Genome-wide analyses identify common variants associated with macular telangiectasia type 2

Authors

Thomas S Scerri[1,2], Anna Quaglieri[1,2], Carolyn Cai[3], Jana Zernant[3], Nori Matsunami[4], Lisa Baird[4], Lea Scheppke[5], Roberto Bonelli[1,2], Larry A Yannuzzi[3,6], MacTel Project Consortium[7], Cathy Egan[8], Marcus Fruttiger[9], Mark Leppert[4], Rando Allikmets[3,10], Melanie Bahlo[1,2,11]*
*Corresponding author

Affiliations

[1]The Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, VIC 3052, Australia. [2]Department of Medical Biology, The University of Melbourne, Parkville, VIC 3010, Australia. [3]Department of Ophthalmology, Columbia University, New York, NY 10032, USA. [4]Department of Human Genetics, University of Utah, Salt Lake City, Utah 84132, USA. [5]The Lowy Medical Research Institute, 3366 N. Torrey Pines Court, La Jolla, CA 92037, USA. [6]Vitreous Retina Macula Consultants of New York, 460 Park Avenue, New York, NY 10022, USA. [7]see **Supplementary Table 1** and **Online Methods** for complete list. [8]Medical Retina Department, Moorfields Eye Hospital NHS Foundation Trust, London, EC1V 2PD, UK. [9]UCL Institute of Ophthalmology, University College London, London, EC1V 9EL, UK. [10]Department of Pathology and Cell Biology, Columbia University, New York, NY 10032, USA. [11]Department of Mathematics and Statistics, The University of Melbourne, Parkville, VIC 3010, Australia. Correspondence should be addressed to M.B. (bahlo@wehi.edu.au).

Idiopathic juxtafoveal retinal telangiectasis type 2 (macular telangiectasia type 2; MacTel) is a rare neurovascular degenerative retinal disease. To identify genetic susceptibility loci for MacTel, we performed a genome-wide association study (GWAS) with 476 cases and 1733 controls of European ancestry. Genome-wide significant associations ($P < 5 \times 10^{-8}$) were identified at 3 independent loci (rs73171800 at 5q14.3, $P = 7.74 \times 10^{-17}$; rs715 at 2q34, $P = 9.97 \times 10^{-14}$; rs477992 at 1p12, $P = 2.60 \times 10^{-12}$) and then replicated ($P < 0.01$) in an independent cohort of 172 cases and 1134 controls. The 5q14.3 locus is known to associate with variation in retinal vascular diameter, and the 2q34 and 1p12 loci have been implicated in the glycine/serine metabolic pathway. We subsequently found significant differences of blood serum levels of glycine ($P = 4.04 \times 10^{-6}$) and serine ($P = 2.48 \times 10^{-4}$) between MacTel cases and controls.

MacTel cases typically present at 40-60 years with abnormal right-angled juxtafoveolar capillaries and parafoveal telangiectasias. It is an uncommon disease with a 0.0045-0.1% population prevalence and no obvious sex bias[1-3]. Retinal lesions typically co-present with MacTel, including retinal transparency, outer retinal and choroidal neovascularization, lamellar holes or foveal cysts, photoreceptor dysfunction, minimal exudation, yellow-white parafoveal crystals, and retinal pigment epithelial (RPE) pigmentation abnormalities and atrophy. Central vision impairment and decreased visual acuity are the usual clinical outcomes. MacTel is a bilateral disease, but asymmetry of the eyes for disease

severity and presence of lesions is possible. The lesions also occur in 0.06-1.18% of the general population[2].

Risk factors for MacTel are largely unknown, however associations have been observed with smoking[2,4], diabetes[5,6], high BMI[6], hypertension[6] and obesity[6].

Observations of MacTel affected monozygotic twins[4,7-9], and multiplex families with vertical transmissions of MacTel[1,5,9-12], suggest a genetic etiology for the disease. The late-age of onset, low penetrance and variable phenotype as exemplified by asymptomatic affected relatives[9], and positive and negative misdiagnoses, complicate the discovery of genetic variants predisposing to MacTel. We previously screened 27 candidate genes in 8 unrelated MacTel cases but found no causative mutations[13]. Linkage analysis of 17 families with MacTel individuals identified a 15.3Mb locus on chromosome 1q41-42.2 (LOD=3.45), however sequencing of the underlying genes revealed no causative mutations[14].

Results

Discovery GWAS stage

The GWAS discovery stage included genotype data for 6,312,048 single nucleotide polymorphisms (SNPs) after quality control and imputation (including 1,093,805 SNPs genotyped on the Illumina Omni SNP chips) in 476 MacTel cases and 1,733 controls (see **Table 1** and **Online Methods**). This sample size was large enough to achieve power of at least 0.90 for risk variants with allele frequencies of 0.10-0.70, assuming a population prevalence of 0.001 for MacTel and an odds ratio (OR) effect size of 2 for the risk allele (**Supplementary Figure 1**), but the power drops to 0.53 for an odds ratio 1.5, as expected from a GWAS of modest size. Samples were of European descent as confirmed by comparative analysis with the 1000 Genomes project samples (**Supplementary Figure 2**). Population substructure was assessed using principal component analysis (PCA; **Supplementary Figures 2-4**). Heritability on the liability scale for MacTel was estimated as 0.21 or 0.74, given prevalence rates of 0.0045% and 0.1%, respectively. The GWAS was performed with logistic regression modeling with the first principal component (PC1) as a covariate. Quantile-quantile plots (**Supplementary Figure 5**) demonstrate the correction of inflation likely due to population substructure (without PC1 genomic λ = 1.141; with PC1 genomic λ = 1.035). Additional PCs increased the genomic inflation factor with no difference to the key results, and is likely due to the small sample size.

SNPs at six loci surpassed the genome-wide significance threshold (**Figure 1** and **Supplementary Tables 2 and 3**). Three of these loci (at 1p36.22, 3p24.1 and 7p21.3) were discounted as false positives after failing technical validation steps (**Online methods**). The remaining three loci (at 5q14.3, 2q34 and 1p12) included 149 SNPs that reached genome-wide significance (**Figure 2**). Seven of the 149 SNPs (**Table 2a**) were genotyped in an independent ethnically European cohort of 172 MacTel cases and 1,134 controls. Selecting SNPs for replication was governed by the ability to create TaqMan assays. We ensured that at least

one actual genotyped, rather than imputed, SNP was selected at each locus. As it was not possible to perform PCA analysis on the replication cohort, the replication analysis was performed without covariates. SNPs at all three loci replicated association ($P < 0.01$; **Table 2a**).

Locus 5q14.3

At 5q14.3, 116 SNPs were genome-wide significant within a 400kb region, including the strongest associated variant in our study, rs73171800 (per-allele OR = 2.41, 95% CI=1.96-2.96; $P = 7.74 \times 10^{-17}$). Two SNPs were selected for replication at this locus, rs17478824 and rs73173548, and both associated with MacTel in our independent cohort (**Table 2a**). Among the other genome-wide significant SNPs at this locus were rs2194025 (per-allele OR = 2.31, 95% CI = 1.88-2.84; $P = 1.97 \times 10^{-15}$) and rs17421627 (per-allele OR = 2.43, 95% CI = 1.93-3.06; $P = 3.51 \times 10^{-14}$). All five of these variants are in strong LD (pair-wise r^2 = 0.73-0.98) and their minor alleles confer risk to MacTel. Similarly, the minor alleles of both rs2194025 and rs17421627 have been associated with increased retinal venular and arterial calibers (**Table 3a**)[15,16]. A broader region encompassing this 5q14.3 locus is implicated in capillary malformation [OMIM 163000] and hereditary benign telangiectasia [OMIM 187260], two cutaneous vasculature abnormality traits that may be variable expressions of the same disorder with possible intra-familial comorbidity[17-19]. Hereditary benign telangiectasia has been linked to a 7Mb region of 5q14.1-5q14.3[17]. Capillary malformations have been linked to a 19Mb region of 5q14.1-5q15[18,20], and germline mutations of the gene RAS p21 protein activator (GTPase activating protein) 1 (*RASA1*) were identified as the likely cause for the disorder, albeit with phenotypic variability[21-24]. Murine knock-out models for *RASA1* are embryonic lethal and display vascular abnormalities[25]. A microdeletion of 5q14.3 that includes the genes *RASA1* and myocyte enhancer factor 2C (*MEF2C*) is also implicated in capillary malformations[26]. Our top associated SNP rs73171800 is intergenic at this locus, and the broader associated region is within the bounds of the 5q14.3 microdeletion, but does not include *RASA1* or *MEF2C*, which are ~1Mb proximal and 100kb distal respectively. The nearest genes are *TMEM161-B*, *TMEM161-B-AS1*, *LINC102546226* and *LINC00461*, which are largely uncharacterised. The role of *RASA1* in angiogenesis and capillary malformations makes it an interesting candidate for MacTel. However, *MEF2C* is closer to our locus, endothelially expressed, and involved in early murine vascular development[27]. *MEF2C* haploinsufficiency causes several neurological disorders, including cortical malformations, epilepsy and mental retardation[28,29]. *Mef2c* is involved in murine synaptic development and regulation of synaptic transmissions[30]. Knock-out *Mef2c* murine models are embryonic lethal[31] and display vascular abnormalities[27], including tissue dependent lumen size variability and defective vascular remodeling[32]. With respect to the late onset characteristic of MacTel, initial vasculogenesis in *Mef2c* null mice is reported normal, but anomalies arise later on in development, albeit still during embryogenesis[32]. Vascular endothelial growth factor A (*VEGFA*)[33], a key player in angiogenesis, controls *MEF2C* expression in retinal epithelial cells. An anti-angiogenic role for *MEF2C* during stress conditions has been suggested[34]. Epithelial cell-specific *Mef2c* null mice suppress retinal vascular degeneration

and promote retinal vessel regrowth in response to oxygen-induced retinopathy, but not during normal unstressed retinal vasculature development[34]. *Mef2c* retinal-specific regulation is controlled by an alternative promoter regulated by the transcription factor neural retina leucine zipper (*Nrl*)[35]. Null *Nrl* mice lack rod function[36], and mutations of *NRL* cause retinitis pigmentosa 27 [OMIM 613750][37].

Locus 2q34

At 2q34 an 83kb region contained 21 variants reaching genome-wide significance and included the gene carbamoyl-phosphate synthase 1, mitochondrial (*CPS1*). This gene produces a rate-limiting mitochondrial enzyme that performs the first step of the urea cycle, converting ammonia and hydrogen carbonate into carbamoyl phosphate (**Figure 3**). Mutations of *CPS1* cause carbamoylphosphate synthetase I deficiency [OMIM 237300] and may lead to death by hyperammonemia[38]. Our most significant SNP at this locus, rs715 (per-allele OR = 0.50, 95% CI = 0.42-0.60; $P = 9.972 \times 10^{-14}$), resides within the 3'-untranslated region (3'-UTR) of *CPS1*. We replicated associations with rs715 and rs4673553, another genome-wide significant SNP at this locus, in our independent cohort (**Table 2a**). We found the major allele of rs715 confers risk towards MacTel. Other GWASs have associated rs715 with several blood metabolite levels[39-42], in particular the major allele with decreased glycine[39,40,42-44] and serine[42] (**Table 3a**). Our estimates for the rs715 effect sizes for MacTel are comparable to those for glycine levels[42,43] (**Supplementary Figure 6**). Another SNP at this locus, rs2216405, has also been associated with blood plasma glycine levels[45,46] and metabolic ratios involving glycine[47], and approached genome-wide significance in our study (per-allele OR = 0.54, 95% CI = 0.43-0.68; $P = 9.426 \times 10^{-8}$). These three SNPs are in moderate LD (pair-wise $r^2 = 0.27$-0.48). The variant rs1047891 (renamed from rs7422339) causes a threonine to asparagine missense mutation within *CPS1* and has been associated with blood plasma levels of glycine[48,49], other metabolites[48,50-58], blood flow[56] and vasodilator response[56]. In our study rs1047891 was imputed, and whilst it was excluded due to a low imputed genotype call rate (93.4%), it would otherwise have been highly significant with the remaining available data ($P = 1.784 \times 10^{-11}$). There is strong LD between rs715 and rs1047891 (pair-wise $r^2 = 0.97$). Several studies report a marked sex effect with *CPS1* associations, with females showing a stronger association for rs715 with glycine levels[43,44], and rs1047891 with homocysteine levels[54,55]. We also observe a sex-interaction with MacTel at this locus (**Supplementary Table 4** and **Online methods**). We find that females are at nearly twice greater risk of MacTel for each additional copy of the rs715 risk allele (female specific per risk allele OR = 2.58, 95% CI = 2.00-3.36; male specific per risk allele OR = 1.40, 95% CI = 1.08-1.81). Our estimates for the rs715 sex specific effect sizes for MacTel are comparable to those for glycine levels[43] (**Supplementary Figure 6**).

Locus 1p12

At 1p12, a region of 46.4kb contained 12 genome-wide significant SNPs. We replicated this signal in our independent cohort with the imputed SNP rs483180

(**Table 2a**). This region includes the genes phosphoglycerate dehydrogenase (*PHGDH*) and 3-hydroxy-3-methylglutaryl-CoA synthase 2, mitochondrial (*HMGCS2*). GWASs have associated blood plasma serine levels with intronic *PHGDH* variants[43,45], including our most significant discovery stage SNP at this locus, rs477992 (per-allele OR = 1.70, 95% CI = 1.47-1.97; $P = 2.60 \times 10^{-12}$), and also rs478093 (per-allele OR = 1.67, 95% CI = 1.44-1.94; $P = 2.89 \times 10^{-11}$). Here, we find the minor alleles of rs477992 and rs478093 confer risk for MacTel, and the same alleles are associated with decreased serine levels in other studies (**Table 3a**)[43,45]. *PHGDH* produces a rate-limiting enzyme that performs the first step of the phosphorylated pathway of serine biosynthesis, converting 3-phosphoglycerate into phosphohydroxy pyruvate (**Figure 3**). *PHGDH* mutations may lead to PHGDH deficiency [OMIM 601815], a severe neuropathological disorder with reduced plasma and cerebrospinal fluid serine levels[59]. *HMGCS2* encodes mitochondrial 3-hydroxy-3-methylglutaryl CoA synthase (HMG-CoA synthase-2), an enzyme integral to ketogenesis. Mutations of *HMGC2* lead to hypoketotic hypoglycemia due to HMG-CoA synthase-2 deficiency[60].

Suggestive loci

There were 25 loci that showed suggestive evidence for association to MacTel ($10^{-5} > P > 5 \times 10^{-8}$; **Supplementary Table 2**). Given our associations with *CPS1* and *PHGDH*, two genes implicated in serine and glycine blood plasma levels, we explored our 25 suggestive loci for overlap with GWAS studies for these metabolite levels[39,40,42,43,45], and identified the loci at 3q21.3 and 7p11.2 (**Figure 4**).

At 3q21.3, rs1107366 has been associated with blood plasma glycine-to-serine ratios and clamp-based insulin sensitivity[43], and rs10934753 with homocysteine levels and incident ischemic stroke[58]. Here, these two specific SNPs are not associated with MacTel (rs1107366, $P = 0.021$; rs10934753, $P = 0.072$), however two other SNPs at this locus are suggestively associated, rs9820286 (per-allele OR = 0.61, 95% CI = 0.49-0.75; $P = 5.38 \times 10^{-6}$) and rs9880406 (per-allele OR = 0.63, 95% CI = 0.52-0.77; $P = 8.98 \times 10^{-6}$). While rs1107366 and rs10934753 are in strong LD ($r^2 = 0.86$) and so are rs9820286 and rs9880406 ($r^2 = 0.88$), the LD between these two pairs is very low ($r^2 < 0.012$), suggesting they are tagging two separate signals. We were able to replicate the association to MacTel with rs9880406 in our independent replication cohort (**Table 2b**). This locus lies between the genes aldehyde dehydrogenase 1 family, member L1 (*ALDH1L1*) and Kruppel-like factor 15 (*KLF15*). *ALDH1L1* is a well characterized gene that encodes the cytosolic enzyme 10-formyltetrahydrofolate dehydrogenase which catalyzes the conversion of 10-formyltetrahydrofolate into tetrahydrofolate and CO_2 or formic acid (**Figure 3**)[61]. *ALDH1L1* is a highly specific antigenic marker of astrocytes[62], a type of glial cell critical for retinal angiogenesis during development[63-65]. Transcriptome analysis suggests that astrocytes have an enrichment of metabolic pathways for amino acids such as serine, glycine and cysteine, and pathways for the production and processing of glutamate into glutamine for transport to neurons[62]. Mutations of another aldehyde dehydrogenase, *ALDH3A2*, may cause Sjögren-Larsson syndrome [OMIM 270200], a multifaceted neurological disorder with juvenile macular

dystrophy[66]. Unlike other maculopathies, MacTel and Sjögren-Larsson syndrome have co-occurring blue light reflectance abnormalities and retinal crystalline deposits. *KLF15* is a transcription factor that regulates the gene low density lipoprotein receptor-related protein 5 (*LRP5*), mutations of which have been associated with the dysplastic retinal vasculature disorder exudative vitreoretinopathy 4 [OMIM 603506][67].

At 7p11.2, rs4947534 and rs4948102 are associated with blood plasma serine levels[42] and homocysteine levels[58], respectively. We find suggestive genome-wide significance for both rs4947534 (per-allele OR = 1.44, 95% CI = 1.23-1.68; $P = 5.38 \times 10^{-6}$) and rs4948102 (per-allele OR = 1.46, 95% CI = 1.25-1.70; $P = 2.25 \times 10^{-6}$) in our discovery stage. Furthermore, our strongest signal at this locus is with rs4948102. We selected the SNPs rs4535700 and rs11238389 for replication testing and both were nominally significant ($P < 0.01$). The variants rs4947534, rs4948102 and rs11238389 are in perfect LD ($r^2 = 1$) and in very high LD with rs4535700 ($r^2 > 0.94$). The minor allele of these SNPs is associated with MacTel risk, and similarly the minor allele of rs4947534 is associated with reduced serine levels[42] and increased homocysteine levels[58] (**Table 3b**). Both rs4947534 and rs4948102 are within the gene phosphoserine phosphatase (*PSPH*), the protein product of which catalyzes the final step in the synthesis of serine by converting phosphoserine into serine (**Figure 3**). *PSPH* mutations may cause phosphoserine phosphatase deficiency, a syndrome with multiple clinical features including reduced plasma levels of serine[68].

These suggestive loci require validation in independent cohorts.

Prediction modeling

Our final model incorporated the three validated and two suggestive loci with additive effects, PC1 and one interaction term between sex and 2q34 (**Online Methods**). Additive modeling was applied since univariate logistic regression analysis showed it had no significant improvement at these 5 loci when compared to the dominant or recessive models (**Supplementary Table 5**). Since the inclusion of PC1 makes only a nominal improvement to the prediction model in the discovery cohort (**Supplementary Table 6**), and because we could not derive PC1 for the replication cohort, we estimated the probability of MacTel for each sample in the training set (discovery cohort; **Supplementary Figures 7-9**) and validation set (replication cohort; AUC = 0.679; **Supplementary Figures 10-12**) without PC1. We produced ROC curves (**Supplementary Figures 13** and **14**), and computed positive predictive values (PPVs) and negative predictive values (NPVs) (**Supplementary Table 7** and **Supplementary Figures 15** and **16**) based on these predictions.

Metabolomics analysis

Blood serum levels of 799 metabolites, including glycine and serine, were compared between 50 MacTel cases and 50 independent controls matched for age, sex, ethnicity and diabetes status (**Online methods**). Student's T-tests were highly significant for both glycine ($T = -6.98$; $P = 3.51 \times 10^{-10}$) and serine ($T = -$

5.39; $P = 4.79 \times 10^{-7}$). Given an observed inflation in our results (**Supplementary Figure 17**), we applied a correction akin to "genomic control" with an inflation factor of 1.753 ($P_{glycine} = 4.04 \times 10^{-6}$; $P_{serine} = 2.48 \times 10^{-4}$). Serum levels were lower in the MacTel cases than the controls (glycine $\log_2$(fold change) = -0.541; serine $\log_2$(fold change) = -0.382). Of all 799 metabolites, glycine and serine were ranked first and third most significant, respectively. Another glycine/serine pathway amino acid, namely threonine, was the second most significant metabolite, thereby reinforcing our glycine and serine results. Sample genotype data was not available.

In-silico variant functional exploration

We utilized the online GTEx Project[69] portal to assess for eQTL signatures the SNPs showing at least suggestive association ($P < 10^{-5}$) at each of the three validated and two suggestive loci (**Online Methods**, **Table 3**, **Supplementary Tables 8 and 9** and **Supplementary Figures 18-22**). The SNPs at 2q34 and 3q21.3 did not yield any *cis* eQTLs.

At 1p12, SNPs with strongest association to MacTel also showed strongest association to gene expression levels, with eQTLs observed in two tissues each for both *ZNF697* and *PHGDH*. Interestingly, for *PHGDH*, the direction of effect was tissue-specific.

At 5q14.3, MacTel risk alleles associated with increased expression of *TMEM161B-AS1*, the antisense RNA to *TMEM161B*, in nearly every tissue tested. The genome-wide significant 5q14.3 MacTel risk SNPs ($P < 5 \times 10^{-8}$) were not however driving the majority of these eQTL associations, instead it was the SNPs showing suggestive association to MacTel ($10^{-5} > P > 5 \times 10^{-8}$). This suggests that while our MacTel risk SNPs are indeed affecting the expression of *TMEM161B-AS1*, they may be having a stronger unobserved functional effect either on *TMEM161-AS1* or other genes.

eQTLs at the suggestive 7p11.2 locus were observed in multiple genes for a limited number of tissues, and the direction of effect was gene specific. Similar to the 1p12 locus, the SNPs with greatest association to MacTel risk also had greatest association to gene expression levels.

Methylation and H3K4me3 histone modification architecture around the five loci was analyzed in the UCSC Genome browser (**Supplementary Figures 18-22**). H3K4me3 was selected due to its known strong association with transcriptional activity. At the 1p12 locus, there is a strong H3K4me3 signature and a moderate unmethylated DNA signature around the strongest associated MacTel SNPs at the start of *PHGDH*. This supports the eQTL results and suggests these SNPs may affect expression of *PHGDH*. At 2q34 and the suggestive 3q21.3 locus there are no remarkable H3K4me3 signatures, but a weak methylation signature is evident at 2q34 indicative of gene silencing. At 5q14.3, there is a small but strong H3K4me3 mark at the most strongly associated MacTel risk SNPs downstream of *TMEM161-AS1*. Finally, at 7p11.2 the suggestive MacTel risk SNPs do not overlay any H3K4me3 modification or methylation signatures, but they do lie

immediately adjacent to such signatures, suggesting they could be influencing the DNA structure by virtue of their proximity.

Interpretation of these results is difficult given that we do not have any strong *a priori* expectation of the particular tissues to be of interest for MacTel, and the absence of eye tissue in the GTEx resource.

Discussion

Here, we identified three robustly validated and two suggestive genetic loci for MacTel. Whilst these loci include several plausible candidate genes, they account for only ~5% of the estimated heritability, suggesting more loci await discovery.

We neither find association to any candidate genes we previously screened nor to any genes within the linkage region we previously reported (here, the most significant discovery stage signal within the linkage region was with rs17352829; $P = 9.83 \times 10^{-5}$)[13,14]. A rat model for MacTel carrying an insertion-deletion mutation within the gene *Crb1* has been described[70], however we find no associated signal at this locus either.

Misdiagnoses can occur between MacTel and age-related macular degeneration (AMD). Hence, we searched for any overlap between our top association signals and those previously reported for AMD[71-73]. Amongst our suggestive association signals we find the TIMP metallopeptidase inhibitor 3 (*TIMP3*)/synapsin III (*SYN3*) locus at 22q12.3 (**Supplementary Table 2**). *TIMP3* is associated with both AMD and another maculopathy, Sorsby fundus dystrophy (SFD) [OMIM 136900], which presents earlier in life than AMD but shares some clinical features[74]. Remarkably, the *TIMP3*/*SYN3* locus has also been suggestively associated with serine and glycine levels (see The Metabolomics GWAS Server[42,45]).

Major cell types of the retina include vascular, neuronal cells (rods and cones) and glial cells (microglia, Müller glia and astrocytes). Glial cells envelop the neurons and provide them with a homeostatic environment, supporting their survival, and serving as an intermediary with blood vessels[75,76]. Müller glial cells remove neurotransmitters, including glutamate and glycine, from synaptic spaces to prevent eventual neurotoxicity and enable continued synaptic functioning. Müller cells also detoxify ammonia by expressing the enzyme glutamate-ammonia ligase (glutamine synthetase)[75,76]. Unlike retinal neuronal cells, healthy Müller cells are particularly resilient to stress, including anoxia, hypoglycemia and ischemia. Dysfunctional Müller cells could therefore lead to the pathological outcomes observed in MacTel. Indeed, we have previously reported abnormalities of Müller cells with MacTel[77,78] and observed MacTel characteristics following Müller cell ablation[79].

Retinal neurovascular coupling is important for the maintenance of local functionality, homeostasis and regulation of local blood supply[80]. The neurovascular unit that facilitates this function includes neurons, pericytes, endothelial cells and glia. Astrocytes and Müller glia regulate blood flow by

vasoconstriction and vasodilation[81], hence aberrant functioning of these cells may affect this regulated control of vascular caliber and thereby blood flow[82]. Since we previously reported an association between MacTel and increased retinal vasculature calibers[82], our findings here linking MacTel risk alleles at 2q34 and 5q14.1 with increased vasodilation, or venular and arterial calibers, respectively (**Table 3a**), may be highly relevant.

We found that the MacTel risk alleles at 1p12, 2q34 and suggestive locus 7p11.2 were previously associated with decreased serine levels (**Table 3**), and we find serine and glycine serum levels significantly lower in MacTel cases. The serine pathway involves ammonia and glutamate substrates, and it is plausible that a perturbation of this system due to defective enzyme activity may prevent Müller glia from performing their multifaceted functions normally. Altered expression of *PHGDH*, *PSHP*, *ALDH1L1* or *CPS1* (at 1p12, 7p11.2, 3q21.3 and 2q34, respectively) could lead to the observed decreases in serine levels, and consequently to hyperammonemia and hyper-glutamate conditions of a neurotoxic level, thereby causing retinal stress and damage. Müller glia and other cells react to retinal damage and stress by undergoing gliosis, a process that includes expression changes of glutamate-ammonia ligase and the induction of *VEGFA* expression. *MEF2C* (at 5q14.3, the strongest genome-wide associated locus) is under the control of VEGFA, and may normally prevent stress-induced angiogenesis. Hence a defective variant of *MEF2C* may be unable to suppress angiogenesis during retinal damage.

Müller glia also remove neuron derived carbon dioxide by expressing carbonic anhydrases (CAs)[75,76] which catalyze the reaction of carbon dioxide and water into hydrogen carbonate. One of the few drugs reported to date with a positive effect on MacTel is acetazolamide, by reducing cystic lesions[83]. Acetazolamide is a CA inhibitor (**Figure 3**). The hydrogen carbonate produced by Müller cells may be released into the vitreous body or blood stream[76]. CPS1 utilizes hydrogen carbonate and ammonia as substrates for the urea cycle (chiefly in the liver).

Previously we showed that glycolytic pathway genes have reduced expression in the MacTel macula compared to either non-macula retinal tissue of the same eye or healthy macula tissue[78], and a reduction in expression of glycolytic pathway genes following Müller cell ablation[79]. The former study observed a difference in expression of the enzyme phosphoglycerate kinase whose substrate is 3-phosphoglycerate. We find an association with *PHGDH*, whose substrate is also 3-phosphoglycerate (**Figure 3**), suggesting that both the glycolytic and serine pathways may be perturbed by *PHGDH* in the etiology of MacTel.

Despite MacTel and diabetes being associated (reflected in our sample ascertainment; **Table 1**), our study finds very limited sharing of genetic risk at the five loci for the two disorders. Only the suggestive 3q21.3 locus has shown marginal association to clamp-based insulin sensitivity, whilst the other four loci have not been associated to any diabetes-related traits (**Table 3**).

Here, we identified three validated and two suggestive loci with maximal observed odds ratio at each loci ranging from 1.43 to 2.46 for each additional

risk allele. Risk factors of similar, and smaller, effect size have been discovered using GWASs for AMD[72]. Further risk factors for MacTel will likely be identified using additional cohorts.

We report the first GWAS for MacTel and find genetic and metabolomic results implicating the glycine/serine metabolic pathway. We recognize that both glycine and serine may not be the actual driver metabolites for MacTel. However, they may still be useful biomarkers for pre-diagnostic screening in at-risk individuals, and our findings may lead to metabolite or enzymatic supplementation as a course for prevention or slowing disease progression. Further studies should validate these associations and determine if targeting those metabolites could become a useful strategy for screening or treatment. We also link MacTel with the glycolytic pathway, and confirm a reported link between MacTel and retinal venular caliber, thereby increasing our understanding of the disease.

URLs. PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/; PLINK1.9, https://www.cog-genomics.org/plink2; Genetic Power Calculator, http://pngu.mgh.harvard.edu/~purcell/gpc; LocusZoom, http://genome.sph.umich.edu/wiki/LocusZoom_Standalone; SHAPEIT, https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html; IMPUTE2, https://mathgen.stats.ox.ac.uk/impute/impute_v2.html; 1000 Genomes Project, http://www.1000genomes.org/; LINKDATAGEN, http://bioinf.wehi.edu.au/software/linkdatagen/; ExPASy[84], http://www.expasy.org/; The Metabolomics GWAS Server, http://mips.helmholtz-muenchen.de/proj/GWAS/gwas/index.php; GTEx Project, http://gtexportal.org/

Data availability statement

The MacTel consortium genotypes for the 704 samples genotyped as part of the discovery stage and that support the findings of this study have been deposited in NCBI's dbGap with accession number #######.

ACKNOWLEDGMENTS

We acknowledge The Genomics Core Facility at the University of Utah School of Medicine for processing the Illumina Human Omni5 Exome Bead chips used in this study. This study was supported by the Lowy Medical Research Institute, La Jolla, CA 92037. We thank and acknowledge all of the participants of the MacTel Project (patients and controls) who have given their time, provided biological samples and undergone extensive testing for this study. The Age-Related Eye Disease Study (AREDS) control dataset used for the discovery GWAS analyses described in this manuscript were obtained from the database at http://www.ncbi.nlm.nih.gov/ through dbGaP accession number [phs000429.v1.p1]. Funding support for AREDS was provided by the National Eye Institute (N01-EY-0-2127). We would like to thank the AREDS participants and the AREDS Research Group for their valuable contribution to this research.

We thank Elvira Agron for providing the AREDS diabetes summary statistics. We thank and acknowledge the staff and participants of the SABRE study who provided metabolomics data. Funding for the control group from the SABRE study was provided by the Wellcome Trust (WT082464), British Heart Foundation (SP/07/001/23603) and Diabetes UK (13/0004774). Funding support for the control cohort at Columbia University was, in part, provided by NIH/NEI grant EY013435. This research was supported in part by the National Institute for Health Research (NIHR) Moorfields Biomedical Research Centre and NIHR Rare Disease TRC, United Kingdom. The views expressed are those of the authors and not necessarily those of the NIHR. This work was supported by the Victorian State Government Operational Infrastructure Support and Australian Government NHMRC IRIISS. MB is supported by an NHMRC Senior Research Fellowship (APP1002098) and an NHMRC Program Grant (APP1054618). Finally, we thank Saskia Freytag, Terry Speed and Gordon K Smyth for useful discussions with respect to statistical analyses, and Kamron Khan for contributing serum samples for the metabolomics study.

AUTHOR CONTRIBUTIONS

TSS designed the study, performed the analyses, interpreted the results, reviewed the literature on MacTel and wrote the manuscript.
AQ performed the prediction modeling and other statistical analyses, and helped write the manuscript.
CC maintained the MacTel genetics database, including DNA isolation and preparation for genotyping from all MacTel subjects and controls, and data organization.
JZ performed TaqMan genotyping of the entire replication cohort and some data analysis.
NM performed the GWAS SNP chip genotyping and helped write the manuscript.
LB performed the GWAS SNP chip genotyping.
LS organized patient databases and helped write the manuscript.
RB performed the heritability and eQTL analyses, and assisted with the metabolomics analysis.
LAY obtained genetic material from the MacTel Project samples.
MacTel consortium members included clinicians and scientists who phenotyped the cohort of MacTel patients and replication controls used in the study.
CE and MF led the metabolomics study, and interpreted the metabolomics data.
ML led the genotyping group, helped design the study, interpreted the results and helped write the manuscript.
RA led the genetics group, including obtaining genetic material from MacTel Project samples and the Columbia University controls, and obtaining replication genotyping data, helped design the study, interpreted the results and helped write the manuscript.
MB led the statistical analysis group, designed the study and helped write the manuscript.

COMPETING FINANICIAL INTERESTS

The authors report no competing financial interests or conflicts of interest.

References

1. Gass, J.D.M. & Blodi, B.A. Idiopathic Juxtafoveolar Retinal Telangiectasis - Update of Classification and Follow-up-Study. *Ophthalmology* **100**, 1536-1546 (1993).
2. Klein, R. *et al.* The Prevalence of Macular Telangiectasia Type 2 in the Beaver Dam Eye Study. *American Journal of Ophthalmology* **150**, 55-62 (2010).
3. Aung, K.Z., Wickremasinghe, S.S., Makeyeva, G., Robman, L. & Guymer, R.H. The Prevalence Estimates of Macular Telangiectasia Type 2 the Melbourne Collaborative Cohort Study. *Retina-the Journal of Retinal and Vitreous Diseases* **30**, 473-478 (2010).
4. Hannan, S.R., Madhusudhana, K.C., Rennie, C. & Lotery, A.J. Idiopathic juxtafoveolar retinal telangiectasis in monozygotic twins. *British Journal of Ophthalmology* **91**, 1729-1730 (2007).
5. Chew, E.Y. Parafoveal Telangiectasis and Diabetic-Retinopathy - Reply. *Archives of Ophthalmology* **104**, 972-972 (1986).
6. Clemons, T.E. *et al.* Medical characteristics of patients with macular telangiectasia type 2 (MacTel Type 2) MacTel project report no. 3. *Ophthalmic Epidemiol* **20**, 109-13 (2013).
7. Menchini, U. *et al.* Bilateral juxtafoveolar telangiectasis in monozygotic twins. *American Journal of Ophthalmology* **129**, 401-403 (2000).
8. Siddiqui, N. & Fekrat, S. Group 2A idiopathic juxtafoveolar retinal telangiectasia in monozygotic twins. *American Journal of Ophthalmology* **139**, 568-570 (2005).
9. Gillies, M.C. *et al.* Familial Asymptomatic Macular Telangiectasia Type 2. *Ophthalmology* **116**, 2422-2429 (2009).
10. Hutton, W.L., Snyder, W.B., Fuller, D. & Vaiser, A. Focal Parafoveal Retinal Telangiectasis. *Archives of Ophthalmology* **96**, 1362-1367 (1978).
11. Oh, K.T. & Park, D.W. Bilateral juxtafoveal telangiectasis in a family. *Retina-the Journal of Retinal and Vitreous Diseases* **19**, 246-247 (1999).
12. Isaacs, T.W. & McAllister, I.L. Familial idiopathic juxtafoveolar retinal telangiectasis. *Eye* **10**, 639-642 (1996).
13. Parmalee, N.L. *et al.* Analysis of candidate genes for macular telangiectasia type 2. *Mol Vis* **16**, 2718-26 (2010).
14. Parmalee, N.L. *et al.* Identification of a potential susceptibility locus for macular telangiectasia type 2. *PLoS One* **7**, e24268 (2012).
15. Ikram, M.K. *et al.* Four Novel Loci (19q13, 6q24, 12q24, and 5q14) Influence the Microcirculation In Vivo. *Plos Genetics* **6**(2010).
16. Sim, X. *et al.* Genetic Loci for Retinal Arteriolar Microcirculation. *Plos One* **8**(2013).
17. Brancati, F. *et al.* Autosomal dominant hereditary benign telangiectasia maps to the CMC1 locus for capillary malformation on chromosome 5q14. *Journal of Medical Genetics* **40**, 849-853 (2003).
18. Breugem, C.C. *et al.* A locus for hereditary capillary malformations mapped on chromosome 5q. *Human Genetics* **110**, 343-347 (2002).

19. Onishi, Y., Ohara, K., Shikada, Y. & Satomi, H. Hereditary benign telangiectasia: image analysis of hitherto unknown association with arteriovenous malformation. *Br J Dermatol* **145**, 641-5 (2001).
20. Eerola, I. *et al.* Locus for susceptibility for familial capillary malformation ('port-wine stain') maps to 5q. *European Journal of Human Genetics* **10**, 375-380 (2002).
21. Eerola, I. *et al.* Capillary malformation-arteriovenous malformation, a new clinical and genetic disorder caused by RASA1 mutations. *Am J Hum Genet* **73**, 1240-9 (2003).
22. Revencu, N. *et al.* RASA1 mutations and associated phenotypes in 68 families with capillary malformation-arteriovenous malformation. *Hum Mutat* **34**, 1632-41 (2013).
23. de Wijn, R.S. *et al.* Phenotypic variability in a family with capillary malformations caused by a mutation in the RASA1 gene. *Eur J Med Genet* **55**, 191-5 (2012).
24. Wooderchak-Donahue, W. *et al.* RASA1 analysis: clinical and molecular findings in a series of consecutive cases. *Eur J Med Genet* **55**, 91-5 (2012).
25. Henkemeyer, M. *et al.* Vascular system defects and neuronal apoptosis in mice lacking ras GTPase-activating protein. *Nature* **377**, 695-701 (1995).
26. Carr, C.W. *et al.* 5q14.3 neurocutaneous syndrome: a novel continguous gene syndrome caused by simultaneous deletion of RASA1 and MEF2C. *Am J Med Genet A* **155A**, 1640-5 (2011).
27. Lin, Q. *et al.* Requirement of the MADS-box transcription factor MEF2C for vascular development. *Development* **125**, 4565-74 (1998).
28. Le Meur, N. *et al.* MEF2C haploinsufficiency caused by either microdeletion of the 5q14.3 region or mutation is responsible for severe mental retardation with stereotypic movements, epilepsy and/or cerebral malformations. *J Med Genet* **47**, 22-9 (2010).
29. Zweier, M. *et al.* Mutations in MEF2C from the 5q14.3q15 microdeletion syndrome region are a frequent cause of severe mental retardation and diminish MECP2 and CDKL5 expression. *Hum Mutat* **31**, 722-33 (2010).
30. Barbosa, A.C. *et al.* MEF2C, a transcription factor that facilitates learning and memory by negative regulation of synapse numbers and function. *Proc Natl Acad Sci U S A* **105**, 9391-6 (2008).
31. Lin, Q., Schwarz, J., Bucana, C. & Olson, E.N. Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C. *Science* **276**, 1404-7 (1997).
32. Bi, W., Drake, C.J. & Schwarz, J.J. The transcription factor MEF2C-null mouse exhibits complex vascular malformations and reduced cardiac expression of angiopoietin 1 and VEGF. *Dev Biol* **211**, 255-67 (1999).
33. Maiti, D., Xu, Z. & Duh, E.J. Vascular endothelial growth factor induces MEF2C and MEF2-dependent activity in endothelial cells. *Invest Ophthalmol Vis Sci* **49**, 3640-8 (2008).
34. Xu, Z. *et al.* MEF2C ablation in endothelial cells reduces retinal vessel loss and suppresses pathologic retinal neovascularization in oxygen-induced retinopathy. *Am J Pathol* **180**, 2548-60 (2012).
35. Hao, H. *et al.* The transcription factor neural retina leucine zipper (NRL) controls photoreceptor-specific expression of myocyte enhancer factor Mef2c from an alternative promoter. *J Biol Chem* **286**, 34893-902 (2011).

36. Mears, A.J. *et al.* Nrl is required for rod photoreceptor development. *Nat Genet* **29**, 447-52 (2001).
37. Bessant, D.A. *et al.* A mutation in NRL is associated with autosomal dominant retinitis pigmentosa. *Nat Genet* **21**, 355-6 (1999).
38. Gelehrter, T.D. & Snodgrass, P.J. Lethal neonatal deficiency of carbamyl phosphate synthetase. *N Engl J Med* **290**, 430-3 (1974).
39. Demirkan, A. *et al.* Insight in genome-wide association of metabolite quantitative traits by exome sequence analyses. *PLoS Genet* **11**, e1004835 (2015).
40. Raffler, J. *et al.* Genome-Wide Association Study with Targeted and Non-targeted NMR Metabolomics Identifies 15 Novel Loci of Urinary Human Metabolic Individuality. *PLoS Genet* **11**, e1005487 (2015).
41. Sabater-Lleal, M. *et al.* Multiethnic Meta-Analysis of Genome-Wide Association Studies in > 100 000 Subjects Identifies 23 FibrinogenAssociated Loci but No Strong Evidence of a Causal Association Between Circulating Fibrinogen and Cardiovascular Disease. *Circulation* **128**, 1310-1324 (2013).
42. Shin, S.Y. *et al.* An atlas of genetic influences on human blood metabolites. *Nat Genet* **46**, 543-50 (2014).
43. Xie, W. *et al.* Genetic variants associated with glycine metabolism and their role in insulin sensitivity and type 2 diabetes. *Diabetes* **62**, 2141-50 (2013).
44. Mittelstrass, K. *et al.* Discovery of sexual dimorphisms in metabolic and genetic biomarkers. *PLoS Genet* **7**, e1002215 (2011).
45. Suhre, K. *et al.* Human metabolic individuality in biomedical and pharmaceutical research. *Nature* **477**, 54-60 (2011).
46. Raffler, J. *et al.* Identification and MS-assisted interpretation of genetically influenced NMR signals in human plasma. *Genome Med* **5**, 13 (2013).
47. Illig, T. *et al.* A genome-wide perspective of genetic variation in human metabolism. *Nat Genet* **42**, 137-41 (2010).
48. Rhee, E.P. *et al.* A genome-wide association study of the human metabolome in a community-based cohort. *Cell Metab* **18**, 130-43 (2013).
49. Yu, B. *et al.* Genetic determinants influencing human serum metabolome among African Americans. *PLoS Genet* **10**, e1004212 (2014).
50. Danik, J.S. *et al.* Novel loci, including those related to Crohn disease, psoriasis, and inflammation, identified in a genome-wide association study of fibrinogen in 17 686 women: the Women's Genome Health Study. *Circ Cardiovasc Genet* **2**, 134-41 (2009).
51. Global Lipids Genetics, C. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet* **45**, 1274-83 (2013).
52. Kleber, M.E. *et al.* Genome-wide association study identifies 3 genomic loci significantly associated with serum levels of homoarginine: the AtheroRemo Consortium. *Circ Cardiovasc Genet* **6**, 505-13 (2013).
53. Kottgen, A. *et al.* New loci associated with kidney function and chronic kidney disease. *Nat Genet* **42**, 376-84 (2010).
54. Lange, L.A. *et al.* Genome-wide association study of homocysteine levels in Filipinos provides evidence for CPS1 in women and a stronger MTHFR effect in young adults. *Hum Mol Genet* **19**, 2050-8 (2010).

55. Pare, G. *et al.* Novel associations of CPS1, MUT, NOX4, and DPEP1 with plasma homocysteine in a healthy population: a genome-wide evaluation of 13 974 participants in the Women's Genome Health Study. *Circ Cardiovasc Genet* **2**, 142-50 (2009).
56. Summar, M.L. *et al.* Relationship between carbamoyl-phosphate synthetase genotype and systemic vascular function. *Hypertension* **43**, 186-91 (2004).
57. van Meurs, J.B. *et al.* Common genetic loci influencing plasma homocysteine concentrations and their effect on risk of coronary artery disease. *Am J Clin Nutr* **98**, 668-76 (2013).
58. Williams, S.R. *et al.* Genome-wide meta-analysis of homocysteine and methionine metabolism identifies five one carbon metabolism loci and a novel association of ALDH1L1 with ischemic stroke. *PLoS Genet* **10**, e1004214 (2014).
59. Klomp, L.W. *et al.* Molecular characterization of 3-phosphoglycerate dehydrogenase deficiency--a neurometabolic disorder associated with reduced L-serine biosynthesis. *Am J Hum Genet* **67**, 1389-99 (2000).
60. Bouchard, L. *et al.* Mitochondrial 3-hydroxy-3-methylglutaryl-CoA synthase deficiency: clinical course and description of causal mutations in two patients. *Pediatr Res* **49**, 326-31 (2001).
61. Hong, M. *et al.* Isolation and characterization of cDNA clone for human liver 10-formyltetrahydrofolate dehydrogenase. *Biochem Mol Biol Int* **47**, 407-15 (1999).
62. Cahoy, J.D. *et al.* A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J Neurosci* **28**, 264-78 (2008).
63. Gariano, R.F. & Gardner, T.W. Retinal angiogenesis in development and disease. *Nature* **438**, 960-6 (2005).
64. Watanabe, T. & Raff, M.C. Retinal astrocytes are immigrants from the optic nerve. *Nature* **332**, 834-7 (1988).
65. Dorrell, M.I., Aguilar, E. & Friedlander, M. Retinal vascular development is mediated by endothelial filopodia, a preexisting astrocytic template and specific R-cadherin adhesion. *Invest Ophthalmol Vis Sci* **43**, 3500-10 (2002).
66. De Laurenzi, V. *et al.* Sjogren-Larsson syndrome is caused by mutations in the fatty aldehyde dehydrogenase gene. *Nat Genet* **12**, 52-7 (1996).
67. Toomes, C. *et al.* Mutations in LRP5 or FZD4 underlie the common familial exudative vitreoretinopathy locus on chromosome 11q. *Am J Hum Genet* **74**, 721-30 (2004).
68. Veiga-da-Cunha, M. *et al.* Mutations responsible for 3-phosphoserine phosphatase deficiency. *Eur J Hum Genet* **12**, 163-6 (2004).
69. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-5 (2013).
70. Zhao, M. *et al.* A new CRB1 rat mutation links Muller glial cells to retinal telangiectasia. *J Neurosci* **35**, 6093-106 (2015).
71. Black, J.R. & Clark, S.J. Age-related macular degeneration: genome-wide association studies to translation. *Genet Med* (2015).
72. Fritsche, L.G. *et al.* Seven new loci associated with age-related macular degeneration. *Nat Genet* **45**, 433-9, 439e1-2 (2013).

73. Fritsche, L.G. *et al.* A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nat Genet* **48**, 134-43 (2016).
74. Weber, B.H., Vogt, G., Pruett, R.C., Stohr, H. & Felbor, U. Mutations in the tissue inhibitor of metalloproteinases-3 (TIMP3) in patients with Sorsby's fundus dystrophy. *Nat Genet* **8**, 352-6 (1994).
75. Giaume, C., Kirchhoff, F., Matute, C., Reichenbach, A. & Verkhratsky, A. Glia: the fulcrum of brain diseases. *Cell Death Differ* **14**, 1324-35 (2007).
76. Bringmann, A. *et al.* Muller cells in the healthy and diseased retina. *Prog Retin Eye Res* **25**, 397-424 (2006).
77. Powner, M.B. *et al.* Perifoveal Muller Cell Depletion in a Case of Macular Telangiectasia Type 2. *Ophthalmology* **117**, 2407-2416 (2010).
78. Len, A.C. *et al.* Pilot application of iTRAQ to the retinal disease Macular Telangiectasia. *J Proteome Res* **11**, 537-53 (2012).
79. Chung, S.H. *et al.* Differential Gene Expression Profiling after Conditional Muller-Cell Ablation in a Novel Transgenic Model. *Investigative Ophthalmology & Visual Science* **54**, 2142-2152 (2013).
80. Usui, Y. *et al.* Neurovascular crosstalk between interneurons and capillaries is required for vision. *J Clin Invest* **125**, 2335-46 (2015).
81. Metea, M.R. & Newman, E.A. Glial cells dilate and constrict blood vessels: a mechanism of neurovascular coupling. *J Neurosci* **26**, 2862-70 (2006).
82. Tikellis, G. *et al.* Retinal vascular caliber and macular telangiectasia type 2. *Ophthalmology* **116**, 319-23 (2009).
83. Chen, J.J. *et al.* Decreased macular thickness in nonproliferative macular telangiectasia type 2 with oral carbonic anhydrase inhibitors. *Retina* **34**, 1400-6 (2014).
84. Artimo, P. *et al.* ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res* **40**, W597-603 (2012).

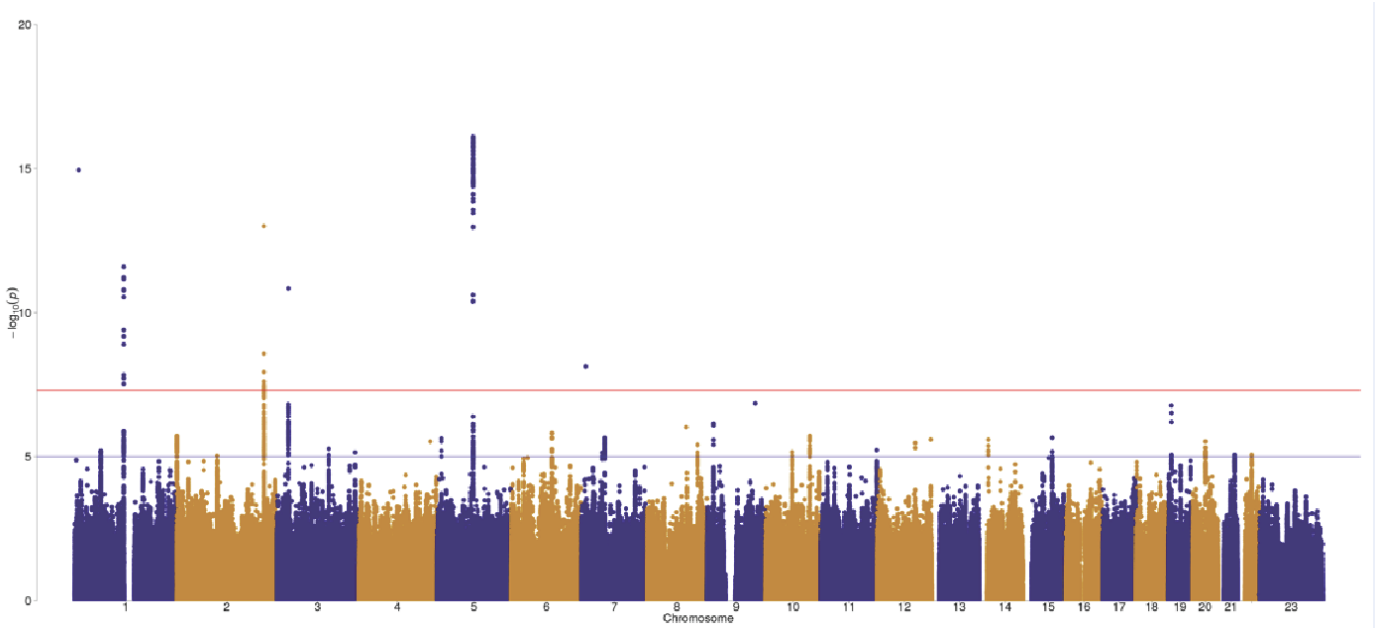


Figure 1 Genome-wide plot of association. SNPs on sequential chromosomes are alternatively colored blue and orange. The x-axis is chromosomal position and the y-axis represents the $-\log_{10}(P \text{ value})$ of association for each SNP to MacTel as tested by logistic regression with PC1 as a predictor to correct for population stratification. Analysis was performed with 476 MacTel cases and 1,733 controls. Red and blue horizontal lines are thresholds for genome-wide and suggestive association.

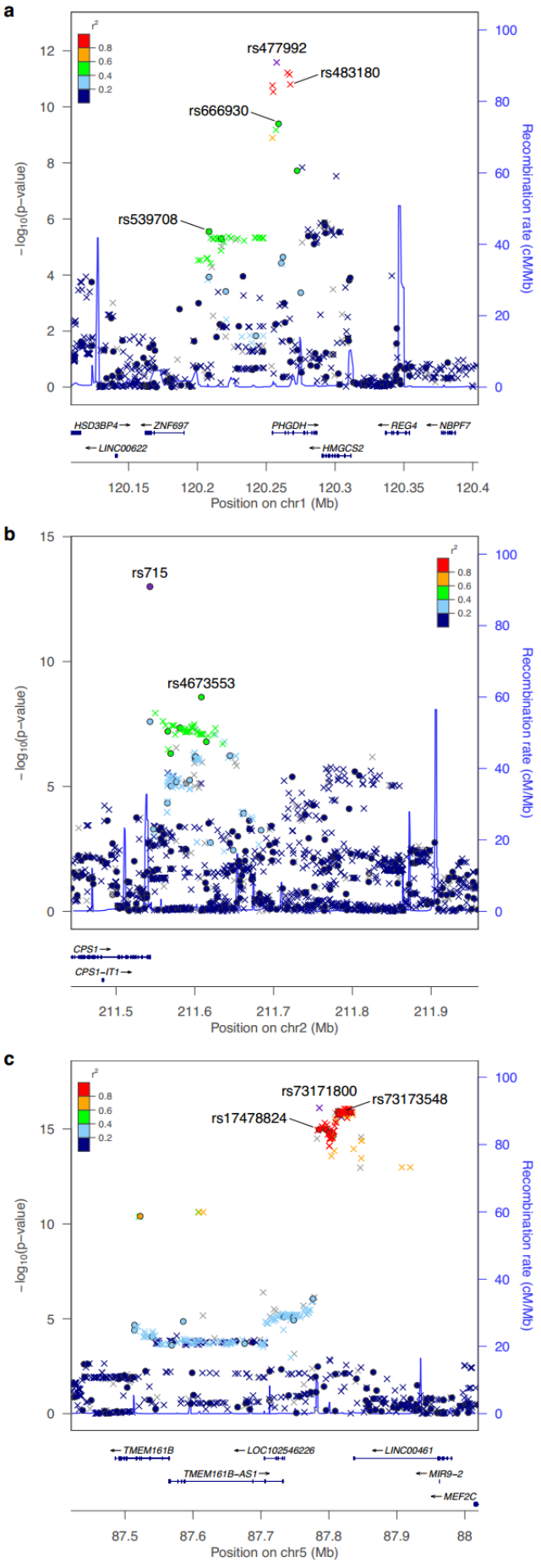
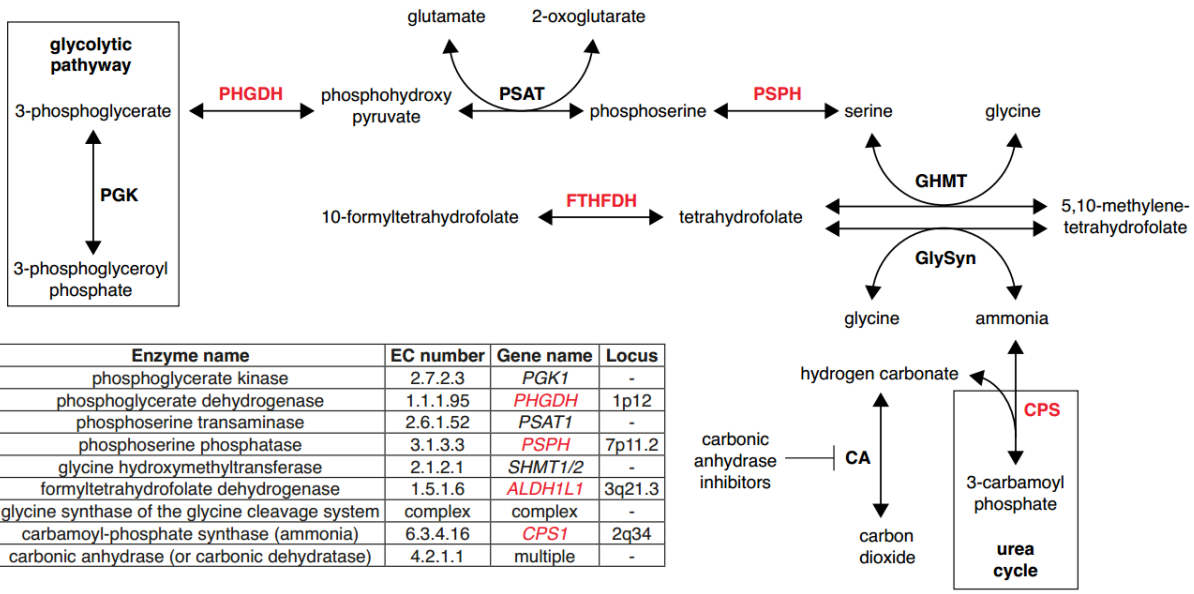
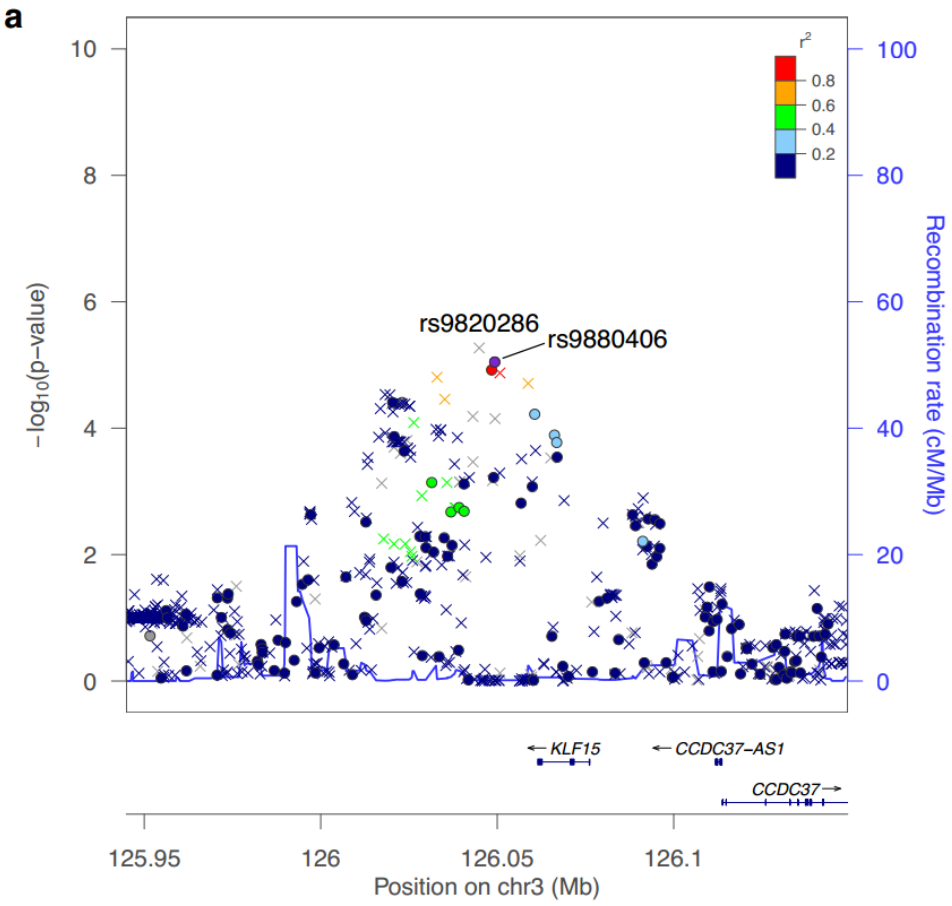

a
rs477992
rs483180
rs666930
rs539708
HSD3B4
ZNF697
PHGDH
REG4
NBPF7
LINC00622
HMGCS2
Position on chr1 (Mb)
b
rs715
rs4673553
CPS1
CPS1-IT1
Position on chr2 (Mb)
c
rs73171800
rs73173548
rs17478824
TMEM161B
LOC102546226
LINC00461
TMEM161B-AS1
MIR9-2
MEF2C
Position on chr5 (Mb)
Recombination rate (cM/Mb)
−log10(p-value)

Figure 2 Regional plots of association. (**a**-**c**) Association results are shown for the analyzed SNPs, along with recombination rates (blue lines) and the location of known genes (named blue horizontal lines), within the three loci associated with genome-wide significance at 1p12 (**a**), 2q34 (**b**), 5q14.3 (**c**). For each plot, the x-axis is scaled such that 100kb flanks each end of the loci as defined by the first and last SNP reaching suggestive significance ($P < 10^{-5}$), and the y-axis represents the $-\log_{10}$(*P* value) of association for each SNP. Individual SNPs are represented as colored circles (genotyped) or crosses (imputed). For (**a**), (**b**) and (**c**), the sole purple circle represents the most significantly associated SNP at that locus in the discovery stage, and the other markers are colored according to their LD with that SNP. Recombination rates and LD are estimated from the EUR 1000 Genomes Project March 2012 release, build hg19.



Reaction ID	Enzyme name	EC number	Gene name	Locus
PGK	phosphoglycerate kinase	2.7.2.3	*PGK1*	-
PHGDH	phosphoglycerate dehydrogenase	1.1.1.95	*PHGDH*	1p12
PSAT	phosphoserine transaminase	2.6.1.52	*PSAT1*	-
PSPH	phosphoserine phosphatase	3.1.3.3	*PSPH*	7p11.2
GHMT	glycine hydroxymethyltransferase	2.1.2.1	*SHMT1/2*	-
FTHFDH	formyltetrahydrofolate dehydrogenase	1.5.1.6	*ALDH1L1*	3q21.3
GlySyn	glycine synthase of the glycine cleavage system	complex	complex	-
CPS	carbamoyl-phosphate synthase (ammonia)	6.3.4.16	*CPS1*	2q34
CA	carbonic anhydrase (or carbonic dehydratase)	4.2.1.1	multiple	-

Figure 3 Pathway analysis. Indicated are the enzymes encoded by genes associated to MacTel on chromosomes 1, 2, 3 and 7 (highlighted in red). A subsection of the glycolytic pathway highlights the enzyme encoded by *PGK*, whose differential expression has been observed in MacTel. The enzymes encoded by both *PHGDH* and *PGK* share the common substrate 3-phosphoglycerate. Also shown is the position that carbonic anhydrase inhibitors, used to treat MacTel, may play in this pathway. For simplicity, common substrates (such as protons, water, NADH, NADPH, ATP and their derivatives) are not shown. Adapted from the ExPASy Bioinformatics Resource Portal[84].


a
r²
0.8
0.6
0.4
0.2
rs9820286
rs9880406
−log10(p−value)
Recombination rate (cM/Mb)
KLF15
CCDC37−AS1
CCDC37
125.95
126
126.05
126.1
Position on chr3 (Mb)

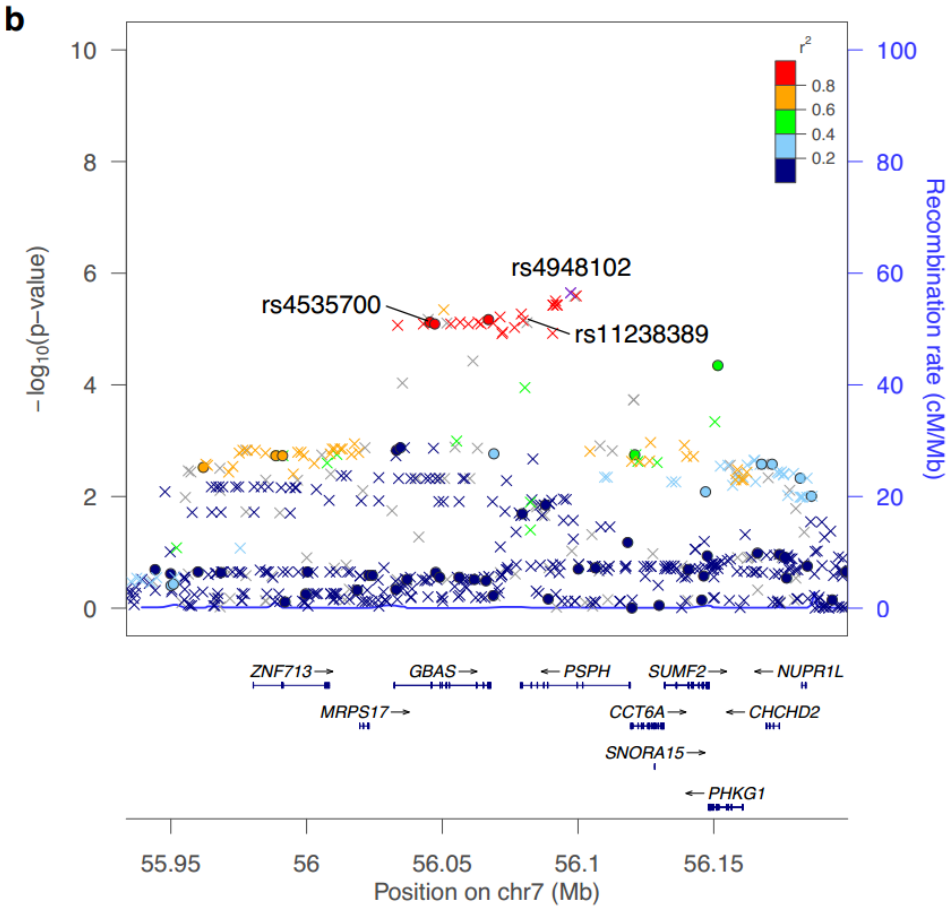

b
r²
0.8
0.6
0.4
0.2
rs4948102
rs4535700
rs11238389
−log10(p−value)
Recombination rate (cM/Mb)
ZNF713
GBAS
PSPH
SUMF2
NUPR1L
MRPS17
CCT6A
CHCHD2
SNORA15
PHKG1
55.95
56
56.05
56.1
56.15
Position on chr7 (Mb)

Figure 4 Regional plots of association. (**a**-**b**) Association results are shown for the analyzed SNPs, along with recombination rates (blue lines) and the location of known genes (named blue horizontal lines), within the two loci associated with suggestive significance at 3q21.3 (**a**) and 7p11.2 (**b**). For each plot, the x-axis is scaled such that 100kb flanks each end of the loci as defined by the first and last SNP reaching suggestive significance ($P < 10^{-5}$), and the y-axis represents the $-\log_{10}(P \text{ value})$ of association for each SNP. Individual SNPs are represented as colored circles (genotyped) or crosses (imputed). For (**b**) the sole purple circle represents the most significantly associated SNP at that locus in the discovery stage, and the other markers are colored according to their LD with that SNP, however for (**a**) the coloring is with respect to rs9880406. Recombination rates and LD are estimated from the EUR 1000 Genomes Project March 2012 release, build hg19.

Table 1 Sample characteristics

Sample	N_{total}	N_{males}	$N_{females}$	Source*	Age range in years[$]	% with diabetes[#]
Discovery cases	476	192	284	MacTel Project	45-86 (mean=63.2, SD=8.04)	33.4
Discovery controls	1,733	735	998	MacTel Project ($N_{subtotal}$=76)	39-84 (mean=62.2, SD=10.64)	6.6
				AREDS ($N_{subtotal}$=1657)	55-80	6.3
Replication cases	172	74	98	MacTel Project	38-85 (mean=62.4, SD=8.45)	33.3
Replication controls	1,134	629	505	Columbia University controls ($N_{subtotal}$=629)	74.8±7.1	-
				CCHMC controls ($N_{subtotal}$=505)	2-52	-

*MacTel Project consortium; AREDS: Age-Related Eye Disease Study; Columbia University controls; CCHMC: Cincinnati Children's Hospital Medical Center; [$]Age extracted from the literature for AREDS, Columbia University controls and CCHMC controls; [#]Diabetes status for the Columbia University and CCHMC controls is unknown but presumed to be at baseline levels.

Table 2 Significant associations of three robustly validated (a) and two suggestive (b) loci with MacTel in the discovery and replication stages

a)

Locus	Most significant discovery stage SNP at locus (imputed?)	OR (95% CI) *P* value#	Gene(s) (genetic location)	SNPs selected for replication							
				SNP selected for replication (imputed?)	Position$ (r^2 LD to most significant SNP at locus)	1000G minor allele (freq.)*	Stage^	Minor allele frequency		OR (95% CI) per minor allele#	*P* value#
								Cases	Controls		
1p12	rs477992 (yes)	1.70 (1.47-1.97) $P = 2.60 \times 10^{-12}$	*PHGDH* (intronic)	rs539708 (no)	120208503 (0.56)	T (0.415)	Discovery	0.4853	0.3904	1.42 (1.22-1.64)	2.82×10^{-6}
							Replication	0.4394	0.3919	1.23 (0.91-1.67)	0.1761
							Meta-analysis	0.4774	0.3910	1.38 (1.21-1.57)	1.54×10^{-6}
				rs666930 (no)	120258970 (0.53)	T (0.458)	Discovery	0.5830	0.4602	1.60 (1.38-1.85)	4.02×10^{-10}
							Replication	0.4949	0.4623	1.14 (0.85-1.54)	0.3744
							Meta-analysis^^	0.5679	0.4610	1.39 (1.00-1.92)	0.047
				rs483180 (yes)	120267505 (0.98)	G (0.315)	Discovery	0.4258	0.3021	1.68 (1.44-1.95)	1.59×10^{-11}
							Replication	0.4142	0.3115	1.64 (1.26-2.13)	0.0002543
							Meta-analysis	0.4227	0.3046	1.67 (1.46-1.90)	1.76×10^{-14}
2q34	rs715 (no)	0.50 (0.42-0.60) $P = 9.97 \times 10^{-14}$	*CPS1* (3'UTR)	rs715 (no)	211543055 (=)	C (0.291)	Discovery	0.1796	0.3090	0.50 (0.42-0.60)	9.97×10^{-14}
							Replication	0.2135	0.3213	0.58 (0.40-0.82)	0.002384
							Meta-analysis	0.1853	0.3138	0.52 (0.44-0.61)	1.14×10^{-15}
				rs4673553 (no)	211608379 (0.41)	G (0.458)	Discovery	0.3466	0.4570	0.63 (0.55-0.74)	2.66×10^{-9}
							Replication	0.3505	0.4641	0.62 (0.45-0.84)	0.002378
							Meta-analysis	0.3473	0.4598	0.63 (0.55-0.72)	2.38×10^{-11}
5q14.3	rs73171800 (yes)	2.41 (1.96-2.96) $P = 7.74 \times 10^{-17}$	*TMEM161-B /LINC00461* (intergenic)	rs17478824 (no)	87785624 (0.97)	T (0.086)	Discovery	0.1775	0.0866	2.34 (1.90-2.87)	1.08×10^{-15}
							Replication	0.2121	0.0855	2.94 (2.00-4.33)	3.90×10^{-8}
							Meta-analysis^^	0.1835	0.0861	2.47 (2.03-300)	1.09×10^{-19}
				rs73173548 (no)	87813971 (0.89)	G (0.080)	Discovery	0.1712	0.0796	2.46 (1.99-3.04)	1.16×10^{-16}
							Replication	0.1667	0.0824	2.21 (1.48-3.32)	0.0001199
							Meta-analysis	0.1704	0.0807	2.40 (1.99-2.90)	7.27×10^{-20}

b)

Locus	Most significant discovery stage SNP at locus (imputed?)	OR (95% CI) *P* value#	Gene(s) (genetic location)	SNPs selected for replication							
				SNP selected for replication (imputed?)	Position$ (r^2 LD to most significant SNP at locus)	1000G minor allele (freq.)*	Stage^	Minor allele frequency		OR (95% CI) per minor allele#	*P* value#
								Cases	Controls		
3q21.3	rs9820286 (yes)	0.61 (0.49-0.75) $P = 5.38 \times 10^{-6}$	*ALDH1L1 /KLF15* (intergenic)	rs9880406 (no)	126049305 (0.88)	A (0.199)	Discovery	0.1345	0.1982	0.63 (0.52-0.77)	8.98×10^{-6}
							Replication	0.1294	0.1898	0.63 (0.45-0.88)	0.00708
							Meta-analysis	0.1331	0.1949	0.63 (0.53-0.75)	2.07×10^{-7}
7p11.2	rs4948102 (yes)	1.46 (1.25-1.70) $P = 2.25 \times 10^{-6}$	*PSPH* (intronic)	rs4535700 (no)	56045448 (0.94)	T (0.258)	Discovery	0.3319	0.2458	1.43 (1.22-1.67)	7.55×10^{-6}
							Replication	0.3500	0.2687	1.44 (1.14-1.83)	0.002477
							Meta-analysis	0.3367	0.2548	1.43 (1.26-1.63)	6.54×10^{-8}
				rs11238389 (yes)	56079744 (1.00)	A (0.273)	Discovery	0.3398	0.2525	1.43 (1.23-1.68)	7.04×10^{-6}
							Replication	0.3550	0.2992	1.28 (0.99-1.64)	0.05517
							Meta-analysis	0.3439	0.2649	1.39 (1.21-1.58)	1.42×10^{-6}

$Position in base pairs (bp) based on the hg19 reference genome, and the r^2 linkage disequilibrium (LD) estimated from the discovery cohort sample genotypes; *Minor allele and frequency determined from the 1000Genomes project European population (http://1000genomes.org/); #ORs and *P* values derived from a logistic regression model, using PC1 as a covariate in the discovery stage; ^Meta-analyses performed with random-effect modeling, although results for only two SNPs (^^) differed from fixed-effects modeling.

Table 3 Directions of association of MacTel risks alleles with respect to metabolite levels, *cis* eQTLs and other traits for the three robustly validated (a) and two suggestive (b) loci

a)

MacTel risk loci	Nearest genes	Direction of effect with respect to MacTel risk haplotypes^	
		Increased	**Decreased**
1p12	*PHGDH*	eQTL: *ZNF697* (2) eQTL: *PHGDH* (1)	serine eQTL: *PHGDH* (1)
2q34	*CPS1*	fibrinogen betaine pyroglutamine glutaroyl carnitine homoarginine nitric oxide metabolites vasodilator response nitropusside induced bloodflow high-density lipoprotein	glycine creatine N-acetylglycine X-08988 serine glycine/threonine ratio homocysteine
5q14.3	*TMEM161-B/ LINC00461*	retinal venular caliber retinal arterial caliber eQTL: *TMEM161B-AS1* (20)	n/a

b)

MacTel risk loci	Nearest genes	Direction of effect with respect to MacTel risk haplotypes^	
		Increased	**Decreased**
*3q21.3	*ALDH1L1/ KLF15*	clamp-based insulin sensitivity* glycine/serine ratio* delta-POST** incident ischemic stroke risk	n/a
7p11.2	*PSPH*	delta-POST** eQTL: *CCT6A* (3) eQTL: *GBAS* (6) eQTL: *NUPR1L* (1) eQTL: *PSPH* (1)	serine eQTL: *SUMF2* (5)

^ Direction of effect for *cis* eQTLs (observed in the given number of tissues) as derived from GTEx Project. Metabolite levels as observed in blood plasma or serum.
* At 3q21.3, the minor alleles of rs1107366 and rs10934753 were marginally associated with increased risk to MacTel.
** Difference between pre- and post-methionine load test for circulating homocysteine levels (delta-POST)

Online Methods

Study population

Our MacTel Project consortium (**Supplementary Table 1**) recruited cases and controls at 23 participating clinical centers in 7 countries (Australia, Germany, France, The UK, Switzerland, Israel, and the USA). Informed written consent was obtained in accordance with ethics protocols for human subjects approved by the appropriate governing body at each site in accordance with the Declaration of Helsinki. Protocols and records of consent were centrally managed by the EMMES Corporation (Rockville, Maryland). The following ethics boards granted approval for human subject enrollment; Quinze-Vingts, Paris, France: Comite De Protection Des Personnes Hopital Saint-Antonie; Centre for Eye Research, Victoria, Australia: The Royal Victorian Eye & Ear Hospital; Clinique Ophtalmolgie de Creteil, Paris, France: Comite De Protection Des Personnes Hopital Saint-Antonie; Hospital Lariboisiere, Paris, France: Comite De Protection Des Personnes Hopital Saint-Antonie; Jules Stein Eye Institute, UCLA,CA, USA: The UCLA Institutional Review Board; Lions Eye Institute, Nedlands, Australia: Sire Charles Gairdner Group Human Research Ethics Committee; Manhattan Eye, Ear & Throat Hospital, NY, USA: Lenox Hill Hospital Institutional Review Board; Moorfields Eye Hospital, London, UK: National Research Ethics Service; Retina Associates of Cleveland, Inc., Cleveland, USA: Sterling Institutional Review Board; Save Sight Institute, Sydney, Australia: South Eastern Sydney Illawarra Area Health Service Human Research Ethics Committee – Northern Hospital Network; Scripps Research Institute, La Jolla, USA: Scripps Institutional Review Board; St. Franziskus Hospital, Munster, Germany: Ethik-Kommission Der Arztekammer Westfalen-Lippe Und der Medizinishchen Fakultat der Westfallschen Wilhelms-Universitat; The Goldschleger Eye Institute, Tel Hashomer, Israel: Ethics Committee The Chaim Sheba Medical Center; The New York Eye and Ear Infirmary, NY, USA: The Institutional Review Board of the New York Eye and Ear Infirmary; The Retina Group of Washington, Olympia, USA: Western Institutional Review Board; University of Bonn, Bonn, Germany: Rheinische Friedrich-Wilhelms-Universitat Ethik-Kommission; University of Chicago, Chicago, USA: The University of Chicago Division of Biological Sciences – The Pritzker School Institutional Review Board; University of Michigan, Ann Arbor, USA: Medical School Institutional Review Board (IRBMED); University of Wisconsin, Madison, USA: Office of Clinical Trials University of Wisconsin School of Medicine and Public Health; The Wilmer Eye Institute of Johns Hopkins University, Baltimore, Maryland: Johns Hopkins School of Medicine Office of Human Subjects Research; Scheie Eye Institute University of Pennsylvania, Philadelphia, USA: University of Pennsylvania Office of Regulatory Affairs; University of Bern, Bern, Switzerland: Kantonale Ethikkommission Bern; John Moran Eye University of Utah, Salt Lake City, USA: The University of Utah Institutional Review Board; Bascom Palmer Eye Institute University of Miami, Miami, USA: The University of Miami Human Subjects Research Office; Columbia University, NY, USA: Columbia University Medical Center Institutional Review Board Category 4 waiver for research involving specimens obtained from de-identified subjects. Participants were given a standardized ophthalmic examination, including best corrected visual

acuity, fundus photography, fluorescein angiography, optical coherence tomography, and blue light reflectance. Images were adjudicated at the Reading Center at Moorfields Eye Hospital, London. Diagnoses were made in accordance with the criteria described by Clemons et al. (2010)[85] based on Gass and Blodi (1993)[1]. Retinal images were assessed for loss of transparency in the perifoveal region, dilated and telangiectatic blood vessels, especially in the temporal retina, and crystalline deposits. Participants are re-evaluated at regular intervals over the course of the study. MacTel Project consortium cases were used in both the GWAS discovery and replication stages.

GWAS discovery stage control samples (N_{total}=1,733) were predominantly sourced from the Age-Related Eye Disease Study (AREDS, $N_{subtotal}$=1,657)[86,87]. The remainder came from the MacTel Project consortium ($N_{subtotal}$=76; most being spouses of MacTel cases) after confirmation as unaffected for MacTel by expert ophthalmologists. AREDS is a long-term natural history and clinical study of AMD and age-related cataracts. Patients were initially enrolled from 55-80 years of age. We used the controls samples that, as described in dbGap (phs000429.v1.p1), "are all Caucasian, do not have age-related macular degeneration (AMD) and were further screened to also exclude individuals with cataracts, retinitis pigmentosa, color blindness, other congenital eye problems, LASIK, artificial lenses, and other eye surgery".

Replication stage control samples were of European ancestry and came from the electronic medical records and genomics (eMERGE) Cincinnati Children's Hospital Medical Center (CCHMC, $N_{subtotal}$=505) study and Columbia University ($N_{subtotal}$=629). The CCHMC controls were diagnosed with eosinophilic esophagitis, had an age range of 2-52, and had Illumina Omni5 SNP chip data available from dbGAP (project "Better Outcomes for Children: GWAS from Cincinnati Children's Hospital Medical Center (CCHMC) - eMERGE Phase II data"; phs000494.v1.p1)[88]. The Columbia University controls were confirmed as being of European ancestry by questionnaire during recruitment, have been used as controls for studies of age-related macular degeneration (AMD) as described previously[72,89,90], were of an advanced age (mean age 74.8±7.1 years), did not exhibit any distinguishing signs of macular/retinal disease after clinical examination by trained ophthalmologists, and had no known family history of retinal disease.

DNA sample preparation and quality control

For MacTel Project consortium samples and Columbia University control samples, DNA was extracted from peripheral venous blood (Qiagen blood maxi kit 51194, Valencia, CA). DNA concentration was determined with the Qubit Fluorometer (Thermo Fisher Scientific, Waltham, MA). DNA samples of low purity were subjected to column purification (Qiagen blood and tissue kit 69504).

GWAS genotyping and quality control

The GWAS discovery stage used a final cleaned dataset of 2,209 samples (N_{cases}=476; $N_{controls}$=1,733; **Table 1**).

AREDS genotypes were obtained from NCBI's dbGap (project "NEI Age-Related Eye Disease Study (AREDS) - Genetic Variation in Refractive Error Substudy"; phs000429 substudy of phs000001). AREDS samples were genotyped with the Illumina Omni2.5 SNP chips. Genotypes for 2,182,680 variants/probes for 1,657 samples were downloaded from dbGap., and those with minor allele frequencies (MAFs) < 1%, or else monomorphic, were removed. Complementary bi-allelic (i.e. AT or CG) SNPs were removed due to strand uncertainty; these would otherwise cause problems when merging with the MacTel Project genotypes. For SNPs with identical chromosome identifiers and base-pair positions, only the one with least missingness was kept. Identity-by-descent (IBD) sharing was assessed using PLINK1.9[91,92] and a 0.04 kinship coefficient threshold was determined to establish independence of samples. Further filtering took place as described below in conjunction with the MacTel Project samples.

From our MacTel Project cohort, a total of 704 samples (including MacTel cases and controls) were genotyped with the Illumina Omni5 SNP chip for the discovery stage. Illumina's GenomeStudio software with default genotyping settings produced genotypes for a total of 4,641,218 variants/probes. DNA samples with <95% genotype call rate or unexpected or unusual sex-chromosome call rates were flagged and removed. Monomorphic probes or variants with an MAF <1% were filtered out. Complementary bi-allelic (i.e. AT or CG) SNPs were removed due to strand uncertainty. For SNPs with identical chromosome identifiers and base-pair positions, only the one with least missingness was kept. Sample independence was tested by assessing IBD sharing using PLINK1.9[91,92] and applying a 0.04 kinship coefficient threshold. When IBD analysis suggested relatedness, a single sample was selected, usually being the sample with least missingness.

AREDS and MacTel project SNP genotypes were merged and differential missingness was assessed between the two cohorts by Fisher's exact tests and an exclusion threshold of $P<0.001$. A mock GWAS was performed on the 76 MacTel Project consortium controls against the 1,657 AREDS controls. A quantile-quantile plot of this analysis shows little deviation between the expected and observed *P* values (**Supplementary Figure 23**). Further IBD analysis was performed to ensure no relatedness between AREDS and MacTel project samples. Principal component analysis (PCA) of the samples was performed with EIGENSOFT and PLINK to identify any population substructure (**Supplementary Figures 2-4**). This was achieved using 29,149 independent genome-wide SNPs selected with PLINK using the "--indep-pairwise" option; this assessed pairwise Pearson's correlations (*r*) within windows of 1000 adjacent SNPs and greedily pruned these SNPs until no such pairs remained with $r^2 > 0.04$, before sliding the window a further 50bp along.

Imputation of SNPs was performed with SHAPEIT (v2.r790)[93] and IMPUTE2 (v2.3.2)[94] following the software's best practice guidelines. Reference genomes were obtained from the 1000 Genomes project[95]. Compare **Figure 1** with **Supplementary Figure 24** to see the effect of imputation on the GWAS results.

Prior to statistical analysis, a final round of filtering removed any SNPs with >5% genotype missingness, a Hardy-Weinberg equilibrium exact test $P < 10^{-6}$, or a minor allele frequency < 5%.

GWAS statistical analysis

For the discovery stage 2,209 samples (476 MacTel cases and 1,733 controls) with genotypes for 6,310,381 SNPs were analyzed by logistic regression with PLINK1.9[92]. Principal component 1 (PC1) from the PCA was added to the logistic regression model as a covariate in the discovery stage. The scree plot indicated that additional PCs would have a negligible contribution (**Supplementary Figure 3**). Quantile-quantile plots (**Supplementary Figure 5**) and Manhattan plots (**Figure 1** and **Supplementary Figure 24**) were created in R with the qqman package[96]. The top associated loci ($P < 10^{-5}$) are summarized in **Supplementary Tables 2 and 3**. Local association plots (**Figures 2** and **4**) were constructed with LocusZoom[97]. The top SNP at each locus was also tested in PLINK using logistic regression with dominant and recessive modeling of the minor allele (A1) (**Supplementary Table 5**).

GWAS power calculations

We maximized the power of our GWAS discovery stage by using all MacTel Project cases available at the time. For GWASs of diseases such as MacTel, which are studied for the first time, it is not known what the expected effect size is. Power calculations were therefore performed post hoc with the Genetic Power Calculator[98], assuming a population prevalence of 0.001, *D'*=1 (between risk marker and true causal marker), *alpha*=5×10^{-8} (significance level), a range of genotypic relative risks and allele frequencies, and using the allelic test (**Supplementary Figure 1**). Given 476 cases and 1,733 controls, genotypic relative risks of 2 for genotype "Aa" and 4 for genotype "AA", relative to the baseline genotype "aa", and risk allele frequencies ranging from 0.20 to 0.60, would have > 0.99 power. However, genotypic relative risks of 1.5 and 2.25 for the genotypes "Aa" and "AA" would only a have peak power of 0.53 with a risk allele frequency of ~0.4, thus demonstrating the limit of our GWAS discovery stage.

Replication analysis

The six loci that reached genome-wide significance were selected for replication (**Figure 2**, **Supplementary Figures 25-27**, and **Supplementary Table 2**). Three of these loci were subsequently ruled as false positives after attempts to technically validate them on an independent genotyping platform (TaqMan assays; Applied Biosystems, Foster City, CA) yielded uncorrelated genotypes (i.e. a low concordance rate). We identified two further loci on chromosomes 3 and 7 that showed suggestive evidence for association and warranted replication. This was based on their reported association to the serine and glycine metabolic pathways in other independent studies. Genotyping of the SNPs for replication at these five loci (three genome-wide significant and two suggestively associated) was performed by TaqMan assays in the new MacTel cases and Columbia

University controls (**Supplementary Figure 28**). Assays were purchased from Applied Biosystems as validated, inventoried SNP assays-on-demand or were submitted to Applied Biosystems Assays-by-design pipeline. The choice of these SNPs for replication was limited by the ability to create TaqMan assays. Furthermore, we ensured that at least one actual genotyped, rather than imputed, SNP was selected at each locus. The genotyping technique used was identical to that described previously[90]. Briefly, 5-10ng of DNA were subjected to 50 cycles on an ABI 9700 384-well thermocycler, and plates were read in an Applied Biosytems 7900 HT Sequence Detection System. CCHMC control genotypes were generated with the Illumina Omni5 platform (data was unavailable for rs483180 and rs11238389).

Analysis in the replication cohort was performed with PLINK using a logistic regression framework and no covariates.

Power calculations were performed for the replication analysis. Given a replication cohort of 172 cases and 1134 controls, a population prevalence of 0.001, *alpha*=0.01, *D'*=1 (between risk marker and true causal marker), and the ORs and allele frequencies of the top SNP at each loci, we estimated >97% power to replicate the top 3 genome-wide significant loci with the allelic test. At the two suggestively significant associated loci, 3q21.3 and 7p11.2, the powers were 0.5943 and 0.7033, respectively.

Meta-analysis

Given the differences in ascertainment of the controls used for the discovery and replication stages, and that a covariate was applied to the discovery stage analysis but not the replication stage, meta-analyses between the discovery and replication stages for the top hits were performed with a random-effects model in PLINK as previously described[99] (**Table 2**). Only two of the 10 SNPs tested showed signs of differing effect sizes between the discovery and replication cohorts, while the results for the other 8 SNPs were identical to modeling by fixed-effects.

Heritability estimates

Heritability on the liability scale was estimated using a linear mixed model method, developed by Yang et al. (2010)[100] and adapted for dichotomous traits by Lee et al. (2011)[101], applied to the 1,093,805 directly genotyped SNPs from the discovery stage (i.e. no imputed SNPs). Applying a population prevalence of 0.0045% and 0.1%, yielded h^2 of 0.21 and 0.74 respectively, with 5% of this h^2 explained by our five replicated loci (for both prevalence rates).

Prediction Modeling

Logistic regression prediction modeling was performed in R with the GWAS discovery stage cohort, using the strongest associated SNP that replicated at each of the five loci as predictors and with PC1 and sex as covariates (**Supplementary Table 6 and Online methods**). Initially, a variable selection procedure

investigated the relevance of the interaction terms between gender and the SNPs and any epistatic effects. Once a stable model was selected its predictive power was assessed. The genotypic effect for all markers was modeled as additive to be consistent with the discovery stage. We have shown that the additive model best describes the loci at chromosome 1, 2, 5 and 7, and there is marginal difference for the chromosome 3 loci between the additive and recessive models (**Supplementary Table 5**). To account for possible population stratification, we included the PC1 using orthogonal polynomials of degree three (PC1_1, PC1_2 and PC1_3). Starting with the "full" model, the variable selection step was performed using backward stepwise regression until convergence was achieved as measured by the Bayesian Information Criterion (BIC). Five fold cross validation (CV) was used to perform the variable selection with random subsets of the initial data, initialized multiple times with 100 different seeds. The case and control status was blinded for the performance testing of the prediction model via an automated procedure in R. In total 11 different models were observed from all 500 fitted models. One model predominated, being selected ~66% of times, that included the main effects for all 5 SNPs, PC1 and the interaction term between gender and the SNP on chromosome two (**column four in Supplementary Table 6**). We did not observe any significant epistatic effect between SNPs. Using the R package ROCR[102], a ROC curve was plotted for this model with an "Area Under the ROC Curve" (AUC) of 0.719 (and without PC1 an AUC of 0.707; **Supplementary Figure 13**).

Since PC1 was unavailable in the replication cohort (due to lack of genotype data), this prediction model was applied to our replication cohort for validation after assessing the estimated effects of each SNP in the discovery cohort without PC1. We achieved an AUC of 0.679 thereby validating our results from the discovery cohort (**Supplementary Figure 14**).

Sex specific ORs at 2q34

At 2q34 we derived sex specific ORs for the top most associated SNP (rs715) by performing logistic regression in R with orthogonal polynomials of degree three for PC1 (**Supplementary Table 4**). Considering the risk allele (which is the major allele), we find a borderline significant effect size in males (per-allele OR = 1.40, 95% CI = 1.08-1.81) and a significant association in females (per-allele OR = 2.58, 95% CI = 2.00-3.36).

Metabolomics analysis

Serum was collected from 50 MacTel cases and 50 controls from the wider MacTel Project cohort. Samples were well matched with regards to gender (25 females and 25 males in both cases and controls), age (average of 64 years in the MacTel cases and 63 years in the controls), diabetic status (38 non-diabetics and 12 type-2 diabetics in both cases and controls) and ethnicity (48 Caucasians MacTel cases and 47 Caucasian controls). All individuals fasted overnight and blood was taken before noon. Around 5ml of blood was collected in a clot activating vacutainer tube (Vacutainer Plastic SST II Advance Tube with Gold Hemogard Closure, Becton Dickinson), left at room temperature for 30min and

then centrifuged for 5 min at 1200g. The supernatant was collected, frozen and stored at -80°C. Levels of 1,281 metabolites in the serum were measured by Metabolon (Durham, USA). Briefly, this involved initial protein precipitation with methanol under vigorous shaking for 2 min (Glen Mills GenoGrinder 2000) followed by centrifugation. The resulting extract was analyzed by reverse phase ultrahigh performance liquid chromatography/tandem mass spectrometry (UPLC-MS/MS) with positive ion mode electrospray ionization (ESI). Peaks were quantified using the area-under-the-curve technique. Quality controls steps were applied to the 1,281 $\log_2$ transformed metabolite levels, thereby removing 403 for high missingness (>10% missing data for either MacTel cases or controls), 30 for high correlation (|Pearson's $r| \geq 0.95$) between any pair of metabolites, and 49 with abnormal distributions, thereby leaving a total of 799 metabolites for analysis. Missing metabolite data was then imputed with the minimum value for each metabolite across all samples. Quantile normalization of the data was performed using the R package limma[103]. Metabolite levels were then compared between the MacTel cases and controls by Student's T-tests.

eQTL analysis

The 419 variants showing at least suggestive association ($P < 10^{-5}$) to MacTel at the five loci were tested for association to *cis* gene expression levels in at least 40 tissues from 572 donors. Here, *cis* is defined as being within +/- 1Mb of the transcription start site of each gene. GTEx returns eQTL association results that are beyond specific gene-and-tissue derived significance thresholds.

Online methods References

85. Clemons, T.E. *et al.* Baseline characteristics of participants in the natural history study of macular telangiectasia (MacTel) MacTel Project Report No. 2. *Ophthalmic Epidemiol* **17**, 66-73 (2010).
86. Bergeron-Sawitzke, J. *et al.* Multilocus analysis of age-related macular degeneration. *Eur J Hum Genet* **17**, 1190-9 (2009).
87. Age-Related Eye Disease Study Research, G. A randomized, placebo-controlled, clinical trial of high-dose supplementation with vitamins C and E and beta carotene for age-related cataract and vision loss: AREDS report no. 9. *Arch Ophthalmol* **119**, 1439-52 (2001).
88. Kottyan, L.C. *et al.* Genome-wide association analysis of eosinophilic esophagitis provides insight into the tissue specificity of this allergic disease. *Nat Genet* **46**, 895-900 (2014).
89. Gold, B. *et al.* Variation in factor B (BF) and complement component 2 (C2) genes is associated with age-related macular degeneration. *Nat Genet* **38**, 458-62 (2006).
90. Hageman, G.S. *et al.* A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proc Natl Acad Sci U S A* **102**, 7227-32 (2005).
91. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

92. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
93. Delaneau, O., Marchini, J., Genomes Project, C. & Genomes Project, C. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat Commun* **5**, 3934 (2014).
94. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G.R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* **44**, 955-9 (2012).
95. Genomes Project, C. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
96. Turner, S.D. qqman: an R package for visualizing GWAS results using QQ and manhattan plots. *bioRxiv*, 005165 (2014).
97. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336-7 (2010).
98. Purcell, S., Cherny, S.S. & Sham, P.C. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* **19**, 149-50 (2003).
99. Borenstein, M., Hedges, L.V., Higgins, J.P. & Rothstein, H.R. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods* **1**, 97-111 (2010).
100. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* **42**, 565-9 (2010).
101. Lee, S.H., Wray, N.R., Goddard, M.E. & Visscher, P.M. Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet* **88**, 294-305 (2011).
102. Sing, T., Sander, O., Beerenwinkel, N. & Lengauer, T. ROCR: visualizing classifier performance in R. *Bioinformatics* **21**, 3940-1 (2005).
103. Ritchie, M.E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).

Supplementary Tables and Figures for manuscript:

Genome-wide analyses identify common variants associated with macular telangiectasia type 2

Supplementary Table 1 Members of the MacTel Project consortium.

Site Name	Principal Investigator	Coordinator & other roles
Coordinating Center: The Emmes Corporation 401 N. Washington Street, Suite 700, Rockville, Maryland 20850	Traci Clemons, PhD Email:	Sarah Duwel, RN MA Stephanie Henson, MA
CENTRE FOR EYE RESEARCH Australia Macula Research Unit Level 1, 32 Gisborne St. East Melbourne VIC 3002	ROBYN GUYMER, PhD Email: rhg@unimelb.edu.au	MELINDA CAIN Email: melindac@unimelb.edu.au EMILY CARUSO Email: emily.caruso@unimelb.edu.au
Moorfields Reading Centre Moorfields Eye Hospital NHS 162 City Road, London EC1V 2PD	Primary Contact: TUNDE PETO, MD, MHealthEd, PhD, FHCO Email: Tunde.Peto@moorfields.nhs.uk	Additional Contact: ROYA NASSAJPOUR ESFAHANI Roya.NassajpourEsfahani@moorfields.nhs.uk
HÔPITAL LARIBOISIÈRE Service d'Ophtalmologie 2, rue Ambroise Paré, Cedex 10 Paris 75475	ALAIN GAUDRIC, MD Email: alain.gaudric@Irb.aphp.fr	ABIR ZUREIK Email: abir.zureik@irb.aphp.fr
JULES STEIN EYE INSTITUTE, UCLA 100 Stein Plaza, Room 2-565 Los Angeles CA, 90095	JEAN-PIERRE HUBSCHMAN, MD Email: hubschman@jsei.ucla.edu	NINA ZELCER Email: Zelcer@jsei.ucla.edu MARGARET HAVUNJIAN Email: Havunjian@jsei.ucla.edu ROSALEEN OSTRICK Email: ostrick@jsei.ucla.edu
LIONS EYE INSTITUTE 2 Verdun Street Nedlands WA 6009	IAN CONSTABLE, MD Email: Ian.constable@uwa.edu.au	TONI BUSBY Email: toni.busby@lei.org.au
MANHATTAN EYE, EAR & THROAT HOSPITAL LuEsther T. Mertz Retinal Research Ctr Manhattan Eye, Ear & Throat Hospital 210 East 64th Street, 8th Floor New York, NY, 10065	LAWRENCE A. YANNUZZI, MD Email: layannuzzi@gmail.com	MARIA SCOLARO Email: mscolaro@retinal-research.org
MOORFIELDS EYE HOSPITAL Clinical Trials Unit 162 City Road London EC1V 2PD	CATHY EGAN, MD Email: catherine.egan@moorfields.nhs.uk	HAYLEY BOSTON Email:Hayley.Boston@moorfields.nhs.uk TANIYA CHOWDHURY, DATA ENTRY taniya.Chowdhury@moorfields.nhs.uk TOYIN ADENUGA, DATA ENTRY Oluwatoyin.Adenuga@moorfields.nhs.uk
RETINA ASSOCIATES OF CLEVELAND, INC. 3401 Enterprise Parkway, Suite 300 Beachwood, OH 44122	LAWRENCE SINGERMAN, MD Email: lsingerman@retina-assoc.com	DIANE WEISS, RN Email: dew@retina-assoc.com
SAVE SIGHT INSTITUTE	MARK GILLIES, MD, PhD GPO Box 4337 (use 2001 zip) 8 MacQuarie Street Sydney NSW 2000 Email: mark.gillies@sydney.edu.au	ROXY MEDINA Email: roxy.medina@sydney.edu.au LUKE SEESINK Email: luke.seesink@sydney.edu.au
SCRIPPS RESEARCH INSTITUTE Scripps Clinical Research 10666 N. Torrey Pines Rd La Jolla, CA 92037	MARTIN FRIEDLANDER, MD, PhD Email: friedlan@scripps.edu	JENNIFER TROMBLEY, RN Email: Trombley.jennifer@scrippshealth.org

Site Name	Principal Investigator	Coordinator & other roles
ST. FRANZISKUS HOSPITAL St. Franziskus Hospital Hohenzollernring 74 Muenster D-48145	DANIEL PAULEIKHOFF, PROF. DR Email: dapauleikhoff@muenster.de	MEIKE ZEIMER, MD Email: Meikezei@web.de
THE GOLDSCHLEGER EYE INSTITUTE The Chaim Sheba Medical Center Tel Hashomer 52621	JOSEPH MOISSEIEV, MD Email: Joseph.moisseiev@sheba.health.gov.il	RINAT RASHTY Email: Rinat.rashty@sheba.health.gov.il
THE NEW YORK EYE AND EAR INFIRMARY Department of Ophthalmology 310 East 14th Street Suite 319 South Building New York NY, 10003	RICHARD ROSEN, MD Email: rrosen@nyee.edu	MELIZA UNSON Email: munson@nyee.edu
UNIVERSITY OF BONN Dept. of Ophthalmology Ernst-Abbe-Str. 2 Bonn, 53127	FRANK HOLZ, MD Email: Frank.Holz@ukb.uni-bonn.de	SIMONE MÜLLER, ASSISTENZÄRZTIN Email: Simone.Mueller@ukb.uni-bonn.de KERSTIN BARTSCH Email: Kerstin.Bartsch@ukb.uni-bonn.de
UNIVERSITY OF MICHIGAN, KELLOGG EYE CENTER 1000 Wall Street Ann Arbor MI, 48105	GRANT COMER, MD Email: gcomer@umich.edu	PAM CAMPBELL, COT, CCRP Email: pamtitus@med.umich.edu
UNIVERSITY OF WISCONSIN 2880 University Ave Madison, WI 53705	BARBARA BLODI, MD Email: bablodi@wisc.edu	JENNIE PERRY-RAYMOND Email: jrperry@ophth.wisc.edu KRISTINE DIETZMAN Email:kadietzman@ophth.wisc.edu
SCHEIE EYE INSTITUTE 51 North 39th Street Philadelphia PA, 19104	ALEXANDER BRUCKER, MD Email: ajbrucke@mail.med.upenn.edu	SHERI DROSSNER Email: Sheri.Grand@uphs.upenn.edu
UNIVERSITY OF BERN Universitätsklinik für Augenheilkunde Inselspital Bern, CH 3010	SEBASTIAN WOLF, MD, PhD Email: Sebastian.wolf@insel.ch	CORINNE STOCKLI Email: Corinne.Stoeckli@insel.ch
UNIVERSITY OF UTAH HEALTH CARE Moran Eye Center University of Utah Health Care 65 Mario Capecchi Drive Salt Lake City, UT 84132	PAUL BERNSTEIN, MD, PhD Email: paul.bernstein@hsc.utah.edu	KIMBERLEY WEGNER, BS, CRC Email: Kimberley.Wegner@hsc.utah.edu ELIZABETH NUTTALL – Regulatory Documents Email: Elizabeth.nuttall@hsc.utah.edu
BASCOM PALMER (MIAMI)	PHILIP ROSENFELD, MD, PhD 900 NW 17th Street Miami FL, 33136 Email: prosenfeld@med.miami.edu	CRISTINA LAGE-RODRIGUEZ, MS, CCRC 1638 NW 10th Avenue, Suite 505B Miami, FL 33136 Email: clage@med.miami.edu
MASSACHUSETTS EYE AND EAR INFIRMARY Retina Research, 12th Floor 243 Charles Street Boston MA, 02114	JOAN MILLER, MD Email: joan_miller@meei.harvard.edu	PATRICIA HOULIHAN Email: patricia_houlihan@meei.harvard.edu
Clinique Opthalmolgie de Creteil Eye University Clinic of Paris XII – Créteil 40 avenue de Verdun Creteil, 94010 France	Gisele Soubrane, Gisele.Soubrane@chicreteil.fr	Karim Atmani karim.atmani@chicreteil.fr
Wilmer Eye Institute, Johns Hopkins University 600 North Wolfe St. Maumanee #749 Baltimore, MD 21287	Diana Do ddo@jhmi.edu	Lisa Greer lgreer3@jhmi.edu
The Retina Group of Washington 5454 Wisconsin Avenue Suite 1540 Chevy Chase, MD 20815	Robert Murphy rpmurphy@comcast.net	Debbie Oliver doliver@retinagroup.org
Department of Ophthalmology, Columbia University, New York, NY 10032, USA		Recruitment of control cohort: John C. Merriam

Supplementary Table 2 All loci with genome-wide significant ($P < 5 \times 10^{-8}$) and otherwise suggestive ($P < 10^{-5}$) associations for MacTel in our discovery stage

Chr	Start	Stop	SNPs / loci*	Lowest p-value	Most significant SNP at loci	Imputed	SNP position
5^	87520793	87919700	176	7.74E-17	rs73171800	Y	chr5:87786004
1#	10815702	10815702	1	1.12E-15	rs205485	.	chr1:10815702
2^	211543055	211859753	143	9.97E-14	kgp1582223	.	chr2:211543055
1^	120208297	120304006	71	2.60E-12	rs477992	Y	chr1:120257576
3#	27680063	28056877	58	1.44E-11	kgp1625553	.	chr3:28056877
7#	9994244	9994244	1	7.43E-09	kgp12086038	.	chr7:9994244
9	116766112	116766112	1	1.43E-07	rs942288	.	chr9:116766112
19	8326904	8788177	5	1.69E-07	rs36259	.	chr19:8326904
9	15298666	15304782	4	7.16E-07	rs1215112	Y	chr9:15303583
8	95047577	95047577	1	9.55E-07	kgp12355311	.	chr8:95047577
6	98963740	98973805	6	1.54E-06	rs72928454	Y	chr6:98973805
10	108375801	108422966	29	1.91E-06	rs11193003	Y	chr10:108418384
2	104197	178487	43	1.93E-06	rs9677393	Y	chr2:141378
15	70668824	70669660	2	2.24E-06	kgp11183162	.	chr15:70669660
7^	56033558	56099352	27	2.25E-06	rs4948102	Y	chr7:56097265
5	11068212	11089528	3	2.37E-06	rs199973235	Y	chr5:11069785
12	131525954	131525954	1	2.56E-06	kgp7438645	.	chr12:131525954
14	22473194	22478961	8	2.59E-06	kgp4994464	.	chr14:22473194
20	31815973	31844770	33	3.00E-06	rs1321417	Y	chr20:31815973
4	174853026	174853026	1	3.01E-06	kgp1280587	.	chr4:174853026
12	94235885	94235902	3	3.37E-06	kgp22814624	.	chr12:94235902
8	122515531	122531475	5	3.87E-06	rs1384188	Y	chr8:122519503
3^	126044895	126049305	2	5.38E-06	rs9820286	Y	chr3:126044895
11	134900054	134900054	1	6.02E-06	rs10585506	Y	chr11:134900054
1	64768149	64803741	14	6.42E-06	rs34557784	Y	chr1:64789648
10	65010626	65363166	2	7.02E-06	rs9415676	Y	chr10:65010626
3	189744368	189744368	1	7.22E-06	rs1829828	Y	chr3:189744368
7	49860530	49863017	2	7.38E-06	rs6957449	.	chr7:49860530
22	33346615	33346615	1	8.86E-06	kgp8323698	.	chr22:33346615
21	39917180	39917180	1	8.89E-06	rs2836502	Y	chr21:39917180
2	97717985	97717985	1	9.66E-06	rs566138585	Y	chr2:97717985

*SNPs / loci: This is the number of significant ($P < 10^{-5}$) SNPs per loci; there are a total of 647 such SNPs (**Supplementary Table 3**), and any within 500kb of each other are grouped into the same locus. ^Loci that replicated. #Three genome-wide significant loci judged to be false positives after attempting to validate the SNP genotypes with TaqMan assays. Only a single SNP was genome-wide significant at the 3p24.1 locus, and this was not in linkage disequilibrium with the other 57 SNPs. Regional plots of association for these loci are shown in Supplementary Figures 20, 21 and 22.

Supplementary Table 3 All 647 SNPs that were either genome-wide significant ($P < 5 \times 10^{-8}$) or otherwise suggestive ($P < 10^{-5}$) associations for MacTel in our discovery stage.

See separate Excel file.

Supplementary Table 4 Sex specific ORs for the top SNP at the 2q34 locus.

Covariates	Risk allele (freq.*)	Males		Females	
		OR (95% CI)	*P* value	OR (95% CI)	*P* value
Intercept	-	0.16 (0.10-0.24)	3.325E-18	0.06 (0.04-0.95)	6.935E-34
PC1_1	-	0.008 (0.00-0.94)	4.264E-02	0.00 (0.00-0.001)	1.856E-08
PC1_2	-	0.05 (0.00-5.58)	2.052E-01	0.003 (0.00-0.29)	1.338E-02
PC1_3	-	0.002 (0.00-0.18)	9.086E-03	0.01 (0.00-1.23)	6.437E-02
Chr2: rs715	T (0.709)	1.40 (1.08-1.81)	1.046E-02	2.58 (2-3.36)	5.290E-13

*Determined from the 1000Genomes project European population (http://1000genomes.org/).