**Article section:** Clinical/Epidemiologic Research

# Evaluation of proteoglycan gene polymorphisms as risk factors in the genetic susceptibility to high myopia

Shea Ping Yip,[1] Kim Hung Leung,[1] Po Wah Ng,[1,2] Wai Yan Fung,[1] Pak Chung Sham,[3] Maurice K.H. Yap[2]

[1]Department of Health Technology and Informatics, and [2]Centre for Myopia Research, School of Optometry, The Hong Kong Polytechnic University, Hong Kong SAR, China; [3]Department of Psychiatry, The University of Hong Kong, Hong Kong SAR, China.

**Running title:** Proteoglycan gene polymorphisms in high myopia

**Corresponding author:** Prof. Shea Ping Yip, Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong SAR, China.
Tel: +852 3400 8571; Fax: +852 2362 4365; E-mail: shea.ping.yip@inet.polyu.edu.hk

**Grant information:** The work was supported by grants from Research Grant Council of Hong Kong (Ref. No. PolyU 5411/06M; account code: B-Q04A) and from The Hong Kong Polytechnic University (J-BB7P, 87MS and 87LV).

**Word count:** 3,250 words (Main text)

**ABSTRACT**

**PURPOSE.** We investigated the relationship between high myopia and single nucleotide polymorphisms (SNPs) in six proteoglycan genes: aggrecan (*ACAN*), fibromodulin (*FMOD*), decorin (*DCN*), lumican (*LUM*), keratocan (*KERA*) and epiphycan (*EPYC*). These genes were selected for study because they are involved in induced myopia in animals and/or are within the human *MYP3* locus identified by linkage analysis of families with high myopia.

**METHODS.** Two groups of Chinese subjects were studied: Group 1 (300 cases and 300 controls) and Group 2 (356 cases and 354 controls). Cases were high myopes with spherical equivalent (SE) ≤-8.00 dioptres and controls had SE between +1.0 and -1.0 dioptre. From these candidate genes, 60 tagging SNPs were selected. First, 12 DNA pools were *each* constructed from 50 samples of the *same* phenotype from Group 1 subjects, and tested for association for the SNPs. Second, putatively positive SNPs were confirmed by individual genotyping of Group 1 subjects. Finally, positive results were replicated in Group 2 subjects.

**RESULTS.** Of the 58 SNPs successfully screened by DNA pooling, 8 *ACAN* SNPs passed the threshold of $P \leq 0.10$ (nested ANOVA) and were then genotyped for individual samples. Haplotypes of rs3784757 and rs1516794 showed significant association with high myopia. However, the positive result could not be replicated in the second subject group.

**CONCLUSION.** These six proteoglycan genes were not associated with high myopia in Chinese and hence were unlikely to be important in the genetic predisposition to high myopia.

(238 words)

## INTRODUCTION

Myopic eyes focus distant objects in front of, instead of on, the retina. Myopia is the most frequent ocular disorder worldwide with a wide range of prevalence in different populations. It is much more frequent in Asian populations (50-70%) than in Caucasians populations (up to 30%).[1] High myopia, usually defined as -5.00 dioptres (D) or worse, is a predisposing factor for many pathological ocular complications such as glaucoma and retinal detachment.[1] It is a multifactorial disease caused by genetic factors, environmental factors and their interactions.[1]

Myopia usually develops as a result of excessive elongation of the eyeball with concomitant scleral remodelling that involves changes in the metabolism of collagen fibrils and proteoglycans.[2,3] Proteoglycans are proteins that are heavily glycosylated, and are each composed of a core protein covalently-linked with at least one glycosaminoglycan chain. They are important in regulating the assembly and interaction of collagen fibrils and scleral hydration although they contribute less than 1% to the dry weight of the sclera.[2,3] Proteoglycans found in the sclera include aggrecan (ACAN), fibromodulin (FMOD), lumican (LUM), decorin (DCN), biglycan (BGN), keratocan (KERA), epiphycan (EPYC) and others.[2,3]

In chicks induced to develop myopia by form deprivation, proteoglycan identified as aggrecan was found to accumulate in increased amounts in the presence of increased turnover rate in the cartilaginous layer of the posterior sclera in parallel with scleral growth of the treated eye when compared to the control eye.[4,5] Interestingly, the changes were in the opposite direction in the fibrous scleral layer of chick, and the sclera (fibrous in nature) of mammals like tree shrew and monkey.[6] In tree shrew, synthesis of glycosaminoglycan (and hence proteoglycan) was decreased in the posterior sclera of form-deprived eye with accompanying axial elongation and scleral thinning when compared to the control eye.[7] This could be explained by reduced *ACAN* gene expression as shown in the sclera of lens-induced myopic eye in tree shrew.[8] In induced myopia, alteration of DCN synthesis has been demonstrated in the sclera of the elongated eye in chick[4] and marmoset,[9] but not in tree shrew.[8,10] BGN and LUM showed little differential regulation in the sclera in response to lens-induced myopia

in tree shrew.[8] Intriguingly, features of high myopia (thin sclera, increased axial length and retinal detachment) were found in the eyes of lumican-fibromodulin double-null mice.[11] Linkage analysis of high myopia families has identified the *MYP3* locus at chromosome 12q21-23, and the proteoglycan genes *DCN*, *LUM*, *KERA* and *EPYC* lie adjacent to each other within this locus.[12]

This biological and positional information justifies our investigation of these proteoglycan genes as candidate genes for high myopia. An efficient stepwise DNA pooling-based case-control study approach[13,14] was used to investigate whether common tagging single nucleotide polymorphisms (tSNPs) of these genes were associated with high myopia in a Chinese population. We examined six proteoglycan genes *ACAN*, *FMOD*, *DCN*, *LUM*, *KERA* and *EPYC* (Table 1). *BGN* is an X-linked gene and hence cannot be studied by DNA pooling-based approach, in which DNA samples from male and female subjects are randomly mixed to construct DNA pools.

## MATERIALS AND METHODS

### Overview of the study approach

The first stage was a screen of separate case and control DNA pools to discover putatively positive SNPs. The second stage was to confirm these putatively positive SNPs by genotyping individual DNA samples that formed the original DNA pools. The third stage was to replicate the confirmed positive SNPs with a new sample set.

### Study subjects

For the first and the second stage of the study, we recruited 600 unrelated ethnic Chinese (**Group 1 subjects**): 300 cases of high myopia, and 300 controls of emmetropia.[15] For the third stage of the study, we recruited 710 unrelated ethnic Chinese (**Group 2 subjects**): 356 cases of high myopia and 354 controls of emmetropia.[16] We used the same recruitment criteria for both subject groups: high myopia was defined as spherical equivalent (SE) ≤-8.00 D for both eyes, and emmetropia as SE between +1.0 D and -1.0 D for both eyes. The characteristics of the subjects have been reported

previously,[15,16] and are summarized here. For cases, the average SE and axial length were -10.53 D and 27.76 mm in Group 1, and -10.30 D and 27.64 mm in Group 2, respectively. For controls, the mean SE and axial length were 0.03 D and 23.85 mm in Group 1, and 0.08 D and 23.73 mm in Group 2, respectively. Group 1 subjects were younger than group 2 subjects (26.1 years vs 33.6 years).

Ethics approval was obtained from the Human Subjects Ethics Subcommittee of the Hong Kong Polytechnic University, and tenets of the Declaration of Helsinki were observed. We recruited all subjects with written informed consent in the Optometry Clinic of our University, collected blood samples and extracted DNA from blood samples as has been reported previously.[15]

**DNA pool construction**

DNA concentration was accurately measured with a commercial PicoGreen Kit (Invitrogen, Carlsbad, CA) for DNA samples (**Group 1 subjects**). DNA samples at 5.0±0.3 ng/μl were mixed in equal amounts to create DNA pools. DNA pools were each created from 50 different subjects of the same disease status: 6 case pools from 300 case samples, and 6 control pools from 300 control samples.

**Genotyping of DNA pools and individual samples**

From the six candidate genes *ACAN*, *FMOD*, *DCN*, *LUM*, *KERA* and *EPYC*, 60 tSNPs were selected using the Tagger software (http://www.broadinstitute.org/mpg/tagger/) with the criteria of pairwise tagging, a correlation coefficient ($r^2$) of at least 0.8 and a minor allele frequency (MAF) of at least 0.1, and based on HapMap Han Chinese data (release 23a, phase II; http://www.hapmap.org/) (Table 1). For each of the selected tSNPs, DNA pools were amplified using touchdown polymerase chain reaction (PCR) as described previously[16] with specific primers and conditions shown in Supplementary Table S1. The PCR products were purified and then used as templates for primer extension (PE).[16] The PE products were injected into a denaturing high performance liquid chromatography (DHPLC) system (WAVE Nucleic Acid Fragment Analysis System; Transgenomic, Omaha, NE) to estimate the relative allele frequencies of the two alleles for each SNP.[16] Each DNA pool was independently tested and analyzed three times for each SNP, and hence each SNP had a total

of 36 sets of readings from 12 DNA pools. Differential incorporation of dideoxynucleotides in PE was corrected by means of a correction factor (called k correction factor),[17] which was obtained as the mean value of three independent analyses of a heterozygous sample for each SNP.

In the second stage of the study, individual samples of Group 1 subjects were genotyped with MassArray iPLEX assays (Sequenom, San Diego, CA) according to the recommended protocols (http://www.sequenom.com/); primer sequences are shown in Supplementary Table S2. Putatively positive SNPs from the first stage were grouped together with genetic markers of other studies, and genotyped using this method by a local genotyping laboratory (http://genome.hku.hk/portal/). One SNP (rs1516794) could not be grouped with other SNPs for iPLEX genotyping, and was genotyped in-house by restriction analysis. In the third stage, two confirmed SNPs (rs3784757 and rs1516794) were further genotyped for individual samples from Group 2 subjects by in-house methods based on restriction analysis (Supplementary Table S2).

**Statistical analysis**

The STATA package (version 8.2; StataCorp, Colleage Station, Texas, USA) was used to analyze the relative allele frequency data from DNA pools. For each SNP, the relative allele frequencies were calculated from the peak intensities of the two PE products with adjustment based on the k correction factor as described previously.[17] *Nested* analysis of variance (ANOVA) was used to compare the relative allele frequencies obtained from the case pools and the control pools because DNA pools were nested separately within the cases and the controls and there was no link between any case pools to any control pools.[16] To avoid excluding potential significant SNP, we used a $P$ value $\leq 0.10$ as the cutoff for following up putatively positive SNPs in the second stage.

Linkage disequilibrium (LD) measures were calculated and plotted using Haploveiw (version 4.2; http://www.broad.mit.edu/mpg/haploview/). Data of individual genotypes (second and third stages) were analyzed by Plink (ver. 1.07; http://pngu.mgh.harvard.edu/~purcell/plink/index.shtml). Exact test was used to test for Hardy-Weinberg equilibrium (HWE)[18] for cases and controls separately.

Logistic regression was employed for testing association between high myopia and single markers and their haplotypes. Exhaustive haplotype analysis was performed with a sliding-window strategy for windows of all possible sizes (i.e. 1 SNP to 7 SNPs per window). To avoid potential confounding, both sex and age were included as covariates for adjustment in all analyses. Wald test gave an asymptotic *P* value ($P_{asym}$) for each test. To correct for multiple comparisons, we used permutation test to permute the phenotype status of the subjects without changing the genotypes across all single markers and all haplotypes for samples individually genotyped within a subject group (Group 1, Group 2 or combined groups, each separately). We generated empirical *P* values ($P_{emp}$) based on 10,000 permutations within each subject group.

## RESULTS

### Stage 1: Analysis of DNA pools (Group 1 subjects)

We failed to find any heterozygous sample after screening 40 samples for two SNPs (rs7174219 and rs7173022) of the *ACAN* gene, which were then dropped from testing. Thus, 58 were successfully analyzed by the DNA pooling approach (Table 2). The mean k correction factor was 0.9488 (range: 0.5645 – 2.6026). The first eluted allele had an estimated frequency ranging from 0.0811 to 0.8991 for cases, and from 0.1130 to 0.9037 for controls. Estimated difference in allele frequencies (case pools – control pools) ranged from -0.0548 to 0.0434. Of the 58 tSNPs tested, 8 SNPs showed a significant difference in allele frequencies at a cutoff of $P \leq 0.10$, all belonging to the *ACAN* gene (Table 2). For confirmation, these 8 SNPs were genotyped for individual samples that formed the original DNA pools (Group 1 subjects). No significant difference in allele frequencies was detected for the remaining 50 SNPs, which were thus not tested any further.

### Stage 2: Confirming pooled DNA results by individual genotyping (Group 1 subjects)

One SNP (rs1516793) genotyped by the MassArray iPLEX assay failed to pass filtering quality checks because of poor assay performance, and was thus not included in subsequent data analysis. The remaining 7 SNPs were also designated as S1 to S7 for easy referencing (Figure 1; Tables 3 and

4). The genotypes of these 7 SNPs were in HWE ($P$ >0.05) for Group 1 subjects. The LD among the SNPs was in general very weak (Figure 2) although two LD blocks could be constructed. Single-marker analysis did not reveal any significant difference in allele frequencies between cases and controls ($P_{emp}$ >0.05, Table 3). However, exhaustive sliding-window-based haplotype analysis identified a 2-SNP window that showed significant difference in haplotype frequencies between cases and controls ($P_{asym}$=0.0002 and $P_{emp}$=0.0017 for S6..S7 or rs3784757-rs1516794; Tables 4 and 5). Of all 28 possible sliding windows, 16 sliding windows gave $P_{asym}$ values ≤0.05, but only the S6..S7 window showed association with high myopia after correction for multiple comparisons (Table 4). Therefore, these two SNPs were further tested in the third stage of the study (replication study) using Group 2 subjects. The remaining five SNPs were dropped from further examination.

**Stage 3: Replication study using Group 2 subjects**

For rs3784757 and rs1516794, the genotypes of Group 2 subjects were in HWE ($P$ >0.05). No significant difference in allele frequencies (Table 3) and haplotype frequencies (Table 5) was revealed between cases and controls. Therefore, the initial positive results obtained with Group 1 subjects were not substantiated in a second group of subjects of the same ethnicity. Further analysis was performed by combining both subject groups (656 cases and 654 controls in total). Significant difference between cases and controls in frequencies of alleles and haplotypes was still not detected (Tables 3 and 5).

## DISCUSSION

Six proteoglycan genes *ACAN*, *FMOD*, *DCN*, *LUM*, *KERA* and *EPYC* were selected for study because of their involvement in induced myopia in animal models and/or being within the *MYP3* interval identified by linkage analysis of families with high myopia.[4-12] In particular, *ACAN* and *FMOD* were selected as biological candidate genes, *KERA* and *EPYC* as positional candidate genes, and *DCN* and *LUM* as biological and positional candidate genes. Association studies of candidate genes selected on the basis of biological and/or positional information have meet with different levels of success.

However, there are indeed examples of myopia susceptibility genes identified from each approach: *TGFB1* from biology-based approach,[17,19,20] *MFN1*, *SOX2OT* and *PSARL* from position-based approach (*MYP8* locus);[21] and *HGF* and *PAX6* from a combined approach based on both biological and positional information.[22-27]

Sixty tSNPs were selected from the six selected genes (Table 1), 58 tSNPs were successfully screened by DNA pooling strategy (Table 2), and 8 tSNPs ($P \leq 0.1$, nested ANOVA, Table 2) were followed up by individual genotyping of 300 cases and 300 controls. Although single-marker analysis did not reveal any significant results (Table 3), a sliding-window-based haplotype analysis identified significant difference between cases and controls in the frequencies of haplotypes consisting of rs3784757 and rs1516794 (S6..S7, Table 4). However, attempt to validate these findings with a second subject group (356 cases and 354 controls) failed to replicate the initial positive results. In other words, these six proteoglycan genes were not associated with high myopia in the Chinese population under study, and are therefore unlikely to make a major contribution to the genetic predisposition to high myopia.

To exclude type II error as a possible explanation for the lack of positive association results, we examine the power of our study. To calculate the power of the first stage of the study (screening tSNPs by DNA pooling approach), an online calculator for power and sample size (http://www.stat.uiowa.edu/~rlenth/Power/) is used. In our nested ANOVA model,[16] the subject group was a fixed-effects factor with 2 levels (case and control) while the DNA pool was a random-effects factor with 6 levels (6 DNA pools per subject group). The technical replicate measurement also assumed a random effect. For random effects and from the analysis output of the STATA package, the effect size expressed as the square root of variance component was *on average* 0.0356 for the factor DNA pool, and 0.0099 for the technical replicate measurements. An allele frequency difference of 0.015 (1.5%) between the subject groups was translated into an effect size of 0.0441 for the fixed-effects factor subject group and calculated as the square root of the sum of squares (from STATA

analysis output) divided by the degree of freedom (df=1). With the significance level set at $\alpha$=0.10 for this screening stage, a SNP showing an allele frequency difference of 1.5% between the subject groups could be detected with a power of 89% (last row of Supplementary Table S3) and hence would be followed up in the second stage by individual genotyping. We tested 7 SNPs in the second stage of the study. We assume a prevalence of 0.05 for high myopia in the general Chinese population of Hong Kong,[28] and a Bonferroni-adjusted significance of 0.0071 for 7 SNPs, which is much more conservative than the permutation tests used in the data analysis. Under a log-additive model as examined using the QUANTO package (version 1.2.4),[29] a sample size of 300 cases and 300 controls would give a power of ~80% for an odds ratio of 1.85 and a minor allele frequency of 0.10 (Supplementary Table S4). We tested 2 SNPs in the third stage of the study. With similar assumptions and a Bonferroni-adjusted significance of 0.025 for 2 SNPs, a sample size of 356 cases and 356 controls would give a power of ~80% for an odds ratio of 1.65 and a minor allele frequency of 0.10 (Supplementary Table S4). In other words, the power is at least 80% for all parts of the study under reasonable assumptions and using the empirical data from the study.

While the role of *ACAN* and *KERA* common polymorphisms in high myopia was investigated for the first time in this study, the other four genes have been examined previously in relation to high myopia in smaller studies. One group examined one *FMOD* SNP (rs7543418), but did not find any association in a study involving 195 Chinese cases (SE ≤-6.5 D) and 94 controls.[30] With a DNA pooling approach, we examined nine tSNPs from *FMOD* and did not find any association for rs7543418 either (Table 2). Another group explored four *DCN* and four *EPYC* polymorphisms with only four (rs2070985 of *DCN*, and rs1920748, rs1920751 and rs1920752 of *EPYC*) being polymorphic in the Chinese population under study, and these four polymorphic markers were found to be *not* associated with high myopia in a study of 120 cases (SE ≤-10.0D) and 137 controls.[31] Two *DCN* and three *EPYC* tSNPs were screened in the present study, and found to be *not* associated with high myopia either (Table 2). We did not examine rs2070985 of *DCN* and rs1920748 of *EPYC* because their MAFs (0.089 and 0.081, respectively) documented in HapMap database for Han

Chinese are less than the selection threshold (MAF ≥0.10) for our study. We examined rs10859081 of *EPYC* (Table 2), which is in perfect linkage disequilibrium (i.e., $r^2$=1; http://www.hapmap.org/) with rs1920751 and rs1920752. One Taiwanese group examined five SNPs located in either the promoter or the 3' untranslated region of the *LUM* gene in 201 cases of high myopia (mean SE ≤-6.0D) and 86 controls (mean SE within ±0.5D), and found that one SNP in 3' untranslated region (c.1567C>T) was associated with high myopia ($P$ = 0.0036 for allelic test and 0.0016 for genotypic test).[32] However, this SNP was not documented in HapMap database and hence not examined in the present study. Another Taiwanese group investigated 8 SNPs in the *LUM* gene in 120 cases of high myopia (≤-10.0D) and 137 controls (-1.5D to +0.5D), and found that rs3759223 showed a significant association with high myopia ($P = 2.83\times10^{-4}$).[31] Nonetheless, this association was not substantiated in two other studies of Chinese subjects.[32,33] We did not examine rs3759223. Instead, rs2300588 was genotyped in the present study, but did not pass the initial screen by DNA pools ($P$ = 0.4250, nested ANOVA; Table 2). Note that rs2300588 is in strong LD with rs3759223 ($r^2$=0.773). This discrepancy may be due to the use of different thresholds for defining high myopia in different studies: -10.0D in the positive study,[31] -8.0D in the present study, and -6.0 D in the other two negative studies.[32,33]

We focused on common polymorphisms of these six candidate genes, but did not explore the role of their rare variants in high myopia. A few studies did search by DNA sequence analysis for rare causal variants in the exons of *FMOD* or *EPYC* in small numbers of high myopes, but without fruitful results.[34-36]

DNA pooling *cannot* be used for screening X-linked candidate genes like *BGN*. In addition, it makes haplotype analysis extremely difficult, if not impossible.[14] However, DNA pooling strategy offers a very cost-effective initial screen of SNPs for follow-up studies.[14] It has also been proposed to be used in the initial phase of genome-wide association studies[37] and in re-sequencing studies for rare variants to make the latter two approaches even more affordable.[38]

In conclusion, we used an efficient three-stage approach to examining the relationship between high myopia and six candidate proteoglycan genes (*ACAN*, *FMOD*, *DCN*, *LUM*, *KERA* and *EPYC*). In the second stage, haplotypes consisting of 2 *ACAN* SNPs (rs3784757 and rs1516794) were found significantly associated with high myopia. However, the initial positive result failed to be replicated in the third stage in a second subject group. Therefore, these six proteoglycan genes are unlikely to play a major role in the genetic susceptibility to high myopia.

**Acknowledgements**

The study is a contribution from the Centre for Myopia Research, The Hong Kong Polytechnic University. The authors would like to thank all subjects taking part in the Myopia Genetics Study.

**REFERENCES**

1. Young TL. Myopia. In: Levin LA, Albert DM, eds. *Ocular Disease: Mechanisms and Management*. Philadelphia: Saunders Elsevier 2010:424-432.
2. McBrien NA, Gentle A. Role of the sclera in the development and pathological complications of myopia. *Prog Retin Eye Res.* 2003;22:307-338.
3. Rada JA, Shelton S, Norton TT. The sclera and myopia. *Exp Eye Res.* 2006;82:185-200.
4. Rada JA, Matthews AL, Brenza H. Regional proteoglycan synthesis in the sclera of experimentally myopic chicks. *Exp Eye Res.* 1994;59:747-760.
5. Rada JA, Achen VR, Rada KG. Proteoglycan turnover in the sclera of normal and experimentally myopic chick eyes. *Invest Ophthalmol Vis Sci.* 1998;39:1990-2002.
6. Marzani D, Wallman J. Growth of the two layers of the chick sclera is modulated reciprocally by visual conditions. *Invest Ophthalmol Vis Sci.* 1997;38:1726-1739.
7. McBrien NA, Lawlor P, Gentle A. Scleral remodeling during the development of and recovery from axial myopia in the tree shrew. *Invest Ophthalmol Vis Sci.* 2000;41:3713-3719.

8. Siegwart JT Jr, Strang CE. Selective modulation of scleral proteoglycan mRNA levels during minus lens compensation and recovery. *Mol Vis.* 2007;13:1878-1886.
9. Rada JA, Nickla DL, Troilo D. Decreased proteoglycan synthesis associated with form deprivation myopia in mature primate eyes. *Invest Ophthalmol Vis Sci.* 2000;41:2050-2058.
10. Siegwart JT Jr, Norton TT. Steady state mRNA levels in tree shrew sclera with form-deprivation myopia and during recovery. *Invest Ophthalmol Vis Sci.* 2001;42:1153-1159.
11. Chakravarti S, Paul J, Roberts L, Chervoneva I, Oldberg A, Birk DE. Ocular and scleral alterations in gene-targeted lumican-fibromodulin double-null mice. *Invest Ophthalmol Vis Sci.* 2003;44:2422-2432.
12. Young TL, Ronan SM, Alvear AB, et al. A second locus for familial high myopia maps to chromosome 12q. *Am J Hum Genet.* 1998;63:1419-1424.
13. Tang WC, Yap MK, Yip SP. A review of current approaches to identifying human genes involved in myopia. *Clin Exp Optom.* 2008;91:4-22.
14. Sham P, Bader JS, Craig I, O'Donovan M, Owen M. DNA Pooling: a tool for large-scale association studies. *Nat Rev Genet.* 2002;3:862-871.
15. Zha Y, Leung KH, Lo KK, *et al*. TGFB1 as a susceptibility gene for high myopia: a replication study with new findings. *Arch Ophthalmol.* 2009;127:541-548.
16. Leung KH, Yiu WC, Yap MK, et al. Systematic investigation of the relationship between high myopia and polymorphisms of the MMP2, TIMP2 and TIMP3 genes by a DNA pooling approach. *Invest Ophthalmol Vis Sci.* 2011;52:3893-3900.
17. Hoogendoorn B, Norton N, Kirov G, et al. Cheap, accurate and rapid allele frequency estimation of single nucleotide polymorphisms by primer extension and DHPLC in DNA pools. *Hum Genet.* 2000;107:488-493.
18. Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. *Am J Hum Genet.* 2005;76:887-893.
19. Lin HJ, Wan L, Tsai Y, et al. The TGFbeta1 gene codon 10 polymorphism contributes to the genetic predisposition to high myopia. *Mol Vis*. 2006;12:698-703.

20. Khor CC, Fan Q, Goh L, et al. Support for TGFB1 as a susceptibility gene for high myopia in individuals of Chinese descent. *Arch Ophthalmol*. 2010;128:1081-1084.
21. Andrew T, Maniatis N, Carbonaro F, et al. Identification and replication of three novel myopia common susceptibility gene loci on chromosome 3q26 using linkage and linkage disequilibrium mapping. *PLoS Genet*. 2008;4:e1000220.
22. Han W, Yap MK, Wang J, Yip SP. Family-based association analysis of hepatocyte growth factor (HGF) gene polymorphisms in high myopia. *Invest Ophthalmol Vis Sci*. 2006;47:2291-2299.
23. Yanovitch T, Li YJ, Metlapally R, Abbott D, Viet KN, Young TL. Hepatocyte growth factor and myopia: genetic association analyses in a Caucasian population. *Mol Vis*. 2009;15:1028-1035.
24. Veerappan S, Pertile KK, Islam AF, et al. Role of the hepatocyte growth factor gene in refractive error. *Ophthalmology*. 2010;117:239-245.
25. Han W, Leung KH, Fung WY, et al. Association of PAX6 polymorphisms with high myopia in Han Chinese nuclear families. *Invest Ophthalmol Vis Sci*. 2009;50:47-56.
26. Jiang B, Yap MK, Leung KH, et al. PAX6 haplotypes are associated with high myopia in Han Chinese. *PLoS One*. 2011;6:e19587.
27. Liang CL, Hsi E, Chen KC, Pan YR, Wang YS, Juo SH. A functional polymorphism at 3'UTR of the PAX6 gene may confer risk for extreme myopia in the Chinese. *Invest Ophthalmol Vis Sci*. 2011;52:3500-3505.
28. Goh WS, Lam CS. Changes in refractive trends and optical components of Hong Kong Chinese aged 19-39 years. *Ophthalmic Physiol Opt*. 1994;14:378-382.
29. Gauderman WJ. Sample size requirements for association studies of gene-gene interaction. *Am J Epidemiol*. 2002;155:478-484.
30. Lin HJ, Wan L, Tsai Y, *et al*. Sclera-related gene polymorphisms in high myopia. *Mol Vis.* 2009;15:1655-1663.
31. Chen ZT, Wang IJ, Shih YF, Lin LL. The association of haplotype at the lumican gene with high myopia susceptibility in Taiwanese patients. *Ophthalmology.* 2009;116:1920-1927.

32. Lin HJ, Kung YJ, Lin YJ, et al. Association of the lumican gene functional 3'-UTR polymorphism with high myopia. *Invest Ophthalmol Vis Sci.* 2010;51:96-102.
33. Wang P, Li S, Xiao X, et al. High myopia is not associated with the SNPs in the TGIF, LUM, TGFB1, and HGF genes. *Invest Ophthalmol Vis Sci.* 2009;50:1546–1551.
34. Paluru PC, Scavello GS, Ganter WR, Young TL. Exclusion of lumican and fibromodulin as candidate genes in MYP3 linked high grade myopia. *Mol Vis.* 2004;10:917-922.
35. Majava M, Bishop PN, Hägg P, *et al*. Novel mutations in the small leucine-rich repeat protein/proteoglycan (SLRP) genes in high myopia. *Hum Mutat.* 2007;28:336-344.
36. Wang P, Li S, Xiao X, Guo X, Zhang Q. An evaluation of OPTC and EPYC as candidate genes for high myopia. *Mol Vis.* 2009;15:2045-2049.
37. Macgregor S, Zhao ZZ, Henders A, Nicholas MG, Montgomery GW, Visscher PM. Highly cost-efficient genome-wide association studies using DNA pools and dense SNP arrays. *Nucleic Acids Res.* 2008;36:e35.
38. Wang T, Lin CY, Rohan TE, Ye K. Resequencing of pooled DNA for detecting disease associations with rare variants. *Genet Epidemiol.* 2010;34:492-501.

**FIGURE LEGEND**

**Figure 1.** The structure of the aggrecan (*ACAN*) gene and the seven single nucleotide polymorphisms (SNPs) tested in the second stage of the study. The top panel shows the alternative exon-intron organizations of the *ACAN* gene taken from UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway) together with their corresponding positions on chromosome 15 based on the GRCh37/hg19 human reference sequence assembly. The physical positions of the seven intronic SNPs tested in the second stage of the study are shown in the bottom panel.

**Figure 2.** The linkage disequilibrium (LD) pattern of seven single nucleotide polymorphisms (SNPs) of the *ACAN* gene. The SNPs are indicated in the 5'>3' direction (left>right) along the sense strand of the gene. Shown are the LD measures expressed as D' and $r^2$ for all subjects of Group 1, and calculated by Haploview. The shades of gray represent the magnitude of the LD measures.

**Table 1.** Features of candidate genes and their tag single nucleotide polymorphisms (tSNPs)

| Gene | Gene name | GeneID | Chromosome location | Region captured* | No. of tSNPs | No. of SNPs captured at $r^2$=? |
|---|---|---|---|---|---|---|
| *ACAN* | Aggrecan | 176 | 15q26.1 | 76.8 kb | 41 | 106 ($r^2$=0.964) |
| *FMOD* | Fibromodulin | 2331 | 1q32 | 16.5 kb | 9 | 23 ($r^2$=0.955) |
| *DCN* | Decorin | 1634 | 12q21.33 | 43.8 kb | 2 | 3 ($r^2$=1.000) |
| *LUM* | Lumican | 4060 | 12q21.3-22 | 14.3 kb | 3 | 15 ($r^2$=0.982) |
| *KERA* | Keratocan | 11081 | 12q22 | 13.9 kb | 2 | 17 ($r^2$=1.000) |
| *EPYC* | Epiphycan | 1833 | 12q21 | 47.3 kb | 3 | 31 ($r^2$=1.000) |

* The region captured includes the gene and its 6-kb adjoining region (3 kb upstream and 3 kb downstream).

**Table 2.** Pooled DNA analysis of tag SNPs in the *ACAN*, *FMOD*, *DCN*, *LUM*, *KERA* and *EPYC* genes

| Candidate gene | SNP* | Alleles† ($1^{st}/2^{nd}$) | k correction factor peak height ratio ($1^{st}/2^{nd}$) | Estimated freq of $1^{st}$ allele in DNA pools | | | Nested ANOVA |
|---|---|---|---|---|---|---|---|
| | | | | Case | Control | Diff (Case - Control) | *P* value |
| ***ACAN*** | rs2203642 | G/A | 0.8991 | 0.4179 | 0.4407 | -0.0228 | 0.3202 |
| | **rs12439075‡** | **C/T** | **1.0880** | **0.4051** | **0.4461** | **-0.0410** | **0.0862** |
| | rs8033375 | C/T | 0.6940 | 0.3178 | 0.329 | -0.0112 | 0.5053 |
| | rs4932429 | G/C | 1.1401 | 0.8991 | 0.8689 | 0.0302 | 0.1916 |
| | rs11858871 | G/T | 0.9799 | 0.1593 | 0.1946 | -0.0353 | 0.1479 |
| | rs17199220 | G/A | 0.9069 | 0.8331 | 0.8122 | 0.0209 | 0.5378 |
| | rs16942248 | T/A | 1.0196 | 0.1721 | 0.1714 | 0.0007 | 0.9739 |
| | rs12905259 | G/A | 0.9024 | 0.5185 | 0.4965 | 0.0220 | 0.2407 |
| | rs12905452 | A/C | 1.0506 | 0.5257 | 0.5448 | -0.0191 | 0.3009 |
| | rs7179602 | T/A | 2.6026 | 0.2653 | 0.2769 | -0.0116 | 0.4223 |
| | rs4932433 | A/T | 1.1017 | 0.3360 | 0.3302 | 0.0058 | 0.8087 |
| | rs16942277 | G/T | 0.9047 | 0.7090 | 0.7004 | 0.0086 | 0.6283 |
| | rs4932434 | C/A | 0.9939 | 0.5989 | 0.5979 | 0.0010 | 0.9605 |
| | rs939586 | G/A | 0.9888 | 0.2369 | 0.2399 | -0.0030 | 0.9134 |
| | rs11073814 | A/T | 1.0536 | 0.8451 | 0.8442 | 0.0009 | 0.9688 |
| | rs8040435 | T/A | 1.0101 | 0.5917 | 0.5908 | 0.0009 | 0.9722 |
| | **rs2280468‡** | **G/A** | **0.9654** | **0.7338** | **0.6974** | **0.0364** | **0.0762** |
| | rs1015081 | G/A | 0.8835 | 0.1868 | 0.1748 | 0.0120 | 0.6409 |
| | rs1015080 | G/A | 0.8955 | 0.1252 | 0.1521 | -0.0269 | 0.1928 |
| | rs883325 | G/T | 1.0082 | 0.2962 | 0.2804 | 0.0158 | 0.4727 |
| | rs4932435 | G/T | 0.9796 | 0.5223 | 0.508 | 0.0143 | 0.5982 |
| | **rs2293087‡** | **G/T** | **0.8570** | **0.3109** | **0.2737** | **0.0372** | **0.0338** |
| | **rs4932438‡** | **C/T** | **1.1072** | **0.2270** | **0.2818** | **-0.0548** | **0.0142** |
| | rs3743398 | C/T | 0.8793 | 0.8990 | 0.9037 | -0.0047 | 0.7613 |
| | rs938608 | G/T | 0.8455 | 0.6948 | 0.7125 | -0.0177 | 0.4717 |
| | rs4932439 | G/A | 0.6748 | 0.4747 | 0.4896 | -0.0149 | 0.5029 |
| | rs1042631 | C/T | 0.9011 | 0.6006 | 0.6211 | -0.0205 | 0.4229 |
| | rs698621 | G/T | 0.9484 | 0.4812 | 0.4461 | 0.0351 | 0.1423 |
| | **rs953065‡** | **C/T** | **0.8001** | **0.4959** | **0.5358** | **-0.0399** | **0.0785** |
| | **rs3784757‡** | **G/A** | **0.9776** | **0.8457** | **0.8201** | **0.0256** | **0.0346** |
| | **rs1516793‡** | **G/A** | **0.9921** | **0.1391** | **0.1606** | **-0.0215** | **0.0982** |
| | **rs1516794‡** | **T/A** | **1.0360** | **0.0811** | **0.113** | **-0.0319** | **0.0542** |
| | rs1516797 | G/T | 0.9280 | 0.329 | 0.3207 | 0.0083 | 0.6185 |
| | rs1879529 | G/T | 0.9867 | 0.7336 | 0.714 | 0.0196 | 0.2597 |
| | rs3817428 | C/G | 1.0224 | 0.8572 | 0.8469 | 0.0103 | 0.5798 |
| | rs16942409 | G/T | 0.8121 | 0.2824 | 0.2786 | 0.0038 | 0.8381 |
| | rs7163146 | A/T | 0.5858 | 0.1443 | 0.1576 | -0.0133 | 0.6233 |
| | rs2280465 | G/A | 0.7680 | 0.8701 | 0.873 | -0.0029 | 0.8892 |
| | rs8031741 | G/A | 0.9224 | 0.1496 | 0.1513 | -0.0017 | 0.9627 |
| ***FMOD*** | rs10920617 | C/T | 0.8611 | 0.7186 | 0.6885 | 0.0301 | 0.1331 |
| | rs10920615 | C/T | 0.6669 | 0.5774 | 0.5628 | 0.0146 | 0.2744 |
| | rs7543148 | G/A | 0.9589 | 0.3008 | 0.3225 | -0.0217 | 0.3377 |
| | rs10800913 | C/T | 0.8889 | 0.2549 | 0.2591 | -0.0042 | 0.8300 |
| | rs2105309 | C/T | 0.9142 | 0.6212 | 0.6244 | -0.0032 | 0.8281 |
| | rs3766913 | G/A | 0.9426 | 0.8716 | 0.8773 | -0.0057 | 0.6797 |
| | rs3820224 | G/A | 0.9440 | 0.8088 | 0.7654 | 0.0434 | 0.1191 |
| | rs2886220 | G/A | 0.8671 | 0.4865 | 0.4965 | -0.0100 | 0.6271 |
| | rs16851319 | C/G | 0.7840 | 0.2402 | 0.1986 | 0.0416 | 0.1072 |
| ***DCN*** | rs3138295 | G/A | 0.6552 | 0.2696 | 0.2428 | 0.0268 | 0.4038 |
| | rs566806 | C/T | 0.9447 | 0.6043 | 0.6039 | 0.0004 | 0.9851 |
| ***LUM*** | rs3759222 | C/A | 0.9089 | 0.6220 | 0.6486 | -0.0266 | 0.4192 |
| | rs10859110 | C/T | 0.9339 | 0.5868 | 0.5793 | 0.0075 | 0.7974 |
| | rs2300588 | G/T | 0.7815 | 0.3538 | 0.3345 | 0.0193 | 0.4250 |
| ***KERA*** | rs2041711 | G/T | 0.9056 | 0.8214 | 0.8109 | 0.0105 | 0.6897 |
| | rs2268579 | G/A | 0.5645 | 0.1430 | 0.1324 | 0.0106 | 0.6827 |
| ***EPYC*** | rs11105899 | T/A | 1.1288 | 0.8827 | 0.8558 | 0.0269 | 0.5534 |
| | rs11105898 | C/T | 1.1958 | 0.4723 | 0.4970 | -0.0247 | 0.5943 |
| | rs10859081 | G/A | 1.0734 | 0.1850 | 0.1826 | 0.0024 | 0.9200 |

* SNPs are arranged down the column in the 5'>3' order along the sense strand of the gene concerned.

† The $1^{st}$ allele is eluted first (shorter elution) while the $2^{nd}$ allele is eluted last (longer elution time). The alleles are named with reference to the sense strand of the respective gene.

‡ SNPs with *P* value ≤0.1 are highlighted in **boldface**, and are followed up by individual genotyping in the second stage of the study.

**Table 3.** Allelic association tests of *ACAN* SNPs genotyped individually

| SNP | Alleles* 1 | 2 | Genotype counts (11/12/22)* Cases | Controls | Minor allele freq Cases | Controls | OR (95%CI) † | Allelic test‡ $P_{asym}$ | $P_{emp}$ |
|---|---|---|---|---|---|---|---|---|---|
| **Group 1 subjects** | | | | | | | | | |
| rs12439075 (S1) | T | C | 118/134/44 | 106/140/52 | 0.3750 | 0.4094 | 0.87 (0.69 - 1.11) | 0.2667 | 0.9319 |
| rs2280468 (S2) | G | A | 162/104/14 | 143/123/15 | 0.2357 | 0.2823 | 0.78 (0.58 - 1.04) | 0.0879 | 0.5690 |
| rs2293087 (S3) | T | G | 147/130/21 | 170/117/10 | 0.2886 | 0.2306 | 1.40 (1.05 - 1.86) | 0.0224 | 0.1936 |
| rs4932438 (S4) | T | C | 198/83/16 | 172/103/22 | 0.1936 | 0.2475 | 0.71 (0.53 - 0.93) | 0.0148 | 0.1333 |
| rs953065 (S5) | C | T | 79/141/68 | 89/138/64 | 0.4809 | 0.4570 | 1.11 (0.88 - 1.41) | 0.3732 | 0.9801 |
| rs3784757 (S6) | G | A | 243/53/3 | 232/62/5 | 0.0987 | 0.1204 | 0.74 (0.51 - 1.08) | 0.1234 | 0.6957 |
| rs1516794 (S7) | A | T | 246/34/0 | 241/47/3 | 0.0607 | 0.0911 | 0.60 (0.37 - 0.96) | 0.0326 | 0.2717 |
| **Group 2 subjects** | | | | | | | | | |
| rs3784757 (S6) | G | A | 281/70/3 | 282/67/4 | 0.1073 | 0.1062 | 1.02 (0.72 - 1.43) | 0.9213 | 0.9916 |
| rs1516794 (S7) | A | T | 304/50/0 | 302/47/1 | 0.0706 | 0.0700 | 1.02 (0.67 - 1.55) | 0.9373 | 0.9960 |
| **Combined (Groups 1 & 2)** | | | | | | | | | |
| rs3784757 (S6) | G | A | 524/123/6 | 514/129/9 | 0.1034 | 0.1127 | 0.90 (0.70 - 1.15) | 0.3998 | 0.6024 |
| rs1516794 (S7) | A | T | 550/84/0 | 543/94/4 | 0.0662 | 0.0796 | 0.82 (0.60 - 1.11) | 0.1976 | 0.3389 |

* Alleles 1 and 2 refer to the major and the minor alleles, respectively; and the genotype counts are indicated in the order of 11, 12 and 22, respectively. There are 300 cases and 300 controls in Group 1 subjects, and 356 cases and 354 controls in Group 2 subjects. Since some samples failed to be genotyped in a random manner, the total counts of genotypes may not add up to the expected numbers.

† The odds ratio (OR) is calculated for allele 2 (minor allele) with allele 1 (major allele) as the reference. Within brackets are the 95% confidence intervals (CI).

‡ Allele frequencies are compared by logistic regression with adjustment for sex and age as covariates to give the asymptotic *P* values ($P_{asym}$). Multiple comparisons are corrected by 10,000 permutations across single marker allelic tests (this table) and omnibus tests of haplotypes for Group 1 subjects (Table 4), for Group 2 subjects (Table 5) or Combined (Groups 1 and 2, Table 5), each separately.

**Table 4.** Summary of exhaustive haplotype analyses based on sex- and age-adjusted omnibus tests for sliding windows of all possible sizes across 7 *ACAN* SNPs genotyped individually for Group 1 subjects*

| Sliding window (SW) | | Most significant omnibus test | | |
|---|---|---|---|---|
| SNPs/SW | No. of SW | SW | $P_{asym}$ | $P_{emp}$ |
| 1 | 7 | S4 | 0.0148 | 0.1333 |
| 2 | 6 | **S6..S7** | **0.0002** | **0.0017** |
| 3 | 5 | S4..S6 | 0.0078 | 0.0753 |
| 4 | 4 | S3..S6 | 0.0088 | 0.0829 |
| 5 | 3 | S3..S7 | 0.0020 | 0.1782 |
| 6 | 2 | S2..S7 | 0.0156 | 0.1400 |
| 7 | 1 | S1..S7 | 0.0152 | 0.1365 |

* Plink was used to compare haplotypes between cases and controls in all sliding windows (SW) of all possible window sizes (the number of SNPs per SW; 1 to 7 SNPs per SW). With logistic regression and adjustment for sex and age as covariates, a single omnibus test of (n-1) degrees was performed for each SW to jointly assess the significance of the haplotype effects for this SW, where n is the number of haplotypes for the specific SW being considered. Hence, a single asymptotic *P* value ($P_{asym}$) is generated for each SW. For a given window size, the omnibus test was conducted for all possible windows of the same size, shifting one SNP at a time. There were 28 windows in total (the sum of numbers in the second column from the left), and multiple comparisons were corrected by performing 10,000 permutations to produce an empirical *P* value ($P_{emp}$). The SW is shown as Sx..Sy, where x is the first SNP and y the last SNP of the SW, and the SW S3-S4-S5-S6, for instance, is indicated as S3..S6. The rightmost columns show the most significant omnibus tests for each fixed-size SW. Note that only S6..S7 remained significant after multiple testing correction ($P_{emp}$ = 0.0017; indicated in **boldface**). Please refer to Table 3 for the identities of the SNPs S1 to S7.

**Table 5.** Haplotype analysis of two *ACAN* SNPs (rs3784757- rs1516794)*

| | Haplotype freq | | Logistic regression adjusted for sex & age | | |
|---|---|---|---|---|---|
| Haplotype | Cases | Controls | OR | $P_{asym}$ | $P_{emp}$ |
| **Group 1 subjects** | | | | | |
| Omnibus test | – | – | – | **0.0002** | **0.0017** |
| AT | 0.0373 | 0.0871 | **0.37** | **0.0005** | |
| GT | 0.0237 | 0.0026 | **12.50** | **0.0119** | |
| AA | 0.0595 | 0.0370 | 1.61 | 0.1150 | |
| GA | 0.8795 | 0.8733 | 1.13 | 0.5100 | |
| **Group 2 subjects** | | | | | |
| Omnibus test | – | – | – | 0.9940 | 1.0000 |
| AT | 0.0697 | 0.0689 | 1.02 | 0.9280 | |
| AA | 0.0370 | 0.0373 | 0.98 | 0.9550 | |
| GA | 0.8933 | 0.8938 | 0.99 | 0.9660 | |
| **Combined (Groups 1 & 2)** | | | | | |
| Omnibus test | – | – | – | 0.2670 | 0.4369 |
| AT | 0.0561 | 0.0776 | 0.69 | 0.0238 | |
| AA | 0.0473 | 0.0368 | 1.26 | 0.2510 | |
| GA | 0.8966 | 0.8856 | 1.03 | 0.8020 | |

* Asymptotic *P* values ($P_{asym}$) are obtained from Wald test based on logistic regression. Multiple comparisons are corrected by 10,000 permutations across single marker allelic tests (Table 3) and omnibus tests of haplotypes (this table) for Group 2 subjects or Combined groups, each separately. For Group 1 subjects, the result is an expansion of the sliding window S6..S7 shown in Table 4, and the correction for multiple comparisons was performed for all single-marker allelic tests (Table 3) and haplotype tests of all sliding windows (Table 4). The empirical *P* value is indicated as $P_{emp}$. Note that Plink does not generate confidence intervals for odds ratios (OR) in haplotype analysis.

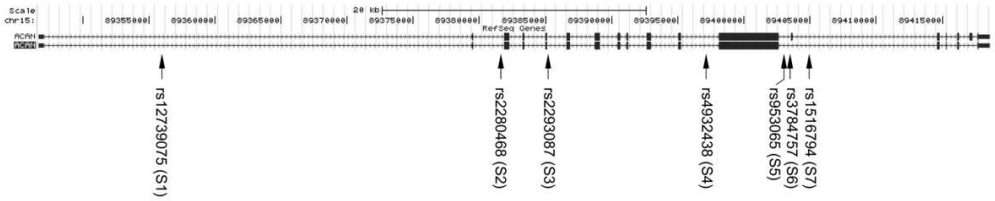


Figure 1. The structure of the aggrecan (*ACAN*) gene and the seven single nucleotide polymorphisms (SNPs) tested in the second stage of the study. The top panel shows the alternative exon-intron organizations of the *ACAN* gene taken from UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway) together with their corresponding positions on chromosome 15 based on the GRCh37/hg19 human reference sequence assembly. The physical positions of the seven intronic SNPs tested in the second stage of the study are shown in the bottom panel.
50x10mm (600 x 600 DPI)

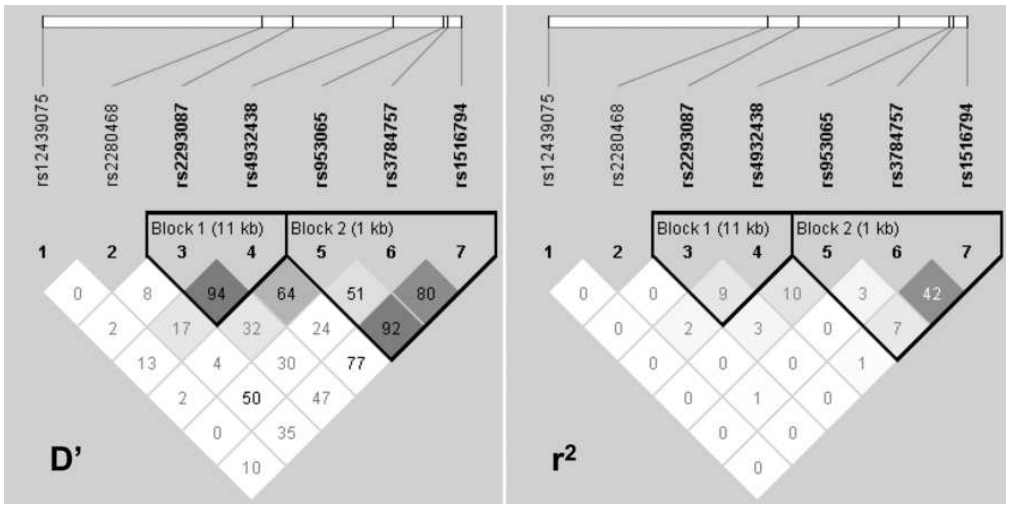


Figure 2. The linkage disequilibrium (LD) pattern of seven single nucleotide polymorphisms (SNPs) of the *ACAN* gene. The SNPs are indicated in the 5'>3' direction (left>right) along the sense strand of the gene. Shown are the LD measures expressed as D' and $r^2$ for all subjects of Group 1, and calculated by Haploview. The shades of gray represent the magnitude of the LD measures.
74x37mm (300 x 300 DPI)