Article section: Clinical/Epidemiologic Research

Systematic investigation of the relationship between high myopia and polymorphisms of the *MMP2*, *TIMP2* and *TIMP3* genes by a DNA pooling approach

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

[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 Hong Kong of University, Hong Kong SAR, China.G

Running title: Myopia candidate genes by a DNA pooling approach

Corresponding author: Shea Ping Yip, Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong SAR, China.

E-mail: shea.ping.yip@inet.polyu.edu.hk

Word count: 4,459 words (Main text)

ABSTRACT

PURPOSE. This study examined the relationship between high myopia and three myopia candidate genes, namely matrix metalloproteinase 2 (*MMP2*) and tissue inhibitor of metalloproteinase 2 and 3 (*TIMP2* and *TIMP3*), involved in scleral remodeling.

METHODS. The study recruited unrelated adult Han Chinese who were high myopes (spherical equivalent ≤-6.0D for both eyes; cases) and emmetropes (within ±1.0D for both eyes; controls). Sample Set 1 had 300 cases and 300 controls while Sample Set 2 had 356 cases and 354 controls. Forty-nine tag single nucleotide polymorphisms (SNPs) were selected from these candidate genes. The first stage was an initial screen of 6 case pools and 6 control pools constructed from Sample Set 1, each pool consisting of 50 distinct subjects of the same affection status. In the second stage, "positive" SNPs from the first stage were confirmed by genotyping individual samples forming the DNA pools. In the third stage, positive SNPs from stage 2 were replicated with Sample Set 2 genotyped individually.

RESULTS. Of the 49 SNPs screened by DNA pooling, 3 passed the lenient threshold of $P < 0.10$ (nested ANOVA) and were followed up by individual genotyping. Of the 3 SNPs genotyped, two *TIMP3* SNPs were found significantly associated with high myopia by single-marker or haplotype analysis. However, the initial positive results could not be replicated by Sample Set 2.

CONCLUSIONS. *MMP2*, *TIPM2* and *TIMP3* genes were not associated with high myopia in Chinese and hence were unlikely to play a major role in the genetic susceptibility to high myopia.

(250 words)

INTRODUCTION

In myopia, the images of distant objects are focused in front of, rather than on, the retina under relaxed accommodation. Myopia is the commonest eye anomaly in the world and imposes a huge impact on the public health care system and the economy.[1] In particular, subjects with high myopia, usually defined as ≤-6.0 diopters (D), are more prone to ocular degenerative changes such as glaucoma and retinal detachment. Myopia is much more frequent in Orientals (60-80%) than in Caucasians (10-25%) although its prevalence varies with time, the age of the subjects, and the ethnic origin of the population concerned.[2] In Hong Kong, the prevalence is highest (70%) for age 19-39, and then drops after age 40.[3]

Both environmental and genetic factors contribute to myopia although the exact cause of myopia remains to be determined.[4-6] Environmental factors such as lifestyle, schooling, near-work and outdoor activities are known to contribute to differences in the prevalence of myopia. Estimates of heritability are high for refractive error and major ocular components, and shared genes between relative pairs could explain the strong correlation between refractive error and axial length.[7-9]

Myopia mainly results from elongated eyeball caused by accelerated postnatal eye growth, rather than changes in corneal or lens power.[10] During myopia development, the sclera undergoes active remodeling, which involves matrix metalloproteinases (MMPs) and tissue inhibitors of metalloproteinases (TIMPs) – the enzymes involved in the degradation of extracellular matrix. Matrix metalloproteinase 2 (MMP2) is increased in the sclera of the myopic eye induced by form deprivation in chicks when compared to the control eye, and the increased expression has been consistently shown for both the protein[11-14] and the mRNA transcript.[14,15] Increased scleral MMP2 expression in form-deprivation myopia has also been shown in tree shrew at both the protein[16] and the mRNA level,[17,18] and in guinea pig at the protein level.[19] Increased *MMP2* transcript level has also been found in human scleral fibroblasts mechanically stretched in an *in vitro* system,[20] in lens-induced myopia in tree shrew,[21] but not in lens-induced myopia in chick.[22] On the other hand, there is less

extensive study of TIMP expression in induced myopia. TIMP2 expression is found to be reduced in form deprivation myopia in chick[14] and in lens-induced myopia in guinea pig,[19] but at comparable levels as the control eye in lens-induced myopia in both tree shrew[21] and chick.[22] Finally, *TIMP3* transcript level is found to be reduced in lens-induced myopia in tree shrew.[21] These studies did not specifically examine any potential interaction among these three genes.

We used a case-control study approach[23] to examine the relationship between high myopia in a Han Chinese population and the tag single nucleotide polymorphisms (SNPs) of three candidate genes. These three candidate genes were selected for this study because their involvement in scleral remodeling has been confirmed by extensive studies of animal myopia models, as has been summarized above. We performed the case-control study in three stages: (1) initial screen of DNA pools to identify putatively positive SNPs, (2) confirmation of "positive" SNPs by genotyping of individual DNA samples forming the original pools, and (3) replication of positive SNPs by an independent sample set (Fig. 1). The initial DNA pooling step served to reduce the cost and time involved in individual genotyping.[23,24] DNA pools were created by mixing equal amounts of DNA from many individuals sharing the same disease status. Thus, "case pools" were constructed from subjects with high myopia (cases) and "control pools" from emmetropic subjects (controls) in this study. Moreover, we adopted an optimal experimental design in DNA pooling step by creating small DNA pools each constructed from 50 distinct individuals of the same disease status.[25]

METHODS

Subjects and DNA samples

In DNA pooling-based initial study, 600 unrelated Southern Han Chinese subjects (**Sample Set 1**) were recruited: 300 cases of high myopes with spherical equivalent (SE≤−8.00D) in both eyes, and 300 emmetropic controls with SE within ±1.0 D in both eyes. Positive SNPs from the DNA pooling-based initial screen were confirmed by individual genotyping of the original Sample Set 1 and, if confirmed, replicated by testing a second sample set (**Sample Set 2**). Sample Set 2 consisted of 710

unrelated Han Chinese subjects with 356 cases and 354 controls. The same entry criteria were used for subject recruitment of both sample sets. This study was approved by the Human Subjects Ethics Subcommittee of the Hong Kong Polytechnic University, and adhered to the tenets of the Declaration of Helsinki. Signed, informed consents were obtained from all participants. All subjects were recruited from the Optometry Clinic of the Hong Kong Polytechnic University, and collection of blood samples and DNA extraction were performed as described previously.[26]

Construction of DNA pools

For the DNA pooling study, all DNA samples were accurately quantified by a PicoGreen method (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions and diluted to a final concentration of 5 ± 0.3 ng/μl. Equal volumes of DNA solutions were mixed to create DNA pools. Six case pools and six control pools were constructed for Sample Set 1, each consisting of 50 *distinct* individuals of the same disease status.

Selection of tag SNPs

Three candidate genes were selected for study: *MMP2*, *TIMP2* and *TIMP3*. Tagger implemented in Haploview (http://www.broadinstitute.org/haploview/haploview) was used to select tag SNPs with the following setting: pairwise tagging algorithm, $r^2 \geq 0.8$ and minor allele frequency (MAF) of more than 0.1. The selection was based on the Han Chinese genotype data from the International HapMap Project database (release 23a, phase II; http://www.hapmap.org/) for these three loci and their flanking regions (3 kb upstream and 3 kb downstream of the genes). In total, 49 tag SNPs were selected from these 3 genes for analysis by the DNA pooling strategy (Table 1).

Allele frequency estimation in DNA pools

The same protocols were used for all 50 SNPs examined unless stated otherwise. Touchdown PCR was performed in a total volume of 15-μl reaction mixture containing 25 ng of genomic DNA template, 0.1 μM of each primer (Supplementary Table S1) and 1.5 mM $MgCl_2$, 0.2 mM of each dNTP and 0.2 U of HotStarTaq Plus DNA Polymerase (Qiagen, Hilden, Germany) in $1\times$ PCR buffer

provided by the manufacturer. There were a few exceptions: 0.3 μM of each primer was used for 3 SNPs (rs243845, rs11639960 and rs12600817) and 2.5 mM of $MgCl_2$ for 2 SNPs (rs11639960 and rs12600817). Amplification was performed in a GeneAmp PCR system 9700 (Applied Biosystems, Foster City, CA). The touchdown thermocycling program included activation at 95ºC for 5 min, followed by 6 cycles of 95ºC for 30 seconds, 64°C (initial annealing temperature) for 45 seconds and decreased by 1ºC per cycle, and 72 ºC for 45 seconds, plus additional 38 cycles of 95ºC for 30 seconds, 58 ºC (final target annealing temperature) for 45 seconds and 72ºC for 45 seconds; and final extension at 72ºC for 7 minutes. There were a few exceptions: the initial and final annealing temperatures were 62 ºC and 56 ºC for 4 SNPs (rs11643630, rs243845, rs11639960 and rs12600817).

PCR products were purified using shrimp alkaline phosphatase and exonuclease I. Primer extension (PE) reactions were performed in a 25-μl reaction volume containing 10 μl purified PCR products, 1.5 μM of the specific primer (Supplementary Table S1), 50 μM of each appropriate ddNTP and/or dNTP (Supplementary Table S1) and 1 unit of Therminator (New England Biolabs, Beverly, MA) in a 1× reaction buffer supplied by the manufacturer. Thermocycling was performed with an initial denaturation step at 96ºC for 1 minute, followed by 55 cycles of 96ºC for 10 seconds, 43ºC for 15 seconds and 60ºC for 1 minute.

Denaturing high performance liquid chromatography (DHPLC) analysis was performed using the WAVE Nucleic Acid Fragment Analysis System (Transgenomic Inc, Omaha, NE). PE products were analyzed with a 6% linear gradient change of the working elution buffer over a 3-minute period and a different starting concentration of buffer B, dependent on the SNP concerned (Supplementary Table S1).[27] Relative allele frequencies in DNA pools were estimated based on the intensity of primer-extended products by DHPLC. For each DNA pool, the analysis included a single PCR followed by a single PE reaction and a single DHPLC analysis. Each DNA pool was analyzed in triplicates (Fig. 2). In other words, there were 36 sets of readings for 6 case pools and 6 control pools (Sample Set 1) for each SNP.

Individual genotyping

The positive findings (3 SNPs) in the DNA pooling-based initial study were confirmed by individual genotyping of the same sample set (Set 1) with MassArray iPLEX chemistry (Sequenom, San Diego, CA; Supplementary Table S2) according to the manufacturer's instructions (http://www.sequenom.com/). These 3 SNPs were grouped together with SNPs of other on-going studies for genotyping using MassArray iPLEX method carried out by a local service provider (http://genome.hku.hk/portal/). The confirmed positive results were tested by a follow-up replication study on Sample Set 2.

For Sample Set 2, two SNPs of the *TIMP3* gene (rs135029 and rs137485) were genotyped by unlabeled probe melting analysis.[28] This method uses asymmetric PCR to generate single-stranded DNA (ssDNA) product, and an unlabeled probe that is 3'-blocked by a phosphate group to prevent probe extension. After PCR, the unlabeled probe and a saturating dsDNA dye are added to ssDNA target for high-resolution melting analysis. Asymmetric PCR reaction was performed in a 10-μL reaction mixture containing 10 ng of genomic DNA, 3.5 mM $MgCl_2$, 0.1 μM forward primer (excess), 0.01 μM reverse primer (limiting) (Supplementary Table S2), 0.2 mM of each dNTP, 1× PCR buffer and 0.2 U of HotStarTaq Plus DNA Polymerase (Qiagen). Amplification was performed in 96-well plates with a GeneAmp 9700 PCR system, including 1 cycle of initial denaturation for 5 minutes at 95°C, 50 cycles of 30 seconds at 95°C, 50 seconds at 55°C for annealing, and 25 seconds at 72°C, plus 1 cycle of final extension for 5 minutes at 72°C. After PCR, a 10-μL reaction mixture containing 8.4 μl PCR product, 0.5 μM unlabeled probe (International DNA Technologies, Coralville, IA) and 2.5 mM SYTO 9 green fluorescent nucleic acid stain (Invitrogen) was prepared in 96-well white plates. Melting was performed in a LightCycler 480 instrument (Roche Diagnostics, Mannheim, Germany) with heating of samples to 95°C for 1 minute and then cooling to 50°C for 1 minute. The melting data were then collected between 50°C and 95°C with a heating rate of 0.11°C/s at 5 acquisitions per °C, using the "melting curves" analysis mode. Samples were again cooled to 40°C for 10 seconds. Melting curves were analyzed with LightCycler 480 Software (Version 1.5, Roche).

Statistical Analysis

Ocular data were analyzed using the STATA package (version 8.2; StataCorp, College Station, Texas). High myopia was examined as a dichotomous trait. Subjects were classified as affected (cases) or unaffected (controls). In PE, unequal representation of the two alleles of a SNP can result from differential incorporation of ddNTPs, and was corrected with a correction factor (known as k correction factor)[29] estimated based on the average of three independent replicate readings from a heterozygous sample. Relative allele frequencies of a given SNP were estimated from the heights (i.e. intensities) of the two peaks representing the two extension products in the DHPLC elution profile with correction by the k correction factor.[29] The relative allele frequencies of a SNP were compared *between* the pools of the case group and the pools of the control group by nested analysis of variance (nested ANOVA; see online supplementary materials for explanation)[30] carried out using the STATA package. SNPs with *P* value ≤0.10 were followed up by genotyping individual samples forming the DNA pools (Sample Set 1). A lenient threshold of *P* ≤0.10 was used in order not to exclude any potentially significant SNPs.

Genotype data of individual samples (Sample Set 1 or 2) were analyzed by the PLINK package (ver. 1.07; http://pngu.mgh.harvard.edu/~purcell/plink/index.shtml).[31] Hardy-Weinberg equilibrium (HWE) testing was performed with exact test for controls and cases separately. Single-marker association analysis was performed with chi-square test or exact test as appropriate. Haplotype analysis was also performed with logistic regression based on Wald test. Multiple comparisons were corrected by permutation tests. Permutation tests are based on re-sampling theory, and widely accepted as the gold standard for correction of multiple comparisons.[32] In each permutation, the genotype data structure and the numbers of cases and controls were kept unchanged while the phenotype status of the subjects was randomly swapped (permutated). The statistic was calculated with each permutation and an empirical *P* value was generated based on 10,000 permutations. Permutation of the phenotype status among study subjects is valid under the assumption of null

hypothesis. In order to control experiment-wise (instead of marker-wise) type I error rates, each permutation involved all individual SNPs and all haplotypes for a given sample set *genotyped individually* (Sample Set 1, Sample Set 2 or Combined Sets, *each separately*), and this was repeated 10,000 times. As such, the generated empirical *P* values controlled the experiment-wise (or more correctly, family-wise) type I error rates for a given sample set.

RESULTS

Analysis of ocular data

This study had two sample sets collected from Han Chinese in Hong Kong. Sample Set 1 consisted of 300 high myopes (cases) and 300 emmetropes (controls). The characteristics of these subjects have been reported previously.[26] Cases (n=356) and controls (n=354) of Sample Set 2 were recruited using the same entry criteria as Sample Set 1, and their characteristics are summarized in Table 2 with the ocular data being shown for the right eye only, as has been done previously.[26] Subjects of Sample Set 2 were on average older than those of Sample Set 1: 34.0 years for cases and 33.1 years for controls of Set 2 (Table 2); and 27.7 years for cases and 24.9 years for controls of Set 1.[26] However, cases and controls of Set 2 had very similar refractive error and axial length as their counterparts in Set 1. The mean SE for the right eye was -10.30 D for cases and 0.08 D for controls of Set 2, and -10.53 D for cases and 0.03 D for controls of Set 1. The mean axial length of the right eye was 27.64 mm for cases and 23.73 mm for controls of Set 2, and 27.76 mm for cases and 23.85 mm for controls of Set 1.

Analysis of pooled DNA results

Results of pooled DNA analysis are summarized in Table 3. The k correction factor ranged from 0.65 to 1.45 with a mean of 1.07. The estimated frequencies of the first eluted allele ranged from 0.0719 to 0.8848 for case pools, and from 0.0939 to 0.8574 for control pools. The difference (case pools – control pools) in estimated allele frequencies ranged from -0.0597 to 0.0401. Of the 49 SNPs tested by the DNA pooling approach, only 3 SNPs showed significant difference in allele frequencies between case pools and control pools: rs2003241 (difference=0.0329, nested ANOVA *P*=0.0119), rs135029 (difference=-0.0597, *P*=0.0010) and rs137485 (difference=0.0351, *P*=0.0727). These three

SNPs were then genotyped for individual samples forming the DNA pools (Sample Set 1) for confirmation. The remaining 46 SNPs did not show significant differences in allele frequencies between case pools and control pools, and hence were not tested any further (Figure 1).

Confirmation of pooled DNA results by individual genotyping

The genotypes of the three follow-up SNPs were in HWE ($P>0.05$, exact test) for Sample Set 1. The only exception was rs135029 for the case group ($P=0.0250$). Deviation from HWE in cases can signify marker-disease association.[33] Single-marker analysis showed that rs135029 of the *TIMP3* gene was associated with high myopia ($P_{asym}=0.0069$, allelic test) while the other two SNPs (rs2003241 and rs137485) showed no significant differences between cases and controls (Sample Set 1, Table 4). In addition, haplotypes consisting of rs135029 and rs137485 (both in the *TIMP3* gene) were also associated with high myopia ($P_{asym}=0.0178$, omnibus test; Sample Set 1, Table 5). These results remained significant after correction of multiple comparisons across single markers and haplotypes by permutation tests: $P_{emp}=0.0162$ (allelic test, Table 4) and $P_{emp}=0.0496$ (omnibus test, Table 5). Therefore, both rs135029 and rs137485 were further tested in a replication study using Sample Set 2. SNP rs2003241 was not tested any further (Figure 1). Note that the asymptotic P value is indicated as P_{aysm}, and empirical P value as P_{emp} (also see footnotes to Tables 4 and 5).

Replication study based on Sample Set 2

The genotypes of both rs135029 and rs137485 were in HWE ($P>0.05$, exact test). Single-marker and haplotype analyses did not show any significant differences in allele or haplotype frequencies between cases and controls (Sample Set 2, Tables 4 and 5). We combined the sample sets (656 cases and 654 controls in total) and re-analyzed the data with adjustment for age as a covariate because the mean age differed very significantly between Sample Sets 1 and 2 (difference=7.46 years, $P <10^{-4}$ for t test). The results remained the same without significant differences in allele or haplotype frequencies between cases and controls (Combined, Tables 4 and 5). In other words, the initial positive results in Sample Set 1 could not be replicated independently by Sample Set 2.

DISCUSSION

We adopted an efficient three-stage approach to investigating the relationship between high myopia and tag SNPs of three candidate genes (*MMP2*, *TIMP2* and *TIMP3*). There are many experimental studies using animal myopia models which suggest the involvement of these genes in myopia development. In the initial stage, 49 tag SNPs were screened using a DNA pooling approach, and 3 SNPs passed the lenient threshold of $P \leq 0.10$ and were followed up. In the second stage, these three "putatively positive" SNPs were genotyped for individual samples forming the original DNA pools. In the third stage, two SNPs from stage 2 were genotyped for individual samples from a second sample set. However, the initial positive results could not be substantiated in the replication study. It is interesting to note that rs135029 of *TIMP3* gave an OR of 1.26 for the combined sample set (P_{asym}=0.0344, Table 4), but did not survive after correction for multiple comparisons (P_{emp}=0.0693, Table 4). In view of this borderline significance, we explored the potential functional role of this SNP in the literature and using a web-based tool (FuncPred; http://manticore.niehs.nih.gov/snpfunc.htm) for prediction of SNP functions, but without success. In other words, *MMP2*, *TIMP2* and *TIMP3* were not associated with high myopia in the Han Chinese population under study, and are thus unlikely to play a major role in the genetic susceptibility to high myopia.

A recent Japanese study examined two functional promoter SNPs of the *MMP2* gene in a case-control study involving 725 high myopes (SE $\leq$-6.0 D) and 546 population-based controls, and found no association of these two SNPs with high myopia.[34] These two promoter SNPs were rs243865 and rs2285053 (named as C -1306T and C -735T, respectively, in the report), and were not examined in the present study. The SNP rs243865 had an MAF of less than 0.10 in Han Chinese and hence did not satisfy the criteria of selecting tag SNPs in our study while the other SNP rs2285053 was not documented in the HapMap database.

A US-based group recently examined 146 tag SNPs from 14 *MMP* and 4 *TIMP* genes for 55 Amish families (358 individuals, mean SE = -1.61 D) and 63 Ashkenazi families (535 individuals, mean SE = -3.56 D).[35] The tag SNPs were selected from the HapMap Caucasian (CEU) database with the criteria of MAF $\geq$ 0.15 and $r^2 \geq$ 0.7. In particular, 6 tag SNPs from *MMP2*, 11 from *TIMP2* and 12 from *TIMP3* were included, which are expectedly less than those examined in our study (Table 1) because of their less restrictive criteria of SNP selection. Two SNPs were found significantly associated with ocular refraction by quantitative trait analysis using family-based association testing in the Amish families only, but not the Ashkenazi families. Both sets of families were sampled from largely endogamous, rapidly expanding, but isolated populations in the USA. The prevalence of refractive errors is high in Jewish populations,[36] but relatively low in the Old Order Amish.[37] The behavioral and environmental factors are more conducive to myopia development in the Jewish populations than in the Amish populations, and could probably explain the discrepancy in the genetic association results, as suggested by the authors.[35] The authors also anticipated that the positive results could not be replicated in South Asian Chinese and Japanese populations with high prevalence of environmentally induced myopia.[35] Indeed, our study could not replicate the findings. One of the positive SNPs in the Amish population was rs9928731 (*P*=0.00026) within the *MMP2* gene.[35] This SNP was also screened by the DNA pooling approach in the present study: the estimated frequency of the C allele was 0.5414 in case pools and 0.5015 in control pools, which were not statistically significant (difference=0.0399, nested ANOVA *P* value = 0.1435, Table 3). The frequency of the C allele in controls is similar to that in Han Chinese documented in HapMap database (0.5015 vs 0.4560). It is worth noting that the phenotype definition was different for these two studies: quantitative measures of refractive errors in the American study, but dichotomous trait of high myopia (affected vs unaffected) in our study.

All three association studies (Japanese, American and our) focus on common polymorphisms in the genes under study and hence assumed the hypothesis of common disease common variants.[38] Strong linkage disequilibrium (LD) between common tag SNPs and common casual variants is critical to the success of this indirect LD mapping approach. Sequence variations must have similar allele

frequencies in order to be highly correlated and in strong LD. However, rare casual variants may also contribute to myopia development – the other side of the story being the hypothesis of common disease rare variants.[38] This indirect approach is of low power in detecting association with rare variants because of the weak LD between common tag SNPs and rare casual variants. Therefore, direct mapping must be performed to detect association with rare causal variants, which must first be identified. Rare variants can be identified for direct association studies by sequencing of good candidate genes or even the whole genome for a very large number of samples.[39]

Our case subjects had extreme refractive errors (mean SE = -10.53 D for Set 1;[26] and -10.30 D for Set 2, Table 2). This would enhance the homogeneity of the myopia phenotype, enrich the genetic components of the contributing factors, and hence increase the power of our study (though in a subtle manner). The three candidate genes were chosen for study because they have been shown to be involved in sclera remodeling in myopia development in many studies.[11-22] Our negative finding might imply that these genes do not carry common sequence variants that are capable of influencing their function and/or regulation in the relevant ocular tissue. However, the contribution of behavioral and environmental effects on high myopia should not be overlooked. Our DNA pooling-based initial screen adopted a lenient threshold of $P \leq 0.10$ in order avoid missing potential SNPs. For rs135029 of *TIMP3*, the power of the third stage study (Sample Set 2) is 73% under an allelic model and 78% under a genotypic model. The power is calculated based on the following assumptions with the online Genetic Power Calculator (http://pngu.mgh.harvard.edu/~purcell/gpc/): OR and allele frequencies obtained for rs135029 for Sample Set 1 (Table 4), a disease prevalence of 0.05 for high myopia in our local Chinese population,[40] and a significance level set at $\alpha=0.025$ because two SNPs were examined in the third stage. One disadvantage of DNA pooling strategy is that it makes haplotype analysis very difficult, if not impossible.[24] Algorithms are available for estimating haplotype frequencies in small DNA pools constructed from a few (<10) individuals. In other words, our current pooling protocol might miss some potential SNPs for follow-up in the second stage analysis if high myopia is associated with certain haplotypes, but not individual SNPs. This is one of the reasons why a lenient

threshold of $P \leq 0.10$ was used to selecting SNPs for follow-up study by individual genotyping in the second stage.

Association testing of DNA pools has been proven to be an effective initial screen of SNPs and candidate genes for subsequent detailed follow-up study.[24,41] The major advantages are tremendous reduction in DNA usage and in the amount of genotyping work when compared to individual genotyping. For example, our study required for each SNP 36 PCRs and subsequent analyses for 6 case pools and 6 control pools (Fig. 2), plus three separate PCRs for heterozygotes to determine the k correction factor. The amount of genotyping work was only about one-fifteenth of that required for genotyping of 600 individual samples. It has also been shown that use of small DNA pools of about 50 individuals is superior to use of fewer, larger DNA pools for candidate gene studies.[25] In addition to the advantages mentioned above, use of small DNA pools allows the use of standard statistical method (nested ANOVA) for data analysis without the need of directly estimating the variance components of the error sources while it properly handles variations arising from sampling of subjects and technical errors, which are due to unequal amounts of DNA being mixed together, errors in PCR and primer extension, and in DHPLC analysis.

The present study used DHPLC analysis to estimate the relative allele frequencies of DNA pools. DHPLC analysis is in fact a rate-limiting step because samples have to be injected and analyzed sequentially. The throughput can be greatly increased if quantitative genotyping is conducted with a mass spectrometer,[42] e.g. using the MassArray iPLEX method (Sequenom). However, the local service provider only entertained request of classical genotyping work based iPLEX method, but not quantitative genotyping. DNA pooling strategy may become less attractive as the *unit* cost of genotyping reduces tremendously with the availability of high-throughput genotyping platforms like whole-genome genotyping arrays. Nevertheless, the total cost of whole-genome genotyping for a large number of samples is still prohibitive for many research groups. In fact, genome-wide association studies can be within the reach of even small- to medium-sized research groups if DNA pooling strategy is applied.[43] Interestingly, errors due to array variations are much

greater than those due to pool construction, and hence it is recommended to have multiple arrays per

DNA pool for a few pools rather than multiple DNA pools with less arrays per pool.[44]

In conclusion, we used a DNA pooling strategy to screen 49 tag SNPs from three candidate genes (*MMP2*, *TIMP2* and *TIMP3*). Three tag SNPs passed the threshold ($P \leq 0.10$) and were tested by individual genotyping of samples forming the DNA pools. Two SNPs from the *TIMP2* gene were found associated with high myopia by single-marker analysis or haplotype analysis. However, the initial positive results could not be replicated by an independent second sample set. Overall, these three candidate genes are unlikely to play a major role in the genetic susceptibility to high myopia in Chinese population.

References

1. Dandona R, Dandona L. Refractive error blindness. *Bull World Health Organ.* 2001;79:237–243.

2. Gilmartin B. Myopia: precedents for research in the twenty-first century. *Clin Experiment Ophthalmol.* 2004;32:305-324.

3. Edwards MH, Lam CS. The epidemiology of myopia in Hong Kong. *Ann Acad Med Singapore.* 2004;33:34-38.

4. Lyhne N, Sjølie AK, Kyvik KO, Green A. The importance of genes and environment for ocular refraction and its determiners: a population based study among 20-45 year old twins. *Br J Ophthalmol*. 2001;85:1470-1476.

5. Chen CY, Scurrah KJ, Stankovich J, et al. Heritability and shared environment estimates for myopia and associated ocular biometric traits: the Genes in Myopia (GEM) family study. *Hum Genet.* 2007;121:511-520.

6. Saw SM, Chua WH, Wu HM, Yap E, Chia KS, Stone RA. Myopia: gene-environment interaction. *Ann Acad Med Singapore.* 2000;29:290-297.

7. Hammond CJ, Snieder H, Gilbert CE, Spector TD. Genes and environment in refractive error: the twin eye study. *Invest Ophthalmol Vis Sci.* 2001;42:1232-1236.

8. Dirani M, Chamberlain M, Shekar SN, et al. Heritability of refractive error and ocular biometrics: the Genes in Myopia (GEM) twin study. *Invest Ophthalmol Vis Sci.* 2006;47:4756-4761.

9. Klein AP, Suktitipat B, Duggal P, et al. (2009) Heritability analysis of spherical equivalent, axial length, corneal curvature, and anterior chamber depth in the Beaver Dam Eye Study. *Arch Ophthalmol.* 2009;127:649-655.

10. McBrien NA, Gentle A. Role of the sclera in the development and pathological complications of myopia. *Retinal Eye Res.* 2003;22:307-328.

11. Rada JA, Brenza HL. Increased latent gelatinase activity in the sclera of visually deprived chicks. *Invest Ophthalmol Vis Sci.* 1995;36:1555-1565.

12. Jones BE, Thompson EW, Hodos W, Waldbillig RJ, Chader GJ. Scleral matrix metalloproteinases, serine proteinase activity and hydrational capacity are increased in myopia induced by retinal image degradation. *Exp Eye Res.* 1996;63:369-381.

13. Liu SZ, Wu WC, Wang JF, Tan XP, Jiang HB. Dynamic changes of MMP-2 activity in the posterior sclera of chicks with form-deprivation myopia. *Zhong Nan Da Xue Xue Bao Yi Xue Ban.* 2005;30:295-298.

14. Rada JA, Perry CA, Slover ML, Achen VR. Gelatinase A and TIMP-2 expression in the fibrous sclera of myopic and recovering chick eyes. *Invest Ophthalmol Vis Sci.* 1999;40:3091-3099.

15. Wu WC, Liu SZ, Wang JF, Tan XP, Jiang HB. Matrix metalloproteinase-2 mRNA expression in the posterior sclera of chick form-deprivation myopia. *Zhong Nan Da Xue Xue Bao Yi Xue Ban.* 2004;29:270-273.

16. Guggenheim JA, McBrien NA. Form-deprivation myopia induces activation of scleral matrix metalloproteinase-2 in tree shrew. *Invest Ophthalmol Vis Sci.* 1996;37:1380-1395.

17. 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.

18. Siegwart JT Jr, Norton TT. The time course of changes in mRNA levels in tree shrew sclera during induced myopia and recovery. *Invest Ophthalmol Vis Sci.* 2002;43:2067-2075.

19. Yang SR, Ye JJ, Long Q. Expressions of collagen, matrix metalloproeinases-2, and tissue inhibitor of matrix metalloproteinase-2 in the posterior sclera of newborn guinea pigs with negative lens-defocused myopia. *Zhongguo Yi Xue Ke Xue Yuan Xue Bao.* 2010;32:55-59.

20. Cui W, Bryant MR, Sweet PM, McDonnell PJ. Changes in gene expression in response to mechanical strain in human scleral fibroblasts. *Exp Eye Res.* 2004;78:275-284.

21. Siegwart JT Jr, Norton TT. Selective regulation of MMP and TIMP mRNA levels in tree shrew sclera during minus lens compensation and recovery. *Invest Ophthalmol Vis Sci.* 2005;46:3484-3492.

22. Schippert R, Brand C, Schaeffel F, Feldkaemper MP. Changes in scleral MMP-2, TIMP-2 and TGFbeta-2 mRNA expression after imposed myopic and hyperopic defocus in chickens. *Exp Eye Res.* 2006;82:710-719.

23. 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.

24. 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.

25. Barratt BJ, Payne F, Rance HE, Nutland S, Todd JA, Clayton DG. Identification of the sources of error in allele frequency estimations from pooled DNA indicates an optimal experimental design. *Ann Hum Genet.* 2002;66:393-405.

26. 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.

27. Han W, Yip SP, Wang J, Yap MK. Using denaturing HPLC for SNP discovery and genotyping, and establishing the linkage disequilibrium pattern for the all-trans-retinol dehydrogenase (RDH8) gene. *J Hum Genet.* 2004;49:16-23.

28. Zhou L, Myers AN, Vandersteen JG, Wang L, Wittwer CT. Closed-tube genotyping with unlabeled oligonucleotide probes and a saturating DNA dye. *Clin Chem.* 2004;50:1328-1335.

29. 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.

30. Zar JH. *Biostatistical Analysis*, 5th ed. Upper Saddle River, NJ: Prentice Hall/Pearson; 2010: 307-327.

31. 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.

32. Westfall PH, Young SS. *Resampling-based Multiple Testing: Examples and methods for p-value adjustment.* New York, NY: John Wiley & Sons Ltd; 1993.

33. Nielsen DM, Ehm MG, Weir BS. Detecting marker-disease association by testing for Hardy-Weinberg disequilibrium at a marker locus. *Am J Hum Genet.* 1998;63:1531-1540.

34. Nakanishi H, Hayashi H, Yamada R, et al. Single-nucleotide polymorphisms in the promoter region of matrix metalloproteinase-1, -2, and -3 in Japanese with high myopia. *Invest Ophthalmol Vis Sci.* 2010;51:4432-4436.

35. Wojciechowski R, Bailey-Wilson JE, Stambolian D. Association of matrix metalloproteinase gene polymorphisms with refractive error in Amish and Ashkenazi families. *Invest Ophthalmol Vis Sci.* 2010;51:4989-4995.

36. Zylbermann R, Landau D, Berson D. The influence of study habits on myopia in Jewish teenagers. *J Pediatr Ophthalmol Strabismus.* 1993;30:319 –322.

37. Peet JA, Cotch MF, Wojciechowski R, Bailey-Wilson JE, Stambolian D. Heritability and familial aggregation of refractive error in the Old Order Amish. *Invest Ophthalmol Vis Sci.* 2007;48:4002–4006.

38. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet. 2008;40:695-701.

39. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet.* 2010;11:415-425.

40. 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.

41. Bansal A, van den Boom D, Kammerer S, et al. Association testing by DNA pooling: an effective initial screen. *Proc Natl Acad Sci U S A.* 2002;99:16871-16874.

42. Mohlke KL, Erdos MR, Scott LJ, et al. High-throughput screening for evidence of association by using mass spectrometry genotyping on DNA pools. *Proc Natl Acad Sci U S A.* 2002;99:16928-16933.

43. 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.

44. Macgregor S. Most pooling variation in array-based DNA pooling is attributable to array error rather than pool construction error. *Eur J Hum Genet.* 2007;15:501-504.

FIGURE LEGEND

Figure 1. A three-stage approach to testing genetic association based on an initial screen of DNA pools.

Figure 2. Nested design of the DNA pooling study. There are two subject groups (case group, G_2 ;and control group, G_1), six DNA pools per group (P_{21} to P_{26} for case group, and P_{11} to P_{16} for control group), and three technical replicates (R_{ij1} to R_{ij3}) for each DNA pool. Note that there is no link from any pools of the case group to any pools of the control group. Therefore, the level of the case group is not cross-classified with the control group, but is nested with the respective group, i.e. the pools are nested within the group. Each DNA pool was constructed by mixing equal amounts of DNA from 50 distinct individuals of the same subject group.

Table 1. Summary of tag SNPs in the *MMP2*, *TIMP2* and *TIMP3* genes

Gene	GeneID	Chromosomal location	Region captioned*	No. of tag SNPs	No. of SNPs captured at mean $r^2 =$?
MMP2	4313	16q13-q21	33.5 kb	17	43 ($r^2 = 0.994$)
TIMP2	7077	17q25	78.4 kb	17	52 ($r^2 = 0.973$)
TIMP3	7078	22q12.3	68.2 kb	15	37 ($r^2 = 0.960$)

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

Table 2. Characteristics of study subjects of Sample Set 2 (for right eyes only)

Characteristic	Cases (n=356)		Controls (n=354)	
Age, mean (SD), y	34.0	(9.1)	33.1	(9.5)
Females, no. (%)	236	(66.3)	209	(59.0)
Spherical equivalent, mean (SD), D	-10.30	(2.46)	0.08	(0.53)
Axial length, mean (SD), mm	27.64	(1.18)	23.73	(0.82)
Corneal power, mean (SD), D	44.38	(3.71)	43.97	(1.51)
Anterior chamber depth, mean (SD), mm	3.34	(0.38)	3.20	(0.41)
Lens thickness, mean (SD), mm	4.30	(0.50)	4.33	(0.57)

Table 3. Pooled DNA analysis of tag SNPs in the *MMP2*, *TIMP2* and *TIMP3* genes

SNP*	Alleles† (1ˢᵗ/2ⁿᵈ)	k correction factor peak height ratio (1ˢᵗ/2ⁿᵈ)	Estimated frequencies of 1ˢᵗ allele Case pools	Control pools	Diff (Case - Control)	nANOVA *P* value
MMP2						
rs11643630	T / G	1.01	0.4573	0.4534	0.0039	0.9029
rs1477017	G / A	1.12	0.3064	0.2958	0.0106	0.5585
rs865094	G / A	0.96	0.3268	0.3335	-0.0067	0.7647
rs11076101	C / T	1.01	0.7637	0.7855	-0.0218	0.3412
rs17301608	C / T	1.06	0.6557	0.6801	-0.0244	0.2008
rs11646643	G / A	1.02	0.1817	0.1702	0.0115	0.6305
rs2241146	G / A	1.01	0.7749	0.7736	0.0013	0.9427
rs9928731	C / T	1.11	0.5414	0.5015	0.0399	0.1435
rs12599775	C / G	1.25	0.1659	0.1819	-0.0160	0.5207
rs243847	C / T	1.09	0.4084	0.4196	-0.0112	0.7028
rs243845	G / A	1.05	0.7187	0.6907	0.0280	0.3367
rs243843	G / A	1.04	0.4285	0.4466	-0.0181	0.4689
rs183112	G / A	1.32	0.6643	0.6718	-0.0075	0.7514
rs1992116	G / A	1.02	0.6671	0.6945	-0.0274	0.2274
rs11639960	G / A	1.07	0.2801	0.2763	0.0038	0.8187
rs243835	C / T	1.05	0.4218	0.3817	0.0401	0.2328
rs1861320	G / T	1.23	0.7969	0.8195	-0.0226	0.5003
TIMP2						
rs4789932	C / T	1.07	0.3519	0.3551	-0.0032	0.9259
rs8080623	C / T	1.15	0.2968	0.2975	-0.0007	0.9770
rs8179091	C / T	0.96	0.4432	0.4465	-0.0033	0.9115
rs7212662	C / A	1.17	0.2912	0.2972	-0.0060	0.8225
rs8066695	G / A	1.03	0.374	0.3907	-0.0167	0.5461
rs12600817	C / T	0.65	0.5475	0.5774	-0.0299	0.4249
rs4789860	G / A	1.09	0.2553	0.2584	-0.0031	0.9410
rs2889529	C / T	1.23	0.3255	0.3202	0.0053	0.8840
rs2376999	C / T	1.01	0.3446	0.3235	0.0211	0.3401
rs2003241	**G / A**	**0.97**	**0.2458**	**0.2129**	**0.0329**	**0.0119**
rs7502935	C / T	1.45	0.6973	0.6914	0.0059	0.8301
rs6501258	A / T	1.20	0.5172	0.5019	0.0153	0.5409
rs6501256	G / A	1.01	0.2361	0.2166	0.0195	0.4293
rs11868442	G / A	1.36	0.6678	0.6783	-0.0105	0.7285
rs2277698	G / A	1.21	0.7445	0.736	0.0085	0.7413
rs9905930	G / T	1.01	0.7625	0.766	-0.0035	0.8442
rs16971783	T / A	1.00	0.1017	0.1042	-0.0025	0.8263
TIMP3						
rs1962223	G / C	1.08	0.6029	0.6245	-0.0216	0.4599
rs9619311	G / A	0.99	0.129	0.152	-0.0230	0.2400
rs242089	C / T	1.04	0.5167	0.5126	0.0041	0.8976
rs80272	G / A	1.01	0.105	0.1306	-0.0256	0.2691
rs8140818	G / A	1.01	0.0719	0.0939	-0.0220	0.2107
rs242076	C / T	0.92	0.5287	0.5623	-0.0336	0.4067
rs715572	C / T	1.04	0.64	0.6452	-0.0052	0.8863
rs242072	G / A	1.01	0.4711	0.4663	0.0048	0.8683
rs135029	**C / T**	**1.03**	**0.7898**	**0.8495**	**-0.0597**	**0.0010**
rs241890	C / A	1.18	0.5981	0.5901	0.0080	0.7674
rs1427385	G / A	1.03	0.5344	0.5391	-0.0047	0.8962
rs9609643	C / T	1.12	0.8848	0.8574	0.0274	0.1511
rs9862	G / A	0.95	0.5242	0.5439	-0.0197	0.5089
rs11547635	A / G	0.98	0.3424	0.327	0.0154	0.5151
rs137485	**A / T**	**0.85**	**0.2440**	**0.2089**	**0.0351**	**0.0727**

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

† The 1ˢᵗ allele has a shorter elution time while the 2ⁿᵈ allele has a longer elution time. The alleles are named with reference to the sense strand of the respective gene.

Table 4. Allelic association tests of *TIMP2* and *TIMP3* SNPs genotyped individually

Gene, SNP	Alleles*		Genotype counts (11/12/22)*		Minor allele freq		OR (95%CI) †	Allelic test‡	
	1	2	Cases	Controls	Cases	Controls		P_{asym}	P_{emp}
Sample Set 1									
TIMP2, rs2003241	A	G	189/93/10	201/84/12	0.1935	0.1818	1.08 (0.81 - 1.45)	0.6078	0.9374
TIMP3, rs135029	C	T	188/106/5	223/69/6	0.1940	0.1359	1.53 (1.12 - 2.09)	**0.0069**	**0.0162**
TIMP3, rs137485	T	A	209/84/4	228/62/7	0.1549	0.1279	1.25 (0.90 - 1.73)	0.1828	0.4466
Sample Set 2									
TIMP3, rs135029	C	T	260/87/8	261/83/9	0.1451	0.1431	1.02 (0.76 - 1.37)	0.9142	0.9966
TIMP3, rs137485	T	A	259/87/6	271/66/9	0.1406	0.1214	1.18 (0.87 - 1.62)	0.2870	0.5231
Combined (Sets 1 & 2)§									
TIMP3, rs135029	C	T	448/193/13	484/152/15	0.1674	0.1398	1.26	0.0344	0.0693
TIMP3, rs137485	T	A	468/171/10	499/128/16	0.1471	0.1244	1.24	0.0917	0.1268

* The major allele is designated as "1" and minor allele as "2"; and the genotype counts are indicated as the counts of the genotypes 11, 12 and 22, respectively. Sample Set 1 has 300 cases and 300 controls while Sample Set 2 has 356 cases and 354 controls. Note that the total genotype counts may not add up to these expected numbers because a few samples failed to be genotyped in a random fashion.

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

Since rs135029 is of interest, its genotypic odds ratios and their 95% CI (in brackets) are given here with CC as the reference genotype: 1.82 (1.27 – 2.61; Set 1), 1.06 (0.75 – 1.50; Set 2), 1.37 (1.08 – 1.77; Combined Set, not adjusted for age here) for genotype CT, and 0.99 (0.30 – 3.29; Set 1), 0.90 (0.34 – 2.37; Set 2) and 0.94 (0.44 – 2.00; Combined Set, not adjusted for age here) for genotype TT. This simply serves to avoid cluttering in the table above.

‡ Allele frequencies are compared by chi-square test to give the asymptotic *P* value (P_{asym}). Multiple comparisons are corrected by 10,000 permutations across single marker allelic tests (this table) <u>and</u> omnibus tests of haplotypes (Table 4) for a given sample set (Set 1, Set 2 or Combined, each separately). The empirical *P* value is indicated as P_{emp}.

§ Single-marker (this table) and haplotype (Table 5) analyses are performed for the combined sample set (Sets 1 and 2) with adjustment for age as a covariate to account for the effect of age. This has to be done using logistic regression, the relevant PLINK command of which does not give the 95% CI for the odds ratio (OR).

Table 5. Haplotype analysis of 2 *TIMP3* SNPs (rs135029-rs137485)*

| | Haplotype freq | | | | |
Haplotype	Cases	Controls	OR	P_{asym}	P_{emp}
Sample Set 1					
Omnibus test	-	-	-	**0.0178**	**0.0496**
TA	0.1431	0.1102	1.38	0.0785	
CA	0.0118	0.0186	0.62	0.3350	
TT	0.0522	0.0254	**2.11**	**0.0201**	
CT	0.7929	0.8458	**0.69**	**0.0172**	
Sample Set 2					
Omnibus test	-	-	-	0.2380	0.4293
TA	0.1028	0.1033	1.00	0.9780	
CA	0.0368	0.0185	2.00	0.0463	
TT	0.0425	0.0387	1.10	0.7260	
CT	0.8179	0.8395	0.86	0.2890	
Combined (Sets 1 & 2) †					
Omnibus test	-	-	-	0.0690	0.1381
TA	0.1219	0.1070	1.20	0.1560	
CA	0.0247	0.0180	1.35	0.2680	
TT	0.0463	0.0320	1.43	0.0829	
CT	0.8071	0.8430	0.77	0.0118	

* 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) <u>and</u> omnibus tests of haplotypes (this table) for a given sample set (Set 1, Set 2 or Combined, each separately). The empirical *P* value is indicated as P_{emp}. Note that PLINK does not generate confidence intervals for odds ratios (OR) in haplotype analysis.

† Single-marker (Table 4) and haplotype (this table) analyses are performed for the combined sample set (Sets 1 and 2) with adjustment for age as a covariate to account for the effect of age. This has to be done using logistic regression, the relevant PLINK command of which does not give the 95% CI for the odds ratio (OR).
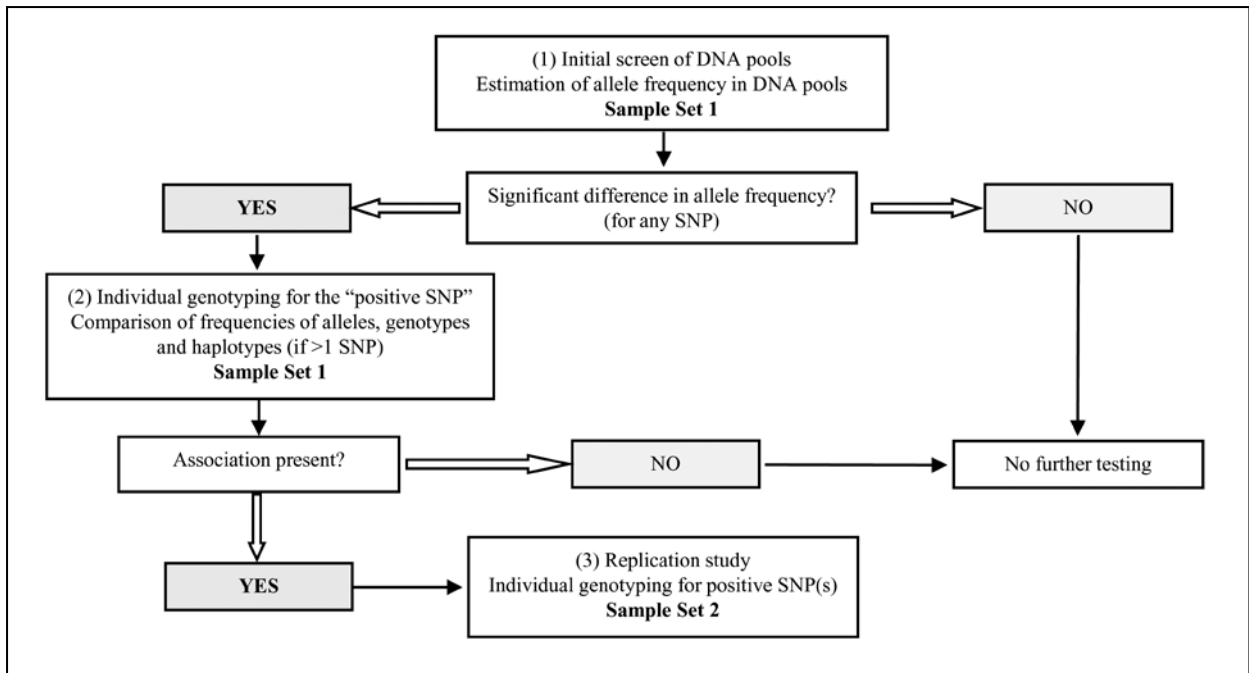
Figure 1

Figure 2
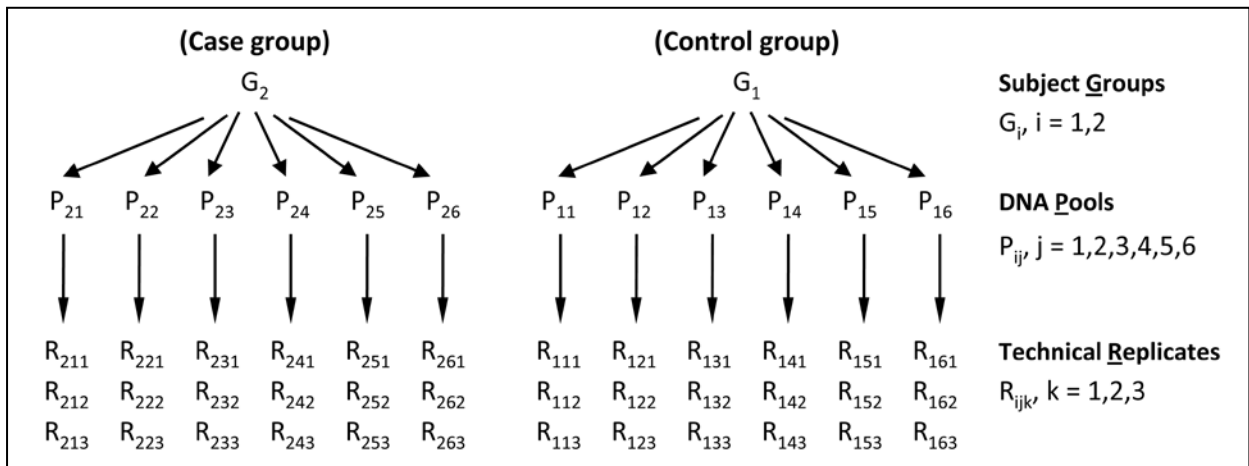
Supplementary Table S1. *MMP2*, *TIMP2* and *TIMP3* SNPs: primers for PCR and primer extension reaction and buffer B concentration ranges for DHPLC

SNP*	Forward PCR primer (5'>3')	Reverse PCR primer (5'>3'')	PE primer (5'>3')	ddGTP	ddCTP	ddTTP	ddATP	Buffer B (%)
MMP2								
rs11643630	GGTAAATTATTCATGCTTCTGCCTTTT	TGGCACAAGATACAGGTCATAAA	ACCAAAACCAAGATGATGATGAG		√		√	23-29%
rs1477017	TCCCTCGCAAGCCCAAGT	GCACTGAAAAGTAACCCAATGTC	TGTGTATTGAAACTCCCTAAGATG		√	√		26-32%
rs865094	CATCTTAGGGAGTTTCAATACACA	GCTACCAATCATCACGTTCCTT	ACAGACTTGAGTTTCATACTTGCT	√			√	27-33%
rs11076101	GGCTTAGATAGGACAGTAGATG	TGCCCAGTGTGACCAGGAAA	CATCTGCCCCCATGTCAAC	√			√	22-28%
rs17301608	GTTGGATGATGCTGGTAGACA	TCCTGTCAATACTGCCCTCTTA	TGTAGACACTAGAGGAAGGGT		√	√		23-29%
rs11646643	GTAAAGAGCAGGGGAAACAATG	TACCACATACCCACATTACCAC	GGATGGAGGATGGCATAGAC	√			√	22-28%
rs2241146	CAGAATGGACAGTGCCCTTG	AGAGAGGGAAGGATGTGAATGA	GTGTCAATCATTTTGTTGTTGGGTG	√			√	27-33%
rs9928731	CAGAACAGACATTGGGCATC	CTGTTGGAGGAAGGGAAGAA	GTCTGTATCTCCACCTTTTAGTC		√	√		26-32%
rs12599775	CGACCACAGCCAACTACGA	GGAGGTGAGTAGCAGCATCT	GACCCTGGACATTGCCCTT	√	√			23-29%
rs243847	AGACACCCACCTACCCAGA	GCAGTGGGAATGTCTCTTAGG	GATGCTGCTACTCACCTCC		√	√		22-28%
rs243845	GTTGGTTCTTACAGTGAGGCTA	GCAGACCTTGGGCTTTTTCC	GAGCCAGGCCAAAATAGTGAAA	√			√	22-28%
rs243843	GCTTGCTTCTTGTCCTCCTT	AGCACACCTGAACCCAAACT	GTCATTCTGGGAAGCATTTGG		√	√		24-30%
rs183112	AGCCTCCACTCACATCTTGT	TCACTCACCACCCATACACA	GTTCCCAATTCTTCCCTCCATTT		√	√		29-35%
rs1992116	TGTGTATGGGTGGTGAGTGAC	GGATGGATGATAGAAGAATGGAAG	GCCCTAGTTCCCTGGAGAAT	√			√	23-29%
rs11639960	GTCTTCACTCTTCACTACTATTC	CTGATTTGGATGCTGGCTTC	TAAGTCTTCTTCATATTTTGTCCAC		√	√		27-33%
rs243835	CACACTGTTTGGCATGTAGTAAG	CATTTTCTTCTTCACCTCATTGTATC	GCCATCAGCTGGGTGCTC		√	√		20-26%
rs1861320	CCCACTCCTAATCTCCTACATA	CAGAAAAAAGAGATAGAAACAATGACA	CACCTGTAGAGTTCACTCCTTA		√		√	25-31%
TIMP2								
rs4789932	CACCTCACTCAGAACAAAATGC	TTCTGTTCCCTCTACCATTATCTC	ACGGTGTCCAGGCTACAG		√	√		21-27%
rs8080623	CTGGAGTGCTGACCTGAGT	ATGTGGGAGCCGAAGGACT	ACCACCCTACACATGTTCATC		√	√		25-31%
rs8179091†	GAGTCGAGCTGAAGGGGAAA	CGCTCGGAGGATTTTCTGCT	CGGGTGGGTCGCCTGGTG		√	√		18-24%
rs7212662	CTTGCTCAGACTTCGGCTTT	GCAGCCAAGAGTAAAAGGAG	GCCCACAGGTGATTTAACC		√		√	22-28%
rs8066695	ATCCGCTCAGACACTACCAC	GGAAGACAACCACAGAACCT	AACCAGCTCTGGGGAAAGG	√			√	20-26%
rs12600817	TGGGTCCGTATTAGGGTTTC	TCACAGACAGCAGATTCCAG	TTCCCACATAAATCCCATCTTAC	√			√	25-31%
rs4789860	TAATGGACGCTGTCGAGGTAT	CATATTGGTCAGACTGGTTACAAA	TGAATCTGTGTGTATCCATATCC		√	√		26-33%
rs2889529	CCTCCTTTGGCTTTTCTTCCTA	TACGAAGTGCCTGTGCTCTA	ACACCTCTCCCTCGTTCAGATT		√	√		25-31%
rs2376999	CTGGGTAGTAAGAGGTGTTCAT	CACAATGGATCATAGACCTTCAT	CACCAAAAGCACAACCAATAAAC	√			√	24-30%

SNP*	Forward PCR primer (5'>3')	Reverse PCR primer (5'>3'')	PE primer (5'>3')	ddGTP	ddCTP	ddTTP	ddATP	Buffer B (%)
rs2003241	CTCAGAAGCAGCCACAGATG	ATTCTGGGTCTTGGGTCTGG	AGACTCTACAAAAATAGGTGGTG	√			√	25-31%
rs7502935	GCCCCATCCTGCTAAAAGAA	TCTGAAAAGGGAGGTGAAGC	TCTAAGCTGTCTTTACTGTGATC		√	√		26-35%
rs6501258	GTGTGATTTGAGGGTGTGGA	TCATTTGGCTCTGGGTGGAA	AGGTGGACTCAGCTTTGTTC			√	√	24-30%
rs6501256	TCTCTTGATTCCCATTTTCCTTCTTT	AAGAGGAGGCTGGTCAAATCA	TGCCCACCCCCAATTTAAAG	√			√	24-30%
rs11868442	AGTGGTAGAATAGTGGGAGCAT	TCTTCTTTCTTCTTCTGTTGGAG	ACAGACGGCTTTGCCCTGA		√	√		22-28%
rs2277698	CCTCCTCCTTGTCTTTCCAG	TACCCTGCCTCACTGTTCCT	TTCTTTCCTCCAACGTCCAG		√	√		25-31%
rs9905930	GCTCTGACACCTGACTCGTTA	AGTGCTGTGCTGCGTGTATG	GGAGTAGGAGACCAGAAATGAT	√			√	23-29%
rs16971783	GGAGAAAAGGACCCGAAGATAAA	TGGAAATGTGGATGGCTTGC	AACAGGAAGTGCGAGCATAAG			√	√	22-28%
TIMP3								
rs1962223	AAGTGGAAACAAAAAATCAAAAGACCT	TGCGAAACTGATGGATGGTCT	TATCTTTCATCCAATAAACACAGAAGT	√	√			28-34%
rs9619311	GCTGAGAAGTGGACAAAGACA	CTTGTGCTTGCTTTTCCTACC	CCTGCCCATGGCTGACAG	√			√	20-26%
rs135025	AGATAGATGATAAATGATAAATAGAAATGGATA	AACAAAACAACAACACGATTCTCC	GCCACCCAGGAGTAGACTA		√	√		23-29%
rs242089	CAATCACAATGGCAGGCAAG	TAGGTGGGAGAATCGCTTGA	TGCTGCATGGTGACCTGTTT	√			√	24-30%
rs80272	ATAGGGAGGAGGCTGGTAGT	TGACCTGCCTGATGCTACCT	CAGCACTGAATCTGGGAAC	√			√	22-28%
rs8140818	ACTTGGGAGGCTGAGGGAAA	CCAGGGCTCTTCTTGCTATT	GAGACTCTGTCTCAGAAAAAGAG	√			√	24-30%
rs242076	AGAAAGCAGGAGGATGGTCA	CTAAAAGATGGGCTAATGACAACA	AAAATCCACGGAAAGCATTTAGC	√			√	24-30%
rs715572	CTGAAATGCTACCTCCTCCA	TCCCACAGCCTCATTCATTCA	CTTTCCTCTCTTTCTTCCAGCA		√	√		25-31%
rs242072	CAGCCACTCTTGATTTCTTCCT	GCCTCATTGGACCCTTAGAA	GATTCCTGGATTCAGGGCC	√			√	21-27%
rs135029	AGCAAGGCAAACAAGGCACA	GAAGGGTCTGGATTTGGTTGA	GTTTCCAGAAAAGCTTACTGGTG		√	√		25-31%
rs241890	TGGAGAAGCACAGATGAAGTAG	AATAGGTGGATCTGGGGTTG	CCCAAGAATTGCCTACTAAGAC		√		√	23-29%
rs1427385	AGGGCAGTATTAGCATCCACAT	CGTAGAAAGAGAAGCACAACCT	GGGAATTGGAATCAGGATTGAAC	√			√	24-30%
rs9609643	TTATCCCCTCAGTTCTCAGC	GAAGAGACACCAATGAAGAGATG	GTTTATAATAGGAAAATGCCTCTACTTTA		√	√		29-35%
rs9862	AGCAGGTACTGGTACTTGTTGA	GATGGGCACTTAGGTTGATTTC	CCCATGTGCAGTACATCCA		√	√		21-27%
rs11547635	AACATTAGAGACTCCTTACAATTTACT	CAGACCCCTCTTCCATATTC	TTAAGGCCACAGAGACTCTC	dGTP‡			√	25-31%
rs137485	GAGCCCACTTGAAACCACTT	TTGCAGCCCTAGAAACATCAG	GGTCTGAGCAGATATAGTAAGGA			√	dATP‡	26-32%

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

† Dimethylsulfoxide (5%) was added as additive in the PCR.

‡ dNTP was used for the primer extension reaction for an optimal discrimination of the extended products.

Supplementary Table S2. Primers and probes for genotyping individual samples

Gene, SNP	Primer sequences (5'>3')*
MassArray iPLEX (Sequenom) for Sample Set 1	
TIMP2, rs2003241	F: ACGTTGGATGAACTATGGCACAAAGGGCAG
	R: ACGTTGGATGAAATGAAAGGGCGTGGCCAG
	PE primer: taagGAGACTCTACAAAAATAGGTGGTG
TIMP3, rs135029	F: ACGTTGGATGCAATGGCAATTGGGATTGAG
	R: ACGTTGGATGGATGAACAGCACATGTGCAA
	PE primer: tgaaaATGTGCAATTTCTGGAGAC
TIMP3, rs137485	F: ACGTTGGATGGAAAGGCATTCTTCCTTCCC
	R: ACGTTGGATGGCAGGTCTGAGCAGATATAG
	PE primer: CTGAGCAGATATAGTAAGGA
Unlabeled probe melting analysis for Sample Set 2	
TIMP3, rs135029	F: GAA GAA GGG TCT GGA TTT GGT T
	R: AGT ATG GAT CAC AGT GGC ACA A
	Probe: AGC TTA CTG GTG CGT CTC AAG AAA TTG C
TIMP3, rs137485	F: CTG ACT TAC AGC CCT AGA AAC A
	R: GAC TCA AGA GCC CTC ATG C
	Probe: TCT GAG CAG ATA TAG TAA GGA TTG TTG CC

* F = forward primer; R = reverse primer; and PE = primer extension

Supplementary materials for statistical analysis (nested ANOVA)

Here, we briefly explain the principle of nested analysis of variance (ANOVA) used to analyze the estimated allele frequencies of DNA pools. Nested ANOVA design is an extension of one-way ANOVA design. All assumptions (e.g. normality of residuals, constant variance, etc) of ANOVA hold true for nested ANOVA. Our nested design had one measurement variable (the estimated allele frequency, R_{ijk}) and two nominal variables (DNA pools and subject groups) with the DNA pools (P_{ij}) *nested within* each of the subject groups (G_i) (Fig. 2 in main text). There were two subject groups (case group and control group, i=1,2), six DNA pools (j=1,2,3,4,5,6) within each subject group, and three replicates (k=1,2,3) per DNA pool. Nested ANOVA assessed the difference between the case group (high myopia) and the control group (emmetropia) to test for allele frequency difference between these two subject groups – the main objective of the pooling experiments. It could also assess the variability of the DNA pools within subject groups, which, however, was not the main objective of the pooling-based screen and was expected to be large because each DNA pool was prepared from the DNA of 50 distinct subjects within the same subject group.