Transethnic Replication of Association of CTG18.1 Repeat Expansion of TCF4 Gene With Fuchs’ Corneal Dystrophy in Chinese Implies Common Causal Variant

Chao Xing, Xin Gong, Imran Hussain, Chiea-Chuen Khor, Donald T. H. Tan, Tin Aung, Jodhbir S. Mehta, Eranga N. Vithana, and V. Vinod Mootha

1University of Texas Southwestern Medical Center, McDermott Center for Human Growth and Development/Center for Human Genetics, Dallas, Texas, United States
2University of Texas Southwestern Medical Center, Department of Ophthalmology, Dallas, Texas, United States
3Division of Human Genetics, Genome Institute of Singapore, Singapore
4Singapore Eye Research Institute, Singapore
5Singapore National Eye Centre, Singapore
6Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
7Department of Clinical Sciences, Duke-NUS Graduate Medical School, Singapore
8Department of Neuroscience, Duke-NUS Graduate Medical School, Singapore
9Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

Correspondence: V. Vinod Mootha, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390-9057, USA; Vinod.Mootha@UTSouthwestern.edu.
Submitted: August 4, 2014
Accepted: September 26, 2014

Purpose. To test the association between the CTG18.1 trinucleotide repeat expansion of TCF4 gene and Fuchs’ endothelial corneal dystrophy (FECD) in a Chinese population.

Methods. The trinucleotide repeat polymorphism CTG18.1 was genotyped using short tandem repeat and triplet repeat primed polymerase chain reaction assays in 57 Chinese subjects with FECD and 121 controls. Statistical association of the expanded CTG18.1 allele and 18 single nucleotide polymorphisms (SNPs) across TCF4 with FECD was evaluated. To investigate the linkage disequilibrium structure of the TCF4 region, haplotype analysis was performed on our study subjects and compared with genotyping data of 97 Han Chinese and 85 Caucasians in the 1000 Genomes Project.

Results. The expanded CTG18.1 allele was associated with FECD ($P = 4.7 \times 10^{-14}$), with the odds ratio of each copy of the expanded allele estimated to be 66.5 (95% confidence interval: 12.6–350.1). Five TCF4 SNPs showed association with FECD at a nominal level ($P < 5.0 \times 10^{-2}$); however, conditional on the expanded CTG18.1 polymorphism, none of the SNPs showed association with FECD. The only haplotype associated with the disease was the one with the expansion at the CTG18.1 locus.

Conclusions. Transethnic replication of the association between the CTG18.1 repeat expansion in the TCF4 gene and FECD suggests it is a common, causal variant shared in Eurasian populations conferring significant risk for the development of FECD. Our data suggest that the expanded CTG18.1 allele is the main, if not sole, causal variant at this gene locus in the Chinese population.

Keywords: Fuchs’ corneal dystrophy, genetic diseases, TCF4, CTG18.1 trinucleotide repeat

Fuchs’ endothelial corneal disorder (FECD) is the most common inherited corneal endothelial disorder in the United States, impacting 1 in 20 Caucasians aged older than 40 years.$^1$ Of the 72,736 corneal transplants performed in the United States in 2013, Fuchs’ dystrophy was the leading indication, accounting for 14,153 cases with an additional 12,356 procedures done for related cases of corneal endothelial cell failure.$^2$ Although large-scale prevalence studies are limited, existing cross-sectional data suggest that the prevalence of FECD is higher in European countries compared with other parts of the world. Population-based studies in Iceland documented corneal guttae in 11% of females and 7% of males.$^3$ The disorder, however, is not uncommon in East Asian populations. A comparative study of the prevalence of guttae found a higher rate in Singapore (8.5%) versus Japan (5.5%).$^4$

In this disorder, the endothelium undergoes accelerated senescence and apoptosis.$^5$ Descemet’s membrane, the underlying basement membrane of the endothelium, becomes diffusely thickened and also develops focal excrescences apparent on slit-lamp biomicroscopy as guttae.$^5$ With progressive decline in the endothelial cell density, the cornea becomes edematous resulting in loss of vision. With advanced disease, bullous keratopathy can cause significant pain and corneal scarring that markedly impacts visual acuity.

Fuchs’ endothelial corneal dystrophy is a common, complex trait with genetic heterogeneity. Rare heterozygous mutations in COL8A2, SLC4A11, TCF8, LOXHD1, CLU, and AGBL1 have been implicated in FECD, but account for a small fraction of the genetic burden.$^{10}$ In 2010, a genome-wide association study (GWAS) identified association between alleles in the transcription factor 4 gene (TCF4), encoding a
member of the E-protein family (E2-2), with typical FECD, whereby the association increased the odds of having FECD by a factor of 30 in homozygous persons, defying the norm for GWAS with common variants. In this study, the most significant signal was detected with an intronic single nucleotide polymorphism (SNP), rs613872. While the association between rs613872 and FECD was replicated in other Caucasian populations, in a Chinese FECD cohort from Singapore, instead of rs613872—which was monomorphic—two other TCF4 intronic SNPs (rs17089887 and rs17089925) were significant experiment-wide (P = 7.34 × 10^−5 and 4.5 × 10^−4, respectively). Interestingly, the vicinal region of TCF4 on chromosome 18 was previously linked to FECD in three large Caucasian pedigrees. However, the risk allele rs613872 was found not to cosegregate with the trait in these three families. Sequencing of TCF4 in 96 late-onset FECD cases also failed to identify any probable pathogenic variants in coding regions of TCF4.

In 2012, a strong association between a CTG trinucleotide repeat locus (CTG18.1) in intron 3 of the TCF4 and FECD was reported, which displayed greater specificity than rs613872. The locus CTG18.1 was initially discovered by the repeat expansion detection assay and was found to be expanded in 3% of subjects in Caucasian pedigrees without known associated phenotype. Expanded alleles with greater than 37 CTG repeats at this locus were found to be unstable. Our group replicated the association with 120 FECD cases and 100 controls, finding that one copy of the expanded CTG18.1 allele increased risk of disease by 32.3-fold (95% confidence interval 11.0–77.6). Additionally, we showed the expanded allele cosegregated with the trait in complete penetrance 11 out of 29 Caucasian kindreds examined and cosegregated with incomplete penetrance in an additional 9 of these 29 families.

Transethnic mapping has been suggested as a powerful tool to detect novel disease susceptibility loci and to locate the disease causal variants. Empirical observations for many traits show high transethnic replicability of GWAS results, which support a model that the underlying causal variants at different loci are shared across ancestry groups. Therefore, transethnic analysis can increase the detection power through meta-analysis with large sample sizes. The companion repeat-primer polymerase chain reaction (TP-PCR) assay was performed to confirm the presence of an expanded CTG18.1 allele. The trinucleotide repeat polymorphism CTG18.1 was genotyped using short tandem repeat (STR) and triplet repeat primed polymerase chain reaction (TP-PCR) assays as previously described. The STR assay was performed on genomic DNA samples from all study subjects. On samples where the STR assay detected only one allele or no alleles, TP-PCR assay was performed to confirm the presence of an expanded CTG18.1 allele.

For the STR assay, a 5' FAM-labeled primer was utilized for the PCR. After polymerase chain reaction, 5 μL DNA was mixed with 10 μL internal lane standard 600 (ILS600; Promega Corp., Madison, WI, USA). Sequencing was carried out using a DNA analyzer (ABI 3730XL; Applied Biosystems, Foster City, CA, USA) and the data were analyzed using genotyping software (ABI GeneMapper 4.0; Applied Biosystems).

We performed the TP-PCR assay to detect the expanded CTG18.1 allele(s) as previously described. Locus-specific fluorescent primer P1 is designed to span an upstream region flanking the CTG18.1 polymorphism. The companion repeat-specific reverse primer P4 on the complementary strand includes five units of the CTG repeat and a 5' tail to serve as an anchor for a second reverse primer P3, which prevents progressive shortening of the amplicons with subsequent cycles. The 5' tail of primer P4 and the “common” flag primer P3 share no homology with the human genome. Polymerase chain reaction was performed with the following parameters: 200 ng of genomic DNA, 1 μmol/L of primer P1, and 0.03 μmol/L of primer P4 and 1 μmol/L of primer P3, 200 μmol/L dNTPs, 1.5 mmol/L MgCl2, and 1 U of Taq DNA polymerase (5 Prime, Gaithersburg, MD, USA). The cycling conditions were an initial denaturation of 9 minutes at 95°C, followed by 10 cycles of 95°C for 30 seconds, 62°C for 30 seconds, and 72°C for 4 minutes; and then 30 cycles of 95°C for 45 seconds, 62°C for 45 seconds, and 72°C for 4 minutes with a 15-second extension at each cycle. The final extension step was 72°C for 10 minutes. We analyzed the TP-PCR amplicons on the DNA analyzer (Applied Biosystems).

Characteristic tracing patterns of the TP-PCR electropherograms were used to distinguish samples that were homozygous for a CTG18.1 allele from those that had an expanded CTG18.1 allele not detected by STR assay. The tracings of TP-PCR were also used to detect the presence of two expanded CTG18.1 alleles in samples where the STR analysis did not detect any allele.

**Genotyping**

Genomic DNA was extracted from leukocytes of peripheral blood samples with a blood extraction kit (Nucleon; Amersham Biosciences, Buckinghamshire, UK). Previously published genotyping results of 18 SNPs across TCF4 from this cohort were reanalyzed in this study in conjunction with the CTG18.1 allele.

Genotyping results of 18 SNPs across TCF4 from this cohort were reanalyzed in this study in conjunction with the CTG18.1 allele.
considered an expanded allele, denoted as X. CTG18.1 allele was defined as P associated with FECD (Table 2). However, conditional on the CTG18.1 polymorphism, none of the SNPs showed association with FECD (Table 2). Among the six variants, rs1319637 was rare in cases (minor allele frequency [MAF] < 0.055) and it was not associated with FECD by the single-variant test (P = 1.4 × 10⁻¹), whereas the other five were common and significantly associated with FECD. Thus, we further performed haplotype-based association analysis based on these five variants (Table 3). There were five common haplotypes with frequencies greater than 0.01 in either cases or controls. The haplotype A-C-T-S-T was most common in both cases (frequency = 0.394) and controls (frequency = 0.387), and was treated as the baseline. The only haplotype associated with FECD was A-C-T-S-T (P = 2.1 × 10⁻⁵) with frequencies of 0.211 and 0.009 in cases and controls, respectively. Note that it was the only common haplotype harboring the expanded CTG18.1 allele, and it was different from the baseline haplotype only at the CTG18.1 locus. A global test resulted in a P value of 1.5 × 10⁻⁹.

**RESULTS**

The demographic information of the subjects was previously described. Specifically there were more females in cases than in controls (79% vs. 58%, P = 7.0 × 10⁻⁴), but there was no age difference between the two groups (Table 1). The dichotomized CTG18.1 trinucleotide repeat polymorphism was in HWE in both cases and controls (P > 5.0 × 10⁻²). It was associated with FECD (P = 4.7 × 10⁻¹⁴) with the odds ratio (OR) of each copy of the expanded allele estimated to be 66.5 (95% CI: 12.6–350.1).

Of the 18 previously genotyped TCF4 SNPs, four were monomorphic, including rs613872. Of the remaining 14 SNPs and the CTG18.1 polymorphism, five showed nominal association with FECD (P < 5.0 × 10⁻²), and the CTG18.1 polymorphism had the most significant signal (P = 4.7 × 10⁻¹⁴). However, conditional on the CTG18.1 polymorphism, none of the SNPs showed association with FECD (Table 2).

Allele frequencies of 16 TCF4 polymorphisms in the four populations are shown in Supplementary Table S1. According to the LD heat maps (Fig.), there was a haplotype block spanning variants rs1452787, rs1319637, rs1708987, rs2123392, CTG18.1, and rs17089925 in all four populations. Among the six variants, rs1319637 was rare in cases (minor allele frequency [MAF] < 0.055) and it was not associated with FECD by the single-variant test (P = 1.4 × 10⁻¹), whereas the other five were common and significantly associated with FECD. Thus, we further performed haplotype-based association analysis based on these five variants (Table 3). There were five common haplotypes with frequencies greater than 0.01 in either cases or controls. The haplotype A-C-T-S-T was most common in both cases (frequency = 0.394) and controls (frequency = 0.387), and was treated as the baseline. The only haplotype associated with FECD was A-C-T-S-T (P = 2.1 × 10⁻⁵) with frequencies of 0.211 and 0.009 in cases and controls, respectively. Note that it was the only common haplotype harboring the expanded CTG18.1 allele, and it was different from the baseline haplotype only at the CTG18.1 locus. A global test resulted in a P value of 1.5 × 10⁻⁹.

**DISCUSSION**

Our results show a strong association of the expanded CTG18.1 allele of the TCF4 gene with FECD in ethnic Chinese from Singapore. The genome-wide association study on the FECD trait in Caucasians found strong associations with SNPs (most strongly SNP rs613872) spanning TCF4. Association of TCF4 polymorphisms with FECD was replicated in this Singapore Chinese cohort, but the highest scoring SNP in Caucasians rs613872 was absent from this study group. Sequencing data from 1000 Genomes Project reveal that this SNP rs613872 is common in Caucasians (CEU MAF = 0.188), but extremely rare in Han Chinese (CHB MAF = 0.005). The disease association originally found in the Chinese with TCF4 SNPs rs1348047, rs2123392, rs17089887, and rs17089925 (generating three independent association signals) was no longer significant when conditioned on the CTG18.1 locus. We also found that the only haplotype associated with the disease is the one with the expansion at the CTG18.1 locus.
ambiguity into account by probabilities; on the other hand, it may indicate that the haplotype analysis did not gain power by tagging nongenotyped causal variants or integrating multiple causal variants, as the method is supposed to do when the causal variant is not genotyped.40 All these data suggest that the expanded CTG18.1 allele is the main, if not sole, functional variant at this gene locus in this population.

It is hypothesized that disease causal variants may have comparable effect size across populations if not masked by other gene-gene or gene-environment interactions.41 In a recent comprehensive survey of GWAS replicability of 28 disorders, the investigators found a "strong and significant correlation of odds ratios across European and East Asian populations, indicating that underlying causal variants are common and shared between the two ancestries."41 We found the OR of each copy of the expanded allele estimated to be 66.5 (95% CI: 12.6–350.1) in the Chinese population, mirroring its extremely high OR in Caucasians of 32.3 (95% CI: 13.4–77.6).27 The large effect sizes of the expanded CTG18.1 allele across both ethnic groups further implies this polymorphism as a causal variant.41

**Table 3.** Haplotype Association* of TCF4 Polymorphisms With FECD in a Chinese Population

<table>
<thead>
<tr>
<th>Haplotype Frequencies</th>
<th>Cases</th>
<th>Control</th>
<th>P Value</th>
<th>Global P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A C T S T</td>
<td>0.394</td>
<td>0.387</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A T C S C</td>
<td>0.272</td>
<td>0.375</td>
<td>2.5 × 10^{-1}</td>
<td>-</td>
</tr>
<tr>
<td>A T T S C</td>
<td>0.050</td>
<td>0.148</td>
<td>1.7 × 10^{-1}</td>
<td>1.5 × 10^{-9}</td>
</tr>
<tr>
<td>A C T X T</td>
<td>0.211</td>
<td>0.009</td>
<td>2.1 × 10^{-5}</td>
<td>-</td>
</tr>
<tr>
<td>A C T S C</td>
<td>0.038</td>
<td>0.055</td>
<td>5.7 × 10^{-1}</td>
<td>-</td>
</tr>
</tbody>
</table>

* Haplotypes with frequencies greater than 0.01 in either cases or controls were considered. The most common haplotype (A-C-T-S-T) was treated as the reference. A generalized linear model was fit adjusting for age and sex. Both haplotype-specific P values and global P values were reported.
The strong association of the expanded CTG18.1 allele in the TCF4 gene with FEDC in Caucasians,\textsuperscript{25-27} demonstration of its cosegregation with the trait in Caucasian pedigrees,\textsuperscript{27} and now transethnic replication of the association with similar effect sizes across populations are all compelling evidence that the repeat expansion is a causal variant. However, functional data are required to prove that the expanded CTG18.1 allele is indeed a causal mutation.

It is known that haploinsufficiency of TCF4 by microdeletions or missense mutations can lead to autosomal dominant Pitt-Hopkins syndrome, a phenotype comprising of microcephaly, encephalopathy, epilepsy, psychomotor delay, and episodic hyperventilation.\textsuperscript{42,43} This phenotypic spectrum is quite dissimilar to FECD. Therefore, we speculate that the repeat expansion may be neither sufficient nor necessary to produce the FECD phenotype. The allele CTG18.1 is likely working via complex molecular mechanisms of expanded repeats.

In summary, the transethnic replication of the association between the CTG18.1 repeat expansion in the TCF4 gene and FEDC suggests it a common, causal variant shared in Eurasian populations conferring significant risk for the development of FEDC.

**Acknowledgments**

The authors thank the patients for their participation in this study. Supported by Grants R01EY022161 (VVM) and P30EY020799 from the National Eye Institute, National Institutes of Health, Bethesda, Maryland, United States, and an unrestricted grant from Research to Prevent Blindness, New York, New York, United States. The authors alone are responsible for the content and writing of the paper.

Disclosure: C. Xing, None; X. Gong, None; I. Hussain, None; C.-C. Khor, None; D.T.H. Tan, None; T. Aung, None; J.S. Mehta, None; E.N. Vithana, None; V.V. Mootha, None

**References**

27. Mootha VV, Gong X, Ku HC, Xing C. Association and familial segregation of CTG18.1 trinucleotide repeat expansion of


