Identification of genetic factors influencing metabolic dysregulation and retinal support for MacTel, a retinal disorder

Bonelli R[1,2]^, Jackson VE[1,2]^, Prasad A[1,2], Munro JE[1,2], Heeren TFC[3,4], Pontikos N[3,4], Scheppke L[5], Friedlander M[5,6], MacTel Consortium[7], Egan CA[3], Allikmets[8,9] R, Ansell BRE[1,2], Bahlo M[1,2]*

Author information

1 Department of Medical Biology, The University of Melbourne, 3052, Parkville, Victoria, Australia.
2 Population Health and Immunity Division, Walter and Eliza Hall Institute of Medical Research, 3052, Parkville, Victoria, Australia.
3 Moorfields Eye Hospital NHS Foundation Trust, EC1V 2PD, London, United Kingdom.
4 University College London Institute of Ophthalmology, London, UK.
5 The Lowy Medical Research Institute, La Jolla, California, USA.
6 Department of Molecular Medicine, The Scripps Research Institute, La Jolla, California, USA
7 A list of members and affiliations is provided in Table S13.
8 Department of Ophthalmology, Columbia University, New York, New York, USA
9 Department of Pathology and Cell Biology, Columbia University, New York, New York, USA

^Joint first authors
* Correspondence: bahlo@wehi.edu.au, +61 03 9345 2630

Abstract

Macular Telangiectasia Type 2 (MacTel) is a rare degenerative retinal disease with complex genetic architecture. We performed a genome-wide association study on 1,067 MacTel patients and 3,799 controls, which identified eight novel genome-wide significant loci ($p<5E-8$), and confirmed all three previously reported loci. Using MAGMA, eQTL and transcriptome-wide association analysis, we prioritised 48 genes implicated in serine-glycine biosynthesis, metabolite transport, and retinal vasculature and thickness. Mendelian randomization indicated a likely causative role of serine (FDR=3.9E-47) and glycine depletion (FDR=0.006) as well as alanine abundance (FDR=0.009). Polygenic risk scoring achieved an accuracy of 0.74 and was associated in UKBiobank with retinal damage ($p=0.009$). This represents the largest genetic study on MacTel to date, and further highlights genetically-induced systemic and tissue-specific metabolic dysregulation in MacTel patients, which impinges on retinal health.

Introduction

Macular telangiectasia type II ('MacTel') is a relatively rare (~1/1000) degenerative disease of the retina[1], with a mean age at diagnosis of 60 years, and known comorbidity with type II diabetes (T2D)[2]. The clinical features of MacTel include macular photoreceptor atrophy, retinal greying, vascular leakage and in some cases, appearance of blunted/angular blood vessels in the normally avascular fovea. MacTel patients experience blurred or lack of peripheral vision, which progresses to complete loss of central vision.

To establish the genetic architecture of MacTel, we previously undertook the first genome-wide association study (GWAS) of 476 patients and 1,733 controls[3], and found three genome-wide significant loci (p<5E-8). The first locus, proximal to transmembrane protein 161B (*TMEM161B*; 5q14.3), was implicated in modifying retinal vascular calibre, consistent with the abnormal retinal vascular morphology observed in MacTel[1]. The other two loci were within the phosphoglycerate dehydrogenase (*PHGDH*; locus 1p12), and carbamoylphosphate synthase 1 (*CPS1*; 2q34) genes. These loci, together with two loci attaining suggestive significance (p<5E-6) within phosphoserine phosphatase (*PSPH*; 7p11.2) and aldehyde dehydrogenase 1L1 (*ALDH1L1*; 3q21.3), suggested systemic perturbation of glycine and serine metabolism in MacTel patients, which was supported by significantly depleted concentrations of both metabolites in patient serum.

To better characterize the systemic and retina-specific phenotypes influenced by MacTel GWAS loci, we recently leveraged GWAS summary statistics to predict genetically-derived retinal vascular calibre[4,5], T2D risk[6], as well as serum concentrations for 174 metabolites[7], in MacTel patients compared to controls. Mendelian randomization analysis revealed that genetically-influenced serine and glycine deficiency, and excessive alanine in patient sera, are likely causal for MacTel[8]. Further, a GWAS analysis performed by conditioning on genetically-predicted T2D risk, glycine and serine concentrations, identified that the genetic contributions from the vasculature-associated locus 5q14.3 was independent of these metabolic traits.

Furthermore, two novel loci were detected at 3p24.1 and 19p13.2. The 3p24.1 locus lies between the retinal bicarbonate symporter gene *SLC4A7* and a neuronal transcriptional factor *EOMES,* while locus 19p13.2 is upstream of *CERS4*, which synthesizes the ceramide precursors of sphingolipids.

The detrimental effect of serine insufficiency on lipid metabolism, and sphingolipid biosynthesis in particular, in MacTel, has since been demonstrated in patients with a rare neuropathy: hereditary sensory neuropathy Type 1 (HSN1; MIM162400). HSN1 is caused by mutations in the *SPTLC1* and *SPTLC2* genes, encoding for the subunits of serine-palmitoyltransferase (SPTLC), which acts upstream of the sphingolipid pathway. These mutations cause excessive deoxy-sphingolipid synthesis. Intriguingly, HSN1 patients develop peripheral neuropathy, and clinical MacTel in many, but not all, cases[9].

Heritability estimates from the original MacTel GWAS suggested that more genetic loci remained to be discovered, which could be achieved with increased sample sizes. In this study we performed a GWAS on MacTel with more than double the cases of the original 2017 study. We aimed to: (1) confirm the originally reported genetic loci, and assess the strength of association with previously suggestive-significant SNPs, (2) identify new disease loci, (3) validate the causal effect of metabolite disruption on MacTel disease and assess genetic correlations with other traits/diseases, (4) prioritise likely causal genes by investigating functional consequences of MacTel SNPs on transcription across relevant tissues and cell types, and (5) assess the predictive value and potential clinical utility of polygenic risk scores. A diagram of the study is presented in **Figure 1.**

Results

After quality control and imputation, genotype data for 7,289,516 SNPs across 1,067 MacTel patients and 3,799 controls were included in genome-wide association testing. The sample was predominantly of European ancestry (3,745 controls, 98%; 931 cases, 87%), with a higher proportion of females (78% in controls; 60% in cases).

We performed two separate GWAS: one for the entire cohort, and one restricted to individuals of European ancestry. Eleven independent Genome-wide (GW) significant disease associations in ten regions were identified in the full-cohort analysis **(Figure 2; Table 1)**, of which seven results in six regions were preserved in the European ancestry GWAS (**Table S1**). There was high concordance in effect estimates for both analyses (**Figure S2**), and minimal inflation of test statistics due to population stratification (λ=1.011 in the full cohort analysis; λ=1.013 in Europeans only, see **Figure S3** for QQ plots and **Supplementary Note** for LD Score Regression (LDSC) results). Given the overall agreement in results from these analyses, we hereafter focus on results from the more powerful full-cohort analysis. After restricting to unrelated individuals (787 cases and 2853 controls) and assuming a population prevalence for MacTel of 0.45%, we estimated the narrow-sense heritability (h^2) for MacTel to be 0.647 (se=0.048). Bayesian fine-mapping[10] was undertaken to identify 95% credible sets of causal SNPs at each GW-significant locus. The number of SNPs in each credible set ranged from 1 to 57. The credible SNP sets are listed in **TableS1,** and LocusZoom plots for all loci are provided in **Figure S4**.

Associations at previously-reported MacTel risk loci and a new rare PHGDH variant

The three GW-significant MacTel loci identified in our original MacTel GWAS[3] again reached GW-significance.The most statistically-significant SNPs at each locus were rs146953046 in *PHGDH* (1p12, P=7.90E-22); rs1047891 in *CPS1* (2q34, P=8.6E-21) and rs17421627 near *TMEM161B* (5q14.3, P=4.7E-17). Whereas the SNPs at 2q34 and 5q14.3 were in linkage disequilibrium (LD) with the previously-identified tagging SNPs (r^2=0.91 with rs715, and r^2=0.67 with rs73171800, respectively), the most significant SNP in

the 1p12 region (rs146953046, MAF in cases / controls: 0.059 / 0.016) was in low LD with, and independent from the previously identified SNP rs477992 (R2=9E-4, MAF=0.342). Conditional analysis at this locus revealed a second signal, whose top SNP rs662602 (P=2.3E-20, MAF in cases / controls: 0.424 / 0.315) was in strong LD with the previously identified SNP rs477992 (R2=0.99). Both signals in *PHGDH* were present in both the European-only and full-cohort analyses.

Novel genome-wide significant loci

Three loci previously implicated in MacTel with suggestive association (P<1E5), reached GW-significance in the present analysis. Specifically, the most significant SNPs were located in the genes *PSPH* (rs6955423 in locus 7p11.2; P=5.9E-09), *TTC39B* (rs677622, 9p22.3; P=3.2E-09), and *REEP3* (rs10995566, 10q21.3; P=4.8E-08). A further two loci tagged by rs9820465 between the genes *SLC4A7* and *EOMES* (3p24.1, P=5.6E-09), and rs139412173 near *CERS4* (19p13.2, P=2.9E-08), were previously implicated in a GWAS of MacTel conditioned on genetically predicted glycine and serine[8]. We identified two novel GW-significant loci tagged by rs2160387 in locus 2p14 (P=5.9E-09), located within the alanine/glycine/serine transporter gene *SLC1A4*, and rs17279437 in locus 3p21.31 (P=6.2E-10) within the glycine transporter gene *SLC6A20*. In addition to the GW-significant association in locus 19p13.2 upstream of *CERS4*, the analysis restricted to European subjects identified a second independent suggestively significant SNP (rs36259) in exon 17 of *CERS4* (P=1.4E-06, MAF in cases/controls =0.294/0.246); however this did not meet the suggestive significance threshold in the full-cohort analysis(P=1.4E-05, MAF in cases/controls =0.286/0.246). A further five novel loci that attained suggestive significance are detailed in **Table 2**.

Gene Prioritisation

Through positional mapping, a total of 52 genes were identified as proximal to or within the 16 MacTel-associated loci described above. To prioritize those genes most likely to be involved in the disease mechanism and identify additional novel genes, we performed gene-wise SNP enrichment analysis with MAGMA[11]; and inferred the transcriptional effects of each locus in multiple tissues using both eQTL analysis and transcriptome-wide association analysis (TWAS)[12].

MAGMA analysis

73 genes were identified by MAGMA to be enriched with MacTel-associated SNPs (FDR<0.05) (**Table S2**), of these, 29 were located within disease-associated loci, and 44 were more distal. The top ten ranked genes are provided in **Table 3**. Of particular interest were genes related to lipid oxidation (*PPARA*, *SUMF2, HMGCS3*), low-density lipoprotein metabolism (*LDLRAD3*) and cholesterol excretion (*SLC51A*), and a second Solute Carrier Family (SLC) gene, monocarboxylic acid transporter *SLC16A8*, previously implicated in age-related macular degeneration. The top ranked gene was *PHGDH*.

eQTL analysis

MacTel SNPs at each locus were investigated for associations with transcription of proximal genes (eQTL) in several tissues via the FUMA portal[13]. Specifically, eQTL for brain (16 regions), arteries (3 sites), tibial nerve, liver, EBV-transformed lymphocytes, and four whole-blood datasets were chosen via the GTEx[13], eQTLGen (eqtlgen.org)[14] and BIOS consortia[15] (**Table S3)**. Recently published retinal eQTLs were manually curated[16] prior to inclusion. In total we identified 99 genes whose expression was significantly altered by MacTel-associated SNP(s), in one or more tissues (**Figure 3; Table S4**). A heatmap of eQTL effect can be found in supplementary material **(Figure S5)**. Positive eQTL effects on the matrix protein gene *MXRA7* (17q25.1) were the most significant results, followed by zinc finger domain protein *ZNF713* (7p11.2) and the mitochondrial respiration-related gene *GBAS* (7p11.2). Conversely, transcriptional suppression of the sphingolipid-related gene *SUMF2* (7p11.2) and differentiation factor *ATRAID (2p23.3)* was inferred in several tissues. Interestingly strong suppression of *PHGDH* (1p12) transcription was inferred in the tibial nerve, brain and vasculature, but not in the retina. The direction of the eQTL transcriptional effect also deviated markedly in the retina compared to other tissues in the genes *NRBF2* (10q21.3) and *TMEM161B* (5q14.3) (negative in retina, positive in brain), and *DCDC1* (11p14.1) (*vice versa*).

At the four genome-wide significant loci which overlapped with significant retina eQTLs, we additionally tested for co-localisation of the GWAS, and eQTL signals, However, no evidence for a shared underlying causal variant was found (posterior probability < 75%) (**Table S5**) for any of the four loci.

Transcriptome-wide association results

TWAS analysis identified seven genes in which SNP-regulated expression significantly affected disease risk across six different tissues (FDR<0.05) (**Table 4**)[12]. A further ten genes achieved suggestive significance (FDR<0.1) (**Table S6**). Increased expression of *PHGDH* and *MXRA7*, and down-regulation of nuclear receptor binding factor 2 (*NRBF2*), was inferred to take place in MacTel patient retinae. In brain tissue we predicted increased expression of *DNAJC24 and SLC1A1*, and suppression of the intraflagellar transport protein-coding gene *IFT172*. Whereas no differentially expressed genes were predicted in the vascular system or peripheral nerves, MacTel risk loci were associated with increased expression of *TTC39B* in the liver. Notably, all functionally-associated genes identified here fall within a locus with GW- or suggestive association with MacTel, with the exception of *SLC1A1*, at 9p24.2.

Gene expression patterns across ocular tissues and retinal cell-types

We identified 153 genes associated with MacTel via at least one of these analyses, either via genomic location, SNP enrichment, eQTL or TWAS (**Table S7**). To investigate MacTel-associated gene expression in ocular tissues and cell types, we prioritized 49 genes supported by at least two sources of evidence, and added the genes *SLC4A7*, *FBN3* and *CERS4,* as these were genes in proximity to two GW-significant loci located in gene deserts but which were not prioritised by any post-GWAS analyses. Analysis of gene expression across temporal, macular and nasal regions of both the neural retina and RPE/choroid from four adult donors[17] revealed 16 genes whose expression was more pronounced in the neural retina than the RPE/choroid, including *TMEM161B, EOMES, MXRA7* and *SLC4A7* (**Table S8**) and **Figure S7**. Among 17 genes showing an opposing trend were *REEP3*, *SLC6A20*, *PHGDH*, *HMGCS2* and *CPS1*. Tests for area specificity within tissues where genes were most expressed, revealed that in the retina, *EOMES* was more highly expressed in the macula than either the nasal or temporal sides.

Using five publicly available scRNA-seq datasets from three independent studies[18–20], we tested for retinal cell-specificity of the prioritised genes. Expression of the genes *PHGDH* and *PSPH* was significantly higher

in Müller glia, whereas *TMEM161B*, *SLC4A7*, *SLC6A20* and *JMJD1C* were more highly expressed in rod photoreceptors. *CPS1* was enriched in amacrine cells, and *EOMES* was most abundant in retinal ganglion cells (**Table S9**).

Genetic correlation between MacTel and other traits

LD score regression between MacTel and 650 selected phenotypes revealed 99 nominally significant associations, of which use of medications for 'pain relief, constipation and heartburn' was the most significant (P<10E-5; **Table S10**). Related to this, pain in the neck, face, limbs and joints, and neuropathy of the upper limbs were also correlated with genetic risk factors for MacTel. In addition, type II diabetes-related phenotypes including insulin resistance, anti-cholesterol medication, increased body fat, and diabetic eye disease, were positively correlated with MacTel. Whereas triglyceride and very-low-density lipoprotein (VLDL) serum concentrations were positively associated, high-density lipoprotein (HDL) concentrations were negatively correlated with MacTel genetic risk factors. Of note, retinal detachment and breast cancer were also negatively correlated with this disease (**Table S10**).

Mendelian Randomization

Using Mendelian randomization analysis we recently demonstrated that low serine and glycine levels, as well as high alanine levels, were likely to have a causal effect on MacTel etiology[8]. Applying Mendelian Randomization analysis to these new GWAS results we found that genetically predicted serum serine concentrations was again the strongest and most significantly associated trait among 141 predicted metabolites (**Table S11**). Indeed, MacTel risk was found to nearly halve for each standard deviation unit increase in serine (OR=0.55, FDR=3.9E-47). We identified an additional 24 metabolites to be significantly associated with MacTel at FDR<0.05. However, conditional analysis allowed the discrimination of independent signals, and only glycine (OR=0.81, FDR=0.006) and alanine (OR=1.16, FDR=0.009) remained significant after inclusion of serine in the regression model, highlighting the high correlation between the genetic determinants of metabolites. Consistent with the previously published results, we found the T2D (Polygenic Risk Score) PRS to be robustly associated with disease risk (OR=1.21, P=7.6E-07), and minimal confounding from pleiotropic effects (**Supplementary Note**). We additionally found a positive

and significant association between retinal arteriolar calibre PRS (b=0.18, p=1.4E-6), retinal vascular calibre PRS (b=0.17, p=1.1E-5) and macular thickness PRS (b=0.08, p=0.042) with MacTel. However, in all three traits these associations were entirely driven by the shared signal in locus 5q14.3 (**Figure S7**). In fact when the 5q14.3 tagging SNP rs17421627 was included in the model as covariate the first two of these associations became non-significant while the third showed negative association **(Supplementary Results)**. This negative association was once again entirely driven by the shared signal in locus 3p21.31 between the two traits **(Supplementary Results)**.

MacTel prediction and Polygenic Risk Scores

We tested the predictive power of two separate polygenic risk scores (PRS) for MacTel, constructed using increasing numbers of SNPs (**Figure 4A;** Methods). A conservative PRS that safeguards against overfitting, comprising only 11 SNPs tagging genome-wide significant loci, achieved an Area Under the Receiver Operating Curve (AUC) of 0.74 (C.I, 0.72-0.76). When trained on a set of European ancestry individuals, this classifier performed similarly in predicting the disease state of both European and non-European subjects, albeit with a modest decrease in accuracy for the latter group (AUC = 0.78 vs AUC=0.76) (**Figure 4B)**. The distribution of scaled MacTel PRS was markedly higher for cases than controls (**Figure 4C)**. Furthermore, the odds ratio for disease risk between different PRS deciles highlighted subjects in the 10th decile of the distribution, who had an OR almost 10 times higher than those in the 5th decile. In contrast the odds ratio for subjects in the 1st PRS decile was 0.25 times that of the 5th decile (**Figure 4D)**.

Investigation of MacTel PRS in UKBiobank data

Given the good prediction power of the MacTel PRS, we applied the 11 SNPs PRS to subjects from the UKBiobank database. The MacTel PRS was associated with a significantly higher probability of self-reported diagnosis with “Other serious eye conditions” (b=0.048, P=2.1E-7) and “Macular degeneration” (b=0.029, P=0.024). Interestingly we found a negative association with self-reported “Diabetes-related eye disorders” (b=-0.034, P=0.018). The MacTel PRS was positively, but non-significantly, associated with reporting any eye disorder (b=0.007, P=0.153) which included those mentioned above, as well as “Cataract”, “Glaucoma” and “Injury or trauma resulting in loss of vision”.

We compared the results of manual, blinded, MacTel grading performed by expert clinicians from retinal scans on 200 randomised patients (100 in the top PRS decile and 100 in the bottom PRS decile) who self-reported absence of any eye conditions. Of these 14 retinal scans were discarded due to low scan quality which could not be scored. We did not identify any definite MacTel patients among these subjects. However, using logistic regression modelling we found the MacTel PRS decile to be positively associated with retinal phenotypes typical of MacTel disease (b=0.54 p=0.003) as well as retinal phenotypes not directly related to MacTel (b=0.17, p=0.036) compared to healthy eyes.

Discussion

MacTel was first clinically defined in 1977, however the associated genetic risk factors would remain elusive for decades. In 2017 the first genetic association signals for MacTel were identified, highlighting glycine and serine metabolism as a dysregulated pathway in MacTel patients. These preliminary genetic clues were discovered with a small scale GWAS of 476 cases and 1733 controls. Subsequent work highlighted

that further GWAS, with increased sample size, would likely increase the number of associated loci[8]. Here we describe results from a second GWAS for MacTel disease with a cohort of 1,067 MacTel patients and 3,799 controls for this relatively rare eye disease (approximately 250 times less common than AMD).

The current study reuses the data from our initial GWAS[3] and is thus not independent from our original study. However, we were able to take advantage of improved GWAS methodologies, which allowed the use of related individuals which were excluded from the previous study, as well as recent post-GWAS techniques allowing extensive follow-up investigations. We made use of publicly available data sources such as GTEx, metabolomics data, retinal GWAS studies and the UKBiobank to extend and contextualize our findings, providing further clues as to biological mechanisms that are perturbed in MacTel. These results confirm previous findings, revealing loci and candidate genes that we expected to observe with an increased sample size based on the current understanding of MacTel, and additionally identified several unanticipated novel associations. We found eleven GW-significant SNPs at ten loci, which collectively explained 0.647% of MacTel heritability, with odds ratios for significant SNPs ranging from 1.39 to 5.47. A PRS comprising all GW-significant SNPs achieved an AUC of 0.74 (C.I. 0.720-0.755), with patients in the top PRS decile harbouring a ten-fold greater risk of MacTel compared to the fifth (middle) decile. As MacTel manifests relatively late in life and is difficult to diagnose, the genetic risk predictor provided here may be useful for both clinical diagnosis and interpretation of clinical trial results[21]. Genetic prediction of MacTel in the independent UKBiobank data demonstrated that the MacTel PRS could be of clinical use to detect early retinal damage. The model could likely be improved through integrating direct metabolic measurements, which are likely to more comprehensively describe the metabolic dysregulation known to be present in MacTel patients but which is tempered by the greater measurement error in metabolites in comparison to genotypes.

Consistent with our previous findings, some of the strongest signals from the present work highlight the involvement of glycine and serine in MacTel disease. The most significant signal observed in our study is located in the 1p12 locus, within the *PHGDH* gene. The enzyme phosphoglycerate dehydrogenase (PHGDH) catalyzes the conversion of 3-phosphoglycerate into 3-phosphohydroxypyruvate, which is the first step in serine biosynthesis. This GWAS signal is driven by a rare risk haplotype with a frequency of 1.6% in controls and 5.9 % in cases (OR = 5.47), which is independent of the previously identified common

risk haplotype at *PHGDH* [3], replicated here with an odds ratio of 0.585. We found overwhelming support for the role of *PHGDH* in disease etiology through several post-GWAS analysis results (MAGMA, eQTL investigations and TWAS). Whilst MacTel risk alleles were associated with decreased *PHGDH* expression in artery, brain and nerve tissues, TWAS results for retina, and blood eQTL results from the eQTLGen cohort (31,684 subjects), suggest that upregulation of PHGDH may be linked to limited serine availability, and is thus a key risk factor for MacTel[14] — a finding that warrants further biochemical investigation.

The 2q34 locus, encompassing the 3' UTR of carbamoyl-phosphate synthase 1 (*CPS1),* is the strongest reported modulator of serum glycine abundance[7] and was the second-most significant MacTel-associated locus. Abundant glycine, which is converted to serine by serine hydroxymethyltransferase, may protect against MacTel[8,9]. The tagging SNP at 2q34 is a coding SNP in *CPS1* (rs1046891), with the minor allele being protective for MacTel. The resulting T1406N substitution reduces the binding affinity between CPS1 and its coactivator N-acetylglutamine (NAG). NAG promotes glycine flux through the urea cycle[22,23], which may contribute to MacTel by drawing down subjects' capacity to buffer serine. Indeed, urea cycle dysregulation is observed at the metabolite level in MacTel patient serum [24]. An allosteric disease mechanism, affecting CPS1 enzyme activity, would be consistent with the lack of eQTL effects for *CPS1* (suggesting stable enzyme abundance). However, the nominally significant TWAS effect linking greater *CPS1* transcription in the retina with MacTel, suggests potentially compounding effects of greater expression and activation of *CPS1* on this disease.

The genome-wide significant signal at 7p11.2 spans the phosphoserine phosphatase (*PSPH)* gene. *PSPH*, along with *PHGDH* and *PSAT1* are essential genes for serine biosynthesis. Whilst we did not identify any GW-significant signals in *PSAT1,* trans-eQTL analysis in blood revealed a possible effect of polymorphisms in *PHGDH* and *CPS1* on *PSAT1* expression, indicating co-regulated expression of genes across the serine biosynthesis pathway. The causal role of serine and glycine depletion on MacTel is also highlighted in previous investigations using Mendelian randomization[7,8]. Replication of those findings in the present study further emphasizes that genetically-driven depletion of these metabolites is to likely cause MacTel.

Our study also suggests that the neural retina may not be the tissue initially affected by serine depletion. Instead our post GWAS transcriptomics analysis of published retina and RPE data show preferential

transcription of glycine/serine biosynthesis genes in the RPE. This is further supported by a murine metabolomics study comparing RPE and the neurosensory retina, where pentose phosphate metabolism, serine and sphingomyelin biosynthesis activity were higher in the RPE than neurosensory retina[25]. Thus it may be that neuroretinal dysfunction is a secondary consequence of the RPE degradation. Other neuronal "support" cells such as Müller glia, may also contribute to retinal dysfunction in MacTel[26].

The present study indicates the potential involvement of sphingolipid and cholesterol metabolism in MacTel disease. Serine deficiency is implicated in retinal degeneration via the accumulation of deoxysphingolipids, which must be degraded via the omega oxidation pathway instead of sphingosine phosphate lyase[9]. Omega oxidation is mediated via PPAR-alpha, which can be stimulated with Fenofibrate, and protects against deoxysphingolipid-induced toxicity in retinal organoids[9]. Interestingly, our study prioritised *PPARA* via MAGMA analysis, and implicated several related genes including *HMGCS2* (GWAS & MAGMA), *ABCA1*, *ABCG1*, and *SREBF1* (trans-eQTL in blood from locus 2p23.2). Furthermore, the 10q21.3 locus, previously found to significantly affect serine and glycine levels[7] also encompasses *NRBF2*, a co-regulator of PPAR-alpha. Two further loci (19p13.2 and 7pp11.2) overlapped with genes involved in sphingolipid metabolism (*CERS4* and *SUMF2,* respectively*).* These complementary lines of evidence support the centrality of perturbed lipid turnover and concomitant deoxy-sphingolipid accumulation in MacTel. The top SNP at the 9p22.3 locus is **l**ocated in an intron of the Tetratricopeptide repeat protein 39B (*TTC39B)*, and is a significant eQTL for this gene in liver. TTC39B is known to regulate HDL in mice [27] and is associated with human circulating HDL concentrations via GWAS[27]. Genetic correlations between MacTel and use of cholesterol-lowering medication, increased body fat, increased HDL cholesterol levels and decreased VLDL cholesterol, provide further support for the importance of perturbed lipid metabolism in MacTel.

The present study implicates several solute carrier (SLC) family proteins in MacTel, within the cholesterol, amino-acid, bicarbonate, and monocarboxylate transporter sub-families. Enrichment of disease-related SNPs in the intestinally-expressed bile acid exporter *SLC51A* (implicated by MAGMA) may affect serum concentrations of cholesterol which is a major bile component, and further supports the likely perturbation in cholesterol metabolism evidenced above. Both *SLC1A1* (implicated in MacTel via TWAS analysis), and *SLC1A4* (implicated by TWAS and MAGMA) are amino acid transporters with detectable transcription in the retina[16]. SLC1A4 transports several amino acids, including serine, whilst SCL1A1 transports aspartate

and glutamate, the latter of which is required for serine biosynthesis by PSAT1[28]. Defective glutamate transport could further affect both neurotransmission and the acidity of the subretinal space. Post-GWAS analyses also highlighted a lactate transporter gene *SLC16A8* (implicated by MAGMA), which is expressed in RPE and regulates lactate concentrations in the subretinal space[29] and *SLC4A7*, a bicarbonate transporter which is robustly expressed in rods in the neural retina. Altered function of these genes could cause pH dysregulation and heighten metabolic stress in this tissue. Indeed, pH regulation between photoreceptors and support cells (Müller glia and RPE) is critical for retinal health, and *SLC16A8* has previously been implicated in retinal disease, via a GWAS of AMD[30] Lastly, of particular interest is the enrichment of MacTel SNPs in the sarcosine/proline/pipecolate transporter *SLC6A20* (GWAS locus 3p21.31), which is detectable in retina. Sarcosine is converted to glycine in a single enzymatic step involving S-Adenosylmethionine, whose metabolic disruption is also reported in MacTel patients [24].

Finally, we observe that MacTel retinae are likely to be affected by genetic mechanisms other than metabolic dysfunction. The strongest signal in the preliminary GWAS for MacTel was an association peak at 5q14.3 located between the genes *TMEM161B* and *LINC00461* (OR = 2.34, p-value = 1.08 e-15). This association is further strengthened in the current study (p-value = 4.7e-17, OR = 2.31. all ancestries). Despite its intergenic location, the 5q14.3 locus is transcriptionally active in human CNS cell cultures, and was previously identified to affect retinal vascular calibre[4,5]. The top SNP at this locus (rs17421627) is highly conserved and modulates the expression of microRNA miR-9-5 in zebrafish (orthologous to human MIR-9-2) [31]. Genomic ablation of 770 bp covering this SNP reduces the number and branching of blood vessels in the zebrafish retina, which is phenocopied by miR-9-5 ablation[31]. This functional link with retinal vasculature formation, combined with shared genetic effects on several human retinal vascular traits, strongly suggests that locus 5q14.3 underpins the structural vascular defects characteristic of MacTel. Also reported to affect retinal vascular calibre is the gene *MXRA7* (encoding a matrix-related protein)[5]; this gene was prioritised in our study by MAGMA, and produced both the largest number of eQTL signals and the strongest TWAS association in the retina. Association between retinal vasculature defects and MacTel comes as no surprise given that the disease pathology and relationship between vascular calibre and MacTel has been previously identified[1]. Our results also shared two signals with a recent GWAS on retinal thickness, reported in UK Biobank subjects by Gao and colleagues[32]. The first signal overlapped the 5q14.3 locus, and the second was identical to our newly identified signal at the 3p21.31 locus, containing the

SLC6A20 gene. MacTel patients often present with decreased macular thickness resulting from degeneration of retinal layers. However, while the signal in the 3p21.31 locus indeed suggests that the MacTel risk allele may reduce macular thickness in UK Biobank subjects, the risk allele at 5q14.3 is in fact associated with increased thickness, and is concordant with our previous results[8]. Lastly, a recently published meta-analysis[33] identified a SNP in the 5q14.3 locus that also affects AMD risk. This SNP (rs17421410) is in strong LD with the top SNP in that locus ($r^2 = 0.725$) and, concordant with our study, also increases disease risk for AMD (OR = 1.16).

Gantner *et al*[9] demonstrated that serine dietary depletion is sufficient to cause a MacTel-like ocular phenotype in a mouse model. This finding coupled with the present study results suggests that sufficient systemic serine abundance is required to maintain retinal health. While dietary supplementation of serine to MacTel patients may be a reasonable therapeutic approach, recent work suggests that local delivery or frequent systemic dosing will be required to maintain adequate L-serine levels [34]. Given the complexities of the genetic basis of MacTel and the diversity of the patient population, we would caution against serine supplementation for MacTel patients at this time, but note that its potential is currently being investigated. Although MacTel is a relatively rare retinal disorder, a better understanding of its genetic aetiology is also informing us of crucial tissue-specific metabolic and structural characteristics of the retina and RPE in general. As this GWAS study also demonstrates, a better understanding of the biological mechanisms involved in MacTel will also assist with our understanding of other, related neurovascular-glial degenerative diseases of the retina.

Authors contribution

RB Conceptualization, Software, Formal analysis, Data Curation, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization. **VEJ** Conceptualization, Software, Formal analysis, Data Curation,

Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization. **AP** Formal analysis, Data Curation, Investigation. **JEM** Formal analysis, Data Curation, Investigation; **CE** and **TFCH** Resources, Data Curation; **LS** Resources. **MF** Resources Funding acquisition. **RA** Resources. **BREA** Conceptualization, Writing - Original Draft, Writing - Review & Editing, Supervision Project administration. **MB** Conceptualization Writing - Review & Editing, Supervision Project administration, Supervision.

Acknowledgements

The authors would like to thank the Lowy Medical Research Institute for funding to conduct this study.

We also thank Professor Nick Martin, Queensland Institute of Medical Research (QIMR), for access to the unpublished Twins Study data for controls; and Professor Stuart McGregor (QIMR) for a discussion on Mendelian Randomisation studies in cancer risk and MacTel. We acknowledge the use of UKBiobank data (project application 36610) and access to the metabolomic PRS (courtesy of Dr Claudia Langenberg, Medical Research Council, Cambridge, UK). Melanie Bahlo was supported by an Australian National Health and Medical Research Council (NHMRC) Fellowship (GNT 1102971). Brendan Ansell was supported by an NHMRC Early Career Fellowship (GNT 1157776). Roberto Bonelli was supported by a Melbourne International Research Scholarship and the John and Patricia Farrant Foundation. This work was also supported by the Victorian Government's Operational Infrastructure Support Program and the NHMRC Independent Research Institute Infrastructure Support Scheme (IRIISS). Finally, the authors wish to thank the MacTel patients and their families for their involvement in this study.

Online Methods

Participants

1126 Mactel cases were recruited from 23 clinical centres, from seven countries. A complete list of participating institutes and approving ethics committees are provided in the **Supplementary Note**. 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* [2] on the basis of Gass and Blodi[35]. 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.

292 unaffected family members and spouses of cases were also recruited via the MacTel Consortium, as controls. The remainder of controls came from the Aged-Related Eye Diseases Study (AREDS, n=1657), and an Australian Study of Twinning known as the QIMR Twinning Genetics Study (n=1892, unpublished). The QIMR Twinning Genetics study was approved by the Human Research Ethics Committee of the QIMR Berghofer Medical Research Institute. The AREDS study includes individuals aged 55–80 years who "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." These controls were accessed via dbGaP (phs000429.v1.p1) A subset of 476 MacTel Cases, 76 controls from the MacTel Consortium, and 1657 controls from AREDs were used in our previous discovery GWAS[3].

Genotyping

MacTel Consortium samples were genotyped on either the Illumina Human Omni5Exome-4 Array (618 cases and 115 controls) or the Illumina Global Screening Array (508 cases and 177 controls). AREDs controls were genotyped on the Illumina Omni2.5 BeadChip array, and the Australian Twinning controls on the Illumina Global Screening Array.

Full details of samples and genotyping platforms are provided in the supplement (**Table S12**). All genomic loci are reported relative to the Hg19 genome build.

Quality control and imputation of genotyping data

Quality control and imputation were carried out in three tranches, determined by genotyping array: Tranche 1 representing the MacTel consortium samples genotyped on Omni5-exome; Tranche 2 the AREDs controls genotyped on Omni-2.5; and Tranche 3 the MacTel Consortium and Twinning samples genotyped on the Global Screening Array.

Within each tranche, initial quality control of the genotype data was undertaken using Plink (v.1.9.3) to remove SNPs with call rates below 98%, and samples with call rates below 98%, outliers in heterozygosity; samples whose sex could not be confidently genetically inferred, and those whose inferred sex did not match recorded sex. Tranches 1 and 3 contained both cases and controls, and included samples genotyped in different batches. In these two tranches additional exclusions were made of SNPs showing differential levels of missingness in cases *versus* controls, and SNPs showing very extreme differences in allele frequency between genotyping batches (difference in frequency between batches $p<1E-10$). Each tranche was merged with the 1000 Genomes reference dataset, and principal components (PCs) of ancestry were estimated using Plink. We identified a subset of our sample as European (931 cases and 3745 controls), whose first two PCs of ancestry were within 4 standard deviations of the mean value for the 1000 Genomes european super-population. No exclusions based on ancestry were made.

The three tranches were separately imputed to the Haplotype Reference Consortium reference panel (v.1.1), using the Michigan Imputation Server v1.0.4 pipeline[36], with phasing using Eagle[37]. The three tranches were merged following imputation.

We examined relatedness and ancestry of our samples in the merged dataset. The data were restricted to variants directly genotyped across all three tranches (230,201 SNPs), and then LD-Pruned to leave 100,804 SNPs. Using this subset of SNPs, we generated an initial genetic relatedness matrix (GRM) using the KING

method[38], implemented in the SNPRelate R package[39]. Using this GRM, we then calculated PCs of ancestry, taking into account relatedness, using the PC-AiR method[40]. We then re-estimated the GRM, taking into account PCs of ancestry, using the PC-Relate method[41], in the GENESIS R package[42]. Eleven individuals were identified as duplicates, and one sample from each duplicate pair was removed, retaining the sample with the highest call rate.

Finally, we identified and removed SNPs which may have been subject to batch effects. After restricting to European samples, we tested for differences in allele frequency, for the following comparisons: tranche 1 cases *versus* tranche 3 cases; tranche 1 controls *versus* all other controls; tranche 2 controls *versus* all other controls; tranche 3 controls *versus* all other controls. SNPs which showed association (P<5E-5) in any tranche comparison were excluded. SNPs with imputation quality <0.9 in any imputation run were also excluded, along with any SNPs deviating from Hardy Weinberg Equilibrium (P<5E-7).

Genome-wide association analyses

Association analyses were undertaken using SAIGE v.0.36.2, which allows incorporation of related individuals, which was not performed in the first MacTel GWAS[3].

Firstly, a null logistic mixed model was fitted, using a subset of 230,085 SNPs, which were directly genotyped across all samples, with sex and 8 PCs of ancestry as covariates. Eight PCs were chosen, justified from the following criteria: 1) Inspection of a scree plot (**Figure S1A**), and through fitting the null model with a varying numbers of PCs, and selecting the best model, based on the Akaike information criterion (**Figure S1B**). Using the parameters estimated under the null model, association testing was undertaken genome-wide on the imputed data. Two genome-wide analyses were carried out: firstly using a subset of European ancestry individuals only; secondly, utilising all samples, regardless of ancestry ('full-cohort' analysis). Results of both analyses were filtered to include only SNPs with MAF>0.25%. Unless otherwise stated, SNPs with uncorrected p-values of association below 5E-8 were considered 'genome-wide significant' (GW-significant), and those with p<5E-6 were considered 'suggestive significant'. Loci with suggestive-significant association with disease but supported by fewer than two SNPs, were discarded.

To verify that our results were not subject to confounding by population stratification, we estimated the genomic inflation factor (λ), to assess overall inflation of the association test statistics. We additionally utilised LD score regression[43], as implemented by LDSC software version 1.1, and using LD-scores estimated from the data. After restricting the results to SNPs with MAF ≥ 1%, we examined the resulting LD Score regression intercept, as a measure of inflation due to confounding, rather than polygenicity of the trait.

SNP-based heritability was estimated using GREML[44,45], as implemented by GCTA[46]. Firstly a GRM was estimated using all SNPs, and a subset of unrelated individuals, (defined as pairwise kinship coefficient < 0.018; 986 cases and 3227 controls). Narrow-sense heritability was estimated using GCTA-GREML, using this GRM and assuming a population prevalence for Mactel of 0.45%.

Identification of secondary signals and fine-mapping of causal variants

Conditional analysis was carried out to identify secondary, independent signals. For each locus with a genome-wide significant association, we performed conditional analyses over a 2Mb region (+/- 1Mb of top SNP), conditional on the top SNP. We identified secondary signals as SNPs with P<5E-8, after conditioning on the original top SNP. To confirm independence where a secondary signal was found, we re-tested for association of the original top SNP, conditional on the secondary top SNP.

For each identified locus, we carried out Bayesian fine mapping, to determine 95% Credible Sets of causal SNPs[10]. This method assumes a single causal SNP at each locus; for loci where a secondary, independent signal was identified, we undertook fine-mapping separately for each signal, using the conditional association results.

Functional mapping and annotation with FUMA

For all genome-wide significant (p< 5E-8), and suggestive significant loci (p< 5E-6), gene mapping was performed using the web-tool FUMA v1.3.5e[13]. The European 1000 Genome Project reference panel was used to define LD between SNPs. A full list of all parameters used for the FUMA analysis is available in **Table S3**. Firstly, we used positional mapping to identify genes containing SNPs in LD (r_2>0.4) with the

most significant SNP et each locus. For intergenic loci, where neither the top SNP, nor any proxy (r^2>0.4) overlapped with any genes, we manually annotated these loci with the closest genes on both sides. Secondly, gene-based analyses were carried out using MAGMA[11]. We corrected for false discovery rate the MAGMA p-values using an ad-hoc Benjamini-Hochberg correction from the function *p.adjust* in R v. 3.6.1. Genes with a corrected p-value less than 0.05 were considered significantly enriched. Thirdly, the most significant SNPs at each locus as well as SNPs in LD (r^2>0.4) with them were mapped to significant eQTLs as defined by the internal FUMA pipeline. In order to maintain results as relevant as possible to MacTel disease only specific tissues were selected for eQTL queries. Such tissues were selected to represent those ones likely affected by the systemic dysregulations previously observed among MacTel patients. Datasets included in the eQTL analysis, spans across brain, blood, liver, nerve and vasculature, specific tissues sub-types can be found among the parameters used in FUMA.

Retinal eQTL analysis

As retinal eQTL results were not available through FUMA, we intersected the GWAS summary results with cis-eQTL data for retinal tissue (retina eQTL)[16], available as an external dataset on the[47] GTEx portal. MacTel effect alleles were first reordered to ensure all disease effects were positive (i.e., increasing risk). Next, the effect alleles and effect estimates were inverted in the eQTL results to be consistent with the MacTel allele order, as required. At each GWAS locus (GW-significant, or suggestive-significant), We identified genes where the top SNP at each GWAS locus, or any SNP in LD (r^2>0.4), was a significant eQTL, based on gene-specific significance thresholds, as defined by[16].

For the implicated genes at each the GW-sig locus, we additionally employed Approximate Bayes Factor colocalization analysis using the R package *coloc*[48]. Using the estimated effect sizes and corresponding variances from the GWAS and the eQTL analyses as input, we estimated the posterior probability that the GWAS signal, and the eQTL signal were the result of a shared causal variant. A posterior probability > 0.75 was taken as strong support of colocalistion of the GWAS and eQTL signal[48].

Transcriptome-wide association analysis

Transcriptome-wide association analysis (TWAS) on gene expression was inferred using the TWAS / FUSION software[12]. Given the previously reported systemic metabolic defects in MacTel patients[9] we downloaded SNPs and relative TWAS weights on gene expression for several tissues from the TWAS / FUSION website (gusevlab.org/projects/fusion/). These tissues included liver, tibial nerve, tibial artery and whole blood from GTEx[14,15,49] as well as retinal weights from Ratnapriya *et al*[16]. For this analysis brain TWAS weights were selected from Gandal *et al*[50] rather than GTEx as these were estimated using a considerably larger sample size. Only genes whose expression was predicted using at least three SNPs were considered in this analysis. Due to the large number of tests, we allowed for a false discovery rate of 5%. Genes with ad-hoc corrected p-values (using the p.adjust function in R) less than 0.05 were thus considered significant. We observed a large number of genes in the retina TWAS analysis to be skipped by the TWAS / FUSION software due to their best predictive model only containing SNPs with zero weight. To this end, all skipped genes in the retina were re-estimated using the best linear unbiased predictor (BLUP) model as defined by the TWAS/FUSION software.

Enrichment of prioritised genes for eye tissues and retinal cell types

We prioritised a set of genes from the GWAS and post-GWAS analysis results supported by at least two of the following criteria: (1) all genes located underneath loci with SNPs reaching genome-wide significant threshold (2) all genes located underneath loci with SNPs reaching suggestive significant threshold, (3) genes whose expression was affected by the identified SNPs in at least one of the analyzed tissues, (4) genes identified by the MAGMA analysis, and (4) genes identified by the TWAS/FUSION software.

We then accessed and analysed previously published human adult retina datasets from: (1) bulk RNA analysis at individual level for four adult donors of temporal, macular and nasal regions of both the neural retina, and the RPE/choroid[17] (2) average gene expression across from four retinal single-cell datasets from 12 samples[18] and 6 samples[19], as well as (3) average retina single-cell expression data from the fovea and peripheral areas of the retina on 5 samples[20]. Genes not expressed in any areas or cell types were excluded. Gene expression in each tissue area or cell type was log transformed to obtain an approximately

normal distribution. We used linear regression modelling to compare expression between cell types, and retinal areas of interest. P-values for area-specific and cell-type-specific tests were independently corrected for false discovery rate as described previously.

Genetic correlation analysis

The genetic correlation between MacTel and 514 UK Biobank traits, 107 metabolites[51], eight glycaemic traits[52], four circulating lipid concentrations [53], and three neurological disorders[54], was assessed using LD-score regression software[55] implemented by the the LDhub database[56] in February 2020. In addition we included as a separate analysis, 13 ocular phenotypes[57] using LDSC standalone software[43]. We have estimated genetic correlations given the strong phenotypic correlations observed between MacTel and many of the examined traits; however we note that genetic correlation analyses performed using small sample size GWAS summary statistics has been previously discouraged[58]. We report all nominally significant correlations ($p<0.05$).

Mendelian Randomization

Mendelian randomization was performed using polygenic risk scores (PRS) obtained through the weighted score method on individual genetic data as described by[59]. Instrument SNPs and relative weights were derived from different studies according to the phenotype of interest. Metabolic instruments were obtained from Lotta *et al*[7]. Type 2 diabetes instruments were obtained from Xue *et al*[60]. Retinal vasculature calibre instruments were obtained from Ikram *et al*[4]; retinal arteriolar calibre instruments from Sim *et al*[5]; and macular thickness instruments from[32]. Causal effects of these phenotypes on MacTel risk was assessed using logistic regression performed on unrelated individuals were the phenotypic PRS scores was regressed against disease status while correcting for sex at birth the the first PC. Metabolic PRS effects on MacTel risk were corrected for multiple testing using the same strategy as described above. Potential pleiotropic effects of SNPs were tested by MR-Egger test[61]. Independent significance of metabolites as well as other phenotypes was assessed by sequential inclusion of less significant MR scores conditioning by more significant MR scores.

MacTel polygenic risk score calculation

Three polygenic risk scores (PRS) were assessed for prediction power. A top SNPs PRS was constructed by using the most significant SNP in each locus with a p-value less than 5E-8. SNPs that remained genome-wide (GW) significant after conditioning for the top SNP, were also included. A less stringent PRS was also constructed, using all top SNPs at loci reaching suggestive significance (5E-6). We compared the PRSs prediction power through a logistic regression model on all unrelated individuals using the Area Under the ROC Curve (AUC) which assess discrimination ability across different disease probability thresholds. We additionally tested for robustness of these prediction models using a leave-one-out approach applied to the entire cohort. Disease odds ratios and their 95% confidence intervals, were determined for each PRS decile comparing each decile to the median. Prediction accuracy differences between European and non-Europeans were also assessed. First, the model was trained on a dataset excluding all non-European subjects as well as a randomly selected group of Europeans of the same sample size as the non-Europeans. Second, AUC of the prediction was compared for both the remaining Europeans, and non-Europeans.

UK Biobank analysis

Prediction of MacTel PRS was performed on 487,311 UK Biobank (UKB) subjects. UKB contains information about hospitalisations and relative primary and secondary ICD codes for admission. However, MacTel disease is unfortunately not uniquely classified by a single ICD10 code and MacTel patients are rarely hospitalised because of their condition. Because of this, we tested instead association between MacTel PRS and self-reported eye conditions using linear regression modelling. This analysis was restricted to 202,213 subjects answering the question “Has a doctor told you that you have any of the following problems with your eyes?”. Subjects reporting “Prefer not to answer” and “Do not know” were excluded from the analysis.

Additionally, UKB retinal imaging grading for MacTel phenotypes was performed on optical coherence tomography by CE and TFCH using a nominal scale. Both CE and TFCH have seen several hundreds of

cases of MacTel both face to face clinically and as part of their role in the Moorfields Eye Hospital Reading Centre as adjudicator of presence of disease. These subjects were randomly selected by the statistical genetics team to equally represent the top and the bottom deciles of the PRS distribution. Grading was performed on 200 subjects reporting no eye conditions as well xxx reporting “other serious eye conditions”. The last were selected since they presented on average higher MaCtel PRS scores compared to all other reported retinal disorders. Ophthalmologists were blinded to the PRS decile for each subject. Grading was performed on the entire OCT volume scan. Each frame of the original Topcon files were converted to single .png files. The .png stacks were converted to gifs using Fiji[62]. The grading was done on the gifs, to facilitate viewing. Both eyes were graded as follows: 0 = normal, 1 = macular changes unrelated to MacTel, 2 = some macular changes resembling MacTel, 3 = eye affected by MacTel disease, NA = ungradable. Logistic regression modelling between MacTel PRS decile and maximum grading value per subject from all imaging technologies and both eyes was performed to assess correlation between MacTel PRS and retinal phenotype grading. Significance of term was assessed using likelihood-ratio tests.

Data availability

Control genotyping data was sourced from dbGaP for the National Eye Institute (NEI) Age-Related Eye Disease Study (AREDS) (dbGaP accession number phs000001.v3.p1). Further control genotyping data was sourced from the QIMR Genetics of Twins study (Professor Nick Martin, available on request). MacTel patient genotyping data and a subset of the controls are available from the European Genome and Phenome archive (EGAS00001002249). Complete & European-only cohort GWAS summary statistics are available at https://doi.org/10.17605/OSF.IO/GEK7B.

Tables

Table 1. Genome wide significant associations with MacTel.

Rsid	Chr	BP	effect/non-effect allele	EAF Cases	EAF Controls	Cytoband	Odds Ratio	95% C.I.	P-value	P-value Eur	Proximal coding genes*	Scerri et al., 2017 GWAS	Bonelli et al., conditional GWAS
rs146953046	1	120278072	G/T	0.059	0.016	p12	5.47	3.87 - 7.74	7.90E-22	1.30E-20	***PHGDH; HMGCS2***	-	-
rs662602	1	120265444	G/A	0.576	0.685	p12	0.585	0.52 - 0.65	9.29E-18*	1.75E-16*	***PHGDH; HMGCS2***	GW	-
rs2160387	2	65220910	C/T	0.351	0.441	p14	0.72	0.65 - 0.81	5.90E-09	3.70E-09	***SLC1A4***	-	-
rs1047891	2	211540507	A/C	0.192	0.306	q34	0.56	0.5 - 0.63	8.60E-21	3.60E-20	***CPS1***	GW	-
rs9820465	3	27706298	C/T	0.145	0.205	p24.1	0.66	0.57 - 0.76	5.60E-09	1.50E-07	*SLC4A7; EOMES*	S	GW
rs17279437	3	45814094	A/G	0.144	0.104	p21.31	1.73	1.45 - 2.06	6.20E-10	9.70E-10	***SLC6A20; SACM1L; LIMD1; LARS2***	-	-
rs17421627	5	87847586	G/T	0.132	0.069	q14.3	2.31	1.9 - 2.81	4.70E-17	1.00E-16	*TMEM161B*	GW	GW
rs6955423	7	56099352	A/G	0.651	0.746	p11.2	0.7	0.62 - 0.79	5.90E-09	2.60E-07	***SEPT14;MRPS17;ZNF713;GBAS;PSPH;CCT6A;SUMF2;PHKG1;CHCHD2;NUPR1L***	S	-
rs677622	9	15302613	G/A	0.821	0.869	p22.3	0.63	0.54 - 0.73	3.20E-09	1.10E-09	***TTC39B***	S	-
rs10995566	10	65363166	T/C	0.231	0.316	q21.3	0.72	0.64 - 0.81	4.80E-08	8.80E-08	***NRBF2;JMJD1C; REEP3***	S	-
rs139412173	19	8235251	G/A	0.024	0.056	p13.2	0.47	0.36 - 0.61	2.90E-08	3.40E-07	*CERS4, FBN3*	S Nearby	GW

Table 1. Top tagging GW significant SNPs for each locus and relevant candidate genes from GWAS analysis. Bold text indicates the gene is covered by the haplotype (as defined by FUMA, based on linkage disequilibrium using the 1000 Genomes Project European cohort). Non-bold text indicates genes proximal to the locus. The Bonelli *et al* GWAS [8] was conditioned on genetically-predicted T2D risk, and serum glycine and serine. *P-values are conditional on SNP rs146953046 (Non-conditional P-values in all ancestry / euro analyses: 2.38E-20 / 5.49E-19). Chr: chromosome number; BP: base pairs; EAF: effect allele frequency; Eur: European-only sample subset. GW: genome-wide significant (p < 5E-8); S: suggestive-significant (p < 5E-6).

Table 2. Novel suggestive-significant MacTel loci

Rsid	Chr	BP	effect/non-effect allele	EAF Cases	EAF Controls	Cytoband	Odds Ratio	95% C.I.	P-value	P-value Eur
rs2120770	1	212509005	C/A	0.179	0.218	q32.3	0.717	0.63 - 0.82	1.55E-06	9.49E-06
rs1260326	2	27730940	C/T	0.511	0.589	p23.2	0.742	0.66 - 0.83	1.32E-07	5.67E-07
rs1973480	6	74511980	G/A	0.342	0.295	q13	1.33	1.18 - 1.5	3.35E-06	3.68E-06
rs2954021	8	126482077	G/A	0.442	0.519	q24.21	0.75	0.67 - 0.84	2.65E-07	2.28E-07
rs2984814	11	31554964	T/G	0.259	0.325	p13	0.755	0.67 - 0.85	2.40E-06	1.87E-06
rs11077850	17	74661436	C/T	0.255	0.18	q25.1	1.38	1.21 - 1.58	1.12E-06	1.58E-06

Table 2. Novel suggestive-significant MacTel loci. The table presents the most significant SNPs reaching suggestive significance (p<5E-6) at each locus. All listed genes are covered by the haplotype (as defined by FUMA, based on linkage disequilibrium using the 1000 Genomes Project Europeans). Chr: chromosome number; BP: base pairs; EAF: effect allele frequency; Eur: European-only sample subset.

Table 3. Top ten genes enriched with SNPs signals (MAGMA)

Symbol	Chr	Start	Stop	ZStat	P-value	FDR
PHGDH	1	120202421	120286838	9.07	5.76E-20	1.06E-15
HMGCS2	1	120290619	120311528	6.17	3.39E-10	3.13E-06
REEP3	10	65281123	65384883	5.97	1.20E-09	7.42E-06
SLC6A20	3	45796942	45838027	5.74	4.82E-09	2.23E-05
TTC39B	9	15163620	15307358	5.28	6.58E-08	2.43E-04
GCKR	2	27719709	27746554	4.96	3.56E-07	1.10E-03
DCDC1	11	30851916	31391357	4.79	8.54E-07	2.25E-03
*ZAP70**	2	98330023	98356325	4.76	9.73E-07	2.25E-03
MXRA7	17	74668633	74707098	4.73	1.10E-06	2.26E-03
PSPH	7	56078744	56119297	4.71	1.23E-06	2.28E-03

Table 3. Top ten genes enriched with SNPs signals (MAGMA) Gene names in bold italic text lie within a MacTel-associated locus. Those in italics lie outside MacTel loci. *lies outside the MacTel-associated loci.

Table 4. TWAS results in tissues relevant for MacTel

Symbol	Chr	start	end	Z	P	FDR	Tissue*
TTC39B	9	15163622	15307360	5.27	1.39E-07	3.92E-04	Liver
MXRA7	17	76672551	76711016	5.08	3.87E-07	3.98E-03	Retina
NRBF2	10	63133247	63155031	-5.01	5.50E-07	3.98E-03	Retina
DNAJC24	11	31369840	31431849	4.54	5.56E-06	2.97E-02	Brain

SLC1A1	9	4490444	4587469	4.27	1.94E-05	3.94E-02	Brain
IFT172	2	27444371	27489789	-4.24	2.21E-05	3.94E-02	Brain
PHGDH	1	119648411	119744226	4.45	8.60E-06	4.15E-02	Retina

Table 4. TWAS results in tissues relevant for MacTel. Z: normalised effect of gene expression on disease risk (positive increases disease risk). FDR: Benjamini-Hochberg corrected p-values. *TWAS weights for tissues accessed from GTEx (Liver) [49], Rantapriya *et al* (Retina) [16], and Gandal *et al* (Brain) [50].

References

1. Issa, P. C. *et al.* Macular telangiectasia type 2. *Progress in Retinal and Eye Research* vol. 34 49–77 (2013).
2. 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).
3. Scerri, T. S. *et al.* Genome-wide analyses identify common variants associated with macular telangiectasia type 2. *Nature Publishing Group* **49**, 559–567 (2017).
4. Ikram, M. K. *et al.* Four Novel Loci (19q13, 6q24, 12q24, and 5q14) Influence the Microcirculation In Vivo. *PLoS Genet.* **6**, e1001184 (2010).
5. Sim, X. *et al.* Genetic Loci for Retinal Arteriolar Microcirculation. *PLoS ONE* vol. 8 e65804 (2013).
6. Morris, A. P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).
7. Lotta, L. A. *et al.* Cross-platform genetic discovery of small molecule products of metabolism and application to clinical outcomes. *bioRxiv* 2020.02.03.932541 (2020) doi:10.1101/2020.02.03.932541.
8. Bonelli, R. *et al.* Genetic Disruption of Serine Biosynthesis is a Key Driver of Macular Telangiectasia Type 2 Etiology and Progression. *bioRxiv* 2020.02.04.934356 (2020) doi:10.1101/2020.02.04.934356.
9. Gantner, M. L. *et al.* Serine and Lipid Metabolism in Macular Disease and Peripheral Neuropathy. *N. Engl. J. Med.* **381**, 1422–1433 (2019).
10. Hutchinson, A., Watson, H. & Wallace, C. Improving the coverage of credible sets in Bayesian genetic fine-mapping. *PLoS Comput. Biol.* **16**, e1007829 (2020).
11. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
12. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).
13. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* vol. 8 (2017).
14. Võsa, U. *et al.* Unraveling the polygenic architecture of complex traits using blood eQTL

metaanalysis. *bioRxiv* 447367 (2018) doi:10.1101/447367.

15. Zhernakova, D. V. *et al.* Identification of context-dependent expression quantitative trait loci in whole blood. *Nat. Genet.* **49**, 139–145 (2017).
16. Ratnapriya, R. *et al.* Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration. *Nat. Genet.* **51**, 606–610 (2019).
17. Whitmore, S. S. *et al.* Transcriptomic analysis across nasal, temporal, and macular regions of human neural retina and RPE/choroid by RNA-Seq. *Exp. Eye Res.* **129**, 93–106 (2014).
18. Lukowski, S. W. *et al.* A single-cell transcriptome atlas of the adult human retina. *The EMBO Journal* vol. 38 (2019).
19. Menon, M. *et al.* Single-cell transcriptomic atlas of the human retina identifies cell types associated with age-related macular degeneration. *Nat. Commun.* **10**, 4902 (2019).
20. Voigt, A. P. *et al.* Single-Cell RNA Sequencing in Human Retinal Degeneration Reveals Distinct Glial Cell Populations. *Cells* **9**, (2020).
21. Chew, E. Y. *et al.* Effect of Ciliary Neurotrophic Factor on Retinal Neurodegeneration in Patients with Macular Telangiectasia Type 2: A Randomized Clinical Trial. *Ophthalmology* **126**, 540–549 (2019).
22. Ahuja, V. & Powers-Lee, S. G. Human carbamoyl-phosphate synthetase: Insight into N-acetylglutamate interaction and the functional effects of a common single nucleotide polymorphism. *J. Inherit. Metab. Dis.* **31**, 481–491 (2008).
23. de Cima, S. *et al.* Structure of human carbamoyl phosphate synthetase: deciphering the on/off switch of human ureagenesis. *Scientific Reports* vol. 5 (2015).
24. Bonelli, R. *et al.* Systemic lipid dysregulation is a risk factor for macular neurodegenerative disease. *Sci. Rep.* **10**, 1–13 (2020).
25. Sinha, T., Naash, M. I. & Al-Ubaidi, M. R. The Symbiotic Relationship between the Neural Retina and Retinal Pigment Epithelium Is Supported by Utilizing Differential Metabolic Pathways. *iScience* **23**, 101004 (2020).
26. Powner, M. B. *et al.* Loss of Müller's cells and photoreceptors in macular telangiectasia type 2. *Ophthalmology* **120**, 2344–2352 (2013).
27. Hsieh, J. *et al.* TTC39B deficiency stabilizes LXR reducing both atherosclerosis and steatohepatitis. *Nature* **535**, 303–307 (2016).
28. Amelio, I., Cutruzzolá, F., Antonov, A., Agostini, M. & Melino, G. Serine and glycine metabolism in

cancer. *Trends Biochem. Sci.* **39**, 191–198 (2014).

29. Daniele, L. L., Sauer, B., Gallagher, S. M., Pugh, E. N., Jr & Philp, N. J. Altered visual function in monocarboxylate transporter 3 (Slc16a8) knockout mice. *Am. J. Physiol. Cell Physiol.* **295**, C451–7 (2008).
30. 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**, 134–143 (2016).
31. Madelaine, R. *et al.* A screen for deeply conserved non-coding GWAS SNPs uncovers a MIR-9-2 functional mutation associated to retinal vasculature defects in human. *Nucleic Acids Res.* **46**, 3517–3531 (2018).
32. Gao, X. R., Raymond Gao, X., Huang, H. & Kim, H. Genome-wide association analyses identify 139 loci associated with macular thickness in the UK Biobank cohort. *Human Molecular Genetics* vol. 28 1162–1172 (2019).
33. Han, X. *et al.* Genome-wide meta-analysis identifies novel loci associated with age-related macular degeneration. *J. Hum. Genet.* (2020) doi:10.1038/s10038-020-0750-x.
34. Miller, D. E., Ferreira, C. R., Scott, A. I. & Chang, I. J. Pharmacokinetics of oral l-serine supplementation in a single patient. *Molecular Genetics and Metabolism Reports* **24**, 100607 (2020).
35. Gass, J. D. & Blodi, B. A. Idiopathic juxtafoveolar retinal telangiectasis. Update of classification and follow-up study. *Ophthalmology* **100**, 1536–1546 (1993).
36. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nature Genetics* vol. 48 1284–1287 (2016).
37. Loh, P.-R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).
38. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
39. Zheng, X. *et al.* A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).
40. Conomos, M. P., Miller, M. B. & Thornton, T. A. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. *Genet. Epidemiol.* **39**, 276–293 (2015).
41. Conomos, M. P., Reiner, A. P., Weir, B. S. & Thornton, T. A. Model-free Estimation of Recent

Genetic Relatedness. *Am. J. Hum. Genet.* **98**, 127–148 (2016).

42. Gogarten, S. M. *et al.* Genetic association testing using the GENESIS R/Bioconductor package. *Bioinformatics* (2019) doi:10.1093/bioinformatics/btz567.
43. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* vol. 47 291–295 (2015).
44. Visscher, P. M., Yang, J. & Goddard, M. E. A commentary on 'common SNPs explain a large proportion of the heritability for human height' by Yang et al. (2010). *Twin research and human genetics: the official journal of the International Society for Twin Studies* vol. 13 517–524 (2010).
45. 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).
46. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
47. GTEx portal. *GTEx API* https://storage.googleapis.com/gtex_external_datasets/eyegex_data/single_tissue_eqtl_data/Retina.nominal.eQTLs.tar).
48. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
49. GTEx Consortium *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
50. Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, (2018).
51. Kettunen, J. *et al.* Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat. Commun.* **7**, 11122 (2016).
52. Saxena, R. *et al.* Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat. Genet.* **42**, 142–148 (2010).
53. Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
54. Simón-Sánchez, J. *et al.* Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat. Genet.* **41**, 1308–1312 (2009).
55. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-

relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).

56. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* vol. 33 272–279 (2017).
57. Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in UK Biobank. *Nature Genetics* vol. 50 1593–1599 (2018).
58. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nature Genetics* vol. 47 1236–1241 (2015).
59. Burgess, S., Dudbridge, F. & Thompson, S. G. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. *Stat. Med.* **35**, 1880–1906 (2016).
60. Xue, A. *et al.* Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nature Communications* vol. 9 (2018).
61. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
62. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

Figures

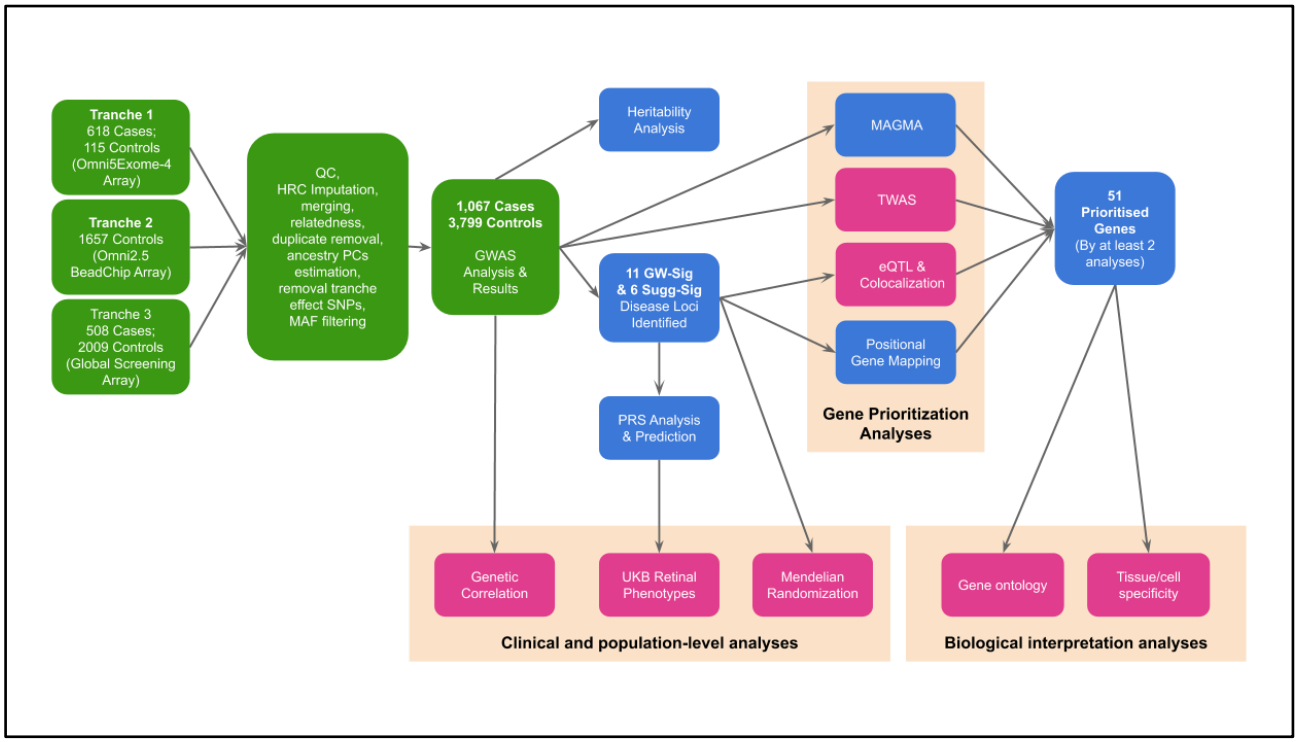

Figure 1. Flowchart of the analyses performed in this study. Each box represents either a set of data or an analysis performed in this study. Green boxes highlight GWAS analyses, while the rest are post-GWAS investigations. The analyses presented in pink boxes employ both data from the current GWAS as well as publicly available data. Orange clusters represent analyses performed to achieve an overall task. GW-Sig: Genome-wide significant, Sugg-Sig: Suggestive significant, TWAS: Transcriptome-Wide Association Analysis. eQTL: expression Quantitative Trait Loci, PRS: Polygenic Risk Score, UKB: UK Biobank.

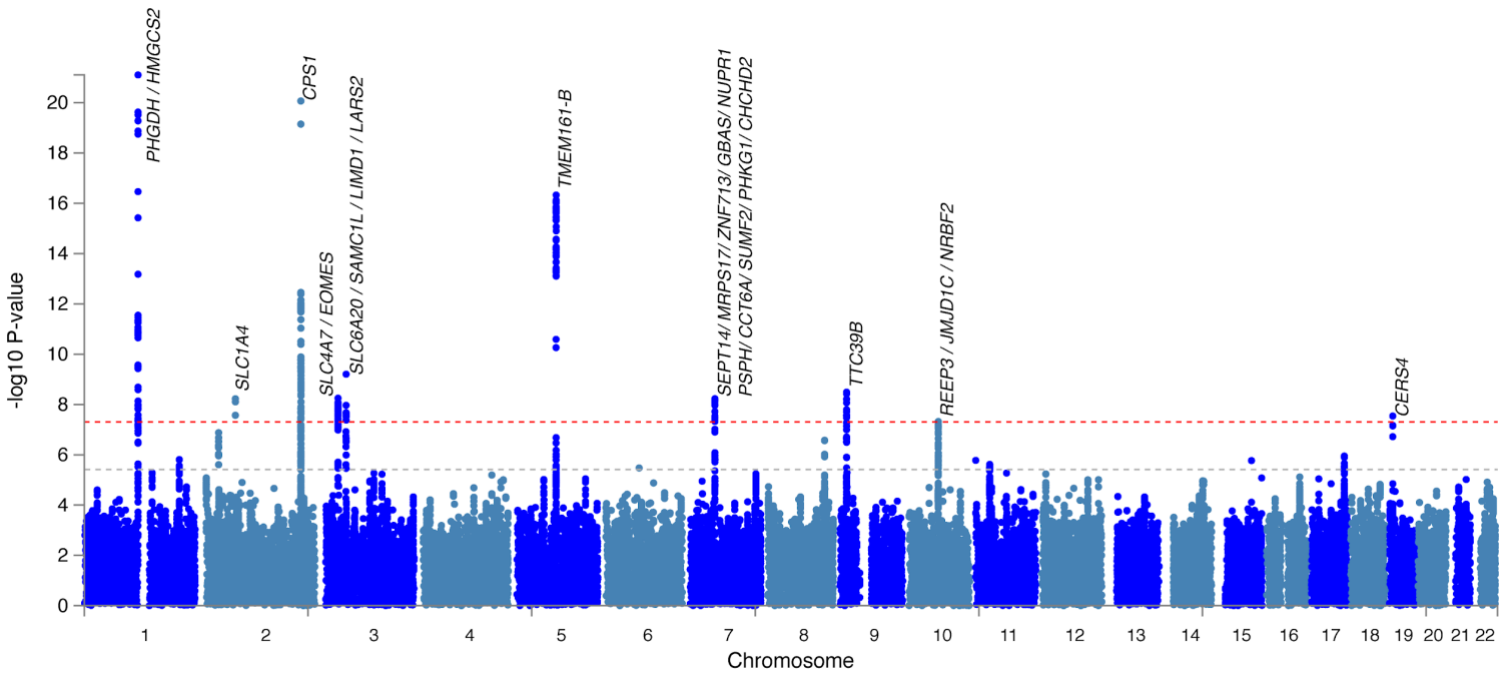

Figure 2. Manhattan plot displaying loci associated with Macular Telangiectasia Type II. Each point denotes a single nucleotide polymorphism (SNP) located on a particular chromosome (x axis). Significance level is presented in the y-axis. The dotted grey line indicates the threshold for genome-wide significance 5E-8 while the dotted orange line indicates the threshold for suggestive genome-wide significance 5E-6. Genes located proximal to or within genome-wide significant loci are displayed above each locus.

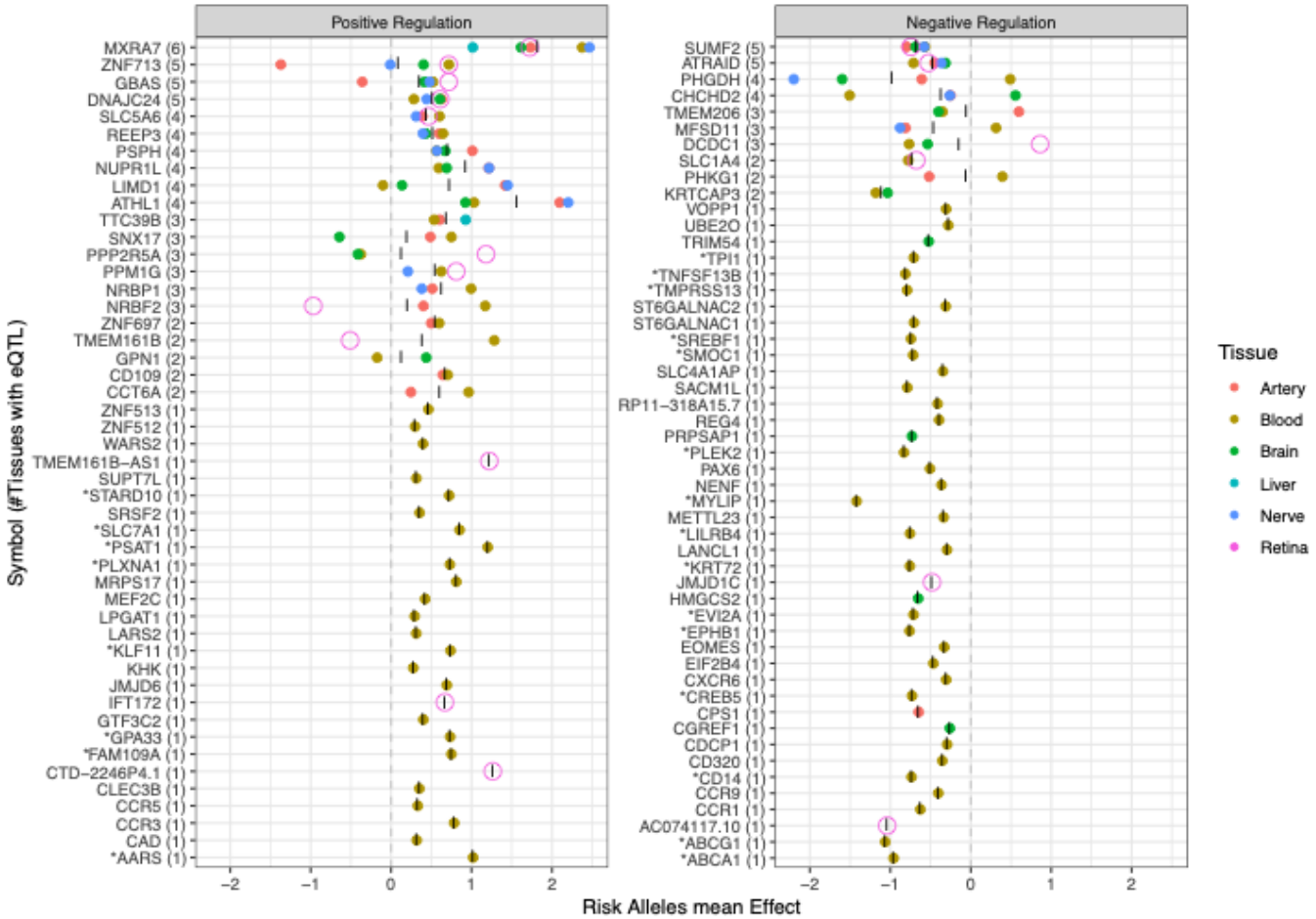

Figure 3. Expression quantitative trait locus transcriptional effects for MacTel risk alleles (SNPs) across six tissue types. The average effect of all SNPs as well as tissue subtypes on gene expression in each tissue is represented as coloured points. As the primary disease site, retina is represented as a larger dot for ease of identification. Mean eQTL effect by tissue type is represented on x-axis while gene symbols are represented on the y-axis. Numbers in brackets denote the number of tissues in which significant eQTL were detected. Genes are ordered by the sum of tissues in which altered expression is inferred in MacTel. Genes are divided in either positive or negative global regulation defined as the average "|" of eQTL effect across the six tissue groups. *Effect on these genes was detected via trans-eQTL analysis in eQTLGen database.

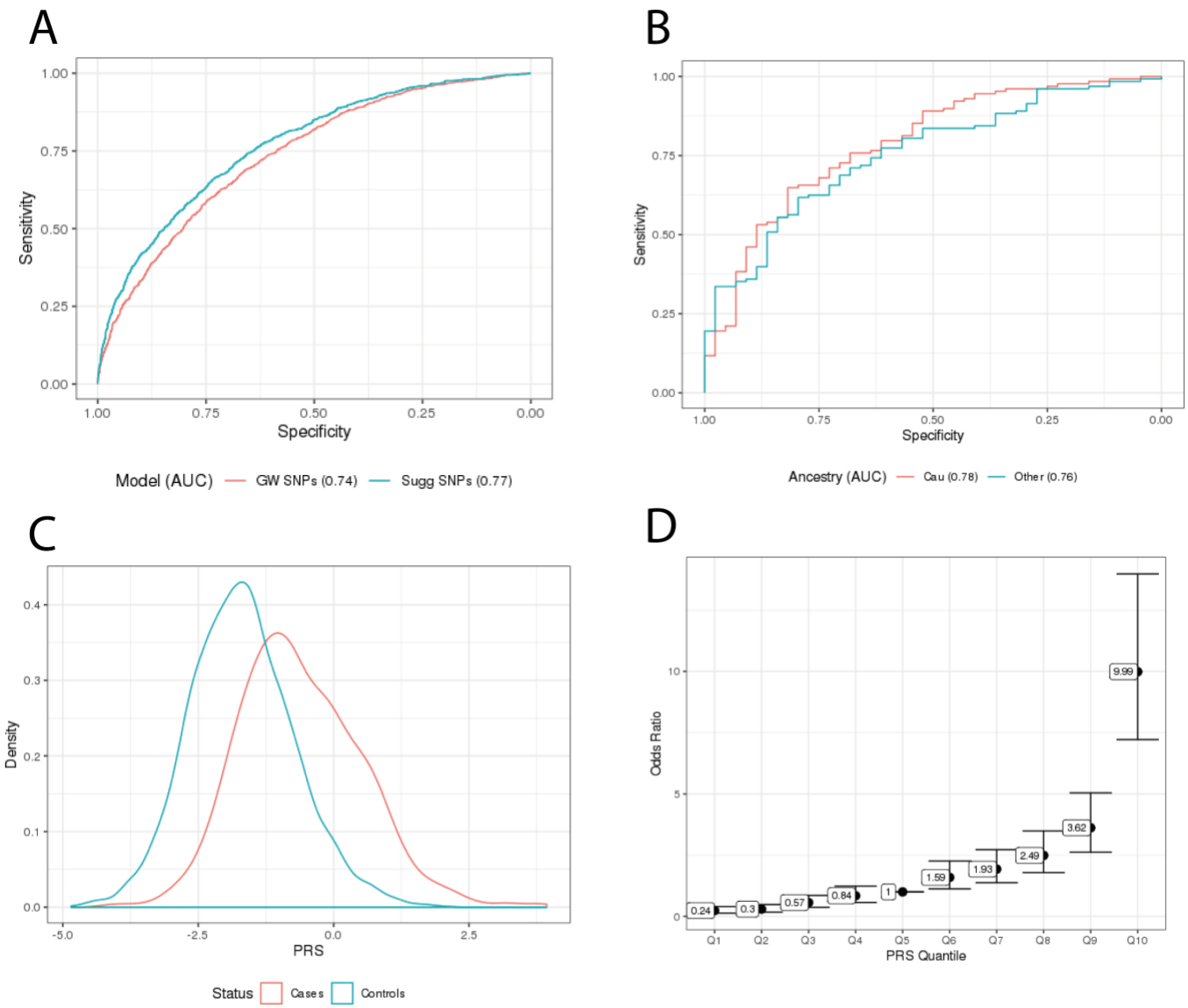

Figure 4. Prediction of MacTel status using polygenic risk score. A) ROC curves of the two tested PRSs. Area under the ROC curve is reported in the figure legend. B) Comparison of ROC curves for prediction accuracy among European and non-European. Area under the ROC curve is reported in the figure legend. C) Density plot comparing PRS distribution of between cases and controls. D) Odds ratio for magnitude MacTel Risk comparing deciles to the median decile.