Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Baratz KH, Tosakulwong N, Ryu E, et al. E2-2 protein and Fuchs's corneal dystrophy. N Engl J Med 2010;363:1016-24. DOI: 10.1056/NEJMoa1007064.

Supplementary Appendix

Table of contents

Supplemental figures

Supplemental Text

Additional subject information and grading

The Krachmer grading scale1 was used to grade the severity of guttae. The scale is as follows: Grade 0 (none), 1 (1-12 central guttae), 2 (> 12 non confluent guttae), 3 (1-2 mm of confluent central guttae), 4 (> 2-5 mm of central confluent guttae), 5 (> 5 mm of central confluent guttae), 6 (grade confluent guttae with stromal or epithelial edema). Increasing Krachmer grade corresponds to worse disease. The scale grades the number and density of clinically visible guttae using specular biomicroscopy and a slit-lamp.

There were 25 cases with Krachmer grade 1, 47 cases with Krachmer grade 2, 17 cases with Krachmer grade 3, 20 cases with Krachmer grade 4, 9 cases with Krachmer grade 5, and 70 cases with Krachmer grade 6 for a total of 188 cases with prospective grading of corneal guttae. The remaining cases were selected from retrospective grading and could not be assigned a exact Krachmer grade; most of these were probably grades 2-5.

Ancestry was obtained by asking subjects which of the following categories best described their race or heritage: Native American/Alaskan, Asian, Native Hawaiian/Pacific Islander, Black or African American, White, More than one, Unknown/Not reported

Population stratification and Genomic inflation

There was minimal population stratification (Figure S1). A quartile-quartile plot shows the absence of genomic inflation (Figure S2).

Replication studies

SNPs from the 11 regions most highly associated loci in the discovery genome-wide association study (GWAS) were selected for replication studies based on a p-value less than 10^{-4} or localization within a candidate gene for FCD (*AJAP1, SLC8A1, CTFR, ACCN4).* Table S1 shows detailed information on these 17 SNPs and Table S2 shows the association with FCD in the discovery and replication groups. Only one locus other than *TCF4* appeared to replicate with association in the same direction. This locus was *PTPRG* and the association is shown in Tables S3 and S4; the association is suggestive and further work is required because it did not reach genome wide significance.

Linkage disequilibrium

There is extensive linkage disequilibrium across the *TCF4* locus (Figures S3 and S4).

Independence of SNPs genotyped on both the discovery and replication groups on FCD risk

Table S5 shows the four SNPs across the *TCF4* locus that were independently associated with FCD.

Construction of the risk model

The risk model was based on the 4 SNP's in Table S5 independently associated with FCD. These four genotyped SNPs (rs17595731, rs613872, rs9954153, and rs2286812) were used first to construct a risk model using the discovery set. The model was based on multiple logistic regression analysis with additive linear effects for each of the 4 SNPs. This model estimates the logit of the probability of being a case as a weighted linear combination of the number of copies of each of the risk alleles. The output of the model is the resulting estimated probability of being a case.

Method for calculating concordance statistic (c-statistic)

To calculate the concordance statistic (c-statistic) for our matched case-control data, we first stratified on matched sets. A matched set refers to the discovery case matched to two controls. Within each set we considered both case-control pairs, and we counted the number of such pairs in which the probability of being a case as predicted by the model was greater for the case than it was for the matched control. Such a pair is defined to be "concordant". We added up the number of concordant pairs across all matched sets, and divided by the total number of case-control pairs. In the case of ties in the model-predicted probabilities (i.e., the probability of being a "case" was the same for both the case and control), we added 0.5 to the numerator count since the chance of concordance is 50%, and 1 to the denominator count. This calculation conforms to the classical concept of the c-statistic, but is adapted to our setting of matched data. The same concept was used for the Replication sample, except that only 1 case-control pair was available per matched set.

Sliding SNP haplotype studies

The best estimate for the genotype of each of the 720 imputed SNPs were imported into haplo.stats and haplotypes formed by the first 3-SNPs across the locus estimated and the global haplotype association level of significance with FCD recorded. The 3-SNP window was shifted by 1 SNP to the right and the process repeated. The result is shown in Figure S5
We and others have found that conditional analyses such as in Figure S4 can provide a parsimonious representation of statistical risk, but may not fully capture the independent association of specific haplotypes later demonstrated using functional studies.[2] Therefore, we inspected the association between individual haplotypes across the *TCF4* locus and FCD. Six of 46 haplotype blocks across this region contained 7 individual haplotypes associated with FCD with a p-value less than 10^{-6} (Figure 4 and Figure S5).

Individual haplotype studies

The best estimate for the genotype of each of the 720 imputed SNPs were imported into Haploview. The haplotype blocks were estimated using confidence intervals as implemented in Haploview. The haplotype blocks were inspected for haplotypes highly associated ($P < 10^{-6}$), and six haplotype blocks with 7 haplotypes were identified (Table S6). The genotype data for SNPs within the six blocks were imported into haplo.stats and the best estimates of the haplotypes for each individual were determined for each of the six haplotype blocks. The 7 highly associated haplotypes were coded 0, 1, or 2 for the number of copies of each haplotype and a correlation matrix created using Pearson correlation coefficients (Table S6). Table S8 summarizes the physical locations of the 6 haplotype blocks highly associated with FCD across the FCD locus. The correlation between these 7 haplotypes showed that at least two regions were independent and probably a third. These results provide further support for multiple independent regions of the *TCF4* locus contributing to FCD.

Population attributable fraction

The population attributable fraction (or "risk") describes the proportion of disease that would be eliminated by removal of a risk factor from a population; it does not describe the proportion of cases with the risk factor, account for the effect of other risks, or interactions between risks. In spite of these limitations, many readers appreciate knowing the impact of a genetic risk in a population. To this end we provide in the next paragraph the population attributable fraction (PAF) for most significant SNP across the TCF4 locus (rs613872). Please note that these calculations may or may not reflect the underlying complexity of the multiple independent haplotypes that appear to underlie the association between FCD and the TCF4 locus.

The population attributable fraction (PAF) for the risk genotypes (SNP 613872 G/T and G/G) was calculated using the formula: p*(1-1/RR), where p is the risk genotype frequency in cases and RR is the relative risk, which can be considered similar to an odds ratio due to the low

prevalence of FCD. The 95% confidence interval was calculated by applying the delta method, after taking into account the correlation between the risk genotype frequency and the odds ratio. We used 1000 bootstrap samples to estimate the correlation. The PAF estimate and calculation of its 95% confidence interval was based on the method for the replication sample (1:1 matching) providing the results very close to the approach taken by Kuritz SJ and Landis JR (Attributable risk ratio estimation from matched-pairs case-control data. American Journal of Epidemiology 1987; 125: 324-328). The PAF for the discovery subjects was 54% (95% CI = 44.4%, 66.6%), for the replication subjects 68% (95% CI = 58%. 79.7%), and for the combined subject groups 61.0% (95% CI = 53.7%, 69.3%).

Expression of *TCF4* in corneal endothelium

The Allen Brain Map (Allen Developing Mouse Brain Atlas [Internet]. Seattle (WA): Allen Institute for Brain Science. ©2009. Available from: http://developingmouse.brain-map.org.) contains images of *in situ* hybridization of a large number of genes in embryos from C57BL/6J mice . The coronal sections show the mouse eye and corneal endothelium. Selected images downloaded from the website are shown in Figure S6 demonstrating expression of *TCF4* in the developing mouse corneal endothelium.

Supplemental Tables

Table S1. Detailed information of the 17 SNPs selected for replication.

SNP	Chromosome	Position	Locus	Function
rs6668002	1	4581964	*AJAP1*	flanking_5UTR
rs9309056	2	40636910	*SLC8A1*	flanking_5UTR
rs746233	2	220092370	*ACCN4*	intron
rs7640737	3	62008393	*PTPRG*	intron
rs10490775	3	62011764	*PTPRG*	intron
rs10484323	6	6217185	*F13A1*	intron
rs2057181	6	153395904	*RGS17*	intron
rs13221882	7	117045502	*CFTR*	intron
rs10492188	12	16565978	*LOC121520*	flanking_3UTR
rs2445770	15	49424450	*GLDN*	intron
rs2446426	15	49428448	*GLDN*	intron
rs7238810	18	37716013	*PIK3C3*	flanking_5UTR
rs17595731	18	51258450	*TCF4*	intron
rs613872	18	51361300	*TCF4*	intron
rs618869	18	51399149	*TCF4*	intron
rs9954153	18	51689255	*FLJ45743**	flanking_3UTR
rs2286812	18	51868462	*FLJ45743**	intron

*Note that this hypothetical transcript does not contain an open reading frame and is unlikely to represent a protein. This region is directly upstream of *TCF4* and is considered the same locus as *TCF4*. See main article for additional information.

Table S2. Genotype distribution of the 17 SNPs selected for replication. The association with FCD in both the discovery group and the independent replication group is shown. Table 1 in the main text shows the demographics of the two subject groups.

Subject group	SNP	Minor allele frequency	Hardy Weinberg equilibrium	Genotype	Number (Frequency) Cases	Number (Frequency) Controls	Log-additive P-value	Odds Ratios (95% CI)
Discovery	rs6668002	0.39	0.74	CC	65 (0.50)	79 (0.30)	$6.72x10^{-5}$	0.51(0.36-0.71)
				CT	55 (0.43)	131 (0.50)		
				TT	9 (0.07)	50 (0.19)		
	rs9309056	0.10	0.20	TT	91 (0.70)	222 (0.85)	$2.01x10^{-4}$	2.57(1.56-4.23)
				CT	36 (0.28)	38 (0.15)		
				CC	3 (0.02)	0 (0.00)		
	rs746233	0.45	0.84	GG	26 (0.20)	90 (0.35)	$3.51x10^{-3}$	1.60(1.17-2.18)
				AG	70 (0.54)	124 (0.48)		
				AA	33 (0.26)	45 (0.17)		
	rs7640737	0.13	0.32	CC	83 (0.64)	211 (0.81)	$9.92x10^{-5}$	2.32(1.52-3.53)
				CT	40 (0.31)	48 (0.18)		
				TT	7 (0.05)	1 (0.00)		
	rs10490775	0.13	0.32	GG	83 (0.64)	211 (0.81)	$9.92x10^{-5}$	2.32(1.52-3.53)

				AG	40 (0.31)	48 (0.18)		
				AA	7 (0.05)	1 (0.00)		
	rs10484323	0.07	0.37	TT	96 (0.74)	238 (0.92)	$1.66x10^{-5}$	3.85(2.09-7.12)
				CT	30 (0.23)	20 (0.08)		
				CC	3 (0.02)	1 (0.00)		
	rs2057181	0.37	0.33	CC	34 (0.26)	115 (0.44)	$8.41x10^{-5}$	1.92(1.39-2.65)
				CT	69 (0.53)	121 (0.47)		
				TT	27 (0.21)	24 (0.09)		
	rs13221882	0.06	1.00	AA	105 (0.81)	239 (0.92)	$1.20x10^{-3}$	2.84(1.51-5.34)
				AG	24 (0.18)	21 (0.08)		
				GG	1 (0.01)	0 (0.00)		
	rs10492188	0.29	0.30	CC	48 (0.37)	147 (0.57)	$1.65x10^{-5}$	2.15(1.52-3.04)
				CT	61 (0.47)	101 (0.39)		
				TT	21 (0.16)	12 (0.05)		
	rs2445770	0.06	1	GG	98 (0.75)	243 (0.93)	$5.65x10^{-6}$	4.51(2.35-8.64)
				AG	32 (0.25)	17 (0.07)		
				AA	0 (0.00)	0 (0.00)		
	rs2446426	0.06	1	CC	98 (0.75)	243 (0.93)	$5.65x10^{-6}$	4.51(2.35-8.64)

				CT	32 (0.25)	17 (0.07)		
				TT	0 (0.00)	0 (0.00)		
	rs7238810	0.44	0.67	AA	24 (0.18)	95 (0.37)	$9.98x10^{-5}$	1.90(1.38-2.63)
				AG	72 (0.55)	127 (0.49)		
				GG	34 (0.26)	38 (0.15)		
	rs17595731	0.05	1.00	GG	102 (0.78)	249 (0.96)	$2.97x10^{-6}$	7.21(3.15-16.51)
				CG	27 (0.21)	10 (0.04)		
				CC	1 (0.01)	0 (0.00)		
	rs613872	0.22	0.38	TT	46 (0.35)	193 (0.74)	$4.25x10^{-10}$	4.22(2.69-6.63)
				GT	73 (0.56)	60 (0.23)		
				GG	11 (0.08)	7 (0.03)		
	rs618869	0.18	0.65	CC	65 (0.50)	199 (0.77)	$1.64x10^{-6}$	2.90(1.88-4.48)
				CT	59 (0.45)	56 (0.22)		
				TT	6 (0.05)	5 (0.02)		
	rs9954153	0.21	0.80	TT	56 (0.43)	181 (0.70)	$2.18x10^{-6}$	2.79(1.82-4.26)
				GT	65 (0.50)	70 (0.27)		
				GG	8 (0.06)	6 (0.02)		
	rs2286812	0.15	0.18	CC	73 (0.56)	204 (0.78)	$4.13x10^{-6}$	2.80(1.81-4.34)

				CT	52 (0.40)	55 (0.21)		
				TT	5 (0.04)	1 (0.00)		
Replication	rs6668002	0.39	0.74	CC	53 (0.36)	60 (0.41)	0.24	1.22(0.88-1.71)
				CT	65 (0.44)	67 (0.45)		
				TT	29 (0.20)	21 (0.14)		
	rs9309056	0.12	0.84	TT	116 (0.80)	110 (0.75)	0.13	0.67(0.40-1.13)
				CT	28 (0.19)	34 (0.23)		
				CC	1 (0.01)	3 (0.02)		
	rs746233	0.47	0.84	GG	50 (0.35)	32 (0.22)	$8.62x10^{-3}$	0.63(0.45-0.89)
				AG	68 (0.47)	74 (0.51)		
				AA	26 (0.18)	40 (0.27)		
	rs7640737	0.14	0.70	CC	101 (0.68)	121 (0.81)	$4.00x10^{-3}$	2.27(1.30-3.96)
				CT	43 (0.29)	27 (0.18)		
				TT	5 (0.03)	1 (0.01)		
	rs10490775	0.14	0.70	GG	100 (0.68)	120 (0.81)	$4.00x10^{-3}$	2.27(1.30-3.96)
				AG	42 (0.29)	27 (0.18)		
				AA	5 (0.03)	1 (0.01)		
	rs10484323	0.09	0.89	TT	119 (0.81)	124 (0.83)	0.48	1.23(0.70-2.16)

				CT	26 (0.18)	24 (0.16)		
				CC	2 (0.01)	1 (0.01)		
	rs2057181	0.35	0.70	CC	51 (0.37)	65 (0.46)	0.21	1.26(0.88-1.80)
				CT	70 (0.50)	60 (0.43)		
				TT	18 (0.13)	16 (0.11)		
	rs13221882	0.04	1.00	AA	135 (0.92)	138 (0.93)	0.57	1.25(0.58-2.68)
				AG	10 (0.07)	11 (0.07)		
				GG	2 (0.01)	0 (0.00)		
	rs10492188	0.34	0.55	CC	74 (0.51)	58 (0.40)	0.21	0.80(0.56-1.14)
				CT	50 (0.35)	70 (0.48)		
				TT	20 (0.14)	17 (0.12)		
	rs2445770	0.06	0.41	GG	131 (0.90)	131 (0.87)	0.45	0.75(0.36-1.59)
				AG	14 (0.10)	19 (0.13)		
				AA	0 (0.00)	0 (0.00)		
	rs2446426	0.05	0.46	CC	131 (0.91)	133 (0.89)	0.56	0.80(0.37-1.71)
				CT	13 (0.09)	17 (0.11)		
				TT	0 (0.00)	0 (0.00)		
	rs7238810	0.39	0.45	AA	53 (0.36)	54 (0.36)	0.92	1.02(0.70-1.48)

				AG	77 (0.52)	75 (0.50)		
				GG	19 (0.13)	20 (0.13)		
	rs17595731	0.06	1.00	GG	115 (0.80)	140 (0.95)	$5.43x10^{-4}$	6.34(2.23-18.05)
				CG	28 (0.19)	7 (0.05)		
				CC	1 (0.01)	0 (0.00)		
	rs613872	0.29	0.13	TT	34 (0.24)	100 (0.70)	$2.51x10^{-9}$	8.90(4.34-18.27)
				GT	95 (0.66)	42 (0.29)		
				GG	15 (0.10)	1 (0.01)		
	rs618869	0.24	0.27	CC	57 (0.39)	106 (0.74)	$1.41x10^{-7}$	5.41(2.89-10.15)
				CT	78 (0.53)	36 (0.25)		
				TT	11 (0.08)	1 (0.01)		
	rs9954153	0.26	0.26	TT	57 (0.40)	93 (0.65)	$7.44x10^{-5}$	2.67(1.64-4.33)
				GT	77 (0.53)	48 (0.33)		
				GG	10 (0.07)	3 (0.02)		
	rs2286812	0.20	0.18	CC	79 (0.53)	109 (0.75)	$2.05x10^{-3}$	2.01(1.29-3.13)
				CT	62 (0.42)	32 (0.22)		
				TT	7 (0.05)	5 (0.03)		

Table S3. Association between Fuchs corneal dystrophy (FCD) and rs7640737 in *PTPRG*

Group	Allele distribution				Genotype distribution				
	Allele	**Case**	**Control**	**P-value**	**Genotype**	**Case**	**Control**	**P-value**	**Odds-ratio***
Discovery	C	0.79	0.90	$1.56x10^{-5}$	CC	0.64	0.81	$9.92x10^{-5}$	2.32(1.52-3.53)
	T	0.21	0.10		CT	0.31	0.18		
					TT	0.05	0.00		
Replication	C	0.82	0.90	0.0043	CC	0.68	0.81	$4.00x10^{-3}$	2.27(1.30-3.96)
	T	0.18	0.10		CT	0.29	0.18		
					TT	0.03	0.01		
Combined	C	0.81	0.90	$3.98x10^{-7}$	CC	0.66	0.81	$1.30x10^{-6}$	2.30(1.64-3.22)
	T	0.19	0.10		CT	0.30	0.18		
					TT	0.04	0.00		

*The odds ratio for one copy of the risk allele (heterozygotes, GT) is shown. See Table 1 in the main article for additional details

Table S4. Association between Fuchs corneal dystrophy (FCD) and rs10490775 in *PTPRG*

Group	Allele distribution				Genotype distribution				
	Allele	**Case**	**Control**	**P-value**	**Genotype**	**Case**	**Control**	**P-value**	**Odds-ratio***
Discovery	G	0.79	0.90	1.56×10^{-5}	GG	0.64	0.81	9.92×10^{-5}	2.32(1.52-3.53)
	A	0.21	0.10		AG	0.31	0.18		
					AA	0.05	0.00		
Replication	G	0.82	0.90	0.0054	GG	0.68	0.81	4.00×10^{-3}	2.27(1.30-3.96)
	A	0.18	0.10		AG	0.29	0.18		
					AA	0.03	0.01		
Combined	G	0.81	0.90	5.06×10^{-7}	GG	0.66	0.81	1.30×10^{-6}	2.30(1.64-3.22)
	A	0.19	0.10		AG	0.30	0.18		
					AA	0.04	0.00		

*The odds ratio for one copy of the risk allele (heterozygotes, GT) is shown. See Table 1 in the main article for additional details

Table S5. Step-wise logistic regression using SAS was performed using 5 SNPs across the *TCF4* locus that were genotyped in both the discovery and replication subjects. Four SNPs were independently associated with FCD as shown below.

Parameter	Chr*	Position	Haplotype block**	Allele	MAF*	Estimate*	P-value
Intercept						-1.7789	<.0001
rs17595731	18	51258450	10	G/C	0.05	1.4909	0.0007
rs613872	18	51361300	17	T/G	0.22	0.7857	0.0010
rs9954153	18	51689255	24	T/G	0.21	0.7277	0.0020
rs2286812	18	51868462	31	C/T	0.15	0.6295	0.0093

*Chr, chromosome; MAF, minor allele frequency; Estimate, regression coefficient

** See the footnote to table S6 for information regarding how the haplotype blocks were determined.. The exact location of these haplotype blocks is shown in Figure S5, except for block 10. The location of block 10 is from rs12963463 to rs9966430 and base pair 51250091 to 51276362. Note that the haplotype block is provided for informational purposes, but should not be used to compare to the haplotype studies using a much larger number of imputed SNPs. Haplotype studies could not be performed with the SNPs genotyped across *TCF4* in both the discovery and replication subjects, because there were just 5 SNPs.

Table S6. Association between individual haplotypes across the *TCF4* locus and FCD. Haplotype blocks highly associated with FCD are shown. The start and stop positions of these haplotypes are shown in Figure S6

Block	**Case:Cnt Freq**	**P Value**
Block 7		
TGG	0.477, 0.629	5.24E-05
TGT	0.345, 0.256	0.0095
CGG	0.066, 0.090	0.2418
CAG	0.112, 0.025	3.80E-07
Block 8		
AGCAACGTGGAGCTGCCCGCTTCG	0.263, 0.363	0.0051
AATAGAACCCGATAATCTATTCTT	0.174, 0.167	0.8035
AGCAACGTGGAGCTGCACGCCTCG	0.132, 0.144	0.6377
AGCAACGTGGAACTGCCCGCTTCG	0.070, 0.090	0.3279
GATAGAGCCCGATAATCTATTCTT	0.097, 0.050	0.0128
AGTGGCGTGGGACTATCTGCTCTG	0.109, 0.013	1.71E-09
AGTGGCGTGGGACTGTCTGCTTTG	0.027, 0.050	0.1362
AATAGAGCCCGATAATCTATTCTG	0.052, 0.012	6.00E-04
AGTGGCGTGGGACTATCTGTTCTG	0.012, 0.031	0.1034
AGCAACGTGGAGCTGCCCGCTTCT	0.011, 0.025	0.2087
AACAACGTGGAGCTGCCCGCTTCG	0.012, 0.015	0.728
AGTAGAGCCCGATAATCTATTCTT	0.012, 0.012	0.9303
Block 11		
CACGCT	0.465, 0.479	0.7181
CCTACC	0.244, 0.285	0.2324
CCTGCC	0.093, 0.137	0.0809
TACGTC	0.070, 0.073	0.8665
CACGCC	0.112, 0.015	2.20E-09
CCCGCC	0.012, 0.012	0.9912
Block 17		
GTAATATGTTATGTCAGATGCCTCAATGGGCGGTAGATTAAGTAGGGCGTTCAGCCATTGTCAC	0.136, 0.202	0.0231
ATAGTTTATATCGCCCAGCGCTCCGAAAGATAATGCGCCGAGGGAAGTACTCGACTAGCGCCAC	0.116, 0.156	0.1368
ATGATATGCTATGTCAGACAGCCGAAAATATAACGCGCCGAGGGAAGTACTCGAATAGCGCCAC	0.132, 0.137	0.851
ATAACATGTTATGTCAGACGCCTCAATGGGCGGTAGATTAAGTAGGGCGTTCAGCCATTGTCAT	0.085, 0.094	0.68
ATGATATGCTATGTCAGACACCCGAAAATATAACGCGCCGAGGGAAGTACTCGAATAGCGCCAC	0.062, 0.101	0.0703

GCAATAGGTTATATTAGACGCCTCAGTGGGCGATAGATTAGATAGGACGTTTAGCCGTTCTTGC	0.093, 0.083	0.6316
GTAATAGGTTATGTTAGACGCCTCAATGGGCGATAGATTAGATAGGACGTATAGCCATTCTTGC	0.171, 0.042	1.55E-09
ATAATATGCTATGTCAGACACCCGAAAATATAACGCGCCGAGGGAAGTACTCGAATAGCGCCAC	0.047, 0.079	0.0913
ATAATAGGTTATGTCAGACGCCTCAATGGGCGATGCGCCGAGGGAAGTACTCGACTAGCGTTGC	0.097, 0.013	3.51E-08
ATAATATGTTATGTCAGACGCCTCAATGGGCGATAGATTAAGTAGGGCGTTCAGCCATTGTCAC	0.019, 0.027	0.5198
ATGATATGCTATGTCAGACACCCGAAAATATAACGCGCCGAGGGAAGTACTCGAATAGCGCTGC	0.008, 0.014	0.4262
ATGATATGCTATGTCAGGCGCTCCGAAAGATAATGCGCCGAGGGAAGTACTCGACTAGCGCCAC	0.008, 0.013	0.4824
Block 31		
AGTCTTTGCCACTGCGCCGCAGGGAGCACTACG	0.097, 0.135	0.1303
GCTCCCCACAACTAAGCTGCAGGCCGCGCCACA	0.105, 0.123	0.4517
AGTCTTTGCCACTGCGCCGCAGGGAGCGCTACG	0.124, 0.100	0.3093
GCTCCTCGTAATTAAGCTGCAGGCAGCGCCACG	0.093, 0.113	0.3846
GCTCCTCGCAACTAAGCTGCAGGCATTGTCACG	0.085, 0.100	0.5094
ACTCCTCGCAACTAAGCTGCATGCATTGTCACG	0.058, 0.106	0.0288
GCCCCTCGCCGCCGAACTACGGACAGCGCTATG	0.097, 0.085	0.5705
GCCTCTCGCCGCCGAAATATGGACAGCGCTCCG	0.105, 0.065	0.0551
GCCTCTCGCCGCCGAAATATGGACAGCGCTACG	0.066, 0.081	0.4605
GCCTCTCGCCGCCGAACTACGGACAGCGCTATG	0.140, 0.025	5.97E-10
AGTCTTTGCCACTGCGCTGCAGGCAGCGCCACG	0.023, 0.038	0.2666
GCTCCCCACAACTAAGCCGCAGGGAGCGCTACG	0.000, 0.015	0.0452
Block 34		
CCA	0.481, 0.481	0.9969
TCA	0.217, 0.256	0.2358
TTA	0.167, 0.238	0.0217
CCC	0.136, 0.025	1.55E-09

The start and stop SNPs for each block are as follows: block 7, rs17594301-rs1348047; block 8, rs17594358-rs2919450; block 11, rs17089826-rs7233312; block 17, rs2123389-rs17089925; block 31, rs1477440-rs17806802; block 34, rs1945737-rs7235583. The table can be recreated by downloading all SNPs between rs17594301and rs7235583 in the GWAS from dbGaP, importing into Haploview, and defining haplotypes using the default settings (confidence intervals).

Table S7. Correlation between the seven haplotypes highly associated ($P < 10^{-6}$) with FCD across the *TCF4* locus is illustrated. Note that the entire 1 megabase locus contained six haplotypes blocks with seven highly associated haplotypes. Two of these haplotypes were located in the block 17.

Block		**7**	**8**	**11**	**17**	**17**	**31**	**34**
	Haplotype	**5**	**16**	**2**	**32**	**27**	**12**	**2**
7	**5**	1.00	0.90	0.83	-0.04	0.75	0.23	0.21
		Ref	<.2*E-16	<.2*E-16	0.3935	<.2*E-16	3.86E-6	2.47E-5
8	**16**		1.00	0.90	-0.02	0.81	0.28	0.26
			Ref	<.2*E-16	0.7361	<.2*E-16	1.90E-8	2.41E-7
11	**2**			1.00	-0.01	0.81	0.26	0.24
				Ref	0.9064	<.2*E-16	1.86E-7	1.76E-6
17	**32**				1.00	-0.03	0.40	0.40
					Ref	0.52	2.22E-16	<.2*E-16
17	**27**					1.00	0.32	0.30
						Ref	8.13E-11	1.73E-9
31	**12**						1.00	0.97
							Ref	<.2*E-16
34	**2**							1.00
								Ref

The top number provides the Pearson correlation, and the bottom number the p-value.
Ref, Reference

Table S8. The physical locations of the six haplotype blocks across the *TCF4* locus are provided.

Block	Start SNPs	End SNPs	BP	Gene Position
7	rs17594301	rs1348047	51199883 - 51201056	Intron6 -
8	rs17594358	rs2919450	51201487 - 51235543	Intron5 (includes exon 6)
11	rs17089826	rs7233312	51277991 - 51294571	Intron5 – Intron3 (includes exons 4 and 5)
17	rs2123389	rs17089925	51352970 - 51551832	Intron3 – 5'region (includes exons 1,2 and 3)
31	rs1477440	rs17806802	51847008 - 51869136	Upstream
34	rs1945737	rs7235583	51895543 - 51898210	Upstream

Supplemental Figures

Figure S1. Population stratification is illustrated by showing the first (PC1) and second (PC2) principal components for the 389 FCD cases and controls. There is minimal population stratification and there is no difference between cases and controls.

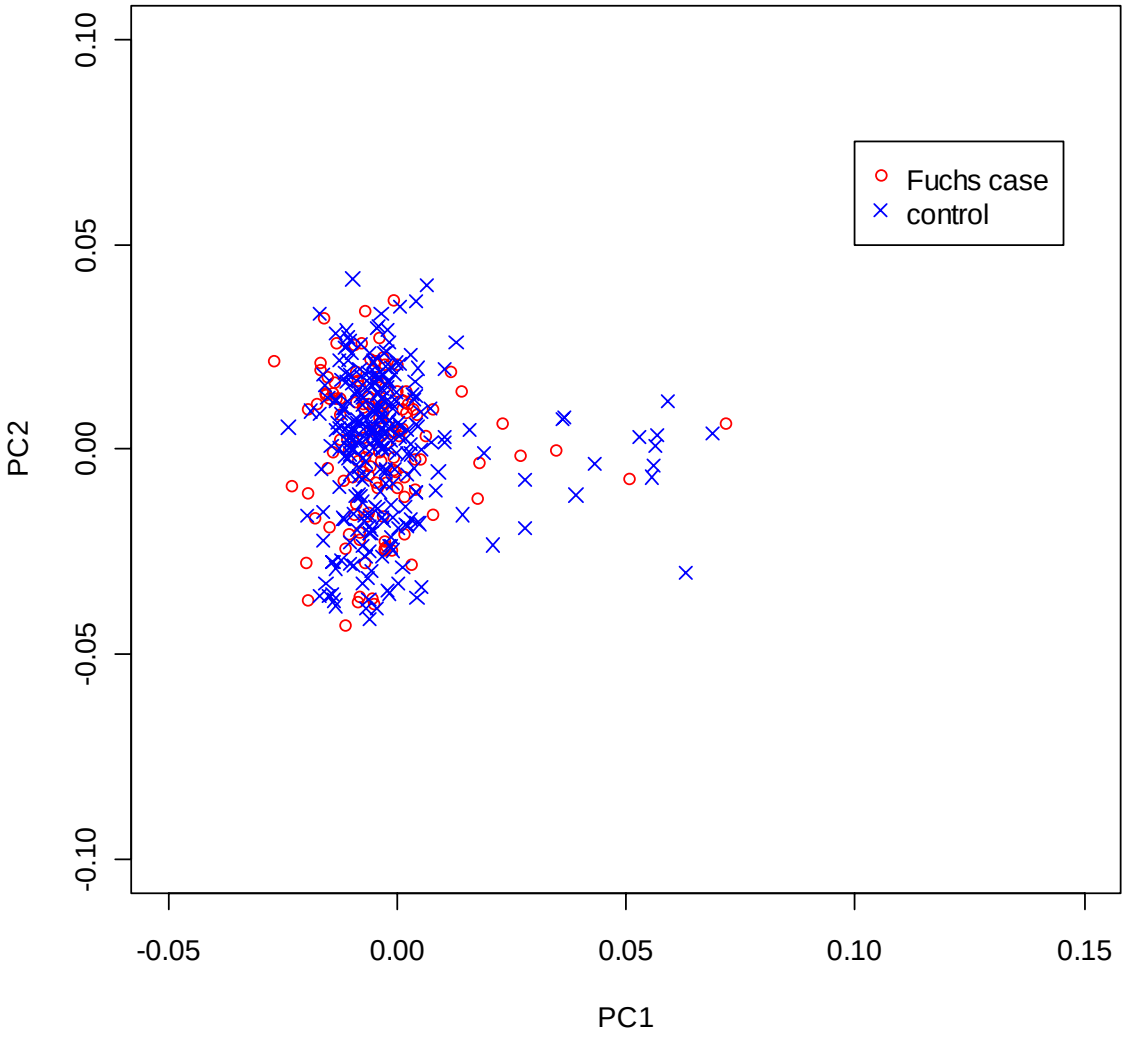


Figure S2. QQ plots of the observed over the expected P values from the discovery GWAS show the absence of genomic inflation. The genomic inflation factor was 0.9945.

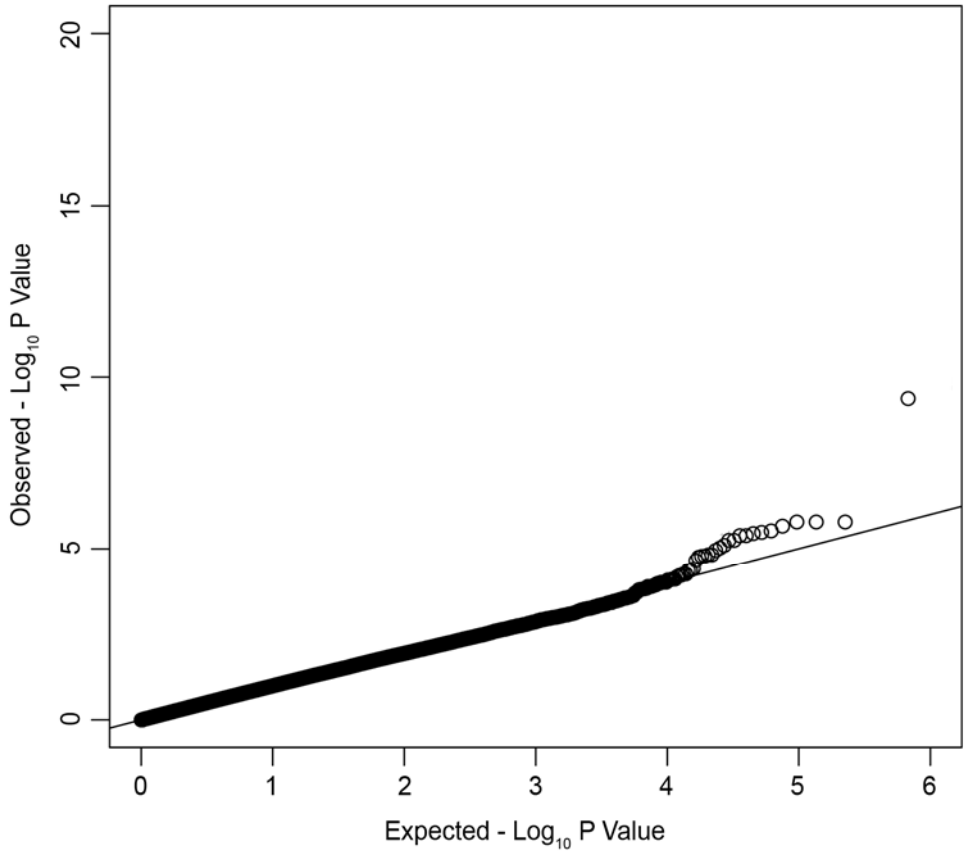


Figure S3. Linkage disequilibrium (r^2) using imputed genotypes for the 720 SNPs across the chromosome 18 locus (51004696-51999558 base pairs). The linkage disequilibrium pattern below was similar to the pattern observed in HapMap subjects. Note the presence of extensive linkage disequilibrium across this region.

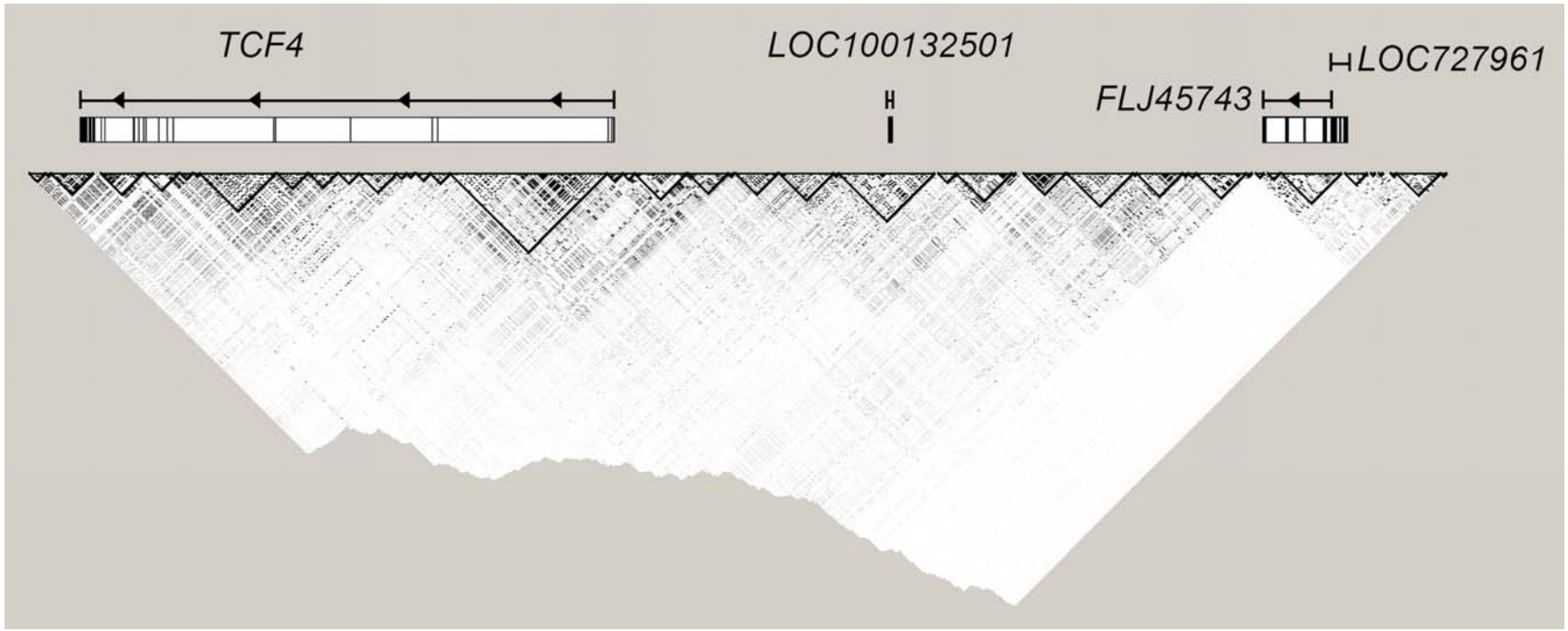


Figure S4. Linkage disequilibrium (r^2) using imputed genotypes for the 720 SNPs across the *TCF4* gene. Note the presence of strong linkage disequilibrium between individual SNPs across the entire gene.

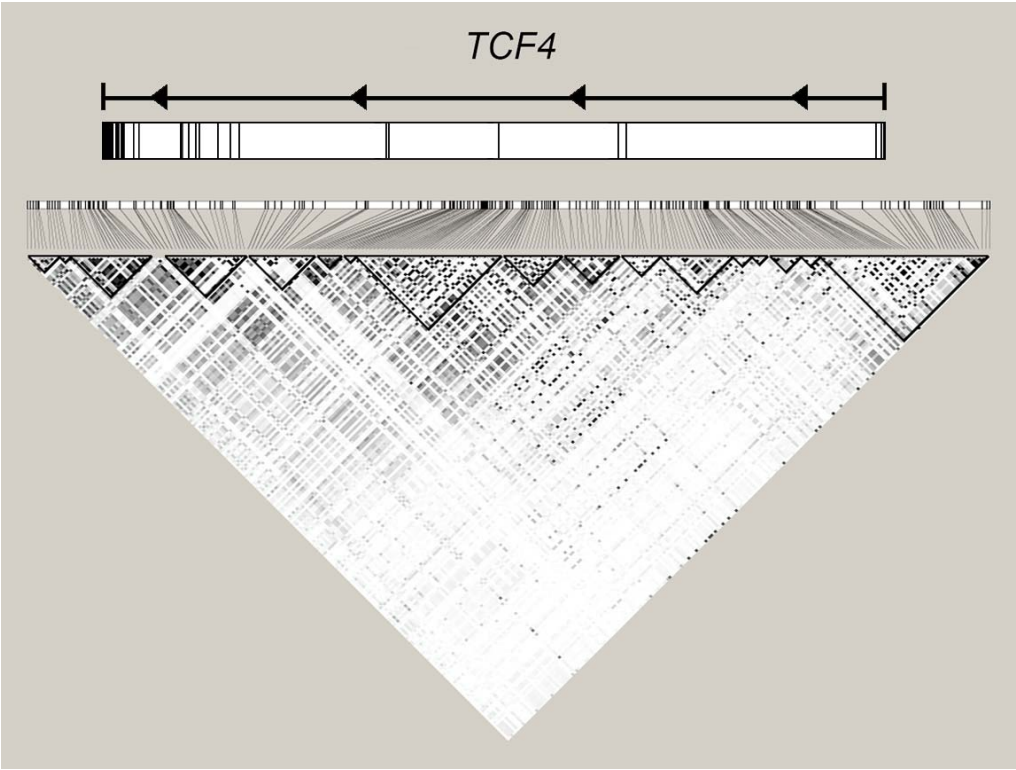


Figure S5. Association between Fuchs corneal dystrophy (FCD) and haplotypes across the *TCF4* locus. The strength of association (y-axis) and position in base-pairs (x-axis) for haplotypes composed of three adjacent imputed SNPs is shown. The 3-SNP haplotypes are shifted one SNP at a time across the 720 imputed SNPs and the global haplotype association with FCD calculated. The closed black circles show that many haplotypes are associated with FCD with the highest association spanning exon 6. After conditioning on the most highly associated haplotype (the green circle, SNPs rs3794891, rs11152363, and rs17594526), many haplotypes across the 5' and upstream regions of *TCF4* remained associated with FCD. The haplotypes containing rs613872 are highlighted as the three overlapping red circles prior to conditioning. The figure suggests that multiple ancestral segments of DNA (haplotypes) contribute to FCD.

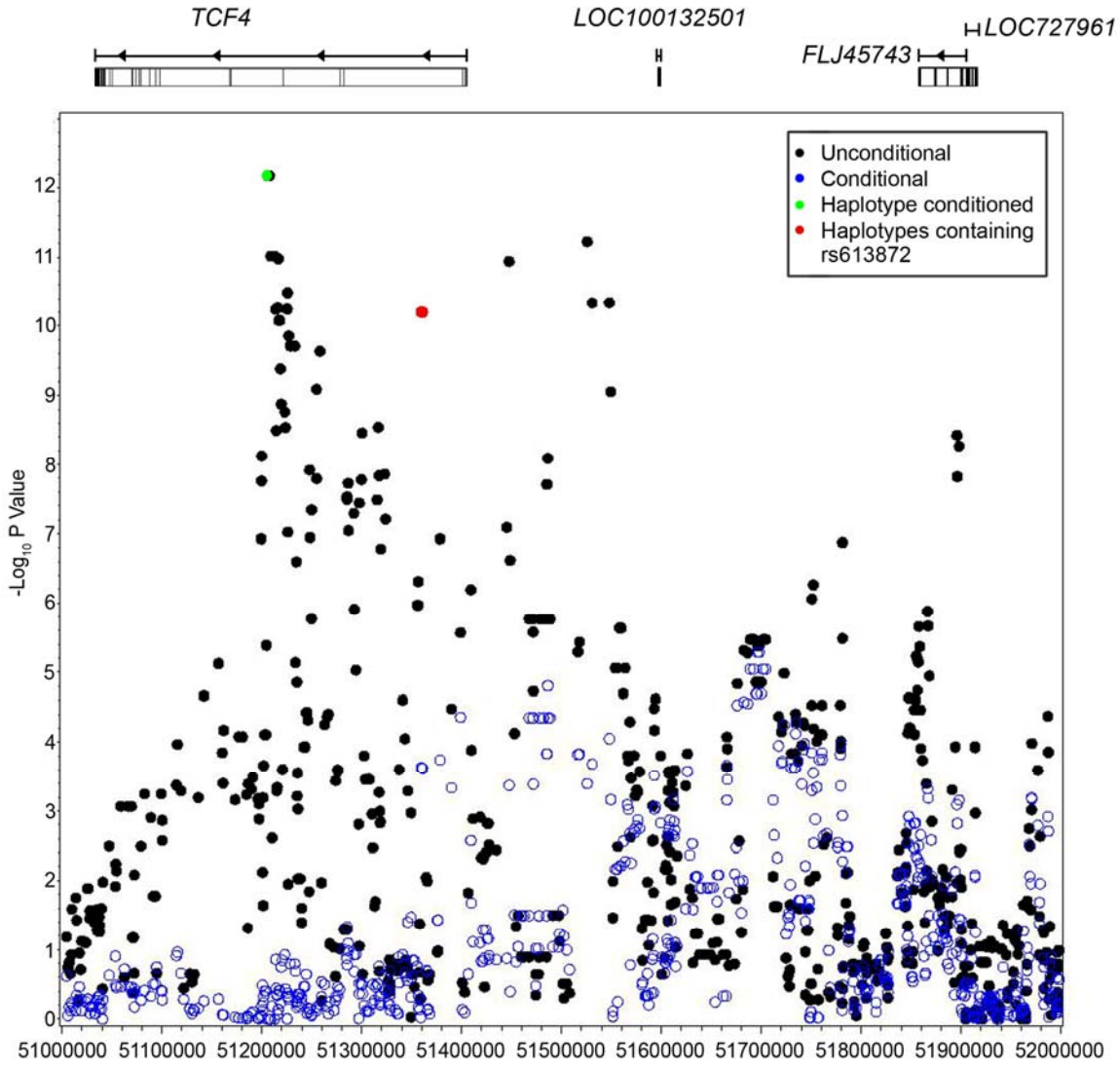
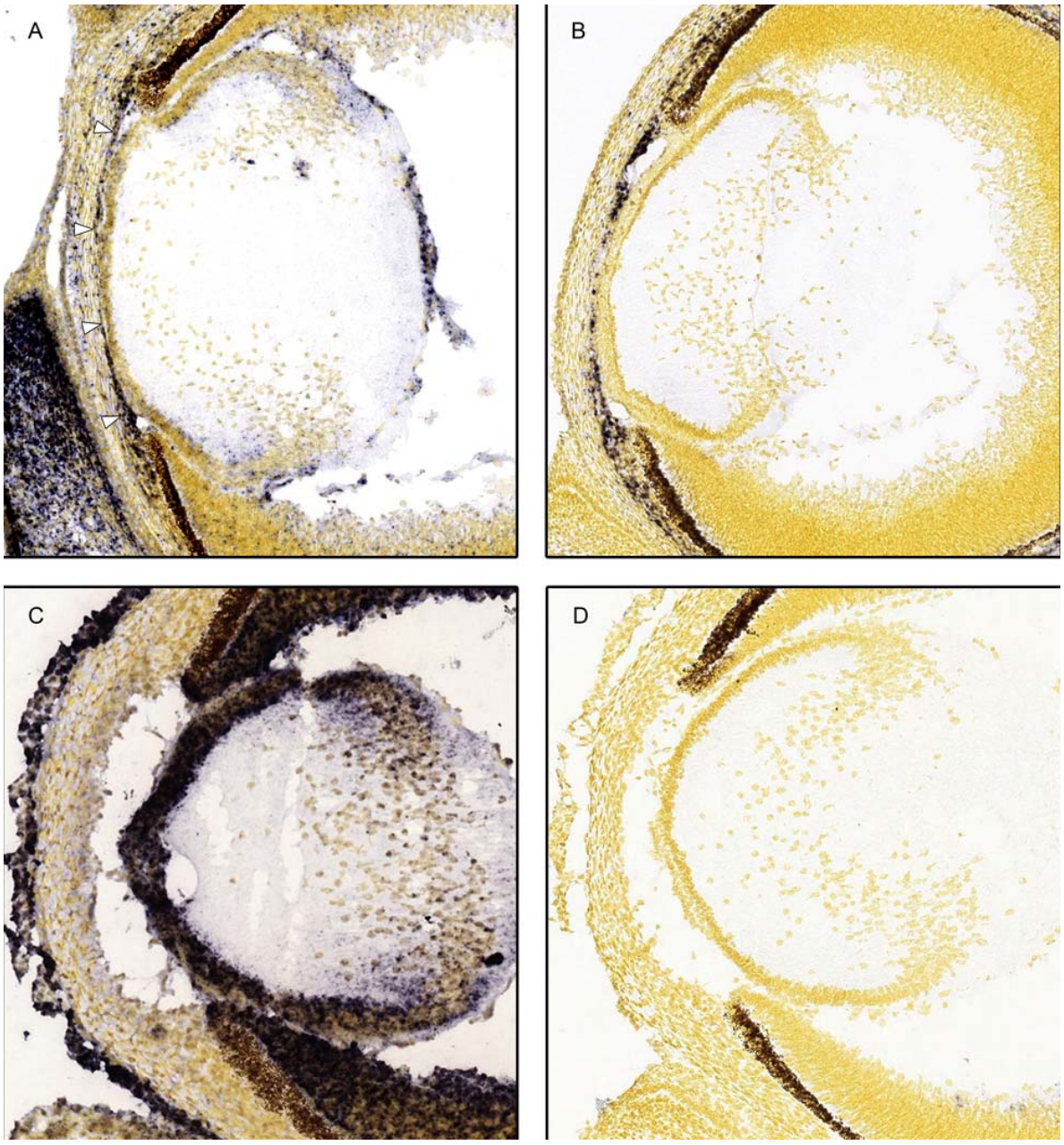

TCF4
LOC100132501
FLJ45743
LOC727961
Unconditional
Conditional
Haplotype conditioned
Haplotypes containing rs613872
-Log10 P Value
0
1
2
3
4
5
6
7
8
9
10
11
12
51000000
51100000
51200000
51300000
51400000
51500000
51600000
51700000
51800000
51900000
52000000

Figure S6. Expression of TCF4 in the developing corneal endothelium of the mouse. Expression of four genes using in-situ hybridization in coronal sections of the developing mouse eye on day E15.5 is shown as follows: A, TCF4 ; B, PITX2; C, PAX6; and D, POU4F1. Arrowheads indicate the corneal endothelium. Note that TCF4 is detected in the corneal endothelium similar to PITX2, which is known to be expressed in the corneal endothelium [3, 4]. PAX6 transcripts are present in the corneal epithelium, but are largely absent from the corneal endothelium [5, 6]. POU4F1 transcripts are detected in the retina, but not the cornea [7, 8]. All sections were obtained from the Developing Mouse Brain Atlas from the Allen Institute for Brain Science as described in the Supplementary text above.


A
B
C
D

References:

1. Krachmer JH, Purcell JJ, Young CW, Bucher KD. A study of 64 families with corneal endothelial dystrophy. Archives of Ophthalmology 1978;96:2036-9.

2. Hecker LA, Edwards AO, Ryu E, et al. Genetic control of the alternative pathway of complement in humans and age-related macular degeneration. Human molecular genetics 2009;19:209-15.

3. Evans AL, Gage PJ. Expression of the homeobox gene Pitx2 in neural crest is required for optic stalk and ocular anterior segment development. Human molecular genetics 2005;14:3347-59.

4. Gage PJ, Qian M, Wu D, Rosenberg KI. The canonical Wnt signaling antagonist DKK2 is an essential effector of PITX2 function during normal eye development. Developmental biology 2008;317:310-24.

5. Davis J, Duncan MK, Robison WG, Jr., Piatigorsky J. Requirement for Pax6 in corneal morphogenesis: a role in adhesion. Journal of cell science 2003;116:2157-67.

6. Collinson JM, Quinn JC, Hill RE, West JD. The roles of Pax6 in the cornea, retina, and olfactory epithelium of the developing mouse embryo. Developmental biology 2003;255:303-12.

7. Quina LA, Pak W, Lanier J, et al. Brn3a-expressing retinal ganglion cells project specifically to thalamocortical and collicular visual pathways. J Neurosci 2005;25:11595-604.

8. Pan L, Yang Z, Feng L, Gan L. Functional equivalence of Brn3 POU-domain transcription factors in mouse retinal neurogenesis. Development (Cambridge, England) 2005;132:703-12.