at position 244 may result in the disruption of secondary structure elements. Leu residues have a very high propensity to form α-helices and can also participate in β-strands. In contrast, Pro residues are unsuited to these structural elements owing to their rigid cyclic structure and inability to donate the amide proton. Finally, the R602W mutation replaces a flexible positively charged Arg with a bulky and hydrophobic Trp. In addition to sterical clashes, this substitution may prevent formation of hydrogen bonds and electrostatic interactions typical of Arg residues and reduce the solubility of the protein as a result of placing a hydrophobic residue on the protein surface where Arg residues commonly reside. ECD-1 also contains two milder mutations (D600E, and D654N) representing more conservative amino acid substitutions. Relatively mild disease caused by the non-conservative S100P substitution may be explained by the fact that both of these residues are often located in loops.

NBD-1 includes several mutations. The severe E1122K mutation replaces the negatively charged Glu with a longer and positively charged Lys that can disrupt the

interactions of this residue with Walker motif B (Suppl. Fig.2C, left) and induce structural changes as a result of electrostatic repulsion. Two surface-located Arg residues of NBD-1 at positions 1098 and 1108 are replaced with smaller Cys residues, producing intermediate disease severity. Conservative changes at residues 965 and 1087 located in Walker motifs A and B, respectively (Suppl. Fig.2C, left), may influence nucleotide binding and result in intermediate disease, whereas the change at residue 1201 results in much milder disease presumably because the location is outside of the enzymatically active part of NBD-1. The complex mutation L541P;A1038V causes severe *ABCA4* disease and has components in both ECD-1 and in NBD-1. The Leu to Pro substitution at residue 541 may have the same negative effects as L244P found in ECD-1 (see above). Replacement of Ala at position 1038 with a Val may induce unfavorable interactions between the side chain of the Val and the tightly packed neighboring residues (Suppl. Fig.2C, left).

Two milder substitutions within ECD-2, V1433I and T1526M, introduce similar residues with respect to charge and solubility, although they differ somewhat in size. The two remaining mutations are more invasive, as reflected by the greater disease severity they cause: a neutral Ala residue at position 1598 is replaced with a much larger acidic Asp residue, while a basic Arg at position 1640 is mutated to a neutral Gln.

NBD-2 contains two severe mutations and five milder mutations. According to the homology model, the Cys at 2150 resides in a β-strand and is accessible to solvent (Suppl. Fig.2C, right). A Tyr at this position may interfere with the correct folding of the neighboring α-helix (residues 2154-2161) and the adjacent loop (residues 2162-2169). The segment of 80 amino acids starting from residue 2158 is highly conserved in ABC transporters and may carry a regulatory function (2). Thus, structural alterations in this region might lead to various aberrations. Also, a solvent-exposed hydrophobic Tyr residue is energetically unfavorable and may provoke protein aggregation. The second mutation causing severe disease, L1940P, may have the same negative effects as the Leu-to-Pro substitutions found in ECD-1 (see above). G1961E and G1961R are two of the mildest mutations included in this study. Both Glu and Arg residues are hydrophilic, suggesting that residue 1961 may be located on the protein surface, where the oppositely charged side chains of these large amino acids face the solvent and,

accordingly, do not create significant structural disturbances. The same conclusion may hold for the R2030Q mutation. The two remaining mild mutations, L2027F and K2172R, replace residues with similar sizes and properties.

Two mutations (L11P and R18W) are located in the short amino-terminal region preceding the first transmembrane α-helix. Relative mildness of the disease caused by these non-conservative mutations suggests that this region of the ABCR molecule does not significantly influence its structure and/or function. The transmembrane regions harbor three mutations. The mildest of these, G818E, is located in the loop connecting transmembrane helices 5 and 6. Such loops are often poorly ordered, which may explain the mild disease associated with this mutation. V1896D, in contrast is a severe mutation located at the cytoplasmic end of helix 12. Introduction of an essentially hydrophilic Asp residue at this position may destabilize the helix and interfere with correct insertion in the membrane. A mutation causing intermediate disease severity, P1380L, is located in the middle of trans-membrane helix 7. From the structures of various membrane-spanning proteins it is well-known that Pro residues often bend helices, which may be important for biological function. Replacing the Pro with a Leu residue may therefore influence the transport of the ABCR substrate across the membrane.

Supplemental References

1. Bungert, S., Molday, L. L. and Molday, R. S. (2001) Membrane topology of the ATP binding cassette transporter ABCR and its relationship to ABC1 and related ABCA transporters: identification of N-linked glycosylation sites. *J. Biol. Chem.*, **276**, 23539-23546.176. Lamb, T.D. and Pugh EN Jr. (2004) Dark adaptation and the retinoid cycle of vision. *Prog. Retin. Eye Res.*, **23**, 307-380.
2. Peelman, F., Labeur, C., Vanloo, B., Roosbeek, S., Devaud, C., Duverger, N., Denèfle, P., Rosier, M., Vandekerckhove, J. and Rosseneu, M. (2003) Characterization of the ABCA transporter subfamily: identification of prokaryotic and eukaryotic members, phylogeny and topology. *J. Mol. Biol.*, **325**, 259-274.
3. Papermaster, D.S., Reilly, P. and Schneider, B.G. (1982) Cone lamellae and red and green rod outer segment disks contain a large intrinsic membrane protein on their margins: an ultrastructural immunocytochemical study of frog retinas. *Vision Res.*, **22**, 1417-1428.

4. Molday, L.L., Rabin, A.R. and Molday, R.S. (2000) ABCR expression in foveal cone photoreceptors and its role in Stargardt macular dystrophy. *Nat. Genet.*, **25**, 257-258.
5. Lamb, T.D. and Pugh EN Jr. (2004) Dark adaptation and the retinoid cycle of vision. *Prog. Retin. Eye Res.*, **23**, 307-380.
6. Nickell, S., Park, P.S., Baumeister, W. and Palczewski, K. (2007) Three-dimensional architecture of murine rod outer segments determined by cryoelectron tomography. *J. Cell Biol.*, **177**, 917-925.

Supplemental Figure 1: Retina-wide disease severity evaluated with standardized (ISCEV) rod and cone electroretinogram (ERG) amplitude parameters as a function of age. (A,B) Rod and cone ERG b-wave amplitude in individuals with *ABCA4* mutations as a function of age. The lower limit (mean-3SD) of normal is shown (horizontal gray line). (C,D) Plot of ERG amplitudes against time after age of disease initiation (ADI) estimated from perimetry data shows a dramatic reduction in variability. Dashed lines represent the normal variability (+/-3SD) extended along the respective exponentials for rod and cone ERG amplitudes.

Supplemental Figure 2: Localization of the ABCR protein in the retina and models of its structure illustrating the locations and severities of the missense mutations. (A) Schematic model of the interface between photoreceptor outer segments (OS) and RPE cells, and location of the ABCR protein at the rims of OS disk membranes. Drawings are based on Refs. 3-6. Scales are approximate. PM, plasma membrane of the OS. (B) Topological model of ABCR illustrating putative transmembrane helices (red cylinders), exocytoplasmic (ECD) and nucleotide binding (NBD) domains (loops). Ellipses encompass the regions of the primary structure that were used for homology modeling of NBD-1 and NBD-2 based on high sequence similarity with ATP-binding domains of other ABC transporters (see Methods). Residues with disease-causing missense mutations are shown (blue circles) and supplemented with mutations and severity estimates (red numbers, greater severity than truncation; blue numbers, less severity than truncation). Overall severity of the complex allele L541P;A1038V is shown duplicated for each allele. Potential N-linked glycosylation sites identified by mutagenesis (1) are shown with green octagons. A putative disulfide bridge between ECD-1 and ECD-2 (1) is shown as two linked green circles. (C) Homology models of the three dimensional structure for NBD-1 (left) and NBD-2 (right) based on the crystal structure of the ABC transporter ATP-binding protein from Thermotoga maritima (PDB ID 1VPL). Walker A and B motifs responsible for ATP binding are colored red and blue, respectively. Protein residues with disease-causing mutations are shown with sticks.

Supplemental Table 1. ABCA4 mutations in the patients studied

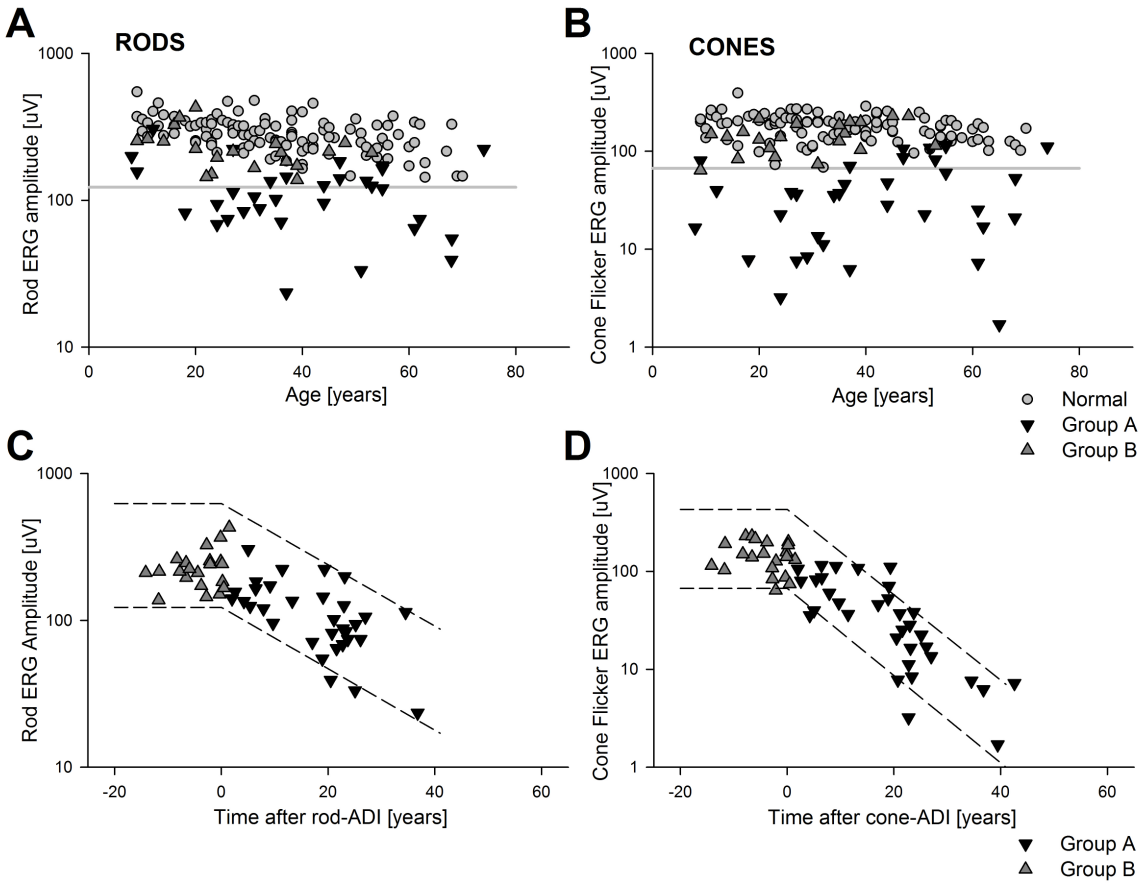
Patient No.	Family No.	Age[a]/Gender	Follow-up [yrs]	Diagnosis[b]	Allele 1[c]	Allele 2[c]	Segregation	Age of Disease Initiation[d] [yrs]
1	1	64/F	4	CRD	L11P	L2027F		47
2	2	53/M	6	STGD	R18W	D600E		48
3	3	19/M	8	CRD	W41X	R1098C	Y	16
4	4	37/F		STGD	C54Y	G863A	Y	>37
5	5	19/F		STGD	C54Y	R1129C	Y	>18
6	6	9/M	6	STGD	C54Y	IVS35+2 T>C	Y	6
7	7	37/M	3	STGD	C54Y	G1961E		>40
8	8	20/F		STGD	A60V	V931M	Y	>20
9	9	55/M	3	CRD	S100P	L1201R		47
10	10	11/F	18	STGD	R152X	IVS38-10 T>C	Y	0
11	11	17/M	4	STGD	R152X	G1961E	Y	>21
12	12	19/F		STGD	R152X	G1961E		>19
13	13	23/M		STGD	A192T	L1525P	Y	>23
14	14	61/F		STGD	R220C	R220C		18
15	15	27/F	20	CRD	L244P	L244P	Y[e]	6
16		19/F	18	CRD	L244P	L244P	Y[e]	1
17		27/M		CRD	L244P	L244P	Y[e]	0
18		18/F		CRD	L244P	L244P	Y[e]	0
19	16	32/M	18	CRD	Y245X	Y245X	Y	9
20		47/M		CRD	Y245X	Y245X	Y	12
21	17	9/M	2	CRD	V256V	T1526M	Y	>11
22	18	43/F	5	STGD	V256V	G1961E	Y	>48
23		39/F	4	STGD	V256V	G1961E	Y	>43
24	19	37/M		CRD	Y362X	IVS38-10 T>C		0
25	20	12/F	5	STGD	R602W	V1896D		7
26	21	63/F		STGD	D654N	W663X		36
27	22	35/F	3	STGD	R681X	P309R	Y	>37
28	23	9/M		CRD	R681X	C2150Y	Y	0
29	24	30/M		STGD	G863A	R1108C	Y	>30
30	25	27/F	12	STGD	G863A	C1490Y	Y	>39
31	26	37/M		STGD	G863A	N1799D	Y	>37
32	27	32/F	2	STGD	V935A	IVS40+5 G>A	Y	>34
33	28	46/F	5	STGD	N965S	N965S		26
34	29	24/M	7	STGD	T1019M	G1961E	Y	>31
35	30	35/M		STGD	A1038V;L541P	G818E	Y	14
36	31	11/F	9	STGD	A1038V;L541P	N965S	Y	>19
37	32	31/M		CRD	A1038V;L541P	P1380L		4
38	33	32/F		STGD	A1038V	P1380L	Y	>32
39	34	26/F		STGD	A1038V;L541P	IVS39+6 T>C	Y	>26
40	35	17/F		STGD	E1087K	G1961E	Y	>17
41	36	44/F		CRD	R1108C	IVS38-10 T>C	Y	21
42	37	53/F	15	STGD	R1108C	IVS40+5 G>A	Y	>67
43	38	68/F		CRD	R1108C	IVS40+5 G>A		49
44	39	47/M	12	STGD	E1122K	G1961E	Y	45
45	40	65/M		CRD	P1380L	P1380L	Y[e]	26
46	41	38/F		STGD	P1380L	V1686M		>38
47	42	45/F	7	STGD	P1380L	G1961E	Y	>53
48	43	55/F	7	STGD	P1380L	K2172R	Y	46
49	44	14/F		STGD	W1408R	IVS40+5 G>A	Y	>14
50	45	44/M	6	STGD	V1433I	L2027F	Y	34
51	46	35/M		STGD	P1511 ins1ccgC	G1961E		>35
52	47	22/M	3	STGD	P1511 ins1ccgC	R1705Q	Y	>25
53	48	52/F		STGD	T1526M	R2030Q	Y	39
54		47/F	6	STGD	T1526M	R2030Q	Y	41
55		39/F	12	STGD	T1526M	R2030Q	Y	>51
56	49	36/F		STGD	A1598D	R1640Q	Y	19
57		37/F		STGD	A1598D	R1640Q	Y	18
58	50	44/F	11	STGD	R1640Q	G1961E	Y	>55
59	51	26/F	15	CRD	IVS38-10 T>C	E1087D	Y	2
60	52[f]	31/M		STGD	L1940P	IVS40+5 G>A	Y	>31
61		34/M	20	STGD	L1940P	IVS40+5 G>A	Y	30
62		64/M		STGD	L1940P	IVS40+5 G>A	Y	43
63		61/F	13	STGD	L1940P	IVS40+5 G>A	Y	39
64		20/F	6	STGD	L1940P	R1129L	Y	>26
65	53	74/F	7	STGD	G1961R	A1739 del11[g]	Y	55
66	54	50/F	11	STGD	G1961E	M669 del2tccAT	Y	49

[a]Age at first visit.
[b]Clinical diagnosis of Stargardt disease (STGD) or cone-rod dystrophy (CRD)
[c]Nucleotide numerations and intron/exon boundaries refer to Genbank NM_000350.2
[d]Earlier of the rod or cone ADI; group B patients are shown with the greater than (>) symbol; negative ages are truncated at zero.
[e]Unaffected heterozygotic individuals were not available for molecular studies. However segregation of homozygous alleles with disease was demonstrated in 4 affected siblings in Family 15 and in 2 affected siblings in Family 40.
[f]Three distinct alleles segregated with disease in members of a large extended pedigree.
[g]A1739 del11gcTGGGCTGGTGG

Supplemental Table 2. Common ABCA4 alleles

Allele	Allele Frequency [Number of Families]	Second Allele (Family)
G1961E	11	C54Y (F7); R152X (F11,F12); V256V (F18); M669 del2tccAT (F54); T1019M (F29) E1087K (F35); E1122K (F39); P1380L (F42); P1511 ins1ccgC (F46); R1640Q (F50)
P1380L	6	A1038V;L541P (F32); A1038V (F33); P1380L (F40); V1686M (F41); G1961E (F42); K2172R (F43)
IVS40+5 G>A	5	V935A (F27); R1108C (F37,F38); W1408R (F44); L1940P (F52)
A1038V;L541P	4	G818E (F30); N965S (F31); P1380L (F33); IVS39+6 T>C (F34)
C54Y	4	G863A (F4); R1129C (F5); IVS35+2 T>C (F6); G1961E (F7)
G863A	4	C54Y (F4); R1108C (F24); C1490Y (F25); N1799D (F26)
R1108C	4	G863A (F24); IVS38-10 T>C (F36); IVS40+5 G>A (F37,F38)
IVS38-10 T>C	4	R152X (F10); Y362X (F19); E1087D (F51); R1108C (F36)
R152X	3	IVS38-10 T>C (F10); G1961E (F11,F12)
V256V	2	T1526M (F17); G1961E (F18)
R681X	2	P309R (F22); C2150Y (F23)
N965S	2	N965S (F28); A1038V;L541P (F31)
P1511 ins1ccgC	2	G1961E (F46); R1705Q (F47)
T1526M	2	V256V (F17); R2030Q (F48)
R1640Q	2	A1598D (F49); G1961E (F50)
L2027F	2	L11P (F1); V1433I (F45)

Supplemental Figure 1



Supplemental Figure 2