

## ARTICLE

# Variation near complement factor I is associated with risk of advanced AMD

Jesen A Fagerness<sup>1,2</sup>, Julian B Maller<sup>1,2,3</sup>, Benjamin M Neale<sup>1,2,4</sup>, Robyn C Reynolds<sup>5</sup>, Mark J Daly<sup>1,2</sup> and Johanna M Seddon<sup>\*,5</sup>

<sup>1</sup>Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; <sup>2</sup>Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; <sup>3</sup>Department of Statistics, University of Oxford, Oxford, UK; <sup>4</sup>SGDP Centre, Institute of Psychiatry, King's College London, London, UK; <sup>5</sup>Ophthalmic Epidemiology and Genetics Service, New England Eye Center, Tufts Medical Center, Boston, MA, USA

**A case–control association study for advanced age-related macular degeneration was conducted to explore several regions of interest identified by linkage. This analysis identified a single nucleotide polymorphism just 3' of complement factor I on chromosome 4 showing significant association ( $P < 10^{-7}$ ). Sequencing was performed on coding exons in linkage disequilibrium with the detected association. No obvious functional variation was discovered that could be the proximate cause of the association, suggesting a noncoding regulatory mechanism.**

*European Journal of Human Genetics* (2009) 17, 100–104; doi:10.1038/ejhg.2008.140; published online 6 August 2008

**Keywords:** AMD; macular degeneration; genetic association; complement pathway

## Introduction

The association of age-related macular degeneration (AMD) with variants on chromosome 1 (complement factor H (CFH)), chromosome 6 (CFB; C2), chromosome 10 (LOC387715/ARMS2), and chromosome 19 (C3) has clearly identified the primary role of the complement pathway in disease pathogenesis, and is consistent with the linkage peaks observed in several whole-genome linkage studies.<sup>1–9</sup> Follow-up of whole-genome linkage regions with fine-mapping has met with limited success in many other complex diseases. However, the effect sizes of the identified risk variants for AMD have been dramatically larger than most late-onset disease associations.<sup>10–12</sup> AMD, the leading cause of irreversible blindness in older adults,<sup>13</sup> has become a model for the genetics of complex disease. In

light of these early successes, we selected 1500 single nucleotide polymorphisms (SNPs) using two different criteria: targeting genes in regions under suggestive linkage peaks from a recent meta-analysis<sup>1</sup> and genes selected from the complement pathway not in these regions.

## Materials and methods

### Sample

The study population consisted of 2053 unrelated Caucasian individuals 60 years of age or older diagnosed based on ocular examination and fundus photography. There were 1228 cases of both dry and neovascular (wet) advanced AMD and 825 controls.<sup>14</sup> The mean age was 74 years for controls (54% women) and 78 years for affected individuals (55% women).<sup>8</sup> Informed consent was obtained in writing from all participants, and procedures were approved by the appropriate institutional review boards. This is largely the same sample set with the same phenotyping criteria that we described in detail previously.<sup>8,9</sup> Importantly, this sample has been previously confirmed to show no inflation of case–control association statistics because of population substructure.<sup>8</sup>

\*Correspondence: Dr JM Seddon, Ophthalmic Epidemiology and Genetics Service, New England Eye Center, 800 Washington Street, 450, Tufts Medical Center, Boston, MA 02111, USA.

Tel: +1 617 636 9000; Fax: +1 617 636 1124;

E-mail: jseddon@tuftsmedicalcenter.org

Received 22 February 2008; revised 25 June 2008; accepted 27 June 2008; published online 6 August 2008

### SNP selection

We selected a total of 1500 SNPs in complement pathway genes and across regions of chromosomes 1, 2, 3, 4, 6, and 16 based on the Fisher *et al*<sup>1</sup> bin rank of a meta-analysis of previous whole-genome linkage studies. We chose SNPs in and around regions of transcription as described by Wiltshire *et al*.<sup>15</sup> However, we built upon the efficiency of this strategy by only selecting SNPs that tag seven or more other SNPs. This SNP selection routine was conducted by using tagger (<http://www.broad.mit.edu/mpg/tagger/>) and HapMap data from the CEPH population (phase II, <http://www.hapmap.org>). We selected SNPs with a minor allele frequency (MAF) > 10% and with a minimum  $r^2$  of 0.8. Within this set of SNPs were nine SNPs in and around the complement factor I (CFI) region. We chose another 20 SNPs in the region to adequately tag the entire region using the same tagging parameters as above. The 29 total SNPs span a 173 kilobase (kb) region and tag 114 out of the 116 HapMap SNPs in the area that have a MAF above 5%. These 29 tag SNPs provide very good information coverage for the 114 HapMap SNPs having a mean  $r^2 = 0.966$ .

### Genotyping

The 1500 SNPs were genotyped at the Center for Inherited Disease Research (CIDR) using an Illumina OPA of which 1409 SNPs passed quality control measures as previously outlined.<sup>8</sup> The follow-up genotyping and sequencing was performed at the Broad and National Center for Research Resources (NCRR) Center for Genotyping and Analysis using the Sequenom MassARRAY system for iPLEX assays and the ABI 3700XL sequencing system, respectively.

### Sequencing

A novel SNP was discovered by sequencing 85 subjects as a subset of our case-control cohort. This SNP has been labeled Broad13981263 and is an A/G SNP, with A being the major allele, and G having an MAF of 7.65%. Broad13981263 is on chromosome 4 just 5' (113bp) of CFI's exon 12 according to dbSNP, and is located at the coordinate 110883313 base pairs on chromosome 4

according to NCBI build 36.1 (Table 1). Broad13981263 has an  $r^2$  of 0.057, 0.006, and 0.003, respectively with rs13117504, rs10033900, and rs11726949. As our two most associated SNPs are not in high correlation with our novel SNP, nor does it have a MAF that is very close to our associated SNPs, we are fairly certain that Broad13981263 is not the causal SNP driving the association in the region between 110787671 and 110961059.

### Analysis

All linkage disequilibrium (LD) calculations were performed with Haploview.<sup>16</sup> We conducted single-locus and two-marker haplotype association analysis using logistic regression tests implemented in PLINK.<sup>17,18</sup>

### Liability estimation

To calculate the percent variance accounted for by any risk alleles, we assume a prevalence of late-stage AMD in this older age group to be 5% and that liability is normally distributed in the population, with a mean of 0 and a variance of 1.

### Results

The most significantly associated SNP in the experiment, rs10033900 ( $P = 9.11 \times 10^{-8}$ ), resides in the chromosome 4 linkage peak region according to Fisher *et al*<sup>1</sup> and was very close to genome-wide significance levels as put forth by Dudbridge and Gusnanto and Pe'er *et al*<sup>19,20</sup> (Supplementary Table 1 shows the full results of the screen). Several nearby SNPs were also associated with  $P < 0.0005$ , suggesting this association was not because of a sporadic genotyping artifact. This SNP is 2781 bp upstream of the 3' UTR of CFI.

Given this compelling result, we decided to genotype a much higher density of SNPs in this region. Our originally associated SNP (rs10033900) remained the most highly associated SNP with a  $P$ -value of  $6.46 \times 10^{-8}$  (OR = 0.7056 referring to lower-risk C allele) (Table 2). This SNP showed a

**Table 1** New SNP found by sequencing the CFI exonic regions

SNP name	Chromosome	hg18	Major allele	Sequence
Broad13981263	4	110883313	A	TACATCTTGACATCTTGGATAAACCACTTGGCACTTACCTGCACATTCCAT TTCTTTTTCATAGAAACGATTTCCGTA AAACTTAGAGCAGTTGCTTATTAGT TTAACTTCACCCCACTGAAGTGAAAAGACTCTTTCGTTATCTAAACAAAGT GAGAAAGCAAACATTTAGAAGTCACAAATGAGAAATCTAAATACATACTCT CATAACTTAAACCACTTGGGATTATGAAAGGGTGATAGTTTTCAATAT [A/G]TAAATTTTTGAGAACTCTTCCCTTAGCGTGTTAATCATATCATATGTC TATTTCTAAAATTGGATGGATAGTCAAGGGGGACTATTGAACATATGGTCT GAAGTCACATTCATTTAATAAAGAGCTGGTACCTTAAGTTGAATGAGATAAA TCTTCCCTTTGGTTGCAGAACTCAGTCAGGCAATTGGATAGGAATCAGT GATAAGGTCTTTTCACATAAATACCTCCTATCTTCTCACAGTCCCTTATTTCC

Abbreviations: SNP, single nucleotide polymorphisms; hg18, Human genome build 18; NCBI build 36.1.

**Table 2** 29 SNPs tested across PLA2G12A and CFI on chromosome 4

CHR	SNP	hg18	SNP spacing	Allele 1 (A1)	Frequency of A1 affected (%)	Frequency of A1 unaffected (%)	Allele 2 (A2)	CHISQ	P-value	Odds ratio	Conditional rs10033900
4	rs13101299 <sup>a</sup>	110787671		A	10.0	8.9	G	1.40	0.237	1.138	0.123
4	rs9990765 <sup>a</sup>	110788160	489	C	30.1	36.5	T	18.15	2.04E-05	0.750	0.114
4	rs4698774 <sup>a</sup>	110791792	3632	G	27.5	27.6	C	0.00	0.953	0.996	0.782
4	rs6830606 <sup>a</sup>	110793498	1706	C	9.7	8.5	T	1.74	0.187	1.158	0.843
4	rs7690921 <sup>a</sup>	110798195	4697	T	35.4	29.5	A	15.59	7.87E-05	1.312	0.224
4	rs17440280 <sup>a</sup>	110801549	3354	C	37.5	36.8	T	0.21	0.648	1.031	0.661
4	rs4698779 <sup>a</sup>	110816366	14817	G	12.9	13.9	C	0.89	0.346	0.916	0.252
4	rs1800627 <sup>a</sup>	110833107	16741	T	45.8	51.4	C	12.53	4.01E-04	0.798	0.747
4	rs768063 <sup>a</sup>	110839905	6798	A	3.8	3.3	G	0.89	0.345	1.180	0.318
4	rs5030535 <sup>a</sup>	110841576	1671	G	34.3	40.3	A	14.82	1.18E-04	0.776	0.610
4	rs2285714 <sup>a</sup>	110858259	16683	T	46.2	38.8	C	21.35	3.84E-06	1.350	0.446
4	rs2107047 <sup>a</sup>	110867224	8965	A	7.4	8.2	G	0.82	0.364	0.898	0.667
4	rs6854876 <sup>b</sup>	110872889	5665	C	40.1	47.3	G	18.86	1.41E-05	0.747	0.118
4	rs2346841 <sup>b</sup>	110874089	1200	A	6.0	6.4	G	0.26	0.607	0.931	0.437
4	rs13117504 <sup>a</sup>	110878305	4216	G	37.7	45.8	C	26.93	2.11E-07	0.715	0.028
4	rs10033900 <sup>a,b</sup>	110878516	211	C	46.0	54.7	T	29.22	6.46E-08	0.706	NA
4	rs11726949 <sup>b</sup>	110884079	5563	T	7.9	5.0	C	11.96	5.44E-04	1.627	0.007
4	rs6848178 <sup>b</sup>	110885620	1541	T	42.8	46.0	A	3.71	0.054	0.879	0.499
4	rs6822976 <sup>a</sup>	110885741	121	A	48.4	49.0	G	0.15	0.698	0.976	0.300
4	rs9998151 <sup>b</sup>	110886794	1053	C	5.2	3.7	T	4.32	0.038	1.410	0.200
4	rs4698784 <sup>a</sup>	110888710	1916	A	32.6	31.3	T	0.72	0.397	1.061	0.475
4	rs11098043 <sup>a</sup>	110891734	3024	G	23.8	24.9	A	0.55	0.457	0.946	0.805
4	rs13129180 <sup>b</sup>	110905862	14128	T	26.3	27.2	A	0.43	0.513	0.952	0.816
4	rs1000954 <sup>b</sup>	110929483	23621	A	28.9	30.6	G	1.36	0.243	0.918	0.639
4	rs4698788 <sup>a</sup>	110938983	9500	C	2.8	2.7	T	0.04	0.847	1.039	0.686
4	rs7675460 <sup>b</sup>	110940037	1054	A	39.1	37.6	C	0.84	0.359	1.065	0.234
4	rs4698792 <sup>a</sup>	110948753	8716	T	29.2	30.3	C	0.61	0.434	0.947	0.943
4	rs1002989 <sup>a</sup>	110953291	4538	C	8.7	5.9	T	11.19	8.25E-04	1.527	0.020
4	rs4422417 <sup>a</sup>	110961059	7768	G	14.1	11.9	A	3.95	0.047	1.213	0.073

Abbreviations: SNP, single nucleotide polymorphism; hg18, Human genome build 18; NCBI build 36.1; CHISQ,  $\chi^2$ -value of allele 1 frequency differences between cases and controls.

Conditional rs10033900, *P*-value based on conditional logistic regression of SNP.

<sup>a</sup>SNP tested on Sequenom iPLEX at Broad and (NCRR) Center.

<sup>b</sup>SNP tested on Illumina OPA at CIDR.

**Table 3** Two-marker haplotype association results for most significant SNPs in CFI region

SNPs	Haplotype	Frequency of affected (%)	Frequency of unaffected (%)	Degrees of freedom	CHISQ	P-value	Odds ratio
rs13117504 rs10033900	OMNIBUS	NA	NA	3	34.61	1.48E-07	NA
rs13117504 rs10033900	CT	50.07	40.92	1	32.52	1.18E-08	1.448
rs13117504 rs10033900	GC	33.64	41.48	1	25.51	4.39E-07	0.715
rs13117504 rs10033900	CC	12.39	13.25	1	0.6447	0.4220	0.926
rs13117504 rs10033900	GT	3.90	4.35	1	0.5068	0.4765	0.891

Abbreviations: CHISQ,  $\chi^2$ -value; NA, Not Applicable (OMNIBUS is a test of all haplotypes)<sup>17</sup>; SNPs, single nucleotide polymorphisms tested as haplotypes.

very high level of genotyping concordance between the Illumina assay and the iPLEX assay designs.

We tested 29 SNPs across this region for association; conditioning on our most associated SNP, rs10033900, and observed no significant independent associations. We did, however, observe modest residual association at two neighboring, highly correlated SNPs (Table 2). This result suggests that rs10033900 may not be the causal variant but may be highly correlated with said variant. Therefore, we

applied multimarker haplotype tests in an attempt to refine and isolate the association signal. We tested the two-marker haplotype of the two closest SNPs to rs10033900, both 5' (rs13117504) and 3' (rs11726949). The two-marker haplotype between rs13117504 and rs10033900 shows a somewhat stronger association to AMD than either SNP alone with a *P*-value of  $1.18 \times 10^{-8}$  (Table 3). None of these three SNPs appear to be functional, although rs11726949 is in intron 11 of CFI. We also tested for differences in

association between the neovascular ('wet') and geographic atrophy ('dry') forms of advanced AMD and found only a 0.2% difference in MAF (46%) between the two groups.

We conservatively defined the span of LD that encompassed the SNPs of interest. All other HapMap SNPs in this region are correlated to rs10033900 and rs13117504 with an  $r^2 > 0.35$ . Multimarker haplotype testing across this region did not show any further association above the levels observed by the two-marker haplotype created between rs13117504 and rs10033900. We then sequenced all of the exons in this region to determine whether an obvious functional variant exists that explains this association. This block of LD spans the last two exonic regions of the 3' end of CFI and all four exons of phospholipase A(2) group 12A (PLA2G12A). We found no SNPs in either gene transcript that could statistically explain the association observed at rs10033900. We did find a novel SNP just 5' of exon 12 in CFI, but this SNP does not appear to be in high  $r^2$  with our associated SNP or haplotype and is, therefore, not the biological source of association.

We next evaluated the role of epistasis between rs10033900 and rs13117504 and the six variants previously established to be associated with AMD.<sup>2-9</sup> Specifically, two variants at CFH, two variants at the CFB/C2 locus, one at the LOC387715/HTRA1 locus, and one at the C3 locus were established as unequivocally associated to AMD risk in this cohort – the typing and analysis of which was described in Maller *et al.*<sup>8,9</sup> Using logistic regression, we observed no statistically significant interaction terms between any pair of these SNPs. Although weak interactions cannot be excluded, this result suggests that despite targeting the same pathway, these variants largely confer risk in an independent, log-additive fashion.

Given the independent action of this new variant, we were able to add it to the multilocus model from Maller *et al.*<sup>9</sup> We estimate that this variant accounts for approximately 1% of the population variance of liability.<sup>21</sup>

## Discussion

We identify one more gene involved in the complement pathway implicated in AMD pathogenesis. Although the complement pathway has been extensively studied, we are only recently learning about its relationship with AMD. The CFI gene spans 63 kb and contains 13 exons, the first 8 of which encode the heavy chain and the last 5 the light chain, which contains the serine protease domain.<sup>22</sup> This serine protease domain is responsible for cleaving and inactivating C4b and C3b.<sup>23</sup> C3b inactivation by CFI is regulated by CFH. CFH acts as a cofactor for CFI-mediated cleavage of C3b and also has decay accelerating activity against the alternative pathway C3 convertase, C3bBb. Membrane cofactor protein (or CD46) also acts as a

cofactor for CFI-mediated cleavage of C3b by downregulating the complement cascade.<sup>24</sup>

Although it would be a remarkable coincidence if CFI is not the associated gene, in the absence of a proven causal variant, we cannot formally exclude PLA2G12A from consideration as the source of this association. Secreted PLA2G12A, expressed by the corneal epithelium, could be involved in the normal antibacterial activity in tears and wound healing.<sup>25</sup> PLA2G12A has been shown to have bactericidal activity against the Gram-negative bacterium *Escherichia coli* and *Helicobacter pylori in vitro*.<sup>26</sup> PLA2G12A might also participate in helper T-cell immune response by the release of immediate second signals and generation of downstream eicosanoids.<sup>27</sup>

With this result we have found two SNPs and their combined haplotype that achieve genome-wide significance.<sup>19,20</sup> It is technically possible that the two intergenic SNPs could form a functional haplotype. However, a more likely explanation is that these SNPs tag an undiscovered, biologically relevant, structural variant. The next step will be to more comprehensively evaluate this region with the aim of uncovering this variation.

## Acknowledgements

Funding was provided by the National Institutes of Health grant EY11309; the Foundation Fighting Blindness; the Massachusetts Lions Research Fund; Research to Prevent Blindness; Macular Degeneration Research Fund of the Ophthalmic Epidemiology and Genetics Service, the New England Eye Center, Tufts Medical Center; Center for Inherited Disease Research; and the Broad Institute Center for Genotyping and Analysis (through grant U54 RR020278 from the National Center for Research Resources). We thank AREDS participants and investigators and the EMMES Corporation for their work on the AREDS Genetic Repository. We also thank Daniel Mirel, Brian Galloway (project management); Robb Onofrio, Rob Borowski, Kat Irzene, Tony Rachupka (genotyping/sequencing), David Turner, Meg Parker (sample management); and Chris Cotsapas (bioinformatics).

## References

- 1 Fisher SA, Abecasis GR, Yashar BM *et al*: Meta-analysis of genome scans of age-related macular degeneration. *Hum Mol Genet* 2005; **14**: 2257–2264.
- 2 Klein RJ, Zeiss C, Chew EY *et al*: Complement factor H polymorphism in age-related macular degeneration. *Science* 2005; **308**: 385–389.
- 3 Haines JL, Hauser MA, Schmidt S *et al*: Complement factor H variant increases the risk of age-related macular degeneration. *Science* 2005; **308**: 419–421.
- 4 Edwards AO, Ritter III R, Abel KJ *et al*: Complement factor H polymorphism and age-related macular degeneration. *Science* 2005; **308**: 421–424.
- 5 Hageman GS, Anderson DH, Johnson LV *et al*: A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proc Natl Acad Sci USA* 2005; **102**: 7227–7232.
- 6 Rivera A, Fisher SA, Fritsche LG *et al*: Hypothetical LOC387715 is a second major susceptibility gene for age-related macular degeneration, contributing independently of complement factor H to disease risk. *Hum Mol Genet* 2005; **14**: 3227–3236.

- 7 Gold B, Merriam JE, Zernant J *et al*: Variation in factor B (BF) and complement component 2 (C2) genes is associated with age-related macular degeneration. *Nat Genet* 2006; **38**: 458–462.
- 8 Maller J, George S, Purcell S *et al*: Common variation in three genes, including a novel non-coding variant in CFH, dramatically influences risk of age-related macular degeneration. *Nat Genet* 2006; **38**: 1055–1059.
- 9 Maller JB, Fagerness JA, Reynolds RC, Neale BM, Daly MJ, Seddon JM: Variation in complement factor 3 is associated with risk of age-related macular degeneration. *Nat Genet* 2007; **39**: 1200–1201.
- 10 Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research. *et al*: Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007; **316**: 1331–1336.
- 11 Easton DF, Pooley KA, Dunning AM *et al*: Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 2007; **447**: 1087–1093.
- 12 Samani NJ, Erdmann J, Hall AS *et al*: Genomewide association analysis of coronary artery disease. *N Engl J Med* 2007; **357**: 443–453.
- 13 Seddon J, Sobrin L: Epidemiology of age-related macular degeneration; In: Albert DM, Miller J (eds): *The Principles and Practice of Ophthalmology*, 3rd edn. W.B. Saunders: Philadelphia, 2007, pp 413–422.
- 14 Seddon JM, Sharma S, Adelman R: Evaluation of the clinical age-related maculopathy staging system. *Ophthalmol* 2006; **113**: 260–266.
- 15 Wiltshire S, de Bakker PI, Daly MJ: The value of gene-based selection of tag SNPs in genome-wide association studies. *Eur J Hum Genet* 2006; **14**: 1209–1214.
- 16 Barrett JC, Fry B, Maller J, Daly MJ: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.
- 17 Purcell S, Neale B, Todd-Brown K *et al*: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
- 18 Sullivan PF, Purcell S: Analyzing genome-wide association study data: a tutorial using PLINK; In: Neale BM, Ferreira MAR, Medland SE, Posthuma D (eds): *Statistical Genetics: Gene Mapping Through Linkage and Association*. Taylor and Table Francis: London, 2007, pp 355–394.
- 19 Dudbridge F, Gusnanto A: Estimation of significance thresholds for genomewide association scans. *Genet Epidemiol* 2008; **32**: 227–234.
- 20 Pe'er I, Yelensky R, Altshuler D, Daly MJ: Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol* 2008; **32**: 381–385.
- 21 Kwan JSH, Purcell S, Sham PC: Introduction to biometrical genetics; In: Neale BM, Ferreira MAR, Medland SE, Posthuma D (eds): *Statistical Genetics: Gene Mapping Through Linkage and Association*. Taylor and Table Francis: London, 2007, pp 355–394.
- 22 Vyse TJ, Bates GP, Walport MJ, Morley BJ: The organization of the human complement factor I gene (*IF*): a member of the serine protease gene family. *Genomics* 1994; **24**: 90–98.
- 23 Catterall CF, Lyons A, Sim RB, Day AJ, Harris TJ: Characterization of primary amino acid sequence of human complement control protein factor I from an analysis of cDNA clones. *Biochem J* 1987; **242**: 849–856.
- 24 Jha P, Bora PS, Bora NS: The role of complement system in ocular diseases including uveitis and macular degeneration. *Mol Immunol* 2007; **44**: 3901–3908.
- 25 Landreville S, Coulombe S, Carrier P *et al*: Expression of phospholipases A2 and C in human corneal epithelial cells. *Invest Ophthalmol Vis Sci* 2004; **45**: 3997–4003.
- 26 Huhtinen HT, Grönroos JO, Grönroos JM *et al*: Antibacterial effects of human group IIA and group XIA phospholipase A2 against *Helicobacter pylori* *in vitro*. *APMIS* 2006; **114**: 127–130.
- 27 Ho IC, Arm JP, Bingham III CO, Choi A, Austen KF, Glimcher LH: A novel group of phospholipase A2s preferentially expressed in type 2 helper T cells. *J Biol Chem* 2001; **276**: 18321–18326.

Supplementary information accompanies the paper on European *Journal of Human Genetics* website (<http://www.nature.com/ejhg>)